To appear in the Journal of Nonparametric Statistics Vol. 00, No. 00, Month 20XX, 1–28

Bias Reduction by Transformed Flat-top Fourier Series Estimator of Density on Compact Support

Liang Wang^{*} and Dimitris N. Politis University of California at San Diego

The problem of nonparametric estimation of a univariate density with rth continuous derivative on compact support is addressed ($r \geq 2$). If the density function has compact support and is non-zero at either boundary, regular kernel estimator will be completely biased at such boundary. Although several correction methods were proposed to improve the bias at the boundary to h^2 in the last decades, this paper initiates a way to further improve bias to higher order (h^r) for interior area of density function support, while remaining the order of bias h^2 at boundary. We will first review flat-top kernel estimator and flat-top series estimator, then propose the Transformed Flat-top Series estimator. The theoretical analysis is supplemented with simulation results as well as real data applications.

Keywords: Nonparametric Density Estimation; Gibbs Phenomenon; Higher Order of Bias; Hypothesis Testing.

AMS Subject Classification: 62G07; 62F03.

1. Introduction

Suppose the observed sample X_1, \ldots, X_n is independent, identically distributed (i.i.d.) having univariate smooth density distribution $f_X(x)$ with compact support; without loss of generality, let $f_X(x)$ has support [0,1] and rth continuous derivative $(r \ge 2)$. The independence assumption is just for ease of presentation. All results in the paper at hand hold true if the sample X_1, \ldots, X_n is only assumed to be strictly stationary obeying some weak dependence assumptions; see Paparoditis and Politis (2016a) and the references therein.

The standard kernel density estimator is

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \tag{1}$$

where $K(\cdot)$ is a kernel function that usually is symmetric and satisfies $\int K(x)dx = 1$; here h > 0 denotes the bandwidth which is assumed to tend to zero as $n \to \infty$. It is well known that such density kernel estimators suffer from boundary effects.

Numerous methods have been developed in the literature to reduce the boundary bias problem. Schuster (1985a), Silverman (1986), Cline and Hart (1991) considered the

^{*}Corresponding author. Email: wangliang6111@gmail.com

reflection method, which will reduce the boundary bias to $O(h^2)$ when the derivative of density at boundary is zero, otherwise the boundary effect is still slight improved with bias O(h). The boundary kernel method that uses a nonsymmetric kernel function at boundary points was developed by Gasser and Muller (1979), Gasser, Muller, and Mammitzsch (1985), Jones (1993), Muller (1991), Zhang and Karunamuni (2000). The local linear method is more general and fast implemented without many assumptions on; it was developed by Cheng, Fan, and Marron (1997), Cheng (1997), Zhang and Karunamuni (1998). The pseudo-data method, proposed by Cowling and Hall (1996), generates pseudo-data by linear interpolation of order statistics, and estimates the density function by using original data together with pseudo-data.

The transformation method, which is employed in the paper at hand, was first proposed by Wand, Marron, and Ruppert (1991), and then further developed by Wen and Wu (2015). Hall and Park (2002) presented the "empirical translation correction" method. The beta kernel estimator that uses the beta density as kernel function was proposed by Chen (1999). Jones and Henderson (2007) discussed a Gaussian copula estimator. Zhang, Karunamuni, and Jones (1999) combined the pseudo-data, transformation and reflection methods. Marron and Ruppert (1994) combined transformation and reflection methods.

In this paper, we propose to combine the above mentioned transformation method and the flat-top kernel series method of Politis (2001), which eventually improves the bias at interior area of density function support to h^r , while remaining bias at boundary h^2 . To give some background on the flat-top kernels: Politis and Romano (1999, 1993) proposed flat-top kernel probability density estimators in the univariate and multivariate case respectively. Later, Politis (2001) proposed flat-top series estimator of a density with compact support while Politis (2003) introduced an adaptive bandwidth choice method for spectral density and probability density functions estimated with flat-top kernels. More details of flat-top kernel estimator and flat-top series estimator are provided in Section 2.

The remainder of the paper is organized as follows. Section 2 gives some background on the transformation method, the flat-top kernel and flat-top series estimators, along with some preliminary results. Section 3 presents the Transformed Rectangular Flat-top Series estimator. Section 4 presents the Transformed Infinitely Differentiable Flat-top Series estimator, and shows the higher order bias in the interior region. The implementation and selection of parameters are discussed in Sections 5 and 6. Section 7 conducts simulation study and Section 8 gives the application to three financial returns time series. Concluding comments are given in section 9; all technical proofs are in the Appendix.

2. Background on the Transformation Method and Flat-top Estimators

2.1. Review of the Transformation Method

Before starting to introduce the spirit of the transformation method, we first briefly review why regular kernel density estimator misbehaves with a density function on compact support. Equation (1) and a Taylor expansion of $\hat{f}_X(x)$ with respect to x near the left boundary gives

$$\mathbf{E}\left[\hat{f}_X(x)\right] = \mu_0(C)f_X(x) - h\mu_1(C)f_X^{(1)}(x) + \frac{h^2}{2}\mu_x(C)f_X^{(2)}(x) + o(h^2);$$

for the proof, see Appendix A of Marron and Ruppert (1994). In the above, x = Ch for some $C \in [0,1], \mu_k(C) = \int_{-1}^C u^k K(u) du$, and $f_X^{(j)}$ is the *j*th derivative of f_X . Since $\mu_0(0) = \frac{1}{2}, \mu_0(1) = 1$ and $\mu_0(C) < 1$ for $C \in [0,1)$, the kernel estimator at zero has bias that does not vanish asymptotically.

To correct this problem, Schuster (1985b) proposed the reflection method which is implemented as:

$$\tilde{f}_X(x) = \begin{cases} \hat{f}_X(x) + \hat{f}_X(-x), & \text{if } x \in [0, h) \\ \hat{f}_X(x), & \text{if } x \in [h, 1-h] \\ \hat{f}_X(x) + \hat{f}_X(2-x), & \text{if } x \in (1-h, 1] \\ 0, & \text{if } x \notin [0, 1]. \end{cases}$$

Then under the same assumptions,

$$\mathbb{E}\left[\tilde{f}_X(x)\right] = f_X(x) + \frac{h^2}{2}\mu_2(1)f_X^{(2)}(x) - 2h\{C\mu_0(-C) + \mu_1(-C)\}f_X^{(1)}(x) + 2h^2\{C^2\mu_0(-C) + C\mu_1(-C)\}f_X^{(2)}(x) + o(h^2)$$

where x = Ch for $C \in [0, 1]$. If $f_X^{(1)}(0) \neq 0$, the bias above has order of h which tends to zero as $n \to \infty$. Furthermore, if $f_X^{(1)}(0) = 0$, the bias will have order of h^2 .

Marron and Ruppert (1994) proposed to construct a new random variable Y = g(X)with probability density f_Y whose first derivative satisfying $f'_Y(0) = 0$. Then, they estimated f_Y near the left boundary via a kernel with the reflection method obtaining a boundary bias of order $O(h^2)$ which can then deliver this property back to estimating f_X . To elaborate, denote the transform function by $g(\cdot)$, a fixed one-to-one and monotonically increasing function; suppose random variable X follows the density $f_X(x)$ with support [0, 1], let Y = g(X), and denote $f_Y(y)$ the density function of Y. Then,

$$f_X(x) = f_Y[g(x)]g'(x)$$

Denote f_Y an estimator of f_Y based on data $\{Y_i = g(X_i)\}_{i=1}^n$; then the transformed estimator employed by Marron and Ruppert (1994) was:

$$\hat{f}_X(x;g) = \hat{f}_Y[g(x)]g'(x),$$
(2)

Assume g(x) is smooth, and that $\sup_{x \in [0,1]} |g'(x)| \leq M$ for some positive value M; then,

$$\left| \operatorname{E}[\hat{f}_X(x;g)] - f_X(x) \right| = \left| \operatorname{E}\left[\hat{f}_Y[g(x)]g'(x) \right] - f_Y[g(x)]g'(x) \right| \\ \leq M \cdot \left| \operatorname{E}\left(\hat{f}_Y[g(x)] \right) - f_Y[g(x)] \right|$$
(3)

Similarly, we also have

$$\operatorname{Var}\left[\hat{f}_X(x;g)\right] \le M^2 \cdot \operatorname{Var}\left[\hat{f}_Y[g(x)]\right] \tag{4}$$

The methodology to be proposed in our paper also uses the transformation idea Y = g(X) with $g(\cdot)$ satisfying some specific conditions but uses the flat-top series estimator of Politis

(2001) instead of the reflection method; more details will be given in Section 3.

2.2. Review of Flat-top Kernel Density Estimator

This subsection reverts to considering the problem where the random variable X has density $f_X(x)$ on the whole of \mathbb{R} . For this problem, higher order kernel density estimator and infinite order kernel estimator were designed to improve the order of bias. Recall that a kernel function $K(\cdot)$ is of order q if $K(\cdot)$ has finite moments up to qth order, and moments of order up to q-1 equal to zero, i.e., $\int x^q K(x) dx \neq 0$ but $\int x^k K(x) dx = 0$ for $\forall k \leq q-1$. Now suppose a density function $f_X(x)$ has r continuous and bounded derivatives, then it is well-known that

Bias
$$\left[\hat{f}_X(x)\right] = \mathrm{E}\hat{f}_X(x) - f_X(x) = c_{f,K}(x)h^k + o(h^k)$$

where $k = \min(q, r)$ and $c_{f,K}(x)$ is a bounded function depending on $K(\cdot)$, f_X and its derivatives. Standard kernel functions are nonnegative and symmetric around zero, and consequently the bias is of order $O(h^2)$ provided $r \ge 2$. In order to get a better bias order of $O(h^k)$ with $k \ge 2$, one must use a kernel of higher order q > 2; this idea dates back to Bartlett (1963) and Parzen (1962). More references on high-order kernels include Silverman (1986); Gasser et al. (1985); Cacoullos (1966); Devroye (1987); Granovsky and Muller (1991); Jones (1995); Jones and Foster (1993); Marron (1994); Marron and Wand (1992); Nadaraya (1989).

The question arises: what order kernel to use to get full advantage of possible smoothness of f_X and obtain maximum bias reduction? One needs to choose order $q \ge r$ but the degree of smoothness r is unknown. To achieve $q \ge r$ for any potential value of r, it is natural to choose $q = \infty$. Indeed, the aforementioned flat-top kernels have infinite order and achieve a bias of order $O(h^r)$ no matter how large is r. Furthermore, Politis (2003) proposed a simple bandwidth selection methods for flat-top kernel estimators that is easier than the bandwidth selection procedures associated with finite-order kernels. In the following definition, we denote the flat-top kernel function as $\Omega(\cdot)$ instead of $K(\cdot)$.

DEFINITION 2.1 Let c be a positive value; the kernel Ω_c is said to be a member of the general family of univariate flat-top kernels of infinite order if

$$\Omega_c(x) = \frac{1}{2\pi} \int \omega_c(s) e^{-isx} ds$$

where the Fourier transform $\omega_c(s)$ satisfies the following properties:

(i) $\omega_c(s) = 1$ for some c > 0, for all $|s| \le c$; (ii) $\int |\omega_c(s)|^2 ds < \infty$; (iii) $\omega_c(s) = \omega_c(-s)$ for any $s \in \mathbb{R}$.

The above three properties will guarantee the infinite order, a finite variance of the resulting estimator, and that $\Omega_c(x)$ is real valued. In this paper, c is a positive value ≤ 1 , and we will assume $\omega_c(s)$ has compact support, i.e.

$$\omega_c(s) = \begin{cases} 1, & |s| \le c \\ \eta_\omega(s), & c < |s| \le 1 \\ 0, & |s| > 1 \end{cases}$$
(5)

where the function $\eta_{\omega}(s)$ determines the shape of $\omega_c(s)$ outside the flat regions. Like regular kernel estimator, with a bandwidth h, we define

$$\Lambda_c(x) = \frac{1}{h} \Omega_c\left(\frac{x}{h}\right) \text{ and } \lambda_c(s) = \int \Lambda_c(x) e^{isx} dx = \omega_c(hs)$$

Denote $\phi_X(s) = \int e^{isx} f_X(x) dx$ the characteristic function, and let $\hat{\phi}_X(s) = (1/n) \sum_{k=1}^n e^{isX_k}$ be the sample characteristic function; then the flat-top kernel estimator of $f_X(x)$ is

$$\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^n \Lambda_c(x - X_i) = \frac{1}{2\pi} \int \lambda_c(s) \hat{\phi}_X(s) e^{-isx} ds$$
(6)

THEOREM 2.2 (Politis and Romano (1999)) Assume there is an r > 0, such that

$$\int |s|^r |\phi_X(s)| ds < \infty \tag{7}$$

Assume that $n \to \infty$, and let $h \sim An^{-1/(2r+1)}$, for some constant A > 0; it follows that

$$\sup_{x \in \mathbb{R}} Bias\left[\hat{f}_X(x)\right] = o(h^r)$$

and

$$\sup_{x \in \mathbb{R}} \mathrm{MSE}\left[\hat{f}_X(x)\right] = O\left(n^{-\frac{2r}{2r+1}}\right)$$

where MSE is short for Mean Squared Error.

Note that condition (7) implies that $f_X(x)$ has r bounded and continuous derivatives.

REMARK 2.3 For the scenario f_X is infinitely differentiable function, although it is less practical, one can show Theorem 2.2 still holds, for any non-negative integer r. Thus if f_X is infinitely differentiable function, with appropriate selection of bandwidth h, Flat-top Kernel Estimator will achieve very fast convergence rate. Selection of bandwidth h will be discussed in Section 6.2.

2.3. Flat-top Series Estimator of Density on Compact Support

We now return to the problem where $f_X(x)$ is a very smooth density function on [0, 1]. Denote $\tilde{f}_X(x)$ the periodic extension of $f_X(x)$ on the whole real line; apparently $\tilde{f}_X(x)$ has period 1. Denote the characteristic function $\phi_X(s) = \int_0^1 e^{isx} f_X(x) dx$, and the sample characteristic function $\hat{\phi}_X(s) = (1/n) \sum_{k=1}^n e^{isX_k}$. Now expand $\tilde{f}_X(x)$ in a Fourier series:

$$\tilde{f}_X(x) = \sum_{s \in \mathbb{Z}} e^{-i2\pi s x} \phi_X(s).$$

Analogously to (6)—but replacing Fourier transform by Fourier series—is the definition of flat-top series estimator of Politis (2001), i.e., let

$$\hat{f}_X(x) = \begin{cases} \sum_{s \in \mathbb{Z}} \lambda_c(s) e^{-i2\pi s x} \hat{\phi}_X(s), & 0 \le x \le 1\\ 0, & x < 0 \text{ or } x > 1 \end{cases}$$
(8)

Apparently $\sum_{s \in \mathbb{Z}} \lambda_c(s) e^{-i2\pi sx} \hat{\phi}_X(s)$ is also a periodic function with period 1 on the whole real line but we only need the part on the interval [0, 1]. First, recall the following theorem:

THEOREM 2.4 (Politis (2001)) If there exists r > 0, such that

$$\sum_{s \in \mathbb{Z}} |s|^r \phi_X(s) < \infty \tag{9}$$

letting $h \sim An^{-1/(2r+1)}$ for some constant A > 0, it follows that

$$\sup_{x \in [0,1]} Bias\left[\hat{f}_X(x)\right] = o(h^r)$$

and

$$\sup_{x \in [0,1]} \text{MSE}\left[\hat{f}_X(x)\right] = O\left(n^{-\frac{2r}{2r+1}}\right)$$
(10)

Notice that condition (9) implies f_X has r bounded and continuous derivatives; in particular, it implies that f_X has r bounded and continuous derivatives that satisfy "wrap-around" continuity, i.e.,

$$f_X(0) = f_X(1), f_X^{(1)}(0) = f_X^{(1)}(1), f_X^{(2)}(0) = f_X^{(2)}(1), \dots, f_X^{(r)}(0) = f_X^{(r)}(1).$$

REMARK 2.5 Similar to Remark 2.3, for the scenario \tilde{f}_X is infinitely differentiable function, one can show Theorem 2.4 still holds, for any non-negative integer r.

In next theorem, we relax (9) it in the following new theorem:

THEOREM 2.6 Let $f_X(x)$ be a differentiable density function defined on [0, 1], and denote $f_X^{(1)}$ its first derivative. Assume $0 < |f_X(1) - f_X(0)| < \infty$, and $f_X^{(1)} \in L^1[0, 1]$, i.e. $\int_0^1 |f_X^{(1)}(x)| dx < \infty$. Then

$$\sup_{x \in [0,1]} \operatorname{Var}\left[\hat{f}_X(x)\right] = O\left[\frac{1}{nh}\log\left(\frac{1}{h}\right)\right]$$
(11)

From the proof of Theorem 2.6 given in the appendix we can also obtain the following Corollary.

COROLLARY 2.7 Let $f_X(x)$ be a twice differentiable density function defined on [0, 1]. Assume $|f_X(1) - f_X(0)| = 0$, $0 < |f_X^{(1)}(1) - f_X^{(1)}(0)| < \infty$ and both of $f_X^{(1)}, f_X^{(2)} \in L^1[0, 1]$. Then

$$\sup_{x \in [0,1]} \operatorname{Var}\left[\hat{f}_X(x)\right] = O\left(\frac{1}{nh}\right)$$
(12)

Lastly, we present the following new theorem under a weaker condition compared to (9) in Theorem 2.4.

THEOREM 2.8 Let $f_X(x)$ denote the density function on [0,1] and $\tilde{f}_X(x)$ be the periodic extension of $f_X(x)$. If $\tilde{f}_X(x)$ has (r-1)th bounded and continuous derivative (for some $r \ge 1$), and $\tilde{f}_X^{(r)}(x)$ is continuous except for jump discontinuities when $x \in \mathbb{Z}$. More specifically, $f_X^{(r)}(x)$ is continuous on [0,1], but $0 < |f_X^{(r)}(0) - f_X^{(r)}(1)| < \infty$. Moreover, assume $f_X^{(r+1)}(x)$ is continuous on [0,1]. Then for $h \sim An^{-1/(2r+1)}$, we still have result (10) and the bias satisfies

$$\sup_{x \in [0,1]} Bias\left[\hat{f}_X(x)\right] = O(h^r)$$

The theorem's proof is provided in the appendix for a special case of $\omega_c(s)$, i.e., $\omega_c(s)$ with c = 1 and $\eta_{\omega} = 0$, and we also call it a rectangular flat-top function; the general case is similar albeit more involved technically.

3. Transformed Rectangular Flat-top Series Estimator

Although Theorem 2.4 and Theorem 2.8 guarantee good properties of flat-top series estimator, they also require very strict conditions. In practical use, the periodic extension \tilde{f}_X of density f_X rarely has up to rth bounded and continuous derivative. The smoothness of $f_X(x)$ on (0, 1) is not issue; the issue is with the boundary values. For example, there is no reason to assume that $f_X(0) = f_X(1)$, and the same is true for the derivatives of f_X . In other words, it is not realistic to assume the periodic extension \tilde{f}_X is smooth or even continuous in which case the flat-top series estimator fails. However, the trasformation method can be used to bring us to a realm where the boundary continuity conditions are satisfied, and hence the flat-top series estimator can be successfully used.

Consequently, our transform function $g(\cdot)$ has the particular target to render

$$f_Y(0) = f_Y(1)$$

where Y = g(X), and $f_Y(\cdot)$ is true density function of Y. Nevertheless, the proof of Theorem 2.8 in appendix shows that if $f_Y^{(1)}(0) \neq f_Y^{(1)}(1)$, then the bias of flat-top series estimator has order at most O(h) universally. Thus, we need to modify our "target" to be:

$$f_Y(0) = f_Y(1); \quad f_Y^{(1)}(0) = f_Y^{(1)}(1)$$
 (13)

in order to achieve order of bias $O(h^2)$.

REMARK 3.1 "Target" (13) is only designed for rectangular flat-top series estimator. In the next section of infinitely differentiable flat-top estimator, we will just need the basic target $f_Y(0) = f_Y(1)$. In what follows, we will pick $g(\cdot)$ from some parametric family of functions. Let $\Theta \subset \mathbb{R}^k$ and $G = \{g_\theta(\cdot) \text{ for } \theta \in \Theta\}$ be a parametric family of transformations where each $g_\theta(x)$ is defined for $x \in [0, 1]$. We will need the following condition throughout:

CONDITION 3.2 Each $g_{\theta}(x) \in G$ is three times continuously differentiable with respect to x and has a strictly positive derivative from [0,1] to [0,1]. Derivatives at 0 and 1 are one sided. In addition there exists a constant K > 0, s.t.

$$\max_{j=0,1,2,3} \{ \sup_{x \in [0,1]} |g_{\theta}^{(j)}(x) - g_{\theta_0}^{(j)}(x)| \} \le K ||\theta - \theta_0||$$
(14)

where $g_{\theta}^{(j)}(x)$ is the *j*th derivative of g_{θ} are with respect to x, and $|| \cdot ||$ denotes the Euclidean norm on \mathbb{R}^k .

One example where condition (3.2) is satisfied is when g_{θ} is polynomial with θ being the vector of coefficients.

Eq. (2) still applies but since g is no longer fixed, we need better notation. Here and for the remainder of the paper, let $f_Y(\cdot; g)$ denote the density function of Y = g(X), let $\hat{f}_Y(\cdot; g)$ denote its flat-top series estimator, let $\phi_Y(\cdot; g)$ denote its characteristic function, and let $\hat{\phi}_Y(\cdot; g)$ denote the estimated characteristic function. We then formally define the Transformed Flat-top Series estimator as

$$\hat{f}_X(x;g) = \hat{f}_Y[g(x);g]g'(x)$$
(15)

for some $g \in G$. Now assume $Y_{\theta_0} = g_{\theta_0}(X)$ satisfies (13); we shall be concerned with the case where θ_n is an estimator of θ_0 based on the sample $\{X_i\}_{i=1}^n$. Let $Y_{\theta_n} = g_{\theta_n}(X)$, the actual transformed flat-top series estimator—using a data-based estimator for g—then is

$$\begin{aligned} \hat{f}_X(x;g_{\theta_n}) &= \hat{f}_{Y_{\theta_n}}[g_{\theta_n}(x);g_{\theta_n}]g'_{\theta_n}(x) \\ &= \hat{f}_Y[g(x);g]\Big|_{g=g_{\theta_n}}g'_{\theta_n}(x) \end{aligned}$$
(16)

where

$$\hat{f}_{Y_{\theta_n}}[g_{\theta_n}(x);g_{\theta_n}] = \sum_{s \in \mathbb{Z}} \lambda_c(s) e^{-i2\pi s g_{\theta_n}(x)} \hat{\phi}_{Y_{\theta_n}}(s;g_{\theta_n}).$$
(17)

Using the same trick as in Marron and Ruppert (1994), we define the expectation of flat-top series estimator of $f_{Y_{\theta_n}}$ given g_{θ_n} as:

which is a sort of modified partial sum of Fourier series of the periodic extension $\tilde{f}_{Y_{\theta_n}}(y; g_{\theta_n})$ of $f_{Y_{\theta_n}}(y; g_{\theta_n})$. Note $\phi_{Y_{\theta_n}}(\cdot; g_{\theta_n})$ is the characteristic function of Y_{θ_n} .

REMARK 3.3 The above definition side-steps the randomness inherent in g_{θ_n} in order to develop tractable expressions for bias and variance. The reason that we can safely ignore the randomness inherent in g_{θ_n} is that g_{θ_n} is estimated at a fast rate. In fact, in Section 5 we will use simple boundary correction method by Jones (1993) to estimate g_{θ_n} of which mean squared error with order $O_p(\sqrt{(nh)^{-1} + h^2})$, smaller (and therefore negligible) compared to $\Delta_n = \frac{1}{\log(1/h)}$, which is the rate of convergence given in the following novel result.

THEOREM 3.4 For the problem f_X is r times continuously differentiable $(r \ge 2)$ on [0,1] (and therefore also bounded) but $f_X(0) \ne f_X(1)$, suppose $\theta_0 \in \Theta$, and Condition 3.2 holds. Define $G_{\triangle_n} = \{g_{\theta_n} : ||\theta_n - \theta_0|| \le \triangle_n \text{ and } \theta_n \in \Theta\}$. Let $h = h(n) \to 0$ as $n \to \infty$ for some function $h(\cdot)$, with large enough n so that h < 1, and let $\triangle_n = \frac{1}{\log(1/h)}$. Then for any $g_{\theta_n} \in G_{\triangle_n}$

$$\sup_{y \in [0,1]} \left| \hat{f}_{Y_{\theta_n}}(y; g_{\theta_n}) - \mathcal{E}\hat{f}_{Y_{\theta_n}}(y; g_{\theta_n}) \right| = O_p\left(\frac{1}{\sqrt{nh}}\right)$$
(19)

Note that if $||\theta_n - \theta_0|| = o_p(\Delta_n)$, then $P(g_{\theta_n} \in G_{\Delta_n}) \to 1$, and the above result allows for g_{θ_n} being random. As mentioned in Remark 3.3, $||\theta_n - \theta_0|| = O_p(\sqrt{(nh)^{-1} + h^2})$ which is indeed $o_p(\Delta_n)$ since typically we let $h = n^{-\beta}$ for some $\beta \in (0, 1)$.

COROLLARY 3.5 If
$$||\theta_n - \theta_0|| = O_p\left(\sqrt{(nh)^{-1} + h^2}\right)$$
, then $P(g_{\theta_n} \in G_{\Delta_n}) \to 1$, and

$$1 \ge P\left[\sup_{y \in [0,1]} \left| \hat{f}_{Y_{\theta_n}}(y; g_{\theta_n}) - \mathcal{E}\hat{f}_{Y_{\theta_n}}(y; g_{\theta_n}) \right| = O\left(\frac{1}{\sqrt{nh}}\right) \right] \ge P(g_{\theta_n} \in G_{\triangle_n}) \to 1$$

In other words,

$$\sup_{y \in [0,1]} \left| \hat{f}_{Y_{\theta_n}}(y; g_{\theta_n}) - \mathcal{E}\hat{f}_{Y_{\theta_n}}(y; g_{\theta_n}) \right| = O_p\left(\frac{1}{\sqrt{nh}}\right)$$
(20)

With equation (15) and Corollary (3.5) it is easy to show variance of Transformationbased Rectangular Flat-top Series Estimator is in order of $O_p\left(\frac{1}{\sqrt{nh}}\right)$ (see proof of Theorem 3.14. The bias of the proposed estimator, will be studied in the rest of this section. Similar to variance, we start from the difference between $\mathrm{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n})$ and $f_{Y_{\theta_n}}(y;g_{\theta_n})$. $\mathrm{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n})$ has been defined in Equation (18), by the Definition 2.1 and the definition of $\omega_c(s)$ in equation (5), we can treat $\mathrm{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n})$ as a partial sum of Fourier series with coefficient $\lambda_c(s)$ at different points s.

REMARK 3.6 If $\theta_n \to \theta_0$ at rate q(n) where $q(n) = n^{-\gamma}$ for some $\gamma > 0$, then Y_{θ_n} does not really satisfy (13). By Condition 3.2, it is obvious that

$$|f_{Y_{\theta_n}}(1;g_{\theta_n}) - f_{Y_{\theta_n}}(0;g_{\theta_n})| = O[q(n)], \quad |f'_{Y_{\theta_n}}(1;g_{\theta_n}) - f'_{Y_{\theta_n}}(0;g_{\theta_n})| = O[q(n)]$$
(21)

Similarly, if $\theta_n \to \theta_0$ in probability at rate q(n), then

$$|f_{Y_{\theta_n}}(1;g_{\theta_n}) - f_{Y_{\theta_n}}(0;g_{\theta_n})| = O_p[q(n)], \quad |f'_{Y_{\theta_n}}(1;g_{\theta_n}) - f'_{Y_{\theta_n}}(0;g_{\theta_n})| = O_p[q(n)] \quad (22)$$

Eq. (21) and (22) imply that Y_{θ_n} fails to satisfy (13). Furthermore, it means the periodic extension $\tilde{f}_{Y_{\theta_n}}(y; g_{\theta_n})$ of $f_{Y_{\theta_n}}(y; g_{\theta_n})$ is not continuous at $y = \ldots, -2, -1, 0, 1, 2, \ldots$

As discussed earlier, $\mathrm{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n})$ is a partial sum of Fourier series of $\tilde{f}_{Y_{\theta_n}}(y;g_{\theta_n})$, which is essentially piecewise continuously differentiable. It is well known the partial sum of Fourier series has oscillations near the discontinuity point, and this peculiar manner is called Gibbs Phenomenon. Fortunately the oscillation between the true function and partial sum of Fourier series is bounded by the size of the jump of the discontinuity at discontinuous points (more details can be found in Lemma 3.10). In our case it means the oscillation between $\mathrm{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n})$ and $f_{Y_{\theta_n}}(y;g_{\theta_n})$ at boundaries 0 and 1, are also bounded by $|f_{Y_{\theta_n}}(1;g_{\theta_n}) - f_{Y_{\theta_n}}(0;g_{\theta_n})| = q(n)$. This statement implies flat-top series estimator $\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n})$ always has bias at the boundaries of order q(n), which is identical to the convergence rate of $\theta_n \to \theta_0$. In this section, the Transformation-based Rectangular Flat-top Series Estimator can retain the same bias of order h^2 in the interior of support. This bias can be improved to higher order by using infinitely differentiable flat-top series estimator introduced in the next subsection.

Now recall Definition (5) of $\omega_c(\cdot)$; when c = 1, $\omega_c(\cdot)$ is the rectangular flat-top function that will be denoted simply as

$$\omega(s) = \begin{cases} 1, & |s| \le 1\\ 0, & |s| > 1. \end{cases}$$

Recall equation (18) and $\lambda_c(s) = \omega_c(hs)$; it follows that

$$\mathbf{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n}) = \sum_{|s| \le 1/h} \phi_{Y_{\theta_n}}(s)e^{i2\pi sy}.$$
(23)

We now define

$$S_{N_h}\tilde{f}_{Y_{\theta_n}}(y;g_{\theta_n}) := \sum_{|s| \le 1/h} \phi_{Y_{\theta_n}}(s) e^{i2\pi sy} = \mathbf{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n})$$

i.e., $S_{N_h} \tilde{f}_{Y_{\theta_n}}(y; g_{\theta_n})$ denotes the partial sum of Fourier series, where $N_h = \lfloor 1/h \rfloor$. $\tilde{f}_{Y_{\theta_n}}(y; g_{\theta_n})$ is the periodic extension of $f_{Y_{\theta_n}}(y; g_{\theta_n})$. The following lemmas are showing the bias at boundary points:

LEMMA 3.7 Define

$$\phi_0(x) = \frac{1}{2} \left(\frac{1}{2} - x\right), 0 \le x \le 1$$

where $\phi(x)$ is periodic extension function of $\phi_0(x)$. Let $\phi_n(x)$ is the nth partial sum of Fourier series:

$$\phi_n(x) = \sum_{k=1}^n \frac{\sin(2k\pi x)}{2k\pi}$$

For any $x \notin \mathbb{Z}$, i.e. x is away from the discontinuity of $\phi(x)$, we have

$$|\phi(x) - \phi_n(x)| = O\left(\frac{1}{n}\right)$$

Notice $\phi(\cdot)$, $\phi_0(\cdot)$ and $\phi_n(\cdot)$ are notations only applied to Lemma 3.7.

REMARK 3.8 Notice that if $x \in [a, b]$ where [a, b] is an interval away from discontinuity, then we can also have:

$$\sup_{x \in [a,b]} |\phi(x) - \phi_n(x)| = O\left(\frac{1}{n}\right)$$
(24)

REMARK 3.9 If f(x) is discontinuous at a point a, then we can define a continuous function

$$f^*(x) = f(x) - 2[f(a^+) - f(a^-)]\phi(x - a)$$

where $f(a^+) = \lim_{x \to a^+} f(x)$ is right hand limit, and $f(a^-)$ is left hand limit. Clearly $f^*(a) = \frac{f(a^+) + f(a^-)}{2}$ and $f^*(x)$ is now continuous, very helpful for showing the next lemma. The reason we construct $f^*(x)$ by $\phi(x)$ instead of by a step function is that $\phi_n(x)$ has a better form and more convenient for the proof.

LEMMA 3.10 (Theorem F, Hewitt and Hewitt (1979)) Let f be a real-valued periodic function on the real line \mathbb{R} with period 1, and suppose that f and its derivative f' are both continuous except for a finite number of finite jump discontinuities in the interval [0,1]. let $S_n f(x)$ be the n^{th} partial sum of the Fourier series of the function f, computed at the point x. Let a be a point of discontinuity of f. The distance between overshoot and undershoot of Gibbs Phenomenon at point a is at most $\frac{2}{\pi}Si(\pi)|f(a^+) - f(a^-)|$, where $Si(x) = \int_0^x \frac{sin(t)}{t} dt$.

REMARK 3.11 Recall (23), i.e., that the expectation of rectangular flat-top series estimator is Fourier series. Lemma 3.10 essentially tells us that, for any $0 < a \le b < 1$:

$$\sup_{y \in [0,a) \bigcup (b,1]} \left| \mathbb{E}\hat{f}_{Y_{\theta_n}}(y;g_{\theta_n}) - f_{Y_{\theta_n}}(y;g_{\theta_n}) \right| \le \frac{2}{\pi} Si(\pi) |\tilde{f}_{Y_{\theta_n}}(0^+;g_{\theta_n}) - \tilde{f}_{Y_{\theta_n}}(0^-;g_{\theta_n})|$$
$$= \frac{2}{\pi} Si(\pi) |f_{Y_{\theta_n}}(1;g_{\theta_n}) - f_{Y_{\theta_n}}(0;g_{\theta_n})|$$

which, together with (22) yields the following theorem:

THEOREM 3.12 If $f_X(x)$ is continuously differentiable function on [0,1], and $\theta_n - \theta_0 = O_p(h^i), i = 1 \text{ or } 2$, then for any $0 < a \le b < 1$:

$$\sup_{y \in [0,a) \bigcup (b,1]} \left| E \hat{f}_{Y_{\theta_n}}(y; g_{\theta_n}) - f_{Y_{\theta_n}}(y; g_{\theta_n}) \right| = O_p(h^i), \quad i = 1 \text{ or } 2.$$

LEMMA 3.13 Let f be a real valued function on [0,1] and \tilde{f} is periodic extension of f. If $\tilde{f}(x)$ is continuous (i.e. f(0) = f(1)), $\tilde{f}^{(1)}(x)$ is continuous except jump discontinuities at $x \in \mathbb{Z}$, in another word, $f^{(1)}(x)$ is continuous on [0,1], but $0 < |f^{(1)}(0) - f^{(1)}(1)| < \infty$.

а

Moreover, $f^{(2)}$ and $f^{(3)}$ is continuous on [0, 1]. Then

$$\sup_{x \in [0,1]} |S_n f(x) - f(x)| \le \frac{1}{n} |f^{(1)}(1) - f^{(1)}(0)| + O\left(\frac{1}{n^2}\right)$$

Notice all the above lemmas and theorems are discussing the rectangular flat-top series estimator for transformed random variable Y_{θ_n} . Recall (16), the following Theorem takes us back to random variable X:

THEOREM 3.14 Assume Condition 3.2 and $h = An^{-1/5}$. If $\theta_n = \theta_0 + O_p(h)$, then for any $0 < a \le b < 1$, the Transformed Rectangular Flat-top Series estimator converges to true density function:

$$\sup_{x \in [a,b]} \left| \hat{f}_X(x;g_{\theta_n}) - f_X(x) \right| = O_p\left(\frac{1}{\sqrt{nh}} + h^2\right) = O_p\left(n^{-2/5}\right)$$

Although Theorem 3.14 only requires $\theta_n = \theta_0 + O_p(h)$, the bias reaches $O(h^2)$ only for a closed interval inside [0, 1]. Looking back at Theorem 3.12, the bias order at boundary is still O(h). To get bias order $O(h^2)$ at boundary, we need

COROLLARY 3.15 Assume Condition 3.2, and if $\theta_n = \theta_0 + O_p(h^2)$ and $h = An^{-1/5}$, then

$$\sup_{x \in [0,1]} \left| \hat{f}_X(x; g_{\theta_n}) - f_X(x) \right| = O_p\left(\frac{1}{\sqrt{nh}} + h^2\right) = O_p\left(n^{-2/5}\right)$$

4. Transformed Infinitely Differentiable Flat-top Estimator

Transformation-based infinitely differentiable flat-top estimator introduced here is our ultimate estimator, as mentioned in Section 1, it is the one improves order of bias at interior area of density function support to higher h^r . Recall (5), i.e., $\lambda_c(s) = \omega_c(hs)$ with c < 1 and some function η_{ω} . Eq. (18) shows that the mean of flat-top series estimator is a modified partial sum of Fourier series. As discussed in previous section, convergence of n^{th} partial sum of Fourier series at the discontinuity point suffers from oscillation, also known as Gibbs Phenomenon. This issue does not only negatively affect convergence at the discontinuity, but also slows global convergence rate.

Now let us introduce a new infinitely differentiable $\omega_c(s)$; e.g. McMurry and Politis (2004) defined

$$\eta_{\omega}(s) = \exp\left(\frac{-b}{(|s|-1)^2} \exp\left[\frac{-b}{(|s|-c)^2}\right]\right)$$
(25)

for some b > 0 and c < |s| < 1; here, c determines the width of region where the flat-top function equals one, and b allows us to alter the shape of η_{ω} . The selection of these two parameters will be discussed in the next section. Function η_{ω} is able to connect the region where $\omega(s)$ is 1 and the region where $\omega(s)$ is 0 in a manner such that $\omega(s)$ is infinitely differentiable for all s, even including where |s| = c and |s| = 1. Recall again that the expectation of a flat-top series estimator is a modified partial sum of Fourier series, and see the following lemma: LEMMA 4.1 (Theorem 3.4, Gottlieb and Shu (1997)) Let f be piecewise continuous function, each continuous part has at least r + 1 order continuous derivative. If f only has one point of discontinuity ξ , and y is one point in [0, 1] away from ξ . Let

$$f_n^{\omega}(y) = \sum_{s \in \mathbb{Z}} \omega_c\left(\frac{s}{n}\right) c_n(f) e^{i2\pi sy}$$

where $c_n(f)$ is the coefficient of Fourier series of f. Then,

$$|f(y) - f_n^{\omega}(y)| \le \frac{C \cdot H(f)}{n^r d(y)^r} + \frac{C||f||_{L^2[0,1]}}{n^{r+1/2}}$$
(26)

where C is some constant, $d(y) = \min_{k=-1,0,1} |y - \xi + k|$, and

$$H(f) = \sum_{l=0}^{r} d(y)^{l} \left[f^{(l)}(\xi^{+}) - f^{(l)}(\xi^{-}) \right] \int_{-\infty}^{\infty} \left| G_{l}^{(r+1-l)}(\eta) \right| d\eta$$

Where $G_l(\eta) = \frac{\omega_c(\eta) - 1}{\eta^l}$.

Notice the condition y is away from ξ is actually defined by $d(y) = \min_{k=-1,0,1} |y-\xi+k|$. If $y = \xi$, the convergence by (26) fails. If $y \neq \xi$, the order of convergence is proportional to $\frac{1}{d(y)}$. In addition $\int_{-\infty}^{\infty} |G_l^{(r+1-l)}(\eta)| d\eta < \infty$ because $\omega_c(\eta) - 1 = 0$ here when $|\eta| \leq c$. Lemma 4.1 shows that convergence rate of this modified partial sum of Fourier series is $O(n^{-r})$ when y is away from discontinuity ξ .

REMARK 4.2 By imitating the proof of this Theorem in Gottlieb and Shu (1997), one can easily show the same result if f does not have discontinuity.

REMARK 4.3 One interesting property of such modified partial sum of Fourier series, by definition of H(f), is that its convergence rate $O(n^{-r})$ is not influenced by the continuity of $f^{(l)}(\cdot)$ at ξ . It implies that when practitioners are looking for transformation function, transformed random variable does not have to satisfy (13). More specifically, suppose $\xi = 0$ and discussing $\tilde{f}_{Y_{\theta_n}}(\cdot; g_{\theta_n})$, the convergent rate $O(N_h^{-r})$ of bias remains, even when $\tilde{f}_{Y_{\theta_n}}(\cdot; g_{\theta_n})$ and $\tilde{f}_{Y_{\theta_n}}^{(1)}(\cdot; g_{\theta_n})$ are discontinuous at 0 (i.e., even when $|f_{Y_{\theta_n}}^{(i)}(1; g_{\theta_n}) - f_{Y_{\theta_n}}^{(i)}(0; g_{\theta_n})| \neq 0$ where i = 0, 1.). Thus the bias of the point away from discontinuity always has order $O(N_h^{-r}) = O(h^r)$. But for the points around discontinuity, it is easy to show that Lemma 3.10 and Theorem 3.12 still apply for infinitely differentiable flat-top function, and therefore we still suggest to employ an estimator $\theta_n = \theta_0 + O_p(h^2)$ so that the bias at boundary has order $O_p(h^2)$.

By Remark 4.3, the target designed for transformation-based infinitely differentiable flat-top estimator is relaxed to

$$f_Y(0) = f_Y(1) \tag{27}$$

REMARK 4.4 The definition of H(f) implies Lemma 4.1 does not work for a density function which is infinite at boundary, or has any order derivative that is infinite at boundary. The following Theorem shows transformation-based infinitely differentiable flat-top estimator improves bias at interior area, and remains order of bias h^2 at boundary. The proof is omitted because it is the same logic as the proof for Theorem 3.14, which employed Corollary 3.5 and Lemma 3.13. The idea of Theorem 4.5 is also employing Corollary 3.5 and Lemma 3.13 for boundaries, and employing Corollary 3.5 and Lemma 4.1 for interior area.

THEOREM 4.5 Assume Condition 3.2 holds, $f_X(x)$ is the density function on [0,1] and $f_X(0) \neq f_X(1)$. Moreover, $f_X(x)$ is rth continuously differentiable and bounded function on [0,1]. If $\theta_n = \theta_0 + O_p(h^2)$, then for any $x \in (0,1)$, the Transformed Infinitely Differentiable Flat-top Series estimator satisfies:

$$\hat{f}_X(x;g_{\theta_n}) = f_X(x) + O_p\left(\frac{1}{\sqrt{nh}} + h^r\right).$$

In addition, if x = 0 or 1, then

$$\hat{f}_X(x;g_{\theta_n}) = f_X(x) + O_p\left(\frac{1}{\sqrt{nh}} + h^2\right).$$

REMARK 4.6 Notice that both rectangular flat-top estimator and infinitely differentiable flat-top estimator might produce negative density values; this is a common finite-sample issue for all higher order kernels that goes away asymptotically. In practical use, we suggest to reset all negative value to zero, and subsequently re-normalize the density estimator so that it has area one; see also Politis and Romano (1993). In Sections 7 and 8, we use this correction method for both simulated and real data.

5. Implementation particulars

5.1. Polynomial Transformation

Here and in what follows, we are more interested in the maximum possible reduction in the bias, and therefore focus on the Transformed Infinitely Differentiable Flat-top Series estimator defined in the previous section. In this section, we propose to find candidate of transformation function from the polynomial function family. By Lemma 4.1, recall the transformation target is just

$$f_Y(0) = f_Y(1).$$

Let $Y_{\theta_0} = g_{\theta_0}(X)$ be the desired transformation random variable. As before, we want to constraint g_{θ_0} to be a one-to-one function from [0,1] to [0,1], so that $X = g_{\theta_0}^{-1}(Y)$. Then by the theorem for derivative of inverse function:

$$f_{Y_{\theta_0}}(y) = \frac{f_X[g_{\theta_0}^{-1}(y)]}{g'_{\theta_0}[g_{\theta_0}^{-1}(y)]}$$
(28)

which implies

$$f_{Y_{\theta_0}}(0) = \frac{f_X(0)}{g'_{\theta_0}(0)} \quad f_{Y_{\theta_0}}(1) = \frac{f_X(1)}{g'_{\theta_0}(1)}$$

A polynomial function and its coefficients will be selected by solving all the following equations:

$$g_{\theta_0}(0) = 0$$

$$g_{\theta_0}(1) = 1$$

$$g'_{\theta_0}(x) > 0, \text{ when } x = 0 \text{ or } x = 1$$

$$g'_{\theta_0}(x) \ge 0, 0 < x < 1$$

$$f_{Y_{\theta_0}}(0) = f_{Y_{\theta_0}}(1)$$
(29)

For example, if we consider a quadratic polynomial $g_{\theta_0}(x) = ax^2 + bx$, then $g_{\theta_0}(0) = 0$ is satisfied. We need a + b = 1 to make $g_{\theta_0}(1) = 1$ holds. In addition, since $f_X(x)$ is always non-negative, the last equation implies

$$\frac{f_X(0)}{b} = \frac{f_X(0)}{g'_{\theta_0}(0)} = \frac{f_X(1)}{g'_{\theta_0}(1)} = \frac{f_X(1)}{2a+b}$$
(30)

and if both $f_X(0)$ and $f_X(1)$ are positive,

$$a = \frac{f_X(1) - f_X(0)}{f_X(1) + f_X(0)} \quad b = \frac{2f_X(0)}{f_X(1) + f_X(0)}.$$
(31)

By a simple calculation, $g'_{\theta_0}(x) \ge 0$ always holds for $x \in [0, 1]$, no matter if $f_X(0) > f_X(1)$ or $f_X(0) < f_X(1)$. By property of convergence in probability, it is easy to see that if

$$\hat{f}_X(0) = f_X(0) + O_p(h^2)$$

$$\hat{f}_X(1) = f_X(1) + O_p(h^2)$$

and letting $g_{\theta_n}(x) = a_n x^2 + b_n$ where

$$a_n = \frac{\hat{f}_X(1) - \hat{f}_X(0)}{\hat{f}_X(1) + \hat{f}_X(0)} \quad b_n = \frac{2\hat{f}_X(0)}{\hat{f}_X(1) + \hat{f}_X(0)}$$
(32)

then $g_{\theta_n}(x) = g_{\theta_0}(x) + O_p(h^2)$ for any fixed x.

REMARK 5.1 In this paper, the simple boundary correction method by Jones (1993), available in R package evmix, is applied to estimate $f_X(0)$ and $f_X(1)$. This method is essentially equivalent to the kernel weighted local linear fitting at the boundary, thus implemented fast and is able to provide very accurate estimate at the boundaries.

5.2. Discussion of Scenarios

Notice that equation (30) can be solved only if $f_X(0)$ and $f_X(1)$ are both positive. On the other side, if any one of them is zero, then the solution equation (31) is invalid, and thus the quadratic polynomial transform introduced in previous section could be also invalid. We summarize all the scenarios below:

- (1) If both of them are zero, first order polynomial function can be applied as transformation function. Solving equation array (29) always results in an identity function as the transformation function.
- (2) If both of them are not zero, quadratic function can be applied as transformation function, as described in Section 5.1.
- (3) If only one of $f_X(0)$ and $f_X(1)$ is zero, a higher order polynomial function needs be applied as transformation function. More details are following

Therefore, if we ascertained if the density function has zero value at the boundaries, then we could select the order of the polynomial for the transformation. For the sake of simplicity, the simulation and real data analysis in this paper we judge if $f_X(0)$ and $f_X(1)$ are zero by visually inspecting histogram and using the estimated $f_X(0)$ and $f_X(1)$ by the method described in Remark 5.1.

REMARK 5.2 In real practice, we propose to use a nonparametric hypothesis testing method, which tests whether a density distribution has zero value at point of interest, by Paparoditis and Politis (2016b).

REMARK 5.3 For scenario (3) discussed above, Without loss of generality, assume $f_X(1) > f_X(0) = 0$. Let $Z = X^k$, where k > 1 is an integer, and first start with k = 2. Then we judge if $f_Z(0) = 0$. If the conclusion is $f_Z(0) = 0$, then we increase k by 1, until to a point $f_Z(0) \neq 0$, denote this k by k_0 . Transformed quadratic polynomial flat-top series estimator could be now applied on datasets $\{Z_i\}_{i=1}^n$, i.e. $Y_{\theta_n} = g_{\theta_n}(Z)$, and

$$\hat{f}_Z(z;g_{\theta_n}) = \hat{f}_{Y_{\theta_n}}[g_{\theta_n}(z);g_{\theta_n}] \cdot g'_{\theta_n}(z)$$

Since $Z = X^{k_0}$ is strictly increasing function on $X \in [0, 1]$, the final estimator of $f_X(x)$ is

$$\hat{f}_X(x;g_{\theta_n},k_0) = \hat{f}_Z(x^{k_0};g_{\theta_n},k_0) \cdot k_0 x^{k_0-1}$$

If $f_X(0) > f_X(1) = 0$ holds, then let $Z = (1 - X)^{k_0}$, which is a decreasing function on [0, 1], thus the true density function is

$$f_X(x) = -f_Z[(1-x)^{k_0}; g_{\theta_n}, k_0] \cdot [(1-x)^{k_0}]'$$

and

$$\hat{f}_X(x;g_{\theta_n},k_0) = \hat{f}_Z[(1-x)^{k_0};g_{\theta_n},k_0] \cdot k_0(1-x)^{k_0-1}$$

REMARK 5.4 For the high order polynomial transformation function in Remark 5.3, recall (28), if $k = k_0$, then

$$f_Z(z) = rac{f_X(z^{1/k_0})}{k_0 z^{(1-k_0)/k_0}}.$$



Figure 1. Plots of $\eta_{\omega}(s)$ with c = 0.05 and different b.

One problem might arise when $z^{(1-k_0)/k_0}$ decay faster than $f_X(z^{1/k_0})$ as $z \to 0$, i.e.

$$\frac{1}{f_Z(z)} = \frac{z^{(1-k_0)/k_0}}{f_X(z^{1/k_0})} = o(1)$$

which implies $f_Z(z) \to \infty$ as $z \to 0$. In our simulations, the estimator still performs well even in this situation but this performance is not theoretically supported, and a small order of bias is not guaranteed.

6. Selection of Parameters

6.1. Selection of shape parameters b and c

The function η_{ω} given in (25) was first introduced by McMurry and Politis (2004) who suggested to use a relatively small value for *b* since a large *b* makes η_{ω} more rectangular-looking which is undesirable, see Fig.1. Similarly we also find that a large value for *c* causes the same problem. In applications, the empirical choices b = 1/4 and c = 0.05 are recommended.

6.2. Selection of Bandwidth h

The problem of the bandwidth selection over the whole compact support needs further study. The first question arises by Theorem 4.5, i.e., the bias orders at the boundary and the interior are different if r > 2. If a global bandwidth $h_1 = An^{-1/(2r+1)}$ is selected, where A_1 is a constant, then in the interior area we have

$$\hat{f}_X(x;g_{\theta_n}) - f_X(x) = O_p\left(n^{-\frac{r}{2r+1}}\right).$$

With same bandwidth h_1 , but at the boundaries,

$$\hat{f}_X(x;g_{\theta_n}) - f_X(x) = O_p\left(n^{-\frac{2}{2r+1}}\right)$$

Compare to boundary correction kernel estimators introduced in Section 1, that achieve order of mean squared error (MSE) $O_p(n^{-2/5})$, MSE of Transformed Flat-top Series estimator at interior area reaches higher order, but worse order at two boundaries. If we select two different bandwidths, one for interior region and one for boundaries, then a discontinuity issue takes place at the border of these two regions.

REMARK 6.1 Another possibility is defining a local bandwidth h = h(t), such that

$$h(t) = \begin{cases} h_{interior}, & \delta \le t \le 1 - \delta \\ h_{boundary}, & t = 0, 1 \\ \gamma_h(t), & 0 < t < \delta, \text{ or } 1 - \delta < t < 1 \end{cases}$$

Where $\gamma_h(t)$ a very smooth function we select to make h(t) continuous, further study on validity of this idea is needed.

For simplicity, all the simulation and real data analysis in this paper, we employ bandwidth $h_{interior}$ that is optimal for the interior area. Note that selection of the bandwidth $h_{interior}$ is nontrivial, Politis (2003) showed the inverse of bandwidth of flat-top series estimator performs as a threshold in the frequency domain. If the periodic extension $\tilde{f}_X(x)$ is very smooth, the characteristic function $\phi_X(s)$ will decay to zero very fast, and be negligible when $|s| > \frac{1}{h}$. The bandwidth thus should be chosen so that the flat-top taper leaves the characteristic function at low frequencies unchanged (due to flat-top region) while damping out the characteristic function at high frequencies. With this in mind, as in Politis (2003), we propose the following adaptive bandwidth choice rule of thumb.

- (1) Pick $h_{boundary}$ to minimize $||\theta_n \theta_0||$ (see (32), and Remark 5.1); obtain transform function $g_{\theta_n}(\cdot)$ and transformed data array $Y_{\theta_n} = g_{\theta_n}(X)$.
- (2) Let $\rho(s) = \phi_{Y_{\theta_n}}(s)/\phi_{Y_{\theta_n}}(0)$, and $\hat{\rho}(s) = \hat{\phi}_{Y_{\theta_n}}(s)/\hat{\phi}_{Y_{\theta_n}}(s)$. Let \hat{m} be the smallest positive real number such that $|\hat{\rho}(\hat{m}+s)| < 2\sqrt{\log n/n}$, for all $s \in (0, K_n)$, where K_n is a positive nondecreasing real-valued function of n such that $K_n = o(\log n)$. Then, let $h_{interior} = 1/(2\hat{m})$.
- (3) Estimate $f_{Y_{\theta_n}}(\cdot; g_{\theta_n})$ by (17) with infinite differentiable flat-top function and local bandwidth $h_{interior}$. Finally obtain estimator of $f_X(\cdot)$ by (16).

Note that the adaptive bandwidth choice method might not always be the most appropriate for Transformed Flat-top Series estimator when the density function is not smooth enough. Moreover, the periodic extension of estimated transformed function $f_{Y_{\theta_n}}(y; g_{\theta_n})$ is not really continuous, and thus its characteristic function might not always decay fast. As an alternative, we may estimate the bandwidth by least squares cross validation, i.e., choose the bandwidth that minimizes

$$\begin{split} \int_{0}^{1} [\hat{f}_{Y_{\theta_{n}}}(y;g_{\theta_{n}}) - f_{Y_{\theta_{n}}}(y;g_{\theta_{n}})]^{2} dy &= \int_{0}^{1} \hat{f}_{Y_{\theta_{n}}}^{2}(y;g_{\theta_{n}}) dy - 2 \int_{0}^{1} \hat{f}_{Y_{\theta_{n}}}(y;g_{\theta_{n}}) f_{Y_{\theta_{n}}}(y) dy \\ &+ \int_{0}^{1} f_{Y_{\theta_{n}}}^{2}(y;g_{\theta_{n}}) dy. \end{split}$$

The last term on the right-hand-side does not dependent on h; the first term is easy to compute because this integral is on closed interval [0, 1]; we will use the leave-one-out method to estimate the second term, i.e.,

$$\int_{0}^{1} \hat{f}_{Y_{\theta_{n}}}(y;g_{\theta_{n}}) f_{Y_{\theta_{n}}}(y;g_{\theta_{n}}) dy \approx \frac{1}{n} \sum_{i=1}^{n} \hat{f}_{Y_{\theta_{n}}}^{-i}(y_{i};g_{\theta_{n}})$$

where $\hat{f}_{Y_{\theta_n}}^{-i}$ is the flat-top series estimator computed by omitting observation y_i from the sample.

7. Simulation

In this section, we simulate data from six different selected densities; see Table 1 below:

| # | Density | Description |
|---|---|-----------------------------|
| 1 | Beta(4,4) | $f_X(0) = f_X(1) = 0$ |
| 2 | $\frac{1}{2}$ Beta(1,1)+ $\frac{1}{2}$ Beta _[1/4,3/4] (4,4) | $f_X(0) = f_X(1) > 0$ |
| 3 | Truncated $N\left(0,\frac{1}{4}\right)$ | $f_X(0) > f_X(1) > 0$ |
| 4 | Beta(1,3) | $f_X(0) > f_X(1) = 0$ |
| 5 | Truncated $N\left(-1,\frac{1}{4}\right)$ | $f_X(0) > f_X(1) \approx 0$ |
| 6 | $\frac{1}{2}$ Beta _[0,1/2] (4,4) + $\frac{1}{2}$ Beta _[1/2,1] (4,4) | Bimodal Distribution |

Table 1. Function Form of 6 Densities for Simulations.

Beta (\cdot, \cdot) is the Beta density distribution, Beta $[a,b](\cdot, \cdot)$ denotes the corresponding Beta density function rescaled to interval [a, b]; "truncated" means the original density function is truncated to interval [0, 1]. Figure 2 shows the shape of the six probability density functions as well as examples of a few corresponding Transformed Flat-top Series estimators. As mentioned in Section 5.2, for density 1 and 6 we use the identity function as transformation function. For Density 2 and 3 we use quadratic function as transformation function introduced in Section 5.1. Density 4 and 5 we use two higher order polynomials as transformation function discussed in Remark 5.3 (density 4 applied 4th order polynomial and density 5 applied 6th order polynomial).

We investigate the finite sample performance of proposed Transformed Flat-top Series estimator using Monte Carlo, and compare its performance to other popular estimators of density with bound support. In addition we apply re-normalization to eliminate any negative value appeared in those estimator of density function, as we brought up earlier in Remark 4.6. The list of estimators are following:

- (1) f_{Beta} denotes the first beta kernel estimator of Chen (1999), and the bandwidth is estimated by rule of thumb bandwidth of Jones and Henderson (2007).
- (2) f_{copula} denotes the Gaussian copula based estimator of Jones and Henderson (2007), and the bandwidth is estimated by rule of thumb bandwidth of Jones and Henderson (2007).
- (3) \hat{f}_{TKE} denotes transformed kernel estimator (TKE) of Wand et al. (1991), and the bandwidth is estimated by plug-in bandwidth of Wen and Wu (2015).

- (4) \hat{f}_{MTKE} denotes modified transformed kernel estimator (MTKE) of Wen and Wu (2015), and the bandwidth is estimated by plug-in bandwidth of Wen and Wu (2015).
- (5) $\tilde{f}_{\rm CV}$ denotes the Transformed Flat-top Series estimator with bandwidth selected by cross-validation.
- (6) \hat{f}_{abc} denotes the Transformed Flat-top Series estimator with bandwidth selected by adaptive bandwidth choice method.



Figure 2. Example of simulations on 6 different styles of densities: Black lines are true density function, each red line is an estimator of density function.

For each density function, we conduct simulations with sample size n = 100 and replicate them 1000 times. Since TKE and MTKE sometimes explode at 0 and 1, we evaluate average of mean integrated square errors (AOMISE) of estimators on an equally spaced grid on [0.001, 0.999] with an increment 0.001. Figure 2 shows the first five estimators. The simulation results are presented in Table 2 where all the values have been multiplied by 1000 to be better readable. For each density, the minimum AOMISE is highlighted in bold font. Estimated standard errors of the listed AOMISE are given in parentheses.

| | Average MISE (Std. Dev.) $\times 10^{-3}$ | | | | | |
|-----------------|---|-----------|-----------|----------|-----------|-----------|
| Methods | Density 1 | Density 2 | Density 3 | Density4 | Density 5 | Density 6 |
| ŕ_ | 61.92 | 82.10 | 23.59 | 20.28 | 42.72 | 193.07 |
| J Beta | (32.30) | (41.89) | (18.77) | (16.41) | (35.80) | (33.59) |
| , | 38.05 | 92.05 | 34.33 | 57.19 | 108.38 | 140.04 |
| J copula | (25.04) | (45.20) | (22.81) | (34.12) | (74.55) | (31.22) |
| \hat{f}_{max} | 42.65 | 98.38 | 60.67 | 85.31 | 121.21 | 158.00 |
| JTKE | (27.23) | (48.47) | (41.01) | (67.13) | (97.87) | (39.61) |
| ĥ | 31.87 | 93.98 | 35.42 | 46.01 | 101.29 | 94.03 |
| JMTKE | (25.48) | (40.30) | (43.84) | (26.30) | (51.72) | (76.94) |
| far | 41.25 | 74.95 | 27.49 | 51.41 | 59.40 | 60.46 |
| JCV | (65.74) | (70.59) | (43.58) | (73.75) | (104.77) | (60.53) |
| , | 18.68 | 124.69 | 17.66 | 39.27 | 45.99 | 53.83 |
| Jabc | (17.98) | (85.08) | (21.30) | (41.33) | (57.76) | (86.02) |

Table 2. Simulation Results with Sample Size n = 100.

REMARK 7.1 A similar simulation was performed on interval $[0, 0.001] \cup [0.999, 1]$, and showed \hat{f}_{copula} , \hat{f}_{TKE} and \hat{f}_{MTKE} all had significantly greater AOMISE than other candidates. \hat{f}_{Beta} , \hat{f}_{CV} and \hat{f}_{abc} had reasonable good results because they all have bias in the order of $O_n(h^2)$ at boundaries.

The simulation results are consistent with the theoretical analysis. Among all the estimators, flat-top estimators provide best AOMISE for 4 out of 6 densities when n = 100, and they are Density 1, 2, 3, 6 (see Table 2 numbers in bold). All these four densities either satisfy $f_X(0) = f_X(1)$ or $f_X(0) \neq f_X(1) > 0$. In these four densities, adaptive bandwidth is powerful as in the previous literature on flat-top based estimators. However, density 2 only has up to 2nd continuous derivative which limits the order of bias that adaptive bandwidth choice method can achieve; also see Politis (2003). Although cross-validation performs in a stable manner, it might not be able to yield $h = A_1 n^{-1/5}$ and is implemented much slower.

For both density 4 and 5, that Beta kernel estimator provides the best AOMISE. Transformation Based Flat-top Estimator with adaptive bandwidth choice and beta kernel give second best AOMISE and work reasonably well. We recommend that the selection of high order polynomial function as transformation function introduced in Section 5.2 deserves more future research work.

8. Real Data Examples

We are looking into two different styles of data, one seems has a density distribution with no boundary effect and the other one has a density distribution with very significant boundary effect. We evaluate Transformed Flat-top Series estimators on those data by comparing to other competitor estimators.

8.1. Data Without Boundary Effect: Financial Returns

In this section, we focus on three representative datasets of daily relative returns taken from a foreign exchange rate, a stock price, and a stock index; a description of our main datasets is as follows.

- Dataset 1: Foreign exchange rate. Daily returns from the Yen vs. Dollar exchange rate from January 1, 1988 to August 1, 2002; the data were downloaded from Datastream. The sample size is 3600 (weekends and holidays are excluded).
- Dataset 2: Stock index. Daily returns of the S&P500 stock index from October 1, 1983 to August 30, 1991; the data are available as part of the garch module in Splus. The sample size is 2000.
- Dataset 3: Stock price. Daily returns of the IBM stock price from February 1, 1984 to December 31, 1991; the data are again available as part of the garch module in Splus. The sample size is 2000.

Letting P_t denote the price of an asset at time t, the relative return at time t is defined as $X_t = (P_t - P_{t-1})/P_{t-1}$. Note that $X_t \ge -1$ by construction since the price P_t can not be negative. In principle, X_t is a weakly dependent and strictly stationary time series random variable, see chapter 10 in Politis (2015), thus its density estimation has the same large-sample variance and bias as i.i.d. random variable, see Adams and Nobel (1998) and Hallin and Tran (1996). The theorems introduced in this paper are based on variance and bias calculation, therefore Transnformed Flat-top Series estimator can be applied to financial returns. In addition X_t is not bounded above; however, when X_t takes on extreme values, it is typically due to crashes in which case the extreme value is towards negative values; see, for example, the extreme value of about -20% associated with the crash of 1987 in the middle of Datasets 2 and 3. Since extreme values are in the negative direction (although it is the bounded direction), it is safe to assume that the positive direction satisfies the symmetric bound, i.e., that $X_t \leq 1$. Therefore, in what follows we assume that X_t has density $f_X(x)$ over the compact support [-1, 1]; We also judge for all three datasets $f_X(-1) = f_X(1) = 0$ by visual inspection on the histogram. Plots and further details on the three datasets can be found in Politis (2015). We apply the same kernel estimators and bandwidth selection as the methods applied in the Section 7. Since sample sizes of all three datasets are over one thousand, all methods are supposed to have small mean integrated squared errors. If we were to plot the estimated densities, it would be difficult visually to tell the differences among these methods. Thus, we propose a k-fold cross validation method to compare methods effectively; the idea is as follows:

- (1) Randomly cut original dataset to K non-overlapping subsets of roughly equal size. Call each subset a partition.
- (2) Use data from partition number i (say) as 'training data' in order to estimate the density function, and use the rest K-1 partitions as testing data; more precisely, use the histogram of the data from all remaining K-1 partitions as the 'true' density function that gives the benchmark to compare the estimated density from partition number i.
- (3) Estimate MISE (mean integrated squared error) by using estimated density function from training data and histogram as true density function.
- (4) Repeat Steps 1 to 3 for all possible i, and obtain K MISE; calculate AOMISE



Figure 3. Example of Transformed Flat-top Series estimator applied on financial return data.

(average of mean integrated squared error) by taking average of K MISE. For the above-mentioned histogram, we take the number of bins M equal

$$M = \lceil \frac{R}{3.49s} N^{1/3} \rceil$$

where R is a fixed range of the data (for all 3 datasets here, R = 2), s is sample standard deviation, N is the sample size of the data to be used; for more details, see Scott (1979). Then, our optimal histogram is built by 9 partitions of data, and becomes the reference close to truth.

For Transformed Flat-top Series estimator, we apply identity function as transformation function. Similar as Section 7, two different bandwidth selection methods were chosen to be applied, the adaptive and cross-validation. An example of Transformed Flat-top Series estimators from 1 partition of Dataset 1 versus histogram of 9 partitions of Dataset 1 is plotted in Figure 3. Notice the density estimator has been renormalized to eliminate negative values described in Remark 4.6. Figure 4 shows the same example but fit by Beta, Gaussian copula, TKE and MTKE. Since all these four methods were designed to handle data with unit support, we normalize all 3 datasets to make them all fall in unit interval, and then transform their density estimator back to original support. All these methods are working well with the exception of the Beta kernel estimator that seriously underestimates the magnitude of the peak.

Table 3 gives the results of the 10-fold cross validation. As already mentioned even by visual inspection, the Beta estimator is the worst. The flat-top estimator \hat{f}_{abc} is the best with Dataset 1 (foreign exchange), and close to the best with Dataset 3 (IBM stock price); the flat-top estimator that uses cross-validated bandwidth is also competitive. Note that the TKE and MTKE estimator perform both well in Table 3. Nevertheless, as mentioned earlier, TKE and MTKE will explode and even produce infinite value at the two boundaries of support, even the boundaries are apparently 0. This explosion is not reflected on AOMISE vs. histogram comparison because the histogram never has a bin centered at the boundary.



Figure 4. Example of competitor estimators applied on financial return data.

output

| Average MISE $\times 10^{-3}$ | | | | |
|-------------------------------|--|--|--|--|
| Dataset 1 | Dataset 2 | Dataset 3 | | |
| 0.231 | 0.219 | 0.182 | | |
| 0.147 | 0.172 | 0.177 | | |
| 0.141 | 0.167 | 0.154 | | |
| 0.148 | 0.168 | 0.159 | | |
| 0.145 | 0.176 | 0.169 | | |
| 0.136 | 0.175 | 0.157 | | |
| | Aver Dataset 1 0.231 0.147 0.141 0.148 0.145 0.136 | Average MISE × Dataset 1 Dataset 2 0.231 0.219 0.147 0.172 0.141 0.167 0.148 0.168 0.145 0.176 0.136 0.175 | | |

Table 3. Real Data Performance by 10-fold cross validation.

8.2. Data With Boundary Effect: Ratio of White Student Enrollment

In this section we evaluate Transformed Flat-top Series estimator by applying to the data set which records the ratio of white student enrollment of 56 public schools in Nassau County, New York for the 1992-1993 school year, and consequently restricted on unit interval. It was an example for illustrating boundary problem by Simonoff (1996), and used to compare among different methods by Geenens (2014) and Wen and Wu (2015). In addition, it is interesting to assess if public schools were segregated by race in the 90's by estimating the density of white student enrollment ratios.

We apply the same kernel estimators and bandwidth selection from Section of simulation and plot them in Figure 5 and 6. The histogram of the data is also presented in the same plots. This data set has mean 0.778, standard deviation 0.249, median 0.867, 80% of the data are greater than 0.707. It is apparently the density function at 0 and 1 are

Flat-top Series Estimate



Figure 5. Example of Transformed Flat-top Series estimator applied on student enrollment data.

not 0 based on histogram, thus a quadratic polynomial function is estimated as transformation function. Notice the result we present here is after re-normalization described in Remark 4.6.

Figure 6 shows the fit by Beta, Gaussian copula, TKE and MTKE. Both Beta and Gaussian copula fail to fit and underestimate the shape of peak. TKE and MTKE perform both good, especially at the area Transformed Flat-top Series estimator is 0 due to renormalization, but did explode and produce infinite value at the two ends of unit support.

9. Conclusion

A Transformed Flat-top Series estimator is proposed for density function in compact support. This method is developed from boundary correction method, transformed method and Flat-top series estimator. We also establish the theoretical properties and show its higher order of bias in the interior region of the support. We use adaptive bandwidth choice and cross-validation as for bandwidth selection and present their performance in the simulation, and compare them to four other popular estimators for density with compact support by simulation and empirical example. In summary, the Transformed Infinite Differentiable Flat-top Series estimator have the best AOMISE among the selected other methods for density with compact support. For densities with same values or positive values at two boundaries, it did not reach to the same high performance for densities with zero at only one boundary. Development of the proposed estimators on density with a pole or infinite derivative, and extension to multivariate density problem will be pursued in our future work.



Figure 6. Example of competitor estimators applied on student enrollment data.

References

- Adams, T., and Nobel, A. (1998), 'On density estimation from an ergodic process', The Annals of Probability, 26, 794–804.
- Bartlett, M. (1963), 'Statistical Estimation of Density Functions', The Indian Journal of Statistics, Series A, 25, 245–254.
- Cacoullos, T. (1966), 'Estimation of a multivariate density', Annals of the Institute of Statistical Mathematics, 18, 179–189.
- Chen, S. (1999), 'Beta Kernel Estimators for Density Function', Computational Statistics and Data Analysis, 31, 131–145.
- Cheng, M.Y. (1997), 'Boundary Aware Estimators of Integrated Density Derivative Products', Journal of the Royal Statistical Society, Series B, 59, 191–203.
- Cheng, M.Y., Fan, J., and Marron, J. (1997), 'On Authomatic Boundary Corrections', The Annals of Statistics, 25, 1691–1708.
- Cline, D.B., and Hart, J.D. (1991), 'Kernel Estimation of Densities with Discontinuities or Discontinuous Derivatives', *Statistics*, 22, 69–84.
- Cowling, A., and Hall, P. (1996), 'On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation', *Journal of the Royal Statistical Society. Series B*, 58, 551–563.
- Devroye, L. (1987), A Course in Density Estimation, Birkhauser Boston.
- Gasser, T., and Muller, H.G. (1979), 'Kernel Estimation of Regression Functions', in Smoothing Techniques for Curve Estimation, lecture Notes in Mathematics, eds. T. Gasser and R. M., Springer Berlin Heidelberg, pp. 23–68.
- Gasser, T., Muller, H.G., and Mammitzsch, V. (1985), 'Kernels for Nonparametric Curve Estimation', Journal of the Royal Statistical Society. Series B, 47, 238–252.
- Geenens, G. (2014), 'Probit Transformation for Kernel Density Estimation on the Unit Interval', Journal of the American Statistical Association, 109, 346–358.
- Gottlieb, D., and Shu, C.w. (1997), 'On the Gibbs Phenomenon and Its Resolution', SIAM Rev., 39, 644–668.
- Granovsky, B.L., and Muller (1991), 'Optimal kernel methods: A unifying variational principle', International Statistical Review, 59, 373–388.
- Hall, P., and Park, B.U. (2002), 'New methods for bias correction at endpoints and boundaries', *The Annals of Statistics*, 30, 1460–1479.
- Hallin, M., and Tran, L. (1996), 'Kernel density estimation for linear processes: asymptotic normality and bandwidth selection', Annals of the Institute of Statistical Mathematics, 48, 429– 449.
- Hewitt, E., and Hewitt, R.E. (1979), 'The Gibbs-Wilbraham phenomenon: An episode in fourier analysis', Archive for History of Exact Sciences, 21, 129–160.
- Jones, M. (1993), 'Simple Boundary Correction for Kernel Density Estimation', Statistics and Computing, 3, 135–146.
- Jones, M. (1995), 'On Higher Order Kernels', Journal of Nonparametric Statistics, 5, 215–221.
- Jones, M., and Foster, P. (1993), 'Generalized Jackknifing and Higher Order Kernels', Journal of Nonparametric Statistics, 3, 81–94.
- Jones, M., and Henderson, D. (2007), 'Kernel-Type Density Estimation on the Unit Interval', Biometrika, 94, 977–984.
- Marron, J. (1994), 'Visual Understanding of Higher-Order Kernels', Journal of Computational and Graphical Statistics, 3, 447–458.
- Marron, J., and Ruppert, D. (1994), 'Transformations to Reduce Boundary Bias in Kernel Density Estimation', *Journal of the Royal Statistical Society. Series B*, 56, 653–671.
- Marron, J., and Wand, M. (1992), 'Exact Mean Integrated Squared Error', The Annals of Statistics, 20, 712–736.
- McMurry, T.L., and Politis, D.N. (2004), 'Nonparametric regression with infinite order flat-top kernels', *Journal of Nonparametric Statistics*, 16, 549–562.
- Muller, H.G. (1991), 'Smooth Optimum Kernel Estimators Near Endpoints', *Biometrika*, 78, 521–530.

- Nadaraya, E.A. (1989), Nonparametric Estimation of Probability Densities and Regression Curves, Springer Netherlands.
- Paparoditis, E., and Politis, D.N. (2016a), 'A Note on the Behaviour of Nonparametric Density and Spectral Density Estimators at Zero Points of their Support', Journal of Time Series Analysis, 37, 182–194.
- Paparoditis, E., and Politis, D.N. (2016b), 'On the behavior of nonparametric density and spectral density estimators at zero points of their support', Journal of time series analysis, 37, 182–194.
- Parzen, E. (1962), 'On Estimation of a Probability Density Function and its Mode', Annals of Mathematical Statistics, 33, 1065–1076.
- Politis, D.N. (2015), Model-free Prediction and Regression: A Transformation-based Approach To Inference, Springer New York.
- Politis, D.N., and Romano, J.P. (1999), 'Multivariate Density Estimation with General Flat-Top Kernels of Infinite Order', Journal of Multivariate Analysis, 68, 1–25.
- Politis, N., Dimitris (2001), 'On nonparametric function estimation with infinite-order flat-top kernels', in Probability and Statistical Models with applications, eds. C.A. Charalambides, K. M.V., and N. Balakrishnan, Boca Raton: Chapman and Hall/CRC, pp. 469–483.
- Politis, N., Dimitris (2003), 'Adaptive bandwidth choice', Journal of Nonparametric Statistics, 15, 517-533.
- Politis, N., Dimitris, and Romano, P., Joseph (1993), 'On a Family of Smoothing Kernels of Infinite Order', in Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface, eds. M. Tarter and M. Lock, San Diego, California: The Interface Foundation of North America, pp. 141–145.
- Schuster, E.F. (1985a), 'Incorporating Support Constraints Into Nonparametric Estimators of Densities', Communications in Statistics - Theory and Methods, 14, 1123–1136.
- Schuster, E.F. (1985b), 'Incorporating support constraints into nonparametric estimators of densities', Communication in Statistics-Theory and Methods, 14, 1123–1136.
- Scott, D.W. (1979), 'On Optimal and Data-Based Histogram', Biometrika, 66, 605–610.
- Silverman, B. (1986), Density Estimation for Statistics and Data Analysis, London: Chapman and Hall.
- Simonoff, J.S. (1996), Smoothing Methods in Statistics, Springer New York.
- Wand, M., Marron, J., and Ruppert, D. (1991), 'Transformations in Density Estimation', Journal of the American Statistical Association, 86, 343–353.
- Wen, K., and Wu, X. (2015), 'An Improved Transformation-based Kernel Estimator of Densities on the Unit Interval', Journal of the American Statistical Association, 110, 773-783.
- Zhang, S., and Karunamuni, R.J. (1998), 'On Kernel Density Estimation Near Endpoints', Journal of Statistical Planning and Inference, 70, 301–316.
- Zhang, S., and Karunamuni, R.J. (2000), 'Smooth Optimum Kernel Estimators Near Endpoints', Journal of Nonparametric Statistics, 12, 197–221.
- Zhang, S., Karunamuni, R., and Jones, M. (1999), 'An Improved Estimator of the Density Function at the Boundary', Journal of the American Statistical Association, 94, 1231–1241.