Optimal HAR Inference*

Liyu Dou

Princeton University

ldou@princeton.edu

Job Market Paper

This version: January 2, 2019

[Click here for latest version]

Abstract

This paper considers the problem of deriving heteroskedasticity and autocorrelation robust (HAR) inference about a scalar parameter of interest. I derive finite-sample optimal tests in the Gaussian location model, under nonparametric assumptions on the underlying spectral density. The optimal test trades off bias and variability, and requires an adjustment of the critical value to account for the maximum bias of the implied long-run variance estimator. I find that with an appropriate adjustment to the critical value, it is nearly optimal to use the so-called equal-weighted cosine (EWC) test, where the long-run variance is estimated by projections onto q type II cosines. The practical implications are an explicit link between the choice of q and assumptions on the underlying spectrum, as well as a corresponding adjustment to the usual Student-t critical value. Simulations show that the suggested new EWC test also performs well outside the Gaussian location model.

Keywords: Heteroskedasticity and autocorrelation robust inference; Long-run variance

JEL Codes: C12; C13; C18; C22; C51

*I am deeply indebted to Ulrich Müller for posing the question and for his continuous help, support, and encouragement. I would like to thank Xu Cheng, Paul Ho, Bo Honoré, Michal Kolesár, Mikkel Plagborg-Møller, Mikkel Sølvsten, James Stock, Mark Watson, and numerous participants in seminars for helpful comments and suggestions.

1 Introduction

This paper considers the problem of deriving appropriate corrections to standard errors when conducting inference with autocorrelated data. The resulting heteroskedasticity and autocorrelation robust (HAR) inference has applications in OLS and GMM settings.[1] Computing HAR standard errors involves estimating the "long-run variance " (LRV) in econometric jargon. Classical references on HAR inference in econometrics include Newey and West (1987) and Andrews (1991), among many others. The Newey-West/Andrews approach is to use t- and F-tests based on consistent LRV estimators and to employ the critical values derived from the normal and chi-squared distributions. The resulting HAR standard errors are asymptotically justified in a large variety of circumstances.

Small sample simulations,[2] however, show that the Newey-West/Andrews approach can lead to tests that incorrectly reject the null far too often. A large subsequent literature (surveyed in Müller (2014)) has proposed many alternative procedures. These procedures employ inconsistent LRV estimators and demonstrate better performance for controlling the null rejection rate. To implement these procedures in practice, however, the user must choose a tuning parameter. One example is the choice of b in the fixed-b scheme,[3] in which a fixed-b fraction of the sample size is used as the bandwidth in kernel LRV estimators. Another example is the choice of q in orthornormal series HAR tests,[4] in which the LRV is estimated by projections onto q mean-zero low-frequency functions of a set of orthonormal functions. The choice of the tuning parameter embeds a tradeoff between bias and variability of the LRV estimator. It subsequently leads to a size-power tradeoff in the resulting HAR inference. Previous studies address this tradeoff by restricting attention to HAR tests that are based on kernel and orthornormal series LRV estimators. They derive the optimal tuning parameter based on second-order asymptotics and under criteria that average the functions

[1]For instance, OLS/GMM with HAR inference has been used in many econometric applications, such as testing long-horizon return predictability in finance (see, e.g., Koijen and Van Nieuwerburgh (2011) and Rapach and Zhou (2013)) and estimating impulse response functions by local projections in macroeconomics (see, e.g., Jordà (2005)).

[2]See, e.g,, den Haan and Levin (1994, 1997) for early Monte Carlo evidence of the large size distortions of HAR tests computed using the Newey-West/Andrews approach.

[3]See pioneering papers by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002, 2005). Also see Jansson (2004); Müller (2004, 2007); Phillips (2005); Phillips, Sun, and Jin (2006, 2007); Sun, Phillips, and Jin (2008); Atchade and Cattaneo (2011); Gonçalves and Vogelsang (2011); Sun and Kaplan (2012); and Sun (2014), among many others.

[4]See, e.g., Müller (2004, 2007); Phillips (2005); Ibragimov and Müller (2010); and Sun (2013), among many others.

of type I and type II errors with different weights.[5] It is not clear, however, whether the resulting HAR tests would remain optimal in finite samples if those restrictions were not imposed.

The purpose of this paper is to provide formal results of finite-sample optimal HAR inference about a scalar parameter of interest, without restricting the class of tests and with no ad hoc loss functions. In particular, I derive finite-sample optimal HAR tests in the Gaussian location model, under nonparametric assumptions on the underlying spectral density. I find that with an appropriate adjustment to the critical value, it is nearly optimal to use the so-called equal-weighted cosine (EWC) test (cf. Müller (2004, 2007); Lazarus, Lewis, Stock, and Watson (2018)), where the LRV is estimated by projections onto q type II cosines.

The main assumption in this paper is that the underlying normalized spectral density is known to lie in a function class $\mathcal{F}$, which possesses a "uniformly minimal" function. By normalized spectral density, I mean its value at the origin is normalized to unity. By "uniformly minimal" function of $\mathcal{F}$, I mean there exists a function $\underline{f}$ in $\mathcal{F}$ such that $\underline{f}(\phi) \leq f(\phi),\ \ \phi \in [-\pi, \pi]$ for all f in $\mathcal{F}$. An explicit stand on possible shapes of the spectrum is necessary, because otherwise there does not exist a nontrivial HAR test (cf. Pötscher (2002)). The notion of "uniformly minimal" function further characterizes the minimal assumption on the spectrum, such that a nontrivial HAR test exists. I stress that the class $\mathcal{F}$ is of a nonparametric nature, as opposed to possibly strong parametric classes.[6] It may contain smoothness restrictions (e.g., bounds on derivatives) and/or shape restrictions (e.g., monotonicity).

This paper makes three main contributions. First, I establish a finite-sample theory of optimal HAR inference in the Gaussian location model. To do so, I follow Müller (2014) and recast HAR inference as a problem of inference about the covariance matrix of a Gaussian vector. The spectrum, as an infinite-dimensional nuisance parameter, complicates solution of the problem. To overcome this obstacle, I use insights from the so-called least favorable approach and identify the "least favorable distribution" over the class $\mathcal{F}$. The resulting optimal test trades off bias and variability, and requires an adjustment of the critical value to account for the maximum bias of the implied LRV estimator. Both the optimal tradeoff and the adjusted critical value are functions of $\mathcal{F}$.

Second, I find that nearly optimal HAR inference can be obtained by using the EWC test, but

[5] See, e.g., Sun, Phillips, and Jin (2008) and Lazarus, Lewis, and Stock (2017).

[6] For examples of parametric classes, Robinson (2005) assumes that the underlying persistence is of the "fractional" type and derives consistent LRV estimators under that class of DGP's; Müller (2014) assumes that the underlying long-run property can be approximated by a stationary Gaussian AR(1) model, with coefficient arbitrarily close to one and derives uniformly valid inference methods that maximize weighted average power.

only after an adjustment to the Student-t critical value. The practical implications are an explicit link between the choice of q and assumptions on the underlying spectrum, as well as a corresponding adjustment to the Student-t critical value. In detail, consider a second-order stationary scalar time series y_t. The spectral density of y_t scaled by 2π is given by the function $f : [-\pi, \pi] \mapsto [0, \infty)$. To test $H_0 : E[y_t] = 0$ against $H_1 : E[y_t] \neq 0$, the EWC test uses a t-statistic

$$t^q_{EWC} = \frac{Y_0}{\sqrt{\sum_{j=1}^q Y_j^2/q}}, \tag{1}$$

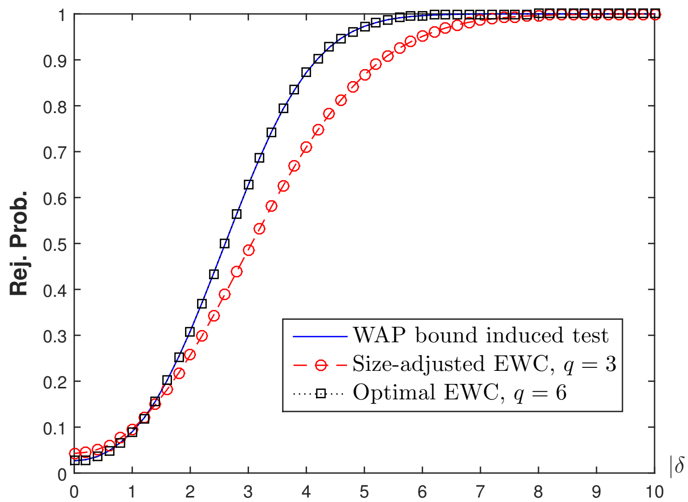
where Y_0 is the sample mean of y_t and $Y_j,\ j = 1, 2, \ldots, q$ are q weighted averages of y_t as $Y_j = T^{-1}\sqrt{2}\sum_{t=1}^T \cos(\pi j(t - 1/2)/T)y_t$. These weighted averages can be approximately thought of as independently distributed, each with variance $T^{-1}f(\pi j/T)$. As mentioned earlier, the choice of q embeds a bias and variance tradeoff of the LRV estimator $\sum_{j=1}^q Y_j^2/q$. The conventional wisdom is to choose q sufficiently small such that $\{Y_j\}_{j=1}^q$ can be treated as independent normal with equal variance. By doing so, one avoid possibly large bias in estimating the LRV, and the resulting EWC test has less size distortions when the Student-t critical value is employed. In contrast, the new EWC test suggests using a larger q and an appropriately enlarged critical value for more powerful inference. Both the choice of q and the critical value adjustment depend on the class $\mathcal{F}$.

Figure 1 illustrates this second contribution in the problem of testing $E[y_t] = 0,\ f \in \mathcal{F}$ against the local alternative $E[y_t] = \delta T^{-1/2}$ for $T = 100$, where y_t follows a Gaussian white noise and the "uniformly minimal" function of $\mathcal{F}$ is the normalized spectrum of an AR(1) model with coefficient 0.8. In this context, to avoid size distortions larger than 0.01, one needs to choose $q = 3$ when the Student-t critical value is employed. The new nearly optimal EWC test, however, has $q = 6$ and inflates the Student-t critical value by a factor of 1.15. This new EWC test nearly achieves a weighted average power bound for all size-controlling scale invariant tests. It has a 38.1% efficiency gain over the size-adjusted EWC test using $q = 3$, in order to achieve the same power of 0.5.[7]

Third, I propose a simple first-order adjustment to the critical value of the EWC test. The adjusted critical value is computed easily, by inverting a one-dimensional numerical integral. For practical convenience, I offer a rule of thumb to adjust the Student-t critical value of the EWC test

[7] By efficiency gain, I mean the increase of δ^2 in percent for the size-adjusted EWC test using $q = 3$ in order to achieve the same power of the new EWC test. I note that one cannot directly appeal to Pitman efficiency measure (the increase of the number of observations required to achieve the same power) in the context of Figure 1, since the sample size T is fixed at 100. A different calculation, however, shows that for $T = 50$ the size-adjusted EWC test using $q = 6$ has power of 0.5, under the same δ such that the EWC test using $q = 3$ yields power of 0.5 for $T = 100$.

Figure 1: Power function plot of a weighted average power (WAP) bound induced test, optimal EWC test and size-adjusted EWC test using $q = 3$.



Notes: Under the alternative, the mean of y_t is $\delta T^{-1/2}$ and y_t follows a Gaussian white noise. Under the null, the "uniformly minimal" function of the class $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

in Table 2, as follows. Under a series of classes $\mathcal{F}$ in which the "uniformly minimal" function is the normalized spectrum of an AR(1) with coefficient ρ, Table 1 lists the optimal choice of q and the adjustment factor of the Student-t critical value for each combination of ρ and nominal level α. Table 2 collapses Table 1 into (α, q) pairs. As a practical matter, if researchers pick a value of q by some other means, then I suggest adjusting the corresponding Student-t critical value according to Table 2.

This paper relates to a large literature. First, unlike the vast majority of the HAR literature, I consider optimal HAR inferences without restricting the class of tests. Second, I do not appeal to ad hoc loss functions in addressing the size-power tradeoff in HAR inference. Rather, I consider that tradeoff in the natural setting of hypothesis testing, where uniform size control is imposed and the most powerful inference is desired. Third, the majority of the literature addresses the sampling variability of LRV estimators via the so-called fixed-b asymptotics, and further accounts for bias by higher-order adjustment to the fixed-b critical value.[8] In contrast, I concurrently tackle bias and

[8] See, e.g., Velasco and Robinson (2001); Sun, Phillips, and Jin (2008); Sun (2011, 2013, 2014); and Lazarus, Lewis, and Stock (2017).

Table 1: Optimal q and adjustment factor of the Student-t critical value of level α EWC test.

ρ	0.5	0.7	0.8	0.9	0.95	0.98	0.999
$\alpha = 0.01$	$(15, 1.07)$	$(10, 1.12)$	$(8, 1.19)$	$(5, 1.34)$	$(3, 1.57)$	$(2, 2.45)$	$(2, 2.45)$
$\alpha = 0.05$	$(12, 1.05)$	$(8, 1.09)$	$(6, 1.13)$	$(4, 1.25)$	$(3, 1.55)$	$(1, 1.85)$	$(1, 31.4)$
$\alpha = 0.10$	$(11, 1.04)$	$(7, 1.07)$	$(5, 1.10)$	$(4, 1.25)$	$(2, 1.36)$	$(1, 1.85)$	$(1, 31.4)$

Notes: Based on a series of classes $\mathcal{F}$, in which the "uniformly minimal" function is the normalized spectrum of an AR(1) with coefficient ρ. Sample size T is 100.

Table 2: Rule of thumb for adjustment factor of the Student-t critical value of level α EWC test.

q	4	6	8	9	10	11	12	13	16	20
$\alpha = 0.01$	1.46	1.32	1.19	1.15	1.12	1.10	1.09	1.08	1.06	1.05
	(0.93)	(0.88)	(0.80)	(0.75)	(0.70)	(0.65)	(0.60)	(0.56)	(0.45)	(0.33)
$\alpha = 0.05$	1.25	1.15	1.09	1.07	1.06	1.06	1.05	1.04	1.04	1.03
	(0.90)	(0.82)	(0.70)	(0.65)	(0.60)	(0.55)	(0.50)	(0.45)	(0.35)	(0.23)
$\alpha = 0.10$	1.25	1.10	1.06	1.05	1.05	1.04	1.04	1.03	1.02	1.02
	(0.90)	(0.78)	(0.65)	(0.60)	(0.55)	(0.50)	(0.45)	(0.40)	(0.28)	(0.18)

Notes: Each q is justified as the optimal choice of level α EWC test, under some class $\mathcal{F}$ and for sample size $T = 100$. An example of the corresponding class $\mathcal{F}$ is the one in which the "uniformly minimal" function is the normalized spectrum of an AR(1) model. Numbers in parentheses are the corresponding AR(1) coefficients.

variance in estimating the LRV by a first-order adjustment. Even so, the resulting adjusted critical value is easily computed without simulations.

The suggestion of using a larger q and enlarged critical values for the EWC test mirrors recent recommendations for nonparametric inference, such as those of Armstrong and Kolesár (2018a,b). In different contexts, Armstrong and Kolesár and I both stress the advantage of accepting bias in estimating a nonparametric function and of then using a suitably adjusted critical value to account for the maximum bias. Our frameworks are, however, different. I consider a Gaussian experiment in which the heteroskedasticity is governed by an unknown nonparametric function, while the main focus in Armstrong and Kolesár (2018a) is an unknown regression function in the mean of a homoskedastic Gaussian experiment.

The remainder of the paper is organized as follows. Section 2 sets up the model and discusses preliminaries. Section 3 derives optimal HAR inference under an essential simplification. Section 4 relaxes the simplification and discusses nearly optimal HAR inference. Section 5 contains simulation results, and Section 6 concludes. Proofs and computational details are provided in the appendices.

2 Model and Preliminaries

Throughout the paper, I mainly focus on inference about μ in the location model,

$$y_t = \mu + u_t,\ t = 1, 2, \ldots, T, \tag{2}$$

where μ is the population mean of y_t and u_t is a mean-zero stationary Gaussian process with absolutely summable autocovariances $\gamma(j) = E[u_t u_{y-j}]$. The spectrum of y_t scaled by 2π is given by the even function $f : [-\pi, \pi] \mapsto [0, \infty)$ defined via $f(\lambda) = \sum_{j=-\infty}^{\infty} \cos(j\lambda)\gamma(j)$. With $y = (y_1, y_2, \ldots, y_T)'$ and $e = (1, 1, \ldots, 1)'$,

$$y \sim \mathcal{N}(\mu e, \Sigma(f)), \tag{3}$$

where $\Sigma(f)$ has elements $\Sigma(f)_{j,k} = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\lambda) \mathrm{e}^{-\mathrm{i}(j-k)\lambda} \mathrm{d}\lambda$ with $\mathrm{i} = \sqrt{-1}$.

The HAR inference problem concerns testing $H_0 : \mu = 0$ (otherwise, subtract the hypothesized mean from y_t) against $H_1 : \mu \neq 0$ based on the observation y. The derivation of powerful tests in this problem is complicated by the fact that the alternative is composite (μ is not specified under H_1), and the presence of the inifinite-dimensional nuisance parameter f. I follow standard approaches to dealing with μ and mainly focus on tackling the nuisance parameter f in this paper.

It is useful to take a spectral transformation of the model (3). In particular, as introduced in the introduction, consider the one-to-one transformation from $\{y_t\}_{t=1}^T$ into the sample mean $Y_0 = T^{-1}\sum_{t=1}^T y_t$ and the $T-1$ weighted averages:

$$Y_j = T^{-1}\sqrt{2}\sum_{t=1}^{T}\cos(\pi j(t-1/2)/T)y_t,\ j=1,2,\ldots,T-1. \tag{4}$$

Define Φ as the $T\times T$ matrix with first column equal to $T^{-1}e$, and $(j+1)$th column with elements $T^{-1}\sqrt{2}\cos(\pi j(t-1/2)/T),\ t=1,\ldots,T$, and ι_1 as the first column of I_T. Then

$$Y = (Y_0, Y_1, \ldots, Y_T)' = \Phi' y \sim \mathcal{N}\left(\mu\iota_1, \Omega_0(f)\right) \tag{5}$$

where $\Omega_0(f) = \Phi'\Sigma(f)\Phi$. The HAR testing problem now becomes $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ based on the observation Y.

A common device for dealing with the composite alternative in the nature of μ is to search for tests that maximize weighted average power over μ. For computational convenience, I follow Müller (2014) to consider a Gaussian weighting function for μ with mean zero and variance η^2. The scalar η^2 governs whether closer or distant alternatives are emphasized by the weighting function. For a given f, the choice $\eta^2 = (\kappa-1)\Omega_0(f)_{1,1}$ (for analytical simplifications later) effectively changes the testing problem to $H_0' : Y \sim \mathcal{N}(0, \Omega_0(f))$ against $H_1' : Y \sim \mathcal{N}(0, \Omega_1(f))$, where

$$\Omega_1(f) = \Omega_0(f) + (\kappa-1)\iota_1\iota_1'\Omega_0(f)_{1,1}. \tag{6}$$

This transforms the problem into one of inference about covariance matrices. The hyperparameter κ specifies a weighted average power criterion. As argued by King (1987), it makes sense to choose κ in a way such that good tests have approximately 50% weighted average power. The choice of $\kappa = 11$ would induce the resulting best 5% level (infeasible) test (reject if $Y_0^2 > 3.84\Omega_0(f)_{1,1}$) to have power of approximately $P(\chi_1^2 > 3.84/11) \approx 56\%$. I thus use $\kappa = 11$ throughout the implementations.

In most applications, it is reasonable to impose that if the null hypothesis is rejected for some observation Y, then it should also be rejected for the observation aY, for any $a > 0$. Due to this scale invariance restriction, it is without loss of generality to normalize all f such that $f(0) \equiv 1$. Furthermore, by standard testing theory (see, e.g., Chapter 6 in Lehmann and Romano (2005)), any test satisfying this scale invariance property can be written as a function of $Y^s = Y/\sqrt{Y'Y}$. The density of Y^s under $H_i',\ i = 0,1$ is equal to (see Kariya (1980) and King (1980))

$$h_{i,f}(y^s) = C|\Omega_i(f)|^{-1/2}\left(y^{s\prime}\Omega_i(f)^{-1}y^s\right)^{-T/2} \tag{7}$$

for some constant C.

By restricting to scale invariant tests, the HAR testing problem has been further transformed into H_0'' : "Y^s has density $h_{0,f}$" against H_1'' : "Y^s has density $h_{1,f}$." The problem remains nonstandard due to the presence of nuisance parameter f. For simplicity, I direct power at flat spectrum $f_1 = 1$. The alternative H_1'' then becomes a single hypothesis H_{1,f_1}'' : "Y^s has density h_{1,f_1}," where $\Omega_1(f_1) = \kappa T^{-1}\text{diag}\left(1, \kappa^{-1}, \ldots, \kappa^{-1}\right)$. Moreover, under the null I assume f belongs to an explicit function class $\mathcal{F}$ and seek scale invariant tests that uniformly control size over $\mathcal{F}$.

The main concern of this paper is to test the composite null H_0'' against H_{1,f_1}''. In this context, a well-known general solution to this type of problem proceeds as follows (cf. Lehmann and Romano (2005)). Suppose Λ is some probability distribution over $\mathcal{F}$, and the composite null H_0'' is replaced by the single hypothesis $H_{0,\Lambda}''$: "Y^s has density $\int h_{0,f} \mathrm{d}\Lambda(f)$." Any ad hoc test φ_{ah} that is known to be of level α under H_0'' also controls size under $H_{0,\Lambda}''$, because $\int \varphi_{\text{ah}}(y^s) \int h_{0,f} \mathrm{d}\Lambda(f) \mathrm{d}y^s = \int\int \varphi_{\text{ah}}(y^s) h_{0,f} \mathrm{d}y^s \mathrm{d}\Lambda(f) \leq \alpha$. As a result, by Neymean-Pearson lemma, the likelihood ratio test of $H_{0,\Lambda}''$ against H_{1,f_1}'', denoted by φ_{Λ,f_1}, yields a bound on the power of φ_{ah}. Furthermore, if φ_{Λ,f_1} also controls size under H_0'', then it must be the best test of H_0'' against H_{1,f_1}'' and the resulting power bound is the lowest possible power bound. In the jargon of statistical testing, the distribution that yields the best test (should it exist) is called the "least favorable distribution," and I denote it by Λ^* throughout the paper.

Unfortunately, there is no systematic way of deriving the least favorable distribution. To make progress, I proceed in the following two steps. First, I consider an approximate "diagonal" model, in which for a given f the joint distribution of Y under the null is

$$Y \sim \mathcal{N}(0, T^{-1}\text{diag}(f(0), f(\pi/T), \ldots, f(\pi(T-1)/T)). \tag{8}$$

In model (8), I analytically derive the least favorable distribution of H_0'' against H_{1,f_1}'', under mild assumptions on the class $\mathcal{F}$. I also find that the "optimal" EWC test is nearly as powerful as the derived optimal test. By optimal EWC test, I mean the EWC test under an optimal choice of q and with an optimal adjustment to the critical value. Second, despite the analytical intractability of the least favorable distribution without approximation (8), it is still feasible to obtain upper bounds on the power of size-controlling tests of H_0'' against H_{1,f_1}''. In particular, I use insights on optimal tests in the diagonal model (8) to establish tight power bounds for all valid tests in the exact model (5). It turns out that the optimal EWC test comes close to achieving this power bound. In light of Lemma 1 in Elliott, Müller, and Watson (2015), this implies that the resulting new EWC test is nearly optimal for HAR inference, and the proposed power bound is essentially the least upper bound. I elaborate on the above analyses in Sections 3 and 4.

Model (8) is in general an approximation of the exact model (5) by ignoring off-diagonal elements and simplifying the diagonal elements in $\Omega_0(f)$. It is motivated by the fact that (8) holds exactly when time series y_t follows a Gaussian white noise or a Gaussian random walk process. For stationary y_t with f falling into other parametric classes, Müller and Watson (2008) find that the covariance matrix of $(Y_0, Y_1, \ldots, Y_q)'$ is nearly diagonal for fixed q and large T. For stationary Gaussian y_t with f being in nonparametric classes, the aforementioned optimality results suggest that (8) is a useful simplification of (5) for HAR inference. I will refer to (8) as the diagonal model and (5) as the exact model hereafter.

3 Optimal HAR Inference in the Diagonal Model

In this section, I derive powerful HAR tests in diagonal model (8). As explained in Section 2, I restrict attention to scale invariant tests that maximize weighted average power over μ and direct power at the flat spectrum f_1. Under the weighted average power criterion, as specified by a given κ, I seek powerful tests as functions of $Y^s = Y/\sqrt{Y'Y}$ in the problem of

$$H_0^d : Y \sim \mathcal{N}\left(0, T^{-1}\text{diag}(1, f(\pi/T), \ldots, f(\pi(T-1)/T)\right), \; f \in \mathcal{F} \tag{9}$$
$$\text{against } H_{1,f_1}^d : Y \sim \mathcal{N}\left(0, \kappa T^{-1}\text{diag}(1, \kappa^{-1}, \ldots, \kappa^{-1})\right),$$

where the superscript d in H_0^d and H_{1,f_1}^d denotes the diagonal model.

The following assumptions are imposed on $\mathcal{F}$ throughout this section.

Assumption 3.1

(a) There exists a $\underline{f} \in \mathcal{F}$ *such that* $\underline{f}(\phi) \leq f(\phi),\; \phi \in [-\pi, \pi]$ *for all* $f \in \mathcal{F}$.

(b) $\underline{f}(\pi j/T) \geq \underline{f}(\pi(j+1)/T),\; j = 0, 1, \ldots, T-2$.

(c) The class $\mathcal{F}$ *contains all kinked functions defined by* $f_a(\phi) = \max\{\underline{f}(\phi), a\}$*, for* $a \in [0, 1]$.

Assumption 3.1(a) states the existence of a "uniformly minimal" function in $\mathcal{F}$, which I will use $\underline{f}$ to denote throughout the paper, while Assumption 3.1(b) requires $\underline{f}$ to be non-increasing at $\lambda = \pi j/T,\; j = 0, 1, \ldots, T-1$. Assumption 3.1(c) is needed to ensure the existence of a point mass least favorable distribution. It is worth noting that only the evaluations of f at frequencies $\pi j/T,\; j = 0, 1, \ldots, T-1$ matter in (9). Assumption 3.1(a) is thus sufficient but not necessary.

Assumption 3.1 is more general than what is assumed in the majority of HAR studies. For example, if $\underline{f}$ is the normalized spectrum of an AR(1) model with coefficient 0.8 and sample size $T = 100$, one is not committing to any parametric classes such as the AR(1) model. Rather, various kinds of parametric classes are covered, as long as the underlying normalized spectrum lies above $\underline{f}$. It includes, but is not limited to, AR(1) models with coefficient less than 0.8 and all MA(1) models and ARMA models whose spectra may oscillate but are above $\underline{f}$. Furthermore, Assumption 3.1 is satisfied by most function classes assumed in the nonparametric inference literature. For example, when $\mathcal{F}$ is the class in which the first derivative of the log spectrum is bounded by a constant C, the corresponding "uniformly minimal" function emerges as $\underline{f}(\phi) = \exp(-C\phi)$.

3.1 Optimal test

The optimal HAR test in the diagonal model is stated in the following theorem.

Theorem 3.2 *Let $\mathcal{F}$ be a set of f satisfying Assumption 3.1 with the "uniformly minimal" function $\underline{f}$, and for a given κ that specifies a weighted average power criterion,*

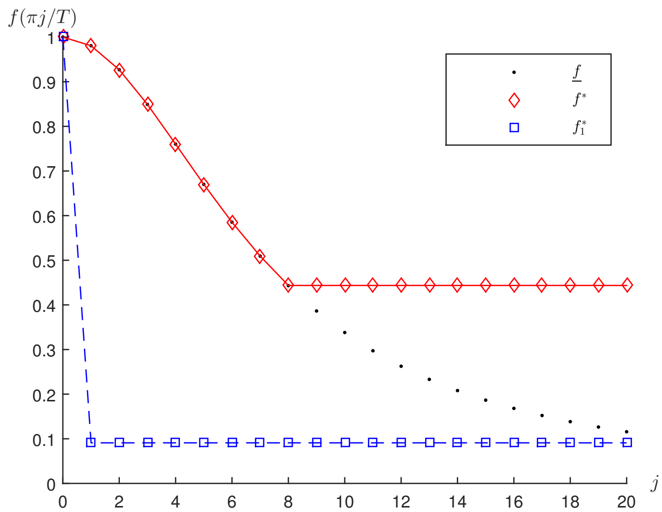
1. *If $\underline{f}(\pi/T) \leq \kappa^{-1}$, then the best weighted average power maximizing scale invariant test of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ is a randomized test.*

2. *If $\underline{f}(\pi/T) > \kappa^{-1}$, then the best level α weighted average power maximizing scale invariant test φ^* of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ rejects for large values of*
$$\frac{Y_0^2 + \sum_{j=1}^{q^*} Y_j^2/\underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2} \tag{10}$$
for a unique $1 \leq q^* \leq T - 1$, and with the critical value cv_{q^*} such that the test is of level α under $f = \underline{f}$.

The proof of part 1 of Theorem 3.2 is simple. Notice that for a given κ, if $\underline{f}(\pi/T) \leq \kappa^{-1}$, then the alternative H''_{1,f_1} is included in the null H_0^d. As a result, any nontrivial size-controlling test cannot be more powerful than a randomized test.

The idea of the proof for part 2 of Theorem 3.2 is to conjecture and verify that the least favorable distribution Λ^* puts a point mass on a function in $\mathcal{F}$. The logic is as follows. Suppose the conjecture is true and Λ^* concentrates on the function f^*. By the Neyman-Pearson lemma, the optimal test of H''_{0,Λ^*} against H''_{1,f_1} in the diagonal model is
$$\varphi_{\Lambda^*,f_1} = 1\left[\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \mathrm{cv}\right],$$

Figure 2: Illustration of the least favorable distribution of H_0^d against H_{1,f_1}^d as a point mass on f^*.



Notes: The "uniformly minimal" function $\underline{f}$ is the normalized spectrum of an AR(1) model with coefficient 0.8. I use $\kappa = 11$ for f_1^*. Sample size T is 100.

for some cv ≥ 0. On the other hand, as discussed in Section 2, for Λ^* to be the least favorable distribution, one needs φ_{Λ^*,f_1} to uniformly control size under H_0''. Intuitively, this requires H_{0,Λ^*}'' to be as indistinguishable as possible from H_{1,f_1}''. This somewhat implies that the function f^* must mimic the discontinuous function of ϕ as $f_1^*(\phi) = \kappa^{-1}\mathbf{1}[\phi \neq 0] + \mathbf{1}[\phi = 0]$. As illustrated by Figure 2, the function f^* must then be kink-shaped, given the presence of $\underline{f}$. I further show that the optimal location of the kink in f^* in conjunction with the resulting cv is equivalent to ignoring Y_j with index $j > q^*$. This then gives rise to the optimal test statistic (10). The formal proof of Theorem 3.2 is given in Appendix A.

3.1.1 Discussion

Comment 1. Part 1 of Theorem 3.2 provides a sharper result than Pötscher (2002). In particular, it characterizes the concrete minimal smoothness assumption on the spectrum such that a nontrivial valid HAR test exists. This is beyond Pötscher's negative result, which only shows that the HAR testing problem is ill-posed if no priori assumptions are imposed on the set of data generating processes.

Comment 2. The optimal test (10) with the resulting critical value cv_{q^*} can be rewritten as

$$\left| \frac{Y_0}{\sqrt{\sum_{j=1}^{q^*} w_j Y_j^2}} \right| > 1, \tag{11}$$

where the weight w_j depends on κ, $\underline{f}(\pi j/T)$ and cv_{q^*}. As can be seen, the implied (inconsistent) LRV estimator $\sum_{j=1}^{q^*} w_j Y_j^2$ does not necessarily fall into the popular kernel and orthonormal series families. This is because the optimal test is not constructed by exploiting flatness of the spectrum close to the origin. Rather, I take an explicit stand on possible shapes of the spectrum and endogenously account for the maximum bias via the weights w_j's.

Comment 3. A possibly small q^* may emerge in Theorem 3.2 for some $\mathcal{F}$. It is, however, worth noting that such q^* is an implication, not an assumption. In particular, I do not start by restricting attention to the class of tests as functions of $\left(Y_1^s, \ldots, Y_{q^*}^s\right)'$. In contrast, the approach taken by Müller (2014) assumes some fixed q as the starting point.

Comment 4. With appropriate modifications of Assumption 3.1, Theorem 3.2 can be adapted to the problem of H_0^d against $H_{1,\tilde{f}_1}^d$ for other fixed alternative $\tilde{f}_1$. In that case, the resulting q^* is also $\tilde{f}_1$-dependent. Furthermore, Theorem 3.2 can be generalized to a minimax result, in which f belongs to a nonparametric class $\mathcal{G} \subset \mathcal{F}$ under H_1. In that case, a "uniformly maximal" function in $\mathcal{F}$ must be properly defined as $\underline{f}$ in Assumption 3.1.

3.1.2 Computational considerations

The existence and uniqueness result of q^* in Theorem 3.2 naturally brings computational convenience in practice. For example, for a given $\mathcal{F}$ satisfying Assumption 3.1, one can appeal to the bisection method to locate q^*. In my implementations, this takes little computing time by using the simple algorithm in Appendix B.1. Moreover, for a given $\mathcal{F}$ and the resulting q^*, the critical value cv_{q^*} can easily be determined due to the following formula of Bakirov and Székely (2006):

$$P\left(Z_0^2 \geq \sum_{j=1}^{n} \zeta_j Z_j^2\right) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(n-1)/2} \mathrm{d}u}{\sqrt{\prod_{j=1}^{n}(1-u^2+\zeta_j)}}, \tag{12}$$

where $\{Z_j\}_{j=0}^n$ are $n+1$ i.i.d. standard normal random variables and $\zeta_j \geq 0,\ j = 1, \ldots, n$. By the t-statistic expression (11), part 2 of Theorem 3.2, and (12), the level α constraint for the optimal test becomes

$$P\left(\frac{Z_0^2}{\sum_{j=1}^{q^*} w_j f(\pi j/T) Z_j^2} > 1\right) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(q^*-1)/2} \mathrm{d}u}{\sqrt{\prod_{j=1}^{q^*}(1-u^2+w_j \underline{f}(\pi j/T))}} = \alpha, \tag{13}$$

where $w_j = \left[\kappa \, \mathrm{cv}_{q^*} - 1/\underline{f}(\pi j/T)\right](1 - \mathrm{cv}_{q^*})^{-1}$ is strictly monotone in cv_{q^*} under Assumption 3.1. The critical value cv_{q^*} is then readily determined by solving equation (13). Computational details are provided in Appendix B.2.

3.2 The optimal EWC test

By using higher-order expansions, Lazarus, Lewis, and Stock (2017) derive a size-power frontier for kernel and orthornormal series HAR tests under an asymptotic framework. The EWC test is shown to achieve that frontier in their context. It is, however, not clear how the EWC test performs in finite-sample contexts and in the unrestricted class of tests. The optimal HAR test derived in the last section provides a natural benchmark to gauge the performance of an ad hoc test. In this section I take up the EWC test as the ad hoc test and discuss its properties.

I have three related goals. The first is to study the (weighted average) power properties of the EWC test relative to the optimal test in Theorem 3.2. As it turns out, the EWC test is close to optimal, under an appropriate choice of q and with the adjusted critical value. Given the efficiency property of this new EWC test, the second goal is to develop simple procedures to implement the test. I discuss the two goals in reverse order, first elaborating on critical value adjustment and optimal choice of q for the EWC test, and then studying the power of the resulting test. My last goal is to compare the practical implications of the new EWC test with the conventional wisdom, that is, to choose a sufficiently small q and use the Student-t critical value. The general takeway from the comparison is: One should use the EWC test with a larger q and appropriately enlarged critical values for more powerful HAR inference.

To clarify ideas and illustrate points in a consistent manner, I use the following running example throughout this section: Under the null, the "uniformly minimal" function $\underline{f}$ of the class $\mathcal{F}$ is AR(1) with coefficient 0.8; sample size T is fixed to be 100. In addition, I will frequently refer to the following two types of classes. For the first type, the "uniformly minimal" function of $\mathcal{F}$ is the normalized spectrum of an AR(1) with coefficient ρ. For the second type, all spectra in $\mathcal{F}$ satisfy a global smoothness assumption, that is, the first derivative of the log-spectrum $\log(f)$ is bounded by a constant C.

3.2.1 Critical value adjustment and choice of q

The diagonal model (8) makes it easy to adjust the critical value for a given class $\mathcal{F}$. In particular, for a given $f \in \mathcal{F}$, the null rejection probability of the EWC test (1) using the critical

Table 3: Size of the 5% level EWC test using Student-t critical values under selected q.

q	**3**	4	6	8	10
size	**0.056**	0.061	0.073	0.089	0.107

Notes: The "uniformly minimal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

value cv is

$$P\left(\left|\frac{Y_0}{\sqrt{\sum_{j=1}^q Y_j^2/q}}\right| \geq \text{cv}\right) = P\left(\frac{Z_0^2}{q^{-1}\,\text{cv}^2\sum_{j=1}^q f(\pi j/T)Z_j^2} \geq 1\right), \tag{14}$$

where $\{Z_j\}_{j=0}^q$ are $q+1$ i.i.d. standard normal random variables. Under Assumption 3.1(a), it is not hard to see that (14) as a functional of f is maximized at $\underline{f}$, regardless of the choice of q and the critical value cv. Two implications are immediate. First, for the testing problem (9) under a given $\mathcal{F}$, it is easy to gauge the size performance of any ad hoc EWC test. In the context of the running example, Table 3 shows the size of the 5% EWC test using the Student-t critical value under selected choices of q. As can be seen, for size distortions less than 0.01, one needs to use $q = 3$ in the usual EWC test.

Second, by Bakirov and Székely's (2006) formula (12), it is easy to adjust the critical value of the EWC test under any ad hoc q. Specifically, as in solving for the critical value of the optimal test in Section 3.1.2, the adjusted critical value cv_q^a of the level α EWC test under given q is obtained by inverting the following level constraint:

$$P\left(\frac{Z_0^2}{q^{-1}(\text{cv}_q^a)^2\sum_{j=1}^q f(\pi j/T)Z_j^2} \geq 1\right) = \frac{2}{\pi}\int_0^1 \frac{(1-u^2)^{(q-1)/2}\text{d}u}{\sqrt{\prod_{j=1}^q(1-u^2+q^{-1}(\text{cv}_q^a)^2\underline{f}(\pi j/T))}} = \alpha. \tag{15}$$

In the context of the running example, the first row in Table 4 summarizes the adjustment factor of the resulting adjusted critical value relative to the Student-t critical value under various q. As can be seen, in order to explicitly account for the resulting downward bias of the LRV estimator $\sum_{j=1}^q Y_j^2/q$, one must inflate the usual Student-t critical value by a factor larger than 1.

Now consider the choice of q in the EWC test. Under given q and using the adjusted critical value cv_q^a, the weighted average power of the resulting EWC test is

$$P\left(\frac{Z_0^2}{q^{-1}\kappa(\text{cv}_q^a)^2\sum_{j=1}^q Z_j^2} \geq 1\right) = \frac{2}{\pi}\int_0^1 \frac{(1-u^2)^{(q-1)/2}\text{d}u}{\sqrt{\prod_{j=1}^q\left(1-u^2+q^{-1}\kappa(\text{cv}_q^a)^2\right)}}. \tag{16}$$

Table 4: Adjustment factor of the Student-t critical value and weighted average power of 5% level EWC test under selected q.

q	3	4	5	**6**	7	8	9	10
adjustment factor	1.044	1.068	1.096	1.126	1.158	1.191	1.225	1.259
weighted average power	0.390	0.422	0.434	**0.438**	0.436	0.431	0.425	0.417

Notes: The "uniformly minimal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

Table 5: Weighted average power (WAP) of the optimal test and the optimal EWC test.

ρ	0.50	0.60	0.70	0.80	0.90	0.95	0.98	0.99
WAP of optimal test	0.506	0.493	0.475	0.441	0.357	0.236	0.089	0.051
WAP of optimal EWC	0.504	0.491	0.472	0.438	0.353	0.233	0.089	0.051

Notes: The "uniformly minimal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient ρ. Nominal level is 5%. Sample size T is 100.

The weighted average power (16) can easily be computed for every q. Under a given $\mathcal{F}$ and nominal level α, the optimal choice of q for the EWC test is then defined as the one such that the resulting EWC test has the largest weighted average power. I refer to the EWC test under the optimal choice of q and with the adjusted critical value as the optimal EWC test. I stress that the notion of "optimality" for this new EWC test is with respect to the assumptions on the underlying spectrum, that is, the class $\mathcal{F}$.

3.2.2 Power of the optimal EWC test

Tables 5 and 6 summarize the weighted average power of the optimal EWC test and the corresponding optimal test under the aforementioned two types of classes $\mathcal{F}$, respectively. As can be seen, the optimal EWC test is nearly as powerful as the optimal test, regardless of the underlying $\mathcal{F}$ within the two types of classes. In unreported numerical results, under various $\mathcal{F}$ of other smoothness types, the near optimality property of the optimal EWC test continues to hold.

Table 6: Weighted average power (WAP) of the optimal test and the optimal EWC test.

C	10.0	5.6	3.2	1.8	1.0	0.6	0.3	0.2	0.1
WAP of optimal test	0.290	0.365	0.419	0.458	0.485	0.504	0.517	0.527	0.534
WAP of optimal EWC	0.286	0.361	0.415	0.454	0.482	0.501	0.515	0.526	0.533

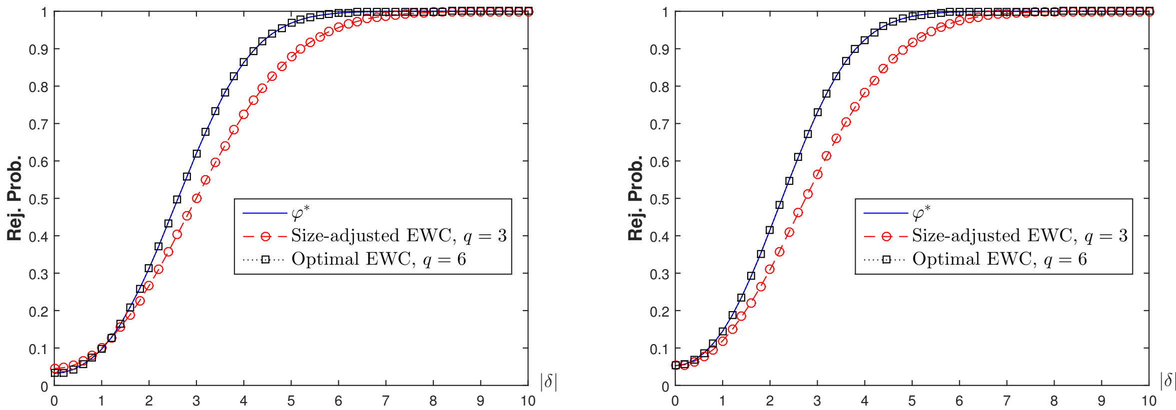
Notes: The "uniformly minimal" function of $\mathcal{F}$ is $\underline{f}(\phi) = \exp(-C\phi)$. Nominal level is 5%. Sample size T is 100.

3.2.3 Practical implications

Recall that the conventional wisdom is to use a sufficiently small q and to employ the Student-t critical value. I find, however, that it is optimal to use a larger q and to employ an enlarged critical value. Take the running example as an illustration. As explained earlier, for size distortions less than 0.01, one needs to use $q = 3$ in the usual EWC test in which the Student-t critical value is employed. However, as highlighted in Table 4, the optimal EWC test has a larger $q = 6$, and the corresponding Student-t critical value must be inflated by a factor of 1.13 for exact size control. To ensure an apples-to-apples comparison, I compute the size-adjusted weighted average power of the usual EWC test using $q = 3$. In contrast to the optimal EWC test, this EWC test has about 11% weighted average power loss.

The superior power property of the optimal EWC test is further evident when local alternatives are considered. In particular, in the context of the running example, I consider $\mu = \delta T^{-1/2}(1-\rho_1)^{-1}$ under the alternative. Panels (a) and (b) of Figure 3 plot the power of the test φ^*, the optimal EWC test, and the size-adjusted EWC test using $q = 3$ for various δ under $\rho_1 = 0$ and $\rho_1 = 0.8$, respectively. As can be seen in panel (a), even though the optimal EWC test underrejects under the null, it is more powerful than the EWC test using $q = 3$ in detecting local deviations from the null. Specifically, by using the optimal EWC test, a 32.0% efficiency improvement is obtained in order to achieve the same power of 0.5. In the case in which $\rho_1 = 0.8$, the efficiency gain is larger (48.7%), since the optimal EWC test then exactly controls size by construction. Furthermore, given that the optimal EWC test is numerically found to be nearly as powerful as the overall optimal test φ^* in terms of weighted average power under the white noise alternative, it is not surprising to see that the power functions of these two tests are almost identical.

Figure 3: Power function plot of the test φ^*, the optimal EWC test, and the size-adjusted EWC test using $q = 3$.



(a) $\rho_1 = 0$. (b) $\rho_1 = 0.8$.

Notes: Under the alternative, the mean of y_t is $\delta T^{-1/2}(1-\rho_1)^{-1}$ and y_t follows a Gaussian AR(1) with coefficient ρ_1. Under the null, the $\underline{f}$ of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

3.3 A Rule of thumb

As a practical matter, one might like to estimate the smoothness class $\mathcal{F}$ from data. Unfortunately, the attempt is not useful. This is because the (nearly) optimal test depends on $\mathcal{F}$, and a "larger" $\mathcal{F}$ leads to a lower power. As a result, one cannot estimate $\mathcal{F}$ and still control size. In implementations of the EWC test, if q is chosen by some other approach, it still makes sense to adjust the Student-t critical value given the previous analysis of the optimal EWC test. As a rule of thumb, I suggest that practitioners implement the EWC test and adjust the Student-t critical value according to Table 2. In detail, the suggested test about the population mean $H_0 : \mu = \mu_0$ of an observed scalar time series $\{y_t\}_{t=1}^T$ is computed as follows.

1. Compute the T cosine weighted averages of $\{y_t\}_{t=1}^T$: $Y_0 = T^{-1}\sum_{t=1}^T (y_t - \mu_0)$ and $Y_j = T^{-1}\sqrt{2}\sum_{t=1}^T \cos(\pi j(t-1/2)/T)y_t,\ j = 1, 2, \ldots, T-1$.

2. For the researcher's choice of q, compute the t-statistic $t_{Y,q} = Y_0/\sqrt{q^{-1}\sum_{j=1}^q Y_j^2}$.

3. Reject the null hypothesis at level α if $|t_{Y,q}| > B_{\alpha,q}\,\mathrm{cv}_q^{na}(\alpha)$, where $\mathrm{cv}_q^{na}(\alpha)$ is the Student-t_q critical value and $B_{\alpha,q}$ is the unparenthesized number in the (α, q)-th entry of Table 2.

As explained in the introduction, the adjustment factors in Table 2 are calibrated based on a series of classes $\mathcal{F}$ in which the "uniformly minimal" function is the normalized spectrum of an AR(1) with coefficient ρ. I make two additional remarks. First, there may be multiple ρ such that the same optimal q emerges, under the respective $\mathcal{F}$. Second, the adjustment factor in each (α, q)-th entry does not change substantially under other types of smoothness classes. Sets of tables similar to Tables 1 and 2 are provided in Appendix C, in which the class $\mathcal{F}$ imposes some global smoothness assumption on the spectrum. For these reasons, one should take Table 2 as rule of thumb. The parenthesized ρ values in Table 2 only serve as a reference for the underlying smoothness class.

4 Nearly Optimal HAR Inference in the Exact Model

The discussions in Section 3 are entirely based on the diagonal model (8). For both theoretical interest and practical relevance, it is natural to ask whether the insights on optimal HAR inference in the diagonal model continue to hold in the exact model (5). This section is devoted to addressing that problem. In particular, I continue restricting attention to scale invariant tests of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ that maximize weighted average power over μ and direct power at the flat spectrum f_1. Under the weighted average power criterion, as specified by a given κ, the goal is to seek powerful tests as functions of $Y^s = Y/\sqrt{Y'Y}$ in the problem of

$$\begin{aligned} H_0^e &: Y \sim \mathcal{N}\left(0, \Omega_0(f)\right), \ f \in \mathcal{F} \\ \text{against } H_{1,f_1}^e &: Y \sim \mathcal{N}\left(0, \kappa T^{-1} \text{diag}(1, \kappa^{-1}, \ldots, \kappa^{-1})\right), \end{aligned} \tag{17}$$

where the superscript e denotes the exact model.

First of all, I note that it is in general difficult to derive the optimal test of (17) under Assumption 3.1. This is mainly due to the complicated manner by which f enters $\Omega_0(f)$. In detail, a direct calculation shows that for a given f and $j, k = 0, 1, \ldots, T-1$,

$$\Omega_0(f)_{j,k} = \int_{-\pi T}^{\pi T} f\left(\frac{\lambda}{T}\right) \left(T^{-1} \sum_{s=1}^{T} \varphi_j\left(\frac{s-1/2}{T}\right) e^{-\frac{\mathrm{i}s\lambda}{T}}\right) \left(T^{-1} \sum_{t=1}^{T} \varphi_k\left(\frac{t-1/2}{T}\right) e^{\frac{\mathrm{i}t\lambda}{T}}\right) d\lambda, \tag{18}$$

where $\varphi_j(\phi) = \left(\sqrt{2}\right)^{\mathbf{1}[j\neq 0]} \cos(\pi j s)$, $0 \leq \phi \leq 1$. In this case, even if it is true that the least favorable distribution of (17) puts a point mass on some function $f^* \in \mathcal{F}$, the determination of f^* seems very diffcult. Alternatively, one may want to impose additional assumptions on $\mathcal{F}$ such that $\Omega_0(f) = T^{-1}\text{diag}\left(f(0), \ldots, f(\pi(T-1)/T)\right)$ holds uniformly in $f \in \mathcal{F}$. The task is also hard, since one must then solve $(T^2+T)/2$ functional constraints, that is, $\Omega_0(f)_{j,k}=0$ for every $j>k$ and $\Omega_0(f)_{i,i}=f(\pi i/T)$ for every i.

Despite the difficulty in analytically deriving the exact optimal test of (17), one still can obtain bounds on the power of any size-controlling test by using the bounding approach of Elliott, Müller, and Watson (2015). Recall from Section 2 that for any probability distribution Λ over $\mathcal{F}$, the likelihood ratio test of $H''_{0,\Lambda}$ against H''_{1,f_1} yields such a power bound. Suppose there exists an ad hoc test φ_{ah} that is known to control size. If the power of φ_{ah} is close to the power bound for some Λ, then φ_{ah} is known to be close to optimal, as no substantially more powerful test exists. It turns out that the insights from the diagonal model are useful in guessing a good Λ and in justifying the near optimality of the EWC test in the exact model. In particular, for a given a in $[0,1]$, let Λ_a be a point mass distribution on the kinked function $f_a(\phi)$, as was defined in Assumption 3.1. It is already known that for every a, the likelihood ratio test of H''_{0,Λ_a} against H''_{1,f_1} yields a power bound. I then numerically search for a such that the resulting power bound is minimized. Denote this a by $a^\dagger$ and the resulting Λ by $\Lambda^\dagger$. The power bound I employ to gauge potential efficiency of the EWC test is then the power of

$$\varphi_{\Lambda^\dagger,f_1}=1\left[\left(Y'\Omega_0(f_{a^\dagger})Y\right)^{-1}\left(Y'\Omega_1(f_1)Y\right)>\mathrm{cv}\right], \tag{19}$$

for some cv such that $E[\varphi_{\Lambda^\dagger,f_1}]=\alpha$ under H''_{0,Λ_a}. In the following subsection, I show that the EWC test essentially achieves this bound, after optimal choice of q and critical value adjustment.

To clarify ideas and illustrate points in a consistent manner, I continue using the running example introduced in Section 3.2. I also use the two types of smoothness classes introduced there, except that for the first type I additionally assume every $f\in\mathcal{F}$ to be non-increasing over $[0,\pi]$.

4.1 The optimal EWC test

I discuss the EWC test in the exact model in the following steps. First, given the aforementioned efficiency property of the EWC test, I elaborate on how to make the critical value adjustment and choose the q for the EWC test in the exact model. Second, I use numerical exercises to study the power of the resulting EWC test. Third, as was done in Section 3, I compare practical implications of the new EWC test with the conventional wisdom. The general takeaway remains: One should use the EWC test with a larger q and appropriately enlarged critical values for more powerful HAR inference. Lastly, as a practical matter, I examine the robustness of the rule of thumb suggested in Section 3.3. I find that there is no substantial change in the adjustment factor of the Student-t

critical value, even if the adjustment is made in the exact model.

4.1.1 Critical value adjustment and choice of q

I first note that, unlike in the diagonal case, there is no analytical expression to adjust the critical value for the EWC test in the exact model. To be precise, at given $f \in \mathcal{F}$, the null rejection probability of the EWC test under given q and with the critical value cv is

$$P\left(\left|\frac{Y_0}{\sqrt{\sum_{j=1}^{q} Y_j^2/q}}\right| \geq \mathrm{cv}\right) = P\left(\frac{Z_0^2}{q^{-1}\,\mathrm{cv}^2 \sum_{j=1}^{q} \lambda_j(f) Z_j^2} \geq 1\right), \tag{20}$$

where $\{Z_j\}_{j=0}^{q}$ are $q+1$ i.i.d. standard normal random variables. The $\lambda_j(f), j = 1, \ldots, q$ in (20) are the normalized positive eigenvalues of $\Omega_{0,q}(f)^{1/2} M(\mathrm{cv}, q)\Omega_{0,q}(f)^{1/2}$ (normalized by the absolute value of the only negative eigenvalue), where $\Omega_{0,q}(f)$ is the upper left $(q+1)\times(q+1)$ block matrix of $\Omega_0(f)$ and $M(\mathrm{cv}, q) = \mathrm{diag}\left(-1, \mathrm{cv}^2/q^2, \mathrm{cv}^2/q^2, \ldots, \mathrm{cv}^2/q^2\right)$. It is known from Section 3 that (20) as a functional of f is maximized when all $\lambda_j(f)$'s are jointly minimized. The opaque mapping from $\lambda_j(f)$ back to f, however, prevents us from explicitly identifying the null rejection probability maximizer(s) like $\underline{f}$ in the diagonal model.

A natural reaction to this obstacle is to search for the null rejection probability maximizer numerically. To render this feasible, I approximate $f \in \mathcal{F}$ as a linear combination of basis functions. The original task is then transformed into a high-dimensional optimization problem. To be more precise, let the $n+1$ node points $\{x_i\}_{i=0}^{n}$ define a partition of the interval $I = [0, \pi]$ into n subintervals $I_i = [x_{i-1}, x_i]$, $i = 1, 2, \ldots, n$, each of length $h_i = x_i - x_{i-1}$, and $x_0 = 0$, $x_n = \pi$. Let $\mathcal{C}^0(I)$ denote the space of continuous functions on I, and $\mathcal{P}_1(I_i)$ denote the space of linear functions on I_i. Let $\{\varsigma_i\}_{i=0}^{n}$ be a set of basis functions for the space $\mathcal{F}_h$ of continuous piecewise linear functions defined by $\mathcal{F}_h = \{f : f \in \mathcal{C}^0(I), f|_{I_i} \in \mathcal{P}_1(I_i)\}$. The basis functions $\{\varsigma_i\}_{i=0}^{n}$ are normalized such that $\varsigma_j(x_i) = \mathbf{1}[i = j]$, $i, j = 0, 1, \ldots, n$. By approximating f via $\hat{f} = \sum_{i=0}^{n} f(x_i)\varsigma_i$ and (12), I approximate the rejection probability (20) by

$$P\left(\frac{Z_0^2}{q^{-1}\,\mathrm{cv}^2 \sum_{j=1}^{q} \lambda_j(\hat{f}) Z_j^2} \geq 1\right) = \frac{2}{\pi}\int_0^1 \frac{(1-u^2)^{(q-1)/2}\mathrm{d}u}{\sqrt{\prod_{j=1}^{q}\left(1-u^2+q^{-1}\,\mathrm{cv}^2\,\lambda_j\left(\hat{f}\right)\right)}}, \tag{21}$$

which is a function of the n-dimensional vector $(f(x_1), f(x_2), \ldots, f(x_n))'$. (By normalization, $f(x_0) = 1$.) With pre-computed $\{\Omega_0(\varsigma_i)\}_{i=0}^{n}$ based on (18), the computation of (21) takes very little computing time for each $\hat{f}$, and it is feasible to obtain a global maximizer of (21) subject to implied constraints on $(f(x_1), f(x_2), \ldots, f(x_n))'$ from a given $\mathcal{F}$. Denote the λ_j's at one of those

Table 7: Diagonal model based cv_q^a, exact model based $\text{cv}_q^{a,e}$, and weighted average power (WAP) of 5% level EWC test using $\text{cv}_q^{a,e}$.

q	3	4	5	**6**	7	8	9	10
cv_q^a	3.322	2.966	2.817	2.756	2.739	2.747	2.772	2.806
$\text{cv}_q^{a,e}$	3.392	3.022	2.868	2.800	2.780	2.783	2.804	2.835
WAP	0.382	0.414	0.427	**0.431**	0.430	0.426	0.420	0.413

Notes: The "uniformly minimal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. All f in $\mathcal{F}$ are non-increasing over $[0, \pi]$. Sample size T is 100.

global maximizers by $\{\lambda_j^*\}_{j=1}^q$. The adjusted critical value $\text{cv}_q^{a,e}$ is then readily determined by inverting

$$\frac{2}{\pi}\int_0^1 \frac{(1-u^2)^{(q-1)/2}\mathrm{d}u}{\sqrt{\prod_{j=1}^q\left(1-u^2+q^{-1}(\text{cv}_q^{a,e})^2\lambda_j^*\right)}} = \alpha,$$

just like solving (15) in the diagonal model. I provide more computational details on numerically locating the null rejection probability maximizer in Appendix D.

In the context of the running example, I additionally assume that the underlying spectrum is non-increasing over $[0, \pi]$. Table 7 lists the resulting $\text{cv}_q^{a,e}$ and cv_q^a under selected q. As can be seen, the difference between these two adjusted critical values is slight. What's more, it is observed that neither $\text{cv}_q^{a,e}$ nor cv_q^a uniformly dominates each other as a function of q. All of these suggest that even if the exact critical value adjustment of the EWC test is complex, the simple rule proposed in Section 3.2.1 is not only practically convenient, but also without loss of generality.

Now consider the choice of q in the EWC test. I note that since the alternative hypothesis of (17) is identical to that of (9), one can proceed as in Section 3.2.1 to choose the optimal q such that the resulting EWC test has the largest weighted average power. The only difference is that one must replace the adjusted critical value cv_q^a by $\text{cv}_q^{a,e}$ in (16). I refer to the EWC test under the optimal choice of q and with the adjusted critical value $\text{cv}_q^{a,e}$ as the optimal EWC test for the rest of this section.

4.1.2 Power of the optimal EWC test

Table 8 and 9 summarize the weighted average power of the optimal EWC test and the weighted average power bound induced by (19), under the two types of classes described in the beginning

Table 8: A bound on weighted average power (WAP) and the WAP of the optimal EWC test.

ρ	0.50	0.60	0.70	0.80	0.90	0.95	0.98	0.99
WAP of optimal EWC	0.502	0.488	0.467	0.431	0.344	0.231	0.096	0.071
WAP bound	0.505	0.492	0.473	0.438	0.361	0.257	0.132	0.088

Notes: The "uniformly minimal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient ρ. All f in $\mathcal{F}$ are non-increasing over $[0, \pi]$. Nominal level is 5%. Sample size T is 100.

Table 9: A bound on weighted average power (WAP) and the WAP of the optimal EWC test.

C	10.0	5.6	3.2	1.8	1.0	0.6	0.3	0.2	0.1
WAP of optimal EWC	0.307	0.372	0.422	0.458	0.484	0.503	0.517	0.527	0.534
WAP bound	0.323	0.382	0.428	0.463	0.488	0.506	0.518	0.528	0.534

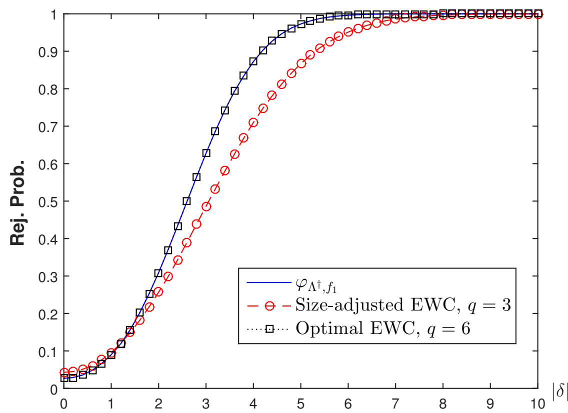
Notes: The "uniformly minimal" function of $\mathcal{F}$ is $\underline{f}(\phi) = \exp(-C\phi)$. Nominal level is 5%. Sample size $T = 100$.

of Section 4, respectively. As can be seen, for most $\mathcal{F}$ under consideration, the optimal EWC test essentially achieves the corresponding weighted average power bound. In the spirit of Lemma 1 in Elliott, Müller, and Watson (2015), the optimal EWC test is then known to be nearly optimal. The numerical findings also imply that the insights from the diagonal model continue to be useful in the exact model, even if the analysis of the overall optimal test is hard. I note that the relatively larger difference between the weighted average power of the optimal EWC test and the corresponding bound (e.g., under large ρ in Table 8 and under large C in Table 9) is not informative about the efficiency of the optimal EWC test, since it can arise either because the bound is far from the least upper bound, or because φ_{ah} is inefficient.

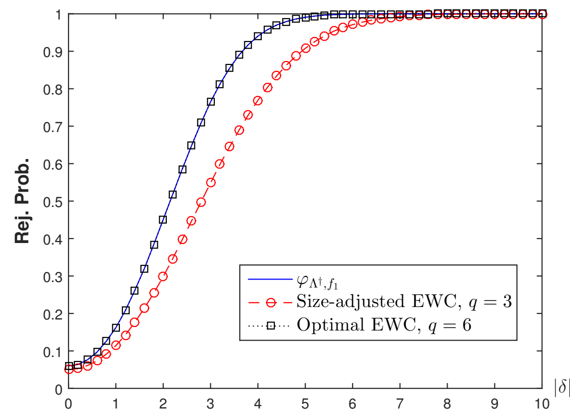
4.1.3 More on the practical implications and the rule of thumb

The practical implication on using the EWC test from the diagonal model continue to hold. In the context of the running example, for size distortions less than 0.01 in the exact model, one must use $q = 3$ in the usual EWC test. As highlighted in Table 7, the optimal EWC test has a larger $q = 6$. Moreover, one must enlarge the corresponding Student-t critical value by a factor of 1.15 for exact size control. In terms of weighted average power, there is a 13% gain by using the optimal EWC test. This efficiency advantage is further evident when the local alternative

Figure 4: Power function plot of the optimal EWC test, the size-adjusted EWC test using $q = 3$, and the weighted average power bound induced test $\varphi_{\Lambda^\dagger, f_1}$.



(a) $\rho_1 = 0$.



(b) $\rho_1 = 0.8$.

Notes: Under the alternative, the mean of y_t is $\delta T^{-1/2}(1 - \rho_1)^{-1}$ and y_t follows a Gaussian AR(1) with coefficient ρ_1. Under the null, The "uniformly minimal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size $T = 100$.

$\mu = \delta T^{-1/2}(1 - \rho_1)^{-1}$ is considered. Panels (a) and (b) of Figure 4 plot the power of $\varphi_{\Lambda^\dagger, f_1}$ as in (19), the optimal EWC test, and the size-adjusted EWC test using $q = 3$ for various δ under $\rho_1 = 0$ and $\rho_1 = 0.8$, respectively. To achieve the same power of 0.5, there is a 38.1% and 71.1% efficiency gain by using the optimal EWC test under $\rho_1 = 0$ and $\rho_1 = 0.8$, respectively.

In Table 10, I recompute the adjustment factor of the Student-t critical value for each (α, q) pair, but use the adjusted critical value $\text{cv}_q^{a,e}$ in the exact model. The calibrations are based on the type of smoothness classes $\mathcal{F}$ in which the "uniformly minimal" function is the normalized spectrum of an AR(1) with coefficient ρ and $f \in \mathcal{F}$ is non-increasing over $[0, \pi]$, and are under sample size $T = 100$. Not surprisingly, there is no substantial change in the adjustment factors from Table 2 to Table 10. As a rule of thumb, I thus recommend that practitioners practice the EWC test in HAR inference, by following the three simple steps in Section 3.3 and adjusting the Student-t critical value according to Table 2.

Table 10: Rule of thumb for adjustment factor of the Student-t critical value of level α EWC test in the exact model.

q	4	6	8	9	10	11	12	13	14	15	16	18	20
$\alpha = 0.01$	1.47	1.34	1.21	1.16	1.13	1.11	1.09	1.09	1.08	1.08	1.07	1.06	1.06
	(0.93)	(0.88)	(0.80)	(0.75)	(0.70)	(0.65)	(0.60)	(0.56)	(0.52)	(0.50)	(0.45)	(0.40)	(0.35)
$\alpha = 0.05$	1.28	1.17	1.10	1.08	1.07	1.06	1.05	1.05	1.04	1.04	1.04	1.03	1.03
	(0.90)	(0.82)	(0.70)	(0.65)	(0.60)	(0.55)	(0.50)	(0.45)	(0.40)	(0.38)	(0.35)	(0.29)	(0.23)
$\alpha = 0.10$	1.28	1.12	1.07	1.06	1.05	1.05	1.04	1.04	1.03	1.03	1.03	1.02	1.02
	(0.90)	(0.78)	(0.65)	(0.60)	(0.55)	(0.50)	(0.45)	(0.40)	(0.35)	(0.30)	(0.28)	(0.23)	(0.19)

Notes: Each q is justified as the optimal choice of level α EWC test, under some class $\mathcal{F}$ and for sample size $T = 100$. An example of the corresponding class $\mathcal{F}$ is the one in which the "uniformly minimal" function is the normalized spectrum of an AR(1) model. Numbers in parentheses are the corresponding AR(1) coefficients.

5 Monte Carlo Simulations

The purpose of this section is twofold. First, I assess size and power performance of the suggested optimal EWC test relative to other approaches to HAR inference in the Gaussian location model. Second, I investigate the extent to which the theory derived in the Gaussian location model generalizes to inference about a scalar parameter of interest in a regression context.

I compare 18 tests in total. For the EWC test using the Student-t critical value, I consider $q = 4, 8, 12, 24$. For the optimal EWC test, labeled OEWC, I consider $q = 3, 6, 8, 10, 20$. According to Table 2, these q's are the optimal choices for the EWC test, under the class $\mathcal{F}$ in which the "uniformly minimal" function is the normalized spectrum of an AR(1) with coefficient 0.95, 0.82, 0.70, 0.60, and 0.23, respectively for $T = 100$. In addition, I consider Müller's (2014) S_q test with $q = 12, 24, 48$; Ibragimov and Müller's (2010) test with 8 and 16 groups, IM_8 and IM_{16}; the classical approach based on two consistent LRV estimators: Andrews's (1991) LRV estimator $\hat{\omega}^2_{\text{A91}}$ with a quadratic spectral kernel and bandwidth selection using an AR(1) model, and Andrews and Monahan's (1992) LRV estimator $\hat{\omega}^2_{\text{AM}}$, in which an AR(1) model is used in prewhitening; and two HAR tests based on inconsistent LRV estimators: Kiefer, Vogelsang, and Bunzel's (2000) Bartlett kernel estimator $\hat{\omega}^2_{\text{KVB}}$ with bandwidth equal to the sample size, and Sun, Phillips, and Jin's (2008) quadratic spectral estimator $\hat{\omega}^2_{\text{SPJ}}$ with a bandwidth that trades off asymptotic type I and type II errors in rejection probabilities, in which the shape of the spectrum is approximated by an AR(1) model and the weight parameter is chosen to be 30.

In all simulations, the sample size is $T = 100$. The first set of simulations concerns inference about the mean of a scalar time series. In the "Gaussian AR(1)" design, the data are generated from a stationary Gaussian AR(1) model with coefficient ρ and unit innovation variance. The second set of simulations concerns inference about the coefficient on a scalar nonconstant regressor. In the "scalar nonconstant regressor" design, the regressions

$$R_t = \beta_0 + x_t\beta_1 + u_t,\ E[u_t|x_{t-1}, x_{t-2}, \ldots] = 0,\ t = 1, \ldots, T$$

contain a constant β_0, and the nonconstant regressor x_t and the regression disturbances u_t are independent zero mean Gaussian AR(1) processes with common coefficient ρ and unit innovation variance. Under the null, the coefficient β_1 is hypothesized to be zero.

Except for the three S_q tests, I compute the test statistics based on

$$\hat{y}_t = b'\hat{\Sigma}_X^{-1} X_t \hat{u}_t, \tag{22}$$

where $\hat{\Sigma}_X = T^{-1}\sum_{t=1}^{T} X_t X_t'$ with $X_t = (1, x_t)'$, $b = (0, 1)'$, and $\hat{u}_t = R_t - \hat{\beta}_0 - x_t\hat{\beta}_1$ with $(\hat{\beta}_0, \hat{\beta}_1)'$ being the ordinary least squares (OLS) estimator for (β_0, β_1). For the three S_q tests, I follow Section 5 in Müller (2014) to use

$$\tilde{y}_t = b'\hat{\Sigma}_X^{-1} X_t \hat{u}_t + \frac{b'\hat{\Sigma}_X^{-1} X_t X_t' \hat{\Sigma}_X^{-1} b}{b'\hat{\Sigma}_X^{-1} b}\hat{\beta}_1,$$

where b, X_t, $\hat{\Sigma}_X$, and $\hat{\beta}_1$ are the same as in (22).

Table 11 reports size and size-adjusted power of the 18 tests in the "Gaussian AR(1)" design. The size adjustment is performed on the ratio of the test statistic and the critical value to ensure that data-dependent critical values are appropriately subsumed in the effective test. Not surprisingly, the optimal EWC test almost exactly controls size in the data generating process (DGP) that coincides with The "uniformly minimal" function of the underlying smoothness class. This can be seen in the cases of OEWC$_3$ under $\rho = 0.95$ and OEWC$_8$ under $\rho = 0.7$. Moreover, though the class of OEWC$_q$ tests is known to essentially maximize weighted average power over μ under white noise, they also have better power performance relative to other tests when the underlying persistence is not negligible. For example, both the OEWC$_3$ test and the S_{24} control size under $\rho = 0.95$, but the size-adjusted power of OEWC$_3$ is 140% larger; the OEWC$_6$ test and the KVB test have roughly the same size distortions under $\rho = 0.95$, but the size-adjusted power of OEWC$_6$ is 36% larger.

In the "scalar nonconstant regressor" design, let $y_t = b'\Sigma_X^{-1} X_t u_t$, where Σ_X is the probability limit of $\hat{\Sigma}_X$ under suitable regularity conditions. The time series y_t is not Gaussian. On the other hand, the optimal EWC tests are based on the observable series $\hat{y}_t$ which, as argued by Müller (2014), behaves like $y_t - T^{-1}\sum_{s=1}^{T} y_s$ asymptotically. Despite the non-Gaussianity of the underlying time series, the optimal EWC tests, as reported in Table 12, continue to control size well and have better power performance relative to most alternative approaches. I note that the exceptional size and power performance of the IM_q test in Table 12 is specific to the design and explained by Müller.

6 Conclusion

This paper considers optimal HAR inference in finite-sample contexts. The driving assumption is that the normalized spectrum of the underlying time series lies in a smoothness class, which possesses a "uniformly minimal" function. Under this assumption, I establish a finite-sample optimal theory of HAR inference in the Gaussian location model. The optimal test trades off bias and variability, and requires an adjustment of the critical value to account for the maximum bias of the implied long-run variance estimator. I find that the EWC test is close to optimal, but one must make an adjustment to the usual Student-t critical value and choose an optimal q accordingly. Both the critical value adjustment and the choice of q depend on assumptions on the spectrum. The main implication of my findings is that when the goal is powerful HAR inference, it is advantageous to allow for bias in LRV estimation and adjust the critical value to explicitly account for the maximum bias.

Table 11: Small sample performance for inference about population mean

ρ	EWC_4	EWC_8	EWC_{12}	EWC_{24}	OEWC_3	OEWC_6	OEWC_8	OEWC_{10}	OEWC_{20}
				Panel A: Size under Gaussian AR(1)					
0.0	5.0	5.0	5.0	5.0	1.6	3.3	3.6	4.0	4.4
0.7	5.7	6.8	8.7	15.4	1.8	4.2	5.1	6.3	12.1
0.9	9.4	17.7	25.1	40.1	2.7	9.9	14.6	19.3	34.7
0.95	17.3	32.4	42.0	57.0	4.9	20.8	28.6	35.0	52.1
0.98	36.3	54.4	62.6	73.5	13.4	42.4	50.9	57.0	70.1
0.999	83.3	89.7	91.9	94.5	68.5	85.9	88.7	90.5	93.7
				Panel B: Size-adjusted power under Gaussian AR(1)					
0.0	34.0	41.9	45.0	47.9	29.0	38.6	41.9	43.9	47.3
0.7	34.6	42.9	45.4	48.8	29.5	39.5	42.9	44.8	47.9
0.9	36.6	44.5	46.9	48.9	31.4	41.7	44.5	45.8	48.5
0.95	39.5	47.1	49.2	51.0	34.3	44.9	47.1	48.5	50.8
0.98	49.1	56.5	58.6	60.3	43.2	54.3	56.5	57.8	60.0
0.999	99.7	99.9	99.9	100.0	99.3	99.9	99.9	99.9	100.0

ρ	S_{12}	S_{24}	S_{48}	$\hat{\omega}^2_{\text{A91}}$	$\hat{\omega}^2_{\text{AM}}$	$\hat{\omega}^2_{\text{KVB}}$	$\hat{\omega}^2_{\text{SPJ}}$	IM_8	IM_{16}
				Panel A: Size under Gaussian AR(1)					
0.0	5.0	4.8	5.0	5.9	6.0	5.1	5.0	4.9	5.0
0.7	5.1	5.0	5.3	13.0	8.6	7.4	6.1	7.6	12.2
0.9	5.1	5.3	5.9	24.9	15.1	12.8	8.5	17.6	31.5
0.95	5.0	5.0	5.7	38.3	22.7	19.8	11.6	31.1	48.3
0.98	4.7	4.7	5.3	59.6	37.8	34.2	18.6	52.3	67.0
0.999	4.7	4.9	5.4	91.6	84.1	79.4	53.8	89.0	93.0
				Panel B: Size-adjusted power under Gaussian AR(1)					
0.0	34.9	43.3	47.4	49.5	49.0	36.7	47.3	40.5	46.2
0.7	31.9	37.9	41.3	44.6	44.4	35.2	34.5	41.3	46.9
0.9	20.7	22.5	24.5	39.8	38.1	33.3	27.4	43.5	47.2
0.95	13.4	14.3	14.4	100.0	36.6	33.1	26.0	45.6	49.4
0.98	8.3	8.8	8.5	100.0	40.6	38.4	28.9	54.4	58.5
0.999	5.4	5.4	5.4	100.0	100.0	93.8	65.2	99.9	99.9

Note: Entries are rejection probability in percent of nominal 5% level tests. Under the alternative, the population mean differs from the hypothesized mean by $2T^{-1/2}(1-\rho)^{-1}$. Based on 100000 simulations.

Table 12: Small sample performance for inference about regression coefficient

ρ	EWC_4	EWC_8	EWC_{12}	EWC_{24}	OEWC_3	OEWC_6	OEWC_8	OEWC_{10}	OEWC_{20}
Panel A: Size under scalar nonconstant regressor									
0.0	5.4	5.5	5.4	5.5	1.7	3.5	4.0	4.3	4.8
0.7	6.3	6.9	7.2	8.0	2.0	4.5	5.1	5.8	6.9
0.9	9.0	10.9	12.3	17.1	3.0	7.2	8.6	9.9	14.4
0.95	12.4	16.0	19.2	28.0	4.2	10.9	13.2	15.5	23.9
0.98	18.3	25.0	30.7	42.8	7.0	17.6	21.8	25.5	38.1
0.999	28.9	39.9	46.9	58.6	12.2	30.3	36.4	41.4	54.5
Panel B: Size-adjusted power under scalar nonconstant regressor									
0.0	46.7	57.7	61.6	65.4	39.4	54.4	57.7	60.1	65.0
0.7	35.3	43.8	46.2	49.3	30.7	40.6	43.8	45.0	48.5
0.9	31.0	37.6	39.7	41.5	27.0	35.3	37.6	38.7	41.2
0.95	31.6	37.9	39.6	42.1	28.1	35.9	37.9	38.9	41.3
0.98	37.7	43.5	45.7	48.4	33.6	41.5	43.5	44.8	47.7
0.999	91.0	94.4	95.2	95.9	87.4	93.4	94.4	94.9	95.8

ρ	S_{12}	S_{24}	S_{48}	$\hat{\omega}^2_{\mathrm{A91}}$	$\hat{\omega}^2_{\mathrm{AM}}$	$\hat{\omega}^2_{\mathrm{KVB}}$	$\hat{\omega}^2_{\mathrm{SPJ}}$	IM_8	IM_{16}
Panel A: Size under scalar nonconstant regressor									
0.0	5.0	4.9	5.1	6.0	6.0	5.4	5.5	5.0	4.9
0.7	5.2	5.0	5.0	10.2	8.0	7.1	7.1	5.5	5.7
0.9	5.3	4.8	4.4	17.7	13.0	11.0	10.2	5.8	5.6
0.95	5.0	4.1	4.1	25.0	18.0	15.4	13.7	5.6	5.3
0.98	4.1	3.5	3.5	36.0	25.6	22.4	19.0	5.0	5.1
0.999	2.8	2.5	2.5	51.3	36.9	33.4	26.1	4.7	4.9
Panel B: Size-adjusted power under scalar nonconstant regressor									
0.0	46.5	58.4	64.2	68.2	68.0	50.8	66.6	53.1	54.2
0.7	35.1	44.0	48.4	49.7	48.8	37.9	44.9	44.4	52.1
0.9	29.7	35.1	37.9	100.0	37.6	32.7	33.9	52.9	73.9
0.95	27.3	31.5	32.8	100.0	35.3	32.5	31.1	70.3	91.6
0.98	25.5	29.2	29.6	100.0	38.8	36.7	33.7	93.5	99.7
0.999	32.0	37.0	37.0	100.0	89.5	87.8	80.3	100.0	100.0

Note: Entries are rejection probability in percent of nominal 5% level tests. Under the alternative, the population regression coefficient differs from the hypothesized mean by $2.5T^{-1/2}(1-\rho^2)^{-1/2}$. Based on 100000 simulations.

Appendix A Proofs in Section 3

Before proving Theorem 3.2, I make some additional notations. Let $\{Z_j\}_{j=0}^n$ be $n+1$ i.i.d. standard normal random variables throughout this section. For $\zeta \in \mathbb{R}_+^n$ and $n \geq 1$, let $J_n(\zeta) = P\left(Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_i\right)$ and define $K_n(\zeta) = P\left(Z_0^2 \leq \sum_{j=1}^n Z_j^2 \zeta_i\right)$. For a given $1 \leq \tilde{q} \leq T-1$, a given $\underline{f}$ that only satisfies Assumption 3.1(a), a given f_1^* (specifically referred to as $\kappa^{-1}\mathbf{1}[\phi \neq 0] + \mathbf{1}[\phi = 0]$ in Section 3), and $0 < \alpha < 1$, define $\mathrm{cv}_{\tilde{q}}$ such that

$$P\left((1-\mathrm{cv}_{\tilde{q}})Z_0^2 > \sum_{j=1}^{\tilde{q}} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right) = \alpha, \tag{23}$$

and a condition $cond_{\tilde{q}}$

$$\max_{j=1,2,\ldots,\tilde{q}} \left\{\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right\} < 0 \text{ or } \min_{j=1,2,\ldots,\tilde{q}} \left\{\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right\} > 0. \tag{24}$$

Further, for $\zeta, \xi \in \mathbb{R}_+^n$, we call ξ majorizes ζ ($\xi \succ \zeta$) if $\sum_{j=1}^n \xi_j = \sum_{j=1}^n \zeta_j$ and $\sum_{j=1}^k \xi_j \geq \sum_{j=1}^k \zeta_j$ for $1 \leq k < n$.

The following lemmas establish some useful properties of $J_n(\cdot)$, $K_n(\cdot)$, $\mathrm{cv}_{\tilde{q}}$, and $cond_{\tilde{q}}$. It is worth noting that if power is directed at a general function $\tilde{f}_1$ under the alternative, modified versions of the following auxiliary lemmas hold for $f_1^*(\phi) = \kappa^{-1}\tilde{f}_1(\phi)\mathbf{1}[\phi \neq 0] + \mathbf{1}[\phi = 0]$. In that case, Assumption 3.1 must be amended accordingly: (b) must be changed to $\underline{f}(\pi j/T)/\tilde{f}_1(\pi j/T) \geq \underline{f}(\pi(j+1)/T)/\tilde{f}_1(\pi(j+1)/T)$, $j = 0, 1, \ldots, T-2$, and for the existence of nontrivial tests we need $\underline{f}(\pi/T) > \kappa^{-1}\tilde{f}_1(\pi)$. I omit details for brevity and state the lemmas and the proof of Theorem 3.2 with f_1^* defined as in the main text.

A.1 Auxiliary lemmas

Lemma A.1 *For any $\zeta^* \in \mathbb{R}_+^n$, define $\mathcal{M}^+(\zeta^*) \equiv \{\zeta \in \mathbb{R}_+^n | \zeta_j \geq \zeta_j^* \ \forall\ 1 \leq j \leq n\}$ and $\mathcal{M}^-(\zeta^*) \equiv \{\zeta \in \mathbb{R}_+^n | \zeta_j \leq \zeta_j^* \ \forall\ 1 \leq j \leq n\}$. Then $J_n(\zeta^*) \geq J_n(\zeta)$ for any $\zeta \in \mathcal{M}^+(\zeta^*)$ and $J_n(\zeta^*) \leq J_n(\zeta)$ for any $\zeta \in \mathcal{M}^-(\zeta^*)$. Both equalities hold only if $\zeta = \zeta^*$.*

Proof. For any $\zeta \in \mathcal{M}^+(\zeta^*)$, the event $\{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\} \subset \{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j^*\}$. Thus $J_n(\zeta^*) \geq J_n(\zeta)$. In the case $\zeta_j > \zeta_j^*$ for some $1 \leq j \leq n$, $\{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\} \subsetneq \{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j^*\}$ and $J_n(\zeta^*) > J_n(\zeta)$. For any $\zeta \in \mathcal{M}^-(\zeta^*)$, the event $\{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\} \supset \{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j^*\}$. Thus $J_n(\zeta^*) \leq J_n(\zeta)$. In the case $\zeta_j < \zeta_j^*$ for some $1 \leq j \leq n$, $\{Z_0^2 \geq \sum_{j=1}^n Z_i^2 \zeta_i\} \supsetneq \{Z_0^2 \geq \sum_{j=1}^n Z_i^2 \zeta_i^*\}$ and $J_n(\zeta^*) < J_n(\zeta)$. ■

Corollary A.2 *For any* $\zeta^* \in \mathbb{R}_+^n$, $\mathcal{M}^+(\zeta^*)$ *and* $\mathcal{M}^-(\zeta^*)$ *are as defined in Lemma A.1. Then* $K_n(\zeta^*) \leq K_n(\zeta)$ *for any* $\zeta \in \mathcal{M}^+(\zeta^*)$ *and* $K_n(\zeta^*) \geq K_n(\zeta)$ *for any* $\zeta \in \mathcal{M}^-(\zeta^*)$. *Both equalities hold only if* $\zeta = \zeta^*$.

Proof. Simply note that $K_n(\zeta) = 1 - J_n(\zeta)$. The conclusions then follow from Lemma A.1. ■

Remark A.3

i) A corrected version of Remark 4 in Bakirov (1996) can lead to the same set of conclusions of Lemma A.1 with a different relationship ($\xi \succ \zeta$ *or* $\zeta \succ \xi$). *The exact statement in that article is not quite right; i.e., the stated relationship* $J_n(\xi) \geq J_n(\zeta)$ *is not necessarily true under weak majorizations. A trivial counterexample is that* $\xi = (2, 0, 0, \ldots, 0) \succ_w \zeta = (0, 0, 0, \ldots, 0)$ *but* $J(\lambda) < J(\mu) = 1$. *The corrected statement is* $J_n(\xi) \geq J_n(\zeta)$ *if* $\xi \succ \zeta$ *and* $\xi_j > 0$, $\zeta_j > 0$ *for all* j. *This is proved by making use of the (Schur) convexity of* $J_n(\cdot)$ *and invoking Caramata inequality.*

ii) The discussion in (i) and Lemma A.1 provide somewhat complementary sufficient conditions to explore the possible monotone properties of $J_n(\cdot)$ *over* $\mathbb{R}_+^n$. *Unfortunately, they have not fully characterized the necessary conditions of* $J_n(\xi) \geq J_n(\zeta)$. *Under the conditions* $(\bar{\xi} - \bar{\zeta})(1, 1, \ldots, 1) + \zeta \succ \xi$ *and* $\sum_{j=1}^n \xi_i \geq \sum_{j=1}^n \zeta_i$ *where* $\bar{\xi} = \sum_{j=1}^n \xi_j / n$ *and* $\bar{\zeta} = \sum_{j=1}^n \zeta_j / n$, *it is less obvious to compare* $J(\xi)$ *and* $J(\zeta)$ *unless* ζ *lies at the boundary.*

Lemma A.4 *For* $\zeta \in \mathbb{R}_+^n$ *and* $n \geq 1$, $J_n(\zeta) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(n-1)/2} \mathrm{d}u}{\sqrt{\prod_{j=1}^n (1-u^2+\zeta_j)}}$ *and* $J_1(\zeta_1) = \frac{2}{\pi} \arcsin \frac{1}{\sqrt{1+\zeta_1}}$.

Proof.

$$J_n(\zeta) = \frac{1}{\pi} \int_0^\infty \frac{dt}{\sqrt{t}(1+t)\sqrt{\prod_{j=1}^n (1 + (1+t)\zeta_j)}} = \frac{1}{\pi} \int_0^1 \frac{s^{(n-1)/2} ds}{\sqrt{(1-s) \prod_{j=1}^n (s + \zeta_j)}}$$

$$= \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(n-1)/2} du}{\sqrt{\prod_{j=1}^n (1 - u^2 + \zeta_j)}}, \tag{25}$$

where the first equality follows from Lemma 2 of Bakirov and Székely (2006), the second equality follows by a change of variable $s = -1/(1+t)$, and (25) follows by another change of variable $u = \sqrt{1-s}$.

$$J_1(\zeta_1) = \frac{2}{\pi} \int_0^1 \frac{du}{\sqrt{(1 - u^2 + \zeta_1)}} = \frac{2}{\pi} \arcsin \frac{1}{\sqrt{1+\zeta_1}},$$

which follows from a change of variable $v = u/\sqrt{1+\zeta_1}$ and the fact that the antiderivative of $(1-v^2)^{-1/2}$ is $\arcsin v$. ■

Lemma A.5 *For* $0 < \alpha < 1$,

(a) cv_1 *exists if and only if* $\underline{f}(\pi/T) \neq \kappa^{-1}$. $\underline{f}(\pi/T) \lesseqgtr \kappa^{-1}$ *if and only if* $1 \lesseqgtr \mathrm{cv}_1$.

(b) $cond_1$ *holds if* $\underline{f}(\pi/T) \neq \kappa^{-1}$.

Proof.

(a) In (23) at $\tilde{q} = 1$, if $\underline{f}(\pi/T) = \kappa^{-1}$, there does not exist a cv_1 and α such that (23) holds. On the other hand, a rearrangement of the event in (23) at $\tilde{q} = 1$ gives

$$P\left(Z_0^2 + Z_1^2 > \left(Z_0^2 + \kappa \underline{f}(\pi/T) Z_1^2\right) \mathrm{cv}_1\right) = \alpha.$$

It follows that $\underline{f}(\pi/T) \lesseqgtr \kappa^{-1}$ if and only if $1 \lesseqgtr cv_1$. Moreover, if $\underline{f}(\pi/T) > \kappa^{-1}$, the second part of Lemma A.4 in conjunction with (23) at $\tilde{q} = 1$ gives

$$cv_1 = \frac{1}{\kappa \underline{f}(\pi/T) \sin^2(\alpha\pi/2) + \cos^2(\alpha\pi/2)},$$

which always exists for every $0 < \alpha < 1$. In a similar vein, if $\underline{f}(\pi/T) < \kappa^{-1}$, we have

$$cv_1 = \frac{1}{\kappa \underline{f}(\pi/T) \cos^2(\alpha\pi/2) + \sin^2(\alpha\pi/2)},$$

which always exists. Thus, cv_1 exists if and only if $\underline{f}(\pi/T) \neq \kappa^{-1}$.

(b) It follows from the above that $cond_1$ holds if $\underline{f}(\pi/T) \neq \kappa^{-1}$.

■

Lemma A.6 *For* $\underline{f}(\pi/T) \neq \kappa^{-1}$ *and* $0 < \alpha < 1$, *if* $cond_{\tilde{q}}$ *is violated for some* $1 < \tilde{q} \leq T - 2$, *then* $cond_q$ *is also violated for any* $\tilde{q} + 1 \leq q \leq T - 1$.

Proof. Suppose $cond_{\tilde{q}}$ is violated while $cond_{\tilde{q}+1}$ holds. We have

$$\max_{j=1,2,\ldots,\tilde{q}} \left\{\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right\} \geq 0 \text{ and } \min_{j=1,2,\ldots,\tilde{q}} \left\{\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right\} \leq 0. \tag{26}$$

Case 1. Consider $\max_{j=1,2,\ldots,\tilde{q}+1} \left\{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right\} < 0$. We must have $\mathrm{cv}_{\tilde{q}+1} > 1$; otherwise, (23) does not hold at $\tilde{q} + 1$. On the other hand, $0 > \max_{j=1,2,\ldots,\tilde{q}+1} \left\{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right\} \geq \max_{j=1,2,\ldots,\tilde{q}} \left\{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right\}$. This, in conjunction with the first part of (26), implies that $\mathrm{cv}_{\tilde{q}} > \mathrm{cv}_{\tilde{q}+1} > 1$, which we now show is impossible. Suppose $\mathrm{cv}_{\tilde{q}} > \mathrm{cv}_{\tilde{q}+1} > 1$ is true. Denote

$A_{\tilde{q}}^- = \{j | 1 \le j \le \tilde{q},\ \mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1 < 0\}$ ($A_{\tilde{q}}^- \neq \emptyset$; otherwise, (23) does not hold for $\tilde{q}$). Now (23) at $\tilde{q}$ gives

$$\alpha = P\left((1-\mathrm{cv}_{\tilde{q}}) Z_0^2 > \sum_{j=1}^{\tilde{q}} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right)$$

$$= P\left(Z_0^2 < \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j=1}^{\tilde{q}} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right)$$

$$= P\left(Z_0^2 < \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^-} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 + \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j \notin A_{\tilde{q}}^-} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right)$$

$$\le P\left(Z_0^2 < \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^-} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right) \tag{27}$$

$$< P\left(Z_0^2 < \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^-} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right) \tag{28}$$

$$\le P\left(Z_0^2 < \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^-} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 + \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j \notin A_{\tilde{q}}^-} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right) \tag{29}$$

$$\le P\left(Z_0^2 < \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j=1}^{\tilde{q}+1} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right) = \alpha,$$

where (27) is due to the fact that $P(A < C - B) \le P(A < C)$ when A, B, C are independent random variables and $B \ge 0$ almost surely. The inequality (28) is due to Corollary A.2 and the fact that for any $j \in A_{\tilde{q}}^-$, $\frac{1}{1-\mathrm{cv}_{\tilde{q}}}\left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1\right] < \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}}\left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right]$ under $\mathrm{cv}_{\tilde{q}} > \mathrm{cv}_{\tilde{q}+1} > 1$. The inequality (29) is due to Corollary A.2.

Case 2. Consider $\min_{j=1,2,\ldots,\tilde{q}+1}\left\{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right\} > 0$. We must have $\mathrm{cv}_{\tilde{q}+1} < 1$; otherwise, (23) does not hold at $\tilde{q}+1$. On the other hand, $0 < \min_{j=1,2,\ldots,\tilde{q}+1}\left\{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right\} \le \min_{j=1,2,\ldots,\tilde{q}}\left\{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{f}(\pi j/T) - 1\right\}$. This, in conjunction with the second part of (26), implies that $\mathrm{cv}_{\tilde{q}} < \mathrm{cv}_{\tilde{q}+1} < 1$, which we next show is impossible. Suppose $\mathrm{cv}_{\tilde{q}} < \mathrm{cv}_{\tilde{q}+1} < 1$ is true. Denote $A_{\tilde{q}}^+ = \{j | 1 \le j \le \tilde{q},\ \mathrm{cv}_{\tilde{q}}\, \kappa \underline{f}(\pi j/T) - 1 > 0\}$ ($A_{\tilde{q}}^+ \neq \emptyset$; otherwise, (23) is violated for $\tilde{q}$). Now (23) at $\tilde{q}$ gives

$$\alpha = P\left((1-\mathrm{cv}_{\tilde{q}})Z_0^2 > \sum_{j=1}^{\tilde{q}} \left[\mathrm{cv}_{\tilde{q}}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right)$$

$$= P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j=1}^{\tilde{q}} \left[\mathrm{cv}_{\tilde{q}}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right)$$

$$= P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j\in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 + \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j\notin A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right)$$

$$\geq P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j\in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right) \tag{30}$$

$$> P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j\in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}+1}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right) \tag{31}$$

$$\geq P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j\in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}+1}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 + \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j\notin A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}+1}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right) \tag{32}$$

$$\geq P\left(Z_0^2 < \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j=1}^{\tilde{q}+1} \left[\mathrm{cv}_{\tilde{q}+1}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right) = \alpha,$$

where (30) is due to the fact that $P(A \geq C+B) \geq P(A \geq C)$ when A, B, C are independent random variables and $B \leq 0$ almost surely. The inequality (31) is due to Lemma A.1 and the fact that for any $j \in A_{\tilde{q}}^+$, $\frac{1}{1-\mathrm{cv}_{\tilde{q}}}\left[\mathrm{cv}_{\tilde{q}}\,\kappa \underline{f}(\pi j/T) - 1\right] < \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}}\left[\mathrm{cv}_{\tilde{q}+1}\,\kappa \underline{f}(\pi j/T) - 1\right]$ under $\mathrm{cv}_{\tilde{q}} < \mathrm{cv}_{\tilde{q}+1} < 1$. The inequality (32) is due to Lemma A.1.

Collecting the results from **Case 1** and **Case 2**, we know that it is impossible to have $cond_{\tilde{q}+1}$ hold, given that $cond_{\tilde{q}}$ does not. By induction, we have for $\underline{f}(\pi/T) \neq \kappa$ and $0 < \alpha < 1$, if $cond_{\tilde{q}}$ is violated for some $1 < \tilde{q} \leq T-2$, then $cond_q$ is also violated for any $\tilde{q}+1 \leq q \leq T-1$. ■

Corollary A.7 *For $\underline{f}(\pi/T) \neq \kappa^{-1}$ and $0 < \alpha < 1$, if $cond_{\tilde{q}}$ holds for some $3 \leq \tilde{q} \leq T-1$, then $cond_q$ also holds for any $2 \leq q \leq \tilde{q}-1$.*

Proof. This is the contrapositive statement of Lemma A.6. ■

Corollary A.8 *For $\underline{f}(\pi/T) \neq \kappa^{-1}$ and $0 < \alpha < 1$, either one of the following will hold:*

(a) there exists a unique $1 \leq q^* \leq T-2$ such that $cond_q$ is satisfied for all $1 \leq q \leq q^*$ and violated for all $q^*+1 \leq q \leq T-1$;

(b) $cond_q$ *is satisfied for all* $1 \le q \le T-1$. *In this case, define* $q^* = T-1$.

Proof. If $cond_{T-1}$ holds, by Corollary A.7 (b) is true. Otherwise if $cond_{T-2}$ holds, then (a) is true with $q^* = T-2$. Otherwise, given that $cond_1$ always holds by Lemma A.5, backward inductions lead (a) to be true for a unique $1 \le q^* \le T-3$. ■

Corollary A.9 *If* $\underline{f}$ *in Assumption 3.1(a) equals* 1 *and* $0 < \alpha < 1$, *then* $cond_q$ *holds for all* $1 \le q \le T-1$.

Proof. $\underline{f}(\pi j/T) = 1,\ j = 1, 2, \ldots, T-1$ reduces $cond_{T-1}$ to $\kappa\,\mathrm{cv}_{T-1} - 1 < 0$ or $\kappa\,\mathrm{cv}_{T-1} - 1 > 0$, which must be true for (23) to hold at $\tilde{q} = T-1$. It follows that $cond_q$ holds for all $1 \le q \le T-1$ by Corollary A.7. ■

Lemma A.10 *For* $\underline{f}(\pi/T) \ne \kappa^{-1}$, $0 < \alpha < 1$, *and* q^* *as defined in Corollary A.8, either one of the following will hold:*

(a) $\mathrm{cv}_q > 1$ *for all* $1 \le q \le q^*$, *and if* $q^* \ge 2$, $\mathrm{cv}_{q+1} > \mathrm{cv}_q,\ q = 1, 2, \ldots, q^*-1$;

(b) $\mathrm{cv}_q < 1$ *for all* $1 \le q \le q^*$, *and if* $q^* \ge 2$, $\mathrm{cv}_{q+1} < \mathrm{cv}_q,\ q = 1, 2, \ldots, q^*-1$.

Proof. Lemma A.5 leads to the conclusions for $q^* = 1$. We now focus on $q^* \ge 2$.

Case 1. Suppose $\kappa\underline{f}(\pi/T) < 1$, then Lemma A.5 implies that $\mathrm{cv}_1 > 1$. Suppose there exists a $\tilde{q} = \min\{q | 2 \le q \le q^*,\ \mathrm{cv}_q < 1\}$. (We note cv_q cannot be 1 for any $q \le q^*$; otherwise, (23) cannot hold at the corresponding q.) Then we must have

$$\min_{j=1,2,\ldots,\tilde{q}-1}\{\mathrm{cv}_{\tilde{q}-1}\,\kappa\underline{f}(\pi j/T) - 1\} > \min_{j=1,2,\ldots,\tilde{q}-1}\{\mathrm{cv}_{\tilde{q}}\,\kappa\underline{f}(\pi j/T) - 1\} \ge \min_{j=1,2,\ldots,\tilde{q}}\{\mathrm{cv}_{\tilde{q}}\,\kappa\underline{f}(\pi j/T) - 1\} > 0.$$

There is a contradiction, because $\max_{j=1,2,\ldots,\tilde{q}-1}\{\mathrm{cv}_{\tilde{q}-1}\,\kappa\underline{f}(\pi j/T) - 1\} < 0$. This subsequently implies that $\mathrm{cv}_q > 1$ for all $1 \le q \le q^*$. Moreover, for each $j = 1, \ldots, q^*$, $R_j(x) = \left[\kappa\underline{f}(\pi j/T)x - 1\right]/(1-x)$ is monotonically decreasing in $(1, \infty)$. For (23) to hold sequentially, we necessarily need $\mathrm{cv}_{q+1} > \mathrm{cv}_q,\ q = 1, 2, \ldots, q^*-1$. (Otherwise, the LHS of (23) would always exceed α by Lemma A.2.)

Case 2. Suppose $\kappa\underline{f}(\pi/T) > 1$, then Lemma A.5 implies that $\mathrm{cv}_1 < 1$. Suppose there exists a $\tilde{q} = \min\{q | 2 \le q \le q^*,\ \mathrm{cv}_q > 1\}$. (We note cv_q cannot be 1 for any $q \le q^*$; otherwise, (23) cannot hold at the corresponding q.) Then we must have

$$\max_{j=1,2,\ldots,\tilde{q}-1}\{\mathrm{cv}_{\tilde{q}-1}\,\kappa\underline{f}(\pi j/T) - 1\} < \max_{j=1,2,\ldots,\tilde{q}-1}\{\mathrm{cv}_{\tilde{q}}\,\kappa\underline{f}(\pi j/T) - 1\} \le \max_{j=1,2,\ldots,\tilde{q}}\{\mathrm{cv}_{\tilde{q}}\,\kappa\underline{f}(\pi j/T) - 1\} < 0.$$

There is a contradiction, because $\min_{j=1,2,\ldots,\tilde{q}-1}\{\text{cv}_{\tilde{q}-1}\,\kappa\underline{f}(\pi j/T)-1\}>0$. This subsequently implies that $\text{cv}_q<1$ for all $1\leq q\leq q^*$. Moreover, for each $j=1,\ldots,q^*$, $R_j(x)=\left[\kappa\underline{f}(\pi j/T)x-1\right]/(1-x)$ is monotonically increasing in $(0,1)$. (To see this, note that $\kappa\underline{f}(\pi j/T)-1>\kappa\underline{f}(\pi j/T)\,\text{cv}_{q^*}-1>0$ for every $1\leq j\leq q^*$.) For (23) to hold sequentially, we necessarily need $\text{cv}_{q+1}<\text{cv}_q$, $q=1,2,\ldots,q^*-1$. (Otherwise, the LHS of (23) would always be below α by Lemma A.1.) ■

Lemma A.11 *For $\underline{f}(\pi/T)\neq\kappa^{-1}$, $0<\alpha<1$, and q^* as defined in Corollary A.8, if additionally $\underline{f}(\pi j/T)\geq\underline{f}(\pi(j+1)/T)$, $j=0,1,\ldots,T-2$ and $q^*<T-1$, then $\kappa^{-1}\left(\underline{f}(\pi j/T)\right)^{-1}\geq\text{cv}_{q^*}$.*

Proof. By Assumption 3.1(b), it suffices to show $\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}\geq\text{cv}_{q^*}$. Suppose not; then $\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}<cv_{q^*}$. Define $Q(x,q)=P\left((1-x)Z_0^2>\sum_{j=1}^{q}\left[x\kappa\underline{f}(\pi j/T)-1\right]Z_j^2\right)$.

Case 1. Suppose $\kappa\underline{f}(\pi/T)<1$, then $\text{cv}_{q^*}>1$ by Lemma A.10. Moreover, by Assumption 3.1(b) we must have $\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}>1$.

$$
\begin{aligned}
&Q(\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1},q^*+1)\\
=&P\left(\left[1-\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}\right]Z_0^2>\sum_{j=1}^{q^*+1}\left[\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}\underline{f}(\pi j/T)-1\right]Z_j^2\right)\\
=&P\left(\left[1-\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}\right]Z_0^2>\sum_{j=1}^{q^*}\left[\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}\underline{f}(\pi j/T)-1\right]Z_j^2\right)\\
=&0<\alpha.
\end{aligned}
$$

Pick ε such that $1<\text{cv}_{q^*}-\varepsilon<\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}<\text{cv}_{q^*}$. Then

$$
\begin{aligned}
&Q(\text{cv}_{q^*}-\varepsilon,q^*+1)\\
=&P\left(\left[1-(\text{cv}_{q^*}-\varepsilon)\right]Z_0^2>\sum_{j=1}^{q^*+1}\left[(\text{cv}_{q^*}-\varepsilon)\kappa\underline{f}(\pi j/T)-1\right]Z_j^2\right)\\
>&P\left(\left[1-(\text{cv}_{q^*}-\varepsilon)\right]Z_0^2>\sum_{j=1}^{q^*}\left[(\text{cv}_{q^*}-\varepsilon)\kappa\underline{f}(\pi j/T)-1\right]Z_j^2\right)\\
=&Q(\text{cv}_{q^*}-\varepsilon,q^*)>Q(\text{cv}_{q^*},q^*)=\alpha,
\end{aligned}
$$

where the last but one inequality follows from the fact that $Q(\cdot,q^*)$ is monotonically decreasing in $(1,\text{cv}_{q^*})$ under $\kappa\underline{f}(\pi/T)<1$.

By the continuity of $Q(\cdot,q^*+1)$ in $(\text{cv}_{q^*}-\varepsilon,\kappa^{-1}\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1})$ and the intermediate value theorem, there must exist a number, denoted by cv_{q^*+1}, such that $Q(\text{cv}_{q^*+1},q^*+1)=\alpha$. There is a contradiction, because $cond_{q^*+1}$ now holds, violating Corollary A.8.

Case 2. Suppose $\kappa \underline{f}(\pi/T) > 1$, then $\mathrm{cv}_{q^*} < 1$ by Lemma A.10.

$$Q(\mathrm{cv}_{q^*}, q^*+1) = P\left(\left[1 - \mathrm{cv}_{q^*}\right] Z_0^2 > \sum_{j=1}^{q^*+1} \left[cv_{q^*} \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right)$$
$$< P\left(\left[1 - \mathrm{cv}_{q^*}\right] Z_0^2 > \sum_{j=1}^{q^*} \left[cv_{q^*} \kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 \right) = Q(cv_{q^*}, q^*) = \alpha.$$

On the other hand, $0 < \kappa^{-1} \left(\underline{f}(\pi(q^*+1)/T)\right)^{-1} < cv_{q^*} < 1$. Then

$$\begin{aligned}
&Q(\kappa^{-1} \left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}, q^*+1) \\
=&P\left(\left[1 - \kappa^{-1} \left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}\right] Z_0^2 > \sum_{j=1}^{q^*+1} \left[\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1} \underline{f}(\pi j/T) - 1\right] Z_j^2 \right) \\
=&P\left(\left[1 - \kappa^{-1} \left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}\right] Z_0^2 > \sum_{j=1}^{q^*} \left[\left(\underline{f}(\pi(q^*+1)/T)\right)^{-1} \underline{f}(\pi j/T) - 1\right] Z_j^2 \right) \\
=&Q(\kappa^{-1} \left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}, q^*) > Q(\mathrm{cv}_{q^*}, q^*) = \alpha,
\end{aligned}$$

where the last but one inequality follows from the fact that $Q(\cdot, q^*)$ is monotonically decreasing in $(0, cv_{q^*})$ under $\kappa \underline{f}(\pi/T) > 1$.

By the continuity of $Q(\kappa^{-1} \left(\underline{f}(\pi(q^*+1)/T)\right)^{-1}, cv_{q^*})$ and the intermediate value theorem , there must exist a number, denoted by cv_{q^*+1}, such that $Q(\mathrm{cv}_{q^*+1}, q^*+1) = \alpha$. There is a contradiction, because $cond_{q^*+1}$ now holds, violating Corollary A.8.

Both **Case 1** and **Case 2** imply that we must have $\kappa^{-1} \left(\underline{f}(\pi(q^*+1)/T)\right)^{-1} \geq cv_{q^*}$, and thus $\min_{q^*+1 \leq j \leq T-1} \kappa^{-1} \left(\underline{f}(\pi j/T)\right)^{-1} \geq cv_{q^*}$ if $q^* < T-1$. ■

A.2 Proof of Theorem 3.2

Part 1 of Theorem 3.2 is already proved in the main text. I now focus on proving part 2 of Theorem 3.2, in which $\underline{f}(\pi/T) > \kappa^{-1}$. Under Assumption 3.1(a) and for $0 < \alpha < 1$, Corollary A.8 shows that there exists a unique q^* such that *either* (i) $cond_q$ holds for $1 \leq q \leq q^*$ and is violated for $q^*+1 \leq q \leq T-1$, *or* (ii) $cond_q$ holds for all $1 \leq q \leq T-1$, where we define $q^* = T-1$. We conjecture that the least favorable distribution Λ^* of (9) puts a point mass on $f^*(\phi) = \underline{f}(\phi)\mathbf{1}[0 \leq \phi \leq \pi q^*/T] + (\kappa\, \mathrm{cv}_{q^*})^{-1}\mathbf{1}[\phi > \pi q^*/T]$. By Lemma A.11 and Assumption 3.1(c), the kinked function f^* is known to be in $\mathcal{F}$.

Using the same notation as in the main text, the best level α test of H^d_{0,Λ^*} against H^d_{1,f_1} is

$$\varphi_{\Lambda^*,f_1} = 1\left[\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \mathrm{cv}\right],$$

for some $\mathrm{cv} \geq 0$ such that $E_{P_{Y,f^*}}[\varphi_{\Lambda^*,f_1}(Y^s)] = \alpha$, where $P_{Y,\tilde{f}}$ denotes the joint distribution of Y at $f = \tilde{f}$ under H_0^d. It follows that

$$\begin{aligned}
\alpha &= P_{Y,f^*}\left(\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \mathrm{cv}\right) \\
&= P_{Y,f^*}\left(Y_0^2 + \sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T) > \mathrm{cv}\left(Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2\right)\right) \\
&= P\left((1-\mathrm{cv})\, Z_0^2 > \sum_{j=1}^{T-1} \left[\mathrm{cv}\,\kappa f^*(\pi j/T) - 1\right] Z_j^2\right) \\
&= P\left((1-\mathrm{cv})\, Z_0^2 > \sum_{j=1}^{q^*} \left[\mathrm{cv}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2 + \sum_{j=q^*+1}^{T-1} \left[\mathrm{cv}\,/\,\mathrm{cv}_{q^*} - 1\right] Z_j^2\right), \qquad (33)
\end{aligned}$$

where the last equality follows from the definition of f^*. Because Y is a continuous random vector, the critical value cv is unique. By matching (33) with (23) at $\tilde{q} = q^*$, we have $\mathrm{cv} = \mathrm{cv}_{q^*}$. Moreover, the events $\left\{\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \mathrm{cv}_{q^*}\right\}$ and $\left\{\frac{Y_0^2 + \sum_{j=1}^{q^*} Y_j^2/\underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2} > \mathrm{cv}_{q^*}\right\}$ are equivalent $P_{Y,\tilde{f}}$-almost surely, and uniformly in $\tilde{f} \in \mathcal{F} \cup \{f_1^*\}$. This then leads to the optimal test statistic in (10).

It remains to check that φ_{Λ^*,f_1} controls size under H_0^d, i.e., $\sup_{\tilde{f}\in\mathcal{F}} E_{P_{Y,\tilde{f}}}[\varphi_{\Lambda^*,f_1}(Y)] \leq \alpha$. For a given $\tilde{f} \in \mathcal{F}$,

$$\begin{aligned}
E_{P_{Y,\tilde{f}}}[\varphi_{\Lambda^*,f_1}(Y)] &= P_{Y,\tilde{f}}\left(\frac{Y_0^2 + \sum_{j=1}^{q^*} Y_j^2/\underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2} > \mathrm{cv}_{q^*}\right) \\
&= P\left([1-\mathrm{cv}_{q^*}]\, Z_0^2 > \sum_{j=1}^{q^*} \left[\mathrm{cv}_{q^*}\,\kappa \underline{f}(\pi j/T) - 1\right] \frac{\tilde{f}(\pi j/T)}{\underline{f}(\pi j/T)} Z_j^2\right) \\
&= P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{q^*}} \sum_{j=1}^{q^*} \left[\mathrm{cv}_{q^*}\,\kappa \underline{f}(\pi j/T) - 1\right] \frac{\tilde{f}(\pi j/T)}{\underline{f}(\pi j/T)} Z_j^2\right) \qquad (34) \\
&\leq P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{q^*}} \sum_{j=1}^{q^*} \left[\mathrm{cv}_{q^*}\,\kappa \underline{f}(\pi j/T) - 1\right] Z_j^2\right) = \alpha,
\end{aligned}$$

where (34) follows from (b) in Lemma A.10 under the condition $\underline{f}(\pi/T) > \kappa^{-1}$, and the inequality follows from the definition of q^* and Lemma A.1 under Assumption 3.1(a).

In sum, we have shown that Λ^* is indeed the least favorable distribution in testing H_0^d against H_{1,f_1}^d, under Assumption 3.1 and $\underline{f}(\pi/T) > \kappa^{-1}$. Further, the statistic (10) is the best test statistic, and the null rejection probability of this test is maximized at $f = \underline{f}$.

Appendix B Computational Details in Section 3

B.1 An algorithm to compute q^* in Theorem 3.2

In Algorithm 1 on the following page, I first use the bisection method three times to zoom into a small set of consecutive integers and then locate q^* by a one-dimensional greedy search. One might be concerned about the sensitivity of the numerically determined q^* to the precision of empirically estimated critical values for different q. As will be explained in the next subsection, this is not an issue in implementations with $B = 50,000$.

B.2 Computational details for cv_{q^*} in Theorem 3.2

Given the q^* delivered via Algorithm 1, I numerically invert (13) by the standard bisection method. Since the integrand in the last integral in (13) is smooth for typical values of cv_{q^*}, I use a 50-point Gaussian quadrature to compute the numerical integration. It suffices for my purpose with numerical errors of the order 1e $-$ 6. Moreover, to check whether the determination of q^* in Algorithm 1 is sensitive to simulation errors in estimating $\text{cv}_{\widetilde{q}}$, I append a sanity check to Algorithm 1. In particular, I substitute the more precisely determined $\text{cv}_{q^*}, \text{cv}_{q^*-1}, \text{cv}_{q^*-2}$ back into $cond_{q^*}, cond_{q^*-1}, cond_{q^*-2}$ and check whether any of them are inconsistent with the definition of q^*. It appears that having $50,000$ realizations of $\{Z_j\}_{j=0}^{T-1}$ suffices for our purpose.

Appendix C More Tables in the Diagonal Model

I follow the procedure of producing Tables 1 and 2 to provide more rule of thumb tables for other types of smoothness classes. See Tables 13 and 14 for the rule of thumb under the type of $\mathcal{F}$, in which the first derivative of $\log(f)$ is bounded by the constant C. See Tables 15 and 16 for the rule of thumb under the type of $\mathcal{F}$, in which the second derivative of $\log(f)$ is bounded by the constant C.

Algorithm 1: Bisection method to produce q^* in Theorem 3.2

Input: $\mathcal{F}$ with a well-defined $\underline{f}$, T, $\alpha = 0.05$ by default, $\kappa = 11$ by default,

$\left\{\{z_j^b\}_{j=0}^{T-1}\right\}_{b=1}^{B} \leftarrow B$ realizations of i.i.d. standard normals for a sufficiently large B.

Output: q^*

Initialization: $\widetilde{q} \leftarrow T-1$, $\hat{\text{cv}}_{\widetilde{q}} \leftarrow \inf\{r \in \mathbb{R}_+ | \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left[\frac{\sum_{j=0}^{\widetilde{q}}(z_j^b)^2}{(z_0^b)^2+\kappa\sum_{j=1}^{\widetilde{q}}(z_j^b)^2\underline{f}(\pi j/T)} > r\right] \geq \alpha\}$;

if $cond_{\widetilde{q}}$ *is true with* $\text{cv}_{\widetilde{q}}$ *replaced by* $\hat{\text{cv}}_{\widetilde{q}}$ **then**

  $q^* \leftarrow \widetilde{q}$;

else

  while $cond_{\widetilde{q}}$ *is false with* $\text{cv}_{\widetilde{q}}$ *replaced by* $\hat{\text{cv}}_{\widetilde{q}}$ **do**

    $\widetilde{q} \leftarrow \left\lfloor \frac{1+\widetilde{q}}{2} \right\rfloor$;

  end

  $\widetilde{\widetilde{q}} \leftarrow \left\lfloor \frac{3\widetilde{q}-1}{2} \right\rfloor$;

  while $cond_{\widetilde{\widetilde{q}}}$ *is false with* $\text{cv}_{\widetilde{\widetilde{q}}}$ *replaced by* $\hat{\text{cv}}_{\widetilde{\widetilde{q}}}$ **do**

    $\widetilde{\widetilde{q}} \leftarrow \left\lfloor \frac{\widetilde{\widetilde{q}}+\widetilde{q}}{2} \right\rfloor$;

  end

  for $q \leftarrow \widetilde{\widetilde{q}}$ **to** $2\widetilde{\widetilde{q}} - \widetilde{q} + 1$ **by** 1 **do**

    if $cond_{q+1}$ *is false with* cv_{q+1} *replaced by* $\hat{\text{cv}}_{q+1}$ **then**

      $q^* \leftarrow q$;

      break ;

    end

  end

end

Table 13: Optimal q and adjustment factor of the Student-t critical value of level α EWC test.

C	10.0	5.6	3.2	1.8	1.0	0.6	0.3	0.2	0.1
$\alpha = 0.01$	$(5, 1.59)$	$(6, 1.35)$	$(9, 1.27)$	$(11, 1.18)$	$(15, 1.13)$	$(20, 1.10)$	$(26, 1.07)$	$(34, 1.05)$	$(45, 1.04)$
$\alpha = 0.05$	$(4, 1.47)$	$(5, 1.30)$	$(7, 1.21)$	$(9, 1.15)$	$(11, 1.10)$	$(15, 1.07)$	$(20, 1.05)$	$(26, 1.04)$	$(34, 1.03)$
$\alpha = 0.10$	$(3, 1.36)$	$(4, 1.24)$	$(6, 1.19)$	$(7, 1.12)$	$(10, 1.09)$	$(12, 1.06)$	$(16, 1.04)$	$(22, 1.03)$	$(29, 1.02)$

Notes: Based on a series of smoothness classes $\mathcal{F}$, in which the "uniformly minimal" function is $\underline{f}(\phi) = \exp(-C\phi)$. Sample size $T = 100$. C is log-spaced between 10 and 0.1.

Table 14: Rule of thumb for adjustment factor of the Student-t critical value of level α EWC test.

q	4	6	8	9	10	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.01$	1.53	1.35	1.26	1.27	1.19	1.18	1.16	1.15	1.14	1.13	1.12	1.12	1.11	1.10	1.10
	(11.0)	(5.6)	(3.3)	(3.2)	(2.1)	(1.8)	(1.5)	(1.3)	(1.1)	(1.0)	(0.9)	(0.8)	(0.7)	(0.62)	(0.6)
$\alpha = 0.05$	1.47	1.30	1.19	1.21	1.11	1.10	1.09	1.09	1.08	1.07	1.07	1.06	1.06	1.05	1.05
	(10.0)	(5.6)	(3.3)	(3.2)	(1.2)	(1.0)	(0.9)	(0.8)	(0.7)	(0.6)	(0.5)	(0.4)	(0.38)	(0.35)	(0.3)
$\alpha = 0.10$	1.24	1.19	1.11	1.10	1.09	1.07	1.06	1.06	1.05	1.05	1.04	1.04	1.04	1.04	1.03
	(5.6)	(3.2)	(1.5)	(1.2)	(1.0)	(0.7)	(0.6)	(0.5)	(0.45)	(0.4)	(0.3)	(0.28)	(0.25)	(0.23)	(0.21)

Notes: Number in parentheses is the C in $\underline{f}(\phi) = \exp(-C\phi)$ that rationalizes the corresponding q as the optimal choice for the EWC test under $T = 100$.

Table 15: Optimal q and adjustment factor of the Student-t critical values of level α EWC test.

C	100.0	56.2	31.6	17.8	10.0	5.6	3.2	1.8	1.0
$\alpha = 0.01$	$(4, 1.43)$	$(5, 1.34)$	$(6, 1.25)$	$(7, 1.18)$	$(8, 1.13)$	$(10, 1.11)$	$(12, 1.09)$	$(14, 1.06)$	$(17, 1.05)$
$\alpha = 0.05$	$(3, 1.25)$	$(4, 1.22)$	$(5, 1.18)$	$(6, 1.14)$	$(7, 1.10)$	$(8, 1.07)$	$(10, 1.06)$	$(12, 1.05)$	$(14, 1.04)$
$\alpha = 0.10$	$(3, 1.25)$	$(3, 1.14)$	$(4, 1.12)$	$(5, 1.10)$	$(6, 1.08)$	$(7, 1.06)$	$(9, 1.05)$	$(10, 1.03)$	$(13, 1.03)$

Notes: Based on a series of smoothness classes $\mathcal{F}$, in which the "uniformly minimal" function is $\underline{f}(\phi) = \exp(-C\phi^2)$. Sample size $T = 100$. C is log-spaced between 100 and 1.

Table 16: Rule of thumb for adjustment factor of the Student-t critical value of level α EWC test.

q	4	6	8	9	10	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.01$	1.43	1.24	1.13	1.13	1.11	1.09	1.09	1.07	1.06	1.06	1.06	1.05	1.05	1.04	1.04
	(100.0)	(31.6)	(10.0)	(8.0)	(5.6)	(4.0)	(3.2)	(2.2)	(1.8)	(1.5)	(1.2)	(1.0)	(0.9)	(0.7)	(0.6)
$\alpha = 0.05$	1.22	1.14	1.07	1.06	1.06	1.05	1.05	1.04	1.04	1.04	1.03	1.03	1.03	1.03	1.02
	(56.2)	(17.8)	(5.6)	(4.0)	(3.2)	(2.0)	(1.8)	(1.2)	(1.0)	(0.9)	(0.7)	(0.6)	(0.5)	(0.4)	(0.35)
$\alpha = 0.10$	1.12	1.08	1.05	1.05	1.03	1.03	1.03	1.03	1.03	1.02	1.02	1.02	1.02	1.02	1.02
	(31.6)	(10.0)	(4.0)	(3.2)	(1.8)	(1.5)	(1.2)	(1.0)	(0.8)	(0.6)	(0.5)	(0.4)	(0.35)	(0.3)	(0.25)

Notes: Number in parentheses is the C in $\underline{f}(\phi) = \exp(-C\phi^2)$ that rationalizes the corresponding q as the optimal choice for the EWC test under $T = 100$.

Appendix D Computational details in Section 4

In this section, I explain in detail how to numerically identify the null rejection probability maximizer of the EWC test in testing (17), as introduced in Subsection 4.1.1.

To make things clearer, we repeat some notations from the main text. Let the $n+1$ node points $\{x_i\}_{i=0}^n$ define a partition of the interval $I=[0,\pi]$ into n subintervals $I_i=[x_{i-1},x_i],\ i=1,2,\ldots,n$ of length $h_i=x_i-x_{i-1}$ with $x_0=0,\ x_n=\pi$. Let $\{\varsigma_i\}_{i=0}^n$ be a set of basis functions for the space $\mathcal{F}_h$ of continuous piecewise linear functions, which is defined in the main text.

In actual implementations, I choose $n=50$, and $\{x_i\}_{i=0}^{50}$ are log-spaced nodes in $[0,\pi]$. The basis functions $\{\varsigma_i\}_{i=0}^n$ are chosen to be the hat functions

$$\varsigma_i(x)=\begin{cases}(x-x_{i-1})/h_i & ,\ \text{if } x\in I_i,\\ (x_{i+1}-x)/h_{i+1} & ,\ \text{if } x\in I_{i+1},\\ 0 & ,\ \text{otherwise.}\end{cases} \tag{35}$$

I pre-compute $\{\Omega_0(\varsigma_i)\}_{i=0}^n$ with a $5,000$-point Gaussian quadrature for each non-zero element. Since each ς_i is compactly supported, these numerical integrations are nearly precise. For every $(f(x_1),f(x_2),\ldots,f(x_n))'$, $\Omega_0(\hat{f})$ is simply a linear combination of these pre-computed covariance matrices.

Note, however, that the ultimate objective function I will optimize is (21), which involves $\Omega_0(f)$ implicitly through $\lambda_j(\hat{f})$. In a separate exercise in which, for a given EWC test and a parametric AR(1) class $\mathcal{F}$ with coefficient varying over a fine grid, I compare the rejection probabilities following the described approximate procedure and an "exact" procedure in which each entry of $\Omega_0(f)$ is evaluated by numerical integrations via Mathematica. The difference in the rejection probabilities are at most of the order 0.0001. I thus hold on to the above choice of n and $\{x_i\}_{i=0}^{50}$ throughout Section 4.

I proceed in three steps to identify the null rejection probability maximizer for a fixed EWC test of (17), in terms of $(f(x_1),f(x_2),\ldots,f(x_n))'$.

1. Program up the null rejection probability at a given $(f(x_1),f(x_2),\ldots,f(x_n))'\in\mathbb{R}_+^n$ from (21), where $\Omega_0(\hat{f})=\sum_{i=0}^n f(x_i)\Omega_0(\varsigma_i)$ with pre-computed $\{\Omega_0(\varsigma_i)\}_{i=0}^n$.

2. Randomly draw $1,000$ n-dimensional vectors such that each vector equals $(f(x_1),f(x_2),\ldots,f(x_n))'$ for some $f\in\mathcal{F}$. This is in general a challenging task, since the number of numerical constraints

to be checked increases exponentially with n for higher order smoothness constraints. For feasibility, I focus on two types of smoothness classes: the class $\mathcal{F}$ in which $\underline{f}$ corresponds is AR(1) with coefficient ρ and $f \in \mathcal{F}$ is non-increasing over $[0, \pi]$; and the class $\mathcal{F}$ in which $f \in \mathcal{F}$ is Lipschitz continuous in logs with Lipschitz constant C. It is not hard to see that for the first type, it suffices to check the monotonicity constraint consecutively and the lower boundedness condition. For the second type, by the result of Beliakov (2006), the complexity of checking the global Lipschitz condition is reduced to consecutive checking of local Lipschitz conditions.

3. Use every n-dimensional vector drawn in Step 2 as the initial condition to optimize the null rejection probability function programmed in Step 1, subject to linear constraints induced by smoothness class $\mathcal{F}$ (as described in Step 2).

Under the above specifications, it takes 5 to 10 minutes to complete the optimization using $fmincon$ in MATLAB via parallel computing in 12 cores.

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Andrews, D. W. K., and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60(4), 953–966.

Armstrong, T. B., and M. Kolesár (2018a): "Optimal Inference in a Class of Regression Models," *Econometrica*, 86(2), 655–683.

——— (2018b): "Simple and Honest Confidence Intervals in Nonparametric Regression," *arXiv:1606.01200 [math, stat]*, arXiv: 1606.01200.

Atchade, Y. F., and M. D. Cattaneo (2011): "Limit Theorems for Quadratic Forms of Markov Chains," *arXiv:1108.2743 [math]*, arXiv: 1108.2743.

Bakirov, N. K. (1996): "Comparison theorems for distribution functions of quadratic forms of gaussian vectors," *Theory of Probability & Its Applications*, 40(2), 340–348.

Bakirov, N. K., and G. J. Székely (2006): "Student's t-test for Gaussian scale mixtures," *Journal of Mathematical Sciences*, 139(3), 6497–6505.

Beliakov, G. (2006): "Interpolation of Lipschitz functions," *Journal of Computational and Applied Mathematics*, 196(1), 20–44.

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

——— (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics*, ed. by G. Maddala, and C. Rao, vol. 15, pp. 299–342. Elsevier.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83(2), 771–811.

Gonçalves, S., and T. J. Vogelsang (2011): "Block Bootstrap HAC Robust Tests: The Sophiscation of the Naive Bootstrap," *Econometric Theory*, 27(04), 745–791.

Ibragimov, R., and U. K. Müller (2010): "t -Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business & Economic Statistics*, 28(4), 453–468.

Jansson, M. (2004): "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72(3), 937–946.

Jordà, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1), 161–182.

Kariya, T. (1980): "Locally Robust Tests for Serial Correlation in Least Squares Regression," *The Annals of Statistics*, 8(5), 1065–1070.

Kiefer, N. M., and T. J. Vogelsang (2002): "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica*, 70(5), 2093–2095.

——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21(06).

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68(3), 695–714.

King, M. L. (1980): "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," *The Annals of Statistics*, 8(6), 1265–1271.

King, M. L. (1987): "Towards a theory of point optimal testing," *Econometric Reviews*, 6(2), 169–218.

Koijen, R. S., and S. Van Nieuwerburgh (2011): "Predictability of Returns and Cash Flows," *Annual Review of Financial Economics*, 3(1), 467–491.

Lazarus, E., D. J. Lewis, and J. H. Stock (2017): "The Size-Power Tradeoff in HAR Inference," Discussion paper, Harvard University Mimeo.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR Inference: Recommendations for Practice," *Journal of Business & Economic Statistics*, 36(4), 541–559.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Springer, New York, 3rd edn.

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

——— (2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141(2), 1331–1352.

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

Müller, U. K., and M. W. Watson (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979 – 1016.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703.

Phillips, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21(1), 116–142.

Phillips, P. C. B., Y. Sun, and S. Jin (2006): "Spectral Density Estimation and Robust Hypothesis Testing Using Steep Origin Kernels without Truncation," *International Economic Review*, 47(3), 837–894.

——— (2007): "Long Run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," *Journal of Statistical Planning and Inference*, 137(3), 985–1023.

Pötscher, B. M. (2002): "Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters," *Econometrica*, 70(3), 1035–1065.

Rapach, D., and G. Zhou (2013): "Forecasting Stock Returns," in *Handbook of Economic Forecasting*, vol. 2, pp. 328–383. Elsevier.

Robinson, P. M. (2005): "Robust Covariance Matrix Estimation: HAC Estimates with Long Memory/Antipersistence Correction," *Econometric Theory*, 21(1), 171–180.

Sun, Y. (2011): "Robust trend inference with series variance estimator and testing-optimal smoothing parameter," *Journal of Econometrics*, 164(2), 345–366.

——— (2013): "A Heteroskedasticity and Autocorrelation robust F test Using an Orthonormal Series Variance Estimator," *The Econometrics Journal*, 16(1), 1–26.

——— (2014): "Fixed-Smoothing Asymptotics in a Two-Step Generalized Method of Moments Framework," *Econometrica*, 82(6), 2327–2370.

Sun, Y., and D. M. Kaplan (2012): "Fixed-Smoothing Asymptotics and Accurate F Approximation Using Vector Autoregressive Covariance Matrix Estimator," Working paper, University of California, San Diego.

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

Velasco, C., and P. M. Robinson (2001): "Edgeworth expansions for spectral density estimates and studentized sample mean," *Econometric Theory*, 17(3), 497–539.