Testing-optimal Kernel Choice in HAR Inference*

Yixiao Sun
Department of Economics
UC San Diego

Jingjing Yang
Department of Economics
University of Nevada, Reno

This Version: January 5, 2020

Abstract

The paper investigates the optimal kernel choice in heteroskedasticity and autocorrelation robust tests based on the fixed-b asymptotics. In parallel with the optimality of the quadratic spectral kernel under the asymptotic mean squared error criterion of the point estimator of the long run variance as considered in Andrews (1991), we show that the optimality of the quadratic spectral kernel continues to hold under the testing-oriented criterion of Sun et al. (2008) which takes a weighted average of the probabilities of type I and type II errors of the fixed-b asymptotic test.

Keywords: heteroskedasticity and autocorrelation robust test, fixed-smoothing asymptotics, optimal kernel choice, testing-optimal smoothing-parameter.

JEL Classification Number: C12, C14, C22

1 Introduction

There is much progress in the last two decades on heteroskedasticity and autocorrelation robust (HAR) inference[1]. Besides the development of the fixed-smoothing asymptotic theory, which

*Email: yisun@ucsd.edu and jingjingy@unr.edu. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-1530592. We thank Oliver Linton (the co-editor), an associate editor, and anonymous reviewers for helpful comments. Address correspondence to Yixiao Sun, Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA or Jingjing Yang, Department of Economics, University of Nevada, Reno, Reno, NV 89557-0030, USA.

[1]The term "HAR" was first introduced by Phillips (2005)

includes as a special case the fixed-b asymptotic theory pioneered by Kiefer and Vogelsang (2002a,b, 2005), there is also some progress in developing testing-optimal rules for selecting the smoothing parameter b for the fixed-b test[2]. In particular, Sun et al. (2008) suggest that one should choose the smoothing parameter to minimize a loss function that is a weighted average of the probabilities of type I and type II errors[3]. This testing-optimal choice of b is, in general, larger than the conventional MSE-optimal b of Andrews (1991). Under the asymptotic MSE of the point estimator of the long run variance (LRV), it is well known that the quadratic spectral (QS) kernel is optimal among the class of positive-definite second-order kernels. An open question is whether the optimality of the QS kernel remains valid under the testing-oriented loss function proposed by Sun et al. (2008).

The answer to this question is not obvious, as the MSE criterion and the testing-oriented criterion are fundamentally different. While the MSE criterion balances the asymptotic variance and the *squared* asymptotic bias of the LRV estimator, the testing-oriented criterion of Sun et al. (2008) balances the asymptotic variance with the asymptotic bias itself. The ways that the asymptotic variance and bias enter the criterion functions are also different. For more discussion on the difference, see Sun (2018). Nevertheless, we show that the QS kernel is still optimal under the testing-oriented criterion.

Recently Lazarus et al. (2018) (hereafter LLSW) have also established the testing-optimality of the QS kernel, but they consider a different loss function. Their loss function is a weighted average of the *squared* size distortion and the *squared* size-adjusted power loss. Such a loss function is closer to the mean-squared loss than the loss function we consider here. For example, like the MSE criterion, LLSW's loss function does not respect the direction of the asymptotic bias. In contrast, the loss function we consider here respects the direction — positive and negative biases of the same magnitude have different effects on our loss function. Indeed, we have to show the optimality of the QS kernel for the case with a positive bias and the case with a negative bias separately.

The testing-optimality of the QS kernel is established via high-order expansions of type I and type II errors. Both types of errors are defined pointwisely under a given data-generating

[2]For a sample size T, $\lfloor bT \rfloor$ is the maximum order of autocovariances included in the kernel long run variance estimator for a kernel function with support on $[-1, 1]$.

[3]We often just say "type I and type II errors" instead of "the probabilities of type I and type II errors."

process (DGP). The (pointwise) type I error is not equal to the (uniform) size of the test, which, by definition, is the supremum of the type I error over a family of DGP's that satisfies the null hypothesis. The testing-optimality of the QS kernel is, therefore, established only in a pointwise sense. In a sequence of papers, Preinerstorfer and Pötscher (2016), Pötscher and Preinerstorfer (2018), and Pötscher and Preinerstorfer (2019) establish the necessary and sufficient conditions under which the size of the commonly used HAR tests can be controlled. Unfortunately, conventional heteroskedasticity and autocorrelation consistent (HAC) tests[4] (e.g., Newey and West (1987) and Andrews (1991)), the fixed-b HAR tests (e.g., Kiefer and Vogelsang (2005) and Sun et al. (2008)), and the related fixed-smoothing HAR tests (e.g., Sun (2011) and Ibragimov and Müller (2010)) can all suffer from extreme size distortion, that is, their size can equal one. The test considered here is a fixed-b HAR test and hence can potentially have the same problem. Whether the size can be controlled or not is problem-specific. Subject to a size control, it is an open question whether the QS kernel is still optimal for HAR testing among the class of positive-definite second-order kernels.

The rest of the paper is organized as follows. In Section 2, we introduce the basic setting and review the basics of the fixed-b HAR test in a simple location model. In Section 3, we construct the loss function and establish the testing-optimality of the QS kernel. Section 4 extends the result to the GMM setting. The next section reports simulation evidence, and the final section concludes.

2 Preliminaries

For simplicity, we start with an m-dimensional location model

$$Y_t = \theta_0 + u_t \in \mathbb{R}^m$$

for $t = 1, \ldots, T$ where the error process $\{u_t \in \mathbb{R}^m\}$ is covariance stationary with mean zero. We do not impose a parametric autocorrelation structure on $\{u_t\}$ so that the error process can exhibit autocorrelation of unknown forms. This simple model abstracts away non-essential nuisance but retains the essence of HAR inference. We will consider an extension to the GMM

[4]We use "HAC" to refer to a test that uses critical values from a normal or chi-square distribution. We use "HAR" to refer to a test that uses critical values from a fixed-b asymptotic distribution or standard F distribution.

setting in Section 4.

We are interested in testing $H_0 : \theta_0 = 0$ against $H_1 : \theta_0 \neq 0$. We can estimate θ_0 by the simple average $\hat{\theta} = \bar{Y} := T^{-1}\sum_{t=1}^{T} Y_t$. Then, under some moment and mixing conditions on $\{u_t\}$, we have

$$\sqrt{T}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \to^d N(0, \Omega),$$

where $\Omega := \sum_{j=-\infty}^{\infty} Eu_t u_{t-j}'$ is the long run variance of the error process $\{u_t\}$. To make statistical inferences on θ_0, we follow the large literature on HAC/HAR inference (e.g., Newey and West (1987) and Andrews (1991)) and estimate Ω by

$$\hat{\Omega}_b = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} k\left(\frac{t-s}{bT}\right)\hat{u}_t\hat{u}_s',$$

where $\hat{u}_t = u_t - \bar{u}$, $\bar{u} = T^{-1}\sum_{s=1}^{T} u_s$, $k(\cdot)$ is a kernel function, and b is the smoothing parameter. We focus on positive-definite kernels in this paper. The positive-definiteness ensures that $\hat{\Omega}_b$ is positive definite almost surely. The Wald statistic for testing H_0 against H_1 is

$$F_T \equiv \left(\sqrt{T}\hat{\theta}\right)' \hat{\Omega}_b^{-1}\left(\sqrt{T}\hat{\theta}\right).$$

To develop an asymptotic approximation of F_T, we can write it as

$$F_T = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\right)' \left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} k\left(\frac{t-s}{bT}\right)\hat{u}_t\hat{u}_s'\right]^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\right).$$

Under the increasing-smoothing asymptotics where $b \to 0, T \to \infty$ and $bT \to \infty$, we can show by a standard argument that $F_T \to^d \chi_m^2$, the chi-square distribution with m degrees of freedom.

It is now well known that the chi-square approximation is not accurate in finite samples, especially when b is relatively large. See, for example, Hansen et al. (1996), Kiefer and Vogelsang (2005), and Sun (2014a). To cope with this problem, the literature has developed the fixed-b asymptotics under which b is fixed as $T \to \infty$. The following assumptions are standard in this strand of literature; see, for example, Assumptions 4 and 5 in Sun (2014a).

Assumption 1 *Define $S_t = \sum_{\tau=1}^{t} u_\tau$. $S_{[Tr]}$ satisfies the functional CLT*

$$T^{-1/2}S_{[Tr]} \to^d \Omega^{1/2}W_m(r), \quad r \in [0,1] \tag{1}$$

where Ω is a positive-definite matrix and $W_m(r)$ is the standard m-dimensional Brownian motion.

Assumption 2 *(i)* $k(x) : \mathbb{R} \to [0,1]$ *is symmetric, piecewise smooth with* $k(0) = 1$, *and* $\int_0^\infty k(x)xdx < \infty$. *(ii) The Parzen characteristic exponent defined by*

$$q = \max\{q_0 : q_0 \in \mathbb{Z}^+,\ k^{(q_0)} = \lim_{x\to 0} \frac{1-k(x)}{|x|^{q_0}} < \infty\} \tag{2}$$

is greater than or equal to 1.

Assumption 1 holds for serially correlated data that satisfy certain moment and mixing conditions. The conditions can be found, for example, in Corollary 2.2 of Phillips and Durlauf (1986) and Theorem 3.15 in Phillips and Solo (1992). In particular, Corollary 2.2 of Phillips and Durlauf (1986) shows that Assumption 1 holds for a mean zero and weakly stationary sequence $\{u_t\}$ if

(i) $E(\|u_t\|^\beta) > \infty$ for some $\beta \geq 2$;

(ii) u_t is φ-mixing with mixing coefficients satisfying $\sum_{\ell=1}^\infty \varphi_\ell^{1-1/\beta} < \infty$;

(iii) $\Omega = \sum_{j=-\infty}^\infty E(u_t u_{t-j}')$ is positive definite.

Assumption 2 is a mild assumption on the kernel function. It is satisfied for commonly used kernel functions in spectrum estimation and HAR inference such as the Bartlett, Parzen, and QS kernels. For the Bartlett kernel, the Parzen characteristic exponent q is 1, and we refer to it as a first-order kernel. For the Parzen and QS kernels, the Parzen characteristic exponent q is 2, and we refer to them as second-order kernels.

Under Assumptions 1 and 2, we have

$$F_T \to^d F_\infty(m,b) := W_m(1)' \left[\int_0^1 \int_0^1 k^*\left(\frac{r-s}{b}\right) dW_m(r) dW_m(s)\right]^{-1} W_m(1),$$

where $k^*((r-s)/b)$ is the demeaned kernel given by

$$k^*\left(\frac{r-s}{b}\right) = k\left(\frac{r-s}{b}\right) - \int_0^1 k\left(\frac{r-s}{b}\right) dr - \int_0^1 k\left(\frac{r-s}{b}\right) ds + \int_0^1 \int_0^1 k\left(\frac{r-s}{b}\right) drds.$$

See equation (5) in Sun (2014a). The fixed-b critical values, i.e., the quantiles of $F_\infty(m,b)$, have been recommended for practical use. There is ample simulation evidence that the fixed-b Wald test is more accurate than the conventional chi-squared test. See, for example, Kiefer and Vogelsang (2005) and Sun (2014a).

The performance of the fixed-b Wald test depends on the choice of b and the kernel function $k(\cdot)$. In this paper, we consider choosing b to optimize a loss function that is oriented towards

the testing problem at hand. Given the optimal b, we show that the QS kernel is optimal among the class of the positive-definite second-order kernels. That is, the optimality of the QS kernel, which has been established under the mean square error of the point estimator of the long run variance, is still valid.

3 Testing-optimal Kernel

For any testing problem, our goal is to minimize the chances of making mistakes, whether they might be type I or type II errors. The ultimate objects of interest are the probabilities of type I and type II errors. It is, therefore, reasonable to select b to minimize the combined probabilities of making type I or type II errors or minimize a weighted average of these two probabilities. We can choose the weights to reflect the relative economic loss under each of the two types of errors and the prior probabilities of the null and alternative hypotheses. In this case, the loss-based objective function has an economic interpretation: it is the expected economic loss from making wrong decisions.

The two types of errors can be approximated by using high-order expansions. To facilitate the high-order Edgeworth expansions, we make the assumption below, which combines Assumption 6 (i&ii) and the additional assumption given in the statement of Theorem 2 in Sun (2014a).

Assumption 3 (*i*) u_t *is a stationary Gaussian process.* (*ii*) *For any* $c \in \mathbb{R}^m$, *the spectral density of* $c'u_t$ *is bounded above and away from zero in a neighborhood around the origin.* (*iii*) $\sum_{h=-\infty}^{\infty} |h|^q \left\| E(u_t u_{t-h}') \right\| < \infty$ *where* q *is the Parzen characteristic exponent of the kernel function used in* $\hat{\Omega}_b$ *and defined in Assumption 2.*

Assumption 3(i) is made to simplify the presentation. When u_t is not a Gaussian process, we have to follow the most general approach to develop Edgeworth expansions for time series data. This often requires highly technical assumptions that are difficult to verify. See, for example, Sun and Phillips (2009) for the technical assumptions and a full-fledged Edgeworth expansion. There are additional terms in the full-fledged expansion. However, these terms do not depend on the kernel functions, and as a result, the testing-optimality of the QS kernel still holds without the Gaussian assumption.

Assumption 3(ii) ensures that the OLS estimator $\hat{\theta}_{\text{OLS}}$ and the GLS estimator $\hat{\theta}_{\text{GLS}}$ are asymptotically equivalent, namely, $E[\sqrt{T}(\hat{\theta}_{\text{GLS}} - \hat{\theta}_{\text{OLS}})]^2 = O(1/T)$. This result is used to facilitate the development of high-order expansions. While this paper focuses on the pointwise result only, the upper bound imposed in Assumption 3(ii) also prevents the variance matrix of $(u_1, ..., u_T)$ from becoming singular, and as a result, the size of our test (uniformly over all stationary AR(1) processes) will be less than one. See example 6.2 of Pötscher and Preinerstorfer (2018) for details.

Assumption 3(iii) is a technical assumption used to characterize the asymptotic bias. It is typically assumed in the literature on spectrum estimation and HAR testing.

Under Assumptions 1–3, Theorem 2 of Sun (2014a) shows that under the increasing-smoothing asymptotics, i.e., $b \to 0, bT \to \infty$ as $T \to \infty$,

$$\begin{aligned}
&\Pr_{H_0}(F_T > z) \\
&= 1 - G_m(z) + \underbrace{A_m(z)\, b}_{\text{variance and demeaning effects}} + \underbrace{G'_m(z)\, z \cdot \rho_m \cdot (bT)^{-q}}_{\text{bias effect}} + o(b + (bT)^{-q}), \qquad (3)
\end{aligned}$$

where $G_m(\cdot)$ is the CDF of χ^2_m, the chi-squared distribution with m degrees of freedom,

$$k_1 = \int_{-\infty}^{\infty} k(x)\, dx, \;\; k_2 = \int_{-\infty}^{\infty} k^2(x)\, dx,$$

$$\begin{aligned}
A_m(z) &= [k_1 + k_2(m-1)]\, G'_m(z) z - k_2 \cdot G''_m(z) z^2, \text{ and} \\
\rho_m &= k^{(q)} \omega^{(q)} \text{ for } \omega^{(q)} = \text{tr}\left[\frac{1}{m}\left(\sum_{h=-\infty}^{\infty} |h|^q\, E u_t u'_{t-h}\right)\Omega^{-1}\right].
\end{aligned}$$

In the above expansion, the term $A_m(z)\, b$ captures three different effects. The first is the variance effect of the LRV estimator when u_t is used in constructing the LRV estimator. More specifically, even if we use $\tilde{\Omega}_b = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} k((t-s)/(bT))\, u_t u'_s$ in place of $\hat{\Omega}_b$ in computing F_T, we will still have the term $-k_2 \cdot G''_m(z) z^2 b$, a component of $A_m(z)\, b$, in the high-order expansion. The second is the demeaning effect, which arises from using the demeaned error $\hat{u}_t$ instead of u_t in the LRV estimator $\hat{\Omega}_b$. This is captured by the term $k_1 G'_m(z) z b$, another component of $A_m(z)\, b$. It is not difficult to see that our argument for the optimality of the QS kernel remains valid if we use u_t, which is available under the null, in constructing the LRV estimator. The third is the dimensionality effect captured in $k_2(m-1)G'_m(z) z b$. This term

disappears when $m = 1$. We will refer to $A_m(z)\, b$ as the variance and demeaning effects for convenience. The term $G'_m(z)\, z \cdot \rho_m \cdot (bT)^{-q}$ captures the bias effect from the nonparametric kernel smoothing underlying the LRV estimator.

For a typical economic time series with positive autocorrelation, the averaged relative bias $\omega^{(q)}$ is positive so that the bias effect is positive and decreases with b. On the other hand, we can show that $A_m(z) > 0$ for all z larger than commonly-used chi-square critical values. As a result, the variance and demeaning effects are positive but decreases with b. There is an opportunity to select b to balance these effects. Here we go one step further by considering the fixed-b test and minimizing its type I and type II errors.

Let $c^{\alpha}_{m,b}$ be the fixed-b critical value such that

$$\Pr\left\{F_\infty(m,b) > c^{\alpha}_{m,b}\right\} = \alpha.$$

Expanding the fixed-b limiting distribution as $b \to 0$, Theorem 1 of Sun (2014a) shows that

$$\Pr\left\{F_\infty(m,b) > z\right\} = 1 - G_m(z) - A_m(z) \cdot b + o(b). \tag{4}$$

As a result, $1 - G_m(c^{\alpha}_{m,b}) - A_m(c^{\alpha}_{m,b})b = \alpha + o(b)$. Then, using (3), we have

$$\begin{aligned}\Pr_{H_0}(F_T > c^{\alpha}_{m,b}) &= 1 - G_m\left(c^{\alpha}_{m,b}\right) + A_m\left(c^{\alpha}_{m,b}\right) \cdot b \\ &\quad + G'_m\left(c^{\alpha}_{m,b}\right) c^{\alpha}_{m,b} \cdot \rho_m \cdot (bT)^{-q} + o(b) + o\left((bT)^{-q}\right) \\ &= \alpha + G'_m\left(c^{\alpha}_{m,b}\right) c^{\alpha}_{m,b} \cdot \rho_m \cdot (bT)^{-q} + o(b) + o\left((bT)^{-q}\right).\end{aligned}$$

Therefore, using the fixed-b critical value removes the variance effect, the demeaning effect, and the dimensionality effect in the probability of type I error. In this sense, the fixed-b critical value is second-order correct. See Sun et al. (2008) and Sun (2014a) for more discussion on the second-order correctness of the fixed-b critical values.

Inverting the expansion in (4), we obtain the Cornish–Fisher expansion:

$$\frac{c^{\alpha}_{m,b} - \chi^{\alpha}_m}{\chi^{\alpha}_m} = \left[k_1 + k_2\left(\frac{m}{2} + \chi^{\alpha}_m\right)\right] b + o(b). \tag{5}$$

So, $c^{\alpha}_{m,b} = \chi^{\alpha}_m + O(b)$ as $b \to 0$, and we can approximate the probability of type I error of the fixed-b test by

$$e^m_{\mathrm{I}}(b) := \alpha + G'_m\left(\chi^{\alpha}_m\right) \chi^{\alpha}_m \cdot \rho_m \cdot (bT)^{-q}.$$

The approximation error is of order $o(b) + o\left((bT)^{-q}\right)$.

To obtain an approximate measure of the probability of the type II error of the fixed-b test, we consider the local alternative:

$$H_1\left(\delta^2\right) : \theta_0 = \frac{\Omega^{1/2}\lambda_m}{\sqrt{T}}$$

for $\lambda_m \in \mathbb{R}^m$ that is uniformly distributed on a sphere with radius δ defined by $\mathcal{S}_m\left(\delta^2\right) = \{\tilde{\lambda}_m \in \mathbb{R}^m : ||\tilde{\lambda}_m||^2 = \delta^2\}$. We can choose δ to reflect a value of scientific interest or economic significance if such a value is available. More specifically, we choose δ such that the size of θ_0 under $H_1\left(\delta^2\right)$, as measured by $\theta_0'\Omega^{-1}\theta_0$, would be considered important to detect if it exists. In the absence of such a value, we recommend choosing a value such that, when Ω is known, the local asymptotic power of the test achieves a specified level, say 50% or 75%. Such a recommendation is often made in the optimal testing literature. For example, Elliott et al. (1996) employ 50% as the power level to determine the local-to-unity parameter in their efficient unit-root test.

Under the above local alternative $H_1 := H_1\left(\delta^2\right)$, we can show that under Assumptions 1–3,

$$\Pr_{H_1}(F_T > z) = 1 - G_{m,\delta^2}(z) + A_{m,\delta^2}(z) \cdot b + G'_{m,\delta^2}(z)\, z \cdot \rho_m \cdot (bT)^{-q} + o(b + (bT)^{-q}),$$

where

$$A_{m,\delta^2}(z) = [k_1 + k_2(m-1)] \cdot G'_{m,\delta^2}(z)z - k_2 \cdot G''_{m,\delta^2}(z)z^2,$$

and $G_{m,\delta^2}(\cdot)$ is the CDF of $\chi^2_m\left(\delta^2\right)$, the noncentral chi-squared distribution with the noncentrality parameter δ^2. See Theorem 5 of Sun (2014a) for a proof. This expansion resembles the expansion under the null but with the CDF and its derivatives for the central chi-squared distribution replaced by the counterparts of the noncentral chi-squared distribution. Of course, the two expansions coincide when $\delta = 0$.

Plugging the fixed-b critical value $c^\alpha_{m,b}$ into the expansion of $\Pr_{H_1}(F_T > z)$ and using (5), we obtain

$$\Pr_{H_1}(F_T < c^\alpha_{m,b}) = e^m_{\mathrm{II}}\left(\delta^2, b\right) + o(b + (bT)^{-q}), \tag{6}$$

where

$$\begin{aligned} e^m_{\mathrm{II}}\left(\delta^2, b\right) = {} & \frac{\delta^2}{2} G'_{(m+2),\delta^2}\left(\chi^\alpha_m\right)\chi^\alpha_m \cdot k_2 \cdot b \\ & - G'_{m,\delta^2}\left(\chi^\alpha_m\right)\chi^\alpha_m \cdot \rho_m \cdot (bT)^{-q} + G_{m,\delta^2}\left(\chi^\alpha_m\right) \end{aligned}$$

is an approximate measure of the probability of the type II error of the fixed-b test. The above result is a variant of Theorem 5 of Sun (2014a), which employs an approximate fixed-b critical value. Note that Sun (2014a) has shown that the difference between the exact fixed-b critical value and the approximate fixed-b critical value is of order $o(b)$. In view of this result, the asymptotic expansion in (6) holds regardless of whether the exact fixed-b critical value or its approximation as given in Sun (2014a) is used.

Let π and $(1-\pi)$ be the prior probabilities of H_0 and H_1, respectively. Let κ_{I} and κ_{II} be the losses associated with type I and II errors, respectively. Then the expected loss from committing type I and type II errors is

$$\ell = \pi \cdot \kappa_{\mathrm{I}} \cdot \Pr_{H_0}(F_T > c^{\alpha}_{m,b}) + (1-\pi) \cdot \kappa_{\mathrm{II}} \cdot \Pr_{H_1}(F_T < c^{\alpha}_{m,b}).$$

Using the high order expansions, we have

$$\begin{aligned}\ell &= \pi \cdot \kappa_{\mathrm{I}} \cdot e^m_{\mathrm{I}}(b) + (1-\pi) \cdot \kappa_{\mathrm{II}} \cdot e^m_{\mathrm{II}}\left(\delta^2, b\right) + o(b + (bT)^{-q}) \\ &= \left[\pi \cdot \kappa_{\mathrm{I}} + (1-\pi) \cdot \kappa_{\mathrm{II}}\right] L\left(a, b, \delta^2\right) + o(b + (bT)^{-q}),\end{aligned}$$

where

$$L\left(a, b, \delta^2\right) = a \cdot e^m_{\mathrm{I}}(b) + (1-a) \cdot e^m_{\mathrm{II}}\left(\delta^2, b\right)$$

is a rescaled and approximate loss function and

$$a = \frac{\pi\kappa_{\mathrm{I}}}{\pi\kappa_{\mathrm{I}} + (1-\pi)\kappa_{\mathrm{II}}} = \left(1 + \frac{(1-\pi)}{\pi}\frac{\kappa_{\mathrm{II}}}{\kappa_{\mathrm{I}}}\right)^{-1}$$

reflects the relative loss weighted by the relative prior probability of the null and alternative hypotheses. The value of a has to be chosen by the user for their specific problem at hand. In the absence of any reliable information concerning the relative loss, namely $\kappa_{\mathrm{II}}/\kappa_{\mathrm{I}}$, and the relative prior probability, namely $(1-\pi)/\pi$, it is not unreasonable to set them equal to one and obtain $a = 1/2$. However, in statistical inferences, it is a convention to put more weights on the type I error, and we can follow this convention and set $a > 1/2$. Note that if $\kappa_{\mathrm{II}}/\kappa_{\mathrm{I}}$ or $(1-\pi)/\pi$ depends on δ, then a will also depend on δ.

Given the rescaled loss function $L\left(a, b, \delta^2\right)$, we choose b according to

$$b^* = \arg\min_b L\left(a, b, \delta^2\right). \tag{7}$$

Such an approach to bandwidth choice was first proposed in Sun et al. (2008). The testing-optimal bandwidth b^* accounts for the relative costs of type I and type II errors. This approach is particularly relevant to empirical applications where the costs of type I and type II errors are different and potentially estimable.

Let

$$D_B = \left[\frac{aG'_m\left(\chi_m^\alpha\right)\chi_m^\alpha}{1-a} - G'_{m,\delta^2}\left(\chi_m^\alpha\right)\chi_m^\alpha\right]\omega^{(q)} \text{ and } D_V = \frac{\delta^2}{2}G'_{(m+2),\delta^2}\left(\chi_m^\alpha\right)\chi_m^\alpha.$$

Except for the Parzen characteristic exponent q, neither D_B nor D_V depends on other aspects of the kernel function. Then the rescaled loss function $L\left(a,b,\delta^2\right)$ becomes

$$L\left(a,b,\delta^2\right) = D_B \cdot k^{(q)} \cdot (bT)^{-q} + D_V \cdot k_2 \cdot b + C_1\left(a,\alpha,m,\delta^2\right),$$

where

$$C_1\left(a,\alpha,m,\delta^2\right) = a\cdot\alpha + (1-a)\,G_{m,\delta^2}\left(\chi_m^\alpha\right)$$

and $k^{(q)}$ is defined in Assumption 2. Here $C_1\left(a,\alpha,m,\delta^2\right)$ captures the first-order loss while $D_B \cdot k^{(q)} \cdot (bT)^{-q} + D_V \cdot k_2 \cdot b$ captures the higher-order loss. Note that $C_1\left(a,\alpha,m,\delta^2\right)$ does not depend on b or the kernel function. The testing-optimal choice of b is given by minimizing the higher-order loss:

$$b^* = \arg\min_b \left[D_B \cdot k^{(q)} \cdot (bT)^{-q} + D_V \cdot k_2 \cdot b\right].$$

We now consider two separate cases and show that the QS kernel is optimal in each case. In the first case with $D_B > 0$, we find that the testing-optimal b is

$$b_{\text{opt,I}} = \left\{\frac{qD_B \cdot k^{(q)}}{D_V \cdot k_2}\right\}^{1/(q+1)} T^{-q/(q+1)}.$$

This choice of b is larger than the MSE-optimal b, which is of order $T^{-2q/(2q+1)}$. For testing problems, we should include autocovariances of higher orders than what is deemed optimal under the MSE criterion given in Andrews (1991).

Plugging $b_{\text{opt,I}}$ into the loss function, we have,

$$L\left(a,b_{\text{opt,I}},\delta^2\right) = C_2\left(q,D_B,D_V\right)\left[(k^{(q)})^{1/q}k_2\right]^{q/(q+1)} T^{-q/(q+1)} + C_1\left(a,\alpha,m,\delta^2\right),$$

where

$$C_2\left(q,D_B,D_V\right) = \left(q^{1/(q+1)} + q^{-q/(q+1)}\right)\left(D_B^{1/q}D_V\right)^{q/(q+1)} > 0$$

depends on the kernel function only via its Parzen characteristic exponent q.

To minimize the above loss function for a given q, we only need to choose the kernel function to minimize $(k^{(q)})^{1/q}k_2$. Such a problem is not well defined unless we restrict the class of the kernel functions under consideration. It is reasonable to impose the restrictions that the kernel function is symmetric with $k(0) = 1$ and positive definite in that its Fourier transform is positive. We thus consider the following class of kernels:

$$\mathcal{K} = \left\{ k(\cdot) : k(0) = 1, k(x) = k(-x), \int_{-\infty}^{\infty} \exp(ix\tau)\, k(x)dx \geq 0 \text{ for all } \tau \in \mathbb{R} \right\}.$$

The restricted minimization problem becomes

$$\min_{k(\cdot)} (k^{(q)})^{1/q}k_2 \text{ such that } k(\cdot) \in \mathcal{K}. \tag{8}$$

Positive definite kernels are either first-order kernels (i.e., $q = 1$) or second-order kernels (i.e., $q = 2$). We focus on the second-order kernels so that $q = 2$. From a theoretical point of view, second-order kernels dominate first-order kernels as long as the spectral density of u_t is twice continuously differentiable at the origin[5]. Note that a kernel function can be rescaled to obtain another kernel function $k_{[c]}(x) := k(x/c)$ for some constant $c > 0$ without changing the loss function $L\left(a, b_{\text{opt}}, \delta^2\right)$. To see this, we note that for the rescaled kernel function $k_{[c]}(x)$, we have $k_{[c]}^{(q)} = c^{-q}k^{(q)}$ and $k_{[c],2} = \int_{-\infty}^{\infty} k_{[c]}^2(x)\, dx = c\int_{-\infty}^{\infty} k^2(x)\, dx$. So $(k_{[c]}^{(q)})^{1/q}k_{[c],2} = (k^{(q)})^{1/q}k_2$.

In view of the relationship $k_{[c]}^{(q)} = c^{-q}k^{(q)}$, we can choose the rescaling constant c appropriately to get any target value for $k_{[c]}^{(q)}$. In particular, we can choose c so that $k_{[c]}^{(q)} = k_{QS}^{(q)}$. Consequently, the restricted minimization problem in (8) is equivalent to solving

$$\min_{k(\cdot)} k_2 := \int_{-\infty}^{\infty} k^2(x)\, dx \text{ such that } k^{(q)} = k_{QS}^{(q)} \text{ and } k(\cdot) \in \mathcal{K}. \tag{9}$$

But the above problem is exactly the same as what we obtain under the MSE of the point estimator of the LRV. See Andrews (1991) and Priestly (1981) (pp. 569–571). Given that the QS kernel is optimal under the MSE criterion, it is also optimal under the criterion we consider here. That is, when $D_B > 0$, the QS kernel is testing-optimal under the criterion that takes a weighted average of the probabilities of type I and type II errors.

[5]The most popular first-order kernel for HAC/HAR inference in econometrics is the Bartlett kernel, which is the kernel used in the Newey-West estimator (Newey and West (1987)). Recently, Kolokotrones and Stock (2019) show that the Bartlett kernel is not optimal among the first-order kernels. They further show that there is no optimal first-order kernel for HAR testing and for minimizing the MSE of the spectrum estimator.

In the second case that $D_B \leq 0$, we have, for *any* b,

$$\begin{aligned} & D_B \cdot k_{QS}^{(q)} \cdot (bT)^{-q} + D_V \cdot k_{2,QS} \cdot b \\ & \leq D_B \cdot k^{(q)} \cdot (bT)^{-q} + D_V \cdot k_2 \cdot b \end{aligned}$$

for any kernel function $k(\cdot) \in \mathcal{K}$ with $k^{(q)} = k_{QS}^{(q)}$. This holds because the QS kernel is the solution to the optimality problem in (9) and $D_V > 0$. Therefore, the QS kernel is still optimal in this case.

In the second case, an aggressive approach to the choice of b is to let $D_B \cdot k^{(q)} \cdot (bT)^{-q} + D_V \cdot k_2 \cdot b = 0$, yielding

$$b_{\text{opt,II}} = \left\{ \frac{|D_B| \cdot k^{(q)}}{D_V \cdot k_2} \right\}^{1/(q+1)} T^{-q/(q+1)}.$$

Under this choice of b, the high order loss is zero and all kernels are asymptotically equivalent. However, in finite samples and with a data-driven bandwidth, we cannot expect to set $D_B \cdot k^{(q)} \cdot (bT)^{-q} + D_V \cdot k_2 \cdot b$ equal to zero exactly.

The optimal bandwidth parameter $b_{\text{opt,II}}$ takes the same form as $b_{\text{opt,I}}$, but there is a difference of factor $q^{1/(q+1)}$. When $q = 1$, this factor takes the value of 1. When $q = 2$, this factor takes the value of $2^{1/3} = 1.2599$, which is not too different from 1. Therefore, it is not unreasonable to use

$$b_{\text{opt}} = b_{\text{opt,I}} = \left\{ \frac{q\,|D_B| \cdot k^{(q)}}{D_V \cdot k_2} \right\}^{1/(q+1)} T^{-q/(q+1)} \tag{10}$$

in both cases. With the above choice, the rescaled loss function in the second case (i.e., when $D_B \leq 0$) becomes

$$L\left(a, b_{\text{opt}}, \delta^2\right) = \tilde{C}_2\left(q, D_B, D_V\right) \left[(k^{(q)})^{1/q} k_2 \right]^{q/(q+1)} T^{-q/(q+1)} + C_1\left(a, \alpha, m, \delta^2\right),$$

where

$$\tilde{C}_2\left(q, D_B, D_V\right) = \left(q^{1/(q+1)} - q^{-q/(q+1)} \right) |D_B|^{1/(q+1)} D_V^{q/(q+1)} > 0.$$

As in the first case with $D_B > 0$, this loss function is an increasing function of $(k^{(q)})^{1/q} k_2$. For the same reason presented before, among the kernels in $\mathcal{K}$, the loss function is still minimized at the QS kernel. This is consistent with the argument above that the QS kernel dominates other kernels in $\mathcal{K}$ in the case that $D_B \leq 0$ regardless of the choice of b.

We can provide an alternative proof of the testing-optimality of the QS kernel under b_{opt} given in (10). With this choice of b, the rescaled loss function $L\left(a, b, \delta^2\right)$ can be represented in a unified way:

$$L\left(a, b_{\text{opt}}, \delta^2\right) = \left[q^{1/(q+1)} + \text{sgn}(D_B) q^{-q/(q+1)}\right] \left[|D_B|^{1/q} D_V\right]^{q/(q+1)} \left[(k^{(q)})^{1/q} k_2\right]^{q/(q+1)} T^{-q/(q+1)} + C_1\left(a, \alpha, m, \delta^2\right).$$

It is easy to see that $\left(q^{1/(q+1)} + \text{sgn}(D_B) q^{-q/(q+1)}\right) \geq 0$ regardless of whether $\text{sgn}(D_B) = -1, 0, 1$. The optimal kernel in $\mathcal{K}$ is thus the kernel that minimizes $(k^{(q)})^{1/q} k_2$ for all $k \in \mathcal{K}$. But the solution to this minimization problem is the QS kernel. That is, when $b = b_{\text{opt}}$, the optimal kernel in $\mathcal{K}$ is the QS kernel.

4 Extension to the General GMM Setting

To extend the optimality of the QS kernel to the GMM setting, we assume that the moment condition

$$Ef\left(v_t, \theta\right) = 0, \ t = 1, 2, \ldots, T$$

holds if and only if $\theta = \theta_0 \in \mathbb{R}^d$ where $f\left(\cdot\right)$ is an $m \times 1$ vector of twice continuously differentiable functions with $m \geq d$ and rank $E\left[\partial f\left(v_t, \theta_0\right) / \partial \theta'\right] = d$. We are interested in testing the null hypothesis $H_0 : \tau(\theta_0) = 0$ against the alternative hypothesis $H_1 : \tau\left(\theta_0\right) \neq 0$, where $\tau\left(\theta\right)$ is a $p \times 1$ vector of twice continuously differentiable functions with first-order derivative matrix $\Gamma(\theta) = \partial \tau(\theta) / \partial \theta'$.

The GMM estimator of θ_0 is given by

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} \left[T^{-1} \sum_{s=1}^{T} f(v_s, \theta)\right]' \mathcal{W}_T T^{-1} \sum_{s=1}^{T} f(v_s, \theta),$$

where $\mathcal{W}_T$ is an $m \times m$ positive semidefinite weighting matrix $\mathcal{W}_T$ and Θ is a compact parameter space.

Let

$$g_t\left(\theta\right) = T^{-1} \sum_{s=1}^{t} f(v_s, \theta), \ G_t(\theta) = \frac{\partial g_t\left(\theta\right)}{\partial \theta'} = \frac{1}{T} \sum_{j=1}^{t} \frac{\partial f(v_j, \theta)}{\partial \theta'} \text{ and } G_0 = E \frac{\partial f(v_j, \theta_0)}{\partial \theta'}.$$

As in Sun (2014a), we maintain the following standard assumptions in the literature on fixed-smoothing asymptotics.

Assumption 4 *$plim_{T\to\infty}\hat{\theta}_T = \theta_0$, θ_0 is an interior point of the compact set Θ, and $\Gamma_0 = \Gamma(\theta_0)$ has a full row-rank p.*

Assumption 5 *$plim_{T\to\infty} G_{[rT]}(\tilde{\theta}_T) = rG_0$ uniformly in r for any $\tilde{\theta}_T$ whose elements are between the corresponding elements of $\hat{\theta}_T$ and θ_0.*

Assumption 6 *$\mathcal{W}_T$ is positive semidefinite, $plim_{T\to\infty}\mathcal{W}_T = \mathcal{W}_\infty$, and $G_0'\mathcal{W}_\infty G_0$ is positive definite.*

These assumptions ensure that

$$\sqrt{T}\left[\tau(\hat{\theta}_T) - \tau(\theta_0)\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t + o_p(1),$$

where

$$u_t := \phi(v_t, \theta_0) = -\Gamma_0\left(G_0'\mathcal{W}_\infty G_0\right)^{-1} G_0'\mathcal{W}_\infty f(v_t, \theta_0).$$

On the basis of the GMM estimator $\hat{\theta}_T$, the Wald statistic for testing H_0 against H_1 is

$$F_T = \left[\sqrt{T}\tau(\hat{\theta}_T)\right]' \hat{\Omega}_T^{-1}\left[\sqrt{T}\tau(\hat{\theta}_T)\right], \tag{11}$$

where, as in the location case, $\hat{\Omega}_b$ takes the form

$$\hat{\Omega}_b = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} k\left(\frac{t-\tau}{bT}\right)\hat{u}_t\hat{u}_\tau'$$

and

$$\hat{u}_t := -\Gamma(\hat{\theta}_T)\left[G_T(\hat{\theta}_T)'\mathcal{W}_T G_T(\hat{\theta}_T)\right]^{-1} G_T(\hat{\theta}_T)'\mathcal{W}_T f(v_t, \theta_0)$$

is the plug-in estimator of u_t. $\hat{\Omega}_b$ is an estimator of the LRV Ω of $\{u_t\}$.

Following Sun (2014a), we can establish the stochastic expansion under H_0 :

$$F_T = F_{T,L} + \psi_T + \psi_T^*$$

where

$$F_{T,L} = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\right]' \tilde{\Omega}_b^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\right]$$

is the dominated linear term in the approximation and

$$\tilde{\Omega}_b = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} k\left(\frac{t-\tau}{bT}\right)\left[u_t - \frac{1}{T}\sum_{s=1}^{T} u_s\right]\left[u_\tau - \frac{1}{T}\sum_{s=1}^{T} u_s\right]'.$$

In this stochastic approximation, $\psi_T = O_p(1/\sqrt{T})$ does not depend on b and $\psi_T^* = O_p([\sqrt{b} + (bT)^{-q}]/\sqrt{T} + 1/T)$.

Using this expansion and maintaining the same additional technical conditions in Theorem 3 of Sun (2014a), we obtain

$$\begin{aligned}\Pr_{H_0}(F_T > z) = 1 - G_p(z) + A_p(z)b + G_p'(z)\, z \cdot \rho_{\text{GMM}} \cdot (bT)^{-q} \\ + O(T^{-1/2}\log T) + o(b) + o\left((bT)^{-q}\right)\end{aligned} \tag{12}$$

where the $O\left(T^{-1/2}\log T\right)$ term does not depend on b,

$$\rho_{\text{GMM}} = k^{(q)}\omega^{(q)} \text{ for } \omega^{(q)} = \text{tr}\left[\frac{1}{p}\left(\sum_{h=-\infty}^{\infty} |h|^q Eu_t u_{t-h}'\right)\Omega^{-1}\right]$$

and $k^{(q)}$ is defined in Assumption 2. This expansion is the same as that in (3) except the difference in the number of restrictions tested. Using the same argument as before, we have

$$\Pr_{H_0}(F_T > c_{p,b}^{\alpha}) = e_{\text{I}}^p(b) + O(T^{-1/2}\log T) + o(b) + o\left((bT)^{-q}\right),$$

where

$$e_{\text{I}}^p(b) := \alpha + G_p'\left(\chi_p^{\alpha}\right)\chi_p^{\alpha} \cdot \rho_{\text{GMM}} \cdot (bT)^{-q}.$$

Consider the local alternative hypothesis $H_1\left(\delta^2\right) : \tau(\theta_0) = \Omega^{1/2}\lambda_p/\sqrt{T}$ where λ_p is uniformly distributed on $\mathcal{S}_p\left(\delta^2\right) = \{\tilde{\lambda}_p \in \mathbb{R}^p : ||\tilde{\lambda}_p||^2 = \delta^2\}$. Under some technical conditions, Theorem 5 of Sun (2014a) has shown that

$$\Pr_{H_1}(F_T \leq c_{p,b}^{\alpha}) = e_{\text{II}}^p\left(\delta^2, b\right) + O(\log T/\sqrt{T}) + o(b) + o\left((bT)^{-q}\right)$$

where the $O(\log T/\sqrt{T})$ term does not depend on b and

$$\begin{aligned}e_{\text{II}}^p\left(\delta^2, b\right) = \frac{\delta^2}{2} G_{(p+2),\delta^2}'\left(\chi_p^{\alpha}\right)\chi_p^{\alpha} \cdot k_2 \cdot b \\ - G_{p,\delta^2}'\left(\chi_p^{\alpha}\right)\chi_p^{\alpha} \cdot \rho_{\text{GMM}} \cdot (bT)^{-q} + G_{p,\delta^2}\left(\chi_p^{\alpha}\right).\end{aligned}$$

Given that the approximate type I and type II errors take the same form as in the case with a location model, the testing-optimality of the QS kernel remains valid in the GMM setting.

5 Simulation Evidence

To evaluate the testing-optimality of the QS kernel in finite samples, we consider two different data generating processes: a location model and a regression model.

5.1 Location Model

We consider a univariate location model with AR(1) errors:

$$Y_t = \theta_0 + u_t \text{ where } u_t = \phi u_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim i.i.d.N(0,1),$$

for $t = 1, 2, \ldots, T$. We set ϕ equal to $0, 0.5, 0.7, 0.9, -0.5, -0.7, -0.9$ and let $u_0 = 0$.

We first consider the finite-sample optimal choice of b. We use the above location model to simulate the exact type I and type II errors for each of the sample sizes $T = 50, 100, 200$ and for a grid of b values ranging from 0.02 to 0.5 with an increment of 0.02. To simulate the type II error, we set δ to a value such that the asymptotic local power of the asymptotic chi-squared test is 75%. That is, δ^2 solves $\Pr\left(\chi_1^2\left(\delta^2\right) > 1.96^2\right) = 0.75$. The number of simulation replications is $10,000$. We set $a = 0.8$ so that we assign more weights toward the type I error. The finite-sample optimal b value is the value of b's on the grid 0.02:0.02:0.5 that achieves the smallest loss function. We consider the fixed-b tests based on the Bartlett, Parzen, and QS kernels. The nominal level of each test is 5%.

Table 1 reports the ratio of the loss function associated with the Bartlett and Parzen kernels to the loss function associated with the QS kernel. Table 2 reports the corresponding finite-sample optimal b's used in Table 1. We can clearly see that the QS kernel dominates the Bartlett and Parzen kernels in terms of having a smaller loss function. This result is consistent with what Andrews (1991) obtains under the MSE criterion.

For completeness, we report in Table 4 the empirical null rejection probabilities underlying Table 1. We note that the empirical null rejection probabilities can be quite different from the nominal level of 5% when the error process is strongly autocorrelated. The difference will be even larger if we use the standard normal or chi-squared critical value. Such a phenomenon has been widely observed in the literature on fixed-b HAR inference. See, for example, Kiefer and Vogelsang (2005) and Sun (2014a). The difference will become larger when ϕ increases to 1. For example, Pötscher and Preinerstorfer (2018) (Appendix J) demonstrates that the empirical null rejection probability can be as large as 60% when ϕ approaches 1. To fix the size distortion problem in the AR(1) design, we can use the fixed-b critical values from Sun (2014b), which develops the fixed-b asymptotics under the local-to-unity specification. We will not pursue this here, as our goal is to examine the testing-optimality of the QS kernel when we are away from

the local-to-unity region.

We next consider the choice of b based on the asymptotic result, that is, we employ the formula given in (10), but we plug in the true values of D_B and D_V based on the model. Such a plug-in implementation is not feasible in practice, but it provides a way to examine the accuracy of the asymptotic measurements and the optimality of the QS kernel in finite samples.

Table 4 provides the ratio of the loss functions and Table 5 provides the corresponding b_{opt} obtained. The results using b_{opt} are similar to those using the finite-sample optimal b. So, with an asymptotically justified bandwidth choice, the superior performance of the QS kernel is still reflected in our finite sample simulations.

5.2 Regression Model

For the regression model, we consider the data generating process:

$$y_t = \gamma_0 + x_t'\beta_0 + \varepsilon_t$$

where x_t is a 4×1 vector process and x_t and ε_t follow AR(1) processes

$$x_{t,j} = \phi_x x_{t-1,j} + e_{t,j},\ \varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + e_{t,0}$$

where $e_{t,j} \sim i.i.d\ N(0,1)$ across t and j and $x_0 = 0$ and $\varepsilon_0 = 0$. We consider the case with the same AR parameter: $\phi_x = \phi_\varepsilon = \phi$ and

$$\phi = 0, \sqrt{0.5}, \sqrt{0.7}, \sqrt{0.9}, -\sqrt{0.5}, -\sqrt{0.7}, -\sqrt{0.9}.$$

In this case, $x_{t,j}\varepsilon_t$ follows AR(1) with AR parameter ϕ^2: $x_{t,j}\varepsilon_t = \phi^2 [x_{t-1,j}\varepsilon_{t-1}] + \tilde{e}_{t,j}$ where $\tilde{e}_{t,j} = \phi_x x_{t-1,j} e_{t,0} + \phi_\varepsilon e_{t,j}\varepsilon_{t-1} + e_{t,j}e_{t,0}$ is a martingale difference sequence. The design here is similar to that in Kiefer and Vogelsang (2005) except that Kiefer and Vogelsang (2005) fix ϕ_x at 0.5. Given that $\{x_t\}$ and $\{\varepsilon_t\}$ are normal and independent, we refer to this design as the independent normal design.

We also consider two variants of the above design: independent *chi-squared* design and *dependent* normal design. In the independent chi-squared design, we let $e_{t,j} \sim i.i.d.\ \left(\chi_1^2 - 1\right)/\sqrt{2}$. This design is used to examine whether the finite sample results are sensitive to the asymmetry of the data distribution. In the dependent normal design, we let

$$x_{t,j} = \phi_x x_{t-1,j} + \left(e_{t,j} + e_{t+1,0}\right)/\sqrt{2} \text{ and } \varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + e_{t,0}$$

where $e_{t,j}$ is still $i.i.d$ $N(0,1)$ across t and j. The only difference from the independent design is that $x_{t,j}$ is now partially determined by the future value of the error innovation $\{e_{t,0}\}$. As a result, the two time series $\{x_t\}$ and $\{\varepsilon_t\}$ are not independent any more. In this case, it is not possible to invoke a conditioning argument to treat the whole sequence $\{x_t\}$ as fixed, and we need to employ a general HAR variance estimator as considered in this paper. Also, the theoretical results of Pötscher and Preinerstorfer (2018) do not cover this case. The dependent design can be justified if the population consists of forward-looking individuals. Pötscher and Preinerstorfer (2018) (Appendix K) consider a similar design.

We are interested in testing the slope parameter $\beta_0 = (\beta_{10}, \beta_{20}, \beta_{30}, \beta_{40})'$. We set the intercept $\gamma_0 = 0$ without the loss of generality. The null hypotheses of interest are

$$H_{0p} : \beta_{10} = \ldots = \beta_{p0} = 0$$

for $p = 1, 2, 3, 4$. Denote $\theta_0 = (\gamma_0, \beta_0')'$ and $R_{0p} = \mathbb{I}_5(2 : p+1, :)$, i.e., row 2 to row $p+1$ of the identity matrix $\mathbb{I}_5$. The above null hypotheses can be written as $H_{0p} : R_{0p}\theta_0 = 0$. Let $\tilde{x}_t = [1, x_t']'$ and $G_0 = -E(\tilde{x}_t\tilde{x}_t')$. We consider the local alternative hypotheses

$$H_{1p}\left(\delta^2\right) : (\beta_1, \ldots, \beta_p) = \Omega^{1/2}\lambda_p/\sqrt{T},$$

where $\Omega^{1/2}$ is the matrix square root of the LRV of $R_{0p}G_0^{-1}\tilde{x}_t\varepsilon_t$ and λ_p is uniformly distributed over the sphere $\mathcal{S}_p\left(\delta^2\right)$. For the parameters not specified under the null, we set $\beta_{p+1} = \ldots = \beta_4 = 0$ under both the null and alternative hypotheses.

As in the case with a location model, for each value of b from 0.02 to 0.5 with increment 0.02, and for each of the three kernels, we simulate the finite-sample type I and type II errors of the 5% fixed-b test. For the type II error, we employ δ^2 that solves $\Pr\left(\chi_p^2\left(\delta^2\right) > \chi_p^2(0.95)\right) = 0.75$ where $\chi_p^2(0.95)$ is the 95% quantile of χ_p^2. We then select the value of b to minimize the finite-sample loss function with $a = 0.8$. We consider three different sample sizes $T = 100, 200, 500$. The number of simulation replications is 10000.

Tables 6 and 7 report the results under the independent normal design. Table 6 provides the relative optimal finite sample loss function when $T = 100$, and Table 7 provides the corresponding finite sample optimal bandwidths. These two tables are representative of the tables for other sample sizes. The basic qualitative message is the same as that in the location model: the QS kernel delivers a smaller loss function than the Bartlett and Parzen kernels. We have

also simulated the finite sample loss when b is set according to (10), but the qualitative message is the same.

Tables 8 and 9 report the relative optimal finite sample loss function under the independent chi-squared design and the dependent normal design, respectively. To save space, we do not report the corresponding bandwidths. These two tables lend further support to the testing-optimality of the QS kernel: the QS kernel dominates the other two kernels when the data distribution is neither normal nor symmetric and when the regressor process and the error process are not independent.

6 Conclusion

We show that the QS kernel is optimal under the testing-oriented loss function that takes a weighted average of type I and type II errors of the fixed-b HAR test. In view of the duality, the QS kernel is still optimal if the bandwidth parameter is chosen to minimize the type II error subject to the constraint that the type I error is less than some pre-specified threshold. Lazarus et al. (2018) show that the QS kernel is optimal under a different testing-oriented criterion. The optimality of the QS kernel appears to be robust to different criterion functions.

Table 1: Ratio of the finite-sample loss function of the Bartlett and Parzen fixed-b tests to that of the QS fixed-b test using the finite-sample optimal b in the location model

		ϕ						
Kernel	T	0	0.5	0.7	0.9	-0.5	-0.7	-0.9
Bartlett	50	1.05	1.20	1.03	1.03	1.04	1.10	1.17
Parzen	50	1.02	1.01	1.01	1.02	1.02	1.01	1.01
QS	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Bartlett	100	1.03	1.08	1.06	1.04	1.03	1.04	1.08
Parzen	100	1.00	1.01	1.01	1.02	1.00	1.00	1.00
QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Bartlett	200	1.01	1.03	1.05	1.05	1.04	1.02	1.02
Parzen	200	1.01	1.00	1.00	1.01	1.00	1.00	1.00
QS	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: The finite-sample optimal b used in Table 1

		ϕ						
Kernel	T	0	0.5	0.7	0.9	-0.5	-0.7	-0.9
Bartlett	50	0.02	0.12	0.15	0.21	0.11	0.10	0.11
Parzen	50	0.02	0.14	0.17	0.17	0.08	0.08	0.08
QS	50	0.02	0.18	0.18	0.30	0.06	0.06	0.06
Bartlett	100	0.02	0.17	0.17	0.21	0.07	0.09	0.09
Parzen	100	0.02	0.16	0.17	0.17	0.05	0.06	0.06
QS	100	0.02	0.18	0.18	0.22	0.04	0.04	0.06
Bartlett	200	0.02	0.13	0.16	0.16	0.06	0.08	0.09
Parzen	200	0.02	0.12	0.17	0.17	0.05	0.06	0.06
QS	200	0.02	0.12	0.18	0.18	0.04	0.06	0.06

Table 3: The empirical null rejection probability underlying Table 1

		ϕ						
Kernel	T	0	0.5	0.7	0.9	-0.5	-0.7	-0.9
Bartlett	50	0.050	0.078	0.098	0.174	0.032	0.017	0.008
Parzen	50	0.050	0.074	0.100	0.211	0.037	0.034	0.032
QS	50	0.048	0.055	0.069	0.106	0.046	0.045	0.043
Bartlett	100	0.052	0.063	0.075	0.123	0.035	0.029	0.016
Parzen	100	0.050	0.059	0.072	0.146	0.039	0.040	0.030
QS	100	0.049	0.050	0.055	0.084	0.047	0.047	0.051
Bartlett	200	0.048	0.056	0.060	0.093	0.039	0.035	0.022
Parzen	200	0.049	0.053	0.055	0.093	0.045	0.046	0.045
QS	200	0.047	0.050	0.049	0.065	0.046	0.048	0.047

Table 4: Ratio of the finite-sample loss function of the Bartlett and Parzen fixed-b tests to that of the QS fixed-b test using b_{opt} in the location model

		ϕ						
Kernel	T	0	0.5	0.7	0.9	-0.5	-0.7	-0.9
Bartlett	50	1.04	1.07	1.03	1.03	1.03	1.08	1.12
Parzen	50	1.01	1.01	1.01	1.02	1.02	1.01	1.01
QS	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Bartlett	100	1.03	1.07	1.06	1.04	1.04	1.05	1.06
Parzen	100	1.01	1.01	1.01	1.02	1.01	1.00	1.00
QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Bartlett	200	1.02	1.04	1.05	1.05	1.04	1.04	1.03
Parzen	200	1.00	1.00	1.00	1.01	1.00	1.00	1.00
QS	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 5: The values of b_{opt} used in Table 4

		ϕ						
Kernel	T	0	0.5	0.7	0.9	-0.5	-0.7	-0.9
Bartlett	50	0.02	0.16	0.19	0.20	0.08	0.10	0.10
Parzen	50	0.02	0.16	0.20	0.21	0.09	0.08	0.08
QS	50	0.02	0.17	0.21	0.22	0.09	0.06	0.06
Bartlett	100	0.02	0.15	0.18	0.20	0.05	0.08	0.08
Parzen	100	0.02	0.15	0.19	0.21	0.05	0.06	0.06
QS	100	0.02	0.16	0.19	0.22	0.05	0.05	0.06
Bartlett	200	0.02	0.11	0.14	0.18	0.04	0.08	0.08
Parzen	200	0.02	0.11	0.15	0.18	0.04	0.04	0.06
QS	200	0.02	0.12	0.15	0.19	0.04	0.04	0.06

Table 6: Ratio of the finite-sample loss function of the Bartlett and Parzen fixed-b tests to that of the QS fixed-b test using the finite-sample optimal b in the regression model under the independent normal design

			$\phi_x = \phi_u = \phi$						
	Kernel	T	0	$\sqrt{0.5}$	$\sqrt{0.7}$	$\sqrt{0.9}$	$-\sqrt{0.5}$	$-\sqrt{0.7}$	$-\sqrt{0.9}$
$p=1$	Bartlett	100	1.00	1.03	1.03	1.04	1.04	1.03	1.01
$p=1$	Parzen	100	1.00	1.01	1.01	1.01	1.01	1.01	1.01
$p=1$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=2$	Bartlett	100	1.00	1.03	1.04	1.05	1.04	1.05	1.04
$p=2$	Parzen	100	1.00	1.01	1.01	1.01	1.01	1.02	1.00
$p=2$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=3$	Bartlett	100	1.00	1.05	1.06	1.06	1.07	1.05	1.03
$p=3$	Parzen	100	1.00	1.01	1.01	1.01	1.02	1.00	1.00
$p=3$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=4$	Bartlett	100	1.00	1.07	1.08	1.09	1.10	1.08	1.08
$p=4$	Parzen	100	1.00	1.01	1.01	1.01	1.01	1.02	1.01
$p=4$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 7: The finite-sample optimal b used in Table 6

			$\phi_x = \phi_u = \phi$						
	Kernel	T	0	$\sqrt{0.5}$	$\sqrt{0.7}$	$\sqrt{0.9}$	$-\sqrt{0.5}$	$-\sqrt{0.7}$	$-\sqrt{0.9}$
$p=1$	Bartlett	100	0.02	0.10	0.14	0.18	0.10	0.10	0.12
$p=1$	Parzen	100	0.02	0.10	0.12	0.18	0.10	0.10	0.12
$p=1$	QS	100	0.02	0.06	0.06	0.10	0.06	0.06	0.06
$p=2$	Bartlett	100	0.02	0.10	0.20	0.22	0.08	0.12	0.14
$p=2$	Parzen	100	0.02	0.10	0.16	0.22	0.10	0.14	0.20
$p=2$	QS	100	0.02	0.06	0.10	0.12	0.06	0.06	0.10
$p=3$	Bartlett	100	0.02	0.10	0.22	0.26	0.08	0.10	0.18
$p=3$	Parzen	100	0.02	0.12	0.22	0.24	0.10	0.12	0.20
$p=3$	QS	100	0.02	0.06	0.12	0.14	0.06	0.06	0.12
$p=4$	Bartlett	100	0.02	0.14	0.18	0.26	0.12	0.14	0.26
$p=4$	Parzen	100	0.02	0.12	0.20	0.22	0.12	0.14	0.22
$p=4$	QS	100	0.02	0.08	0.08	0.12	0.06	0.08	0.12

Table 8: Ratio of the finite-sample loss function of the Bartlett and Parzen fixed-b tests to that of the QS fixed-b test using the finite-sample optimal b in the regression model under the independent chi-squared design

			$\phi_x = \phi_u = \phi$						
	Kernel	T	0	$\sqrt{0.5}$	$\sqrt{0.7}$	$\sqrt{0.9}$	$-\sqrt{0.5}$	$-\sqrt{0.7}$	$-\sqrt{0.9}$
$p=1$	Bartlett	100	1.00	1.01	1.03	1.02	1.07	1.02	1.02
$p=1$	Parzen	100	1.00	1.00	1.00	1.01	1.01	1.01	1.01
$p=1$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=2$	Bartlett	100	1.00	1.05	1.04	1.02	1.06	1.03	1.02
$p=2$	Parzen	100	1.00	1.00	1.00	1.01	1.01	1.01	1.01
$p=2$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=3$	Bartlett	100	1.00	1.10	1.12	1.08	1.07	1.10	1.04
$p=3$	Parzen	100	1.00	1.01	1.01	1.01	1.01	1.01	1.01
$p=3$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=4$	Bartlett	100	1.00	1.11	1.12	1.07	1.11	1.11	1.06
$p=4$	Parzen	100	1.00	1.00	1.01	1.01	1.01	1.00	1.00
$p=4$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 9: Ratio of the finite-sample loss function of the Bartlett and Parzen fixed-b tests to that of the QS fixed-b test using the finite-sample optimal b in the regression model under the dependent normal design

			$\phi_x = \phi_u = \phi$						
	Kernel	T	0	$\sqrt{0.5}$	$\sqrt{0.7}$	$\sqrt{0.9}$	$-\sqrt{0.5}$	$-\sqrt{0.7}$	$-\sqrt{0.9}$
$p=1$	Bartlett	100	1.01	1.02	1.03	1.18	1.01	1.00	1.05
$p=1$	Parzen	100	1.00	1.01	1.00	1.15	1.01	1.00	1.05
$p=1$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=2$	Bartlett	100	1.00	1.01	1.02	1.09	1.00	1.01	1.03
$p=2$	Parzen	100	1.00	1.00	1.00	1.05	1.00	1.00	1.00
$p=2$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=3$	Bartlett	100	1.00	1.02	1.01	1.07	1.02	1.02	1.02
$p=3$	Parzen	100	1.00	1.00	1.01	1.00	1.00	1.00	1.02
$p=3$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=4$	Bartlett	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=4$	Parzen	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p=4$	QS	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–836.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics*, 14(3):262–280.

Ibragimov, R. and Müller, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics*, 28(4):453–468.

Kiefer, N. M. and Vogelsang, T. J. (2002a). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, 18:1350–1366.

Kiefer, N. M. and Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70:2093–2095.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21:1130–1164.

Kolokotrones, T. and Stock, J. (2019). Is Newey-West optimal among q = 1 kernels? Working Paper, Department of Economics, Harvard University.

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

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Phillips, P. C. B. (2005). Automated discovery in econometrics. *Econometric Theory*, 21(1):3–20.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple Time Series Regression with Integrated Processes. *The Review of Economic Studies*, 53(4):473–495.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, 20(2):971–1001.

Pötscher, B. M. and Preinerstorfer, D. (2018). Controlling the size of autocorrelation robust tests. *Journal of Econometrics*, 207(2):406–431.

Pötscher, B. M. and Preinerstorfer, D. (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.

Preinerstorfer, D. and Pötscher, B. M. (2016). On size and power of heteroskedasticity and autocorrelation robust tests. *Econometric Theory*, 32(2):261–358.

Priestly, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press.

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

Sun, Y. (2014a). 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. (2014b). Fixed-smoothing asymptotics and asymptotic f and t tests in the presence of strong autocorrelation. *Advances in Econometrics*, 33:23–63.

Sun, Y. (2018). Comment on "HAR inference: Recommendations for practice". *Journal of Business & Economic Statistics*, 36(4):565–568.

Sun, Y., Philips, P. C. B., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, 76:175–194.