Is Newey-West Optimal Among First-Order Kernels?

Thomas Kolokotrones* James H. Stock† Christopher D. Walker‡

June 22, 2022

Abstract

Newey-West [1987] standard errors are the dominant standard errors used for heteroskedasticity and autocorrelation robust (HAR) inference in time series regression. The Newey-West estimator uses the Bartlett kernel, which is a first-order kernel, meaning that its characteristic exponent, q, is equal to 1, where q is defined as the largest value of r for which the quantity $k^{[r]}(0) = \lim_{t\to 0} |t|^{-r}(k(0) - k(t))$ is defined and finite. This raises the apparently uninvestigated question of whether the Bartlett kernel is optimal among first-order kernels. Here, we demonstrate that, for $q < 2$, there is no optimal q^{th}-order kernel for HAR testing in the Gaussian location model or for minimizing the MSE in spectral density estimation. In fact, for any $q \leq 2$, the space of q^{th}-order positive-semidefinite kernels is not closed and, moreover, all continuous q^{th}-order kernels can be decomposed into a weighted sum of q^{th} and second-order kernels, which suggests that there is no meaningful notion of 'pure' q^{th}-order kernels for $q < 2$. We additionally show that the classical result that nonconstant, positive-semidefinite kernels have $q \leq 2$ holds without the traditional regularity conditions. Thus, our results apply to all nonconstant, symmetric, positive-semidefinite kernels for any value of q. Nevertheless, it is possible to rank any given collection of q^{th}-order kernels using the functional $I_q[k] = \left(k^{[q]}(0)\right)^{\frac{1}{q}} \int k^2(t)dt$ with smaller values corresponding to better asymptotic performance. We examine a wide variety of first-order estimators, including all those commonly encountered in the literature and ones newly developed here. None improve upon the Bartlett kernel.

1 Introduction

In time series regression, when the product of the error term and the regressor is serially correlated and generalized least squares is not possible, one must use heteroskedasticity and au-

*Harvard University, Departments of Biostatistics and Epidemiology.
†Harvard University, Department of Economics.
‡Harvard University, Department of Economics.

tocorrelation robust (HAR) standard errors (SEs). In econometrics, the dominant method for computing HAR SEs entails computing the Newey-West estimator of the Long Run Variance (LRV) matrix [Newey and West, 1987]. The Newey-West estimator uses the Bartlett, or triangle, kernel, which is a first-order kernel, meaning that the Parzen characteristic exponent, q, which is defined as the largest value of r such that the quantity $k^{[r]}(0) = \lim_{t\to 0} |t|^{-r}(k(0) - k(t))$ is defined and finite, is equal to 1.[1] It is well known that, for the problem of spectral estimation, the mean squared error (MSE) of first-order ($q = 1$) kernels is asymptotically dominated by that of second-order ($q = 2$) kernels, which are approximately quadratic at the origin, when the spectral density is smooth enough at the origin [Priestley, 1981]. In the HAR testing problem for the Gaussian location model (the model for which the HAR Edgeworth literature is most developed), a similar asymptotic dominance of second-order kernels has been shown if one focuses on a Type I/Type II error tradeoff (Sun et al. [2008]) or on a size-power tradeoff (Lazarus, Lewis, and Stock [2021], LLS). However, calculations in Lazarus, Lewis, Stock, and Watson [2018] suggest that, for sample sizes typically used in time series econometric applications, neither first-order nor second-order kernels dominate; typically, the size-power frontiers cross, so that, in some regions, the Bartlett kernel is preferred in finite samples to the optimal second-order kernel (the Quadratic Spectral (QS) kernel). Since even the optimal second-order kernel does not necessarily outperform first-order kernels, in particular, the Bartlett kernel, in finite samples, this raises the question of whether the Bartlett kernel is optimal among first-order kernels or, if not, what first-order kernel improves upon it.

It appears that the question of optimality among first-order kernels has received little attention, either in the classical spectral estimation literature or, more recently, in the HAR inference literature. Other first-order kernels, or tests that have first-order implied mean kernels (in the sense of LLS), include the split-sample (or 'batch mean') estimator of the LRV (Ibragimov and Müller [2010]) and the LRV estimator obtained by projecting the product of the covariates and the regression residuals onto the first m Legendre polynomials.

Classical results in spectral estimation show that the Mean Squared Error (MSE) of a kernel estimator of the spectral density, evaluated at the optimal rate for the sequence of truncation parameters, S, is increasing in $I_q[k] = \left(k^{[q]}(0)\right)^{\frac{1}{q}} \int k^2(t)dt$, where k is the kernel of interest, $k^{[r]}(0)$ was defined in the previous paragraph, and q is the largest value for which $k^{[r]}(0)$ is defined and finite. LLS show that, in the Gaussian location model, both the size-power and Type I/II error tradeoffs of HAR tests are increasing in $I_q[k]$ when Kiefer-Vogelsang Fixed-b inference is used (where $b = T^{-1}S$). Thus, kernels of the same order can be ranked, for both estimation and testing, by their values of $I_q[k]$, with smaller values preferred. Among second-order kernels, $I_q[k]$ is minimized by the so-called Quadratic Spectral (QS) kernel (Epanechnikov [1969]). Minimization of $I_q[k]$ over the class of first-order kernels appears to be unad-

[1]The use of brackets in the superscript is nonstandard, but the typical use of parentheses is easily confused with the more standard notation for the r^{th} derivative.

dressed.

In this paper, we demonstrate that, for $q < 2$, there is no optimal q^{th}-order kernel in the sense of minimizing $I_q[k]$. We further show that, for $q \leq 2$, the set of q^{th}-order, positive semidefinite kernels is not closed as a subset of the space of all positive-semidefinite kernels and that, indeed, all continuous q^{th}-order kernels, including the Bartlett kernel, can be decomposed into a weighted sum of q^{th} and second-order kernels, which suggests that there is no meaningful notion of 'pure' q^{th}-order kernels for $q < 2$. We additionally prove that the classical result that nonconstant, positive-semidefinite kernels have $q \leq 2$ holds without the traditional regularity conditions. Thus, our results apply to all nonconstant, symmetric, positive-semidefinite kernels for any value of q.

We provide a restricted family of first-order kernels among which the Bartlett kernel is optimal and we also show (by analytical calculations) that the Bartlett kernel produces HAR size-power tradeoffs that dominate other selected first-order kernels that do not fit into this class. Finally, we explore a collection of orthogonal series estimators and compare their performance to both their corresponding limiting implied mean kernel estimators and the Bartlett kernel. Despite considering a wide variety of first-order estimators, including all those commonly encountered in the econometric literature, we do not find any that dominate the Bartlett kernel, although a limiting implied mean kernel estimator based on the Haar system of wavelets is able to achieve parity asymptotically. These results suggest that the Newey-West estimator may, in fact, achieve a form of optimality among those using first-order kernels.

This research builds on a vast literature on HAR estimation and inference in models with time series variables. The seminal paper in econometrics is Newey and West [1987], which introduced the Newey-West LRV estimator in the context of HAR inference. Drawing on classical results in the literature on spectral density estimation (e.g., Grenander and Rosenblatt [1953], Brillinger [1975], and Priestley [1981]), Andrews [1991] characterized the optimal rate for the truncation parameter for minimizing the estimator mean squared error and, along with Newey and West [1994], 1994), proposed feasible LRV estimators to achieve the optimal estimation rate. Early Monte Carlo evidence, notably Newey and West [1994], showed, however, that LRV estimators with optimal estimation rates resulted in large size distortions. The asymptotic expansions of Velasco and Robinson [2001] and Sun et al. [2008] show that the leading higher order terms of the null rejection rate of the test are a weighted sum of the variance and the bias, not the squared bias, which enters the MSE. Accordingly, the size distortion can be reduced by using larger truncation parameters (Kiefer et al. [2000]) and by using the so-called fixed-b critical values of Kiefer and Vogelsang [2005] to account for the increased variability of these estimators. Jansson [2004], Sun et al. [2008], and Sun [2014] show that using fixed-b critical values provides a higher-order refinement to the null rejection rate of HAR test statistics in the Gaussian location model. In general, the fixed-b distributions of HAR test statistics are nonstandard, however, if the LRV estimator is computed as a projection onto a low-frequency orthogonal series, the HAR t- and F-statistics have fixed-b t- and F-distributions (Brillinger [1975], Phillips [2005], Müller [2007], and Sun [2013]). Müller [2014] and Lazarus, Lewis, Stock, and Watson [2018] provide additional surveys of this literature.

2 Kernels and Positive-Semidefinite Functions

Consider the problem of estimating the time series regression

$$y_t = x_t'\beta + u_t,\ t = 1, ..., T$$

where x_t and u_t are second-order stationary stochastic processes with $\mathbb{E}[u_t|x_t] = 0$. Inference for β require an estimator for the long-run variance Ω of $z_t = x_t u_t$. Formally, we need to estimate

$$\Omega = \sum_{i=-\infty}^{+\infty} \Gamma(i)$$

where $\Gamma(i) = Cov(z_t, z_{t-i})$ for $i = 0, \pm 1, \pm 2, ...$. This is challenging because there are infinitely-many autocovariances $\Gamma(i)$ but we have $T < \infty$ observations.

Kernel estimators of the long-run variance are widely used because, under certain assumptions on the kernel, the resulting estimators are guaranteed to have desirable properties such as nonnegativity. Letting k be the kernel function and S be a scaling (truncation) parameter, such estimators take the form

$$\hat{\Omega} = \sum_{i=1-T}^{T-1} k\left(\frac{i}{S}\right)\hat{\Gamma}(i),\ \text{where},\ \hat{\Gamma}(i) = \frac{1}{T}\sum_{t=\max(1,i+1)}^{\min(T,T+i)} \hat{z}_t\hat{z}_{t+i},$$

where $\hat{z}_t = x_t\hat{u}_t$ with $\hat{u}_t$ denoting the regression residual.

In variance estimation, kernels are required to be positive-semidefinite (psd), symmetric about 0, and normalized so that $k(0) = 1$. They are typically characterized by their behavior near zero. In particular, a kernel's *order* is determined by its Parzen characteristic exponent, q, which is defined as the largest r for which the quantity $k^{[r]}(0) = \lim_{t\to 0} |t|^{-r}(k(0) - k(t))$ is defined and finite. If $q = 1$ then we say that k is a first-order kernel, if $q = 2$ then then k is a second-order kernel, and so on. A special case is when the kernel k admits a series expansion of the form $k(t) = k(0) + \sum_{i=1}^{\infty} c_i|t|^i$ in which case the order of the kernel corresponds to the index of the first nonzero coefficient.[2] For convenience, and consistent with the standard notation for smoothness, we will let $q = \infty$ if $k^{[r]}(0)$ is defined and finite for all r.

[2]To see why, observe that $k(t) = k(0) - k^{[q]}(0)|t|^q + h(t)|t|^q$ for some $h(t)$, where $\lim_{t\to 0} h(t) = 0$, which implies that if k admits a series expansion, as above, $k^{[r]}(0) = \lim_{t\to 0}|t|^{-r}(k(0) - k(t)) = -\sum_{i=1}^{\infty} c_i \lim_{t\to 0}|t|^{i-r}$, so q will be the smallest i for which $c_i \neq 0$.

Positive-semidefiniteness is an important property for the kernel to satisfy because it ensures that the estimator long-run variance is nonnegative. A function f is called positive-semidefinite (psd) if, for any $n \in \mathbb{N}$ and any $t \in \mathbb{R}^n$, the matrix M with entries $M_{ij} = f(t_i - t_j)$ is also positive-semidefinite. We now derive a number of facts about such functions, which will be of use in characterizing positive-semidefinite kernels.[3]

Lemma 1. *Let f be a symmetric, positive-definite function, then $f(0) \geq 0$, $|f(t)| \leq f(0)$, and $|f(t+h) - f(t)| \leq \sqrt{2f(0)|f(t) - f(0)|}$. Thus is f is continuous at 0, then it is everywhere uniformly continuous. If $f^{[2]}(0)$ and $f'(t)$ exist, then $|f'(t)| \leq \sqrt{2f(0)f^{[2]}(0)}$; if $f''(t)$ exists as well, then $|f''(t)| \leq 2f^{[2]}(0)$.*

Lemma 1 reveals that a psd kernel satisfies $k(t) \leq k(0)$ and that $k^{[q]}(0) \geq 0$. Next, we use the previous lemma to establish a variety of important facts about the connection between the characteristic exponent of a symmetric, positive-semidefinite function and its global structure.

Lemma 2. *Let f be a symmetric, positive-semidefinite function with characteristic exponent q. Then,*

1. *For all $r < q$, $f^{[r]}(0) = 0$*
2. *f is constant if and only if $f^{[2]}(0) = 0$*
3. *If $q > 0$, then f is uniformly continuous.*
4. *If $q = 2$, then $f^{[2]}(0) > 0$*

Proposition 3. *If f is a symmetric, positive-semidefinite function with characteristic exponent, q, then either $q \leq 2$ or $q = \infty$. Further, f is constant if and only if $q = \infty$*

An immediate consequence of Proposition 3 is the following theorem.

Theorem 4. *If f is a nonconstant, symmetric, positive-semidefinite function with characteristic exponent, q, then $q \leq 2$.*

This result is well-known in the literature, but is traditionally proven under strong regularity conditions, typically involving the existence and integrability of products of the Fourier Transform of the kernel and quadratic polynomials. One can remove the existence assumptions, at the cost of requiring continuity, by invoking Bochner's Theorem, which, specialized to our purposes, states that the class of continuous, symmetric, positive-semidefinite functions with $f(0) = 1$, is exactly the set of (inverse) Fourier Transforms of symmetric Borel probability measures (it is more generally formulated with respect to Radon measures on locally compact topological groups). In this setting, the integrability conditions can be interpreted as second-moment conditions, however, they cannot be easily removed.

[3] All proofs are contained in the Online Appendix except that of Theorem 4, which is an immediate corollary of Proposition 3.

Before proceeding further, we define some notation. A function, f, will be said to be integrable with respect to some measure, μ, if f is μ-measurable and $\int |f| d\mu < +\infty$. L^p spaces are defined in the standard way, so that $f \in L^p(\mu)$ means that $|f|^p$ is μ-integrable. If μ is not specified, the measure is assumed to be the Lebegesgue measure. $\mathcal{F}$ denotes the Fourier transform with $\hat{k}(\omega) = (\mathcal{F}k)(\omega) = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-i\omega t} k(t) dt$ if $k \in L^1$, so that, if $\hat{k} \in L^1$, $k(t) = \left(\mathcal{F}^{-1}\hat{k}\right)(t) = \int_{-\infty}^{\infty} e^{i\omega t} \hat{k}(\omega) d\omega$. When $\hat{k}(\omega) \geq 0$ and $\int_{-\infty}^{\infty} \hat{k}(\omega) d\omega = 1$, $\hat{k}$ can be regarded as a probability density function with $k(t)$ as its characteristic function. Additionally, we define the (inverse) Fourier transform of a finite (possibly signed or complex) Borel measure by $(\mathcal{F}^{-1}\mu)(t) = \int_{-\infty}^{\infty} e^{i\omega t} d\mu$. Note that this convention is somewhat unusual, but is standard in the kernel literature because the (inverse) Fourier Transform of a probability measure is then simply its characteristic function. It also ensures that, if μ is a probability measure and $k = \mathcal{F}^{-1}\mu$, then $k(0) = 1$. We will use $D_x^r f$ and $\partial_x^r f$ to denote the r^{th} derivative of f with respect to x.

3 Non-Existence of Optimal Kernels of Less Than Second-Order

In this section, we establish that there is no optimal first-order ($q = 1$), positive-semidefinite kernel that minimizes $I_q[k]$. Let $q_1 < q_2 \leq 2$ and, for $i = 1, 2$, let $k_i(t)$ be a symmetric, positive-semidefinite kernel with characteristic exponent q_i, $k_i(0) = 1$, $k_i^{[q_i]}(0) = p_i > 0$, and $\int k_i^2(t) dt = M_i < 1$. Then, by Lemma 2, $k_2^{[q_1]}(0) = 0$. For $\varepsilon \in (0, 1]$, let $k_\varepsilon = \varepsilon k_1 + (1-\varepsilon) k_2$. Since the kernel k_i is normalized so that $k_i(0) = 1$ for $i = 1, 2$, it follows that $k_\varepsilon(0) = 1$. This implies that

$$\begin{aligned} k_\varepsilon^{[r]}(t) &= \lim_{t \to 0} |t|^{-r}(1 - k_\varepsilon(t)) \\ &= \varepsilon \lim_{t \to 0} |t|^{-r}(1 - k_1(t)) + (1 - \varepsilon) \lim_{t \to 0} |t|^{-r}(1 - k_2(t)) \\ &= \varepsilon k_1^{[r]}(0) + (1 - \varepsilon) k_2^{[r]}(0) \end{aligned}$$

and therefore $k_\varepsilon^{[q_1]}(0) = \varepsilon p_1$ because $k_1^{[q_1]}(0) = p_1$ and $k_2^{[q_1]}(0) = 0$. Note that, since $k_i^{[q_i]}(0) = p_i > 0$, for $r > q_i$, $k_i^{[r]}(0) = \lim_{t \to 0} |t|^{-r}(k_i(0) - k_i(t)) = \lim_{t \to 0} |t|^{q_i - r} |t|^{-q_i}(k_i(0) - k_i(t)) = \infty \cdot p_i = \infty$. Thus, for $r > q_1$, $k_1^{[r]}(0) = 1$ and, since $0 \leq k_2^{[r]}(0)$, $k_\varepsilon^{[r]}(0) = \varepsilon k_1^{[r]}(0) + (1 - \varepsilon) k_2^{[r]}(0) = \infty$ and k_ε has characteristic exponent $q = q_1$ and so $k_\varepsilon^{[q]}(0) = \varepsilon p_1$. The (weighted) sum of symmetric, positive-semidefinite functions is again symmetric and positive-semidefinite, since the sum of symmetric, positive-semidefinite matrices is also symmetric and positive-semidefinite. Therefore, we can produce valid symmetric, positive-semidefinite kernels with characteristic exponent $q = q_1$ possessing arbitrarily small values of $k^{[q]}(0)$. In order to simplify notation in what follows, let $p = p_1$. Then,

$$I_q[k_\varepsilon] = \left(k_\varepsilon^{[q]}(0)\right)^{\frac{1}{q}} \int k_\varepsilon^2(t)dt = \varepsilon^{\frac{1}{q}} p^{\frac{1}{q}} \int \left[\varepsilon^2 k_1(t)^2 + 2\varepsilon(1-\varepsilon)k_1(t)k_2(t) + (1-\varepsilon)^2 k_2^2(t)\right] dt$$

We then use Hölder's inequality and that $\varepsilon \in (0,1]$, $\varepsilon(1-\varepsilon) \leq \frac{1}{4}$, and $\int k_i^2(t)dt = M_i < +\infty$ to conclude that

$$I_q[k_\varepsilon] \leq \varepsilon^{\frac{1}{q}} p^{\frac{1}{q}} \left(M_1^{\frac{1}{2}} + M_2^{\frac{1}{2}}\right)^2.$$

Therefore, for any $q < 2$, we can make $I_q[k]$ arbitrarily small by mixing sufficiently small amounts of a square-integrable, symmetric, positive-semidefinite kernel with characteristic exponent q with any higher-order, square-integrable, symmetric, positive-semidefinite kernel (which is necessarily nonconstant, since then $k = 1$, which is not square-integrable). Specifically, we can choose ε so that $I_q[k_\varepsilon]$ takes any value in $(0, p^{\frac{1}{q}} M_1]$, since $I_q[k_1] = p^{\frac{1}{q}} M_1$ and $I_q[k_\varepsilon]$ is continuous in ε. Specializing to the Bartlett kernel, $k_1(t) = (1-|t|)\mathbf{1}_{[-1,1]}(t)$, which is used by the Newey-West variance estimator, $M_1 = \frac{2}{3}$ and $k^{[1]}(0) = 1$, so $I_q[k_\varepsilon]$ can take any value in $(0, \frac{2}{3}]$. Therefore, for $q < 2$, there can be no optimal kernel using the same optimality criterion used for second-order kernels (ensuring that it provides the optimal asymptotic size-power tradeoff). The reason that this argument does not apply to second-order, symmetric, positive-semidefinite kernels is because, as shown in Theorem 4, any nonconstant, symmetric, positive-semidefinite kernel must have $q \leq 2$. Thus, there are no nonconstant, symmetric, positive-semidefinite kernels with $q > 2$ with which to mix. (If one were to mix a second-order, symmetric, positive-semidefinite kernel, k_1, with the constant function, $k_2 = 1$, then, for $\varepsilon \in [0,1)$, the resulting kernel, k_ε would fail to be square-integrable, and, in particular, $\int k_\varepsilon^2(t)dt = \infty$ Thus, for $\varepsilon \in (0,1]$, $I_q[k_\varepsilon] = 1$, since, from Lemma 2, $k_1^{[2]}(0) > 0$, so $k_\varepsilon^{[2]}(0) = \varepsilon k_1^{[2]}(0) > 0$, as well.)

This also gives us a topological interpretation of why there can be no optimal kernel with characteristic exponent $q < 2$: the set of q^{th}-order, symmetric, positive-semidefinite kernels is not a closed subset of the set of all symmetric, positive-semidefinite kernels, so sequences of q^{th}-order kernels may have limits that are instead higher-order (consider k_ε as ε goes to 0).[4] Said differently, the closure of the (open) set of first-order kernels is the set of all positive-semidefinite kernels.

It is also true that there exist sequences of second-order kernels which have lower-order kernels as their limits. Let k be a continuous, symmetric, positive-semidefinite kernel with $k(0) = 1$ and characteristic exponent q. Then, by Bochner's Theorem, $k = \mathcal{F}^{-1}(\mu)$for some Borel probability measure μ. Clearly, there exists some $a' > 0$ such that $\int_{-a}^{a} d\mu > 0$ for all $a \geq$

[4]Specifically, if k_1, k_2 are elements of some normed vector space $(V, ||\cdot||)$, so is k_ε and $||k_\varepsilon - k_2|| = \varepsilon||k_1 - k_2||$ so $\lim_{\varepsilon\to 0} k_\varepsilon = k_2$. Since, by Lemma 1 $|k(t)| \leq k(0)$, positive-semidefinite kernels are bounded so we can take the norm to be the supremum norm, and, thus, $k_\varepsilon \to k_2$ uniformly.

a'. Now, consider the family of kernels $k_a = \mathcal{F}^{-1}(\mu_a)$, where $d\mu_a = (\mu([-a,a]))^{-1} I_{[-a,a]} d\mu$ and $a \geq a'$. Since μ_a is also a Borel probability measure, by the converse of Bochner's Theorem, the k_as are all continuous, positive-semidefinite as well, but by Lemma 10 of Section 4, they are second-order. Using the total variation norm,

$$||\mu_a - \mu||_{TV} = \left(\mu([-a,a])^{-1} - 1\right)\mu([-a,a]) + \mu([-a,a]^c) = 2(1 - \mu([-a,a])).$$

Thus, $\lim_{a\to\infty} ||\mu_a - \mu||_{TV} = 0$, so $\lim_{a\to\infty} \mu_a = \mu$ in the total variation topology. Since, for any finite (possibly signed or complex) Borel measure, ν, $|(\mathcal{F}^{-1}\nu)(\omega)| = |\int_{-\infty}^{\infty} e^{it\omega} d\nu(t)| \leq \int_{-\infty}^{\infty} |e^{it\omega}| d|\nu|(t) = ||\nu||_{TV}$, we have that $||(\mathcal{F}^{-1}\nu)||_\infty \leq ||\nu||_{TV}$. As the inverse Fourier transform is linear, it follows that $\lim_{a\to+\infty} ||k_a - k||_\infty = \lim_{a\to} ||\mathcal{F}^{-1}\mu_a - \mathcal{F}^{-1}\mu||_\infty = \lim_{a\to\infty} ||\mathcal{F}^{-1}(\mu_a - \mu)||_\infty \leq \lim_{a\to\infty} ||\mu_a - \mu||_{TV} = 0$, so k_a converges to k uniformly as $a \to +\infty$. If μ has a density $f \in L^2$, then, since the Fourier Transformation is a unitary transformation on L^2, one can also show that $k_a \to k$ in the topology induced by the L^2 norm, by a similar argument. A simple example of a collection of second-order kernels with a limiting first-order kernel is the family of kernels $k_a = \mathcal{F}^{-1}\left(c_a \omega^{-2}(1 - \cos(\omega)) I_{[-a,a]}(\omega)\right)$ with $c_a = \left(\int_{-a}^{a} \omega^{-2}(1 - \cos(\omega)) d\omega\right)^{-1}$. For any finite $a > 0$, this yields a second-order kernel. However, since $\mathcal{F}\left((1 - |t|) I_{[-1,1]}(t)\right)(\omega) = \pi^{-1}\omega^{-2}(1 - \cos(\omega))$, as $a \to +\infty$,the limiting kernel (with respect to both the supremum and L^2 norms) is the Bartlett kernel, which is first-order.

Taken together, these results tell us that, for any $q \leq 2$, the space of q^{th}-order, symmetric, positive-semidefinite kernels is not a closed subset of the set of all symmetric, positive-semidefinite kernels.

We summarize the above results in the following Theorem.

Theorem 5. *Let k_0 be a square-integrable, symmetric, positive-semidefinite kernel with characteristic exponent q, $k_0(0) = 1$, $\int k_0^2(t) dt = M < \infty$, and $k_0^{[q]}(0) = p$. Then, for any $I \in (0, p^{\frac{1}{q}} M]$, there exists a q^{th}-order positive-semidefinite kernel k, with $I_q[k] = \left(k^{[q]}(0)\right)^{\frac{1}{q}} \int k^2(t) dt = I$. In particular, if $q = 1$, there exists such a k for any $I \in (0, \frac{2}{3}]$. For any $q \leq 2$, the set of q^{th}-order positive-semidefinite kernels is not a closed subset of the set of all positive-semidefinite kernels.*

Proof. As above. $\square$

Several remarks are in order. We start with a discussion of higher order ($q > 2$) kernels. As the only valid normalized positive-semidefinite function with $q > 2$ is $k = 1$, by Proposition 3, it is necessarily optimal among all kernels with $q > 2$ and $k(0) = 1$, even though we cannot define $I_q[k]$ in the standard way. Further, $k = 1$ cannot be decomposed as a weighted sum involving lower-order kernels, since, as shown above if $k_\varepsilon = \varepsilon k_1 + (1-\varepsilon)k_2$, its characteristic exponent is equal to the smallest of the characteristic exponents of k_1 and k_2. Finally, while it is easy to find a sequence of q^{th}-order kernels that converges to $k=1$, say $k_\varepsilon = \varepsilon k_1 + (1-\varepsilon)$, so that $k_\varepsilon \to 1$ uniformly as $\varepsilon \to 0$, clearly, no sequence of constant functions can converge to a nonconstant function and, thus, to a kernel with $q \leq 2$.

Next,we discuss the requirement that $k^{[q]}(0) > 0$. There are several reasons for this. The first is that the characteristic exponent is too coarse to completely characterize the behavior of a kernel at the origin, as we will see in Section 4.1. In particular, it assumes that, at the origin, all kernels must behave like $|t|^q$ to leading order. However, this ignores the possibility of intermediate rates of convergence like $|t|^q|\log(|t|)|^p$ or $|t|^q|log(|t|)|^{-p}$, which go to $k(0)$ slightly slower or faster than $|t|^q$, respectively. The examples in Section 4.1 have such rates of convergence and, so, should not be regarded as a true q^{th}-order kernels. Additionally, our interest in $I_q[k]$ arises from the fact that it occurs in the coefficient of the leading term in certain expansions, such as the Edgeworth expansion in LLS. If $I_q[k] = 0$ because $k^{[q]}(0) = 0$, then, by definition, it cannot be part of the coefficient of the leading term (unless the function is identically 0). Thus, it is logical to consider only kernels for which the characteristic exponent, q, is well defined and $k^{[q]}(0) > 0$ to be true q^{th}-order kernels.

It turns out that we can use Fourier analysis to derive a general decomposition of continuous, symmetric, and positive-semidefinite kernels. Although the derivations focus on continuous kernels, this is a relatively unrestrictive assumption because Lemmas 1 and 2 establish that this will occur whenever $q > 0$ or the kernel is continuous at the origin. Moreover, continuous kernels dominate empirical practice. The results substantially generalize the discussion in this section in that they establish a similar weighted-sum representation for any nonconstant q^{th}-order, continuous, symmetric, positive-semidefinite kernel k that satisfise $k(0) = 1$ and $k^{[q]}(0) > 0$ and show that there are infinitely-many of these decompositions. Precisely, the theory states that any continuous, nonconstant, and symmetric q^{th}-order positive-semidefinite kernel can be decomposed into another q^{th}-order positive-semidefinite and a second-order positive-semidefinite kernel. Consequently, there is no natural notion of a minimal q^{th}-order kernel for $q < 2$. As these results are more of theoretical interest, they are included in the Online Appendix.

Finally, the reader may have noticed by now that we have always been very careful to require that our kernels have defined values of q. Is this actually necessary? It turns out that one can construct a kernel that does not possess a characteristic exponent, meaning that there is no maximal value of r for which $k^{[r]}(0)$ is finite, but $k^{[r]}(0)$ is not defined and finite for all r (which would mean that its characteristic exponent is, by definition, infinity). We view this as being of theoretical interest and, for that reason, also defer it to the Online Appendix.

4 Other Estimators of the Long-Run Variance

4.1 Other First-Order Kernels

A natural starting point for further exploration of first-order kernels is other Bartlett-like kernels. Two straightforward generalizations of the Bartlett kernel are the families $k_p(t) = \left(1 - \left|\frac{t}{b}\right|\right)^p \mathrm{I}_{[-1,1]}\left(\frac{t}{b}\right)$ and $k_r(t) = \left(1 - \left|\frac{t}{b}\right|^r\right) \mathrm{I}_{[-1,1]}\left(\frac{t}{b}\right)$. The k_p kernel has been referred to in the econometrics literature as the sharp origin kernel and has been studied extensively in Phillips, Sun, and Jin [2006, 2007], and Sun, Phillips, and Jin [2011]. It is somewhat surprising that the first choice gives positive definite kernels for $p \geq 1$, but that the second class does not. However, it is easy to see why $k_p(x)$ gives a positive-semidefinite kernel, at least for integer values of p, since it is simply the p^{th} power of the Bartlett kernel so, using the fact that the Fourier Transform turns multiplication into convolution, $\hat{k}_p = \mathcal{F}\left(k_1^p\right) = \hat{k}_1^{*p} \geq 0$, since $\hat{k}_1 \geq 0$, where $\hat{k}_1^{*p}$ is the p^{th} convolutional power of $\hat{k}_1$. For k_p, we compute $k_p^{[1]}(0) = \lim_{t\to 0} \frac{1-\left(1-\left|\frac{t}{b}\right|\right)^p}{|t|} = \lim_{t\to 0} \frac{p\left|\frac{t}{b}\right|}{|t|} = \frac{p}{b}, \int k_p(t)^2 dt = \int_{-b}^{b} \left(1 - \left|\frac{t}{b}\right|\right)^{2p} dt = \frac{2b}{2p+1}$, so $I_q[k] = \frac{p}{b} \cdot \frac{2b}{2p+1} = \frac{2p}{2p+1} \geq \frac{2}{3}$. Therefore, within this family, the Bartlett kernel offers the best performance.

It is also interesting to consider what happens in the limit as $b \to \infty$ if p also scales with b. Let $p = cb$, then $k_{p(b)}(t) = \left(1 - \left|\frac{t}{b}\right|\right)^{cb} \mathrm{I}(|t| \leq b)$ and $\lim_{b\to\infty} k_{p(b)}(t) = e^{-c|t|}$, which has the Fourier Transform $\frac{c}{c^2+\omega^2} \geq 0$. This gives another natural first-order, positive-semidefinite kernel, which seems to be used very uncommonly. Perhaps this is because $k^{[1]}(0) = c$ and $\int k(t)^2 dt = c^{-1}$ so $I_q[k] = c \cdot c^{-1} = 1 > \frac{2p}{2p+1}$ for any finite value of $p \geq 1$.

However, it is surprisingly difficult to construct families of first-order positive-semidefinite functions. An alternative approach has been to instead consider orthogonal series estimators, which we discuss next.

4.2 Orthogonal Series Estimators

4.2.1 Basic Principles

Weighted Orthogonal Series (WOS) estimators take a different approach to estimating the long run variance from traditional kernel estimators. Starting with some orthonormal basis in the Hilbert space of square integrable functions on $[0,1]$, $\{\phi_i\}_{i=0}^{\infty}$, such estimators project the sequence $\{\hat{z}_t\}_{t=1}^{T}$ onto one of the basis functions (excepting $\phi_0 = 1$) and then compute the empirical variance $\hat{\Omega}_i$ from the projection (recall that $\hat{z}_t = x_t \hat{u}_t$ where $\hat{u}_t$ is the OLS residual from the time series regression $y_t = \beta x_t + u_t$). These estimates are then combined via a weighted sum to give a final estimate $\hat{\Omega}$. Formally,

$$\hat{\Omega} = \sum_{i=1}^{B} w_i \hat{\Omega}_i,\ \sum_{i=1}^{B} w_i = 1,\ w_i \geq 0\,\forall\, i,\ \hat{\Omega}_i = \hat{\Lambda}_i \hat{\Lambda}_i',\ \hat{\Lambda}_i = T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_i\left(\frac{t}{T}\right) \hat{z}_t$$

As $\hat{\Omega}_i$ is positive-semidefinite for all i with probability 1, it follows $\hat{\Omega}$ is positive-semidefinite with probability 1. We remark that a researcher may restrict attention to $\{w_i\}_{i=b_0}^{B}$ for some $b_0 > 1$ and the idea remains the same except that the adding up constraint is now $\sum_{i=b_0}^{B} w_i = 1$; for ease of exposition, we set $b_0 = 1$ but generalize the discussion in the Online Appendix.

Associated with each orthogonal series estimator, and sequence of weights $\{w_i\}_{i=1}^{B}$ is a limiting implied mean kernel,

$$k(t) = \sum_{i=1}^{B} w_i k_i\left(\frac{t}{B}\right),\ k_i(t) = \int_{\max(0,t)}^{\min(1,1+t)} \phi_i(u)\phi_i(u-t)du$$

(The nonlimiting implied mean kernel is a sum over discrete time points; see LLS for more details.)

For finite T, this provides a connection between orthogonal series estimators and kernel estimators. Specifically, LLS show that, similarly to the case for kernel estimators, for both estimation and testing, the performance of Weighted Orthogonal Series estimators is characterized by the quantity $I_q[k] = \left(k^{[q]}(0)\right)^{\frac{1}{q}} \sum_i w_i^2$, where $w_i s$ are the weights. Note that, compared to the case of kernel estimators, $\sum_{i=1}^{B} w_i^2$ has replaced $\int k^2(t)dt$. This results in important differences in performance between the two classes of estimators. For any sequence $k_i^{[q]}(0)$, we can solve

$$\min\left\{\left(k^{[q]}(0)\right)^{\frac{1}{q}} \sum_{i=1}^{B} w_i^2 : \sum_{i=1}^{B} w_i = 1,\ w_i \geq 0\ \forall\, i \in \{1, ..., B\}\right\}$$

to determine the optimal weights $\{w_i\}_{i=1}^{B}$. For first-order kernels, we obtain a relatively simple closed form expression:

$$w_i = \frac{1}{B} - \left(k_i^{[1]}(0) - \frac{1}{B}\sum_{j=1}^{B} k_j^{[1]}(0)\right) A\left(k^{[1]}(0)\right) \tag{1}$$

where A is a known function of the $k_i^{(1)}(0)$s, which shows that the weights decrease linearly with increasing i. The derivation is presented in detail in the Online Appendix and considers the general case $\{w_i\}_{i=b_0}^{B}$ for some $b_0 \geq 1$.

4.2.2 Legendre Polynomials

We discuss two important classes of Weighted Orthogonal Series estimators. The first is the Legendre Polynomials, which form an orthogonal series on $[-1, 1]$. After remapping their domains to $[0, 1]$ and normalizing by multiplying by $\sqrt{2i+1}$ so that $\int \phi_i(x)\phi_j(x) = \delta_{ij}$, we can compute the limiting implied mean kernels $k_i(t) = \int_{\max(0,t)}^{\min(1,1+t)} \phi_i(u)\phi_i(u-t)du$ and $k^{[1]}(0)$. Based on an asymptotic expansion, LLS compute the limiting value of a normalized version of $k^{[1]}(0)$ for the Weighted Orthogonal Series estimator. However, we need to compute $k^{[1]}(0)$, for a single limiting implied mean kernel, which we do directly from the definition and summarize with a proposition (see Online Appendix for proof).

Proposition 6. *If* $k(t) = \int_{\max(0,t)}^{\min(1,1+t)} \phi(u)\phi(u-t)du$, $k(0) = 1$, *and* $\phi \in \mathcal{C}^1[0,1]$ *then,* $k^{[1]}(0) = \frac{1}{2}\left(\phi(0)^2 + \phi(1)^2\right)$*. Further, if* $\phi \in \mathcal{C}^n[0,1]$*, then, if* n *is odd,*

$$\lim_{t\to 0^+} \frac{d^n k}{dt^n}(t) = \frac{1}{2}\sum_{i=1}^{n}(-1)^i \left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) + \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] = -\lim_{t\to 0^-} \frac{d^n k}{dt^n}(t),$$

while, if n *is even,*

$$\lim_{t\to 0^+} \frac{d^n k}{dt^n}(t) = \lim_{t\to 0^-} \frac{d^n k}{dt^n}(t) = \int_0^1 \phi(u)\phi^{(n)}(u)du.$$

Additionally, if $k \in \mathcal{C}^{n-1}[-1,1]$*, these equalities hold with the one-sided derivatives at* 0*,* $\frac{d^n k}{dt^{n\pm}}(0)$ *in place of the limits* $\lim_{t\to 0^\pm} \frac{d^n k}{dt^n}(t)$*, so* $\frac{d^n k}{dt^{n\pm}}(0) = \lim_{t\to 0^\pm} \frac{d^n k}{dt^n}(t)$ *and* $k \in \mathcal{C}^n[-1,1]$ *if the limits are equal. Finally, if* ϕ *generates a second-order kernel, then,* $k \in \mathcal{C}^2[-1,1]$ *and* $k^{[2]}(0) = -\frac{1}{2}\frac{d^2 k}{dt^2}(0) = -\frac{1}{2}\int_0^1 \phi(u)\phi^{(2)}(u)du$*.*

Note that this agrees with the expression given for the limiting implied mean kernel derived by LLS. Then, for the orthonormal series on $[0,1]$ generated by the Legendre Polynomials, $k_i^{[1]}(0) = 2i+1$. Using the optimal weights, as given in equation (1) gives

$$\lim_{B\to\infty} B^{-1}k^{[1]}(0) = \lim_{B\to\infty} B^{-1}\sum_{i=1}^{B} w_i k_i^{[1]}(0) = \frac{2}{3}$$

and $\lim_{B\to\infty} B\sum_{i=1}^{B} w_i^2 = \frac{4}{3}$, so $I_q[k] = \left(k^{[q]}(0)\right)^{\frac{1}{q}}\sum_{i=1}^{B} w_i^2 \to \frac{8}{9}$ as $B\to\infty$ (see appendix for details). If we had, instead, used equal weights, we would get $\lim_{B\to\infty} B^{-1}k^{[1]}(0) = B^{-2}\sum_{i=1}^{B}(2i+1) = \lim_{B\to\infty} B^{-2}[(B+1)B + B] = 1$ and $B\sum_{i=1}^{B} w_i^2 = B\sum_{i=1}^{B} B^{-2} = 1$ so $I_q[k] \to 1$ as $B \to \infty$. Interestingly, the equal-weighted orthogonal series estimator is asymptotically equivalent to (i.e. has the same $I_q[k]$ as) the exponential kernel, $k(t) = e^{-c|t|}$, in terms of performance for estimation and testing. Also, the optimal WOS Legendre polynomial series estimator is asymptotically equivalent to the k_p kernel with $p = 4$, that is $k_4(t) = \left(1 - \left|\frac{t}{b}\right|\right)^4 I_{[-1,1]}\left(\frac{t}{b}\right)$. Figure 1 shows the limiting implied mean kernels associated with the first few Legendre polynomials as well as their Fourier Transforms and optimally weighted sums. As we mentioned previously, due to the fact that $I_q[k] = \left(k^{[q]}(0)\right)^{\frac{1}{q}}\sum_{i=1}^{B} w_i^2$ for Weighted Orthogonal Series but $I_q[k] = \left(k^{[q]}(0)\right)^{\frac{1}{2}}\int k^2(t)dt$ for positive-semidefinite kernels, using an WOS estimator will result in a different size-power tradeoff than using the limiting implied mean kernel associated with the WOS. Table 1 shows numerically computed values of $I_q[k]$ using the limiting implied mean kernel for the first B Legendre polynomials (excepting $\phi_0 = 1$) for both equal and numerically estimated optimal weights. It is somewhat surprising that, when using the limiting implied mean kernel, we require only the first three Legendre Polynomials (LPs) with optimal weights, and the first two LPs, with equal weights, to achieve values of $I_q[k]$ that are lower than the asymptotic limiting values of the corresponding WOS estimators. At least in this case, using the limiting implied mean kernels, instead of using the orthogonal series directly, leads to a superior asymptotic bias-variance or size-power tradeoff. This result is similar to a classical result of Grenander and Rosenblatt [1957].

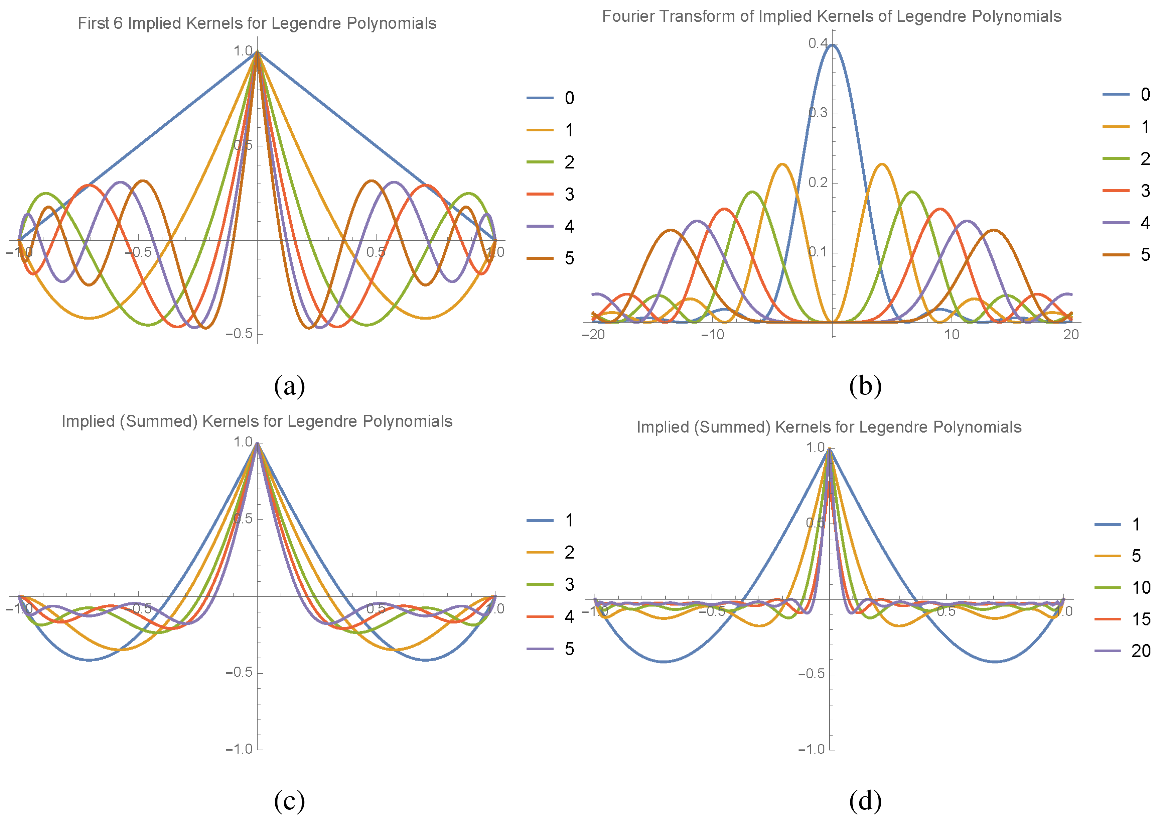

Figure 1: (A) Limiting implied mean kernels for each of the first 6 Legendre polynomials (including $\phi_0 = 1$). (B) Fourier Transforms of each of the first 6 Legendre polynomials (including $\phi_0 = 1$). (C) Optimally weighted sums using the first (1, 2, 3, 4, 5) Legendre polynomials. (D) Optimally weighted sums using the first (1, 5, 10, 15, 20) Legendre polynomials.

4.2.3 Haar Induced Kernels

Expanding on the idea of using an orthonormal series for variance estimation, we now consider the so-called Haar system (of wavelets). The Haar wavelets are defined by the wavelet function $\psi(x) = I_{[0,\frac{1}{2})}(x) - I_{[\frac{1}{2},1)}(x)$, so that the Haar basis functions are given by 1 and $\psi_{n,\ell}(x) = 2^{\frac{n}{2}}\psi(2^n x - \ell)$ with $n, \ell \in Z_+,\ 0 \leq \ell < 2^n$. This collection forms a complete orthonormal basis for $L^2[0,1]$. Their induced kernels are given by $k_n(t) = (1-3{\cdot}2^n|t|)I_{[0,2^{-(n+1)})}(|t|) - (1-2^n|t|)I_{[2^{-(n+1)},2^{-n})}(|t|)$ (Figure 2). Note that $k_n(t)$ is independent of ℓ because, within each level of the hierarchy, the basis functions are translations of each other. Thus, it is clear that k_n is a first-order kernel with $k_n^{[1]}(0) = 3 \cdot 2^n$. Let $n' > n$, then we also have, $\int k_n^2(t)dt = \frac{1}{3}2^{-n}$, $\int k_{n'}(t)k_n(t)dt = 2^{n-2n'-1}$ (see Online Appendix for derivations). Using these expressions, for any sequence of weights $\{w_i\}_{i=1}^N$, we can construct the limiting implied mean kernel $k(t) = \sum_{i=1}^N w_i k_i(t)$, which will also be a first-order kernel, since it is the sum of a finite number of first-order kernels. We can use $k(t)$ to construct a kernel estimator, which we can then optimize with respect to the weights. Somewhat surprisingly, weighting each basis function equally, so

	$I_q[k]$	
B	Equal	Optimal
1	1.02857	1.02857
2	0.98355	0.974026
3	0.928701	0.876429
4	0.891171	0.855479
5	0.86494	0.825379
6	0.845773	0.814168
7	0.831214	0.799687
8	0.8198	0.792678
9	0.810619	0.784183
10	0.80308	0.779378
15	0.779385	0.76074
20	0.766907	0.751449
25	0.759211	0.745529
30	0.753991	0.741595
40	0.747362	0.736528
50	0.743329	0.733434

Table 1: Values of $I_q[k]$ using the limiting mean kernel implied by the first B Legendre Polynomials (excepting $\phi_0 = 1$) using either equal and numerically estimated optimal weights.

that $w_i = \frac{2^n}{2^{N+1}-1}$, where N is the highest level of the Haar system used (since there are 2^n basis functions at each level of the hierarchy), performs almost as well as using optimal weights (Table 2). Denoting the equal weighted limiting implied mean kernel by $k_e(t)$, we find that $k_e^{[1]}(0) = 2^{N+1} + 1$, $\int k_e^2(t)dt = \frac{2}{3} \cdot 2^{-(N+1)}$, and, thus, $I_q[k_e] = \frac{2}{3}(1 + 2^{-(N+1)})$ (see Online Appendix). Therefore, the kernel estimator constructed using the equal-weighted limiting implied mean kernel for the Haar system asymptotically achieves the same size-power tradeoff as the Bartlett kernel.

However, using equal weights for an Orthogonal Series estimator gives $\sum_{i=1}^{N} w_i^2 = (2^{N+1} - 1)^{-1}$, where the weights here are for each individual basis function $\psi_{n,\ell}$ and not for the limiting implied mean kernels k_n, as in the previous expression. This gives $I_q[k] = k^{[1]}(0) \sum_i w_i^2 = \frac{2^{N+1}+1}{2^{N+1}-1}$, which converges to 1 as $N \to +\infty$. This finding agrees with the fact that $I_q[k] = 1$ asymptotically for Ibragimov and Müller's Split Sample estimator, which has been shown to be equivalent to an orthogonal series estimator using the Haar system as the basis functions when $B = 2^n - 1$ for some $n \in N$. If we instead use optimal weights for the Haar orthogonal

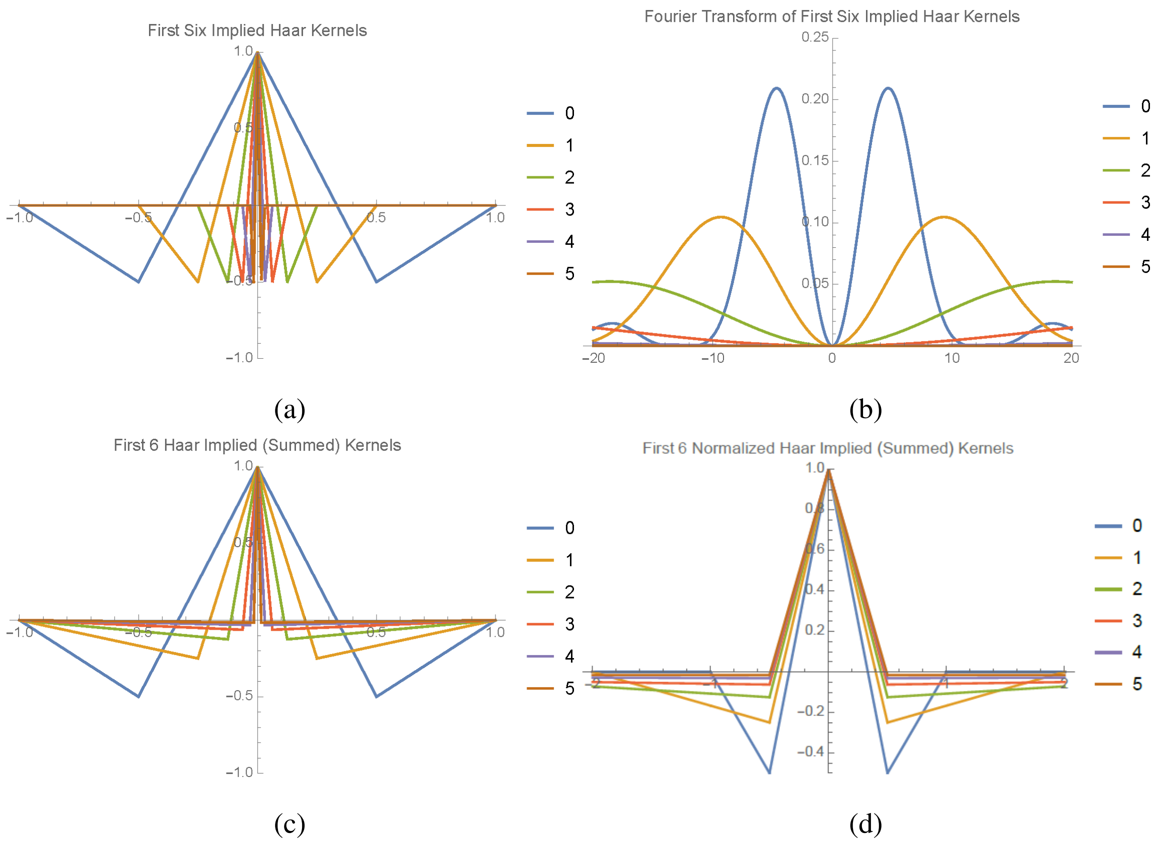

Figure 2: (A) Limiting implied mean kernels for each of the first 6 levels of the Haar system. (B) Fourier Transforms of each of the first 6 levels of the Haar system. (C) Optimally weighted sums using the first 6 levels of the Haar system. (D) Optimally weighted sums using the first 6 levels of the Haar system, normalized to allow for comparison of shape

	$I_q[k]$	
B	Equal	Optimal
0	1	1
1	0.833333	0.831539
2	0.75	0.742463
3	0.708333	0.703763
4	0.6875	0.684976
5	0.677083	0.675771

Table 2: Values of $I_q[k]$ based on the number of levels of the Haar hierarchy used (B). Note that $I_q[k]$ is bounded from below by $\frac{2}{3}$.

series estimator, we obtain a somewhat lower limiting value of $I_q[k] \approx .91$, which is a modest improvement over the equal weighted case, but is still far from the limiting value of $I_q[k] = \frac{2}{3}$

attained by the Haar limiting implied mean kernel or Bartlett kernel.

We see that the kernel estimator that uses the equal-weighted limiting implied mean kernel based on the Haar system approaches the performance of the Bartlett kernel used in Newey-West after only a few steps of the hierarchy. It is noteworthy, however, that, even for very large B, the $I_q[k]$ of the limiting implied Haar mean kernel never drops below $\frac{2}{3}$, the value of $I_q[k]$ for the Bartlett kernel.

4.3 Summary

The values of $I_q[k]$ for the estimators discussed in this paper are summarized in Table 3. One clear observation is that, among the estimators considered, the Bartlett kernel achieves the minimum value of $I_q[k]$. For q fixed, one can rank the quality of the estimators using $I_q[k]$ because the mean-squared error of estimator of the spectral density is increasing in $I_q[k]$ at the optimal rate for the sequence of truncation parameters, this suggests that the use of the Bartlett kernel should not be interpreted as an arbitrary choice. Instead, the numerical results indicate that Bartlett is asymptotically superior to many other natural and novel kernels encountered in practice. These first-order kernels include the Ibragimov-Müller Split-Sample estimator and the one implied by projections onto Legendre polynomials.

With that said, we caution the reader against interpreting the Bartlett kernel as being optimal in a global sense. The main result of the paper (i.e. Theorem 5) implies that Bartlett cannot be optimal in the sense of minimizing $I_q[k]$ because we can find an alternative kernel that dominates Bartlett by mixing small amounts of the Bartlett kernel with large amounts of a second-order, square integrable, continuous, positive-semidefinite kernel of choice. A potentially reasonable second-order kernel to mix is the Quadratic Spectral (QS) kernel because it achieves the minimum value of $I_q[k]$ over the class of second-order ($q = 2$) kernels [Epanechnikov, 1969]. Conditional on this choice, there is still the question of the optimal weights to place on the first-order and second-order parts of the mixture kernel unresolved. We believe that a full investigation of the practical implications of the theorem in Section 3 is an important avenue for future research. Nevertheless, a middle-ground takeaway is that that the Newey-West estimator of the long run variance matrix combined with the fixed-b bandwidth and critical values offers performs well against other estimators encountered in practice but researchers should be careful to interpret the estimator as being optimal in the sense of minimizing $I_1[k]$ as such a kernel does not exist.

A second observation is that the values of $I_q[k]$ for the Bartlett kernel and the limiting implied mean kernel for the Haar system are identical. Based on the discussion in Section 4.2.3, this result is unsurprising because the performance of the Haar system approaches that of the Bartlett kernel after only a few steps of the hierarchy.

Finally, a curious feature of Table 3 is that there are multiple, seemingly very different,

Kernel/Series	$I_q[k]$
Newey-West (Bartlett)	$\frac{2}{3}$
Bartlett-Like Kernels	
$(1-\lvert t\rvert)^p I_{[-1,1]}(t)$	$\frac{2p}{2p+1}$
$e^{-c\lvert t\rvert}$	1
Legendre Polynomials	
Optimal Limiting Implied Mean Kernel (Numerical)	< 0.734 (at $B = 50$)
Optimal Weighted Series	$\frac{8}{9}$
Equal Weighted Series	1
Haar Wavelets	
Limiting Implied Mean Kernel (Optimal/Equal Basis Weights)	$\frac{2}{3}$
Optimal Weighted Series	≈ 0.91
Equal Weighted Series	1
Split-Sample Step Function (Ibragimov and Müller)	1

Table 3: Summary of values of $I_q[k]$ for the first-order, positive-semidefinite kernel, and Weighted Orthogonal Series estimators discussed in this paper.

estimators that are asymptotically equivalent in the sense that $I_q[k] = 1$: the exponential kernel, equal-weighted projection onto Legendre polynomials, and the split-sample (batch mean) estimator (which is the equal-weighted Haar orthogonal series estimator when the number of basis functions in the split-sample estimator, $B = 2^n - 1$ for some $n \in N$). We do not have an interpretation for this coincidence.

5 Conclusion

In this work, we have explored the optimality of first-order, positive-semidefinite kernels in depth. We have shown that, in fact, the question in the title of this paper is ill-posed: there is no optimal first-order, positive-semidefinite kernel, in the sense that it minimizes $I_q[k]$, the quantity that LLS have shown determines the asymptotic size-power tradeoff for kernel and Weighted Orthogonal Series estimators. Indeed, we can produce first-order, positive-semidefinite kernels with arbitrarily small values of $I_q[k]$ by mixing small amounts of the Bartlett kernel, which is used in the Newey-West estimator, with large amounts of any second-order, square integrable, continuous, positive-semidefinite kernel of choice. Further, we have shown that neither the set of first-order, nor the set of second-order, kernels are closed as subsets of the space of all positive-semidefinite kernels. Thus, sequences of first-order kernels may have a limiting second-order kernel and vice-versa. We are therefore left with a delimited answer to the title's question: while optimality is ill-posed, Bartlett cannot be beaten by a large variety of first-order kernels including some novel ones. These results provide additional justification for the continuing use of the popular Newey-West estimator (with fixed-b critical values).

Given these impossibility results, we then explored a variety of families of first-order estimators. In every case, we found that the Bartlett kernel possessed a value of $I_q[k]\left(\frac{2}{3}\right)$ that is at least as low as that of any family we considered. Notably, limiting implied mean kernel estimators based on the Haar system were able to achieve this value of $I_q[k]$ asymptotically. This appears to be because the forms of the limiting implied mean kernels go to the Bartlett kernel asymptotically.

In light of these findings, the next steps appear to be the derivation of higher order Edgeworth expansions for first-order estimators and careful numerical study of the performance of various first-order estimators in finite sample settings. This includes such "artificial" estimators as the mixtures of the Bartlett kernel and QS kernels. Even though we have shown that every first-order, positive-semidefinite kernel that satisfies mild regularity conditions can be represented as an infinite family of nontrivial mixtures of first and scoond-order kernels, one might expect, based on higher order Edgeworth expensions, that artificial mixtures containing large amounts of second-order kernels with the Bartlett kernel will behave differently from the Bartlett kernel itself. Despite the apparent lack of 'pure' first-order kernels, perhaps differences in higher order terms may yield additional insights into how first-order kernels differ. In any case, given the surprisingly good performance of first-order, positive-semidefinite kernel estimators in finite samples, particularly Newey-West, it seems that further exploration of their finite sample behavior is in order.

References

D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, pages 817–858, 1991.

D. R. Brillinger. *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston, 1975.

V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969.

U. Grenander and M. Rosenblatt. Statistical spectral analysis of time series arising from stationary stochastic processes. *The Annals of Mathematical Statistics*, pages 537–558, 1953.

U. Grenander and M. Rosenblatt. *Statistical Analysis of Stationary Time Series*. New York: Wiley, 1957.

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

M. Jansson. The error in rejection probability of simple autocorrelation robust tests. *Econometrica*, 72(3):937–946, 2004.

N. M. Kiefer and T. J. Vogelsang. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6):1130–1164, 2005.

N. M. Kiefer, T. J. Vogelsang, and H. Bunzel. Simple robust testing of regression hypotheses. *Econometrica*, 68(3):695–714, 2000.

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

E. Lazarus, D. J. Lewis, and J. H. Stock. The size-power tradeoff in har inference. *Econometrica*, 89(5):2497–2516, 2021.

U. K. Müller. A theory of robust long-run variance estimation. *Journal of Econometrics*, 141 (2):1331–1352, 2007.

U. K. Müller. Hac corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, 32(3):311–322, 2014.

W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987. ISSN 00129682, 14680262.

W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653, 1994.

P. C. B. Phillips. Hac estimation by automated regression. *Econometric Theory*, 21(1):116–142, 2005.

P. C. B. Phillips, Y. Sun, and S. Jin. Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review*, 47(3):837–894, 2006.

P. C. B. Phillips, Y. Sun, and S. Jin. 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, 2007.

M. B. Priestley. *Spectral analysis and time series*. London: Academic Press, 1981.

Y. Sun. A heteroskedasticity and autocorrelation robust f test using an orthonormal series variance estimator. *The Econometrics Journal*, 16(1):1–26, 2013.

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

Y. Sun, P. C. B. Phillips, and S. Jin. Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, 76(1):175–194, 2008.

Y. Sun, P. C. B. Phillips, and S. Jin. Power maximization and size control in heteroskedasticity and autocorrelation robust tests with exponentiated kernels. *Econometric Theory*, 27(6): 1320–1368, 2011.

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