# IS NEWEY-WEST OPTIMAL AMONG $q = 1$ KERNELS?

THOMAS KOLOKOTRONES* AND JAMES STOCK†
*DEPARTMENTS OF BIOSTATISTICS AND EPIDEMIOLOGY AND †DEPARTMENT OF ECONOMICS, HARVARD UNIVERSITY

NOVEMBER 12, 2019

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 $q = 1$ kernel, meaning that 1 is the largest value of $q$ for which the quantity, $k^{(q)}(0) = \lim_{t\to 0} |t|^{-q}(1 - k(t))$, is finite. This raises the apparently uninvestigated question of whether the Bartlett kernel is optimal among $q = 1$ kernels. Here, we demonstrate that there is no optimal $q = 1$ kernel for HAR testing in the Gaussian location model or for minimizing the MSE of the spectral estimator. In fact, the space of q = 1 positive semidefinite kernels is not closed and, moreover, all $q = 1$ kernels (satisfying mild regularity conditions) can be decomposed into a weighted sum of $q = 1$ and $q = 2$ kernels, which suggests that there is no meaningful notion of "pure" $q = 1$ kernels. Nevertheless, it is possible to rank any given collection of $q = 1$ 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 $q = 1$ 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 autocorrelation 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 $q = 1$ kernel, where $q$ is the Parzen characteristic exponent, which is defined as the largest value of $q$ such that the quantity, $k^{(q)}(0) = \lim_{t\to 0} |t|^{-q}(1 - k(t))$ is finite. It is well known that, for the problem of spectral estimation, the mean squared error (MSE) of $q = 1$ kernels is asymptotically dominated by that of $q = 2$ kernels, which are approximately quadratic 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 $q = 2$ kernels has been shown if one focuses on a Type I/Type II error tradeoff (Sun, Phillips, and Jin (2008)) or on a size-power tradeoff (Lazarus, Lewis, and Stock (2017), LLS). However, calculations in Lazarus, Lewis, Stock, and Watson (2018) suggest that for sample sizes typically used in time series econometric applications, neither $q = 1$ nor $q = 2$ kernels dominate; typically, the size-power frontiers cross, so that, in some regions, the Bartlett kernel is preferred in finite samples to the optimal $q = 2$ kernel (the Quadratic Spectral (QS) kernel). Since even the optimal $q = 2$ kernel does not necessarily outperform $q = 1$ kernels, in particular, the Bartlett kernel, in finite samples, this raises the question of whether the Bartlett kernel is optimal among $q = 1$ kernels or, if not, what $q = 1$ kernel improves upon it.

It appears that the question of optimality among $q = 1$ kernels has received little attention, either in the classical spectral estimation literature or, more recently, in the HAR inference literature. Other $q = 1$ kernels, or tests that have $q = 1$ implied mean kernels (in the sense of LLS), include the split-sample (or "batch mean") estimator of the LRV (Ibragimov and Muller (2010)) and the LRV estimator obtained by projecting the product series $\hat{z}_t = x_t \hat{u}_t$ ($x_t$ being the regressors and $\hat{u}_t$ the regression residual) 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, $k^{(q)} = \lim_{t\to 0} |t|^{-q}(1-k(t))$, and $q$ is the largest value for which $k^{(q)}(0)$ is finite. Lazarus, Lewis, and Stock 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 Keifer-Vogelstein Fixed-b inference is used (where $b = T^{-1}S$). Thus, kernels with the same value of $q$ can be ranked, for both estimation and testing, by their values of $I_q[k]$, with smaller values preferred. Among $q = 2$ 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 $q = 1$ kernels appears to be unaddressed.

Here, we demonstrate that there is no optimal $q = 1$ kernel in the sense of minimizing $I_q[k] = \left(k^{(q)}(0)\right)^{\frac{1}{q}} \int k^2(t)dt$. We further show that the set of $q = 1$ psd kernels is not closed as a subset of the space of all psd kernels and that, indeed, all $q = 1$ kernels satisfying mild regularity conditions, which include the Bartlett kernel, can be decomposed into a weighted sum of $q = 1$ and $q = 2$ kernels, which suggests that there is no meaningful notion of "pure" $q = 1$ kernels.

We provide a restricted family of $q = 1$ 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 selected other $q = 1$ 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 $q = 1$ 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 $q = 1$ 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 (1957), 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), 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, Phillips and Jin (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, Vogelsang, and Bunzel (2000)) and by using Kiefer and Vogelsang's (2005) so-called fixed-b critical values to account for the increased variability of these estimators. Jansson (2004), Sun, Phillips and Jin (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), Mller (2007), and Sun (2013)). Muller (2014) and Lazarus, Lewis, Stock, and Watson (2018) provide additional surveys of this literature.

## 2. Kernels and Known Results

Kernel functions are widely used for long-run variance estimation in time series applications since, under certain additional assumptions on the kernel, namely that it is positive semidefinite (psd) and appropriately normalized, kernel variance estimators are guaranteed to have desirable properties such as nonnegativity. Letting $k$ be the kernel function, $T$ be the number of time points, $S$ be a scaling (truncation) parameter, $\Gamma$ be the autocovariance function, and $\Omega$ be the long-run variance, such estimators take the form

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

In variance estimation, kernels are required to be symmetric about 0 and are normalized to have $k(0) = 1$. They are typically characterized by their behavior near zero. If a kernel admits a series expansion of the form $k(t) = 1 + \sum_{i=1}^{\infty} c_i|t|^i$, the order of the kernel corresponds to the index of the first nonzero coefficient. More generally, a kernel's order is defined as the largest $q$ for which the quantity, $k^{(q)}(0) = \lim_{t\to 0}|t|^{-q}(1 - k(t))$, is finite. Note that, for any psd function, $|k(t)| \leq k(0)$, so $k^{(q)}(0) \geq 0$.

In addition to being symmetric with $k(0) = 1$, we will also require that kernels be positive semidefinite and continuous. The restriction that $k$ be psd means that, for any $n \in N$ and any $t \in R^n$, the matrix $K$, with entries $K_{ij} = k(t_i - t_j)$, is also psd. By Bochner's Theorem, the class of continuous psd functions with $k(0) = 1$, is exactly the set of (inverse) Fourier Transforms of probability measures (we say inverse due to the convention that we use for the Fourier Transform). This strong result, combined with the properties of the Fourier Transform itself, and the definition of $k^{(q)}(0)$, tells us that, under some additional regularity conditions, all continuous psd kernels must have an order, $q$, that is at most 2.

The assumption that psd kernels have nonnegative Fourier transforms is common in the literature, with Andrews (2001) requiring this of kernels in his class $\mathcal{K}_2$ and Priestly (1981) similarly assuming that the discrete Fourier Transform of his kernels be nonnegative. Thus, the assumption of continuity is not burdensome and, indeed, all widely used psd kernels are, in fact, continuous.

We now define some notation. Integrability will be with respect to Lebesgue measure on $R$, so that $f$ is integrable if it is a Lebesgue measurable function and $\int |f(x)|dx < \infty$, with $\mathcal{L}^p$ spaces defined in the standard way, so that $f \in \mathcal{L}^p$ means that $\int |f(x)|^p dx < \infty$. $\mathcal{F}$ will denote the Fourier Transform with $\hat{f}(\omega) = (\mathcal{F}f)(\omega) = (2\pi)^{-1} \int e^{-i\omega t} f(t)dt$, so that $\left(\mathcal{F}^{-1}\hat{f}\right)(t) = \int e^{i\omega t}\hat{f}(\omega)d\omega$. Additionally, we define the (inverse) Fourier Transform of a measure by $\left(\mathcal{F}^{-1}\mu\right)(t) = \int e^{i\omega t} d\mu$. Note that this convention is somewhat unusual, but is standard in the kernel literature because it ensures that the (inverse) Fourier transformation of a probability measure has $k(0) = 1$.

## 3. Non-Existence of an Optimal $q = 1$ kernel

For $q = 1$, $I_q[k] = k^{(1)}(0) \int k^2(t)dt$. Let $k_1(t) = (1 - |t|)\mathrm{I}_{[-1,1]}(t)$, the Bartlett kernel, which is used by the Newey-West variance estimator, and let $k_2(t)$ be any fixed $q = 2$, continuous psd kernel with $\int k_2(t)^2 dt = M < \infty$ (and $k_2(0) = 1$). Then $k_1^{(1)}(0) = \lim_{t\to 0} \frac{1-(1-|t|)}{|t|} = 1$, $k_2^{(1)}(0) = \lim_{t\to 0} \frac{1-k_2(t)}{|t|} = 0$. Now, let $k_\epsilon = \epsilon k_1 + (1-\epsilon)k_2, \epsilon \in (0, 1]$. Then,

$$\begin{aligned}
k_\epsilon^{(q)}(0) &= \lim_{t\to 0} \frac{1 - k_\epsilon(t)}{|t|^q} = \lim_{t\to 0} \frac{1 - (\epsilon k_1(t) + (1-\epsilon)k_2(t))}{|t|^q} \\
&= \epsilon \lim_{t\to 0} \frac{1 - k_1(t)}{|t|^q} + (1-\epsilon) \lim_{t\to 0} \frac{1 - k_2(t)}{|t|^q} \\
&= \epsilon k_1^{(q)}(0) + (1-\epsilon)k_2^{(q)}(0)
\end{aligned} \tag{2}$$

Thus, $k_\epsilon^{(1)}(0) = \epsilon k_1^{(1)}(0) + (1-\epsilon)k_2^{(1)}(0) = \epsilon \cdot 1 + (1-\epsilon) \cdot 0 = \epsilon$. Since $k_1^{(q)}(0) = \infty$ for $q > 1$ and $0 \le k_2^{(q)}(0)$, $k_\epsilon^{(q)}(0) = \epsilon k_1^{(q)}(0) + (1-\epsilon)k_2^{(q)}(0) = \infty$ for $q > 1$ as well, so $k_\epsilon$ is a $q = 1$ kernel. By Bochner's Theorem, the class of continuous positive semidefinite functions with $k(0) = 1$, is composed of exactly those functions that are the (inverse) Fourier transform of a probability measure. Letting $k_i = \mathcal{F}^{-1}(\mu_i)$, where $\mu_i$ is a probability measure, and using the linearity of the Fourier Transform, we have, $k_\epsilon = \epsilon k_1 + (1-\epsilon)k_2 = \epsilon\mathcal{F}^{-1}(\mu_1) + (1-\epsilon)\mathcal{F}^{-1}(\mu_2) = \mathcal{F}^{-1}(\epsilon\mu_1 + (1-\epsilon)\mu_2)$. So, since the weighted average of two probability measures is again a probability measure, $k_\epsilon$ is also the (inverse) Fourier Transform of a probability measure, and, thus, a valid continuous psd kernel. Therefore, we can produce valid $q = 1$, continuous positive semidefinite kernels with arbitrarily small values of $k_1^{(1)}(0)$. Then,

$$I_q[k_\epsilon] = \left(k_\epsilon^{(q)}(0)\right)^{\frac{1}{q}} \int k_\epsilon^2(t)dt = \epsilon \cdot \int \epsilon^2 k_1(t)^2 + 2\epsilon(1-\epsilon)k_1(t)k_2(t) + (1-\epsilon)^2 k_2(t)^2 dt \tag{3}$$

So, for small values of $\epsilon$, $I_q[k_\epsilon] \approx \epsilon \int k_2(t)^2 dt$ or, more formally, using Holder's Inequality and that $0 < \epsilon \leq 1$, $\epsilon(1-\epsilon) \leq \frac{1}{4}$, $\int k_2(t)^2 dt = M < \infty$, and $\int_{-1}^{1} (1-|t|)^2 \, dt = \frac{2}{3}$,

$$I_q[k_\epsilon] \leq \epsilon \cdot \left(\frac{2}{3} + \frac{1}{2}\left(\frac{2M}{3}\right)^{\frac{1}{2}} + M\right) \tag{4}$$

Therefore, we can make $I_q[k]$ arbitrarily small by mixing sufficiently small amounts of the Bartlett kernel with any $q = 2$, square integrable, continuous psd kernel. Specifically, since $I_q[k_1] = \frac{2}{3}$ and $I_q[k_\epsilon]$ is continuous in $\epsilon$, we can choose $\epsilon$ so that $I_q[k_\epsilon]$ takes any value in $(0, \frac{2}{3}]$. Therefore, there can be no optimal $q = 1$ kernel using the same optimality criterion used for $q = 2$ kernels (ensuring that it provides the optimal asymptotic size-power tradeoff). The reason that this argument does not apply to $q = 2$, continuous psd kernels is because, as discussed above, any continuous psd kernel must have $q \leq 2$. Thus, there are no continuous psd kernels with $q > 2$ with which to mix.

This also gives us a topological interpretation of why there can be no optimal $q = 1$ kernel: the set of $q = 1$ psd kernels is not a closed subset of the set of all psd kernels, so sequences of $q = 1$ kernels may have limits that are instead $q = 2$ (consider $k_\epsilon$ as $\epsilon$ goes to 0). It is also true that there exist sequences of $q = 2$ kernels which have $q = 1$ kernels as their limits. A simple example is the family of kernels $k_b = \mathcal{F}^{-1}\left(c_b \omega^{-2}(1-\cos(\omega)) I_{[-b,b]}(\omega)\right)$ with $c_b = \left(\int_{-b}^{b} \omega^{-2}(1-\cos(\omega))d\omega\right)^{-1}$. For any finite $b > 0$, this yields a $q = 2$ kernel. However, since $\mathcal{F}\left((1-|t|)\, I_{[-1,1]}(t)\right)(\omega) = \pi^{-1}\omega^{-2}\left(1-\cos(\omega)\right)$, as $b$ goes to infinity, the limiting kernel is the Bartlett kernel, which is $q = 1$. Taken together, these results tell us that neither the set of $q = 1$, nor the set of $q = 2$, psd kernels are closed subsets of the set of all psd kernels. This has another interesting implication: since the linear subspaces of a Banach space are closed, there do not exist (Banach) bases for the set of all scalar multiples of the $q = 1$ psd kernels or the set of all scalar multiples of the $q = 2$ psd kernels (even if we restrict to only the continuous psd kernels). This makes it more difficult to study their structures separately.

We summarize the above results in the following Theorem.

**Theorem 1.** *For any $I \in (0, \frac{2}{3}]$ there exists a $q = 1$ psd kernel, $k$, with $I_q[k] = k^{(1)}(0) \int k^2(t)dt = I$. Neither the set of $q = 1$, nor the set of $q = 2$, psd kernels, is a closed subset of the space of all psd kernels. Thus, neither the set of all scalar multiples of the $q = 1$ psd kernels, nor the set of all scalar multiples of the $q = 2$ psd kernels, possess (Banach) bases.*

*Proof.* As above. □

## 4. Structure of $q = 1$ PSD Functions

### 4.1. General Decomposition of $q = 1$ Kernels.

We now use Fourier Analysis in order to more fully characterize the structure of $q = 1$ psd kernels. We begin with several simple lemmas. Recall that $\mathcal{F}$ denotes the Fourier Transform with $\hat{f}(\omega) = (\mathcal{F}f)(\omega) = (2\pi)^{-1} \int e^{-i\omega t} f(t)dt$, so that $\left(\mathcal{F}^{-1}\hat{f}\right)(t) = \int e^{i\omega t}\hat{f}(\omega)d\omega$. Several basic results from Fourier analysis are widely used in the classical spectral estimation literature (e.g. Priestley, 1981), including the fact that, if $f, f' \in \mathcal{L}^1$, then $\mathcal{F}(f')(\omega) = i\omega\hat{f}(\omega)$, which can be easily shown using integration by parts and can be iterated to give $\mathcal{F}\left(f^{(n)}\right)(\omega) = (i\omega)^n\hat{f}(\omega)$. We will make use of a similar result.

**Lemma 2.** *Let $K(\omega)$ and $\omega K(\omega)$ be integrable functions, then, if $k = \mathcal{F}^{-1}K$, $k$ is continuously differentiable and $k' = i\mathcal{F}^{-1}(\omega K(\omega))$*

*Proof.*

$$\begin{aligned}
\frac{k(t+h) - k(t)}{h} &= \frac{\left(\mathcal{F}^{-1}K\right)(t+h) - \left(\mathcal{F}^{-1}K\right)(t)}{h} \\
&= h^{-1}\left[\int e^{i\omega(t+h)}K(\omega)d\omega - \int e^{i\omega t}K(\omega)d\omega\right] \\
&= h^{-1}\int e^{i\omega t}\left(e^{i\omega h} - 1\right)K(\omega)d\omega
\end{aligned}$$

$$\begin{aligned}
\left|h^{-1}e^{i\omega t}\left(e^{i\omega h}-1\right)\right| &= |h|^{-1}\left|e^{i\omega t}\right|\left|e^{i\omega h}-1\right| \\
&= |h|^{-1}\cdot 1\cdot\left(1-e^{i\omega h}-e^{-i\omega h}+1\right)^{\frac{1}{2}} \\
&= |h|^{-1}\left(2-2\cos(\omega h)\right)^{\frac{1}{2}} \\
&= |h|^{-1}\left(2-2\left(1+0-\frac{1}{2}\cos(\omega h^*)(\omega h)^2\right)\right)^{\frac{1}{2}} \\
&\leq |h|^{-1}\left(1\cdot(\omega h)^2\right)^{\frac{1}{2}} = |h|^{-1}|\omega h| \\
&= |\omega|
\end{aligned}$$

where the fourth equality is due to Taylor's Theorem, with $h^* \in [0,h]$, and the inequality is due to the fact that $|\cos(x)| \leq 1$. Thus, $|h^{-1}e^{i\omega t}\left(e^{i\omega h}-1\right)K(\omega)| \leq |\omega K(\omega)|$. Since $k'(t) = \lim_{h\to 0} h^{-1}\left(k(t+h)-k(t)\right)$, $\lim_{h\to 0} h^{-1}e^{i\omega t}\left(e^{i\omega h}-1\right) = \frac{d}{dh}e^{i\omega(t+h)}|_{h=0} = i\omega e^{i\omega t}$, and, by assumption, $\omega K(\omega)$ is integrable, we can apply the Dominated Convergence Theorem to get

$$k'(t) = (2\pi)^{-1}\int i\omega e^{i\omega t}K(\omega)d\omega = i\mathcal{F}^{-1}\left(\omega K(\omega)\right)(t)$$

Since $\omega K(\omega)$ is integrable, $\mathcal{F}^{-1}(\omega K(\omega))$ is uniformly continuous, so $k$ is continuously differentiable.

□

**Corollary 3.** *Let $K(\omega)$ and $\omega^2 K(\omega)$ be integrable functions, $K \geq 0$, and $\int K(\omega)d\omega = 1$. Then $k = \mathcal{F}^{-1}K$ is a $q = 2$, continuous psd kernel.*

*Proof.* Since $|\omega K(\omega)| \leq |K(\omega)|$ for $|\omega| \leq 1$, $|\omega K(\omega)| \leq |\omega^2 K(\omega)|$ for $|\omega| \geq 1$, and $K(\omega), \omega^2 K(\omega) \in \mathcal{L}^1$, $\omega K(\omega) \in \mathcal{L}^1$ as well. Then, by applying the lemma twice, we get $k = \mathcal{F}^{-1}K \in \mathcal{C}^2$ and $k''(t) = -\mathcal{F}^{-1}\left(\omega^2 K(\omega)\right)$. Since $K$ is a probability distribution, by Bochner's Theorem, $k$ is a continuous psd kernel. Also, note that, since $\int K(\omega)d\omega = 1$, $K$ is positive on a set of positive measure and $k(0) = 1$. Since, $k \in \mathcal{C}^2$, by Taylor's Theorem, $k(t) = k(0) + k'(0)t + 2^{-1}k''(t^*)t^2$ for some $t^* \in [0,t]$. Since $k' \in \mathcal{C}^1$ and $k$ is symmetric, then $k'(0) = 0$. Thus, $k(t) = 1 + 2^{-1}k''(t^*)t^2$ so $k^{(2)}(0) = \lim_{t\to 0} t^{-2}(1-k(t)) = -2^{-1}\lim_{t^*\to 0} f''(t^*) = -2^{-1}k''(0) = -2^{-1}\cdot -\int \omega^2 K(\omega)d\omega = 2^{-1}\int \omega^2 K(\omega)d\omega$. Since $0 < 2^{-1}\int \omega^2 K(\omega)d\omega < \infty$, k is a $q = 2$, continuous psd kernel. □

**Corollary 4.** *Let $K \geq 0$, $\int K(\omega)d\omega = 1$, and $K$ have compact support, then $k = \mathcal{F}^{-1}K$ is a $q = 2$, continuous psd kernel.*

*Proof.* Let $M = \sup \omega$ over the support of $K$. Then, $\int \omega^n K(\omega)d\omega \leq M^n \int K(\omega)d\omega = M^n < \infty$, so the conditions of the above lemma are satisfied and $k$ is a $q = 2$, continuous psd kernel. □

We can now prove the following, somewhat surprising, result.

**Theorem 5.** *If $k$ is a $q = 1$, continuous psd kernel with $k^{(1)}(0) > 0$, such that there exists an integrable function $K$ such that $k = \mathcal{F}^{-1}K$, then $k$ can be decomposed as the (nontrivial) weighted sum of $q = 1$ and $q = 2$, uniformly continuous psd kernels. Further, there are infinitely many such decompositions.*

*Proof.* If such a $K$ exists, then, since it is integrable and $1 = k(0) = \int K(\omega)d\omega$, it is a probability distribution (as we would expect from Bochner's Theorem, the subtlety is that Bochner's theorem only guarantees the existence of a probability measure, not a density, so the existence assumption is not vacuous). Let $b$ be sufficiently large that $\int_{-b}^{b} K(\omega)d\omega > 0$ and define $K_{1,b}(\omega) = \left(2\int_b^\infty K(\omega)d\omega\right)^{-1} K(\omega)I[|x| > b]$, $K_{2,b}(\omega) = \left(\int_{-b}^{b} K(\omega)d\omega\right)^{-1} K(\omega)I[|\omega| \leq b]$. Then $K_{i,b} \geq 0$ and $\int K_{i,b}(\omega)d\omega = 1$, so, by Bochner's Theorem, $k_{i,b} = \mathcal{F}^{-1}K_{i,b}$ are continuous psd kernels (in fact, they are uniformly continuous since $K_{i,b} \in \mathcal{L}^1$, since $K \in \mathcal{L}^1$, by assumption). Further, by Corollary 4, $k_{2,b}$ is $q = 2$. Since $1 = \int K(\omega)d\omega$, $K(\omega) = \left(2\int_b^\infty K(\omega)d\omega\right) K_{1,b}(\omega) + \left(\int_{-b}^{b} K(\omega)d\omega\right) K_{2,b}(\omega)$, and $k(t) = \left(2\int_b^\infty K(\omega)d\omega\right) k_{1,b}(t) + \left(\int_{-b}^{b} K(\omega)d\omega\right) k_{2,b}(t)$, we have, $k_{1,b}^{(1)}(0) = \left(2\int_b^\infty K(\omega)d\omega\right)^{-1} k^{(1)}(0) > 0$, so $k_{1,b}$ is $q = 1$. Thus, as claimed, $k$ can be decomposed into the (nontrivial) sum of $q = 1$ and $q = 2$, uniformly continuous psd kernels. Since $b$ is arbitrary, there are infinitely many such decompositions. □

We can now prove a useful corollary which follows from a form of the Fourier Inversion Theorem typically encountered in abstract harmonic analysis. Specialized to our setting, it says that, if $k$ is continuous, positive semidefinite, and normalized so that $k(0) = 1$, and $k \in \mathcal{L}^1$, then, $\hat{k} \in \mathcal{L}^1$ as well, and the probability measure $\mu_k$, of which $k$ is the inverse Fourier Transform, has density $K = \hat{k}$, so that, $\mu_k(\omega) = \hat{k}(\omega)d\omega$. Thus, we can use the Fourier Transform to recover $K = \hat{k}$, which will be nonnegative and integrate to 1.

**Corollary 6.** *If $k$ is a $q = 1$, continuous psd kernel with $k^{(1)}(0) > 0$, and $k \in \mathcal{L}^1$, then it can be decomposed as the (nontrivial) weighted sum of $q = 1$ and $q = 2$, continuous psd kernels. Further, there are infinitely many such decompositions.*

*Proof.* If $k \in \mathcal{L}^1$, then, using the version of the Fourier Inversion Theorem quoted above, $K = \hat{k}$ is a probability density and is, thus, integrable. The result then follows directly from the theorem. Since any $k$ with compact support is integrable, the result holds in that important special case. $\square$

The above Theorem tells us that any $q = 1$ psd kernel that satisfies very mild regularity conditions, can be (nontrivially) decomposed into another $q = 1$ psd kernel and a $q = 2$ psd kernel, so there is no natural notion of a minimal or irreducible $q = 1$ kernel. This means that apparently artificial kernels, such as the one we exhibited to show that no optimal $q = 1$ kernel exists, are, in some sense, no more unnatural than the Bartlett kernel, which appears at first glance to be a "pure" $q = 1$ kernel in a way that such a composite kernel does not. Also of note is the fact that it is the tails of $K = \hat{k}$ that determine the order of the kernel. The central portion has no effect.

### 4.2. **Kernel Decompositions.**

As a demonstration of this decomposition, we will separate the Bartlett kernel into a $q = 1$ piece, which comes from the tails of $K = \hat{k}$, and a $q = 2$ piece, which is derived from the center of $K$. We begin by computing the Fourier transform of $k(t) = (1-|t|)I_{[-1,1]}(t)$, $K(\omega) = \frac{1-\cos(\omega)}{\pi\omega^2}$. We then split $K$ at some $b \geq 0$ into two functions $K_{1,b}(\omega) = \frac{1-\cos(\omega)}{\pi\omega^2} I[|\omega| > b]$ and $K_{2,b}(\omega) = \frac{1-\cos(\omega)}{\pi\omega^2} I[|\omega| \leq b]$ and apply the inverse Fourier Transform to obtain $k_{1,b}, k_{2,b}$. Note that these kernels are not properly normalized, so $k_i(0) \neq 1$, but they instead satisfy $k = k_{1,b} + k_{2,b}$, which makes it easier to see how each kernel contributes to the whole. As seen in Figure 1, the Tail Kernel, $k_{1,b}$, is a scaled $q = 1$ kernel and contributes the sharp point at 0, while the Central Kernel, $k_{2,b}$, is a smooth, scaled $q = 2$ kernel. Also of note, as $b$ increases, the relative contribution of $k_{1,b}$ decreases; additionally, $k_{2,b}$ becomes narrower and more pointed at its central peak, although, even for arbitrarily large $b$, $k_{2,b}$ will still remain $q = 2$ with a rounded, rather than sharp, peak.

## 5. Other $q = 1$ Kernels

A natural starting point for further exploration of $q = 1$ kernels is other Bartlett-like kernels. Two straightforward generalizations of the Bartlett kernel are the families $k_p(t) = \left(1 - \left|\frac{t}{b}\right|\right)^p \mathrm{I}_{[-1,1]}\left(\frac{t}{b}\right)$

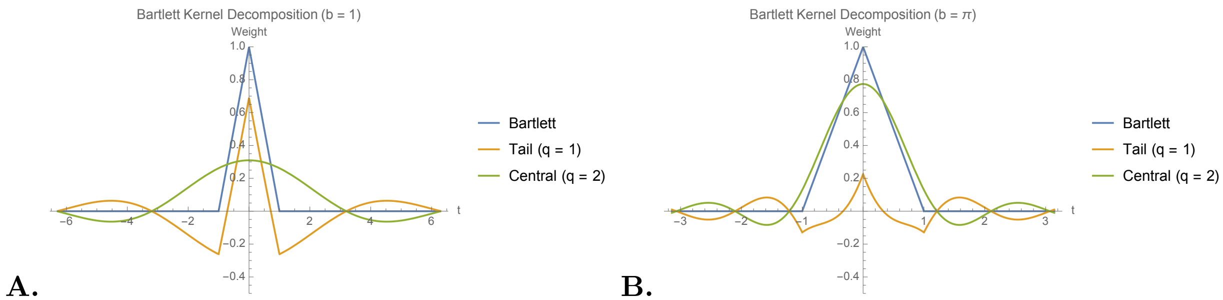


FIGURE 1. Decompostion of the Bartlett Kernel into (nonunique) $q = 1$ and $q = 2$ components. **A.** $b = 1$. **B.** $b = \pi$.

and $k_r(t) = \left(1 - \left|\frac{t}{b}\right|^r\right) \mathrm{I}_{[-1,1]}\left(\frac{t}{b}\right)$. 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 psd 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 $q = 1$ psd 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 q = 1 psd functions. An alternative approach has been to instead consider orthogonal series estimators, which we discuss next.

## 6. Orthogonal Series Estimators

### 6.1. **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 for $\mathcal{L}^2[0,1]$, $\{\phi_i\}_{i=0}^{\infty}$ such estimators first project the sequence $\{\hat{z}_t\}_{t=1}^{T}$ onto one of the basis functions (excepting $\phi_0 = 1$) and then compute an empirical variance $\widehat{\Omega}_i$ from the projection. These estimates are then combined via a weighted sum to give the final estimate, $\widehat{\Omega}$.

$$\widehat{\Omega} = \sum_{i=1}^{B} w_i \widehat{\Omega}_i, \quad \sum_{i=1}^{B} w_i = 1, w_i \geq 0 \quad \widehat{\Omega}_i = \widehat{\Lambda}_i \widehat{\Lambda}_i', \quad \widehat{\Lambda} = T^{-\frac{1}{2}} \sum_{t=1}^{T} \phi_i \left( \frac{t}{T} \right) \hat{z}_t \tag{5}$$

By construction, each $\widehat{\Omega}_i$, and, thus, $\widehat{\Omega}$, are psd w.p. 1.

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

$$k_w(t) = \sum_{i=1}^{B} w_i \tilde{k}_i(B^{-1}t), \quad \tilde{k}_i(t) = \int_{\max(0,t)}^{\min(1,1+t)} \phi_i(u)\phi_i(u-t)du \tag{6}$$

(The nonlimiting implied mean kernel is a sum over discrete time points; see Lazarus, Lewis, and Stock (2017) for more details.)

For finite $T$, this provides a connection between orthogonal series estimators and kernel estimators. Specifically, Lazarus, Lewis, and Stock 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 $I_q[k] = (k^{(q)}(0))^{\frac{1}{q}} \sum_i w_i^2$, where the $w_i$s are the weights. Note that, compared to the case of kernel estimators, $\sum_i w_i^2$ has replaced $\int k^2(x)dx$. This results in important differences in performance between the two classes of estimators. For any sequence $k_i^{(q)}(0)$, we can minimize this expression subject to the constraint $\sum_i w_i = 1$, in order to determine the optimal weights. For $q = 1$, we obtain a relatively simple closed form expression:

$$w_i = (B - b_0 + 1)^{-1} - \left( k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) A \left( k^{(1)}(0) \right) \tag{7}$$

where $b_0$ is the index of the first term used (typically 1) and $A$ 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 appendix.

## 6.2. **Legendre Polynomials.**

We begin by considering the well known Legendre Polynomials, which form an orthogonal series on $[-1, 1]$. After remapping their domains to $[0, 1]$ and normalizing by multiplying by

$(2i+1)^{\frac{1}{2}}$ so that $\int_0^1 \phi_i(x)\phi_j(x)dx = \delta_{ij}$, 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, Lazarus, Lewis, and Stock 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. We show the following,

**Proposition 7.** *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* $\phi$ *generates a* $q = 2$ *kernel, then,* $k \in \mathcal{C}^2[-1,1]$ *and* $k^{(2)}(0) = -\frac{1}{2}\frac{d^2 k}{dt^2}(0) = -\frac{1}{2}\int_0^1 \phi(u)\phi^{(2)}(u)du$.

*Proof.* See appendix. □

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 7, gives $\lim_{B\to\infty} B^{-1}k^{(1)}(0)$ $= \lim_{B\to\infty} B^{-1}\sum_i w_i k_i^{(1)}(0) = \frac{2}{3}$ and $\lim_{B\to\infty} B\sum_i w_i^2 = \frac{4}{3}$, so $I_q[k] = (k^{(q)}(0))^{\frac{1}{q}}\sum_i w_i^2 = \frac{8}{9}$ as $B \to \infty$ (see appendix for details). If we had, instead, used equal weights, we would get $\lim_{B\to\infty} B^{-1}k^{(1)}(0) = B^{-2}\sum_i 2i+1 = \lim_{B\to\infty} B^{-2}\left((B+1)^2 - b^2\right) = 1$ and $B\sum_i w_i^2 = B\sum_i B^{-2} = 1$ so $I_q[k] = 1$ 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 $k_4(t) = \left(1 - \left|\frac{t}{b}\right|\right)^4 \mathrm{I}_{[-1,1]}\left(\frac{t}{b}\right)$. Figure 2 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 $I_q[k] = (k^{(q)}(0))^{\frac{1}{q}}\sum_i w_i^2$ for Weighted Orthogonal Series, but $I_q[k] = \left(k^{(q)}(0)\right)^{\frac{1}{2}}\int k^2(t)dt$ for psd 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 2 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

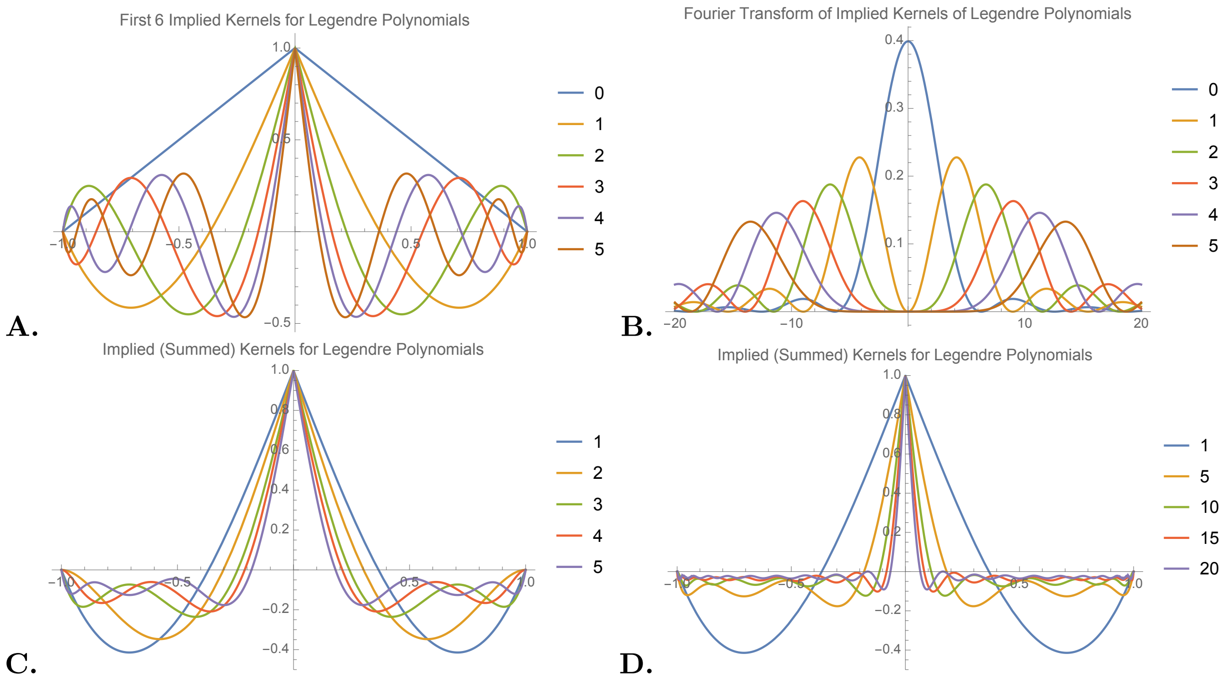


FIGURE 2. **A.** Limiting implied mean kernels for each of the first 6 Legendre polynomials (including $\phi_0 = 1$). **B.** Fourier Transforms of each of the first 6 Legendre polynomials (including $\phi_0 = 1$). **C.** Optimally weighted sums using the first (1, 2, 3, 4, 5) Legendre polynomials. **D.**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

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 from 1957.

### 6.3. **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). Recall that 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

| | $I_q[k]$ | |
|---|---|---|
| $B$ | 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 (except $\phi_0 = 1$) using either equal or (numerically estimated) optimal weights.

$\psi_{n,\ell}(x) = 2^{\frac{n}{2}}\psi(2^n x - \ell)$ 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 3). Note that $k_n(t)$ is independent of $\ell$, since, within each level of the hierarchy, the basis functions are simply translations of each other. Thus, it is clear that $k_n$ is a $q = 1$ kernel with $k_n^{(1)}(0) = \lim_{t\to 0}|t|^{-1}(1 - (1 - 3 \cdot 2^n|t|)) = 3 \cdot 2^n$. Let $n' > n$, then we also have, $\int k_n^2(t)dt = \frac{1}{3}2^{-n}$, $\int k_{n'}(t)k_n(t)dx = 2^{n-2n'-1}$ (see appendix for derivations).

Using these expressions, for any sequence of weights $\{w_i\}_{i=1}^N$, we can construct the limiting implied mean kernel $k_w(t) = \sum_{i=1}^N w_i k_i(t)$, which will also be a $q = 1$ kernel, since it is the sum of a finite number of $q = 1$ kernels. We can use $k_w(t)$ to construct a kernel estimator, which we can then optimize with respect to the weights. Somewhat surprisingly, weighting each basis function equally, so that $w_n = \frac{2^n}{2^{N+1}-1}$, 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, $I_q[k_e] = \frac{2}{3}\left(1 + 2^{-(N+1)}\right)$ (see 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 $\sum_i w_i^2 = \sum_i \left(2^{N+1}-1\right)^{-2} = \left(2^{N+1}-1\right)^{-1}$, 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 goes to 1 asymptotically. This finding agrees with the fact that $I_q[k]=1$ asymptotically for Ibragimov and Muller'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 series estimator, we obtain a somewhat lower limiting value of $I_q[k] \approx .91$, which is a modest improvement vs. 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.

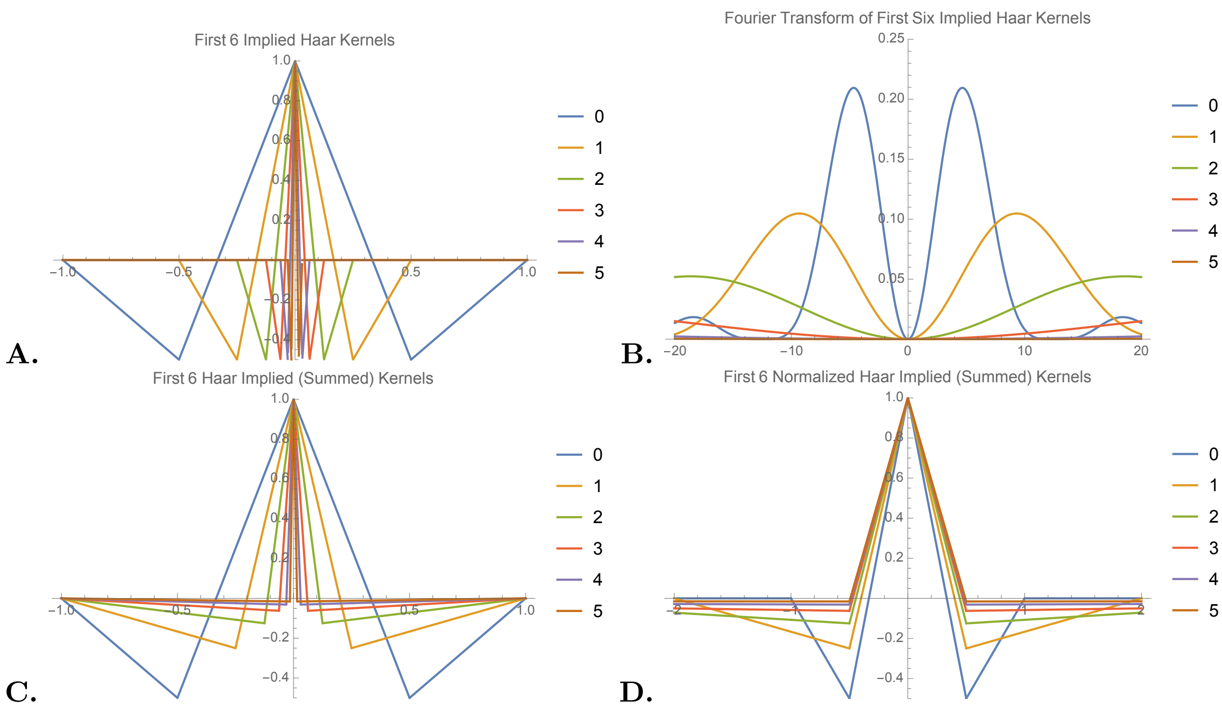


FIGURE 3. **A.** Limiting implied mean kernels for each of the first 6 levels of the Haar system. **B.** Fourier Transforms of each of the first 6 levels of the Haar system. **C.** Optimally weighted sums using the first 6 levels of the Haar system. **D.** Optimally weighted sums using the first 6 levels of the Haar system, normalized to allow for comparison of shape.

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 $I_q[k]$

| | $I_q[k]$ | |
|---|---|---|
| $B$ | 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 below by $\frac{2}{3}$.

of the limiting implied Haar mean kernel never drops below $\frac{2}{3}$, the value of $I_q[k]$ for the Bartlett kernel.

## 7. Summary and Conclusion

We summarize the values of $I_q[k]$ for the estimators discussed above in Table 3.

A curious feature of Table 3 is that there are multiple, seemingly very different, 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.

In this work, we have explored the optimality of $q = 1$ positive semidefinite kernels in depth. We have shown that, in fact, there is no optimal $q = 1$ psd kernel, in the sense that it minimizes $I_q[k]$, the quantity that Lazarus, Lewis, and Stock have shown determines the asymptotic size-power tradeoff for kernel and Weighted Orthogonal Series estimators. Indeed, we can produce $q = 1$ psd 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 $q = 2$, square integrable, continuous psd kernel of choice. Further, we have shown that neither the set of $q = 1$, nor the set of $q = 2$, kernels are closed as subsets of the space of all psd kernels. Thus, sequences of $q = 1$ kernels may have a limiting $q = 2$ kernel and vice-versa. Next, we demonstrated that, under mild regularity

$q = 1$ Kernels

| Kernel/Series | $I_q[k]$ |
|---|---|
| Newey-West (Bartlett) | $\frac{2}{3}$ |
| Bartlett-Like Kernels | |
| $(1-\lvert x\rvert)^p \, \mathrm{I}_{[-1,1]}(x)$ | $\frac{2p}{2p+1}$ |
| $e^{-c\lvert x\rvert}$ | 1 |
| Legendre Polynomials | |
| Optimal Limiting Implied Mean Kernel (Numerical) | $< .734$ (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 | $\approx .91$ |
| Equal Weighted Series | 1 |
| Split-Sample Step Function (Ibragimov and Muller) | 1 |

Table 3. Summary of values of $I_q[k]$ for the $q = 1$ psd kernel and Weighted Orthogonal Series estimators discussed in this paper.

conditions, there is no class of "pure" or irreducible $q = 1$ psd kernels, in the sense that they cannot be expressed as nontrivial mixtures of a $q = 1$ kernel and a kernel of higher order. Indeed, we showed that there are infinitely many such decompositions. These families of decompositions demonstrate that it is the tails of a kernel's Fourier Transform that control the order of the kernel and that, in particular, the tails of $q = 1$ kernels must not fall off too fast.

Given these impossibility results, we then explored a variety of families of $q = 1$ estimators. In every case, we found that the Bartlett kernel possessed a value of $I_q[k]$ $(\frac{2}{3})$ 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 $q = 1$ estimators and careful numerical study of the performance of various $q = 1$ estimators in finite sample settings. This includes such "artificial" estimators as the mixtures of the Bartlett kernel with $q = 2$ psd kernels that we used to show the nonexistence of an optimal $q = 1$ psd kernel. Even though we have shown that every $q = 1$ psd kernel that satisfies mild regularity conditions can be represented as an infinite family of nontrivial mixtures of $q = 1$ and $q = 2$ kernels, one might expect, based on higher order Edgeworth expansions, that artificial mixtures containing large amounts of $q = 2$ kernels with the Bartlett kernel will behave differently from the Bartlett kernel itself. Despite the apparent lack of "pure" $q = 1$ kernels, perhaps differences in higher order terms may yield additional insights into how $q = 1$ kernels differ. In any case, given the surprisingly good performance of $q = 1$, positive semidefinite kernel estimators in finite samples, particularly Newey-West, it seems that further exploration of their finite sample behavior is in order.

## References

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

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

[3] Kiefer, N., T.J. Vogelsang, and H. Bunzel (2000). Simple Robust Testing of Regression Hypotheses, Econometrica, 69, 695-714.

[4] Kiefer, N.M. and T.J. Vogelsang (2005). A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests, Econometric Theory, 21, 2093-2095.

[5] Lazarus, E., D.J. Lewis, J.H. Stock, and M.W. Watson (2018). HAR Inference: Recommendations for Practice, Journal of Business and Economic Statistics 36, 541-559.

[6] Muller, U. (2007). A Theory of Robust Long-Run Variance Estimation, Journal of Econometrics, 141, 1331-1352.

[7] Muller, U. (2014), HAC Corrections for Strongly Autocorrelated Time Series, Journal of Business and Economic Statistics 32, 311-322.

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

[9] Newey, W.K. and K.D. West (1994). Automatic Lag Selection in Covariance Matrix Estimation, Review of Economic Studies, 61, 631-653.

[10] Phillips, P. C. B. (2005). HAC Estimation by Automated Regression, Econometric Theory 21, 116142.

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

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

[13] Sun, Y. (2013). Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator, The Econometrics Journal, 16, 126.

[14] Sun, Y. (2014). Lets Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference, Journal of Econometrics 178, 659-677.

# 8. Appendix

## 8.1. Proof of Proposition 7.

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 $\phi$ generates a $q = 2$ kernel, then, $k \in \mathcal{C}^2[-1,1]$ and $k^{(2)}(0) = -\frac{1}{2}\frac{d^2 k}{dt^2}(0) = -\frac{1}{2}\int_0^1 \phi(u)\phi^{(2)}(u)du$.

*Proof.* Let $\frac{d}{dt^\pm}$ denote the right and left sided derivatives, respectively, so that $\frac{df}{dt^\pm}(t) = \lim_{h\to 0^\pm} \frac{f(t+h)-f(t)}{h}$ then,

$$\begin{aligned}
\lim_{t\to 0^+} |t|^{-1}(1-k(t)) &= \lim_{t\to 0^+} |t|^{-1}(k(0)-k(t)) = -\lim_{t\to 0^+} t^{-1}(k(t)-k(0)) = -\frac{dk}{dt^+}(0) \\
&= -\frac{d}{dt^+}\left[\int_{\max(0,t)}^{\min(1,1+t)} \phi(u)\phi(u-t)du\right]_{t=0} \\
&= -\frac{d}{dt^+}\left[\int_t^1 \phi(u)\phi(u-t)du\right]_{t=0} \\
&= -\left[-\phi(t)\phi(0) + \int_t^1 -\phi(u)\phi'(u-t)du\right]_{t=0} \\
&= \phi(0)^2 + \int_0^1 \phi(u)\phi'(u)du = \phi(0)^2 + \frac{1}{2}\left.\phi(u)^2\right|_0^1 \\
&= \phi(0)^2 + \frac{1}{2}\left(\phi(1)^2 - \phi(0)^2\right) \\
&= \frac{1}{2}\left(\phi(0)^2 + \phi(1)^2\right)
\end{aligned}$$

$$\begin{aligned}
\lim_{t\to 0^-} |t|^{-1}(1-k(t)) &= \lim_{t\to 0^-} |t|^{-1}(k(0)-k(t)) = \lim_{t\to 0^-} (-t)^{-1}(k(0)-k(t)) \\
&= \lim_{t\to 0^-} t^{-1}(k(t)-k(0)) = \frac{dk}{dt^-}(0) \\
&= \frac{d}{dt^-}\left[\int_{\max(0,t)}^{\min(1,1+t)} \phi(u)\phi(u-t)du\right]_{t=0} \\
&= \frac{d}{dt^-}\left[\int_0^{1+t} \phi(u)\phi(u-t)du\right]_{t=0} \\
&= \left[\phi(1+t)\phi(1) + \int_0^{1+t} -\phi(u)\phi'(u-t)du\right]_{t=0} \\
&= \phi(1)^2 - \int_0^1 \phi(u)\phi'(u)du = \phi(1)^2 - \frac{1}{2}\left.\phi(u)^2\right|_0^1 \\
&= \phi(1)^2 - \frac{1}{2}\left(\phi(1)^2 - \phi(0)^2\right) \\
&= \frac{1}{2}\left(\phi(0)^2 + \phi(1)^2\right)
\end{aligned}$$

So the limit exists and $k^{(1)}(0) = \lim_{t\to 0} |t|^{-1}(1-k(t)) = \frac{1}{2}\left(\phi(0)^2 + \phi(1)^2\right)$.

The proof of the second part of the theorem is by induction. By hypothesis $\phi \in \mathcal{C}^n[0,1]$. First assume that $t > 0$ and consider the hypothesis that $\frac{d^n k}{dt^n}(t) = \sum_{i=1}^n (-1)^i \phi^{(n-i)}(t)\phi^{(i-1)}(0) + (-1)^n \int_t^1 \phi(u)\phi^{(n)}(u-t)du$. From the above, this clearly holds for $n = 1$. Now assume that it holds for all $n' < n$, then,

$$
\begin{aligned}
\frac{d^n k}{dt^n}(t) = \frac{d}{dt}\frac{d^{n-1}k}{dt^{n-1}} &= \frac{d}{dt}\left[\sum_{i=1}^{n-1}(-1)^i\phi^{(n-1-i)}(t)\phi^{(i-1)}(0) + (-1)^{n-1}\int_t^1 \phi(u)\phi^{(n-1)}(u-t)du\right] \\
&= \sum_{i=1}^{n-1}(-1)^i\phi^{(n-i)}(t)\phi^{(i-1)}(0) + (-1)^{n-1}\left[\frac{d}{dt}\int_t^1 \phi(u)\phi^{(n-1)}(u-t)du\right] \\
&= \sum_{i=1}^{n-1}(-1)^i\phi^{(n-i)}(t)\phi^{(i-1)}(0) + (-1)^{n-1}\left[-\phi(t)\phi^{(n-1)}(0) - \int_t^1 \phi(u)\phi^{(n)}(u-t)du\right] \\
&= \sum_{i=1}^{n-1}(-1)^i\phi^{(n-i)}(t)\phi^{(i-1)}(0) + (-1)^n\phi^{(n-n)}(t)\phi^{(n-1)}(0) + (-1)^n\int_t^1 \phi(u)\phi^{(n)}(u-t)du \\
&= \sum_{i=1}^{n}(-1)^i\phi^{(n-i)}(t)\phi^{(i-1)}(0) + (-1)^n\int_t^1 \phi(u)\phi^{(n)}(u-t)du
\end{aligned}
$$

Thus, the hypothesis holds for $n$, so, by induction, the theorem holds for all $n \in N$, with $t > 0$.

Now assume that $t < 0$ and consider the hypothesis that $\frac{d^n k}{dt^n}(t) = \sum_{i=1}^n (-1)^{i-1}\phi^{(n-i)}(1+t)\phi^{(i-1)}(1) + (-1)^n\int_0^{1+t}\phi(u)\phi^{(n)}(u-t)du$. From the above, this clearly holds for $n = 1$. Now assume that it holds for all $n' < n$, then,

$$
\begin{aligned}
\frac{d^n k}{dt^n}(t) = \frac{d}{dt}\frac{d^{n-1}k}{dt^{n-1}} &= \frac{d}{dt}\left[\sum_{i=1}^{n-1}(-1)^{i-1}\phi^{(n-1-i)}(1+t)\phi^{(i-1)}(1) + (-1)^{n-1}\int_0^{1+t} \phi(u)\phi^{(n-1)}(u-t)du\right] \\
&= \sum_{i=1}^{n-1}(-1)^{i-1}\phi^{(n-i)}(1+t)\phi^{(i-1)}(1) + (-1)^{n-1}\left[\frac{d}{dt}\int_0^{1+t} \phi(u)\phi^{(n-1)}(u-t)du\right] \\
&= \sum_{i=1}^{n-1}(-1)^{i-1}\phi^{(n-i)}(1+t)\phi^{(i-1)}(1) + (-1)^{n-1}\left[\phi(1+t)\phi^{(n-1)}(1) - \int_0^{1+t} \phi(u)\phi^{(n)}(u-t)du\right] \\
&= \sum_{i=1}^{n-1}(-1)^{i-1}\phi^{(n-i)}(1+t)\phi^{(i-1)}(1) + (-1)^{n-1}\phi^{(n-n)}(1+t)\phi^{(n-1)}(1) + (-1)^n\int_0^{1+t} \phi(u)\phi^{(n)}(u-t)du \\
&= \sum_{i=1}^{n}(-1)^{i-1}\phi^{(n-i)}(1+t)\phi^{(i-1)}(1) + (-1)^n\int_0^{1+t} \phi(u)\phi^{(n)}(u-t)du
\end{aligned}
$$

Thus, the hypothesis holds for $n$, so, by induction, it holds for for all $n \in N$ with $t < 0$.

Note that, if $k \in \mathcal{C}^{n-1}$, in a neighborhood of 0, then the above results also hold for the one-sided $n^{th}$ derivatives at 0, as well. In particular, since $\phi \in \mathcal{C}^n[0,1]$, then $\frac{d^n k}{dt^n}(t)$ will be continuous on $[-1,1]$, except, perhaps, at 0; since $\phi$ and its first $n$ derivatives are continuous, and, thus uniformly bounded on $[0,1]$, this follows from an application of the Dominated Convergence Theorem to the two expressions above. If the left and right derivatives are also equal at 0, then $k \in \mathcal{C}^n[-1,1]$.

It is now useful to compute $\frac{d^n k}{dt^n}(0)$. We first compute $\lim_{t\to 0^\pm} \frac{d^n k}{dt^n}(t)$.

$$\begin{aligned}
\lim_{t\to 0^+} \frac{d^n k}{dt^n}(t) &= \lim_{t\to 0^+} \left[ \sum_{i=1}^{n} (-1)^i \phi^{(n-i)}(t)\phi^{(i-1)}(0) + (-1)^n \int_t^1 \phi(u)\phi^{(n)}(u-t)du \right] \\
&= \sum_{i=1}^{n} (-1)^i \phi^{(n-i)}(0)\phi^{(i-1)}(0) + (-1)^n \int_0^1 \phi(u)\phi^{(n)}(u)du
\end{aligned}$$

$$\begin{aligned}
\lim_{t\to 0^-} \frac{d^n k}{dt^n}(t) &= \lim_{t\to 0^-} \left[ \sum_{i=1}^{n} (-1)^{i-1} \phi^{(n-i)}(1+t)\phi^{(i-1)}(1) + (-1)^n \int_0^{1+t} \phi(u)\phi^{(n)}(u-t)du \right] \\
&= \sum_{i=1}^{n} (-1)^{i-1} \phi^{(n-i)}(1)\phi^{(i-1)}(1) + (-1)^n \int_0^1 \phi(u)\phi^{(n)}(u)du
\end{aligned}$$

We will use integration by parts in order to reexpress $\int_0^1 \phi(u)\phi^{(n)}(u)du$.

$$\begin{aligned}
\int_0^1 \phi(u)\phi^{(n)}(u)du &= \phi(u)\phi^{(n-1)}(u)|_0^1 - \int_0^1 \phi^{(1)}(u)\phi^{(n-1)}(u)du \\
&= \phi(1)\phi^{(n-1)}(1) - \phi(0)\phi^{(n-1)}(0) - \int_0^1 \phi^{(1)}(u)\phi^{(n-1)}(u)du \\
&= \phi(1)\phi^{(n-1)}(1) - \phi(0)\phi^{(n-1)}(0) - \phi^{(1)}(u)\phi^{(n-2)}(u)|_0^1 + \int_0^1 \phi^{(2)}(u)\phi^{(n-2)}(u)du \\
&= \sum_{i=1}^{2}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] + (-1)^2\int_0^1 \phi^{(2)}(u)\phi^{(n-2)}(u)du \\
&\dots \\
&= \sum_{i=1}^{n}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] + (-1)^n\int_0^1 \phi^{(n)}(u)\phi(u)du
\end{aligned}$$

$$(1-(-1)^n)\int_0^1 \phi(u)\phi^{(n)}(u)du = \sum_{i=1}^{n}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right]$$

For $n$ odd,

$$\int_0^1 \phi(u)\phi^{(n)}(u)du = \frac{1}{2}\sum_{i=1}^{n}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right]$$

so

$$
\begin{aligned}
\lim_{t\to 0^+}\frac{d^n k}{dt^n}(t) &= \sum_{i=1}^{n}(-1)^i\phi^{(n-i)}(0)\phi^{(i-1)}(0) + (-1)^n\int_0^1 \phi(u)\phi^{(n)}(u)du \\
&= \sum_{i=1}^{n}(-1)^i\phi^{(n-i)}(0)\phi^{(i-1)}(0) - \frac{1}{2}\sum_{i=1}^{n}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] \\
&= \sum_{i=1}^{n}(-1)^i\phi^{(n-i)}(0)\phi^{(i-1)}(0) + \frac{1}{2}\sum_{i=1}^{n}(-1)^{i}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] \\
&= \frac{1}{2}\sum_{i=1}^{n}(-1)^{i}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) + \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right]
\end{aligned}
$$

$$
\begin{aligned}
\lim_{t\to 0^-}\frac{d^n k}{dt^n}(t) &= \sum_{i=1}^{n}(-1)^{i-1}\phi^{(n-i)}(1)\phi^{(i-1)}(1) + (-1)^n\int_0^1 \phi(u)\phi^{(n)}(u)du \\
&= \sum_{i=1}^{n}(-1)^{i-1}\phi^{(n-i)}(1)\phi^{(i-1)}(1) - \frac{1}{2}\sum_{i=1}^{n}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] \\
&= \frac{1}{2}\sum_{i=1}^{n}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) + \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] \\
&= -\frac{1}{2}\sum_{i=1}^{n}(-1)^{i}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) + \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] \\
&= -\lim_{t\to 0^+}\frac{d^n k}{dt^n}(t)
\end{aligned}
$$

For $n$ even,

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}(-1)^{i-1}\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) - \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] \\
&= \sum_{i=1}^{n}(-1)^{i-1}\phi^{(i-1)}(1)\phi^{(n-i)}(1) + \sum_{i=1}^{n}(-1)^{i}\phi^{(i-1)}(0)\phi^{(n-i)}(0)
\end{aligned}
$$

so

$$
\sum_{i=1}^{n}(-1)^{i-1}\phi^{(i-1)}(1)\phi^{(n-i)}(1) = -\sum_{i=1}^{n}(-1)^{i}\phi^{(i-1)}(0)\phi^{(n-i)}(0)
$$

Note that, for $n$ even,

$$\begin{aligned}
&\sum_{i=1}^{n}(-1)^i\phi^{(i-1)}(0)\phi^{(n-i)}(0) \\
&\quad= \sum_{i=1}^{\frac{n}{2}}(-1)^i\phi^{(i-1)}(0)\phi^{(n-i)}(0) + \sum_{i=\frac{n}{2}+1}^{n}(-1)^i\phi^{(i-1)}(0)\phi^{(n-i)}(0) \\
&\quad= \sum_{i=1}^{\frac{n}{2}}(-1)^i\phi^{(i-1)}(0)\phi^{(n-i)}(0) + \sum_{i=\frac{n}{2}+1}^{n}(-1)(-1)(-1)^n(-1)^{-i}\phi^{(n-(n-i+1))}(0)\phi^{((n-i+1)-1)}(0) \\
&\quad= \sum_{i=1}^{\frac{n}{2}}(-1)^i\phi^{(i-1)}(0)\phi^{(n-i)}(0) - \sum_{i=\frac{n}{2}+1}^{n}(-1)^{n-i+1}\phi^{(n-(n-i+1))}(0)\phi^{((n-i+1)-1)}(0) \\
&\quad= \sum_{i=1}^{\frac{n}{2}}(-1)^i\phi^{(i-1)}(0)\phi^{(n-i)}(0) - \sum_{j=1}^{\frac{n}{2}}(-1)^j\phi^{(n-j)}(0)\phi^{(j-1)}(0) \\
&\quad= 0
\end{aligned}$$

where $j = n - i + 1$.

Thus, $\sum_{i=1}^{n}(-1)^{i-1}\phi^{(i-1)}(1)\phi^{(n-i)}(1) = -\sum_{i=1}^{n}(-1)^i\phi^{(i-1)}(0)\phi^{(n-i)}(0) = 0$, so $\lim_{t\to0^+}\frac{d^nk}{dt^n}(t) = \lim_{t\to0^-}\frac{d^nk}{dt^n}(t) = \int_0^1\phi(u)\phi^{(n)}(u)du.$

Therefore, for $n$ odd, $\lim_{t\to0^+}\frac{d^nk}{dt^n}(t) = \frac{1}{2}\sum_{i=1}^{n}(-1)^i\left[\phi^{(i-1)}(1)\phi^{(n-i)}(1) + \phi^{(i-1)}(0)\phi^{(n-i)}(0)\right] = -\lim_{t\to0^+}\frac{d^nk}{dt^n}(t)$ and, for $n$ even, $\lim_{t\to0^+}\frac{d^nk}{dt^n}(t) = \lim_{t\to0^-}\frac{d^nk}{dt^n}(t) = \int_0^1\phi(u)\phi^{(n)}(u)du.$

Thus, if $n$ is even, $\phi \in \mathcal{C}^n[0,1]$, and $k \in \mathcal{C}^{n-1}[-1,1]$, then, from the above, $\frac{d^nk}{dt^{n+}}(0) = \frac{d^nk}{dt^{n-}}(0)$ so $k \in \mathcal{C}^n[-1,1]$ and $\frac{d^nk}{dt^n}(0) = \int_0^1\phi(u)\phi^{(n)}(u)du.$

Finally, if $\phi$ generates a $q = 2$ kernel, then $k^{(1)}(0) = 0$, so the left and right derivatives of $k$ are both 0 at 0, so $k \in \mathcal{C}^1[-1,1]$. Then, $k \in \mathcal{C}^2[1,-1]$ and $k^{(2)}(0) = -\frac{1}{2}\frac{d^2k}{dt^2}(0) = -\frac{1}{2}\int_0^1\phi(u)\phi^{(n)}(u)du.$

□

## 8.2. **Optimal Weights for $q = 1$ Orthogonal Series.**

In this appendix, we will derive the optimal weights for a $q = 1$ orthogonal series. Recall that every Weighted Orthogonal Series (WOS) has a limiting implied mean kernel $k_w = \sum_{i=b_0}^{B} w_ik_i$, where $\{w_i\}_{i=b_0}^{B}$ is a sequence of weights and $k_i$ is the limiting implied mean kernel of the $i^{th}$ term in the series. Then, $k_w^{(q)} = \sum_{i=b_0}^{B} w_i k_i^{(q)}$, so, for $q = 1$ WOS estimators, $I_q[k_w] = \left(k_w^{(q)}(0)\right)^{\frac{1}{q}} \sum_i w_i^2 = \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) \cdot \sum_{i=b_0}^{B} w_i^2$. We will now minimize this quantity under the constraint $\sum_{i=b_0}^{B} w_i = 1$, in order to obtain the optimal sequence of weights. We begin by constructing the Lagrangian, setting its first derivative equal to zero, and then using this to solve for $\lambda$.

$$\mathcal{L} = \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) \cdot \sum_{i=b_0}^{B} w_i^2 + \lambda \left(1 - \sum_{i=b_0}^{B} w_i\right)$$

$$0 = \frac{\partial \mathcal{L}}{\partial w_i} = k_i^{(1)}(0) \cdot \sum_{i=b_0}^{B} w_i^2 + 2w_i \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) - \lambda$$

$$\lambda = k_i^{(1)}(0) \cdot \sum_{i=b_0}^{B} w_i^2 + 2w_i \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0)$$

$$\begin{aligned}
\lambda &= (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} \left[k_i^{(1)}(0) \cdot \sum_{i=b_0}^{B} w_i^2 + 2w_i \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0)\right] \\
&= (B - b_0 + 1)^{-1} \left(\sum_{i=b_0}^{B} k_i^{(1)}(0) \cdot \sum_{i=1}^{B} w_i^2 + 2 \sum_{i=b_0}^{B} w_i \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0)\right) \\
&= (B - b_0 + 1)^{-1} \left(\sum_{i=b_0}^{B} k_i^{(1)}(0) \cdot \sum_{i=1}^{B} w_i^2 + 2 \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0)\right)
\end{aligned}$$

We can then use this expression to eliminate $\lambda$ from the first order condition giving,

$$
\begin{aligned}
0 &= k_i^{(1)}(0) \cdot \sum_{i=b_0}^{B} w_i^2 + 2w_i \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) - \lambda \\
&= k_i^{(1)}(0) \cdot \sum_{i=b_0}^{B} w_i^2 + 2w_i \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \cdot \sum_{i=1}^{B} w_i^2 + 2 \cdot \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) \right) \\
&= \left( k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) \cdot \sum_{i=b_0}^{B} w_i^2 + \left( w_i - (B - b_0 + 1)^{-1} \right) \cdot 2 \sum_{i=b_0}^{B} w_i k_i^{(1)}(0)
\end{aligned}
$$

We can then rewrite this as,

$$
w_i = (B - b_0 + 1)^{-1} - \left( k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) \cdot \sum_{i=b_0}^{B} w_i^2 \cdot \left( 2 \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) \right)^{-1}
$$

Let $A = \frac{1}{2} \left( \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) \right)^{-1} \sum_{i=b_0}^{B} w_i^2$, then

$$
w_i = (B - b_0 + 1)^{-1} - \left( k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) A
$$

We can feed this expression back into the definition of $A$ in order to obtain a quadratic equation, in terms of $A$, which we can then solve to obtain $A$ as a function of the $k_i^{(1)}$s.

$$\begin{aligned}
2A &= \left(\sum_{i=b_0}^{B}\left(\left((B-b_0+1)^{-1}-\left(k_i^{(1)}(0)-(B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)A\right)k_i^{(1)}(0)\right)\right)^{-1} \\
&\quad\times\sum_{i=b_0}^{B}\left((B-b_0+1)^{-1}-\left(k_i^{(1)}(0)-(B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)A\right)^2 \\
&= \left(\sum_{i=b_0}^{B}\left(\left((B-b_0+1)^{-1}-\left(k_i^{(1)}(0)-(B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)A\right)k_i^{(1)}(0)\right)\right)^{-1} \\
&\quad\times\sum_{i=b_0}^{B}\left((B-b_0+1)^{-2}-2(B-b_0+1)^{-1}\left(k_i^{(1)}(0)-(B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)A\right. \\
&\qquad\left.+\left(k_i^{(1)}(0)-(B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2A^2\right) \\
&= \left((B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A\right)^{-1} \\
&\quad\times\left((B-b_0+1)^{-1}-2(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)-\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)A\right. \\
&\qquad\left.+\sum_{i=b_0}^{B}\left(k_i^{(1)}(0)^2-2k_i^{(1)}(0)(B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)+(B-b_0+1)^{-2}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A^2\right) \\
&= \left((B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A\right)^{-1} \\
&\quad\times\left((B-b_0+1)^{-1}-2(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)-\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)A\right. \\
&\qquad\left.+\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-2(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2+(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A^2\right) \\
&= \left((B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A\right)^{-1} \\
&\quad\times\left((B-b_0+1)^{-1}+\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A^2\right)
\end{aligned}$$

$$2A\left((B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A\right)$$
$$=(B-b_0+1)^{-1}+\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A^2$$

$$0=3\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-(B-b_0+1)^{-1}\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A^2-2(B-b_0+1)^{-1}\sum_{i=b_0}^{B}k_i^{(1)}(0)\cdot A$$
$$+(B-b_0+1)^{-1}$$
$$=(B-b_0+1)^{-1}\left[3\left((B-b_0+1)\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)A^2-2\sum_{i=b_0}^{B}k_i^{(1)}(0)\cdot A+1\right]$$

We can now use the quadratic formula to write $A$ in terms of the weights and $k^{(1)}(0)_i$s.

$$A=\frac{2\sum_{i=b_0}^{B}k_i^{(1)}(0)\pm\sqrt{\left(2\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2-4\cdot 3\left((B-b_0+1)\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)}}{2\cdot 3\left((B-b_0+1)\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)}$$
$$=\frac{\sum_{i=b_0}^{B}k_i^{(1)}(0)\pm\sqrt{\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2-3\left((B-b_0+1)\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)}}{3\left((B-b_0+1)\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)}$$
$$=\frac{\sum_{i=b_0}^{B}k_i^{(1)}(0)\pm\sqrt{4\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2-3(B-b_0+1)\sum_{i=b_0}^{B}k_i^{(1)}(0)^2}}{3\left((B-b_0+1)\sum_{i=b_0}^{B}k_i^{(1)}(0)^2-\left(\sum_{i=b_0}^{B}k_i^{(1)}(0)\right)^2\right)}$$

Using this, we can now write the weights as functions of the $k^{(1)}(0)_i$s.

$$
\begin{aligned}
w_i &= (B - b_0 + 1)^{-1} - \left( k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) A \\
&= (B - b_0 + 1)^{-1} - \left( k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) \\
&\quad \times \frac{\sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3(B - b_0 + 1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2}}{3 \left( (B - b_0 + 1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right)}
\end{aligned}
$$

Note that there is an unresolved $\pm$ in this expression. Thus, it is necessary to compute the weights in both cases, check that they are nonnegative, and then use the set of (valid) weights which gives the lowest value of $I_q[k_w]$. We can use these weights to compute both $k^{(1)}(0)$ and $\sum_{i=b_0}^{B} w_i^2$,

$$
\begin{aligned}
k^{(1)}(0) &= \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) = \sum_{i=b_0}^{B} \left( (B-b_0+1)^{-1} - \left( k_i^{(1)}(0) - (B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) A \right) k_i^{(1)}(0) \\
&= (B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) - \left( \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - (B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right) A \\
&= (B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) - \left( \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - (B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right) \\
&\quad \times \frac{\sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3(B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2}}{3 \left( (B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right)} \\
&= (B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \\
&\quad - \frac{1}{3}(B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3(B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2} \right) \\
&= \frac{2}{3}(B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \\
&\quad \mp \frac{1}{3}(B-b_0+1)^{-1} \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3(B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2}
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=b_0}^{B} w_i^2 &= \sum_{i=b_0}^{B} \left( (B-b_0+1)^{-1} - \left( k_i^{(1)}(0) - (B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) A \right)^2 \\
&= \sum_{i=b_0}^{B} \left( (B-b_0+1)^{-2} - 2(B-b_0+1)^{-1} \left( k_i^{(1)}(0) - (B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) A \right. \\
&\quad \left. + \left( k_i^{(1)}(0) - (B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 A^2 \right) \\
&= (B-b_0+1)^{-1} - 2(B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) - \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) A \\
&\quad + \sum_{i=b_0}^{B} \left( k_i^{(1)}(0)^2 - 2(B-b_0+1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \cdot k_i^{(1)}(0) + (B-b_0+1)^{-2} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right) A^2 \\
&= (B-b_0+1)^{-1} - 2 \cdot 0 \cdot A \\
&\quad + \left( \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - 2(B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 + (B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right) A^2 \\
&= (B-b_0+1)^{-1} + \left( \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - (B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right) A^2 \\
&= (B-b_0+1)^{-1} + \left( \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - (B-b_0+1)^{-1} \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right) \\
&\quad \times \left( \frac{\sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3(B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2}}{3 \left( (B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right)} \right)^2 \\
&= (B-b_0+1)^{-1} + \frac{\left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3(B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2} \right)^2}{9(B-b_0+1) \left( (B-b_0+1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right)}
\end{aligned}
$$

$$
\begin{aligned}
k^{(1)}(0)\sum_{i=b_0}^{B} w_i^2 = &\left(\frac{2}{3}(B-b_0+1)^{-1}\sum_{i=b_0}^{B} k_i^{(1)}(0)\right.\\
&\left.\mp\frac{1}{3}(B-b_0+1)^{-1}\sqrt{4\left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2 - 3(B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2}\right)\\
&\times\left((B-b_0+1)^{-1} + \frac{\left(\sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4\left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2 - 3(B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2}\right)^2}{9(B-b_0+1)\left((B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2\right)}\right)
\end{aligned}
$$

8.3. **Legendre Polynomials.** The Legendre Polynomials, $\{P_i\}$, are a set of orthogonal polynomials, typically defined on $[-1,1]$ that satisfy, $P_i(0) = (-1)^i$, $P_i(1) = 1$ and are normalized to have square norm $2(2i+1)^{-\frac{1}{2}}$, for the $i^{th}$ Legendre polynomial. Remapping their domain to the interval $[0,1]$ will result in dividing their norms by 2, but will not alter their orthogonality. Therefore, if we define $\phi_i(x) = (2i+1)^{\frac{1}{2}} P_i\left(\frac{1}{2}(x+1)\right)$, then the $\phi_i$s will define an orthonormal series on $[0,1]$.

Using Proposition 7, $k_i^{(1)}(0) = \frac{1}{2}\left((2i+1)+(2i+1)\right) = 2i+1$. Then,

$$
\sum_{i=b_0}^{B} k_i^{(1)}(0) = \sum_{i=1}^{B} 2i+1 = 2\cdot\frac{1}{2}B(B+1)+B = B(B+1)+B = B(B+2)
$$

$$
\begin{aligned}
\sum_{i=1}^{B} k_i^{(1)}(0)^2 &= \sum_{i=1}^{B}(2i+1)^2 = \sum_{i=1}^{B} 4i^2+4i+1 = 4\cdot\frac{1}{6}B(B+1)(2B+1) + 4\cdot\frac{1}{2}B(B+1)+B\\
&= \frac{2}{3}B(B+1)(2B+1) + 2B(B+1) + B = \frac{1}{3}\left(4B^3+12B^2+11B\right)
\end{aligned}
$$

Using the formula for optimal weights (with $b_0 = 1$) gives,

$$
\begin{aligned}
w_i &= (B - b_0 + 1)^{-1} - \left( k_i^{(1)}(0) - (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \right) \\
&\quad \times \frac{\sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3B \sum_{i=b_0}^{B} k_i^{(1)}(0)^2}}{3 \left( B \sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 \right)} \\
&= B^{-1} - \left( k_i^{(1)}(0) - B^{-1} B(B+2) \right) \\
&\quad \times \frac{B(B+2) \pm \sqrt{4 \left( B(B+2) \right)^2 - 3B \cdot \frac{1}{3} \left( 4B^3 + 12B^2 + 11B \right)}}{3 \left( B \cdot \frac{1}{3} \left( 4B^3 + 12B^2 + 11B \right) - \left( B(B+2) \right)^2 \right)} \\
&= B^{-1} - \left( k_i^{(1)}(0) - B - 2 \right) \frac{B(B+2) \pm \sqrt{4B^3 + 5B^2}}{3 \cdot \frac{1}{3} \left( B^4 - B^2 \right)} \\
&= B^{-1} - \left( k_i^{(1)}(0) - B - 2 \right) \frac{B + 2 \pm \sqrt{4B + 5}}{B \left( B^2 - B \right)}
\end{aligned}
$$

$$
\begin{aligned}
k^{(1)}(0) &= \frac{2}{3} (B - b_0 + 1)^{-1} \sum_{i=b_0}^{B} k_i^{(1)}(0) \\
&\qquad \mp \frac{1}{3} (B - b_0 + 1)^{-1} \sqrt{4 \left( \sum_{i=b_0}^{B} k_i^{(1)}(0) \right)^2 - 3(B - b_0 + 1) \sum_{i=b_0}^{B} k_i^{(1)}(0)^2} \\
&= \frac{2}{3} B^{-1} \cdot B(B+2) \mp \frac{1}{3} B^{-1} \sqrt{4B^3 + 5B^2} \\
&= \frac{2}{3} (B+2) \mp \frac{1}{3} \sqrt{4B + 5}
\end{aligned}
$$

$$
\begin{aligned}
\lim_{B \to \infty} B^{-1} k^{(1)}(0) &= \frac{2}{3} \lim_{B \to \infty} B^{-1} (B+2) \mp \frac{1}{3} \lim_{B \to \infty} B^{-1} \sqrt{4B + 5} \\
&= \frac{2}{3}
\end{aligned}
$$

$$\begin{aligned}\sum_{i=1}^{B} w_i^2 &= B^{-1} + \frac{\left(B(B+2) \pm \sqrt{4B^3 + 5B^2}\right)^2}{9B \cdot \frac{1}{3}\left(B^4 - B^2\right)} \\ &= B^{-1} + \frac{\left(B + 2 \pm \sqrt{4B+5}\right)^2}{3B\left(B^2 - 1\right)}\end{aligned}$$

$$\begin{aligned}\lim_{B\to\infty} B \sum_{i=1}^{B} w_i^2 &= 1 + \lim_{B\to\infty} \frac{\left(B + 2 \pm \sqrt{4B+5}\right)^2}{3\left(B^2 - 1\right)} = 1 + \frac{1}{3} \\ &= \frac{4}{3}\end{aligned}$$

Then, the asymptotic value of $I_q[k]$ is $\lim_{B\to\infty} I_q[k] = \lim_{B\to\infty} k^{(1)}(0) \sum_{i=1}^{B} w_i^2 = \frac{2}{3} \cdot \frac{4}{3} = \frac{8}{9}$. If we instead used equal weights, so that $w_i = B^{-1}$ for $i \leq B$, we would get $k^{(1)}(0) = \sum_{i=1}^{B} B^{-1}(2i+1) = B^{-1} \cdot B(B+2) = B+2$, $\sum_{i=1}^{B} w_i^2 = \sum_{i=1}^{B} B^{-2} = B^{-1}$, so $I_q[k] = 1 + 2B^{-1}$ and $\lim_{B\to\infty} I_q[k] = 1 > \frac{8}{9}$.

### 8.4. **Haar Wavelets.**

We begin by computing the limiting implied kernel for each term in the system. 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 $n, \ell \in Z_+, 0 \leq \ell < 2^n$. Then, the limiting implied kernel corresponding to the wavelet $\psi_{n,\ell}$ is given by,

$$
\begin{aligned}
k_{n,\ell}(t) &= \int_{\max(0,t)}^{\min(1,1+t)} \psi_{n,\ell}(u)\psi_{n,\ell}(u-t)du = \int_{\max(0,t)}^{\min(1,1+t)} 2^{\frac{n}{2}}\psi\left(2^n u-\ell\right)2^{\frac{n}{2}}\psi\left(2^n(u-t)-\ell\right)du \\
&= 2^n \int_{\max(0,t)}^{\min(1,1+t)} \left[I_{[0,\frac{1}{2})}\left(2^n u-\ell\right) - I_{[\frac{1}{2},1)}\left(2^n u-\ell\right)\right]\left[I_{[0,\frac{1}{2})}\left(2^n(u-t)-\ell\right) - I_{[\frac{1}{2},1)}\left(2^n(u-t)-\ell\right)\right]du \\
&= 2^n \int_0^1 I_{[0,\frac{1}{2})}\left(2^n u-\ell\right)I_{[0,\frac{1}{2})}\left(2^n(u-t)-\ell\right) - I_{[0,\frac{1}{2})}\left(2^n u-\ell\right)I_{[\frac{1}{2},1)}\left(2^n(u-t)-\ell\right) \\
&\quad - I_{[\frac{1}{2},1)}\left(2^n u-\ell\right)I_{[0,\frac{1}{2})}\left(2^n(u-t)-\ell\right) + I_{[\frac{1}{2},1)}\left(2^n u-\ell\right) + I_{[\frac{1}{2},1)}\left(2^n(u-t)-\ell\right)du \\
&= 2^n \int_0^1 I\left[2^{-n}\ell \le u, u-t < 2^{-n}\left(\ell+2^{-1}\right)\right] \\
&\qquad - I\left[\left(2^{-n}\ell \le u < 2^{-n}\left(\ell+2^{-1}\right)\right)\wedge\left(2^{-n}\left(\ell+2^{-1}\right) \le u-t < 2^{-n}\left(\ell+1\right)\right)\right] \\
&\qquad - I\left[\left(2^{-n}\left(\ell+2^{-1}\right) \le u < 2^{-n}\left(\ell+1\right)\right)\wedge\left(2^{-n}\ell \le u-t < 2^{-n}\left(\ell+2^{-1}\right)\right)\right] \\
&\qquad + I\left[2^{-n}\left(\ell+2^{-1}\right) \le u, u-t < 2^{-n}\left(\ell+1\right)\right]du \\
&= 2^n\left[2^{-(n+1)}\left(1-2^{n+1}|t|\right)I_{\left(-2^{-(n+1)},2^{-(n+1)}\right)}(t) - 2^{-(n+1)}\left(1-\left|1+2^{n+1}t\right|\right)I_{(-2^{-n},0]}(t)\right. \\
&\quad \left. -2^{-(n+1)}\left(1-\left|1-2^{n+1}t\right|\right)I_{[0,2^{-n})}(t) + 2^{-(n+1)}\left(1-2^{n+1}|t|\right)I_{\left(-2^{-(n+1)},2^{-(n+1)}\right)}(t)\right] \\
&= 2^{-1}\cdot 2\left(1-2^{n+1}|t|\right)I_{\left[0,2^{-(n+1)}\right)}(|t|) - 2^{-1}\left(1-\left|1-2^{n+1}|t|\right|\right)I_{[0,2^{-n})}(|t|) \\
&= \left(1-2^{n+1}|t|\right)I_{\left[0,2^{-(n+1)}\right)}(|t|) - 2^{-1}\left(2^{n+1}|t|\right)I_{\left[0,2^{-(n+1)}\right)}(|t|) \\
&\quad - 2^{-1}\left(2-2^{n+1}|t|\right)I_{\left[2^{-(n+1)},2^{-n}\right)}(|t|) \\
&= \left(1-3\cdot 2^n|t|\right)I_{\left[0,2^{-(n+1)}\right)}(|t|) - \left(1-2^n|t|\right)I_{\left[2^{-(n+1)},2^{-n}\right)}(|t|)
\end{aligned}
$$

where the fourth equality is due to the fact that the indicators in the wavelet function automatically enforce that the integrand is 0 unless $\max(0,t) \le u \le \min(1,1+t)$, as can be seen explicitly when the indicator functions are rewritten in the fifth equality.

Note that this is independent of $\ell$, since, within each level of the hierarchy, the basis functions are simply translations of each other, so we will simply write,

$$k_n(t) = (1-3\cdot 2^n|t|)I_{\left[0,2^{-(n+1)}\right)}(|t|) - (1-2^n|t|)I_{\left[2^{-(n+1)},2^{-n}\right)}(|t|)$$

Then, it is clear that $k_n$ is a $q=1$ kernel with

$$k_n^{(1)}(0) = \lim_{t \to 0} |t|^{-1}(1 - (1 - 3 \cdot 2^n |t|)) = 3 \cdot 2^n$$

$$\begin{aligned}
\int k_n^2(t)dt &= \int (1 - 3 \cdot 2^n |t|)^2 \, I_{[0,2^{-(n+1)})}(|t|) + (1 - 2^n |t|)^2 \, I_{[2^{-(n+1)},2^{-n})}(|t|)dt \\
&= 2\int_0^{2^{-(n+1)}} (1 - 3 \cdot 2^n t)^2 \, dt + 2\int_{2^{-(n+1)}}^{2^{-n}} (1 - 2^n t)^2 \, dt \\
&= 2\left[\left[t - \frac{1}{2} 2 \cdot 3 \cdot 2^n t^2 + \frac{1}{3} 9 \cdot 2^{2n} t^3\right]_0^{2^{-(n+1)}} + \left[t - \frac{1}{2} 2 \cdot 2^n t^2 + \frac{1}{3} \cdot 2^{2n} t^3\right]_{2^{-(n+1)}}^{2^{-n}}\right] \\
&= 2\left[\left[t - 3 \cdot 2^n t^2 + 3 \cdot 2^{2n} t^3\right]_0^{2^{-(n+1)}} + \left[t - 2^n t^2 + \frac{1}{3} \cdot 2^{2n} t^3\right]_{2^{-(n+1)}}^{2^{-n}}\right] \\
&= 2\left[2^{-(n+1)} - 3 \cdot 2^n 2^{-2(n+1)} + 3 \cdot 2^{2n} 2^{-3(n+1)}\right] \\
&\quad + 2\left[2^{-n} - 2^n 2^{-2n} + \frac{1}{3} \cdot 2^{2n} 2^{-3n} - \left(2^{-(n+1)} - 2^n 2^{-2(n+1)} + \frac{1}{3} \cdot 2^{2n} 2^{-3(n+1)}\right)\right] \\
&= 2\left[2^{-n-1} - 3 \cdot 2^{-n-2} + 3 \cdot 2^{-n-3}\right] \\
&\quad + 2\left[2^{-n} - 2^{-n} + \frac{1}{3} 2^{-n} - \left(2^{-n-1} - 2^{-n-2} + \frac{1}{3} \cdot 2^{-n-3}\right)\right] \\
&= 2^{-n}\left[1 - 3 \cdot 2^{-1} + 3 \cdot 2^{-2}\right] + 2^{-n}\left[\frac{2}{3} - \left(1 - 2^{-1} + \frac{1}{3} 2^{-2}\right)\right] \\
&= 2^{-n}\frac{1}{4} + 2^{-n}\left[\frac{2}{3} - \frac{1}{2} - \frac{1}{12}\right] = 2^{-n}\left[\frac{1}{4} + \frac{1}{12}\right] \\
&= \frac{1}{3} \cdot 2^{-n}
\end{aligned}$$

It is interesting to note that, for an individual kernel, $I_q[k_n] = k_n^{(1)}(0) \cdot \int k_n^2(t)dt = 3 \cdot 2^n \cdot \frac{1}{3} \cdot 2^{-n} = 1$. Let $n' > n$, then,

$$
\begin{aligned}
\int k_n(t)k_{n'}(t)dt &= \int \Big( (1 - 3\cdot 2^n|t|)\, I_{\left[0,2^{-(n+1)}\right)}(|t|) - (1-2^n|t|)\, I_{\left[2^{-(n+1)},2^{-n}\right)}(|t|) \Big) \\
&\quad \times \left( \left(1 - 3\cdot 2^{n'}|t|\right) I_{\left[0,2^{-(n'+1)}\right)}(|t|) - \left(1-2^{n'}|t|\right) I_{\left[2^{-(n'+1)},2^{-n'}\right)}(|t|) \right) dt \\
&= \int \left[ \left(1 - 3\cdot\left(2^n + 2^{n'}\right)|t| + 9\cdot 2^{n+n'}t^2\right) I_{[0,2^{-(n'+1)})}(|t|) \right. \\
&\quad \left. - \left(1 - \left(3\cdot 2^n + 2^{n'}\right)|t| + 3\cdot 2^{n+n'}t^2\right) I_{\left[2^{-(n'+1)},2^{-n'}\right)}(|t|) \right] dt \\
&= 2\int_0^{2^{-(n'+1)}} 1 - 3\cdot\left(2^n + 2^{n'}\right)t + 9\cdot 2^{n+n'}t^2 dt \\
&\quad - 2\int_{2^{-(n'+1)}}^{2^{-n'}} 1 - \left(3\cdot 2^n + 2^{n'}\right)t + 3\cdot 2^{n+n'}t^2 dt \\
&= 2\left[ t - \frac{3}{2}\cdot\left(2^n + 2^{n'}\right)t^2 + 3\cdot 2^{n+n'}t^3 \right]_0^{2^{-(n'+1)}} \\
&\quad - 2\left[ t - \frac{1}{2}\left(3\cdot 2^n + 2^{n'}\right)t^2 + 2^{n+n'}t^3 \right]_{2^{-(n'+1)}}^{2^{-n'}} \\
&= 2\left[ 2^{-(n'+1)} - \frac{3}{2}\cdot\left(2^n + 2^{n'}\right)2^{-2(n'+1)} + 3\cdot 2^{n+n'}2^{-3(n'+1)} \right] \\
&\quad - 2\left[ 2^{-n'} - \frac{1}{2}\left(3\cdot 2^n + 2^{n'}\right)2^{-2n'} + 2^{n+n'}2^{-3n'} \right] \\
&\quad + 2\left[ 2^{-(n'+1)} - \frac{1}{2}\left(3\cdot 2^n + 2^{n'}\right)2^{-2(n'+1)} + 2^{n+n'}2^{-3(n'+1)} \right] \\
&= \left(2^{-n'} - 2^{-(n'-1)} + 2^{-n'}\right) - \left(3\cdot 2^{-2} - 3 + 3\cdot 2^{-2}\right)2^n 2^{-2n'} \\
&\quad - \left(3\cdot 2^{-2} - 1 + 2^{-2}\right)2^{n'}2^{-2n'} + \left(3\cdot 2^{-2} - 2 + 2^{-2}\right)2^{n'+n}2^{-3n'} \\
&= 0 - 3\left(2^{-1} - 1\right)2^{n-2n'} - 0\cdot 2^{-n'} + (1-2)2^{n-2n'} \\
&= \frac{3}{2}\cdot 2^{n-2n'} - 2^{n-2n'} = \frac{1}{2}\cdot 2^{n-2n'} = 2^{n-2n'-1} \\
&= 2^{(n-n')-n'-1}
\end{aligned}
$$

Summing this over $n' > n$ gives,

$$
\begin{aligned}
\sum_{n'>n} 2^{(n-n')-n'-1} &= 2^{-n-3} \sum_{n'>n} 2^{-2(n'-n-1)} = 2^{-n-3} \sum_{i=0}^{\infty} 2^{-2i} = 2^{-n-3} \left(1-2^{-2}\right)^{-1} \\
&= 2^{-n-3} \left(\frac{3}{4}\right)^{-1} = \frac{4}{3} \cdot 2^{-n-3} = \frac{1}{3} \cdot 2^{-(n+1)} = \frac{1}{2} \cdot \frac{1}{3} \cdot 2^{-n} \\
&= \frac{1}{2} \int k_n^2(t) dt
\end{aligned}
$$

Since $\sum_{n,n'} \int k_n(t) k_{n'}(t) dt$ contains 2 copies of $\int k_n(t) k_{n'}(t) dt$ for $n' \neq n$, the sum of all terms of the form $\int k_n(t) k_{n'}(t) dt$ with $n' > n$ is $2 \cdot \frac{1}{2} \int k_n^2(t) dt = \int k_n^2(t) dt = \frac{1}{3} \cdot 2^{-n}$, so the sum of the cross terms is equal to the sum of the diagonal terms $\int k_n^2(t) dt$. We can now compute $k_w^{(1)}(0)$, $\int k_w^2(t) dt$, and $I_q[k_w]$

$$
\begin{aligned}
k_w^{(1)}(0) &= \lim_{t \to 0} |t|^{-1} (1 - k_w(t)) = \lim_{t \to 0} |t|^{-1} \left(1 - \sum_{n=1}^{N} w_n k_n(t)\right) = \lim_{t \to 0} |t|^{-1} \sum_{n=1}^{N} w_n (1 - k_n(t)) \\
&= \sum_{n=1}^{N} w_n \left(\lim_{t \to 0} |t|^{-1} (1 - k_n(t))\right) = \sum_{n=1}^{N} w_n k_n^{(1)}(0) = \sum_{n=1}^{N} w_n \cdot 3 \cdot 2^n \\
&= 3 \sum_{n=1}^{N} w_n \cdot 2^n
\end{aligned}
$$

$$
\begin{aligned}
\int k_w^2(t) dt &= \int \left(\sum_{n=1}^{N} w_n k_n(t)\right)^2 dt = \sum_{n,n'=1}^{N} \int w_n w_{n'} k_n(t) k_{n'}(t) dt \\
&= \sum_{n=1}^{N} \left(w_n^2 \int k_n^2(t) dt + 2 w_n \sum_{n'<n} w_{n'} \int k_n(t) k_{n'}(t) dt\right) \\
&= \sum_{n=1}^{N} \left(\frac{1}{3} \cdot 2^{-n} \cdot w_n^2 + 2 w_n \sum_{n'<n} w_{n'} \cdot 2^{n'-2n-1}\right) \\
&= \sum_{n=1}^{N} 2^{-n} \left(\frac{1}{3} \cdot w_n^2 + w_n \cdot 2^{-n} \sum_{n'<n} w_{n'} \cdot 2^{n'}\right)
\end{aligned}
$$

$$I_q[k_w] = k_w^{(1)}(0) \int k_w^2(t)dt = 3 \sum_n w_n \cdot 2^n \cdot \sum_n 2^{-n} \left( \frac{1}{3} \cdot w_n^2 + w_n \cdot 2^{2n+1} \sum_{n'>n} w_n' \cdot 2^{-(2n'+1)} \right)$$

Using equal weights for each basis function, and noting that for each level, $n$, there are $2^n$ basis functions, $w_n = \left(\sum_{n=1}^{N} 2^n\right)^{-1} 2^n = \left(2^{N+1} - 1\right)^{-1} 2^n$. Then we get:

$$\begin{aligned} k_w^{(1)}(0) &= \sum_{n=0}^{N} w_n k_n^{(1)}(0) = 3 \sum_n w_n \cdot 2^n = 3 \sum_{n=0}^{N} \left(2^{N+1} - 1\right)^{-1} 2^n \cdot 2^n = 3 \left(2^{N+1} - 1\right)^{-1} \sum_{n=0}^{N} 2^{2n} \\ &= 3 \left(2^{N+1} - 1\right)^{-1} \sum_{n=0}^{N} 4^n = 3 \left(2^{N+1} - 1\right)^{-1} \frac{4^{N+1} - 1}{4 - 1} = 3 \left(2^{N+1} - 1\right)^{-1} \frac{4^{N+1} - 1}{3} \\ &= \frac{4^{N+1} - 1}{2^{N+1} - 1} = \frac{\left(2^{N+1}\right)^2 - 1}{2^{N+1} - 1} = 2^{N+1} + 1 \end{aligned}$$

$$\begin{aligned}
\int k_w^2(x)dx &= \int\left(\sum_{n=0}^{N} w_n k_n(x)\right)^2 dx = \sum_{n=0}^{N} 2^{-n}\left(\frac{1}{3}\cdot w_n^2 + w_n\cdot 2^{-n}\sum_{n'<n} w_{n'}\cdot 2^{-n'}\right) \\
&= \sum_{n=0}^{N} 2^{-n}\left(\frac{1}{3}\cdot\left(\left(2^{N+1}-1\right)^{-1}2^n\right)^2 + \left(2^{N+1}-1\right)^{-1}2^n\cdot 2^{-n}\sum_{n'=0}^{n-1}\left(2^{N+1}-1\right)^{-1}2^{n'}\cdot 2^{n'}\right) \\
&= \left(2^{N+1}-1\right)^{-2}\sum_{n=0}^{N} 2^{-n}\left(\frac{1}{3}\cdot 2^{2n} + 2^n\cdot 2^{-n}\sum_{n'=0}^{n-1} 2^{2n'}\right) \\
&= \left(2^{N+1}-1\right)^{-2}\sum_{n=0}^{N}\left(\frac{1}{3}\cdot 2^n + 2^{-n}\sum_{n'=0}^{n-1} 4^{n'}\right) \\
&= \left(2^{N+1}-1\right)^{-2}\sum_{n=0}^{N}\left(\frac{1}{3}\cdot 2^n + 2^{-n}\frac{4^n-1}{4-1}\right) = \left(2^{N+1}-1\right)^{-2}\sum_{n=0}^{N}\left(\frac{1}{3}\cdot 2^n + \frac{1}{3}\cdot 2^{-n}\left(4^n-1\right)\right) \\
&= \frac{1}{3}\cdot\left(2^{N+1}-1\right)^{-2}\sum_{n=0}^{N}\left(2^n + \left(2^n - 2^{-n}\right)\right) = \frac{1}{3}\cdot\left(2^{N+1}-1\right)^{-2}\sum_{n=0}^{N}\left(2^{n+1}-2^{-n}\right) \\
&= \frac{1}{3}\left(2^{N+1}-1\right)^{-2}\left(2\cdot\frac{2^{N+1}-1}{2-1} - \frac{1-2^{-(N+1)}}{1-2^{-1}}\right) \\
&= \frac{2}{3}\left(2^{N+1}-1\right)^{-2}\left(2\cdot\left(2^{N+1}-1\right) - 2\cdot 2^{-(N+1)}\left(2^{N+1}-1\right)\right) \\
&= \frac{2}{3}\cdot\frac{1-2^{-(N+1)}}{2^{N+1}-1} \\
&= \frac{2}{3}\cdot 2^{-(N+1)}
\end{aligned}$$

$$I_q[k_w] = k_2^{(1)}(0)\int k_w^2(x)dx = \left(2^{N+1}+1\right)\cdot\frac{2}{3}\cdot 2^{-(N+1)} = \frac{2}{3}\left(1+2^{-(N+1)}\right)$$

Finally, we can also compute the sum of squared weights, for the implied kernel,

$$\begin{aligned}
\sum_{n=0}^{N} w_n^2 &= \sum_{n=0}^{N}\left(2^{N+1}-1\right)^{-2}2^{2n} = \left(2^{N+1}-1\right)^{-2}\sum_{n=0}^{N}4^n = \left(2^{N+1}-1\right)^{-2}\frac{4^{N+1}-1}{4-1} \\
&= 3^{-1}\frac{2^{N+1}+1}{2^{N+1}-1}
\end{aligned}$$

Note that this is actually not the value that is used when computing $I_q[k]$ for the orthogonal series estimator. For that, we count each basis function separately, which gives,

$$\begin{aligned}\sum_{n=0}^{N} w_{n,s}^2 &= \sum_{n=0}^{N} 2^n \left(2^{N+1}-1\right)^{-2} = \left(2^{N+1}-1\right)^{-2} \sum_{n=0}^{N} 2^n = \left(2^{N+1}-1\right)^{-2} \frac{2^{N+1}-1}{2-1} \\ &= \left(2^{N+1}-1\right)^{-1}\end{aligned}$$

Then, for the equal weighted orthogonal series, we get,

$$\begin{aligned}I_q[k_s] &= k_w^{(1)}(0) \cdot \sum_{n=0}^{N} w_{n,s}^2 = \left(2^{N+1}+1\right)\left(2^{N+1}-1\right)^{-1} \\ &= \frac{2^{N+1}+1}{2^{N+1}-1}\end{aligned}$$

To summarize, for the orthogonal series generated by the Haar system using equal weights for each basis function,

$$\begin{aligned}k_w^{(1)}(0) &= 2^{N+1}+1 \\ \int k_w^2(t)dt &= \frac{2}{3} \cdot 2^{-(N+1)} \\ I_q[k_w] &= \frac{2}{3}\left(1+2^{-(N+1)}\right) \\ \sum_{n=0}^{N} w_n^2 &= 3^{-1}\frac{2^{N+1}+1}{2^{N+1}-1}\end{aligned}$$

$$\begin{aligned}\sum_{n=0}^{N} w_{n,s}^2 &= \left(2^{N+1}-1\right)^{-1} \\ I_q[k_s] &= \frac{2^{N+1}+1}{2^{N+1}-1}\end{aligned}$$

We can also use this to compute the optimal weights for the Haar orthogonal series. Since each level $n$, has $2^n$ wavelets within it, we need to include this factor in all sums,

$$\sum_{n=0}^{N} 2^n k_n^{(1)}(0) = \sum_{n=0}^{N} 2^n \cdot 3 \cdot 2^n = 3 \sum_{n=0}^{N} 4^n = 3 \cdot \frac{4^{N+1}-1}{4-1} = 3 \cdot \frac{1}{3}\left(4^{N+1}-1\right)$$
$$= 4^{N+1} - 1$$

$$\sum_{n=0}^{N} 2^n \left(k_n^{(1)}(0)\right)^2 = \sum_{n=0}^{N} 2^n \left(3 \cdot 2^n\right)^2 = \sum_{n=0}^{N} 2^n \cdot 9 \cdot 4^n = 9 \sum_{n=0}^{N} 8^n = 9 \cdot \frac{8^{N+1}-1}{8-1}$$
$$= \frac{9}{7}\left(8^{N+1}-1\right)$$

We can now compute the weights,

$$\begin{aligned}
w_n &= (B-b_0+1)^{-1} - \left(k_i^{(1)}(0) - (B-b_0+1)^{-1}\sum_{i=b_0}^{B} k_i^{(1)}(0)\right) \\
&\quad \times \frac{\sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4\left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2 - 3(B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2}}{3\left((B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2\right)} \\
&= \left(2^{N+1}-1\right)^{-1} - \left(k_n^{(1)}(0) - \left(2^{N+1}-1\right)^{-1}\left(4^{N+1}-1\right)\right) \\
&\quad \times \frac{\left(4^{N+1}-1\right) \pm \sqrt{4\left(4^{N+1}-1\right)^2 - 3\left(2^{N+1}-1\right)\frac{9}{7}\left(8^{N+1}-1\right)}}{3\left(\left(2^{N+1}-1\right)\frac{9}{7}\left(8^{N+1}-1\right) - \left(4^{N+1}-1\right)^2\right)} \\
&= \left(2^{N+1}-1\right)^{-1} - \left(k_n^{(1)}(0) - \left(2^{N+1}-1\right)^{-1}\left(4^{N+1}-1\right)\right) \\
&\quad \times \frac{\left(4^{N+1}-1\right) \pm \sqrt{4\left(4^{N+1}-1\right)^2 - \frac{27}{7}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right)}}{3\left(\frac{9}{7}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right) - \left(4^{N+1}-1\right)^2\right)}
\end{aligned}$$

$$\begin{aligned}
k^{(1)}(0) &= \sum_{i=b_0}^{B} w_i k_i^{(1)}(0) \\
&= \frac{2}{3}(B-b_0+1)^{-1}\sum_{i=b_0}^{B} k_i^{(1)}(0) \\
&\quad \mp \frac{1}{3}(B-b_0+1)^{-1}\sqrt{4\left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2 - 3(B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2} \\
&= \frac{2}{3}\left(2^{N+1}-1\right)^{-1}\left(4^{N+1}-1\right) \\
&\quad \mp \frac{1}{3}\left(2^{N+1}-1\right)^{-1}\sqrt{4\left(4^{N+1}-1\right)^2 - 3\left(2^{N+1}-1\right)\cdot\frac{9}{7}\left(8^{N+1}-1\right)} \\
&= \frac{2}{3}\frac{\left(2^{N+1}\right)^2-1}{2^{N+1}-1} \mp \frac{1}{3}\sqrt{4\left(\frac{4^{N+1}-1}{2^{N+1}-1}\right)^2 - \frac{27}{7}\cdot\frac{8^{N+1}-1}{2^{N+1}-1}} \\
&= \frac{2}{3}\left(2^{N+1}+1\right) \mp \frac{1}{3}\sqrt{4\left(2^{N+1}+1\right)^2 - \frac{27}{7}\cdot\frac{8^{N+1}-1}{2^{N+1}-1}}
\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty}\left(2^{N+1}-1\right)^{-1}k^{(1)}(0) &= \lim_{n\to\infty}\left[\frac{2}{3} \mp \frac{1}{3}\sqrt{4\left(\frac{2^{N+1}+1}{2^{N+1}-1}\right)^2 - \frac{27}{7}\cdot\frac{8^{N+1}-1}{\left(2^{N+1}-1\right)^3}}\right] \\
&= \frac{2}{3} \mp \frac{1}{3}\sqrt{4\cdot 1^2 - \frac{27}{7}\cdot 1} \\
&= \frac{2}{3} \mp \frac{1}{3}\sqrt{7^{-1}} \\
&= \frac{2 \mp 7^{-\frac{1}{2}}}{3}
\end{aligned}$$

$$\sum_{i=b_0}^{B} w_i^2 = (B-b_0+1)^{-1} + \frac{\left(\sum_{i=b_0}^{B} k_i^{(1)}(0) \pm \sqrt{4\left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2 - 3(B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2}\right)^2}{9(B-b_0+1)\left((B-b_0+1)\sum_{i=b_0}^{B} k_i^{(1)}(0)^2 - \left(\sum_{i=b_0}^{B} k_i^{(1)}(0)\right)^2\right)}$$

$$= \left(2^{N+1}-1\right)^{-1} + \frac{\left(\left(4^{N+1}-1\right) \pm \sqrt{4\left(4^{N+1}-1\right)^2 - 3\left(2^{N+1}-1\right)\cdot\frac{9}{7}\left(8^{N+1}-1\right)}\right)^2}{9\left(2^{N+1}-1\right)\left(\left(2^{N+1}-1\right)\cdot\frac{9}{7}\left(8^{N+1}-1\right) - \left(4^{N+1}-1\right)^2\right)}$$

$$= \left(2^{N+1}-1\right)^{-1} + \frac{\left(\left(4^{N+1}-1\right) \pm \sqrt{4\left(4^{N+1}-1\right)^2 - \frac{27}{7}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right)}\right)^2}{9\left(2^{N+1}-1\right)\left(\frac{9}{7}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right) - \left(4^{N+1}-1\right)^2\right)}$$

$$= \left(2^{N+1}-1\right)^{-1} + \frac{\left(1 \pm \sqrt{4 - \frac{27}{7}\left(4^{N+1}-1\right)^{-2}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right)}\right)^2}{9\left(2^{N+1}-1\right)\left(\frac{9}{7}\left(4^{N+1}-1\right)^{-2}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right) - 1\right)}$$

$$\lim_{N\to\infty}\left(2^{N+1}-1\right)\sum_{i=b_0}^{B} w_i^2 = \lim_{N\to\infty}\left[1 + \frac{\left(1 \pm \sqrt{4 - \frac{27}{7}\left(4^{N+1}-1\right)^{-2}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right)}\right)^2}{9\left(\frac{9}{7}\left(4^{N+1}-1\right)^{-2}\left(2^{N+1}-1\right)\left(8^{N+1}-1\right) - 1\right)}\right]$$

$$= 1 + \frac{\left(1 \pm \sqrt{4 - \frac{27}{7}\cdot 1}\right)^2}{9\left(\frac{9}{7}\cdot 1 - 1\right)} = 1 + \frac{\left(1 \pm 7^{-\frac{1}{2}}\right)^2}{9\left(2\cdot 7^{-1}\right)} = 1 + \frac{\left(7^{\frac{1}{2}} \pm 1\right)^2}{18}$$

Then,

$$\lim_{N\to\infty} I_q\left[k_{w,s}\right] = \left(\frac{2 \mp 7^{-\frac{1}{2}}}{3}\right)\left(1 + \frac{\left(7^{\frac{1}{2}} \pm 1\right)^2}{18}\right)$$

$$\approx .94, .91$$