Optimal HAR Inference*

Liyu Dou†

January 2024

Abstract

This paper considers the problem of deriving heteroskedasticity and autocorrelation robust (HAR) inference about a scalar parameter of interest. The main assumption is that there is a known upper bound on the degree of persistence in data. I derive finite-sample optimal tests in the Gaussian location model and show that the robustness-efficiency tradeoffs embedded in the optimal tests are essentially determined by the maximal persistence. 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 persistence, as well as a corresponding adjustment to the usual Student-t critical value. I illustrate the results in two empirical examples.

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

JEL Codes: C12; C18; C22

*I am deeply indebted to Ulrich Müller for posing the question and for his continuous help, support, and encouragement. I would like to thank the Co-Editors and two anonymous referees for constructive comments and suggestions, which have substantially improved the paper. I also thank Xu Cheng, Paul Ho, Bo Honoré, Michal Kolesár, Jia Li, Mikkel Plagborg-Møller, Mikkel Sølvsten, James Stock, Mark Watson, Ke-Li Xu, Jun Yu, and numerous participants in seminars for helpful comments and suggestions. I gratefully acknowledge financial support from the National Natural Science Foundation of China through Grant 72103176.

†*Address*: Singapore Management University and The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen). *Email*: liyudou@smu.edu.sg.

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 false rejections of the null far too often. A large subsequent literature (surveyed in Müller (2014)) employs alternative asymptotics that is often more accurate in finite samples and thus demonstrates 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 orthonormal series HAR tests,[4] in which the LRV is estimated by projections onto q mean-zero low-frequency 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 orthonormal series LRV estimators. They derive the optimal tuning parameter based on second-order asymptotics

[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); Atchadé and Cattaneo (2011); Gonçalves and Vogelsang (2011); Sun and Kaplan (2012); Sun (2014a); and Sun (2014b), among many others.

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

and under criteria that average functions 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 finite-sample efficiency results of HAR inference about a scalar parameter of interest, without restricting the class of tests and with commonly used notions of optimality in hypothesis testing. Specifically, I derive optimal (weighted average power maximizing scale invariant) HAR tests in the Gaussian location model, under nonparametric assumptions on the underlying spectral density. I also 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 there exists an upper bound on the degree of persistence in data. In time series terminology and from a spectral perspective, this amounts to specifying a worst-case steepest, or "uniformly maximal" spectral density function $\underline{f}$ in class $\mathcal{F}$, which is the collection of all plausible spectra and 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 under a simplifying approximation. 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 the solution of the problem. To make progress, 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 embeds robustness-efficiency tradeoffs in hypothesis testing. This optimal tradeoff is a function of the underlying primitive $\mathcal{F}$, namely the serial correlations one is willing to correct for under the null, and the alternative dependency that one desires the test to orient power toward.

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

[5] See, e.g., Sun, Phillips, and Jin (2008) and Lazarus, Lewis, and Stock (2021).

[6] For parametric examples, Robinson (2005) assumes that the underlying persistence is of the "fractional" type and derives consistent LRV estimators under that class; 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.

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

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

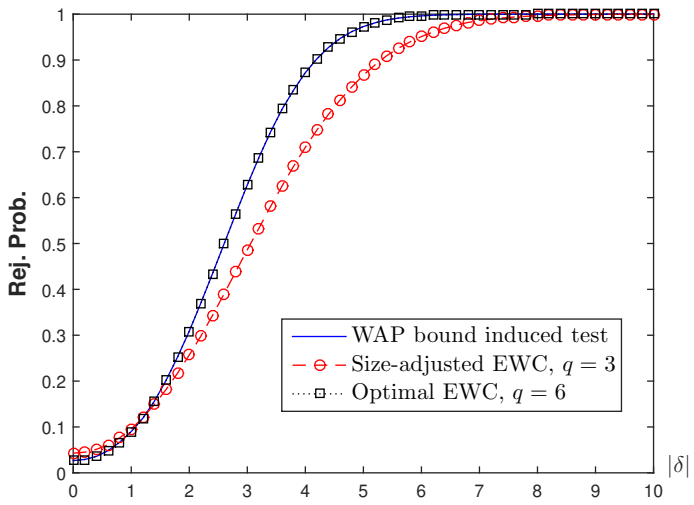
where Y_0 is the sample mean of y_t and $Y_j,\ j = 1, 2, \ldots, q$ are q weighted averages of y_t as $Y_j = T^{-1}\sqrt{2}\sum_{t=1}^T \cos(\pi j(t - 1/2)/T)y_t$. These weighted averages can be approximately thought of as independently normally distributed, each with variance $T^{-1}f(\pi j/T)$. As 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 i.i.d. normals. 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 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 maximal" function of $\mathcal{F}$ corresponds to 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 EWC test, however, has $q = 6$ and inflates the Student-t critical value by a factor of 1.15. Moreover, it is nearly as powerful as a weighted average power bound induced test. 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 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 in Table 2, as follows. Under a series of classes $\mathcal{F}$ where the largest persistence

[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 $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size $T = 100$.

is parameterized as an AR(1) with coefficient $\rho = 1 - c/T$, Table 1 lists the optimal choice of q and the adjustment factor of the Student-t critical value for selected c (and ρ for fixed T). It turns out that the resulting optimal choice of q (almost) remains unchanged as T varies, for fixed $c \leq 55$. More interestingly, for fixed q and T, the adjustment factor does not change substantially under other types of $\mathcal{F}$. Table 2 collects the adjustment factors in Table 1 for selected q. 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. Otherwise, Section 5.1 provides guidance on determining a reasonable $\underline{f}$, the subsequent critical value adjustment, and the selection of q.

This paper relates to a large literature. First, unlike the majority of the HAR literature, I consider optimal HAR inferences without restricting the class of tests. Second, 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 variance in estimating the LRV by a first-order adjustment in the spirit of employing fixed smoothing asymptotics under strong

[8]See, e.g., Velasco and Robinson (2001); Sun, Phillips, and Jin (2008); Sun (2011, 2013, 2014c); and Lazarus, Lewis, and Stock (2021).

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

c	5	10	18	30	40	50	77
ρ	0.95	0.9	0.82	0.7	0.6	0.5	0.23
$\alpha = 0.05$	$(3, 1.55)$	$(4, 1.25)$	$(6, 1.15)$	$(8, 1.09)$	$(10, 1.06)$	$(12, 1.05)$	$(20, 1.03)$

Notes: Based on a series of classes $\mathcal{F}$, in which the "uniformly maximal" function corresponds an AR(1) with coefficient $\rho = 1 - c/T$. Sample size T is 100, but the resulting optimal choice of q (almost) remains unchanged for fixed $c \leq 55$ and for $T = 200, 500, 1000$.

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

q	3	4	6	8	9	10	11	12	16	20
$\alpha = 0.05$	1.55	1.25	1.15	1.09	1.07	1.06	1.06	1.05	1.04	1.03

Notes: Each q is justified as the optimal choice of level α EWC test, under some class $\mathcal{F}$ and for sample size T. An example of the corresponding class $\mathcal{F}$ is the one in which the "uniformly maximal" function corresponds an AR(1) model with coefficient $\rho = 1 - c/T$ as in Table 1.

persistence in Sun (2014a). Even so, the resulting adjusted critical value is easily computed without simulations under a simplifying structure. Third, this paper contributes to the uniform size control literature developed by Müller (2014), Preinerstorfer and Pötscher (2016), Pötscher and Preinerstorfer (2018, 2019), and Müller and Watson (2022). I analytically derive powerful tests that uniformly control size over arguably large classes of models, while Müller (2014) numerically determines powerful tests under a possibly restricted parametric class of models. Preinerstorfer and Pötscher (2016) and Pötscher and Preinerstorfer (2018, 2019) focus on size distortions and power deficiencies of given HAR tests, allowing for general classes of models. Müller and Watson (2022) derive finite-sample size control results in general spatial settings without the simplifying structure considered in this paper but offer limited analytical results on the efficiency side.

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 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 and is thus more in the spirit of Lehmann and Stein (1948), 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 efficiency results under an essential simplification, which are theoretically investigated in more general settings in Section 4. Section 5 converts the theoretical insights into practical guidance and discusses the implementation in regression models. Section 6 provides empirical illustrations with a self-contained guide to implementation. An interested practitioner can skip the theoretical discussions and read Section 6 directly. Proofs and computational details are provided in the appendices.

2 Model and Preliminaries

The paper concerns 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 location model (2) is often considered a stylized setting to provide theoretical insights into HAR inference.[9] As simple as it is, this model is empirically relevant in a number of situations. For example, the statistical study of unconditional equal predictive ability (UEPA) concerning competing forecasts in financial and macroeconomic contexts amounts to testing an unconditional mean-zero condition in (2) with y_t being the produced loss differential series.

Throughout the paper, I mainly focus on presenting and analyzing new efficiency results for HAR inference in the univariate Gaussian location model (3), and discuss the implications for conducting inference about a nonconstant regressor in regression settings. The ideas and methods explored in the simplest model (3) can often be used as a foundation for studying

[9] For HAR studies based on the Gaussian location model, see, e.g., Velasco and Robinson (2001); Jansson (2004); Sun, Phillips, and Jin (2008); Sun (2011, 2013, 2014c); Müller (2014); Lazarus, Lewis, and Stock (2021), among many others.

HAR inference in multivariate location models and general GMM settings, potentially involving additional complications.[10] Formal generalizations of this paper's efficiency results along those lines are, however, not straightforward and beyond the scope of the paper.

The HAR inference problem in (3) 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 infinite-dimensional nuisance parameter f. I follow standard approaches to deal 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 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 analytical tractability, 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 and thus known $\Omega_0(f)_{1,1}$, the choice $\eta^2 = (\kappa-1)\Omega_0(f)_{1,1}$ 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 $\Omega_1(f) = \Omega_0(f) + (\kappa-1)\iota_1\iota_1'\Omega_0(f)_{1,1}$. 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.

[10] For formal discussions of HAR inference in general GMM settings, see, e.g., Sun (2014b); Hwang and Sun (2017, 2018).

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 $a > 0$. By standard testing theory,[11] 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))

$$h_{i,f}(y^s) = C|\Omega_i(f)|^{-1/2} \left(y^{s\prime}\Omega_i(f)^{-1}y^s\right)^{-T/2} \tag{6}$$

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. To make progress, I first consider directing power at a flat spectrum $f_1 = 1$ (white noise) when demonstrating how to derive efficiency bounds in Section 3. In this case, the alternative H_1'' then becomes a single hypothesis H_{1,f_1}'' : "Y^s has density h_{1,f_1}," where $\Omega_1(f_1) = \kappa T^{-1}\text{diag}\,(1, \kappa^{-1}, \ldots, \kappa^{-1})$. 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 $\mathcal{F}$. In Section 4.1, I robustify insights from the white noise case to ones where power is directed at a non-flat spectrum $\tilde{f}_1$, or when minimax bounds are concerned if f belongs to a class $\mathcal{G} \subset \mathcal{F}$ under H_1.

The testing problem is now reduced to distinguishing the composite null H_0'' from a single alternative. 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 $\int \varphi_{\text{ah}}(y^s) \int h_{0,f}(y^s)\mathrm{d}\Lambda(f)\mathrm{d}y^s = \int\int \varphi_{\text{ah}}(y^s)h_{0,f}(y^s)\mathrm{d}y^s\mathrm{d}\Lambda(f) \leq \alpha$. 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 one. 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 such a distribution. I make progress along this line in the following sections.

[11] See, e.g., Chapter 6 in Lehmann and Romano (2005)).

3 Finite-sample Efficiency Results

In this section, I impose a simplifying Whittle-type diagonal structure on the implied covariance matrices Ω_0 of the effective observation Y, specify *a priori* that $\mathcal{F}$ possesses a most persistent spectrum, and analytically derive the resulting least favorable distribution and thus obtain the optimal test. More specifically, I make the following assumptions.

Assumption 3.1 *For all* $f \in \mathcal{F}$, $\Omega_0(f) = T^{-1}\text{diag}(f(0), f(\pi/T), \ldots, f(\pi(T-1)/T)$.

Assumption 3.1 specifies an approximation of $\Omega_0(f)$ in the spirit of Whittle (1951). It holds exactly when y_t follows a Gaussian white noise or a demeaned Gaussian random walk process. For stationary y_t with f falling into popular 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 $\mathcal{F}$ being a nonparametric class, Choudhuri, Ghosal, and Roy (2004) has shown the mutual contiguity of the Whittle measure with the actual distribution of the data for a Gaussian time series; Golubev, Nussbaum, and Zhou (2010) has established asymptotic equivalence in the sense of Le Cam's deficiency measure between an experiment of observing a stationary centered Gaussian sequence and another of observing Y under a diagonal covariance structure analogous to Assumption 3.1.

Under Assumption 3.1 and the weighted average power criterion, as specified by a given κ, I now seek powerful tests as functions of $Y^s = Y/\sqrt{Y'Y}$ in the problem of

$$\begin{aligned} H_0^d : Y &\sim \mathcal{N}\left(0, T^{-1}\text{diag}(f(0), f(\pi/T), \ldots, f(\pi(T-1)/T))\right), \ f \in \mathcal{F} \\ \text{against } H_{1,f_1}^d : Y &\sim \mathcal{N}\left(0, \kappa T^{-1}\text{diag}(1, \kappa^{-1}, \ldots, \kappa^{-1})\right), \end{aligned} \tag{7}$$

where the superscript d denotes the Whittle-type diagonal structure in Assumption 3.1.

Assumption 3.2

(a) There exists a $\underline{f} \in \mathcal{F}$ *such that* $\frac{f(0)}{f(\phi)} \leq \frac{\underline{f}(0)}{\underline{f}(\phi)}$, *for all* $\phi \in [-\pi, \pi]$ *and* $f \in \mathcal{F}$.

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

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

Assumption 3.1(a) states the existence of a "uniformly maximal" function in $\mathcal{F}$ in terms of persistence. Loosely speaking, more persistent processes have higher values of $f(0)/f(\phi)$. As such, one could understand it as restricting the degree of persistence *a priori*. When additional parametric assumptions are imposed on $\mathcal{F}$, such as the local-to-unity parametrization

in Müller (2014), Assumption 3.1(a) amounts to bounding the degree of mean reversion from below. However, I stress that one is not committing to any parametric classes here, even if $\underline{f}$ is defined from a parametric model. For example, suppose $\underline{f}$ is the spectrum of an AR(1) model with coefficient 0.8. Then, the resulting $\mathcal{F}$ not only covers all stationary AR(1) models with coefficient less than 0.8, but it also contains many other empirically relevant models, such as ARMA models whose spectra may oscillate but the implied persistence is bounded above by that of $\underline{f}$. Assumption 3.1 (a) can often be implied by global shape restrictions 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 in magnitude by a positive constant C, the implied "maximal persistence" function emerges as $\underline{f}(0)/\underline{f}(\phi) = \exp(C\phi)$.

Assumptions 3.1(b) and (c) regularize $\mathcal{F}$, such that the least favorable distribution over $\mathcal{F}$ exists and puts a point mass on an element in it. Inspection of the proof of Theorem 3.3 shows that Assumptions 3.1(b) is sufficient and may not be necessary. But I stress that, even so, one is not committing to a shape restriction for all spectra in $\mathcal{F}$. Furthermore, a stronger monotonicity assumption on $\underline{f}$, that is, it is non-increasing not only at $\lambda = \pi j/T$, $j = 0, 1, \ldots, T-1$, but also over $[-\pi, \pi]$, is often satisfied in the smoothness classes $\mathcal{F}$ discussed above. Thus, given Assumption 3.1(a), Assumptions 3.1(b) is arguably fairly mild.

3.1 Optimal test

The optimal HAR test under Assumptions 3.1 and 3.2 is stated in the following theorem.

Theorem 3.3 *Let $\mathcal{F}$ be a set of f satisfying Assumption 3.2 with the "uniformly maximal" function $\underline{f}$. Under Assumption 3.1 and for a given κ that specifies a weighted average power criterion,*

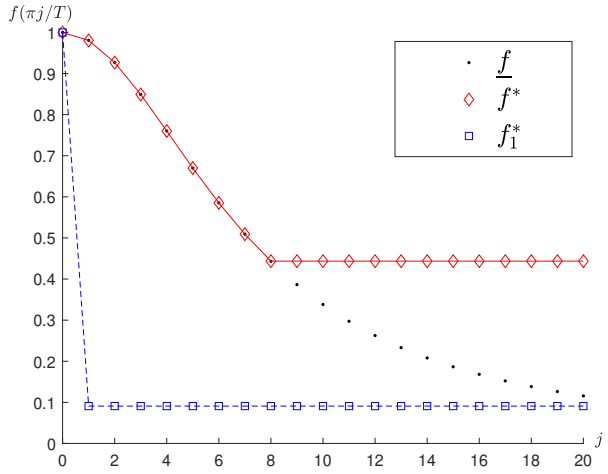
1. *If $\kappa \leq \underline{f}(0)/\underline{f}(\pi/T)$, then the best weighted average power maximizing scale invariant test of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ is the trivial randomized test.*

2. *If $\kappa > \underline{f}(0)/\underline{f}(\pi/T)$, 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*

$$\frac{Y_0^2 + \underline{f}(0)\sum_{j=1}^{q^*} Y_j^2/\underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2} \tag{8}$$

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 $f = \underline{f}$.

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



Notes: The "uniformly maximal" 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.

The proof of part 1 of Theorem 3.3 is simple. Notice that if $1/\kappa^{-1} \leq \underline{f}(0)/\underline{f}(\pi/T)$, then the alternative H_{1,f_1}^d is included in the null H_0^d. Any nontrivial size-controlling test thus cannot be more powerful than the trivial randomized test in this case.

The idea of the proof for part 2 of Theorem 3.3 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^*}^d against H_{1,f_1}^d is

$$\varphi_{\Lambda^*,f_1} = 1\left[\frac{Y_0^2 + f^*(0)\sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T)}{Y_0^2 + \kappa\sum_{j=1}^{T-1} Y_j^2} > \text{cv}\right],$$

for some $\text{cv} \geq 0$. 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^d. Intuitively, this requires H_{0,Λ^*}^d to be as indistinguishable as possible from H_{1,f_1}^d. This somewhat implies that, after appropriate normalization, 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 (8). The formal proof of Theorem 3.3 is given in Appendix A.

Comment 1. Theorem 3.3 shows that we can only be confident about a HAR test if

we are willing to make *a priori* assumptions about the persistence properties of the data. Furthermore, it gives an explicit recipe for devising optimal tests as functions of the underlying primitive, namely the largest persistence one is willing to consider (without such assumptions, informative inference is impossible, as shown by Pötscher (2002)).

Comment 2. It is known from Preinerstorfer and Pötscher (2016), Pötscher and Preinerstorfer (2018, 2019) that if AR(1) structures are allowed in the Gaussian location model, typical HAR tests will have size equal to one. This is essentially caused by the so-called singularity points that correspond to AR(1) processes with roots near -1 and 1. In the current context, Assumption 3.1 implicitly precludes AR(1) processes with a coefficient close to -1: Their averaged absolute correlations for the implied Ω_0, as Figure 3 illustrates for $(Y_0, Y_1, \ldots, Y_q)'$ with various choices of q, are not negligible so that the Whittle diagonal structures are no longer appropriate approximations; Assumption 3.1(a) implies an upper bound on the AR(1) coefficient so that the very close vicinity of 1 is also precluded.

In samples of size $T = 100$ and when $\underline{f}$ corresponds to an AR(1) with coefficient 0.8, Theorem 3.3 implies the optimal q^* to be 7. The null rejection probabilities of this optimal test in AR(1) models are plotted against the coefficient ρ in Figure 3. The size-controlling properties of the optimal test are well demonstrated for processes less persistent than $\underline{f}$. For more persistent processes on the right end, the inferior size performance is well expected and consistent with the insights of Preinerstorfer and Pötscher (2016). For processes on the left end, it turns out that the optimal test also controls size well. But because these models are precluded, the explanation for this size property is beyond the scope of Theorem 3.3.

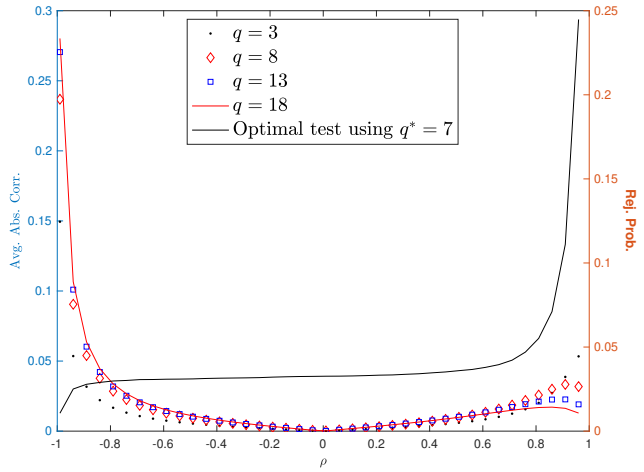
More interestingly, the above size-controlling features of the optimal test extend to more complicated error dynamics than AR(1). For example, Figure 4 reports the average absolute correlations for Ω_0 and null rejection probabilities of the optimal test using $q^* = 7$ in AR(2) Gaussian location models with roots ρ_1 and ρ_2. An AR(2) with at least one root near -1 is again incompatible with Assumption 3.1 yet the optimal test controls size well in that model. In addition, the size distortions are not severe as long as one root is not near 1.

Comment 3. The optimal test (8) can be rewritten as

$$\left| \frac{Y_0}{\sqrt{\sum_{j=1}^{q^*} w_j^* Y_j^2}} \right| \geq 1, \tag{9}$$

which is of a familiar t-statistic form where the implied LRV estimator is a weighted orthonormal series one with weights $w_j^* = \left[\kappa \, \mathrm{cv}_{q^*} - \underline{f}(0)/\underline{f}(\pi j/T)\right](1 - \mathrm{cv}_{q^*})^{-1}$, $j = 1, 2, \ldots, q^*$ for an "endogenously" determined number q^* of weighted averages of y_t. Note that neither does

Figure 3: Average Absolute Corrrelations for Ω_0 and size performance of the optimal test using $q^* = 7$ in AR(1) Gaussian location model with root ρ.



Notes: The $\underline{f}$ function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size $T = 100$. Average absolute correlation on the left y-axis; size on the right y-axis. Size computed based on 50000 simulations.

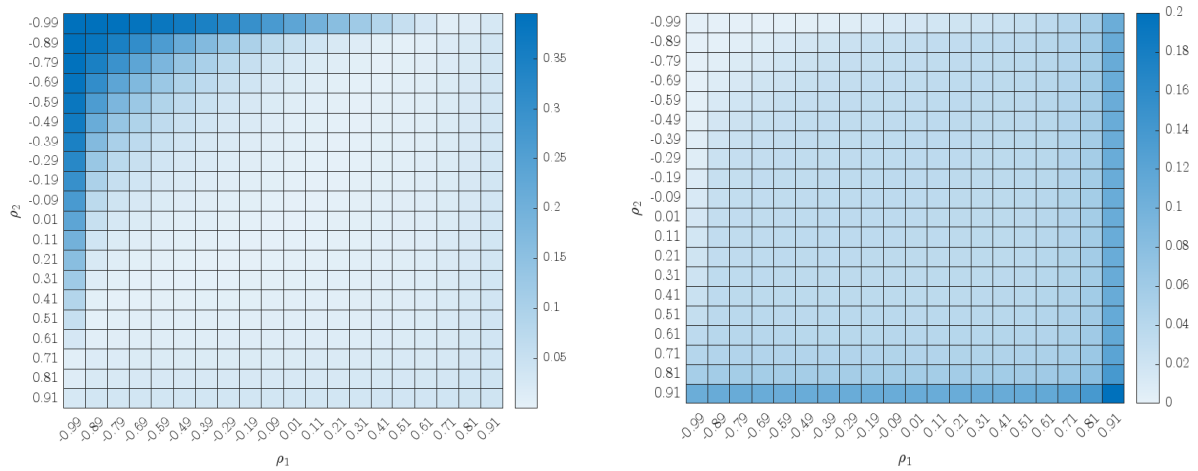
this statistic form emerge by restricting the class of LRV estimators *a priori* nor does it start by considering functions of $\left(Y_1^s, \ldots, Y_q^s\right)'$ for some fixed q as in Müller (2014).

Inspection of the proof of Theorem 3.3 reveals that the determination of the implied weights $\{w_j^*\}$ and q^* embeds classical robustness-efficiency tradeoffs in optimal testing. On the one hand, for a given $f \in \mathcal{F}$, the null rejection probability of test (9) is

$$P\left(\left|\frac{Y_0}{\sqrt{\sum_{j=1}^{q^*} w_j^* Y_j^2}}\right| > 1\right) = P\left(\frac{f(0)Z_0^2}{\sum_{j=1}^{q^*} f(\pi j/T) w_j^* Z_j^2} \geq 1\right), \tag{10}$$

where $\{Z_j\}_{j=0}^{q^*}$ are $q^* + 1$ i.i.d. standard normals. Under Assumption 3.2(a), it is not hard to see that (10) as a functional of f is maximized at $\underline{f}$, as long as the weights $\{w_j^*\}$ are all positive. (And this observation breaks down when negative weights emerge.) As such, one needs these implied weights to be positive such that the optimal test in its original form ϕ_{Λ^*, f_1} uniformly controls the null rejection frequencies under H_0^d and then the "least favorable" argument can be applied. For given κ and $\underline{f}$, suppose one starts by considering $\tilde{q} = 1, 2, \ldots$ and thus uses the accordingly determined $\text{cv}_{\tilde{q}}$, it is found that the positiveness criterion is less likely to be satisfied as $\tilde{q}$ increases. In this sense, a smaller $\tilde{q}$ is preferred to achieve better robustness properties. On the other hand, one may desire to use a larger $\tilde{q}$ for efficiency purposes. It is these tradeoffs that essentially pin down q^* and the implied weights:

Figure 4: Average Absolute Correlations for Ω_0 and size performance of the optimal test using $q^* = 7$ in AR(2) Gaussian location model with roots ρ_1 and ρ_2.



(a) Average Absolute Correlations.

(b) Null Rejection Probabilities.

Notes: The "uniformly maximal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100. Panel (b) based on 10000 simulations.

One shall include the largest possible number of weighted averages such that the positiveness conditions are satisfied; The accordingly implied weights are optimal in the sense that the resulting statistic is equivalent to a likelihood ratio one based on all weighted averages and inherits its efficiency properties in Theorem 3.3.[12]

The above robustness-efficiency tradeoffs continue to hold when the testing primitives (κ and $\underline{f}$) vary. Intuitively, a larger κ or a less persistent $\underline{f}$ implies that it is easier to distinguish H_0^d from H_{1,f_1}^d and a larger q^* is expected. It is instructive to illustrate it in cases where $\underline{f}$ is parameterized as an AR(1) with coefficient ρ, and Table 3 reports the corresponding q^*. Unsurprisingly, q^* is monotonically decreasing in ρ, reflecting the need to satisfy more stringent size constraints (robust to larger sets of data generating processes). It is also worth noting that if the sequence of ρ is understood as fixed in asymptotic thought experiments, then the notion of persistence under consideration is only a relative concept, and the $O(T^{-1})$ local-to-zero spectra for all $f \in \mathcal{F}$ will be effectively flat and correspond to the white noise

[12]One may wonder why t tests based on $q(< q^*)$ Y_j's and the accordingly derived weights are not least favorable as well. As it turns out, the corresponding kinked-shaped functions, as plausible least favorable candidates, all lie outside $\mathcal{F}$ due to Assumption 3.2(b). This observation, together with Assumption 3.2 (c), ensures that the identified least favorable function f^* associated with q^* exists and is unique.

Table 3: q^* in the optimal test and $\underline{f}$ parameterized as AR(1) with coefficient ρ.

ρ	0	0.05	0.15	0.3	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
q^*	99	51	31	22	15	13	12	11	10	9	7	6	5	3

Notes: The "uniformly maximal" function corresponds to an AR(1) with coefficient ρ. Sample size T is 100.

case in large samples. In that event, Table 3 suggests that $q^* \to \infty$ as $T \to \infty$, and one has to appeal to asymptotic expansions to further analyze finite-sample size-power tradeoffs and the optimal rate of q^*. Said differently, one shall understand Theorem 3.3 as in thought experiments in which strong autocorrelations or large persistence is still of concern in large samples and interpret the curvature of $\underline{f}$ at frequencies close to zero, for say $\rho_T = 1 - c/T$, as the primitive.

Comment 4. For t-statistic using a conventional weighted cosine LRV estimator with predetermined positive weights $\{w_j\}$, referred to as weighted cosine tests hereafter, the null rejection probability at a given f and critical value cv is

$$P\left(\left|\frac{Y_0}{\sqrt{\sum_{j=1}^q w_j Y_j^2}}\right| > \text{cv}\right) = P\left(\frac{f(0)Z_0^2}{\text{cv}^2 \sum_{j=1}^q f(\pi j/T) w_j Z_j^2} \geq 1\right), \tag{11}$$

which is always maximized at $\underline{f}$. The critical value cv at level α can readily be obtained by equating α to the following formula of Bakirov and Székely (2006):

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

where $\zeta_j = \text{cv}^2\, \underline{f}(\pi j/T)/\underline{f}(0) w_j$. The robustness-efficiency tradeoffs now manifest in terms of choosing q and using its associated adjusted critical value cv_q^a such that (12) is maximized at $\{\tilde{\zeta}_j = (\text{cv}_q^a)^2 \kappa^{-1} w_j\}_{j=1}^q$ and thus the maximum power at H_{1,f_1}^d is achieved. If one desires to direct power at say $\tilde{f}_1$ instead of f_1, then the critical value adjustments will not change for any given q, and the optimal selection of q now makes (12) reach its largest value at $\{\tilde{\zeta}_j = (\text{cv}_q^a)^2 \kappa^{-1} \tilde{f}_1(\pi j/T)/\tilde{f}_1(0) w_j\}_{j=1}^q$. In this sense, robustness and efficiency constraints are sequentially addressed. In contrast, the optimal test $\varphi_{\Lambda^*, \tilde{f}_1}$ (see details in Corollary 4.4) and the resulting optimal weights $\{w_j^*\}$ in the t-statistic form (9) vary simultaneously with $\tilde{f}_1$, rendering it less operationally straightforward as compared to the weighted cosine tests.

The above discussion, of course, applies to the class of EWC tests, for which $w_j = q^{-1}$ for any given q. In that case, if $\underline{f}$ is further parameterized using the limiting local-to-zero spectra

under local-to-unity asymptotics, that is, $1/(\pi^2 j^2 + c^2)$ for some $c > 0$, $(\text{cv}_q^a)^2$ mirrors the critical value of an F statistic (with $p = 1$) from fixed-smoothing asymptotics under strong (local-to-unity) persistence in Sun (2014a). Moreover, $(\text{cv}_q^a)^2$ is of a hump shape as a function of q for any given and finite c (see Figure 3 in Sun (2014a)). The optimal choice of q then amounts to exploiting that hump shape and picking the q such that the vector that stacks q replicates of $(\text{cv}_q^a)^2\kappa^{-1}q^{-1}$ and $T-1-q$ zeros majorizes all other candidate vectors of the same form. In face of $c = \infty$, cv_q^a monotonically decreases in q. As a result, one shall utilize all information in the data and optimally choose $q = T - 1$.

Intuitively, this q shall also balance the worst absolute bias $(|\underline{f}(0) - q^{-1}\sum_{j=1}^{q}\underline{f}(\pi j/T)|)$ and variability of the LRV estimator $q^{-1}\sum_{j=1}^{q} Y_j^2$ in a testing-optimal sense. Accordingly, the adjusted critical value cv_q^a accounts for this maximum bias in the spirit of Armstrong and Kolesár. However, since cv_q^a is determined in a rather more complicated way than in their settings, I choose not to discuss further these bias-variance tradeoffs. Furthermore, I note that because the implied weights $\{w_j^*\}_{j=1}^{q^*}$ in the optimal test do not depend on the associated critical value cv_{q^*} in a straightforward way and thus may bring another layer of complications, neither will I formally compare q^* in the optimal test with the MSE or testing optimal q in the EWC class in this paper.

Comment 5. One may wonder whether Theorem 3.3 is limited by only orienting power toward f_1. As it turns out, the insights can be generalized to accommodate possibly more complex alternatives. I relegate the formal analysis along this line to Section 4.1.

3.2 The optimal EWC test

By using higher-order expansions, Lazarus, Lewis, and Stock (2021) derive a size-power frontier for kernel and orthonormal 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 the current context. The efficiency bounds derived in the last section provide 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 two related goals. The first is to study the (weighted average) power properties of the EWC test relative to the optimal test in Theorem 3.3. As it turns out, the EWC test is close to optimal, under an appropriate choice of q and with the adjusted critical value, which is just discussed in *Comment* 4 above. I refer to this new EWC test as the optimal EWC hereafter. My second goal is to draw the following practical implications of the optimal EWC test via its comparison with the conventional EWC test: One should use the EWC test with

Table 4: 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
q in optimal EWC	13	9	7	**6**	4	3	1
critical value adj. factor	1.04	1.05	1.07	**1.13**	1.26	1.55	1.87
WAP of optimal EWC	0.507	0.492	0.472	**0.438**	0.351	0.233	0.088
WAP of optimal test	0.507	0.492	0.474	0.442	0.356	0.239	0.089

Notes: The "uniformly maximal" function corresponds to an AR(1) with coefficient ρ. Nominal level is 5%. Sample size T is 100.

Table 5: 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	0.2	0.1
WAP of optimal EWC	0.286	0.363	0.415	0.454	0.483	0.501	0.524	0.533
WAP of optimal test	0.290	0.365	0.416	0.456	0.483	0.501	0.526	0.535

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

a larger q and appropriately enlarged critical values for more powerful HAR inference.

3.2.1 Power of the optimal EWC test

Consider the type of $\mathcal{F}$ in Table 3, that is, the "uniformly maximal" function $\underline{f}$ of the class $\mathcal{F}$ corresponds to an AR(1) with coefficient ρ. As ρ varies, Tables 4 displays the crucial ingredients (q and adjustment factor relative to Student-t critical value) in the optimal EWC test, its resulting weighted average power and the efficiency bound induced by the optimal test. Two observations are immediate. First, the optimal EWC test is nearly as powerful as the optimal test. This observation remains when a different type of class $\mathcal{F}$ is considered in Table 5. Second, the selections of q in the optimal EWC test and q^* in the optimal test are not necessarily equal. After all, as explained in *Comments* 3 *and* 4 above, they are determined in arguably different robustness-efficiency tradeoff mechanisms. It is worth noting that q in the optimal EWC test can be somewhat sensitive to numerical errors in evaluating (12) when $\underline{f}$ is relatively flat. This, however, does not have any substantial consequence in theory.

3.2.2 Practical implications

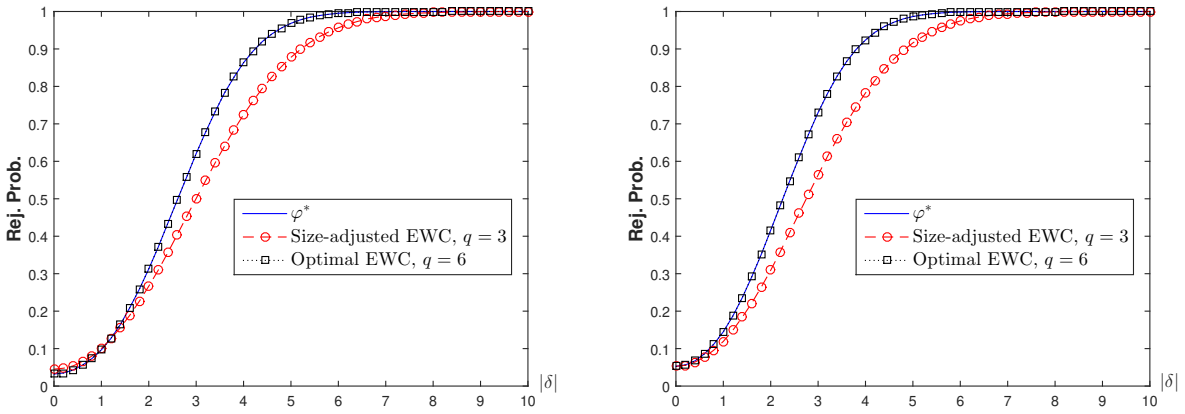
Recall that the conventional wisdom in implementing the EWC test is to use a sufficiently small q and to employ the Student-t critical value. I find, however, that it is better to use a larger q and to employ an enlarged critical value. Take the example from Figure 2 as an illustration: The "uniformly maximal" function $\underline{f}$ corresponds to an AR(1) with coefficient 0.8 and the sample size is fixed to be 100. According to conventional wisdom, one needs to use $q = 3$ in the usual EWC test to obtain size distortions less than 0.01. The optimal EWC test, however, selects a larger $q = 6$ (highlighted in Table 4), and the corresponding Student-t critical value must be inflated by a factor of 1.13 for exact size control. An apple-to-apple comparison then reveals that the size-adjusted weighted average power of the usual EWC test (0.39) has about 12% loss as compared to that of the optimal EWC test (0.438, as highlighted in Table 4).

The superior power property of the optimal EWC test is further evident when local alternatives are considered. In particular, in the context of the above example, I consider $\mu = \delta T^{-1/2}(1-\rho_1)^{-1}$ under the alternative. Panels (a) and (b) of Figure 5 plot the power of the optimal 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 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.

4 Theoretical Generalizations

The finite-sample efficiency results in Section 3 are derived under seemingly restrictive assumptions. In particular, when power is directed at the white noise alternative *a prior*, the optimal test φ^* possesses a precise sense of optimality, and the optimal EWC is numerically found to be nearly as powerful as φ^*. More importantly, the existing efficiency results are entirely based on the Whittle-type diagonal structure. It is natural to ask how limited these simplifying assumptions are in eliciting theoretical insights in HAR inference. Said

Figure 5: Power function plot of the test φ^*, the optimal and conventional EWC tests.



(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.

differently, can the insights on efficiency in Section 3 be generalized to more general settings? In this section, I take up these questions and discuss the theoretical generalizations.

4.1 Power directions and minimax efficiency results

First of all, I devise optimal tests that direct power at a non-flat $\tilde{f}_1$, and, more generally, a non-singleton class $\mathcal{G} \subset \mathcal{F}$ under H_1. For analytical tractability, I maintain Assumption 3.1 and also impose a Whittle-type structure on $\Omega_1(f)$, which automatically holds under f_1.

Assumption 4.1 *For all* $f \in \mathcal{G}$, $\Omega_1(f) = T^{-1}\mathrm{diag}(\kappa f(0), f(\pi/T), \ldots, f(\pi(T-1)/T)$.

Adopting the conventional scale invariance and weighted average power maximizing criteria, I now seek powerful tests as functions of $Y^s = Y/\sqrt{Y'Y}$ in the problem of

$$H_0^d : Y \sim \mathcal{N}\left(0, T^{-1}\mathrm{diag}(f(0), f\left(\pi/T\right), \ldots, f\left(\pi(T-1)/T\right))\right), \ f \in \mathcal{F} \qquad (13)$$
$$\text{against } H_{1,\mathcal{G}}^d : Y \sim \mathcal{N}\left(0, T^{-1}\mathrm{diag}\left(\kappa f(0), f\left(\pi/T\right), \ldots, f\left(\pi(T-1)/T\right)\right)\right), \ f \in \mathcal{G},$$

under the following assumption that nests Assumption 3.2 as a special case ($\mathcal{G} = \{f_1\}$).

Assumption 4.2

(a) There exists a $\underline{f} \in \mathcal{F}$ *such that* $\frac{f(0)}{f(\phi)} \leq \frac{\underline{f}(0)}{\underline{f}(\phi)}$, *for all* $\phi \in [-\pi, \pi]$ *and* $f \in \mathcal{F}$.

(b) There exists a $\overline{f} \in \mathcal{G}$ *such that* $\frac{f(0)}{f(\phi)} \geq \frac{\overline{f}(0)}{\overline{f}(\phi)}$*, for all* $\phi \in [-\pi, \pi]$ *and* $f \in \mathcal{G}$*.*

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

(d) $\mathcal{F}$ *contains all kinked functions defined by* $f_\theta(\phi) = r_\theta(\phi)\underline{f}(\phi)$*, in which, for* $\theta \in [0, \pi]$*,* $r_\theta(\phi) = 1$ *if* $[|\phi| \leq \theta]$ *and* $r_\theta(\phi) \in [1, \infty)$ *if* $|\phi| > \theta$*.*

(e) $\mathcal{G}$ *contains all kinked functions defined by* $g_\theta(\phi) = r'_\theta(\phi)\overline{f}(\phi)$*, in which, for* $\theta \in [0, \pi]$*,* $r'_\theta(\phi) = 1$ *if* $[|\phi| \leq \theta]$ *and* $r'_\theta(\phi) \in [0, 1]$ *if* $|\phi| > \theta$*.*

Intuitively, relative persistence between the null and alternative hypotheses matters in distinguishing them. In Section 3, because the alternative is fixed at a constant singleton, the "uniformly maximal" spectral density plays a vital role in deriving "least favorable" results. Now, under Assumption 4.2, $\underline{f}/\overline{f}$ posits the largest possible relative persistence and is thus expected to be an important primitive in eliciting similar efficiency results. It turns out that this intuition is correct and I formalize it below as a minimax result.

I follow Lehmann and Romano (2005) to define the necessary notations in the HAR context. Let Λ_0 and Λ_1 denote distributions of f over $\mathcal{F}$ and $\mathcal{G}$, respectively. Let $\varphi_{\Lambda_0,\Lambda_1}$ be the most powerful level α weighted average power maximizing scale invariant test for testing H_{0,Λ_0} against H_{1,Λ_1}, in which H_{0,Λ_0} and H_{1,Λ_1} are simple hypotheses defined analogously to $H''_{0,\Lambda}$ in Section 2, and let $\beta_{\Lambda_0,\Lambda_1}$ be its weighted average power for a given κ. Suppose Λ_0 and Λ_1 are such that $\sup_{f\in\mathcal{F}} E[\varphi_{\Lambda_0,\Lambda_1}(y)] \leq \alpha$ and $\inf_{f\in\mathcal{G}} E[\varphi_{\Lambda_0,\Lambda_1}(y)] = \beta_{\Lambda_0,\Lambda_1}$, then $\varphi_{\Lambda_0,\Lambda_1}$ maximizes $\inf_{f\in\mathcal{G}} E[\varphi(y)]$ among all valid level α tests φ of H_0. The following theorem makes further optimal statements about the minimax weighted average power bound $\beta_{\Lambda_0,\Lambda_1}$ among all possible candidates of Λ_0 and Λ_1.

Theorem 4.3 *Under Assumptions 3.1, 4.1, and 4.2,*

1. *If* $\kappa\overline{f}(0)/\underline{f}(0) \leq \overline{f}(\pi/T)/\underline{f}(\pi/T)$*, then the smallest* $\beta_{\Lambda_0,\Lambda_1}$ *among all possible pairs of* Λ_0 *and* Λ_1 *is* α *and is attained by the trivial randomized test.*

2. *If* $\kappa\overline{f}(0)/\underline{f}(0) > \overline{f}(f)(\pi/T)/\underline{f}(\pi/T)$*, then the following test* φ^* *identifies the least favorable pair of distributions* $(\Lambda_0^*, \Lambda_1^*)$ *in the sense that*

$$\beta_{\Lambda_0^*,\Lambda_1^*} \leq \beta_{\Lambda_0,\Lambda_1}$$

for all possible pairs of Λ_0 *and* Λ_1*, in which* φ^* *rejects for large values of*

$$\frac{Y_0^2 + \underline{f}(0)\sum_{j=1}^{q^*} Y_j^2/\underline{f}(\pi j/T)}{Y_0^2 + \kappa\overline{f}(0)\sum_{j=1}^{q^*} Y_j^2/\overline{f}(\pi j/T)} \tag{14}$$

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* $f = \underline{f}$ *and attains its minimax power* $\beta_{\Lambda_0^*,\Lambda_1^*}$ *at* $f = \overline{f}$*.*

The proof strategy of Theorem 4.3 is similar to that of Theorem 3.3. I prove both of them and the two immediate corollaries within a coherent framework in Appendix A.

Corollary 4.4 *Let* $\mathcal{G} = \{\tilde{f}_1\}$ *for some* $\tilde{f}_1$ *that is not necessarily equal to the flat* f_1 *so that* $\overline{f} = \tilde{f}_1$ *in the sense of Assumption 4.2(b). Under Assumptions 3.1, 4.1, 4.2,*

1. *If* $\kappa\tilde{f}_1(0)/\underline{f}(0) \leq \tilde{f}_1(\pi/T)/\underline{f}(\pi/T)$*, then the best weighted average power maximizing scale invariant test of* $H_0 : \mu = 0$ *against* $H_1 : \mu \neq 0$ *is the trivial randomized test.*

2. *If* $\kappa\tilde{f}_1(0)/\underline{f}(0) > \tilde{f}_1(\pi/T)/\underline{f}(\pi/T)$*, 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*

$$\frac{Y_0^2 + \underline{f}(0)\sum_{j=1}^{q^*} Y_j^2/\underline{f}(\pi j/T)}{Y_0^2 + \kappa\tilde{f}_1(0)\sum_{j=1}^{q^*} Y_j^2/\tilde{f}_1(\pi j/T)} \tag{15}$$

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* $f = \underline{f}$*.*

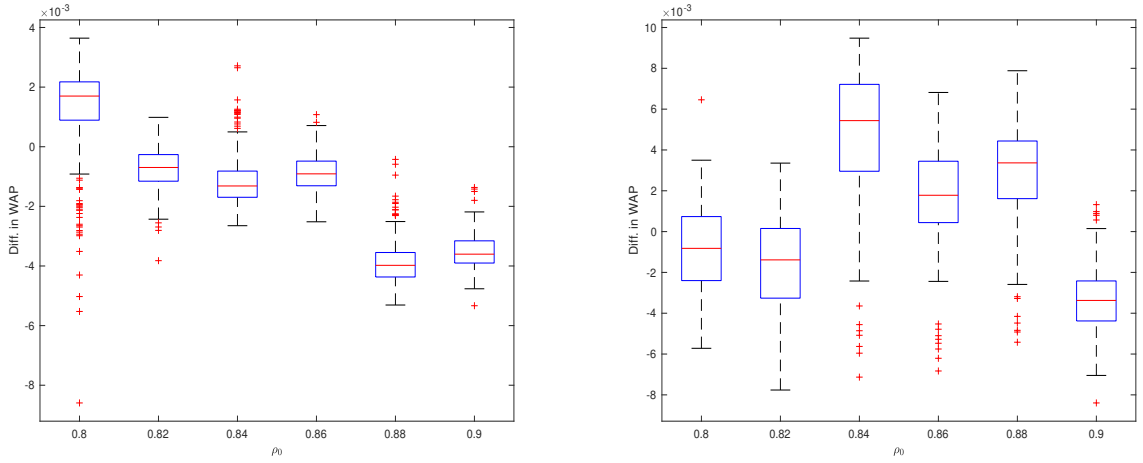
Corollary 4.5 *Let* $\mathcal{F}$ *and* $\mathcal{G}$ *be sets of* f *satisfying Assumption 4.2 and with* $\overline{f}_1 = f_1$*. Under Assumptions 3.1 and 4.1, the weighted average power of the optimal test in Theorem 3.3 gives the minimax power bound in the sense of Theorem 4.3.*

4.2 Near optimality of EWC tests

It is found in Section 3.2 that the so-called optimal EWC test nearly archives the efficiency bound when power is directed at the white noise (f_1). It is tempting to ask whether such findings remain in more realistic situations, especially given that efficiency bounds are derived in Section 4.1 when power is oriented toward non-flat alternatives.

I first maintain the Whittle diagonal structure with AR(1) $\underline{f}$ as before (coefficient ρ_0) but consider power directions at possibly non-monotonic $\tilde{f}_1$ of AR(2) processes with roots ρ_1 and ρ_2. In Panel (a) of Figure 6, I do not endogenize the test statistics for each (ρ_0, ρ_1, ρ_2) combination. More precisely, the boxplots are for the difference in weighted average powers of the φ^* and the new EWC tests that are initially designed to be (nearly) optimal at f_1 (as in Section 3) but are now at $\tilde{f}_1$ for various (ρ_1, ρ_2)'s. In contrast, Panel (b) leverages Corollary 4.4 to endogenize $\tilde{f}_1$ in obtaining the efficiency bound and in selecting q for the optimal EWC test. I note that for the existence of the optimal test in those cases I restrict

Figure 6: Boxplots of weighted average power differences between φ^* and EWC tests.



(a) Non-endogenized tests.

(b) Endogenized tests.

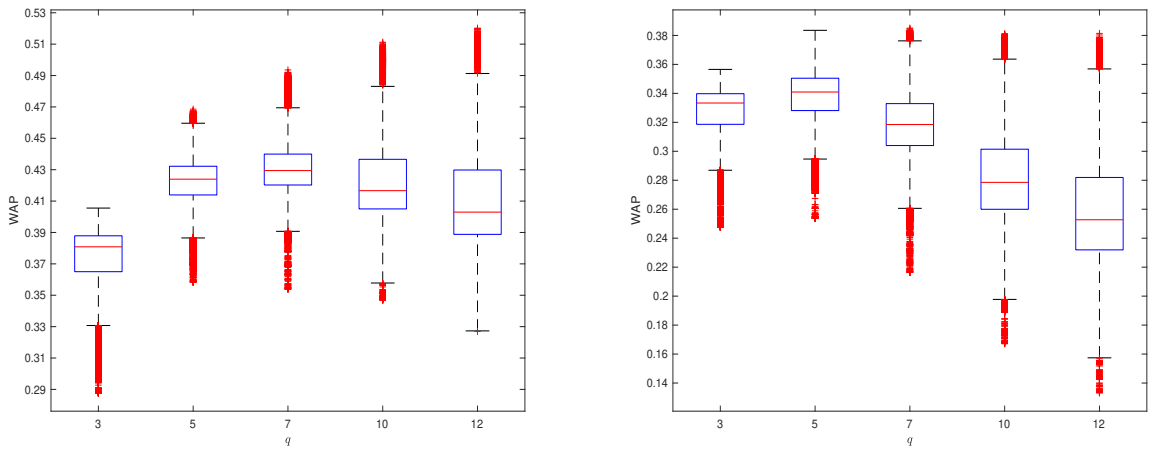
Notes: The "uniformly maximal" function of $\mathcal{F}$ corresponds to an AR(1) with coefficient $\rho_0 \in [0.8, 0.9]$. Powers are directed at AR(2) alternatives with roots ρ_1 and ρ_2 ($-0.8 \leq \rho_1(\rho_2) \leq 0.8$ in Panel (a); $0 \leq \rho_1 \leq 0.8$ and $-0.8 \leq \rho_2 \leq 0$ in Panel (b)). Sample size T is 100 and level of significance is 5%.

the ranges of ρ_1 and ρ_2 such that Assumption 4.2 (c) is satisfied, but the corresponding spectral density $\tilde{f}_1$ may still be non-monotone. Displayed results in Panel (a) suggest that the new EWC test is nearly as powerful as the test φ^* at AR(2) alternatives, even if both are derived under different rationales and neither possesses a well-defined sense of optimality by construction. Panel (b) corroborates the near optimality feature of the optimal EWC test when φ^* is by design optimal according to Corollary 4.4 at a restricted yet considerably large set of AR(2) alternatives. I note that the appearance of negative values in Figure 6 might be attributed to numerical errors in calculating the critical values for both tests.[13]

It is noted in *Comment* 3 of Section 3.1 that both the optimal and EWC tests belong to the so-called weighted cosine tests. I investigate whether the near optimality is a unique

[13]For numerical stability, I set the tolerance level to be 0.003 around 0.05 in order to obtain the critical values, so it is not surprising to have numerical errors in approximating rejection probabilities to be of the order 10^{-3}. With a smaller tolerance level, the numerical integration of (12) using 2000-point Gaussian quadrature may still produce complex-valued numbers. Moreover, in obtaining cv_{q^*} for φ^*, because the positiveness conditions and thus the integral expression of (12) may not hold *a priori* for a given $\tilde{q}$, I first simulate 100000 φ^* under $\underline{f}$ to determine a preliminary q^* and then use a bisection method to obtain a more precise cv_{q^*} while checking the positiveness conditions. In the above senses, the plots in Figure 6 are bound by small numerical and simulation errors, and negative values are not evidence to dismiss the theories.

Figure 7: Boxplots of weighted average power for weighted cosine tests.



(a) $\rho_0 = 0.8$. (b) $\rho_0 = 0.9$.

Notes: The "uniformly maximal" function $\underline{f}$ corresponds to an AR(1) with root ρ_1. Powers are directed at AR(2) alternatives with roots ρ_1 and ρ_2 ($-0.7 \leq \rho_1(\rho_2) \leq 0.7$). 5% Tests using q weighted cosines are considered, in which 1000 sets of random positive weights are used for each (ρ_0, ρ_1, ρ_2). Sample size T is 100.

feature of EWC tests in that class. Specifically, I consider such tests using q weighted cosines but with 1000 sets of random positive weights that are normalized to sum to one. The critical values are obtained such that these tests exactly control size at AR(1) $\underline{f}$ with coefficient ρ_0. Figure 7 displays boxplots of the resulting weighted average power at AR(2) alternatives with roots ρ_1 and ρ_2. Note that q^* is by construction 7 and 5 when power is directed at f_1 in Panels (a) and (b), respectively. It is found that the dispersion of weighted average power is relatively small at q^* across all tests and all AR(2) alternatives, suggesting that there may exist a set of weighted cosine tests, including the EWC test, that nearly achieve the efficiency bounds. In fact, the EWC test included is not even the best one, as its q is not optimized yet. Furthermore, the comparisons of boxplots across different q hint that one might need to choose weights (test statistic) more judiciously when relatively large q is used, and the theoretical insights so far recommend the EWC test as a good candidate so long as the adjusted critical value is adopted.

4.3 Relaxation of the Whittle-type appoximation

The theoretical discussions so far are entirely based on the Whittle-type diagonal structure. For both theoretical interest and practical relevance, it is natural to ask whether the above insights on optimal HAR inference continue to hold without that structure. To that end, I maintain the criteria of weighted average power maximizing and scale invariance and still direct power at the flat spectrum f_1. The goal is to seek powerful tests as functions of $Y^s = Y/\sqrt{Y'Y}$ in the problem of

$$\begin{aligned} H_0^e : Y \sim \mathcal{N}\left(0, \Omega_0(f)\right), \ f \in \mathcal{F} \\ \text{against } H_{1,f_1}^e : Y \sim \mathcal{N}\left(0, \kappa T^{-1}\text{diag}(1, \kappa^{-1}, \ldots, \kappa^{-1})\right), \end{aligned} \tag{16}$$

where the superscript e denotes the exact model. Note that H_{1,f_1}^e is identical to H_{1,f_1}^d, because Ω_1 exactly becomes diagonal under f_1.

First of all, I note that it is, in general, difficult to derive the optimal test of (16). This is mainly due to the complicated manner by which f enters $\Omega_0(f)$. In this case, even if it is true that the least favorable distribution puts a point mass on some function $f^* \in \mathcal{F}$, its determination seems rather difficult. Despite so, 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. If the power of a valid ad hoc test 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 suggesting 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 3.2 (c). For every a, the likelihood ratio test of H_{0,Λ_a}'' against H_{1,f_1}'' yields a power bound. I 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 the efficiency of ad hoc tests is then the power of

$$\varphi_{\Lambda^\dagger,f_1} = 1\left[\left(Y'\Omega_0(f_{a^\dagger})Y\right)^{-1}\left(Y'\Omega_1(f_1)Y\right) > \text{cv}\right], \tag{17}$$

for some cv such that $E[\varphi_{\Lambda^\dagger,f_1}] = \alpha$ under H_{0,Λ_a}''. It turns out that the EWC test essentially achieves this bound, after appropriate critical value adjustment and optimally choosing q.

For an EWC test with a given q, the null rejection probability at a given f and critical

value cv now becomes

$$P\left(\left|\frac{Y_0}{\sqrt{\sum_{j=1}^{q} Y_j^2/q}}\right| \geq \text{cv}\right) = P\left(\frac{Z_0^2}{q^{-1}\,\text{cv}^2 \sum_{j=1}^{q} \lambda_j(f) Z_j^2} \geq 1\right), \tag{18}$$

in which $\{\lambda_j(f)\}_{j=1}^{q}$ are positive eigenvalues of $M(\text{cv}, q)\Omega_{0,q}(f)$ (normalized by the magnitude of the only negative eigenvalue),[14] where $\Omega_{0,q}(f)$ is the upper left $(q+1)\times(q+1)$ block matrix of $\Omega_0(f)$ and $M(\text{cv}, q) = \text{diag}\left(-1, \text{cv}^2/q^2, \text{cv}^2/q^2, \ldots, \text{cv}^2/q^2\right)$. By the same arguments in Section 3, (18) is maximized at f such that 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 $\underline{f}$ under Assumption 3.2, rendering the critical value adjustment generally infeasible. In spatial settings, Müller and Watson (2022) consider a parametric $\mathcal{F}$ with a well-defined bound on the parameter as a benchmark model to obtain feasible adjustment and then robustify the size-controlling property of the resulting EWC-type test in more general model classes using the above eigenvalue insights.

I, however, take a numerical approach: I approximate f as a linear combination $\hat{f}$ of basis functions, numerically search the weights such that resulting $\hat{f}$ maximizes (18) under an additional assumption that f is non-increasing over $[0, \pi]$,[15] and obtain the critical value $\text{cv}_q^{a,e}$ accordingly in the same way as in Section 3. See Appendix B for the computational details. I then proceed as in Section 3.2 to select q optimally. In the context of Tables 3 and 4 (AR(1) $\underline{f}$), it is found that the difference between cv_q^a and $\text{cv}_q^{a,e}$ is considerably small and the largest size distortion of 5% level EWC test using cv_q^a is 0.003 in the exact model. These numerical findings are considerably robust when different $\mathcal{F}$'s are considered.[16]

Tables 6 and 7 summarize the weighted average power of the optimal EWC test and the weighted average power bound induced by (17), paralleling the exercises in Tables 4 and 5, respectively. As can be seen, for most $\mathcal{F}$ considered, the optimal EWC test essentially achieves the corresponding weighted average power bound and thus possesses a notion of near optimality in the spirit of Lemma 1 in Elliott, Müller, and Watson (2015). 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 6 and under large C in Table 7) 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 the ad hoc test is inefficient.

[14] See Lemma 1(i) in Müller and Watson (2022) for the proof that there is only one negative eigenvalue. Note the sign difference between $M(\text{cv}, q)$ here and $\mathbf{D}(\text{cv})$ in their context.

[15] The benchmark models considered in Müller and Watson (2022) all satisfy this shape restriction.

[16] I refer interested readers to Tables 7, 12, 13, 14, and 15 in an earlier version of this paper (cf. Dou (2020)) for numerical details. I choose not to include them in this version for the sake of space.

Table 6: Weighted average power (WAP) bound 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 $\underline{f}$ 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 7: Weighted average power (WAP) bound and the WAP of the optimal EWC test.

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

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

The practical implications of using the EWC test from Section 3.2.2 remain. In the exact model and in the context of Figure 5, the conventional wisdom and the optimal EWC test continue to suggest using 3 and 6 for q, respectively, but the corresponding Student-t critical value has to be enlarged by a slightly higher factor of 1.15 in 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 $\mu = \delta T^{-1/2}(1 - \rho_1)^{-1}$ is considered with power directed at $\rho_1 = 0.8$, as in Figure 1. I reiterate the general takeaway here: One should use the EWC test with a larger q and appropriately enlarged critical values for more powerful HAR inference.

5 Practical Implementation

In this section, I discuss the practical implementation of the optimal EWC test in the location model and extend it to inference about a scalar parameter in regression models.

5.1 Location model

Recall that the above theoretical discussions suggest the advantage of using a larger q and adjusted critical value when implementing the EWC test, and this new test possesses a notion of near optimality under pre-specified efficiency criteria and a smoothness class

$\mathcal{F}$. As a practical matter, one might like to estimate the smoothness class $\mathcal{F}$, in particular the associated $\underline{f}$, from data. Unfortunately, the attempt is not useful in theory. This is because the (nearly) optimal tests depend on $\mathcal{F}$, and a "larger" $\mathcal{F}$ incorporating sampling uncertainties potentially leads to a lower power. Put differently, one cannot estimate $\mathcal{F}$ and still control size (cf. Pötscher (2002)).

But how to determine a reasonable $\underline{f}$ in actual implementations? Given the analogy between the optimal EWC test and the test considered in Sun (2014a) when $\underline{f}$ is parameterized in the local-to-unity form (see *Comment* 4 in Section 3), I follow Sun (2014a) and suggest the practitioners calibrate $\underline{f}$ in the following way when testing about the population mean of an observed scalar time series $\{y_t\}_{t=1}^T$. One first computes the OLS estimator $\hat{\rho} = \left(\sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}\right)^{-1} \left(\sum_{t=2}^T \hat{u}_{t-1}^2\right)$, where $\hat{u}_t = y_t - T^{-1}\sum_{t=1}^T y_t$ in the location model, and takes $\hat{c}$ that satisfies $\hat{c} = T(1-\hat{\rho})$. To reduce the randomness of $\hat{\rho}$ and hence $\hat{c}$, one can use $\tilde{\rho}$ that is closest to $\hat{\rho}$ in a discretized grid over $[0,1]$, and thus let $\hat{c} = T(1-\tilde{\rho})$. The $\underline{f}$ is then calibrated by $\tilde{\underline{f}}$ such that $\tilde{\underline{f}}(\pi j/T) = 1/(\pi^2 j^2 + \hat{c}^2)$.[17]

After obtaining a reasonable $\underline{f}$, one proceeds to appropriately choose q that maximizes (12) with $\{\zeta_j = (\mathrm{cv}_q^a)^2 \kappa^{-1} q^{-1}\}_{j=1}^q$, in which, for 5% testing, cv_q^a is computed by $B_{\alpha,q}\,\mathrm{cv}_q^{na}(\alpha)$ with cv_q^{na} being the Student-t_q critical value and the adjustment factor $B_{\alpha,q}$ read directly from Table 2. For a generic α level of significance, one solves for the corresponding $\mathrm{cv}_q^a(\alpha)$ by equating α to (12) with $\{\zeta_j = (\mathrm{cv}_q^a(\alpha))^2 \tilde{\underline{f}}(\pi j/T)/\tilde{\underline{f}}(0) q^{-1}\}_{j=1}^q$. The evaluations of (12) involve numerical integration but are straightforward to compute. In case the researcher comes with a choice of q *a priori* and concerns 5% testing, she shall simply appeal to Table 2 to obtain cv_q^a and proceed to the following step without further considerations of q.

With μ_0 being the hypothesized population mean under the null, the researcher computes $q+1$ cosine weighted averages $\{Y_0\}_{j=0}^q$ of $\{y_t\}_{t=1}^T$: $Y_0 = T^{-1}\sum_{t=1}^T (y_t - \mu_0)$ and $Y_j = T^{-1}\sqrt{2}\sum_{t=1}^T \cos(\pi j(t-1/2)/T) y_t$, $j = 1, 2, \ldots, q$, and reject the null hypothesis at level α if $|Y_0/\sqrt{q^{-1}\sum_{j=1}^q Y_j^2}| > \mathrm{cv}_q^a(\alpha)$.

5.2 Extension to regression models

Consider the linear regression model

$$R_t = x_t\beta + \mathbf{z}_t'\delta + \varepsilon_t,\ t = 1, \ldots, T,$$

[17] This calibration is also analogous to that of Müller and Watson (2022) in spatial settings via a local-to-unity type parametrization. Alternatively, one could choose the regularity of $\underline{f}$ via a rule of thumb similar to Fan-Gijbels or Silverman rules of thumb for local polynomial and density estimation, as considered in Armstrong and Kolesár (2020), and formally justify the procedure under auxiliary assumptions.

where β is the scalar parameter of interest, and $\mathbf{z}_t$ are additional controls in the regression. Let $\tilde{x}_t = x_t - \left(\sum_{t=1}^T x_t \mathbf{z}_t\right)\left(\sum_{t=1}^T \mathbf{z}_t\mathbf{z}_t\right)^{-1}\mathbf{z}_t$ be the residual from regressing x_t on $\mathbf{z}_t$. Let $\hat{\beta}(\hat{\varepsilon}_t)$ be the coefficient estimate (residual) from regressing R_t on $\tilde{x}_t$. Suppose $T^{-1}\sum_{t=1}^T \tilde{x}_t^2 \xrightarrow{p} \sigma^2_{\tilde{x}\tilde{x}}$ and $T^{-1/2}\sum_{t=1}^T \tilde{x}_t\varepsilon_t \Rightarrow \mathcal{N}(0,\sigma^2_{\tilde{x}\varepsilon})$, then $\sqrt{T}\left(\hat{\beta}-\beta\right) \Rightarrow \mathcal{N}(0,\omega^2)$, where $\omega^2 = \sigma^2_{\tilde{x}\varepsilon}/\sigma^4_{\tilde{x}\tilde{x}}$. It is not hard to see that the generated series

$$y_t = \hat{\beta} + \frac{\tilde{x}_t\hat{\varepsilon}_t}{T^{-1}\sum_{t=1}^T \tilde{x}_t^2} \approx \beta + \frac{\tilde{x}_t\varepsilon_t}{\sigma^2_{\tilde{x}\tilde{x}}} \tag{19}$$

behave approximately like observations from the location model (2) with β replacing μ and the persistence of $\tilde{x}_t\varepsilon_t$ being the main source of serial correlation that is concerned in HAR inference about β.

For conducting inference using OLS estimates with HAR standard errors in the above regression models, I suggest first generating the series (19) from preliminary OLS estimates and proceeding as in the location model in Section 5.1 to implement the optimal EWC test, in which the primitive $\underline{f}$ is calibrated using the persistence of $\tilde{x}_t\varepsilon_t$. It is, however, worth noting that the extension of the theoretical insights from the location model to regression models requires that $\tilde{x}_t\varepsilon_t$ is stationary and the partial sum of $\tilde{x}_t$ is roughly linear ($T^{-1}\sum_{t=1}^{rT}\tilde{x}_t^2 \approx r\sigma^2_{\tilde{x}}$ for $0 \leq r \leq 1$),[18] which may be implausible in some applications. In those situations, I refer the readers to, for instance, Müller and Watson (2023) and Ibragimov and Müller (2010), respectively, for valid inference approaches.

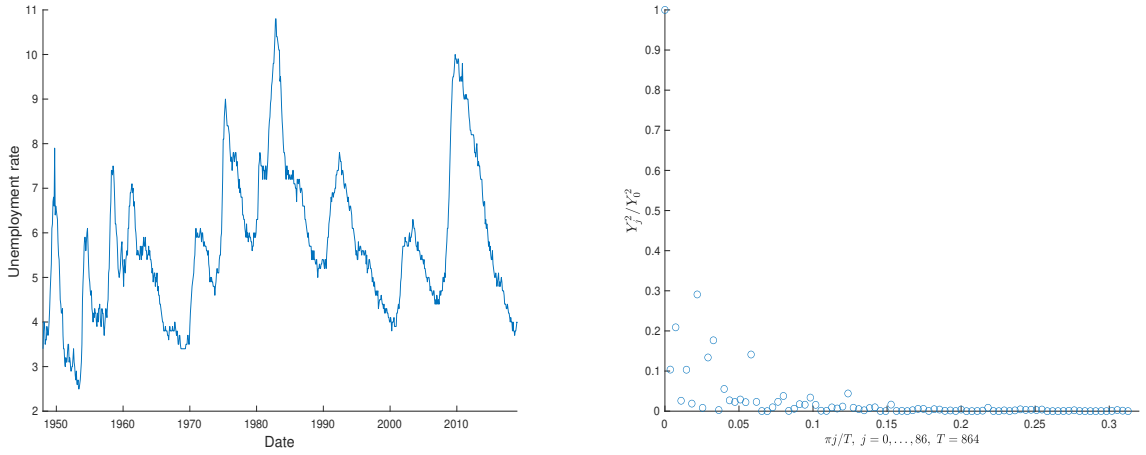
6 Empirical Illustrations

The purpose of this section is twofold. First, I demonstrate the empirical relevance of the choice of LRV estimator or HAR test statistic in confidence interval construction and hypothesis testing with autocorrelated data, even in simple settings. Second, I illustrate the implementation of the proposed optimal EWC test in these empirical examples.

I focus on intervals of the familiar "estimator plus and minus a standard error times a critical value" form. Relatedly, the HAR test statistic of concern is in the usual t-statistic form. In this context, the choice of LRV estimator/HAR test statistic amounts to selecting a combination of the standard error and the critical value. In particular, I consider Andrews's (1991) LRV estimator $\hat{\omega}^2_{\text{A91}}$ with a quadratic spectral kernel and bandwidth selection using an AR(1) model and critical value from standard normal distribution; Kiefer, Vogelsang, and Bunzel's (2000) Bartlett kernel estimator $\hat{\omega}^2_{\text{KVB}}$ with bandwidth equal to the sample

[18] See Section 5 in Müller (2014) for details.

Figure 8: U.S. unemployment rate.



(a) Time series plot.

(b) Squared weighted averages.

Notes: Series LNS14000000 of the Bureau of Labor Statistics from 1948:1 to 2019:12. The cosine weighted averages are normalized by Y_0^2 that is constructed using $\mu_0 = 4.5$ for illustration purpose only.

size and nonstandard critical value derived from their fixed-b asymptotics; 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; Lazarus, Lewis, Stock, and Watson's (2018) status quo recommendation for practice, EWC test using their suggested rule of thumb $q = 0.4T^{2/3}$ and Student-t_q critical value; EWC tests using ad hoc choices of q as 3, 9, and 12 and respective Student-t critical value; and my proposal in this paper, the optimal EWC test using adjusted critical value and optimally selected q.

In the first empirical example, I consider the simple task of constructing a 95% confidence interval for the population mean of U.S. unemployment. Figure 8 plots the monthly observations from January 1948 to December 2019, and the square of their 87 cosine weighted averages: $Y_0 = T^{-1}\sum_{t=1}^{T}(y_t - \mu_0)$ and $Y_j = T^{-1}\sqrt{2}\sum_{t=1}^{T}\cos(\pi j(t-1/2)/T)y_t,\ j = 1, 2, \ldots, 86$ using $\mu_0 = 4.5$ for illustration purpose only. Intuitively, most of the existing HAR inference approaches treat the first few averages as independent and identically distributed normals. But, as in Panel (b), it may be implausible in the current example. The theory developed in this paper, however, finds it advantageous to exploit the potential heteroskedasticity of these averages, which, to some extent, reflects the underlying persistence in the original series.

Table 8: 95% confidence intervals for U.S. unemployment population mean.

	$\hat{\omega}^2_{\text{A91}}$	$\hat{\omega}^2_{\text{KVB}}$	$\hat{\omega}^2_{\text{SPJ}}$	$\hat{\omega}^2_{\text{LLSW}}$	$\hat{\omega}^2_{\text{Y},3}$	$\hat{\omega}^2_{\text{Y},9}$	$\hat{\omega}^2_{\text{Y},12}$	$\hat{\omega}^2_{\text{OEWC}}$
m.e.	0.658	1.125	1.956	0.535	1.319	0.963	0.834	1.517

Notes: Same data as in Figure 8. All confidence intervals are symmetric around the sample mean $\hat{\mu} = 5.73$ with endpoints $\hat{\mu} \pm m.e.$, where the margin of error m.e. is reported in the table.

To proceed, I suggest practitioners fit an AR(1) to the series. The resulting OLS estimate $\hat{\rho}$ provides a reasonable measure of persistence, which is 0.9918 for U.S. unemployment. To reduce the effect of randomness, one then finds $\tilde{\rho}$ that is closest to $\hat{\rho}$ in a reasonably fine grid over $[0, 1]$, for which I choose 0.99. After that, one deploys a framework commonly used to model large sample persistence, namely the local-to-unity parameterization as $\rho = 1 - c/T$, to obtain $\hat{c} = T(1 - \tilde{\rho}) = 8.64$, a key ingredient to adjust the Student-t critical value to accommodate the above heteroskedasticity issue in the presence of strong autocorrelation. Lastly, one selects the q such that the resulting EWC test with the adjusted critical value is most powerful according to a well-defined efficiency criterion. The numerical determination of the adjustment and q are straightforward and detailed in Section 5.1 and I omit them here for brevity. In case the researcher comes with a choice of q *a priori*, she shall simply refer to Table 2 to obtain the adjusted critical value and proceed without considering q further.

Table 8 lists the margin of error in symmetric confidence intervals of the U.S. unemployment population mean using the above eight methods. Interestingly, the lengths of these intervals vary by more than a factor of three, which demonstrates the empirical relevance of the choice of LRV estimator/HAR test with strongly autocorrelated time series data. Moreover, as q increases, the margin of error for conventional EWC tests shrinks. Thus, it is not surprising to see that Lazarus, Lewis, Stock, and Watson's (2018) EWC test using $q = 36$ leads to the shortest length. Yet, the resulting confidence interval may have a coverage rate below 95%, since their test is primarily designed to accommodate persistence no larger than 0.7. In contrast, the optimal EWC test uses $q = 4$ but inflates the Student-t_4 critical value by a factor of 1.33, which, to some extent, adapts to the underlying persistence and achieves efficiency in some testing-optimal sense.

Now, let's consider the problem of comparing the predictive accuracy of inflation forecasts. In particular, consider a benchmark forecast sequence $(F_{0,t})_{t=1,\dots,T}$ and J competing forecasts $(F_{j,t})_{t=1,\dots,T}$, $1 \leq j \leq J$. The measure of performance of the benchmark method relative to the jth competitor is usually defined as $y_{j,t} \equiv L(F^\dagger_t, F_{j,t}) - L(F^\dagger_t, F_{0,t})$ with $L(\cdot, \cdot)$ be-

ing a pre-defined loss function. The problem of testing unconditional equal predictive ability (UEPA) for these two forecasts is formalized as testing the hypothesis $H_0^{\text{UEPA}} : E[y_{j,t}] = 0$. The arguably most popular method to do the task is Diebold and Mariano's (1995) test, which is essentially a t-test with an appropriate choice of LRV estimator and critical value, as in the first empirical example.

I follow Li, Liao, and Quaedvlieg (2021) to select eight inflation forecasting methods from those constructed by Medeiros, Vasconcelos, Veiga, and Zilberman (2021) for the one-month-ahead and twelve-month-ahead forecasts of Consumer Price Index (CPI). It includes four traditional methods (random walk, AR(1), Bayesian vector-autoregression (BVAR), and factor model) and four other machine-learning-based methods (LASSO, elastic net (ElNet), bagging, and random-forecast regression (RF-OLS)). Table 9 lists the counts of null rejections using the sample period January 1990 to December 2004 and differentials in quadratic losses among the eight methods, with each column being the benchmark. It turns out the choice of test again matters empirically in evaluating inflation forecasts.

For the implementation of the optimal EWC test, take the comparison of twelve-month-ahead forecasts between LASSO and RF-OLS as an example. A similar calculation as in the unemployment example first finds that the underlying persistence is only 0.125, yet the four popular tests under consideration have not reached a unanimous conclusion. Moreover, even if one dogmatically follows the practical guide (and uses the local-to-unity parameterization that is otherwise meant for strong persistence modeling), the corresponding optimal choice of $q = 14$ is close to Lazarus, Lewis, Stock, and Watson's (2018) rule of thumb suggestion ($q = 12$) and the critical value adjustment factor is close to 1 (1.006). This demonstrates the usefulness of the optimal EWC test even with weakly dependent data, especially given that Lazarus, Lewis, Stock, and Watson's (2018) EWC test is also theoretically justified for such data in Lazarus, Lewis, and Stock (2021).

A further investigation of the twelve-month-ahead forecasts between LASSO and RF-OLS using a rolling window (of the same width 180 months) over other time periods, however, reveals that the testing conclusions may not be time consistent for the same test, including Lazarus, Lewis, Stock, and Watson's (2018) EWC test and the optimal EWC test proposed in this paper. This somewhat raises the necessity to study finer testing problems to evaluate competing inflation forecasts, such as testing *conditional equal predictive ability (CEPA)* as in Giacomini and White (2006) and *conditional superior predictive ability (CSPA)* as in Li, Liao, and Quaedvlieg (2021). Alternatively, one can insist on testing UEPA but need valid HAR inference methods even under possible nonstationarity. I leave it for future research.

Table 9: Rejection counts of UEPA among competing inflation forecasts.

	RW	AR	BVAR	LASSO	ElNet	Factor	Bagging	RF-OLS
$\hat{\omega}^2_{\text{A91}}$	$(6,6)$	$(4,3)$	$(2,3)$	$(4,3)$	$(4,3)$	$(4,6)$	$(1,2)$	$(5,6)$
$\hat{\omega}^2_{\text{KVB}}$	$(6,1)$	$(1,1)$	$(0,2)$	$(1,1)$	$(1,0)$	$(1,3)$	$(1,0)$	$(1,4)$
$\hat{\omega}^2_{\text{SPJ}}$	$(6,3)$	$(3,3)$	$(0,2)$	$(3,2)$	$(3,2)$	$(4,6)$	$(1,2)$	$(4,6)$
$\hat{\omega}^2_{\text{LLSW}}$	$(5,3)$	$(3,2)$	$(0,2)$	$(2,2)$	$(3,2)$	$(3,4)$	$(1,1)$	$(3,6)$

Notes: Monthly inflation forecasts are constructed by Medeiros, Vasconcelos, Veiga, and Zilberman (2021). The period used for illustration is 1990:1 to 2004:12. Each cell contains $(\#1, \#2)$ where #1 and #2 are counts of rejections of UEPA between the benchmark (column) and other 7 forecasting methods, using one-month-ahead and twelve-month-ahead forecasts.

Appendix A Proofs in Sections 3.1 and 4.1

I first note that due to the scale invariance restriction, it is without loss of generality to impose the normalization on f such that $f(0) = 1$ throughout the proofs. Moreover, I note that, for generic f and f' and borrowing the definition of H'' from Section 2, the likelihood ratio test of $H''_{0,f}$: "Y^s has density $h_{0,f}$" against $H''_{1,f'}$: "Y^s has density $h_{1,f'}$" only depends on $\{f(\pi j/T)/f'(\pi j/T)\}_{j=1}^{T-1}$ under Assumption 3.1, where the density form of h is in (6). Furthermore, define function $m(\cdot)$ such that $m(\pi j/T) = f(\pi j/T)/f'(\pi j/T)$. For $f \in \mathcal{F}$ and $f' \in \mathcal{G}$ and under Assumption 4.2(a,b), denotes the set of such m by $\mathcal{M}$, and note that the function $\underline{m} = \underline{f}/\overline{f}$ traces out the largest relative persistence among all pairs of f and f' or is the "uniformly maximal" m of $\mathcal{M}$ in the sense of Assumption 3.2(a) or 4.1(a). Also, Assumption 4.2(c) assumes the monotonicity of $\underline{m}$ in the exact sense of Assumption 3.2(b). Having these features of $\underline{m}$ in the background, I prove some useful auxiliary lemmas below.

I introduce some additional notations. Let $\{Z_j\}_{j=0}^n$ be i.i.d. standard normals 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 each $1 \leq \tilde{q} \leq T-1$, and $0 < \alpha < 1$, define $\text{cv}_{\tilde{q}}$ such that

$$P\left((1-\text{cv}_{\tilde{q}})Z_0^2 > \sum_{j=1}^{\tilde{q}} \left[\text{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2\right) = \alpha, \tag{20}$$

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

$$\max_{j=1,2,\ldots,\tilde{q}} \left\{\text{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right\} < 0 \text{ or } \min_{j=1,2,\ldots,\tilde{q}} \left\{\text{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right\} > 0. \tag{21}$$

Further, for $\zeta, \xi \in \mathbb{R}^n_+$, we call ξ weakly majorizes ζ ($\xi \succ_w \zeta$) iff $\sum_{j=1}^k \xi_j \geq \sum_{j=1}^k \zeta_j$ for $1 \leq k \leq n$. If, in addition, $\sum_{j=1}^n \xi_j = \sum_{j=1}^n \zeta_j$, we say that ξ majorizes ζ ($\xi \succ \zeta$).

A.1 Auxiliary lemmas

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

Proof. For any $\zeta \in \mathcal{D}^+(\zeta^*)$, 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 $\zeta \in \mathcal{D}^-(\zeta^*)$, 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* $\zeta^* \in \mathbb{R}^n_+$, $\mathcal{D}^+(\zeta^*)$ *and* $\mathcal{D}^-(\zeta^*)$ *are as defined in Lemma A.1. Then* $K_n(\zeta^*) \leq K_n(\zeta)$ *for any* $\zeta \in \mathcal{D}^+(\zeta^*)$ *and* $K_n(\zeta^*) \geq K_n(\zeta)$ *for any* $\zeta \in \mathcal{D}^-(\zeta^*)$. *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 flawed; 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* $J(\xi) < J(\zeta) = 1$. *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.

$$J_n(\zeta) = \frac{1}{\pi}\int_0^\infty \frac{dt}{\sqrt{t}(1+t)\sqrt{\prod_{j=1}^n (1+(1+t)\zeta_j)}} = \frac{1}{\pi}\int_0^1 \frac{s^{(n-1)/2}ds}{\sqrt{(1-s)\prod_{j=1}^n (s+\zeta_j)}}$$
$$= \frac{2}{\pi}\int_0^1 \frac{(1-u^2)^{(n-1)/2}du}{\sqrt{\prod_{j=1}^n (1-u^2+\zeta_j)}}, \tag{22}$$

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

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

Proof. (a) In (20) at $\tilde{q} = 1$, if $\underline{m}(\pi/T) = \kappa^{-1}$, there does not exist a cv_1 and α such that (20) holds. On the other hand, a rearrangement of the event in (20) at $\tilde{q} = 1$ gives

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

It follows that $\underline{m}(\pi/T) \lesseqgtr \kappa^{-1}$ if and only if $1 \lesseqgtr \mathrm{cv}_1$. Moreover, if $\underline{m}(\pi/T) > \kappa^{-1}$, the second part of Lemma A.4 in conjunction with (20) at $\tilde{q} = 1$ gives

$$\mathrm{cv}_1 = \frac{1}{\kappa\underline{m}(\pi/T)\sin^2(\alpha\pi/2) + \cos^2(\alpha\pi/2)},$$

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

$$\mathrm{cv}_1 = \frac{1}{\kappa\underline{m}(\pi/T)\cos^2(\alpha\pi/2) + \sin^2(\alpha\pi/2)},$$

which always exists. Thus, cv_1 exists if and only if $\underline{m}(\pi/T) \neq \kappa^{-1}$. (b) follows from the above that $cond_1$ holds if $\underline{m}(\pi/T) \neq \kappa^{-1}$. ■

In what follows, I fold most of the parts corresponding to $\mathrm{cv}_q > 1$ in the proofs. Partly, this is because they can be worked out by exactly symmetric arguments. Also, the most relevant results from these auxiliary lemmas in establishing Part(2) of Theorems 3.3, 4.3 and Corollary 4.4 are when $\kappa > \underline{m}(\pi/T)^{-1}$, which is equivalent to $\mathrm{cv}_q < 1$ for all q by Lemmas A.5 and A.10.

Lemma A.6 *For $\underline{m}(\pi/T) \neq \kappa^{-1}$ 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

$$\max_{j=1,2,\ldots,\tilde{q}} \{\mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\} \geq 0 \text{ and } \min_{j=1,2,\ldots,\tilde{q}} \{\mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\} \leq 0. \tag{23}$$

Consider $\min_{j=1,2,\ldots,\tilde{q}+1} \{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\} > 0$. We must have $\mathrm{cv}_{\tilde{q}+1} < 1$; otherwise, (20) does not hold at $\tilde{q}+1$. On the other hand, $0 < \min_{j=1,2,\ldots,\tilde{q}+1} \{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\} \leq \min_{j=1,2,\ldots,\tilde{q}} \{\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\}$. This, in conjunction with the second part of (23), 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 $A_{\tilde{q}}^+ = \{j | 1 \leq j \leq \tilde{q},\ \mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1 > 0\}$ ($A_{\tilde{q}}^+ \neq \emptyset$; otherwise, (20) is violated for $\tilde{q}$). Now (20) at $\tilde{q}$ gives

$$\begin{aligned}
\alpha &= P\left((1-\mathrm{cv}_{\tilde{q}}) Z_0^2 > \sum_{j=1}^{\tilde{q}} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 \right) \\
&= P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 + \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j \notin A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 \right) \\
&\geq P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 \right) \qquad (24) \\
&> P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 \right) \qquad (25) \\
&> P\Bigg(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 \\
&\qquad\qquad + \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j \notin A_{\tilde{q}}^+} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 \Bigg) \qquad (26) \\
&> P\left(Z_0^2 > \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \sum_{j=1}^{\tilde{q}+1} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\right] Z_j^2 \right) = \alpha,
\end{aligned}$$

where (24) 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 (25) is due to Lemma A.1 and the fact that for any $j \in A_{\tilde{q}}^+$, $\frac{1}{1-\mathrm{cv}_{\tilde{q}}} \left[\mathrm{cv}_{\tilde{q}}\, \kappa \underline{m}(\pi j/T) - 1\right] < \frac{1}{1-\mathrm{cv}_{\tilde{q}+1}} \left[\mathrm{cv}_{\tilde{q}+1}\, \kappa \underline{m}(\pi j/T) - 1\right]$ under $\mathrm{cv}_{\tilde{q}} < \mathrm{cv}_{\tilde{q}+1} < 1$. The inequality (26) is due to Lemma A.1.

The other half of (23) corresponds to $\text{cv}_{\tilde{q}+1} > 1$ and the proof is exactly symmetric to the above. Overall, we have for $\underline{m}(\pi/T) \neq \kappa$ 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 $\tilde{q}+1 \leq q \leq T-1$ by inductions. ■

Corollary A.7 *For $\underline{m}(\pi/T) \neq \kappa^{-1}$ 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 $\underline{m}(\pi/T) \neq \kappa^{-1}$ 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 $1 \leq q \leq T-1$. 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 $1 \leq q^* \leq T-3$. ■

Corollary A.9 *If $\underline{m}$ equals 1 and $0 < \alpha < 1$, then $cond_q$ holds for all $1 \leq q \leq T-1$.*

Proof. In this case, $\mathcal{F} = \mathcal{G} = \{\underline{m}\} = \{f_1\}$ and $cond_{T-1}$ is trivially satisfied. It follows that $cond_q$ holds for all $1 \leq q \leq T-1$ by Corollary A.7. ■

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

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

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

Proof. Lemma A.5 leads to the conclusions for $q^* = 1$. We now focus on $q^* \geq 2$. Suppose $\kappa\underline{m}(\pi/T) > 1$, then Lemma A.5 implies that $\text{cv}_1 < 1$. Suppose there exists a $\tilde{q} = \min\{q | 2 \leq q \leq q^*,\ \text{cv}_q > 1\}$. (I note that cv_q cannot be 1 for any $q \leq q^*$; otherwise, (20) cannot hold at the corresponding q.) Then we must have

$$\begin{aligned}\max_{j=1,2,\ldots,\tilde{q}-1}\{\text{cv}_{\tilde{q}-1}\,\kappa\underline{m}(\pi j/T) - 1\} &< \max_{j=1,2,\ldots,\tilde{q}-1}\{\text{cv}_{\tilde{q}}\,\kappa\underline{m}(\pi j/T) - 1\} \\ &\leq \max_{j=1,2,\ldots,\tilde{q}}\{\text{cv}_{\tilde{q}}\,\kappa\underline{m}(\pi j/T) - 1\} < 0.\end{aligned}$$

This is a contradiction, because $\min_{j=1,2,\ldots,\tilde{q}-1}\{\mathrm{cv}_{\tilde{q}-1}\,\kappa\underline{m}(\pi j/T)-1\}>0$. It subsequently implies that $\mathrm{cv}_q<1$ for all $1\le q\le q^*$. Moreover, for each $j=1,\ldots,q^*$, $R_j(x)=\left[\kappa\underline{m}(\pi j/T)x-1\right]/(1-x)$ is monotonically increasing in $(0,1)$. (To see this, note that $\kappa\underline{m}(\pi j/T)-1>\kappa\underline{m}(\pi j/T)\,\mathrm{cv}_{q^*}-1>0$ for every $1\le j\le q^*$.) For (20) to hold sequentially, we necessarily need $\mathrm{cv}_{q+1}<\mathrm{cv}_q,\ q=1,2,\ldots,q^*-1$. (Otherwise, the LHS of (20) would always be below α by Lemma A.1.) Part (b) is proved, and Part (a) holds by symmetric arguments. ■

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

Proof. Define $Q(x,q)=P\left((1-x)Z_0^2>\sum_{j=1}^{q}\left[x\kappa\underline{m}(\pi j/T)-1\right]Z_j^2\right)$. Given $\underline{m}(\pi j/T)\ge\underline{m}(\pi(j+1)/T),\ j=q^*+1,\ldots,T-2$, it suffices to show $\kappa^{-1}\left(\underline{m}(\pi(q^*+1)/T)\right)^{-1}\ge\mathrm{cv}_{q^*}$. Suppose not; then we must have $\kappa^{-1}\left(\underline{m}(\pi(q^*+1)/T)\right)^{-1}<\mathrm{cv}_{q^*}$. Suppose $\kappa\underline{m}(\pi/T)>1$, then $\mathrm{cv}_{q^*}<1$ by Lemma A.10.

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

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

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

where the last but one inequality follows from the fact that $Q(\cdot,q^*)$ is monotonically decreasing in $(0,\mathrm{cv}_{q^*})$ under $\kappa\underline{m}(\pi/T)>1$. By the continuity of $Q(\cdot,q^*+1)$ 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. Suppose $\kappa\underline{m}(\pi/T)<1$ instead, the proof follows by almost symmetric arguments as above. Overall, we have $\kappa^{-1}\left(\underline{m}(\pi(q^*+1)/T)\right)^{-1}=\min_{q^*+1\le j\le T-1}\kappa^{-1}\left(\underline{m}(\pi j/T)\right)^{-1}\ge\mathrm{cv}_{q^*}$. ■

A.2 Proof of Theorem 4.3

Part 1 holds by the definition of $\beta_{\Lambda_0,\Lambda_1}$ and by simply recognizing that the alternative $H^d_{1,\underline{f}}$ (defined analogous to H^d_{1,f_1}) is included in the null H^d_0. Any nontrivial size-controlling test thus cannot be more powerful than the trivial randomized test, which does not depend on any Λ_0 nor Λ_1.

In part 2, $\underline{m}(\pi/T) > \kappa^{-1}$. Under Assumption 4.2(a,b,c) 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 $q^* = T-1$. I conjecture that the pair of least favorable distributions $(\Lambda_0^*, \Lambda_1^*)$ put probability masses on functions $\{f^*\} \subset \mathcal{F}$ and $\{g^*\} \subset \mathcal{G}$, in which $f^*(\phi) = \underline{f}(\phi)\mathbf{1}[|\phi| \leq \pi q^*/T] + a^*(\phi)\underline{f}(\phi)\mathbf{1}[|\phi| > \pi q^*/T]$ and $g^*(\phi) = \overline{f}(\phi)\mathbf{1}[|\phi| \leq \pi q^*/T] + b^*(\phi)\overline{f}(\phi)\mathbf{1}[|\phi| > \pi q^*/T]$, for $a^*(\phi) \geq 1$ and $b^*(\phi) \leq 1$ such that $a^*(\phi)\underline{f}(\phi)/(b^*(\phi)\overline{f}(\phi)) = (\kappa\,\mathrm{cv}_{q^*})^{-1}$ for $|\phi| > \pi q^*/T$. Assumption 4.2 (d,e) ensures that such two sets of functions are non-empty as long as the pair of functions (a^*, b^*) exists. This true by Lemma A.11 ($\underline{m}(\phi) \leq (\kappa\,\mathrm{cv}_{q^*})^{-1}$ for $|\phi| > \pi q^*/T$).

Now first let $b^*(\phi) = 1$ for all ϕ and $a^*(\phi) = (\kappa\,\mathrm{cv}_{q^*}\,\underline{m}(\phi))^{-1}$. The best level α test of H^d_{0,f^*} against H^d_{1,g^*} is

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

for some $\mathrm{cv} \geq 0$ such that $E_{P_{Y,f^*}}[\varphi_{f^*,g^*}(Y^s)] = \alpha$, where $P_{Y,\tilde{f}}$ denotes the joint distribution of Y at $f = \tilde{f}$ under H^d_0. It follows that

$$\begin{aligned}
\alpha &= P_{Y,f^*}\left(\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T)}{Y_0^2 + \kappa\sum_{j=1}^{T-1} Y_j^2/g^*(\pi j/T)} > \mathrm{cv}\right) \\
&= P_{Y,f^*}\left(Y_0^2 + \sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T) > \mathrm{cv}\left(Y_0^2 + \kappa\sum_{j=1}^{T-1} Y_j^2/g^*(\pi j/T)\right)\right) \\
&= P\left((1-\mathrm{cv})\,Z_0^2 > \sum_{j=1}^{T-1}\left[\mathrm{cv}\,\kappa f^*(\pi j/T)/g^*(\pi j/T) - 1\right]Z_j^2\right) \\
&= P\left((1-\mathrm{cv})\,Z_0^2 > \sum_{j=1}^{q^*}\left[\mathrm{cv}\,\kappa\underline{m}(\pi j/T) - 1\right]Z_j^2 + \sum_{j=q^*+1}^{T-1}\left[\mathrm{cv}\,/\,\mathrm{cv}_{q^*} - 1\right]Z_j^2\right), \qquad (27)
\end{aligned}$$

where the last equality follows from the definition of f^* and g^*. Because Y is a continuous random vector, the critical value cv is unique. By matching (27) with (20) at $\tilde{q} = q^*$, we have $\mathrm{cv} = \mathrm{cv}_{q^*}$. Also, the events $\left\{\frac{Y_0^2+\sum_{j=1}^{T-1} Y_j^2/f^*(\pi j/T)}{Y_0^2+\kappa\sum_{j=1}^{T-1} Y_j^2/g^*(\pi j/T)} > \mathrm{cv}_{q^*}\right\}$ and $\left\{\frac{Y_0^2+\sum_{j=1}^{q^*} Y_j^2/\underline{f}(\pi j/T)}{Y_0^2+\kappa\sum_{j=1}^{q^*} Y_j^2/\overline{f}(\pi j/T)} > \mathrm{cv}_{q^*}\right\}$

are equivalent $P_{Y,\tilde{f}}$-almost surely, uniformly in $\tilde{f} \in \mathcal{F}$. The rejection regions defined by φ_{f^*,g^*} and the optimal test statistic in (14) are thus identical.

It remains to check the following conditions: (1) φ_{f^*,g^*} is also the best level α test of H^d_{0,Λ^*_0} against H^d_{1,Λ^*_1}; (2) φ_{f^*,g^*} uniformly controls size H^d_0 and has its largest size distortion at $\underline{f}$; (3) φ_{f^*,g^*} attains $\beta_{\Lambda^*_0,\Lambda^*_1}$ at $\overline{f}$; and (4) for any other (Λ_0, Λ_1), $\beta_{\Lambda^*_0,\Lambda^*_1} \leq \beta_{\Lambda_0,\Lambda_1}$.

For (1), note that φ_{f^*,g^*} is of exact size α under H_{0,Λ^*_0}: $\int \varphi_{f^*,g^*}(y^s) \int h_{0,f}(y^s) d\Lambda^*_0(f) dy^s = \int \varphi_{f^*,g^*}(y^s) h_{0,f^*}(y^s) dy^s \int h_{0,f} d\Lambda^*_0(f) = E_{P_{Y,f^*}}[\varphi_{f^*,g^*}(Y^s)] = \alpha$, where the second equality holds because the set of distributions of φ_{f^*,g^*} is degenerate under Λ^*_0. By the same logic, the rejection probabilities of φ_{f^*,g^*} under H_{1,g^*} and H_{1,Λ^*_1} are identical. Because the best level α test of H^d_{0,Λ^*_0} against H^d_{1,Λ^*_1} is unique, (1) holds.

For (2), consider a given $\tilde{f} \in \mathcal{F}$,

$$
\begin{aligned}
E_{P_{Y,\tilde{f}}}[\varphi_{f^*,g^*}(Y)] &= P_{Y,\tilde{f}}\left(\frac{Y_0^2 + \sum_{j=1}^{q^*} Y_j^2 / \underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2 / \overline{f}(\pi j/T)} > \mathrm{cv}_{q^*} \right) \\
&= P\left([1 - \mathrm{cv}_{q^*}] Z_0^2 > \sum_{j=1}^{q^*} [\mathrm{cv}_{q^*} \kappa \underline{m}(\pi j/T) - 1] \frac{\tilde{f}(\pi j/T)}{\underline{f}(\pi j/T)} Z_j^2 \right) \\
&= P\left(Z_0^2 > \frac{1}{1 - \mathrm{cv}_{q^*}} \sum_{j=1}^{q^*} [\mathrm{cv}_{q^*} \kappa \underline{m}(\pi j/T) - 1] \frac{\tilde{f}(\pi j/T)}{\underline{f}(\pi j/T)} Z_j^2 \right) \qquad (28) \\
&\leq P\left(Z_0^2 > \frac{1}{1 - \mathrm{cv}_{q^*}} \sum_{j=1}^{q^*} [\mathrm{cv}_{q^*} \kappa \underline{m}(\pi j/T) - 1] Z_j^2 \right) = \alpha,
\end{aligned}
$$

where (28) follows from (b) in Lemma A.10 under the condition $\underline{m}(\pi/T) > \kappa^{-1}$, and the inequality follows from the definition of q^* and Lemma A.1 under Assumption 4.2 (a).

For (3), consider a given $\tilde{g} \in \mathcal{G}$ and let $P^1_{Y,\tilde{g}}$ denote the joint distribution of Y under $H^d_{1,\tilde{g}}$

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

where the inequality follows from Corollary A.2 under Assumption 4.2 (b). In this sense, $\beta_{\Lambda^*_0,\Lambda^*_1} = E_{P^1_{Y,\overline{f}}}[\varphi_{f^*,g^*}(Y)]$.

For (4), because φ_{f^*,g^*} uniformly controls size under H_0^d, it also controls size under H_{0,Λ_0}^d, then by the definition of $\beta_{\Lambda_0,\Lambda_1}$,

$$\beta_{\Lambda_0,\Lambda_1} \geq \int \varphi_{f^*,g^*}(y^s) \int h_{1,\tilde{g}}(y^s) d\Lambda_1(\tilde{g}) dy^s \geq \inf_{\tilde{g}\in\mathcal{G}} E_{P^1_{Y,\tilde{g}}}\left[\varphi_{f^*,g^*}(Y)\right] = \beta_{\Lambda_0^*,\Lambda_1^*}.$$

A.3 Proofs of Theorem 3.3 and Corollary 4.5

Proof of Corollary 4.5 follows immediately from those in Section A.2 with $\overline{f} = f_1$. Theorem 3.3 is a special case of Corollary 4.5 with $\mathcal{G} = \{f_1\}$, and its results follow by realizing that $\beta_{\Lambda_0^*,\Lambda_1^*}$, in that case, is simply the weighted average power of test (8) at f_1.

A.4 Proof of Corollary 4.4

Corollary 4.4 is a special case of those considered in Theorem 4.3 with $\mathcal{G} = \{\tilde{f}_1\}$. The proof thus follows directly from those in Section A.2.

Appendix B Computational details in Section 4.3

In this section, I explain in detail how to numerically identify the null rejection probability maximizer of the EWC test in testing (16).

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 by (12), I approximate (18) by

$$P\left(\frac{Z_0^2}{q^{-1}\,\mathrm{cv}^2 \sum_{j=1}^q \lambda_j(\hat{f}) Z_j^2} \geq 1\right) = \frac{2}{\pi}\int_0^1 \frac{(1-u^2)^{(q-1)/2}\mathrm{d}u}{\sqrt{\prod_{j=1}^q \left(1 - u^2 + q^{-1}\,\mathrm{cv}^2\,\lambda_j\left(\hat{f}\right)\right)}}, \tag{29}$$

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 $\{\Omega_0(\varsigma_i)\}_{i=0}^n$, the computation of (29) takes very little computing time for each $\hat{f}$, and it is feasible to obtain a global maximizer of (29) subject to implied constraints on $(f(x_1), f(x_2), \ldots, f(x_n))'$ from a given $\mathcal{F}$. I additionally assume that the underlying spectrum is non-increasing over $[0,\pi]$.

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

$$\varsigma_i(x) = \begin{cases} (x - x_{i-1})/h_i & , \text{ if } x \in I_i, \\ (x_{i+1} - x)/h_{i+1} & , \text{ if } x \in I_{i+1}, \\ 0 & , \text{ otherwise.} \end{cases} \tag{30}$$

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 (29), which involves $\Omega_0(f)$ implicitly through $\lambda_j(\hat{f})$. In an unreported exercise, 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 differences in the rejection probabilities are at most of the order 0.0001. I thus hold on to the above choice of n and $\{x_i\}_{i=0}^{50}$.

I proceed in three steps to identify the null rejection probability maximizer for a fixed EWC test of (16), 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 (29), 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 $(f(x_1), f(x_2), \ldots, f(x_n))'$ such that each vector corresponds to 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.

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.

Armstrong, T. B., and M. Kolesár (2020): "Simple and honest confidence intervals in nonparametric regression," *Quantitative Economics*, 11(1), 1–39.

Atchadé, 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.

Choudhuri, N., S. Ghosal, and A. Roy (2004): "Contiguity of the Whittle measure for a Gaussian time series," *Biometrika*, 91(1), 211–218.

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.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3), 134–144.

Dou, L. (2020): "Optimal HAR Inference," Working Paper, CUHK-Shenzhen.

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.

Giacomini, R., and H. White (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74(6), 1545–1578.

Golubev, G. K., M. Nussbaum, and H. H. Zhou (2010): "Asymptotic equivalence of spectral density estimation and Gaussian white noise," *The Annals of Statistics*, 38(1), 181–214.

Gonçalves, S., and T. J. Vogelsang (2011): "Block Bootstrap HAC Robust Tests: The Sophiscation of the Naive Bootstrap," *Econometric Theory*, 27(4), 745–791.

Hwang, J., and Y. Sun (2017): "Asymptotic F and t tests in an efficient GMM setting," *Journal of Econometrics*, 198(2), 277–295.

——— (2018): "Should we go one step further? An accurate comparison of one-step and two-step procedures in a generalized method of moments framework," *Journal of Econometrics*, 207(2), 381–405.

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(6), 1130–1164.

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 (2021): "The Size-Power Tradeoff in HAR Inference," *Econometrica*, 89(5), 2497–2516.

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.

Lehmann, E. L., and C. Stein (1948): "Most Powerful Tests of Composite Hypotheses. I. Normal Distributions," *The Annals of Mathematical Statistics*, 19(4), 495 – 516.

Li, J., Z. Liao, and R. Quaedvlieg (2021): "Conditional Superior Predictive Ability," *The Review of Economic Studies*, 89(2), 843–875.

Medeiros, M. C., G. F. R. Vasconcelos, A. Veiga, and E. Zilberman (2021): "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods," *Journal of Business & Economic Statistics*, 39(1), 98–119.

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 & Economic Statistics*, 32, 311–322.

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

Müller, U. K., and M. W. Watson (2022): "Spatial Correlation Robust Inference," *Econometrica*, 90(6), 2901–2935.

——— (2023): "Spatial Correlation Robust Inference in Linear Regression and Panel Models," *Journal of Business & Economic Statistics*, 41(4), 1050–1064.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

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.

Preinerstorfer, D., and B. M. Pötscher (2016): "On Size and Power of Heteroskedasticity and Autocorrelation Robust Tests," *Econometric Theory*, 32(02), 261–358.

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.

Pötscher, B. M., and D. Preinerstorfer (2018): "Controlling the Size of Autocorrelation Robust Tests," *Journal of Econometrics*, 207(2), 406–431.

——— (2019): "Further results on size and power of heteroskedasticity and autocorrelation robust tests, with an application to trend testing," *Electronic Journal of Statistics*, 13(2), 3893–3942.

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.

——— (2014a): "Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation," in *Advances in Econometrics*, ed. by Y. Chang, T. B. Fomby, and J. Y. Park, vol. 33, pp. 23–63. Emerald Group Publishing Limited.

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

——— (2014c): "Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference," *Journal of Econometrics*, 178, 659–677.

Sun, Y., and D. M. Kaplan (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.

Whittle, P. (1951): "Hypothesis testing in time series analysis," Uppsala: Almqvist & Wiksells Boktryckeri AB.