KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.1081

"Smoothed bootstrapping kernel density estimation under higher order kernel"

Kun Yi, Yoshihiko Nishiyama

September 2022

KYOTO UNIVERSITY

KYOTO, JAPAN

Smoothed bootstrapping kernel density estimation under higher order kernel

Kun Yi [*] Yoshihiko Nishiyama [†]

September, 2022

Abstract

Smoothed bootstrap method is a useful method to approximates the bias of Kernel density estimation. However, it can only be applied when the kernel function is of second order. In this study, we propose a novel method to generalize the smoothed bootstrap method to higher order kernel for estimating the bias and construct bias corrected estimator based on it.Theoretical formulation and numerical simulation demonstrate that the proposed method achieve better performance compared to the traditional bias correction method.
keywords **kernel density estimation, smoothed bootstrap, bias estimation, higher order kernel.**

1 Introduction

Kernel Density Estimation (KDE) is a nonparametric method to estimate the probability density function of a random variable without assuming that the underlying density function is from a particular family [16]. This flexibility makes KDE one of the most popular method to estimate probability density function, especially for data drawn from a complicated distribution. However, KDE generally contains a bias which is difficult to be handled because it often involves higher-order derivatives of the underlying function and cannot be easily captured by resampling methods, such as the bootstrap [3].

To handle the bias, two main approaches have been proposed: undersmoothing and bias corrected. The undersmoothing method is to choose a bandwidth smaller than the optimal bandwidth for point estimation. then the bias is of smaller order than the estimator asymptotically [6, 18]. However, the conventional bandwidth selector does not give an undersmoothing bandwidth so it remains a question how to practically implement this method.

The second approach is to bias correct the estimator with the explicit goal of removing the bias [1, 2, 4, 5, 9, 10, 15]. Because the bias term involves higher-order derivative of the targeted function,

[*]Graduate Shool of Economics, Kyoto University, yi.kun.52n@st.kyoto-u.ac.jp

[†]Institue of Economic Research, Kyoto University, nishiyama@kier.kyoto-u.ac.jp

most existing debiased methods introduce another estimator of the derivatives to estimate the bias and obtain a consistent bias estimator based on it.

Besides estimating derivatives of objective density function, an alternative method for bias estimation, namely smoothed bootstrap [11], is to resample from empirical distribution and sample from kernel function to construct smoothed bootstrap estimator. However, the smoothed bootstrap method can not be applied for higher order kernels because higher order kernels are not always non-negative. Higher-order kernels have been widely used in nonparametric estimation because they often give faster asymptotic rates of convergence [12, 14]. Several studies have attempted to remove or correct the negative part of higher order kernels [8, 13].

In this study, we propose a generalized smoothed bootstrap method for bias estimation under higher order kernels and construct bias correction estimator based on it. We also construct confidence bands/intervals via bootstrapping the bias correction estimator [2, 7, 10]. The major contributions of this work are summarized as follows.

- We propose a novel bias correction method and theoretically prove that the proposed method outperform traditional bias correction method in terms of bias, variance and Mean Squared Error (MSE).
- We construct confidence bands/intervals based on the proposed bias correction method. Simulation studies reveal that our proposed bias correction estimator and confidence bands/intervals outperform the traditional bias correction method.

The remainder of this paper is organized as follows.

In Section 2, we give a brief review of the debiased estimator and smoothed bootstrap method under second order kernel. In Section 3, we propose a generalized smoothed bootstrap estimator for bias estimation and construct bias correction estimators based on it. Then we theoretically compare the proposed method with traditional bias correction method. In Section 4, We use simulations to demonstrate that bias correction estimation and confidence bands/intervals based on our proposed method are valid compared to traditional method. Finally, we conclude this paper and discuss some possible future directions in Section 5.

2 Kernel density and smoothed bootstrap

2.1 Kernel density estimator and the bias

Here we briefly review the debiased estimator of the KDE proposed in [1]. Assume that the observed data obey the following assumption.

Assumption 1. *$\mathcal{X} = \{X_1, \dots, X_n\}$ is a random sample with an absolutely continuous distribution with Lebesgue density f. For a fixed interior point x, in a neighborhood of x, f is L-times continuously differentiable with bounded derivatives $f^{(l)}, l = 1, 2, \dots, L$.*

The classical KDE of f with kernel function K and bandwidth h is,

$$\hat{f}_{K,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{X_i - x}{h}\Big) \tag{1}$$

We say kernel function K is of order L if

$$\int u^s K(u)du = \begin{cases} 1 & s = 0 \\ 0 & s = 1, ..., L-1 \\ C\,(\neq 0, \pm\infty) & s = L. \end{cases} \tag{2}$$

Assumption 2. *The kernel K is a bounded, even function with a compact support, and of order $L \geq 2$, where L is an even integer. The bandwidth h follows $h \to 0$ and $nh \to \infty$ as $n \to \infty$.*

Remark 1. *When $L = 2$, the kernel is a density function. For $L \geq 4$, K is said to be a higher order kernel function of order L. Note that the $K(u) \geq 0$ does not necessarily hold when $L \geq 4$.*

Estimator (1) is a biased estimator of $f(x)$ and the bias is

$$\begin{aligned} E\{\hat{f}_{K,h}(x) - f(x)\} &= E\{\frac{1}{h}K(\frac{X_1 - x}{h})\} - f(x) \\ &= \int K(u)\{f(x + hu) - f(x)\}du. \end{aligned} \tag{3}$$

If we use a kernel of order $L(\geq 2)$, and $f(x)$ is at least L times differentiable,

$$E\{\hat{f}_{K,h}(x) - f(x)\} = \frac{h^L f^{(L)}(x)}{L!}\int u^L K(u)du + o(h^L). \tag{4}$$

An intuitive method to estimate and correct the bias is to use the estimation of L-th derivative of f. From (4), the leading term of the bias is,

$$B_1(x) = \frac{h^L f^{(L)}(x)}{L!}\int u^L K(u)du \tag{5}$$

which can be estimated by

$$\hat{B}(x) = \frac{h^L \hat{f}^{(L)}_{H,b}(x)}{L!}\int u^L K(u)du \tag{6}$$

where

$$\hat{f}^{(L)}_{H,b}(x) = \frac{1}{nb^{L+1}}\sum_{i=1}^{n} H^{(L)}(\frac{X_i - x}{b})$$

and $H(\cdot)$ and b are a kernel function and a bandwidth.

Assumption 3. *The kernel H is a bounded, even function with a compact support, and of order $L' \geq 2$, where L' is an even integer. The bandwidth b follows $b \to 0$ and $nb \to \infty$ as $n \to \infty$. H is L-times continuously differentiable with bounded derivatives $H^{(l)}, l = 1, 2, \ldots, L$.*

If we assume additional conditions on the smoothness of the objective function,

Assumption 4. *For a fixed interior point x, in a neighborhood of x, f is $(L+L')$-times continuously differentiable with bounded derivatives $f^{(l)}, l = 1, 2, \ldots, (L+L')$.*

then, we have the following theorem which is proved in Appendix B.1.

Theorem 1. *Under Assumptions 1, 2, 3, and 4, the bias and variance of $\hat{B}(x)$ is*

$$Bias\{\hat{B}\} = \frac{h^L b^{L'} \mu_{L1}(K)\mu_{L'1}(H) f^{(L+L')}(x)}{(L!)(L'!)} - \frac{\mu_{L+1,1}(K) h^{L+1} f^{(L+1)}(x)}{(L+1)!} + O(h^{L+2}) + O(h^L b^{L'+1}) \tag{7}$$

$$Var\{\hat{B}\} = \frac{h^{2L} f(x) R(H^{(L)}) \mu_{L1}(K)^2}{n b^{2L+1} (L!)^2} + O(\frac{h^{2L}}{n}) \tag{8}$$

Remark 2. *The bias of $\hat{B}$ is at rate $O(\max\{h^{L+1}, h^L b^{L'}\})$ and the variance of $\hat{B}$ is at rate $O(\frac{h^{2L}}{nb^{2L+1}})$.*

This result is a little different from that in [1] because Assumption 4 assume more smoothness.

2.2 Smoothed bootstrap under second order kernel

Let $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leq x)$ be the empirical cumulative distribution function and $\{X_1^*, \cdots, X_n^*\}$, $X_i^* \sim i.i.d.\hat{F}$ be a nonparametric bootstrap sample. We can construct the bootstrap analogue of the estimator,

$$\hat{f}_{K,h}(x)^* = \frac{1}{nh}\sum_{i=1}^n K(\frac{X_i^* - x}{h}).$$

This does not work for bias approximation because

$$\begin{aligned} E^*\{\hat{f}_{K,h}(x)^* - \hat{f}_{K,h}(x)\} &= E^*\{\frac{1}{h}K(\frac{X_1^* - x}{h})\} - \hat{f}_{K,h}(x) \\ &= \frac{1}{nh}\sum_{i=1}^n K(\frac{X_i - x}{h}) - \hat{f}_{K,h}(x) \\ &= 0 \end{aligned}$$

in view of the bias (3), where $E^*(\cdot) = E(\cdot|\mathcal{X})$.

For a second order kernel function K, we can estimate the bias by smoothed bootstrapping [11]. Smoothed bootstrap sample is obtained as follows.

Let $\{X_1^*, \cdots, X_n^*\}$ be a nonparametric bootstrap sample and $\{\varepsilon_1, \cdots, \varepsilon_n\}$ be a random sample from $K(\cdot)$ independent of $\{X_1^*, \cdots, X_n^*\}$. Compute $\{X_1^+, \cdots, X_n^+\}$ where $X_i^+ = X_i^* + h\varepsilon_i$. Smoothed bootstrap estimator is defined as

$$\tilde{f}^+_{K,h}(x) = \frac{1}{nh}\sum_{i=1}^n K(\frac{X_i^+ - x}{h}). \tag{9}$$

We evaluate its expectation conditional on $\mathcal{X}$. For this purpose, we derive the conditional distribution of X_1^+,

$$\begin{aligned}
P(X_1^+ \le x|\mathcal{X}) &= P(X_1^* + h\varepsilon_1 \le x|\mathcal{X}) \\
&= E\{I(X_1^* + h\varepsilon_1 \le x)|\mathcal{X}\} \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} I(X_i + h\varepsilon_i \le x)|\mathcal{X}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[I(X_i + h\varepsilon_i \le x)|X_i] \\
&= \frac{1}{n}\sum_{i=1}^{n} P(\varepsilon_i \ge \frac{x - X_i}{h}|X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} P(\varepsilon_i \le \frac{X_i - x}{h}|X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\frac{x-X_i}{h}} K(t)dt.
\end{aligned}$$

The second equality from the bottom holds because of the symmetry of $K(\cdot)$. This implies that the density fucntion of X_1^+ is

$$f_{X_1^+}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{X_i - x}{h}) = \hat{f}_{K,h}(x).$$

Therefore,

$$\begin{aligned}
E^*\{\tilde{f}^+_{K,h}(x)\} &= E^*\{\frac{1}{h}K(\frac{X_1^+ - x}{h})\} = E\{\frac{1}{h}K(\frac{X_1^+ - x}{h})|\mathcal{X}\} \\
&= \int \frac{1}{h}K(\frac{y - x}{h})\hat{f}_{K,h}(y)dy \\
&= \int K(u)\hat{f}_{K,h}(x + hu)du
\end{aligned}$$

and thus

$$\begin{aligned}
E^*\{\tilde{f}^+_{K,h}(x) - \hat{f}_{K,h}(x)\} &= \int K(u)\hat{f}_{K,h}(x + hu)du - \hat{f}_{K,h}(x) \\
&= \int K(u)\{\hat{f}_{K,h}(x + hu) - \hat{f}_{K,h}(x)\}du \qquad (10)
\end{aligned}$$

Or equivalently,

$$\begin{aligned}
E^*\{\tilde f^+_{K,h}(x) - \hat f_{K,h}(x)\} &= E[\{\frac{1}{h}K(\frac{X_1^*-x}{h}+\varepsilon_1)\}|\mathcal{X}] - \hat f_{K,h}(x)\\
&= E[\frac{1}{h}\int K(\frac{X_1^*-x}{h}+u)K(u)du|\mathcal{X}] - \hat f_{K,h}(x)\\
&= \frac{1}{nh}\sum_{i=1}^n \int K(v)K(v-\frac{X_1^*-x}{h})dv - \hat f_{K,h}(x)\\
&= \int \frac{1}{nh}\sum_{i=1}^n K(\frac{(x+hv)-X_i}{h})K(v)dv - \hat f_{K,h}(x)\\
&= \int K(u)\{\hat f_{K,h}(x+hu) - \hat f_{K,h}(x)\}du.
\end{aligned}$$

Comparing (3) and (10), we see the smoothed bootstrap bias approximates the bias of KDE.

3 Main result

3.1 Generalized smoothed bootstrap under higher order kernels

The above method can be used when K is a second order kernel or a density. When using a higher order kernel, we cannot generate ε_i as above because K is not a density. We consider the following modification of the bootstrapping method to handle this issue. Let $\varepsilon_i \sim i.i.d.\, g(\varepsilon)$ where $g(\cdot)$ is a user-determined density. We construct the following generalized smoothed bootstrap estimator,

$$\tilde f^+_{K;H,b}(x) = \frac{1}{nh}\sum_{i=1}^n \frac{H(\varepsilon_i)}{g(\varepsilon_i)}K\Big(\frac{X_i^*+b\varepsilon_i-x}{h}\Big)$$

We note that this reduces to (9) when $g=H, b=h$. Also, it is a density only asymptotically because

$$\int \tilde f^+_{K;H,b}(x)dx = \frac{1}{n}\sum_{i=1}^n \frac{H(\varepsilon_i)}{g(\varepsilon_i)} \to E\{\frac{H(\varepsilon_1)}{g(\varepsilon_1)}\} = \int H(u)du = 1.$$

If we want to make it a density, we can simply construct

$$\tilde{\tilde f}^+_{K;H,b}(x) = \frac{1}{nh}\sum_{i=1}^n \frac{H(\varepsilon_i)}{\delta_n g(\varepsilon_i)}K(\frac{X_i^+-x}{h})$$

with $\delta_n = \frac{1}{n}\sum_{i=1}^n \frac{H(\varepsilon_i)}{g(\varepsilon_i)}$. Obviously, if $g=H$, this reduces to the standard smoothed bootstrap.

The expectation of $\tilde{f}^{+}_{K;H,b}(x)$ conditional on $\mathcal{X}$ is

$$\begin{aligned}
E^*\{\tilde{f}^{+}_{K;H,b}(x)\} &= E^*\{\frac{1}{h}\frac{H(\varepsilon_i)}{g(\varepsilon_i)}K(\frac{X_i^*+b\varepsilon_i-x}{h})\} \\
&= E^*\int \frac{1}{h}\frac{H(z)}{g(z)}K(\frac{X_i^*+bz-x}{h})g(z)dz \\
&= E^*\int \frac{1}{h}H(z)K(\frac{X_i^*+bz-x}{h})dz \\
&= E^*\int \frac{1}{h}\frac{h}{b}H(hy-\frac{X_i^*+x}{b})K(y)dy \qquad (11)\\
&= \int \frac{1}{b}K(y)\frac{1}{n}\sum_{i=1}^{n}H(\frac{x+hy-X_i}{b})dy \\
&= \int K(u)\frac{1}{nb}\sum_{i=1}^{n}H(\frac{x+hu-X_i}{b})du \\
&= \int K(u)\hat{f}_{H,b}(x+hu)du. \qquad (12)
\end{aligned}$$

Therefore, we see that it potentially approximates $E\hat{f}_{K,h}(x) = \int K(u)f(x+hu)du$. Then

$$\hat{B}^{+}(x) = E^*\{\tilde{f}^{+}_{K;H,b}(x) - \hat{f}_{H,b}(x)\} = \int K(u)\{\hat{f}_{H,b}(x+hu) - \hat{f}_{H,b}(x)\}du \qquad (13)$$

3.2 Comparison of $\hat{B}(x)$ and $\hat{B}^{+}(x)$

We compare the performance of $\hat{B}(x)$ in (6) and $\hat{B}^{+}(x)$ in (13). The following theorem reveals the bias and variance of $\hat{B}^{+}(x)$, which demonstrates that $\hat{B}^{+}(x)$ performs very well as an estimator of $E\{\hat{f}_{K,h}(x) - f(x)\}$.

Theorem 2. *Under Assumptions 1, 2, 3, and 4, $\hat{B}^{+}(x) = E^*\{\tilde{f}^{+}_{K;H,b}(x) - \hat{f}_{H,b}(x)\}$ approximates the bias in (3) with the following bias and variance, where E^* indicates expectation conditional on $\mathcal{X}$.*

$$Bias\{\hat{B}^{+}\} = \frac{h^L b^{L'}\mu_{L1}(K)\mu_{L'1}(H)f^{(L+L')}(x)}{(L!)(L'!)} + O(h^L b^{L'+1}) \qquad (14)$$

$$Var\{\hat{B}^{+}\} = \begin{cases} \frac{R(H)f(x)}{nb} + O(\frac{1}{n}) & (h/b\to\infty) \\ \frac{\int[\int K(u)H(v+Mu)du - H(v)]^2 dv f(x)}{nb} + O(\frac{1}{n}) & (h/b\to M<\infty) \\ \frac{h^{2L}f(x)R(H^{(L)})\mu_{L1}(K)^2}{nb^{2L+1}(L!)^2} + O(\frac{h^{2L+1}}{nb^{2L+2}}) & (h/b\to 0) \end{cases} \qquad (15)$$

Remark 3. *The bias of $\hat{B}^{+}$ is at rate $O(h^L b^{L'})$ and the variance of $\hat{B}^{+}$ is at rate $O(\max\{\frac{1}{nb}, \frac{h^{2L}}{nb^{2L+1}}\})$.*

The rate of bias and variance of $\hat{B}$ depends on the choice of h and b. Comparing the bias and variance of $\hat{B}^{+}$ with $\hat{B}$, we have

- When $h/b\to\infty$, $\hat{B}^{+}$ has smaller bias and variance compared to $\hat{B}$.

- When $h/b \to M < \infty$, $\hat{B}^+$ has smaller bias and the variances of $\hat{B}^+$ and $\hat{B}$ are at the same rate.
- When $h/b \to 0$ and $h/(b^{L'}) \to \infty$, $\hat{B}^+$ has smaller bias, and the variances of $\hat{B}^+$ and $\hat{B}$ have the same leading term.
- When $h/(b^{L'}) \to M' < \infty$, the biases of $\hat{B}^+$ and $\hat{B}$ are at the same rate, and the variances of $\hat{B}^+$ and $\hat{B}$ have the same leading term.

Hence, the generalized smoothed bootstrap estimator, $\hat{B}^+$ approxiamtes the bias better than the traditional bias estimator, $\hat{B}$.

We demonstrate the optimal bandwidth bandwidth for $\hat{B}^+$ in Appendix C.

4 Simulation studies

In this section, we conduct a simulation study to compare generalized smoothed bootstrap based bias correction and traditional bias correction. We evaluate our approach by answering the following questions.

Q1. How well does the proposed method perform in bias correction estimation compared to traditional methods?

Q2. How well does the proposed method perform in constructing confidence bands/intervals compared to those based on traditional methods?

4.1 Simulation settings

We consider the number of observations $n = 100, 300$, and 500 with the underlying distributions being Gaussian mixture, $0.6 \times N(0, 1) + 0.4 \times N(3, 0.25)$. The Gaussian kernel of order 4 is used as $K = H$.

For simulation of generalized smoothed bootstrap, we drew 100 random samples of size n from f. Based on each random sample we constructed the kernel density estimator $\hat{f}$. Then 100 bootstrap samples of size n were drawn from $\hat{f}$. The generalized smoothed bootstrap estimate of the bias was approximate from these 100 bootstrap samples.

The bandwidths b, h are set to be the same. [1] argues that b/h should be bounded and positive. If $b/h \to 0$, i.e. b vanishes much faster than h, although more bias will be removed, the variance is no longer controlled. If $b/h \to \infty$, the bias rate is not improved. Moreover, leaving only one bandwidth to select is convenient from the practitioner's point of view.

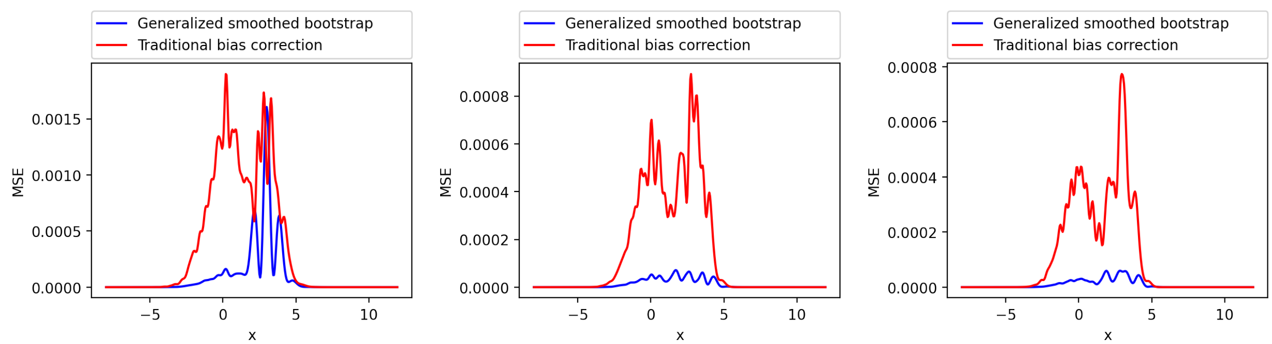
4.2 Q1: Bias estimation

We firstly compare the Mean Integrated Squared Error(MISE) of generalized smoothed bootstrap based bias correction and traditional bias correction using several pre-determined bandwidths.

As shown in Table 1, our method outperforms the traditional bias correction in terms of MISE under most bandwidths. Moreover, by comparing the best result of each row, we see that the best result of our method outperforms the best result of traditional bias correction.

Table 1: MISE of estimators with several pre-determined bandwidths. The † denotes best result of each row.

	$h(=b)$	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$n = 100$	Traditional bias correction	0.016088	0.011101	0.008315	0.007264	0.006010	0.004663†	0.004928
	Proposed method	**0.001030**	**0.000720**	**0.000595†**	**0.000690**	**0.001043**	**0.001970**	**0.003647**
$n = 300$	Traditional bias correction	0.005490	0.003784	0.002932	0.002334	0.002298†	0.002359	**0.002436**
	Proposed method	**0.000351**	**0.000251**	**0.000225†**	**0.000318**	**0.000815**	**0.001826**	0.003439
$n = 500$	Traditional bias correction	0.003565	0.002169	0.001796	0.001443	0.001397†	**0.001483**	**0.001788**
	Proposed method	**0.000224**	**0.000141†**	**0.000150**	**0.000295**	**0.000790**	0.001812	0.003428



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Figure 1: Bias estimation using order 4 Gaussian kernel and bandwidth selected by cross-validation.

We also compare pointwise Mean Squared Error (MSE) using bandwidth selected by Least Squares Cross-Validation (LSCV) [17].

As shown in Figure 1, when $n = 100$, our method outperforms traditional bias correction at most points; when $n = 300, 500$, our method outperforms traditional bias correction for any point.

4.3 Q2: Confidence bands/intervals

We follow the approach proposed by [2] to construct confidence bands/intervals via bootstrapping the generalized smoothed bootstrap based bias correction estimator and traditional bias correction estimator.

We firstly compare the coverage and (average) length of confidence bands/intervals using several pre-determined bandwidths. As shown in Table 2, 3, and 4, our method achieves similar coverage with much shorter length compared to traditional bias correction.

We also compare the coverage and (average) length of confidence bands/intervals using bandwidth selected by LSCV. As shown in Figure 2, 3, and 4, our method achieves similar coverage and much

Table 2: Length of 95% confidence bands (average length of 95% confidence intervals) with several pre-determined bandwidths. The † denotes best result of each row.

	$h(=b)$	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$n = 100$	Traditional bias correction	0.2858	0.2336	0.2024	0.1823	0.1672	0.1540	0.1394†
	Proposed method	**0.2089**	**0.1702**	**0.1465**	**0.1326**	**0.1182**	**0.1085**	**0.1001†**
$n = 300$	Traditional bias correction	0.1692	0.1377	0.1200	0.1067	0.0973	0.0904	0.0838†
	Proposed method	**0.1233**	**0.1001**	**0.0866**	**0.0769**	**0.0690**	**0.0639**	**0.0589†**
$n = 500$	Traditional bias correction	0.1372	0.1099	0.0947	0.0852	0.0780	0.0699	0.0673†
	Proposed method	**0.1005**	**0.0805**	**0.0676**	**0.0607**	**0.0551**	**0.0493**	**0.0475†**

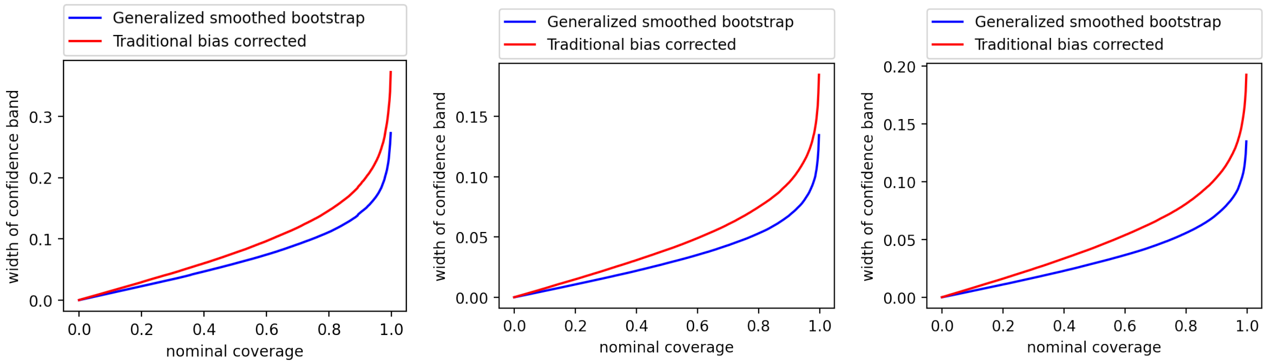
Table 3: Coverage of 95% confidence bands with several pre-determined bandwidths. The † denotes best result of each row.

	$h(=b)$	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$n = 100$	Traditional bias correction	**0.8867**	**0.8467†**	**0.7667**	0.7267	0.7467	**0.7800**	0.7067
	Proposed method	0.8400	0.7733	0.7467	**0.7800**	**0.7600**	0.7667	**0.8467†**
$n = 300$	Traditional bias correction	0.8867	0.9200	**0.9600†**	0.9467	**0.9333**	**0.9533**	**0.8867**
	Proposed method	**0.9200**	**0.9533†**	0.9533	**0.9467**	0.8667	0.8200	0.7867
$n = 500$	Traditional bias correction	0.8867	0.8867	0.9133	0.9533	**1.0000†**	**0.9733**	**0.9733**
	Proposed method	**0.9000**	**0.9400**	**0.9733**	**1.0000†**	**1.0000†**	0.8133	0.7333

Table 4: Coverage of 95% confidence intervals with several pre-determined bandwidths. The † denotes best result of each row.

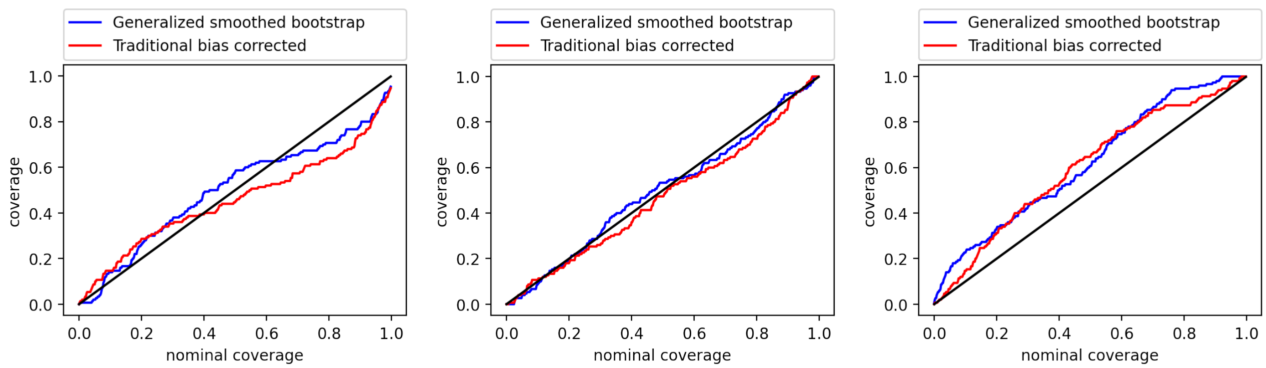
	$h(=b)$	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$n = 100$	Traditional bias correction	**0.8267**	**0.8333†**	0.7800	**0.7733**	**0.8000**	0.7733	0.6667
	Proposed method	0.7333	0.7600	**0.7800**	0.7667	0.7667	**0.7867**	**0.8600†**
$n = 300$	Traditional bias correction	0.7933	0.8267	0.9000	0.9267	0.8933	0.9200	**0.9533†**
	Proposed method	**0.8000**	**0.8600**	**0.9333**	**0.9467**	**0.9533†**	**0.9267**	0.8467
$n = 500$	Traditional bias correction	0.9200	0.9467	0.9933	**1.0000†**	**1.0000†**	**1.0000†**	**1.0000**
	Proposed method	**0.9467**	**0.9867**	**1.0000†**	**1.0000†**	**1.0000†**	0.8400	0.7333

shorter length compared to traditional bias correction.



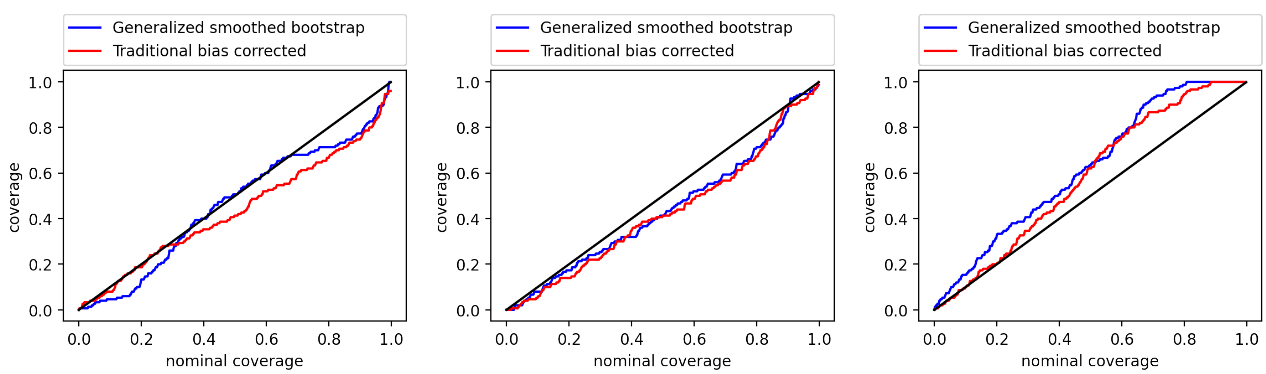
(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Figure 2: Length of confidence bands using bandwidth selected by cross-validation.



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Figure 3: Confidence bands using bandwidth selected by cross-validation.



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Figure 4: Confidence intervals using bandwidth selected by cross-validation.

5 Conclusions

In this study, we propose a generalized smoothed bootstrap method to estimate the bias of KDE. We prove theoretically that the proposed estimator has smaller bias and the same order variance compared to traditional bias correction. We construct bias corrected estimator based on the proposed method. Simulation results reveal that the proposed estimator outperforms traditional bias corrected estimator in terms of MISE. We also construct confidence bands/intervals based on the proposed generalized smoothed bootstrap. The confidence bands/intervals based on the proposed method achieves similar coverage and shorter length compared to traditional method. In addition to its applicability in KDE, the proposed bias estimator could be used in other kernel based methods. A future study to explore the effectiveness of the proposed approach in other domains would be worthwhile.

References

[1] Sebastian Calonico, Matias D Cattaneo, and Max H Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.

[2] Gang Cheng and Yen-Chi Chen. Nonparametric inference via bootstrapping the debiased estimator. *Electronic Journal of Statistics*, 13(1):2194–2256, 2019.

[3] Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer, 1992.

[4] Randall L Eubank and Paul L Speckman. Confidence bands in nonparametric regression. *Journal of the American Statistical Association*, 88(424):1287–1301, 1993.

[5] Peter Hall. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics*, pages 675–694, 1992.

[6] Peter Hall. On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics*, pages 695–711, 1992.

[7] Peter Hall and Joel Horowitz. A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, pages 1892–1921, 2013.

[8] Peter Hall and Robert David Murison. Correcting the negativity of high-order kernel density estimators. *Journal of Multivariate Analysis*, 47(1):103–122, 1993.

[9] Wolfgang Härdle and Adrian W Bowman. Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83(401):102–110, 1988.

[10] Wolfgang Härdle, Sylvie Huet, and E Jolivet. Better bootstrap confidence intervals for regression curve estimation. *Statistics: A Journal of Theoretical and Applied Statistics*, 26(4):287–306, 1995.

[11] Yvonne HS Ho and Stephen MS Lee. Iterated smoothed bootstrap confidence intervals for population quantiles. *The Annals of statistics*, 33(1):437–462, 2005.

[12] M Chris Jones and DF1482137 Signorini. A comparison of higher-order bias kernel density estimators. *Journal of the American Statistical Association*, 92(439):1063–1073, 1997.

[13] Stephen Lee and G. Alastair Young. Practical higher-order smoothing of the bootstrap. *Statistica Sinica*, 4(2):445–459, 1994.

[14] JS Marron. Visual understanding of higher-order kernels. *Journal of Computational and Graphical Statistics*, 3(4):447–458, 1994.

[15] Michael H Neumann. Automatic bandwidth choice and confidence intervals in nonparametric regression. *The Annals of Statistics*, 23(6):1937–1959, 1995.

[16] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[17] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Monographs on Statistics and Applied Probability 60. Springer US, 1st ed edition, 1995.

[18] Larry Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.

A The variance of $\hat{f}_{K,h}$ and $\tilde{f}^{+}_{K;H,b}$

The variance of $\hat{f}_{K,h}(x)$ is

$$
\begin{aligned}
Var\{\hat{f}_{K,h}(x)\} &= \frac{1}{n}\left(E\{\frac{1}{h^2}K(\frac{X_1-x}{h})^2\} - [E\{\hat{f}_{K,h}(x)\}]^2\right) \\
&= \frac{1}{n}\left(\int \frac{1}{h^2}K(\frac{y-x}{h})^2 f(y)dy - [E\{\hat{f}_{K,h}(x)\}]^2\right) \\
&= \frac{1}{n}\left(\frac{1}{h}\int K(u)^2 f(x+hu)du - [E\{\hat{f}_{K,h}(x)\}]^2\right) \\
&= \frac{1}{n}\left(\frac{1}{h}\{f(x)\int K(u)^2du + \frac{h^2}{2}f''(\bar{x})\int K^2(u)u^2du\} - [E\{\hat{f}_{K,h}(x)\}]^2\right) \\
&= \frac{1}{n}\left(\frac{f(x)}{h}R(K) + \frac{h}{2}f''(\bar{x})\mu_{22}(K) - [E\{\hat{f}_{K,h}(x)\}]^2\right) \qquad (16) \\
&= \frac{f(x)}{nh}R(K) + O(\frac{1}{n}) \qquad (17)
\end{aligned}
$$

where $R(K) = \int K(u)^2du$ and $\mu_{lk}(K) = \int u^l K(u)^k du$.

We examine if the variance of $\tilde{f}^{+}_{K;H,b}(x)$ well approximates that of $\hat{f}_{K,h}(x)$,

$$Var^*\{\tilde{f}^{+}_{K;H,b}(x)\} = \frac{1}{n}\left(E^*[\{\frac{1}{h}\frac{H(\varepsilon_i)}{g(\varepsilon_i)}K(\frac{X_i^*+b\varepsilon_i-x}{h})\}^2] - [E^*\{\frac{1}{h}\frac{H(\varepsilon_i)}{g(\varepsilon_i)}K(\frac{X_i^*+b\varepsilon_i-x}{h})\}]^2\right)$$

where the second term is easily seen to be $f(x)^2+o_p(1)$ from (12), while the first term equals to

$$\begin{aligned}
&E^*\{\frac{1}{h^2}\frac{H(\varepsilon_i)^2}{g(\varepsilon_i)^2}K(\frac{X_i^*+b\varepsilon_i-x}{h})^2\} \\
&= E^*\int\frac{1}{h^2}\frac{H(z)^2}{g(z)^2}K(\frac{X_i^*+bz-x}{h})^2 g(z)dz \\
&= E^*\int\frac{1}{h^2}\frac{H(z)^2}{g(z)}K(\frac{X_i^*+bz-x}{h})^2 dz \\
&= \int\frac{1}{h^2}\frac{H(z)^2}{g(z)}\frac{1}{n}\sum_{i=1}^{n}K(\frac{X_i+bz-x}{h})^2 dz \\
&= \frac{1}{nh^2}\sum_{i=1}^{n}\int\frac{H(z)^2}{g(z)}K(\frac{X_i+bz-x}{h})^2 dz.
\end{aligned}$$

We compute its unconditional expectation.

$$\begin{aligned}
&E\left\{\frac{1}{h^2}\int\frac{H(z)^2}{g(z)}K(\frac{X_i+bz-x}{h})^2dz\right\} \\
=&\int\frac{1}{h^2}\int\frac{H(z)^2}{g(z)}K(\frac{y+bz-x}{h})^2 f(y)dzdy \\
=&\frac{1}{h}\int\int\frac{H(z)^2}{g(z)}K(u+\frac{bz}{h})^2 f(x+hu)dzdu \\
=&\frac{1}{h}\int\int\frac{H(z)^2}{g(z)}K(u+\frac{bz}{h})^2\{f(x)+f'(x)hu+\frac{f''(\tilde{x})}{2}(hu)^2\}dzdu \\
=&\frac{f(x)}{h}\int\int\frac{H(z)^2}{g(z)}K(u+\frac{bz}{h})^2dzdu + f'(x)\int\int\frac{H(z)^2}{g(z)}K(u+\frac{bz}{h})^2udzdu \\
&+\frac{h}{2}\int\int f''(\tilde{x})\frac{H(z)^2}{g(z)}K(u+\frac{bz}{h})^2u^2dzdu \\
=&\frac{f(x)}{h}R(K)\int\frac{H(z)^2}{g(z)}dz + f'(x)\int\int\frac{H(z)^2}{g(z)}K(u+\frac{bz}{h})^2udzdu \\
&+\frac{f''(\tilde{x})}{2}h\int\int\frac{H(z)^2}{g(z)}K(u+\frac{bz}{h})^2u^2dzdu. \qquad (18)
\end{aligned}$$

Note that $g(z)$ is an arbitrary density, and if we can choose $g(z)$ such that

$$\int\frac{H(z)^2}{g(z)}dz = 1, \qquad (19)$$

the variance is asymptotically equal to that of $\hat{f}_{K,h}(x)$. When $L=2$ and $g=H$, it obviously holds. For $L'\geq 4$, (19) does not hold, but $g(u)=|H(u)|/\int|H(u)|du$ gives the smallest value of $\int\frac{H(z)^2}{g(z)}dz$.

Furthermore, if

$$\int\int \frac{H(z)^2}{g(z)} K(u+\frac{bz}{h})^2 u dz du = 0, \tag{20}$$

$$\int\int \frac{H(z)^2}{g(z)} K(u+\frac{bz}{h})^2 u^2 dz du = \mu_{22}(K), \tag{21}$$

then $Var^*\{\tilde{f}^+_{K;H,b}(x)\}$ well mimics $Var\{\hat{f}_{K,h}(x)\}$ in a higher order. For the first condition (20),

$$\begin{aligned}\int\int \frac{H(z)^2}{g(z)} K(u+\frac{bz}{h})^2 u dz du &= \int\int \frac{H(z)^2}{g(z)} K(u+\frac{bz}{h})^2 (u+\frac{bz}{h}) dz du - \int\int \frac{H(z)^2}{g(z)} K(u+\frac{bz}{h})^2 \frac{bz}{h} dz du \\ &= \int \frac{H(z)^2}{g(z)} \{\int K(u+\frac{bz}{h})^2 (u+\frac{bz}{h}) du\} dz - \int \frac{H(z)^2}{g(z)} \frac{bz}{h} \{\int K(u+\frac{bz}{h})^2 du\} dz \\ &= \mu_{12}(K) \int \frac{H(z)^2}{g(z)} dz - R(K) \int \frac{H(z)^2}{g(z)} \frac{bz}{h} dz = 0\end{aligned}$$

noting $\mu_{12}(K) = 0$, if K is symmetric. And $\int \frac{H(z)^2}{g(z)} \frac{bz}{h} dz = 0$, if $\frac{H(z)^2}{g(z)}$ is symmetric (or simply both H and g are symmetric). For the second relationship (21),

$$\begin{aligned}\int\int \frac{H(z)^2}{g(z)} K(u+\frac{bz}{h})^2 u^2 dz du &= \int \frac{H(z)^2}{g(z)} \int K(u+\frac{bz}{h})^2 \{(u+\frac{bz}{h}) - \frac{bz}{h}\}^2 du dz \\ &= \int \frac{H(z)^2}{g(z)} \int K(u+\frac{bz}{h})^2 \{(u+\frac{bz}{h})^2 - 2(u+\frac{bz}{h})\frac{bz}{h} + (\frac{bz}{h})^2\} du dz \\ &= \int \frac{H(z)^2}{g(z)} \{\mu_{22}(K) - 2\frac{bz}{h}\mu_{12}(K) + (\frac{bz}{h})^2 R(K)\} dz \\ &= \mu_{22}(K) \int \frac{H(z)^2}{g(z)} dz + R(K) \int (\frac{bz}{h})^2 \frac{K(z)^2}{g(z)} dz \\ &\geq \mu_{22}(K) + R(K) \int (\frac{bz}{h})^2 \frac{H(z)^2}{g(z)} dz \\ &> \mu_{22}(K),\end{aligned}$$

where the second last inequality uses (19). The second term of the second last line comes from the smoothing component $\{\varepsilon_i\}$ in view that $\int (\frac{bz}{h})^2 \frac{H(z)^2}{g(z)} dz = \int (\frac{bz}{h})^2 H(z) dz = (\frac{b}{h})^2 Var(\varepsilon_i)$ when $g = H$. Also, if we use standard nonparametric bootstrap, not smoothed bootstrap, namely $Var(\varepsilon_i) = 0$, this term disappears and the result coincides with the second term of (16).

B Proof of Results

B.1 Proof of Theorem 1

We have

$$\begin{aligned}E\{\hat{B}(x)\} &= \frac{h^L E\{f^{(L)}_{H,b}(x)\}}{L!} \int u^L K(u) du \\ &= \frac{h^L \mu_{L1}(K)}{L!} E\{f^{(L)}_{H,b}(x)\}\end{aligned}$$

and

$$
\begin{aligned}
E\{\hat{f}^{(L)}_{H,b}(x)\} &= \frac{1}{b^{L+1}} E\{H^{(L)}(\frac{X_1 - x}{b})\} \\
&= \frac{1}{b^{L+1}} \int H^{(L)}(\frac{z-x}{b}) f(z) dz \\
&= \frac{1}{b^{L}} \int H^{(L)}(u) f(x+bu) du \\
&= \int H(u) f^{(L)}(x+bu) du \\
&= \int H(u) \sum_{k=0}^{L'} \frac{(bu)^k}{k!} f^{(L+k)}(x) du + O(b^{L'+1}) \\
&= f^{(L)}(x) + \frac{\mu_{L'1}(H) b^{L'}}{L'!} f^{(L+L')}(x) + O(b^{L'+1}).
\end{aligned}
$$

Therefore

$$
E\{\hat{B}(x)\} = \frac{h^L \mu_{L1}(K)}{L!} f^{(L)}(x) + \frac{h^L \mu_{L1}(K)}{L!} \frac{\mu_{L'1}(H) b^{L'}}{L'!} f^{(L+L')}(x) + O(h^L b^{L'+1}),
$$

while the bias is

$$
\begin{aligned}
E\{\hat{f}_{K,h}(x)\} - f(x)] &= \int K(u)\{f(x+hu) - f(x)\} du \\
&= \sum_{k=1}^{L+L'} \frac{\mu_{k1}(K) h^k}{k!} f^{(k)}(x) + O(h^{2L+1}) \\
&= \sum_{k=L}^{L+L'} \frac{\mu_{k1}(K) h^k}{k!} f^{(k)}(x) + O(h^{2L+1}).
\end{aligned}
$$

This yields,

$$
E\{\hat{B}(x)\} - E\{\hat{f}_{K,h}(x)\} - f(x)] = O(h^{L+1} + h^L b^{L'}). \tag{22}
$$

The variance of $\hat{B}(x)$ is

$$\begin{aligned}
Var\{\hat{B}(x)\} &= \frac{h^{2L}\mu_{L1}(K)^2}{(L!)^2} var\{f_{H,b}^{(L)}\} \\
&= \frac{h^{2L}\mu_{L1}(K)^2}{nb^{2L+2}(L!)^2} var\{H^{(L)}(\frac{X_i - x}{b})\} \\
&= \frac{h^{2L}\mu_{L1}(K)^2}{n(L!)^2}\Big(\frac{1}{b^{2L+2}}E\{H^{(L)}(\frac{X_i - x}{b})^2\} - (\frac{1}{b^{L+1}}E\{H^{(L)}(\frac{X_i - x}{b})\})^2\Big) \\
&= \frac{h^{2L}\mu_{L1}(K)^2}{n(L!)^2}\Big(\frac{1}{b^{2L+2}}\int H^{(L)}(\frac{z-x}{b})^2 f(z)dz - f^{(L)}(x) + o(1)\Big) \\
&= \frac{h^{2L}\mu_{L1}(K)^2}{n(L!)^2}\frac{1}{b^{2L+1}}\int H^{(L)}(u)^2 f(x+bu)du + O(\frac{h^{2L}}{n}) \\
&= \frac{h^{2L}\mu_{L1}(K)^2}{n(L!)^2}\frac{f(x)}{b^{2L+1}}\int H^{(L)}(u)^2 du + O(\frac{h^{2L}}{n}) \\
&= \frac{h^{2L}\mu_{L1}(K)^2 R(H^{(L)}) f(x)}{nb^{2L+1}(L!)^2} + O(\frac{h^{2L}}{n}) \qquad (23)
\end{aligned}$$

B.2 Proof of Theorem 2

We have

$$\begin{aligned}
& E\{\hat{B}^+(x)\} - [E\{\hat{f}_{K,h}(x)\} - f(x)] \\
=& E\Big[\int K(u)\{\hat{f}_{H,b}(x+hu) - \hat{f}_{H,b}(x)\}du\Big] - \Big[E\{\hat{f}_{K,h}(x)\} - f(x)\Big] \\
=& \int K(u)\int \frac{1}{b}H(\frac{z-(x+hu)}{b})f(z)dzdu - \int K(u)\int \frac{1}{b}H(\frac{z-x}{b})f(z)dzdu - [\int \frac{1}{h}K(\frac{z-x}{h})f(z)dz - f(x)] \\
=& \int\int K(u)H(v)f(x+hu+bv)dvdu - \int K(u)H(v)f(x+bv)dvdu - [\int K(u)f(x+hu)du - f(x)] \\
=& \int\int K(u)H(v)\{f(x+hu+bv) - f(x+bv)\}dvdu - \int K(u)\{f(x+hu) - f(x)\}du \\
=& (A) - (B).
\end{aligned}$$

We first evaluate (B) by Taylor expansion as

$$\begin{aligned}
(B) &= \int K(u)\sum_{l=1}^{L+L'}\frac{h^l u^l}{l!}f^{(l)}(x)du + O(h^{L+L'+1}) \\
&= \int K(u)\sum_{l=L}^{L+L'}\frac{h^l u^l}{l!}f^{(l)}(x)du + O(h^{L+L'+1}) \\
&= \sum_{l=L}^{L+L'}\frac{\mu_{l1}(K)h^l}{l!}f^{(l)}(x) + O(h^{L+L'+1})
\end{aligned}$$

where second equality holds because $K(\cdot)$ is an L-th order kernel function. Similarly,

$$\begin{aligned}
(A) &= \int\int K(u)H(v)\sum_{l=0}^{L+L'}\frac{h^l u^l}{l!}f^{(l)}(x+bv)dvdu + O(h^{L+L'+1}) \\
&= \int\int K(u)H(v)\sum_{l=L}^{L+L'}\frac{h^l u^l}{l!}f^{(l)}(x+bv)dvdu + O(h^{L+L'+1}) \\
&= \int\int K(u)H(v)\sum_{l=L}^{L+L'}\frac{h^l u^l}{l!}\sum_{k=0}^{L'}\frac{b^k v^k f^{(l+k)}(x)}{k!}dvdu + O(h^L b^{L'+1}) + O(h^{L+L'+1}) \\
&= \int\int K(u)H(v)\sum_{l=L}^{L+L'}\frac{h^l u^l}{l!}\Big(f^{(l)}(x) + \frac{b^{L'}v^{L'}f^{(l+L')}(x)}{L'!}\Big)dvdu + O(h^L b^{L'+1}) \\
&= \int\int K(u)H(v)\Big(\sum_{l=L}^{L+L'}\frac{h^l u^l}{l!}f^{(l)}(x) + \frac{h^L u^L b^{L'} v^{L'} f^{(L+L')}(x)}{L'!}\Big)dvdu + O(h^L b^{L'+1}) \\
&= \sum_{l=L}^{L+L'}\frac{\mu_{l1}(K)h^l}{l!}f^{(l)}(x) + \frac{\mu_{L1}(K)\mu_{L'1}(H)h^L b^{L'} f^{(L+L')}(x)}{L!L'!} + O(h^L b^{L'+1})
\end{aligned}$$

Therefore,

$$E\{\hat{B}^+(x)\} - [E\{\hat{f}_{K,h}(x)\} - f(x)] = \frac{\mu_{L1}(K)\mu_{L'1}(H)h^L b^{L'} f^{(L+L')}(x)}{L!L'!} + O(h^L b^{L'+1}). \qquad (24)$$

Next we compute the variance of $\int K(u)\{\hat{f}_{H,b}(x+hu) - \hat{f}_{H,b}(x)\}du$.
Write

$$\begin{aligned}
\int K(u)\{\hat{f}_{H,b}(x+hu) - \hat{f}_{H,b}(x)\}du &= \int K(u)\{\frac{1}{nb}\sum_{i=1}^{n}H(\frac{X_i-(x+hu)}{b}) - \frac{1}{nb}\sum_{i=1}^{n}H(\frac{X_i-x}{b})\}du \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{b}\int K(u)\{H(\frac{X_i-(x+hu)}{b}) - H(\frac{X_i-x}{b})\}du.
\end{aligned}$$

then, using Eq. (24), we have

$$\begin{aligned}
Var(\hat{B}^+(x)) &= \frac{1}{n}Var[\frac{1}{b}\int K(u)\{H(\frac{X_1-(x+hu)}{b}) - H(\frac{X_1-x}{b})\}du] \\
&= \frac{1}{nb^2}E[\int K(u)\{K(\frac{X_1-(x+hu)}{b}) - H(\frac{X_1-x}{b})\}du]^2 - \frac{1}{n}[E\{\hat{B}^+(x)\}]^2 \\
&= \frac{1}{nb^2}\int[\int K(u)\{H(\frac{z-x+hu}{b}) - H(\frac{z-x}{b})\}du]^2 f(z)dz - O\big(\frac{h^{2L}}{n}\big) \\
&= \frac{1}{nb}\int[\int K(u)\{H(\frac{bv+hu}{b}) - H(v)\}du]^2 f(x+bv)dv - O\big(\frac{h^{2L}}{n}\big)
\end{aligned}$$

If $h/b \to \infty$, $H([bv+hu]/b) \to 0$, hence

$$Var(\hat{B}^+(x)) = \frac{1}{nb}\int[\int K(u)H(v)du]^2 f(x+bv)dv = \frac{R(H)f(x)}{nb} + O(\frac{1}{n}) \qquad (25)$$

If $h/b \to M < \infty$,

$$Var(\hat{B}^+(x)) = \frac{1}{nb}\int [\int K(u)H(v+Mu)du - H(v)]^2 f(x+bv)dv$$
$$= \frac{\int [\int K(u)H(v+Mu)du - H(v)]^2 dv f(x)}{nb} + O(\frac{1}{n}) \quad (26)$$

If $h/b \to 0$,

$$\int K(u)\{H(v+hu/b) - H(v)\}du = \int K(u)\{H(v) + \frac{huH'(v)}{b} + \cdots + \frac{h^L u^L H^{(L)}(v)}{b^L(L!)} + O(\frac{h^{L+1}}{b^{L+1}}) - H(v)\}du$$
$$= \int K(u)\{\frac{h^L u^L H^{(L)}(v)}{b^L(L!)} + O(\frac{h^{L+1}}{b^{L+1}})\}du$$
$$= \frac{h^L H^{(L)}(v)\mu_{L1}(K)}{b^L(L!)} + O(\frac{h^{L+1}}{b^{L+1}})$$

Therefore,

$$Var(\hat{B}^+(x)) = \frac{1}{nb}\int [\int K(u)\{H(v+hu/b) - H(v)\}du]^2 f(x+bv)dv - O(\frac{h^{2L}}{n})$$
$$= \frac{1}{nb}\int [\frac{h^L H^{(L)}(v)\mu_{L1}(K)}{b^L(L!)} + O(\frac{h^{L+1}}{b^{L+1}})]^2 f(x+bv)dv - O(\frac{h^{2L}}{n})$$
$$= \frac{1}{nb}\int [\frac{h^L H^{(L)}(v)\mu_{L1}(K)}{b^L(L!)}]^2 f(x+bv)dv + O(\frac{h^{2L+1}}{nb^{2L+2}}) - O(\frac{h^{2L}}{n})$$
$$= \frac{h^{2L} f(x) R(H^{(L)})\mu_{L1}(K)^2}{nb^{2L+1}(L!)^2} + O(\frac{h^{2L+1}}{nb^{2L+2}}) \quad (27)$$

Hence,

$$Var(\hat{B}^+(x)) = O(\max\{\frac{1}{nb}, \frac{h^{2L}}{nb^{2L+1}}\})$$

C AMISE and optimal bandwidth

As shown in Appendix B.2, for a fixed b, the bias and variance of $\hat{B}^+(x)$ under $h/b \to 0$ is much smaller than that under $h/b \to \infty$ or $h/b \to M < \infty$. Hence, we only consider the $h/b \to 0$ here.

The mean squared error of $\hat{B}^+(x)$ is

$$MSE(\hat{B}^+(x)) = \big(Bias(\hat{B}^+(x))\big)^2 + Var(\hat{B}^+(x))$$
$$= \frac{h^{2L} b^{2L'} \mu_{L1}(K)^2 \mu_{L'1}(H)^2 \big(f^{(L+L')}(x)\big)^2}{\big(L!L'!\big)^2} + \frac{h^{2L} f(x) R(H^{(L)})\mu_{L1}(K)^2}{nb^{2L+1}(L!)^2}$$
$$+ O(max\{h^{2L} b^{2L'+1}, \frac{h^{2L+1}}{nb^{2L+2}}\})$$
$$\simeq \frac{h^{2L} b^{2L'} \mu_{L1}(K)^2 \mu_{L'1}(H)^2 \big(f^{(L+L')}(x)\big)^2}{\big(L!L'!\big)^2} + \frac{h^{2L} f(x) R(H^{(L)})\mu_{L1}(K)^2}{nb^{2L+1}(L!)^2}$$
$$= AMSE(\hat{B}^+(x))$$

Hence, the asymptotic mean integrated squared error,

$$\begin{aligned}
AMISE(\hat{B}^{+}(x)) &= \int \Big(\frac{h^{2L} b^{2L'} \mu_{L1}(K)^2 \mu_{L'1}(H)^2 \big(f^{(L+L')}(x)\big)^2}{\big(L!L'!\big)^2} + \frac{h^{2L} f(x) R(H^{(L)}) \mu_{L1}(K)^2}{nb^{2L+1}(L!)^2} \Big) dx \\
&= \frac{h^{2L} b^{2L'} \mu_{L1}(K)^2 \mu_{L'1}(H)^2 R\big(f^{(L+L')}\big)}{\big(L!L'!\big)^2} + \frac{h^{2L} R(H^{(L)}) \mu_{L1}(K)^2}{nb^{2L+1}(L!)^2} \\
&= \frac{h^{2L} \mu_{L1}(K)^2}{(L!)^2} \Big(\frac{b^{2L'} \mu_{L'1}(H)^2 R\big(f^{(L+L')}\big)}{(L'!)^2} + \frac{R(H^{(L)})}{nb^{2L+1}} \Big)
\end{aligned}$$

Take the derivative of the $AMISE$ with respect to b and setting it equal to zero,

$$\begin{aligned}
\frac{d}{db} AMISE(\hat{B}^{+}(x)) &= \frac{d}{db} \Big[\frac{h^{2L} \mu_{L1}(K)^2}{(L!)^2} \Big(\frac{b^{2L'} \mu_{L'1}(H)^2 R\big(f^{(L+L')}\big)}{(L'!)^2} + \frac{R(H^{(L)})}{nb^{2L+1}} \Big) \Big] \\
&= \frac{h^{2L} \mu_{L1}(K)^2}{(L!)^2} \Big(\frac{2L' b^{2L'-1} \mu_{L'1}(H)^2 R\big(f^{(L+L')}\big)}{(L'!)^2} - \frac{(2L+1) R(H^{(L)})}{nb^{2L+2}} \Big) \\
&= 0
\end{aligned}$$

The solution is

$$\tilde{b}_{opt} = \Big(\frac{(L'!)^2 (2L+1) R(H^{(L)})}{2nL' \mu_{L'1}(H)^2 R\big(f^{(L+L')}\big)} \Big)^{\frac{1}{2L+2L'+1}} \tag{28}$$

which coincides with the optimal bandwidth selection for dnsity derivative estimation.