

Asymptotic validity of bootstrap confidence intervals in nonparametric regression without an additive model

Liang Wang and Dimitris N. Politis

*Department of Mathematics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, USA 92093*
e-mail: wangliang6111@gmail.com; dpolit@ucsd.edu

Abstract: Bootstrap for nonparametric regression has been around for more than 30 years. Nevertheless, most results are based on assuming an additive regression model with respect to independent and identical (i.i.d.) errors. An exception is the Local Bootstrap of Shi [23] for which, however, no bootstrap consistency results are available. We attempt to remedy this here while at the same time showing bootstrap consistency for a more general class of methods that fall under the heading of Model-free Bootstrap of Politis [18].

Keywords and phrases: Non-additive regression model, model-free bootstrap, local bootstrap, bootstrap confidence interval, heteroscedasticity.

Received May 2019.

1. Introduction

Consider regression data $\{(Y_i, X_i)\}, i = 1, \dots, n$ which are independent and identical (i.i.d.) pairs. The response Y_i is real-valued; for simplicity, we also suppose that X_i is univariate and random with density function $f(x)$, but the method works in the same way when X_i is multivariate. Just assuming that the pairs $\{(Y_i, X_i)\}, i = 1, \dots, n$ are i.i.d. can be considered a “Model-free” regression setup where interest lies on nonparametric estimation of the conditional expectation $m(x) = E(Y_i|X_i = x)$.

Nevertheless, often practitioners assume a model equation that relates Y_i to X_i ; this would be a “model-based” setup despite that fact that the model could be of nonparametric nature. A typical model for nonparametric regression with i.i.d. errors is:

$$Y_i = m(X_i) + \sigma \cdot \epsilon_i, \quad i = 1, \dots, n,$$

or, more generally,

$$Y_i = m(X_i) + \sigma(X_i) \cdot \epsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

where the ϵ_i 's are assumed i.i.d. with mean 0 and variance 1. As in the Model-free case, $m(x) = E(Y|X = x)$ is the target of interest, and $\sigma^2(x) = \text{Var}(Y|X =$

x) allows for heteroscedasticity; the functions $m(\cdot)$ and $\sigma^2(\cdot)$ will be assumed unknown but smooth, i.e., a nonparametric setup.

Since Efron [3] pioneering paper, several resampling methods for nonparametric regression problem were developed. One of the first was the paper by Härdle and Bowman [5]; see also the book by Hall [4]. These methods are very powerful on additive model such as (1.1). An interesting procedure called the Local bootstrap was proposed early on by Shi [23]. The Local bootstrap was designed for the heteroscedastic model (1.1), and it was based on the estimator of Priestley and Chao [19]. Unfortunately its consistency was never proven. We attempt to remedy this here; in fact, we will show that the Local bootstrap is consistent under the Model-free setup of i.i.d. pairs without resort to a restrictive model equation such as (1.1). We will further show the consistency of a more general class of methods that fall under the heading of Model-free Bootstrap of Politis [18].

Before describing the notion of Model-free Bootstrap it is important to note that Efron's [3] well-known method of resampling pairs has a chance does not work appropriately in nonparametric regression. The caveat has to do with nonparametric estimation of $m(x) = E(Y_i|X_i = x)$ for some particular x of interest. Although there may be many points with regressor value near x on the original scatterplot, there is no guarantee that there will be *any* points with regressor value near x on the bootstrap scatterplot, rendering nonparametric estimation (which is a local method) vacuous in the bootstrap world. More discussion see Remark 1.1 below.

Remark 1.1. Regarding to bootstrap pairs, as one example, suppose $X_i = i/n, i = 1, \dots, n$ (discrete grid), in addition suppose we use a bandwidth that is equivalent to using the m nearest neighbors, i.e. estimating $f(a)$ by a weighted average of the $2m + 1$ values in a window centered at a , where a is a value of interest on the grid. Assume that a is an interior point (i.e., $0 < a - m/n$ and $a + m/n < 1$), and that $2m + 1 < n$; then using the pairs bootstrap to create (Y_i^*, X_i^*) for $i = 1, \dots, n$, the probability that not one of the X_i^* for $i = 1, \dots, n$ falls in the local window $[a - m/n, a + m/n]$ is $((n - (2m + 1))/n)^n$ which, for large n , can be approximated by $\exp(-2m - 1)$. When this situation occurs, we can not construct the estimate of $f(a)$ in the bootstrap world, as there are no data in the local neighborhood of a . This might appear extreme, but a similar problems occurs when only a few X^* fall in the local window. Notably, if m becomes large, $\exp(-2m - 1)$ will become smaller but it is always nonzero. To perform bootstrap, we typically draw B resamples, each of size n , where B is a large number. It is easy to see that the probability of the extreme situation (no X^* in the local window) occurring in at least one of the B resamples (and thus rendering the whole simulation vacuous) is: $1 - [1 - \exp(-2m - 1)]^B$, which can be appreciable even for m large.

The notion of Model-free Bootstrap was proposed in a regression context by Politis [13], and further developed by Sperlich [24], Politis [14, 17, 16]. The basic idea is to transform a non-i.i.d. dataset such as the set of responses Y_1, \dots, Y_n to an i.i.d. dataset that can be resampled in the usual way. The Model-free Boot-

strap is described in detail in the recent book by Politis [18] but no consistency results have been proven so far. In this paper we will provide the first theoretical results on Model-free bootstrap consistency, and we will further compare to the Local bootstrap by simulation. We also reveal a close relationship between the two methods under specified assumptions, i.e., the Local bootstrap is essentially a special case of the Model-free bootstrap.

Although both the Local bootstrap and the Model-free bootstrap can be implemented using any form of nonparametric regression estimators, to fix ideas in this paper we focus on the Nadaraya-Watson (N-W) kernel estimator [11]:

$$\hat{m}_{n,h}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i-x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)}, \quad (1.2)$$

whose properties have been very well studied [2, 7, 10, 8]. E.g., the conditional cumulative distribution function (CDF) and the quantile estimator based on N-W estimator were studied by Li and Racine [9] and Qu and Yoon [20].

The following is a brief introduction to the spirit of Local bootstrap and Model-free method in regression problem; more details can be found in sections 3 and 4.

- (i) **Local bootstrap.** Instead of resampling data in an i.i.d. fashion, the Local bootstrap method resamples data independently but not identically distributed. More specifically, the probability mass function (p.m.f.) that the Local bootstrap uses is derived by the N-W estimator first. WLOG, suppose we have n i.i.d. pairs $(X_i, Y_i), i = 1, \dots, n$ at our disposal. To each X_i , the bootstrap will assign a bootstrap response Y_i^* which is selected from one of the n responses, i.e., Y_1, \dots, Y_n , according to the p.m.f.
- (ii) **Model-free bootstrap.** The idea of Model-free bootstrap is transforming the non-i.i.d. data to i.i.d. data first, and then applying i.i.d. resampling. More particularly for regression problem, by probability integral transform theory, we can transform each observation Y_i to approximately $U(0,1)$ random variable U_i . Ideally U_i 's are i.i.d., thus we can resample U_i independently to obtain bootstrap data U_i^+ . The last step is using estimated quantile function of $Y|x$ to transform U_i^+ 's back, and denote this final bootstrap data at different X_i by Y_i^+ .

The rest of the paper is organized as follows. In section 2, we address all the necessary assumptions for the theoretical analysis. Section 3 and 4 introduce formal definition of local bootstrap and model free bootstrap as well as important lemmas. Main results and theorems are presented in section 5. In section 6 we discuss the details of parameter selection, like the choice of transformed function and bandwidth. Simulation results are presented in section 7. Proofs of most lemmas and theorems can be found in the Appendix.

2. Assumptions and problem set up

This paper is motivated to prove consistency of Model-free bootstrap method and Local bootstrap method without the assumption of an additive model; instead we assume that the data $\{(Y_i, X_i)\}, i = 1, \dots, n$ are i.i.d. pairs. Recall definition of N-W estimator (1.2), and denote the density function of independent random variables X_1, \dots, X_n by $f(x)$, and the function of interest $m(x) = E(Y|X = x)$. Let $f_{Y|X}(y|x)$ be the conditional probability density function (PDF) of Y given $X = x$, $D(y|x)$ is the conditional cumulative density function (CDF) of Y given $X = x$, i.e. $D(y|x) = \mathbb{P}(Y \leq y|X = x)$.

Consider the following assumptions:

Assumption A for r.v. X :

- (i) EX_i^2 is finite;
- (ii) $f(x)$ is twice differentiable, and both first and second order derivative functions satisfy the Lipschitz condition $|g(x) - g(y)| \leq C|x - y|$ for some $C > 0$; here g denotes either f' or f'' ;
- (iii) $m(x)$ is twice differentiable, and both first and second order derivative functions satisfy the Lipschitz condition $|g(x) - g(y)| \leq C|x - y|$ for some $C > 0$; here g denotes either f' or f'' ;
- (iv) $\sigma^2(x) = \text{Var}(Y_i|X_i = x)$ is a continuous function of x ;
- (v) $\forall x \in \text{support of } X$, there exists a compact set $\mathcal{S} \subset \text{support of } X$, such that $\inf_{x \in \mathcal{S}} f(x) \geq \delta > 0$.

Assumption B for r.v. Y :

- (i) $f_{Y|X}(y|x)$ is twice differentiable with respect to x ;
- (ii) There exist $B_1(y) > 0, B_2(y) > 0$, such that $\frac{\partial f(y|x)}{\partial x} \leq B_1(y), \frac{\partial^2 f(y|x)}{\partial x^2} \leq B_2(y)$, and $\int B_i(y)dy < \infty, i = 1, 2$;
- (iii) $E(|Y|^3|X = x) < \infty$ for all x ;
- (iv) $\text{Var}(|Y_i|^3|X_i = x) < \infty$ for all x ;
- (v) $D(y|x) > 0$ and is twice differentiable with respect to both y and x in \mathcal{S} .

Assumption C for kernel function $K(\cdot)$:

- (i) $K(\cdot)$ is a symmetric, bounded and twice differentiable function;
- (ii) $K(x) \geq 0, \int K(x)dx = 1$;
- (iii) $\int x^2 K(x)dx < \infty$;
- (iv) $K(\cdot)$ is Lipschitz continuous. Denote the Lipschitz constant by C_1 .

Assumption D for sample size n and bandwidth h :

- (i) $n \rightarrow \infty, h \rightarrow 0, nh \rightarrow \infty$;
- (ii) $h = o(n^{-\frac{1}{5}})$.

Li and Racine [8] proved the following theorems under the assumption of i.i.d. pairs, Assumption A (i)-(iv), Assumption C (i)-(iii), and Assumption D (i) for estimator (1.2):

$$E[\hat{m}_{n,h}(x)] = m(x) + h^2 B_s(x) + O(h^3), \quad (2.1)$$

$$\text{Var}[\hat{m}_{n,h}(x)] = \frac{1}{nh} \frac{\kappa\sigma^2(x)}{f(x)} + O\left(\frac{h}{n}\right), \quad (2.2)$$

$$\sqrt{nh}(\hat{m}_{n,h}(x) - m(x) - h^2 B_s(x)) \xrightarrow{d} N\left(0, \frac{\kappa\sigma^2(x)}{f(x)}\right), \quad (2.3)$$

where

$$B_s(x) = \frac{\kappa_2}{2} \frac{2f'(x)m'(x) + f(x)m''(x)}{f(x)}, \quad (2.4)$$

and $\kappa_2 = \int v^2 K(v)dv$, $\kappa = \int K^2(v)dv$.

Remark 2.1. Motivations of some assumptions:

- a. In the following sections, we need the same assumptions Li and Racine [8] used, only with Assumption A (iii) and (iv) excluded. Additional assumptions, which are slightly stronger and play similar role in proof, are introduced in Assumption B (i) and (ii). The motivation is to show validity of bootstrap methods without additive model, we need to not only assume $m(x) = E(Y|X = x)$ and $\sigma^2(x) = \text{Var}(Y|X = x)$ are smooth with respect to x , but also assume $f_{Y|X}(y|x)$ is smooth with respect to x . Assumption B (ii) will allow us to apply Dominated Convergence Theorem.;
- b. Assumption A (iv), B (v) and C (iv) help to show uniform almost surely convergence for an estimator of $D(y|x)$, more details please see Section 4;
- c. Assumption B (iii) and (iv) are essentially conditions of Lyapunov central limit theorem applied in Section 5;
- d. Assumption D (ii) is essentially under smoothing technique.

3. Local bootstrap

The Local bootstrap method proposed by Shi [23] was originally for the scenario that the regressors x_i 's are deterministic design points in $[0, 1]$. In what follows, we address the condition that regressors X_i are random variables. The procedure of local bootstrap can be described as follows:

- (i) For each observation X_i , we create an estimator of conditional probability mass function (p.m.f.) of Y given X_i :

$$\hat{G}_{X_i} : \begin{pmatrix} Y_1 & Y_2 & \dots & Y_n \\ w_{1i} & w_{2i} & \dots & w_{ni} \end{pmatrix}, \quad (3.1)$$

i.e. we estimate the probability distribution of Y given X_i as a discrete dis-

tribution taking value Y_j with probability w_{ji} , where $w_{ji} = \frac{K\left(\frac{x_j - x_i}{h}\right)}{\sum_{l=1}^n K\left(\frac{x_l - x_i}{h}\right)}$

for $i, j = 1, 2, \dots, n$, so $\sum_j w_{ji} = 1$.

- (ii) At each X_i , we resample Y_i^* from distribution \hat{G}_{X_i} .

(iii) Compute bootstrap kernel estimator:

$$\hat{m}_{n,h}^*(x) = \frac{\sum_{i=1}^n Y_i^* K\left(\frac{X_i-x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)}. \quad (3.2)$$

Shi [23] provided theoretical analyses on Y_i^* but not the asymptotic properties of $\hat{m}_{n,h}^*(x)$; This gap is filled in section 5 but needs the following lemma and theorem presented in rest of current section.

Lemma 3.1. *Under Assumption B (i)-(iii) and Assumption C (i)-(iii),*

$$\mathbb{E}_{X_i, Y_i} \left[g(Y_i) K^l \left(\frac{X_i - x}{h} \right) \right] = h \left\{ A(x) \mathbb{E}[g(Y)|X = x] + O(h^2) \right\}, \quad (3.3)$$

where $g(\cdot)$ is a continuous function (in this paper, $g(x) = x^k$ or $g(x) = |x|^k$, $k = 1, 2, 3$), $A(x) = f_X(x) \int K^l(u) du$, and l is a fixed positive integer, $f_X(x)$ is density function of X_1, \dots, X_n . Also, \mathbb{E}_{X_i, Y_i} denote the joint expectation of (X_i, Y_i) .

Notice i disappears on the right side of (3.3) and thus $O(h^2)$ is uniform for $i = 1, \dots, n$. This lemma is also valid if replacing Y_i by Y_i^* , and $O(\cdot)$ by $O_p(\cdot)$ under the same assumptions:

$$\mathbb{E}_{X_i, Y_i^*} \left[g(Y_i^*) K^l \left(\frac{X_i - x}{h} \right) \right] = h \left\{ A(x) \mathbb{E}^*[g(Y^*)|X = x] + O_p(h^2) \right\}, \quad (3.4)$$

where \mathbb{E}_{X_i, Y_i^*} is the expectation in bootstrap world, and thus is conditional expectation in real world. Lemma 3.1 and (3.4) then lead to the following preliminary result:

Theorem 3.2. *Under Assumption B (i)-(iv) and Assumption C (i)-(iii),*

$$|\mathbb{E}^*[g(Y^*)|X = x] - \mathbb{E}[g(Y)|X = x]| = O_p \left(h^2 + \frac{1}{\sqrt{nh}} \right), \quad (3.5)$$

where $g(\cdot)$ is the same continuous function as the one in lemma 3.1, and similarly to equation (3.4), \mathbb{E}^* is the expectation in bootstrap world, and thus is conditional expectation in real world.

Notice Shi [23] has a similar result for the Priestley and Chao Estimator [19] for the scenario regressors x_i 's are deterministic design points in $[0, 1]$. Theorem 3.2 here holds for the Nadaraya-Watson estimator under the condition $(X_i, Y_i), i = 1, 2, \dots$ are i.i.d. pairs.

4. Model-free bootstrap

The principle of all bootstrap methods is imitating the original sampling that occurred in real world. If the data are sampled i.i.d., then bootstrap method resamples data i.i.d.. The problem local bootstrap method solves is the response

data are collected independently but not identically distributed. As a solution, the local bootstrap assigns different probabilities to the observed responses.

The Model-free bootstrap of Politis [13] is using a different technique: transforming the non-i.i.d. data to i.i.d. (or approximately i.i.d.), and then performing i.i.d. resampling. In the Model-free setup of i.i.d. pairs, the transformation could be achieved by the probability integral transform [1]. More specifically, denote the conditional distribution function $D_x(\cdot)$ of Y as

$$D_x(y) = P(Y \leq y | X = x), \quad (4.1)$$

and assume it is continuous with respect to y ; then, $D_x(Y)$ follows Unif(0,1) distribution. Since the true conditional distribution $D_x(y)$ is unknown, it can be estimated by a N-W estimator, i.e.,

$$\widehat{D}_x(y) = \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i \leq y\}} K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}. \quad (4.2)$$

Estimator (4.2) is consistent (see Li and Racine [8]) but not continuous in y . A doubly smoothed version of (4.2) is defined by

$$\widetilde{D}_x(y) = \frac{\sum_{i=1}^n \Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}, \quad (4.3)$$

where $\Lambda(\cdot)$ is a continuous and strictly increasing CDF over its support. Now define two estimators of quantile-inverse function:

$$\widehat{D}_x^{-1}(u) = \min_{i=1, \dots, n} \{Y_i : u \leq \widehat{D}_x(Y_i)\}, \quad (4.4)$$

and

$$= \inf_y \{y : u \leq \widetilde{D}_x(y)\}. \quad (4.5)$$

Both $\widehat{D}_x^{-1}(\cdot)$ and $\widetilde{D}_x^{-1}(\cdot)$ have good properties, see [13, 14]; more discussions are given in section 6. To specify the ideas, we apply $\widetilde{D}_x^{-1}(\cdot)$ defined in (4.5) in all our theoretical results; the same results are expected to hold if one uses $\widehat{D}_x^{-1}(\cdot)$ instead of $\widetilde{D}_x^{-1}(\cdot)$.

In this section, we denote Model-free bootstrap data by Y^+ , to distinguish from Local bootstrap data Y^* . Politis [13] proposed three ways to generate Model-free bootstrap data Y_1^+, \dots, Y_n^+ associated with the same original regressors X_1, \dots, X_n .

Method 1. Obtain transformed data U_1, \dots, U_n by computing $U_i = \widetilde{D}_{X_i}(Y_i)$. Then sample randomly from the transformed data U_1, \dots, U_n to get bootstrap pseudo-data u_1^+, \dots, u_n^+ . Finally $Y_i^+ = \widetilde{D}_{X_i}^{-1}(u_i^+)$, for $i = 1, \dots, n$.

Method 2. Generate bootstrap pseudo-data u_1^+, \dots, u_n^+ i.i.d. from $\text{Unif}(0,1)$ distribution, then $Y_i^+ = \hat{D}_{X_i}^{-1}(u_i^+)$, for $i = 1, \dots, n$.

Method 3. Let $\tilde{D}_x^{(t)}$ denote the estimator $\tilde{D}_x(\cdot)$ as computed from the delete- (X_t, Y_t) dataset, i.e., $\{(Y_i, X_i), i = 1, \dots, t - 1, t + 1, \dots, n\}$. Now define the “predictive” residuals as

$$u_t^{(t)} = \tilde{D}_{X_t}^{(t)}(Y_t), \quad \text{for } t = 1, \dots, n.$$

Then sample randomly from predictive u -data $u_1^{(1)}, \dots, u_n^{(n)}$ to create bootstrap pseudo-data u_1^+, \dots, u_n^+ , which leads to $Y_i^+ = \tilde{D}_{X_i}^{-1}(u_i^+)$, for $i = 1, \dots, n$.

Remark 4.1. In the rest of this paper, we will follow the notation given by Politis [18], i.e., denote Method 1 by the term Model-free (MF), Method 2 by the term Limit Model-free (LMF), and Method 3 by the term Predictive Model-free (PMF).

Remark 4.2. For LMF with $Y_i^+ = \hat{D}_{X_i}^{-1}(u_i^+)$, notice $\hat{D}_{X_i}(\cdot)$ and $\hat{D}_{X_i}^{-1}(\cdot)$ are both step functions. It implies LMF is essentially a method resampling Y_i^+ from Y_i with a p.m.f. \hat{H}_{X_i} . Further discussion please see theorem 4.4 and its proof. We can also modify LMF by using $Y_i^+ = \tilde{D}_{X_i}^{-1}(u_i^+)$, and expect it has similar property, but the discussion is omitted due to space limit.

Remark 4.3. Let x_f is the point of interest, and denote $\tilde{m}_{n,h}(x_f)$ the estimator proposed by Politis [13], i.e.

$$\tilde{m}_{n,h}(x_f) = \frac{1}{n} \sum_{i=1}^n \hat{D}_{x_f}^{-1}(u_i). \tag{4.6}$$

We noticed $\tilde{m}_{n,h}(x_f)$ and $\hat{m}_{n,h}(x_f)$ are numerically identical. In fact, by the definition of $\hat{D}_x(\cdot)$ (4.2) and $\hat{D}_x^{-1}(u)$ (4.4), the internal relationship of $\tilde{m}_{n,h}(x_f)$ and $\hat{m}_{n,h}(x_f)$ can be derived:

$$\begin{aligned} \hat{m}_{n,h}(x_f) &= \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x_f}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x_f}{h}\right)} \\ &= \int y d\hat{D}_{x_f}(y) \\ &= \int_0^1 \hat{D}_{x_f}^{-1}(u) du \\ &= E_U[\hat{D}_{x_f}^{-1}(U)], \end{aligned} \tag{4.7}$$

where u is a variable for integration, and U is random variable, whose expectation is denoted by E_U , conditional on x_f , and U follows uniform $[0,1]$ distribution. Notice (4.6) is approximation of (4.7), consequently $\tilde{m}_{n,h}(x_f)$ is also an approximation of $\hat{m}_{n,h}(x_f)$ if u_i 's are random variables from $\text{Unif}(0,1)$.

The following describe resampling algorithm to establish Model-free confidence intervals for $m(x_f)$:

- (a) Pick one Model-free method from MF, LMF and PMF, to obtain bootstrap data u_1^+, \dots, u_n^+ and Y_1^+, \dots, Y_n^+
- (b) Based on the pseudo data $\{(Y_i^+, X_i), i = 1, \dots, n\}$, re-estimate the conditional CDF $D_x(\cdot)$; Denote the bootstrap estimates by $\widehat{D}_x^+(\cdot)$ and $\widetilde{D}_x^+(\cdot)$ (which are unsmoothed and smoothed CDF estimator described earlier).
- (c) Give it a replicate of the bootstrap confidence interval root: $\widehat{m}_{n,h}(x_f) - \widehat{m}_{n,h}^+(x_f)$, or $\widetilde{m}_{n,h}(x_f) - \widetilde{m}_{n,h}^+(x_f)$, where

$$\widehat{m}_{n,h}^+(x_f) = \frac{\sum_{i=1}^n Y_i^+ K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}, \quad (4.8)$$

and

$$\widetilde{m}_{n,h}^+(x_f) = \frac{1}{n} \sum_{i=1}^n \widehat{D}_{x_f}^{+,-1}(u_i^+). \quad (4.9)$$

- (d) Steps (a)-(c) in the above are repeated B times, and the B bootstrap root replicates are collected in the form of an empirical distribution with α -quantile denoted by $q(\alpha)$.
- (e) Then, the Model-free $(1 - \alpha)100\%$ equal-tailed confidence interval for $m(x_f)$ is

$$[\widehat{m}_{n,h}(x_f) + q(\alpha/2), \widetilde{m}_{n,h}(x_f) + q(1 - \alpha/2)].$$

The following lemmas and theorems reveal MF, LMF and PMF method has the same rate of convergence to true $m(x)$ in probability as local bootstrap method. In fact, LMF is closely related to the local bootstrap.

Theorem 4.4. *LMF and local bootstrap method are equivalent. More specifically, for i.i.d. pairs $(X_i, Y_i), i = 1, 2, \dots$, bootstrapped $Y^*|X = X_i$ by local bootstrap method and bootstrapped $Y^+|X = X_i$ by LMF have exactly the same distribution.*

Notice the major difference from LMF to MF and PMF, are the algorithm how u_i^+ are resampled. LMF resamples u_i^+ directly from $\text{Unif}(0,1)$ distribution, while MF and PMF resample u_i^+ from transformed data U_1, \dots, U_n . The following theorem shows the intuition behind of MF and PMF by supposing U_1, \dots, U_n are $\text{Unif}(0,1)$ distributed. The setting of Theorem 4.5 is a hybrid between MF and LMF.

Theorem 4.5. *Recall Model free bootstrap method 1, i.e., MF method, but replace $U_i = \widehat{D}_{X_i}(Y_i)$ by random variable from $\text{Unif}(0,1)$. Still denote u_i^+ the bootstrapped data after each bootstrap iteration, and let $Y_i^+ = \widehat{D}_{X_i}^{-1}(u_i^+)$. Then under Assumption B (i)-(iv) and Assumption C (i)-(iii), a similar result as Theorem 3.2 holds, that is:*

$$|\mathbb{E}^+[g(Y^+)|X = x] - \mathbb{E}[g(Y)|X = x]| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right), \quad (4.10)$$

where $g(\cdot)$ is the same continuous function in lemma 3.1.

Lemma 4.6. Under Assumption A (v), Assumption B (i)-(v) and Assumption C (i)-(iv), we have:

$$\sup_{x \in \mathcal{S}} |\tilde{D}_x(y) - D_x(y)| = O\left(h_0^2 + h^2 + \left[\frac{\ln(n)}{nh}\right]^{\frac{1}{2}}\right), \quad \text{a.s.}$$

where $D_x(\cdot)$ is the real conditional CDF of Y , $\tilde{D}_x(y)$ is the estimated conditional CDF of Y , h and h_0 are two bandwidth parameters, and \mathcal{S} is any fixed compact set.

Lemma 4.6 and Polya’s Theorem lead to the following theorems:

Corollary 4.7. Under Assumption A (ii) (v), Assumption B (i)-(v) and Assumption C (i)-(iv), we have:

$$\sup_{x \in \mathcal{S}} \sup_y |\tilde{D}_x(y) - D_x(y)| = o(1) \text{ a.s.}$$

Remark 4.8. Recall MF is using $\tilde{D}_x(y)$ and PMF is using $\tilde{D}_x^{(t)}(y)$ where t th pair (X_t, Y_t) is deleted. By lemma 4.6, $\tilde{D}_x(y)$ and $\tilde{D}_x^{(t)}(y)$ have different orders of almost surely convergence. Fortunately, order of almost surely convergence is not necessary to show pointwise convergence of MF and PMF. We only need corollary 4.7 which holds and supports main results for both MF and PMF.

Lemma 4.9. $\forall \epsilon > 0$ and $q_0 \in (0, 1)$, $\exists N = N(\epsilon, q_0) > 0$ and $\delta = \delta(\epsilon, q_0)$, such that $\forall n > N$, if $|q - q_0| < \delta$, then

$$\left| \tilde{D}_x^{-1}(q) - \tilde{D}_x^{-1}(q_0) \right| < \epsilon \text{ a.s.}$$

Moreover, for a compact support $\mathcal{C} \subset (0, 1)$, $\forall q_0 \in \mathcal{C}$, $\exists N = N(\epsilon) > 0$ and $\delta = \delta(\epsilon)$, such that $\forall n > N$, $|q - q_0| < \delta$, then

$$\left| \tilde{D}_x^{-1}(q) - \tilde{D}_x^{-1}(q_0) \right| < \epsilon \text{ a.s.}$$

Theorem 4.10. For MF method, under Assumption A (ii) (v), Assumption B (i)-(v) and Assumption C (i)-(iv),

$$|\mathbb{E}^+[g(Y^+)|X = x] - \mathbb{E}[g(Y)|X = x]| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right),$$

where $g(\cdot)$ is the same continuous function in lemma 3.1.

Remark 4.11. Theorem 4.10 is valid for PMF as well. The proof is exactly the same, thus omitted.

Remark 4.12. The proof of the theorems above are inspired by [6, 10, 15, 21]. The theorems, lemmas and corollaries in this section are the cornerstone of the consistency results in the next section.

5. Main results

Although we provided theorem 3.2 and theorem 4.10 which show the consistency of bootstrapped sample mean, the main interest in this paper is consistency of bootstrapped N-W estimator (4.8) by local bootstrap method, LMF, MF and PMF. Before moving forward, the following notations and lemmas are needed:

$$\tilde{K}\left(\frac{X_i - x}{h}\right) = \frac{K\left(\frac{X_i - x}{h}\right)}{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right)} = \frac{K\left(\frac{X_i - x}{h}\right)}{\hat{f}_X(x)} = \frac{K\left(\frac{X_i - x}{h}\right)}{f_X(x) + o_p(1)}, \quad (5.1)$$

where $f_X(x)$ is the density function of X_1, \dots, X_n and $\hat{f}_X(x)$ is kernel estimator of $f_X(x)$, i.e.,

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right).$$

By equation (5.1), we have the corollary of Lemma 3.1:

Lemma 5.1. *Under Assumption B (i)-(iii) and Assumption C (i)-(iii),*

$$\mathbb{E}_{X_i, Y_i} \left[g(Y_i) \tilde{K}^l \left(\frac{X_i - x}{h} \right) \right] = h \left\{ \tilde{A}(x) \mathbb{E}[g(Y)|X = x] + O(h^2) \right\}, \quad (5.2)$$

where $g(\cdot)$ is the same function in lemma 3.1, $\tilde{A}(x) = \int \tilde{K}^l(u) du$, and l is any fixed positive integer.

Similarly to (3.4), this lemma is also valid if replacing Y_i by local bootstrapped Y_i^* or model free bootstrapped Y_i^+ . For example, for local bootstrapped Y_i^* :

$$\mathbb{E}_{X_i, Y_i^*} \left[g(Y_i^*) \tilde{K}^l \left(\frac{X_i - x}{h} \right) \right] = h \left\{ \tilde{A}(x) \mathbb{E}^*[g(Y^*)|X = x] + O(h^2) \right\},$$

where \mathbb{E}_{X_i, Y_i^*} is the expectation in bootstrap world, and thus is conditional expectation in real world. Let us re-write (1.2) and (4.8):

$$\hat{m}_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^n Y_i \tilde{K} \left(\frac{X_i - x}{h} \right), \quad (5.3)$$

$$\hat{m}_{n,h}^*(x) = \frac{1}{nh} \sum_{i=1}^n Y_i^* \tilde{K} \left(\frac{X_i - x}{h} \right). \quad (5.4)$$

The new form is equivalent to original N-W estimator. For the rest of this paper, we use superscript $*$ to denote local bootstrap and model free bootstrap for simplicity. Recall lemma 5.1, we have the following theorems:

Theorem 5.2. *Under Assumption A (ii) (v), Assumption B (i)-(v) and Assumption C (i)-(iv),*

$$|\mathbb{E}^*[\hat{m}_{n,h}^*(x)] - \hat{m}_{n,h}(x)| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right). \quad (5.5)$$

Theorem 5.3. Under Assumption A (ii) (v), Assumption B (i)-(v) and Assumption C (i)-(iv),

$$nh|\text{Var}^*[\hat{m}_{n,h}^*(x)] - \text{Var}[\hat{m}_{n,h}(x)]| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right). \tag{5.6}$$

Finally, by Lyapunov CLT, we can get:

Theorem 5.4. Under Assumption A (i)-(v), Assumption B (i)-(v), Assumption C (i)-(iv) and Assumption D (i), for any fixed x ,

$$\sqrt{nh}(\hat{m}_{n,h}^*(x) - E^*[\hat{m}_{n,h}^*(x)]) \xrightarrow{d} N\left(0, \frac{\kappa\sigma^2(x)}{f(x)}\right). \tag{5.7}$$

Recall equations (2.1) and (2.3), and by Theorem 5.4, we have:

$$\sup_u \left| P(\sqrt{nh}(\hat{m}_{n,h}^*(x) - E^*\hat{m}_{n,h}^*(x)) \leq u) - P(\sqrt{nh}(\hat{m}_{n,h}(x) - E\hat{m}_{n,h}(x)) \leq u) \right| \xrightarrow{P} 0.$$

In order to build bootstrap confidence intervals for $m(x)$, according to equation (2.3), we either have to estimate the bias $h^2B_s(x)$, where $B_s(x)$ is quite complex defined in equation (2.4), or use an under-smoothed bandwidth as given by Assumptions (viii); the latter is recommended in this paper:

Corollary 5.5. Under Assumption A (i)-(v), Assumption B (i)-(v), Assumption C (i)-(iv) and Assumption D (i)-(ii),

$$\sup_u \left| P(\sqrt{nh}(\hat{m}_{n,h}^*(x) - E^*\hat{m}_{n,h}^*(x)) \leq u) - P(\sqrt{nh}(\hat{m}_{n,h}(x) - m(x)) \leq u) \right| \xrightarrow{P} 0.$$

Then an equal-tailed $(1 - \alpha)100\%$ bootstrap confidence interval can be constructed:

$$[\hat{m}_{n,h}(x) + q^*(\alpha/2), \hat{m}_{n,h}(x) + q^*(1 - \alpha/2)],$$

where $q^*(\alpha)$ denote the α -quantile of empirical distribution function of $\hat{m}_{n,h}(x) - E^*\hat{m}_{n,h}^*(x)$. In the next two sections, we go over the parameters setup and simulation results.

6. Practical issues

6.1. The transform and transform-back functions

In this section, we discuss the selection of transform & transform back functions of model free method between discrete and smooth conditional CDF kernel estimation. Politis applied the smooth kernel estimator in [13, 14], i.e. $\tilde{D}_x(\cdot)$, see (4.3). The discrete estimation, $\hat{D}_x(\cdot)$ (4.3), was not suggested because it makes the transformed data unnatural and too equidistant in application.

For the selection of transform back function, that transforms U_i^* (resampled from U_1, \dots, U_n , where U_i is transformed data for MF or PMF method, true Unif(0,1) data for LMF method), to bootstrapped data $Y^+|x$. As mentioned earlier, all theorems in this paper are proven based on smooth transform back function $\tilde{D}_x^{-1}(\cdot)$. The validity of these theorems if replacing the smooth transform back function by discrete transform back function $\hat{D}_x^{-1}(\cdot)$ requires further study but very likely remains, due to the fact two transform back functions share many same properties. In addition, this selection problem between $\tilde{D}_x^{-1}(\cdot)$ and $\hat{D}_x^{-1}(\cdot)$ is similar to the question discussed by Efron [3]. He was interested the performance of resampling from regular empirical distribution function and from a kernel distribution estimation, and so compared the mean squared error (MSE) of $\int g(x)d\hat{F}(x)$ and $\int g(x)d\tilde{F}(x)$, see [22]. For the sake of simplicity, suppose u_i^* are distributed Unif(0,1), then $\tilde{D}_x^{-1}(u_i^*)$ is equivalent to a random variable from $\tilde{D}_x(\cdot)$, $\hat{D}_x^{-1}(u_i^*)$ is equivalent to a random variable from $\hat{D}_x(\cdot)$. Define:

$$\tilde{g} = \int g(y)d\tilde{D}_x(y),$$

and

$$\hat{g} = \int g(y)d\hat{D}_x(y).$$

If we assume x_1, \dots, x_n are deterministic, and do some simple computation, can obtain:

$$\text{MSE}[\tilde{g}] = \text{MSE}[\hat{g}] + h_0^2 \cdot \frac{\sum_{i=1}^n \left\{ \int g(y_i)g''(y_i)dD_{x_i}(y_i) \right\} K\left(\frac{x_i-x}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)} + O(h_0^4).$$

Denote the second term on the left side by Δ . It is apparently Δ determines which MSE is greater. For example, if $\Delta > 0$, $\tilde{D}_x^{-1}(\cdot)$ is better with respect to MSE. Although it is extremely difficult to estimate Δ , when $g(\cdot)$ is identical function:

$$\text{MSE}\left[g(\tilde{E})\right] = \text{MSE}\left[g(\hat{E})\right] = O(h_0^4).$$

The MSE difference is trivial when $h_0 \rightarrow 0$ (requires $n \rightarrow \infty$). More similar discussions can be found in [8, 9].

6.2. Selection of bandwidth

Since the smooth kernel estimation of conditional CDF has two different bandwidths h and h_0 , see lemma 4.6, the whole procedure of model free method requires to optimize both of them. For \hat{h} , the cross validation method proposed by Politis [13] is employed in this paper. For h_0 , Politis [13] employed R software function `bw.nrd0()` which is a fast bandwidth selection for density estimation.

Li and Racine [8] showed that $h_0 = O(n^{-\frac{2}{5}})$ and $h = O(n^{-\frac{1}{5}})$ by minimizing weighted IMSE. Therefore running another cross validation for h_0 does not guarantee the correct order but definitely increase the overall complexity of

algorithm. We suggest the same idea by Pan and Politis [12], that is simply making $\hat{h}_0 = \hat{h}^2$. In our simulation, we find that both two selections of h_0 work well, and give numerical close outputs. The work on bandwidth h and especially h_0 require further study.

7. Simulation

In this section we present simulation results for both additive model and non-additive model. The results are tabulated in the following two sub-sections.

7.1. Additive model

The additive model in this simulation has functional defined in (1.1), with $m(x) = \sin(x)$, $\sigma(x) = 1/2$, and errors ϵ_i i.i.d. $N(0,1)$ or two sided exponential (Laplace) rescaled to unit variance.

$$Y_i = \sin(x_i) + \frac{1}{2}\epsilon_i, i = 1, \dots, n.$$

For each model, 500 datasets are simulated. Each dataset has sample size $n = 100$, where regressors x_1, \dots, x_n are randomly drawn from normal distribution with mean π and standard deviation π . Kernel function employed by N-W estimator, $\hat{D}_x(\cdot)$ and $\tilde{D}_x(\cdot)$ is always a normal density function. In each cell of the following tables, first line gives estimated coverage probability, second line gives mean of length of confidence interval, third line gives standard error of length. “Norm” denote normal approximation by (2.3). “MB” is model-based bootstrap and “PRMB” is predictive residual model-based bootstrap, both of them were proposed by Politis [13] for nonparametric regression. “LB” stands for local bootstrap. “LMF”, “MF”, “PMF” are using statistics defined in equation (4.9). “LMF using N-W” is using a different statistics defined in equation (4.8). All the intervals have confidence level 95%.

- The standard error of the reported coverage probability levels over the 500 replications is $\sqrt{0.05 \times 0.95/500} \approx 0.0097$
- Since the true model is $m(x) = \sin(x)$, this simulation has some symmetry that helps us to adjust the CVRs. To elaborate, note that for any $x \in [0, \pi]$, we have $|m(x)| = |m(2\pi - x)|$, and the same symmetry holds for the derivatives of $m(x)$ as well due to the sinusoidal structure. Due to this symmetry, for any good confidence interval method, $\hat{m}_{n,h}(x)$ at each two symmetric points are supposed to have exactly same limit distribution, and hence numerically very close CVRs.
- PRMB method show outstanding CVR, while had very high average length and standard error. After diving into the stored bootstrap data, it turned out in some bootstrap iteration, the bootstrap ϵ were very high, and resulted in a wide confidence interval. Such confidence intervals are apparently outliers, and deserve further root cause analysis.

- Theorem 4.4, the equivalence between the local bootstrap and LMF is verified in the simulation. Notice that only “LMF using N-W” produces numerically similar results as local bootstrap. Finite sample difference exists when using “LMF”, which can be relieved by larger sample size.
- In table 2, CDF of laplace distribution violate assumption (vii). More precisely, CDF of Laplace distribution is not continuously differentiable at 0, so lemma 4.6 might fail or the convergence has a slower rate. It implies that U_1, \dots, U_n calculated from $(X_i, Y_i), i = 1, \dots, n$ might not be approximately $\text{Unif}(0,1)$ distributed, see Fig 1. This is the primary reason at some points LMF, MF and PMF show over-coverage problem for Laplace error.

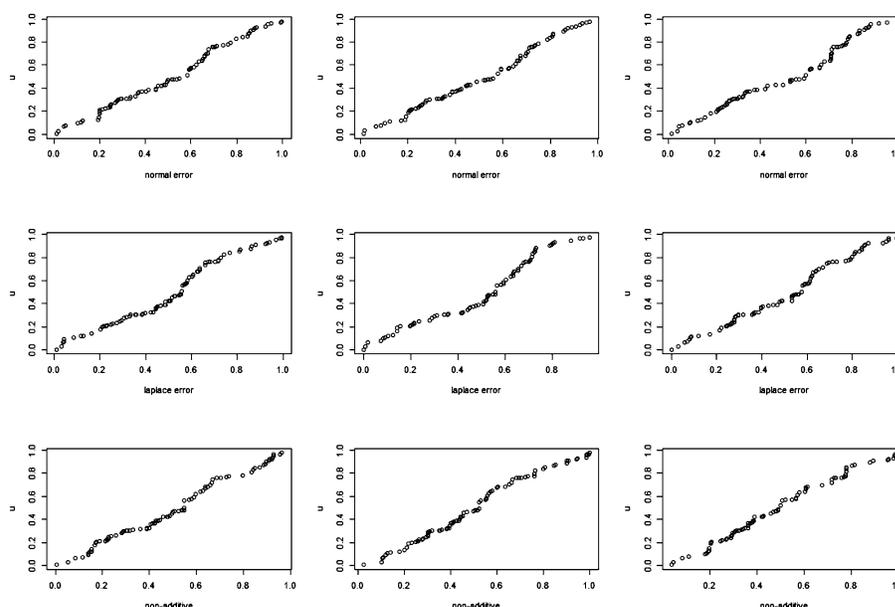


FIG 1. Examples of QQ plots from all three models

7.2. Non-additive model

In this subsection we present the simulation result of a non-additive model. For the sake of comparison to section 7.1, we use the following model:

$$Y_i = \sin(x_i) + \frac{1}{2}\epsilon_{x_i}, i = 1, \dots, n,$$

where $x_i \in [0, 2\pi]$, $\epsilon_x = \frac{c_x Z + (1-c_x)W}{\sqrt{c_x^2 + (1-c_x)^2}}$, and $c_x = x/2\pi$, $Z \sim N(0, 1)$ independent of W that will be distributed as $\frac{1}{2}\chi_2^2 - 1$. Thus ϵ_x has mean 0 and variance 1. $E(Y|X = x) = \sin(x)$ and $\text{Var}(Y|X = x) = 1/4$, and regressors x_1, \dots, x_n are randomly drawn from normal distribution with mean π and standard deviation

TABLE 1

Simulation results of 95% level C.I. with additive model and i.i.d. normal distributed error

Methods	Metrics	$x_f/p_i =$								
		0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
Norm	CVR	88.6%	86.0%	79.4%	78.8%	87.6%	81.0%	78.6%	84.8%	86.0%
	Avg Length	0.520	0.487	0.458	0.432	0.427	0.434	0.462	0.485	0.514
	Std. Err. Length	0.005	0.004	0.004	0.003	0.003	0.003	0.004	0.004	0.005
MB	CVR	83.4%	80.8%	81.0%	83.4%	86.6%	86.2%	82.2%	81.6%	83.4%
	Avg Length	0.509	0.454	0.413	0.431	0.451	0.427	0.406	0.452	0.504
	Std. Err. Length	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.005
PRMB	CVR	94.2%	92.8%	91.6%	92.8%	95.0%	93.6%	91.8%	91.6%	92.8%
	Avg Length	1.791	1.559	1.379	1.437	1.478	1.420	1.344	1.551	1.649
	Std. Err. Length	0.137	0.114	0.105	0.109	0.105	0.102	0.100	0.130	0.122
LB	CVR	88.4%	88.0%	88.4%	89.8%	92.4%	89.0%	88.2%	89.4%	89.8%
	Avg Length	0.612	0.543	0.494	0.517	0.543	0.513	0.491	0.544	0.610
	Std. Err. Length	0.006	0.005	0.005	0.004	0.004	0.004	0.005	0.006	0.006
LMF using N-W	CVR	90.4%	89.0%	86.6%	88.2%	93.8%	88.0%	85.4%	89.8%	90.2%
	Avg Length	0.587	0.523	0.472	0.492	0.520	0.494	0.471	0.520	0.580
	Std. Err. Length	0.005	0.004	0.004	0.003	0.004	0.004	0.004	0.005	0.005
LMF	CVR	89.8%	91.8%	92.0%	93.4%	94.2%	93.6%	91.4%	93.6%	90.8%
	Avg Length	0.719	0.646	0.595	0.642	0.686	0.642	0.592	0.643	0.710
	Std. Err. Length	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.005
MF	CVR	88.6%	88.6%	87.0%	91.0%	92.6%	88.6%	87.4%	91.4%	89.6%
	Avg Length	0.582	0.522	0.483	0.511	0.547	0.512	0.482	0.518	0.573
	Std. Err. Length	0.005	0.004	0.003	0.003	0.003	0.003	0.003	0.004	0.004
PMF	CVR	95.2%	95.2%	93.8%	96.2%	98.8%	96.6%	93.4%	95.2%	96.2%
	Avg Length	0.728	0.662	0.617	0.649	0.686	0.649	0.616	0.660	0.721
	Std. Err. Length	0.006	0.005	0.005	0.004	0.005	0.004	0.005	0.006	0.006

TABLE 2

Simulation results of 95% level C.I. with additive model and i.i.d. Laplace distributed error

Methods	Metrics	$x_f/p_i =$								
		0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
Norm	CVR	89.2%	89.0%	80.6%	83.2%	87.4%	82.8%	78.4%	87.8%	88.6%
	Avg Length	0.518	0.488	0.459	0.434	0.426	0.434	0.460	0.489	0.507
	Std. Err. Length	0.005	0.004	0.004	0.004	0.003	0.003	0.004	0.004	0.004
MB	CVR	87.0%	84.0%	80.6%	85.8%	87.2%	83.6%	81.2%	85.0%	84.6%
	Avg Length	0.503	0.444	0.401	0.428	0.445	0.411	0.391	0.440	0.492
	Std. Err. Length	0.006	0.005	0.005	0.005	0.004	0.004	0.005	0.005	0.006
PRMB	CVR	94.6%	95.4%	94.4%	95.4%	96.6%	95.8%	92.8%	95.2%	95.2%
	Avg Length	1.610	1.415	1.274	1.309	1.347	1.251	1.201	1.407	1.646
	Std. Err. Length	0.113	0.100	0.094	0.094	0.092	0.085	0.083	0.095	0.121
LB	CVR	92.4%	92.0%	88.6%	91.0%	94.0%	91.0%	88.0%	91.0%	91.2%
	Avg Length	0.615	0.547	0.496	0.522	0.546	0.502	0.479	0.538	0.599
	Std. Err. Length	0.007	0.007	0.006	0.006	0.005	0.005	0.006	0.006	0.007
LMF using N-W	CVR	93.6%	91.0%	88.0%	90.4%	93.2%	90.0%	87.2%	90.8%	92.8%
	Avg Length	0.584	0.523	0.479	0.500	0.521	0.489	0.464	0.514	0.575
	Std. Err. Length	0.006	0.006	0.005	0.005	0.004	0.004	0.005	0.006	0.006
LMF	CVR	93.2%	95.0%	96.4%	96.0%	95.0%	96.2%	94.0%	95.0%	93.6%
	Avg Length	0.715	0.646	0.602	0.652	0.688	0.637	0.584	0.636	0.703
	Std. Err. Length	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.006	0.006
MF	CVR	92.4%	91.0%	88.8%	90.4%	93.6%	92.6%	85.6%	90.0%	90.2%
	Avg Length	0.547	0.483	0.446	0.479	0.519	0.475	0.438	0.478	0.538
	Std. Err. Length	0.005	0.004	0.004	0.003	0.003	0.003	0.003	0.004	0.004
PMF	CVR	97.4%	98.2%	95.6%	96.8%	98.8%	96.4%	93.4%	96.2%	97.0%
	Avg Length	0.712	0.649	0.610	0.642	0.680	0.633	0.599	0.643	0.703
	Std. Err. Length	0.006	0.006	0.005	0.005	0.004	0.005	0.005	0.006	0.006

TABLE 3
Simulation results of 95% level C.I. with non-additive model

Methods	Metrics	$x_f/p_i =$								
		0.15	0.3	0.5	0.75	1	1.25	1.5	1.7	1.85
Norm	CVR	88.4%	86.2%	81.0%	78.2%	86.4%	80.4%	78.8%	85.8%	86.0%
	Avg Length	0.513	0.483	0.456	0.429	0.425	0.430	0.459	0.482	0.506
	Std. Err. Length	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.005
MB	CVR	82.2%	81.0%	81.0%	80.6%	86.0%	83.0%	79.0%	80.6%	83.6%
	Avg Length	0.486	0.435	0.391	0.417	0.449	0.420	0.398	0.442	0.497
	Std. Err. Length	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.005	0.005
PRMB	CVR	92.8%	90.0%	87.6%	91.2%	96.6%	95.2%	93.4%	92.4%	94.6%
	Avg Length	1.581	1.404	1.254	1.240	1.373	1.353	1.285	1.406	1.606
	Std. Err. Length	0.112	0.100	0.088	0.080	0.096	0.100	0.093	0.096	0.112
LB	CVR	88.4%	86.4%	85.0%	88.0%	92.6%	90.6%	87.8%	89.2%	89.8%
	Avg Length	0.592	0.531	0.474	0.511	0.549	0.514	0.485	0.539	0.607
	Std. Err. Length	0.007	0.007	0.006	0.005	0.006	0.006	0.006	0.006	0.006
LMF using N-W	CVR	90.8%	88.0%	83.8%	86.6%	93.6%	90.8%	85.6%	88.4%	92.2%
	Avg Length	0.573	0.512	0.462	0.491	0.525	0.495	0.467	0.519	0.582
	Std. Err. Length	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.005	0.005
LMF	CVR	92.0%	91.8%	88.6%	95.0%	95.4%	94.8%	92.8%	93.2%	92.2%
	Avg Length	0.701	0.633	0.584	0.640	0.694	0.644	0.588	0.641	0.712
	Std. Err. Length	0.007	0.007	0.006	0.005	0.005	0.005	0.005	0.005	0.005
MF	CVR	88.4%	86.4%	81.2%	86.8%	92.8%	90.0%	86.4%	89.4%	90.6%
	Avg Length	0.533	0.475	0.437	0.468	0.523	0.488	0.458	0.500	0.557
	Std. Err. Length	0.005	0.004	0.003	0.003	0.004	0.004	0.004	0.004	0.004
PMF	CVR	95.6%	94.6%	89.6%	94.2%	98.2%	97.4%	95.6%	96.0%	96.6%
	Avg Length	0.685	0.622	0.576	0.613	0.673	0.641	0.603	0.643	0.705
	Std. Err. Length	0.007	0.006	0.005	0.005	0.005	0.006	0.005	0.006	0.006

π with sample size $n = 100$. However, ϵ_x has skewness depending on x that violating i.i.d. assumption. The result is presented in table 3. It seems that PMF performs the best among the different methods although it yields some over-coverage, e.g. at $x = \pi$ and $x = 1.25\pi$.

Acknowledgements

We thank Srinjoy Das for assistance with code review of simulation, and Yiren Wang for comments that greatly improved the manuscript. We are also grateful to three anonymous reviewers for their constructive suggestions. Research was partially supported by NSF grant DMS 19-14556.

Appendix: Proof

Proof of lemma 3.1. Since l is a fixed positive integer, by Assumption C(i), $K^l(\cdot)$ is also symmetric, bounded and twice differentiable function. Then,

$$\begin{aligned}
 & E_{X_i, Y_i} \left[g(Y_i) K^l \left(\frac{X_i - x}{h} \right) \right] \\
 &= \iint g(y_i) K^l \left(\frac{x_i - x}{h} \right) f_{X_i, Y_i}(x_i, y_i) dx_i dy_i \\
 &= h \iint g(y_i) K^l(u_i) f_{X_i, Y_i}(x + u_i h, y_i) du_i dy_i
 \end{aligned}$$

$$\begin{aligned}
&= h \iint g(y_i) K^l(u_i) \left[f_{X_i, Y_i}(x, y_i) + \frac{\partial f_{X_i, Y_i}(x, y_i) u_i h}{\partial x} \right. \\
&\quad \left. + \frac{1}{2} \frac{\partial^2 f_{X_i, Y_i}(x, y_i) u_i^2 h^2}{\partial x^2} \right] du_i dy_i \\
&= h \left\{ \int K^l(u_i) du_i \int g(y_i) f_{X_i, Y_i}(x, y_i) dy_i \right. \\
&\quad + h \int u_i K^l(u_i) du_i \int g(y_i) \frac{\partial f_{X_i, Y_i}(x, y_i)}{\partial x} dy_i \\
&\quad \left. + \frac{h^2}{2} \int u_i^2 K^l(u_i) du_i \int g(y_i) \frac{\partial^2 f_{X_i, Y_i}(x, y_i)}{\partial x^2} dy_i + O(h^3) \right\} \\
&= h \left\{ \int K^l(u_i) du_i \int g(y_i) f_{Y_i|X_i}(y_i|x) f_{X_i}(x) dy_i + \frac{h^2}{2} \int u_i^2 K^l(u_i) du_i \int g(y_i) \right. \\
&\quad \left. \frac{\partial^2 f_{X_i, Y_i}(x, y_i)}{\partial x^2} dy_i + O(h^3) \right\} \\
&= h \left\{ f_X(x) \int K^l(u) du \cdot E[g(Y)|X = x] \right. \\
&\quad \left. + \frac{h^2}{2} \int u^2 K^l(u) du \int g(y) \frac{\partial^2}{\partial x^2} f_{X, Y}(x, y) dy + O(h^3) \right\} \\
&= h \left\{ A(x) \cdot E[g(Y)|X = x] + O(h^2) \right\}
\end{aligned}$$

where $A(x) = f_X(x) \int K^l(u) du$. The very last step we can get $O(h^2)$ is because of Assumption B(ii), by which and dominated convergence theorem, we can interchange order of differentiation and integration. Then by Assumption B(iii), the integration is finite. Notice the original term of $O(h^2)$ is not related to i anymore, so they have the common bound. \square

Proof of theorem 3.2. By definition of Big O notation in probability, we just need to show:

$$E^*[g(Y^*)|X = x] - E[g(Y)|X = x] = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

First, by the definition of local bootstrap algorithm,

$$\begin{aligned}
E^*[g(Y^*)|X = x] &= \frac{\sum_{i=1}^n g(Y_i) K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \\
&= \frac{1}{\widehat{f}_X(x)} \cdot \frac{1}{nh} \sum_{i=1}^n g(Y_i) K\left(\frac{X_i - x}{h}\right) \\
&= \frac{1}{f_X(x) + o_p(1)} \cdot \frac{1}{nh} \sum_{i=1}^n g(Y_i) K\left(\frac{X_i - x}{h}\right)
\end{aligned}$$

By lemma 3.1, we have:

$$\begin{aligned}
& \mathbb{E}_{X,Y} \left[\frac{1}{nh} \sum_{i=1}^n g(Y_i) K \left(\frac{X_i - x}{h} \right) \right] \\
&= \frac{1}{nh} \sum_{i=1}^n \mathbb{E}_{X_i, Y_i} \left[g(Y_i) K \left(\frac{X_i - x}{h} \right) \right] \\
&= \frac{1}{nh} \sum_{i=1}^n h \left\{ f_X(x) \int K(u) du \cdot \mathbb{E}[g(Y)|X = x] + O(h^2) \right\} \\
&= f_X(x) \cdot \mathbb{E}[g(Y)|X = x] + O(h^2)
\end{aligned}$$

As $\{(X_i, Y_i)_{i=1}^n$ are independent and under Assumption B(iv), we also have

$$\begin{aligned}
& \text{Var}_{X,Y} \left[\frac{1}{nh} \sum_{i=1}^n g(Y_i) K \left(\frac{X_i - x}{h} \right) \right] \\
&= \frac{1}{(nh)^2} \sum_{i=1}^n \text{Var} \left[g(Y_i) K \left(\frac{X_i - x}{h} \right) \right] \\
&= \frac{1}{(nh)^2} \sum_{i=1}^n \left\{ \mathbb{E} \left[g(Y_i) K \left(\frac{X_i - x}{h} \right) \right]^2 - \left(\mathbb{E} \left[g(Y_i) K \left(\frac{X_i - x}{h} \right) \right] \right)^2 \right\} \\
&= \frac{1}{(nh)^2} \sum_{i=1}^n \left(h \left\{ f_X(x) \int K^2(u) du \cdot \mathbb{E}[g^2(Y)|X = x] + O(h^2) \right\} \right. \\
&\quad \left. - h^2 \left\{ f_X(x) \int K(u) du \cdot \mathbb{E}[g(Y)|X = x] + O(h^2) \right\}^2 \right) \\
&= \frac{1}{nh} \left[\int K^2(u) du f_X(x) \mathbb{E}[g^2(Y)|X = x] + O(h^2) \right] \\
&\quad - \frac{1}{n} \left[f_X(x) \mathbb{E}[g(Y)|X = x] + O(h^2) \right]^2 \\
&= O\left(\frac{1}{nh}\right)
\end{aligned}$$

By Markov's Inequality, it is easy to get

$$\frac{1}{nh} \sum_{i=1}^n g(Y_i) K \left(\frac{X_i - x}{h} \right) - f_X(x) \mathbb{E}[g(Y)] = O_p \left(h^2 + \frac{1}{\sqrt{nh}} \right)$$

Finally, we get:

$$\begin{aligned}
& \mathbb{E}^*[g(Y)|X = x] - \mathbb{E}[g(Y)|X = x] \\
&= \frac{1}{f_X(x) + o_p(1)} \cdot \frac{1}{nh} \sum_{i=1}^n g(Y_i) K \left(\frac{X_i - x}{h} \right) - \mathbb{E}[g(Y)|X = x] \\
&= \frac{1}{f_X(x) + o_p(1)} \cdot \left[\frac{1}{nh} \sum_{i=1}^n g(Y_i) K \left(\frac{X_i - x}{h} \right) - \right.
\end{aligned}$$

$$\begin{aligned}
& \left. [f_X(x) + o_p(1)]\mathbb{E}[g(Y)|X = x] \right] \\
&= \frac{O_p(h^2 + 1/\sqrt{nh}) - o_p(1)\mathbb{E}[g(Y)|X = x]}{f_X(x) + o_p(1)} \\
&= O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right) \quad \square
\end{aligned}$$

Proof of theorem 4.4. For the resampling method of local bootstrap, at X_i , recall (3.1), \widehat{G}_{X_i} is the p.m.f. to resample data and generate Y_i^* . On the other side, remark 4.2 explains LMF also uses a p.m.f. \widehat{H}_{X_i} to generate Y_i^* . To show the equivalence, we will prove $\widehat{G}_{x_i} = \widehat{H}_{x_i}$.

$$\widehat{H}_{x_i} : \begin{pmatrix} Y_1 & Y_2 & \dots & Y_n \\ \tilde{w}_{1i} & \tilde{w}_{2i} & \dots & \tilde{w}_{ni} \end{pmatrix}$$

The reason we can define such a function for model free, is that no matter what u_i^* we get from $\text{Unif}(0,1)$, by the transformed function $\widehat{D}_x^{-1}(\cdot)$, Y^+ will be always one point from $\{Y_1, \dots, Y_n\}$. Thus for each Y_j , there is a probability we choose it as Y_i^+ . For each i, j , we define $Y_j^- = \max_{k=1, \dots, n} \{Y_k, Y_k < Y_j\}$. Then, we have:

$$\begin{aligned}
\tilde{w}_{ij} &= P(Y_i^+ = Y_j) \\
&= P(\widehat{D}_{x_i}^{-1}(u_i^*) = Y_j) \\
&= P(\mathbb{1}_{\{\exists k, Y_k < Y_j\}} \cdot \widehat{D}_{x_i}(Y_j^-) < u_i^* \leq \widehat{D}_{x_i}(Y_j)) \\
&= \frac{\sum_{t=1}^n \mathbb{1}_{\{Y_t \leq Y_j\}} K\left(\frac{X_i - X_t}{h}\right)}{\sum_{t=1}^n K\left(\frac{X_i - X_t}{h}\right)} - \mathbb{1}_{\{\exists k, Y_k < Y_j\}} \cdot \frac{\sum_{t=1}^n \mathbb{1}_{\{Y_t \leq Y_j^-\}} K\left(\frac{X_i - X_t}{h}\right)}{\sum_{t=1}^n K\left(\frac{X_i - X_t}{h}\right)} \\
&= \frac{K\left(\frac{X_i - X_j}{h}\right)}{\sum_{t=1}^n K\left(\frac{X_i - X_t}{h}\right)} \quad \square
\end{aligned}$$

Proof of Theorem 4.5. By the algorithm of MF, but notice each $u_i, i = 1 \dots, n$ is assumed to be $\text{Unif}(0,1)$ distributed. Then,

$$\mathbb{E}^+[g(Y^+)|X = x] = \frac{1}{n} \sum_{i=1}^n g[\widehat{D}_x^{-1}(u_i)]$$

Then, consider

$$\begin{aligned}
& |\mathbb{E}^+[g(Y^+)|x] - \mathbb{E}[g(Y)|X = x]| \leq \\
& |\mathbb{E}^+[g(Y^+)|x] - \mathbb{E}^*[g(Y^*)|X = x]| + |\mathbb{E}^*[g(Y^*)|X = x] - \mathbb{E}[g(Y)|X = x]| \quad (\text{A.1})
\end{aligned}$$

Let \mathbb{E}^* denote expectation in local bootstrap world, and \mathbb{E}^+ denote expectation in model free bootstrap world. For the right side of inequality above, the second

term is proved in theorem 3.2. For the first term, by equation (4.7), we know $E^*[g(Y^*)|X = x] = E_u g[\widehat{D}_x^{-1}(u)]$, where u comes from Unif(0,1) distribution. Let random vector $U = (u_1, \dots, u_n)$, independent from u . Then we have:

$$\begin{aligned}
& E\{E^+[g(Y^+)|X = x] - E^*[g(Y^*)|X = x]\} \\
&= E_{X,Y,U} \left[\frac{1}{n} \sum_{i=1}^n g[\widehat{D}_x^{-1}(u_i)] - E_u g[\widehat{D}_x^{-1}(u)] \right] \\
&= E_{X,Y} \left\{ E_U \left[\frac{1}{n} \sum_{i=1}^n g[\widehat{D}_x^{-1}(u_i)|X, Y] \right] - \right. \\
&\quad \left. E_u g[\widehat{D}_x^{-1}(u)|X, Y] \right\} \\
&= 0 \tag{A.2}
\end{aligned}$$

Then consider the variance:

$$\begin{aligned}
& \text{Var}_{X,Y,U} \{E^+[g(Y^+)|X = x] - E^*[g(Y^*)|X = x]\} \\
&= E_{X,Y,U} \left\{ E^+[g(Y^+)|X = x] - E^*[g(Y^*)|X = x] \right\}^2 \\
&= E_{X,Y} \left(E_U \{E^+[g(Y^+)|X = x] - E^*[g(Y^*)|X = x]|X, Y\}^2 \right) \\
&= E_{X,Y} \left(E_U \left\{ \frac{1}{n} \sum_{i=1}^n g[\widehat{D}_x^{-1}(u_i)] - E_u g[\widehat{D}_x^{-1}(u)] \middle| X, Y \right\}^2 \right) \\
&= E_{X,Y} \left(\text{Var}_U \left\{ \frac{1}{n} \sum_{i=1}^n g[\widehat{D}_x^{-1}(u_i)] \middle| X, Y \right\} \right) \\
&= E_{X,Y} \left(\frac{1}{n^2} \sum_{i=1}^n \text{Var}_U \{g[\widehat{D}_x^{-1}(u_i)]|X, Y\} \right) \\
&= \frac{1}{n} E_{X,Y} \left(\text{Var}^*[g(Y^*)|X = x] \right) \\
&= \frac{1}{n} E_{X,Y} \left(E^*[g(Y^*)|X = x]^2 - \{E^*[g(Y^*)|X = x]\}^2 \right)
\end{aligned}$$

For the second term, in theorem 3.2, we already show that

$$E^*[g(Y^*)|X = x] - E[g(Y)|X = x] = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

and

$$\text{Var}_{X,Y} \left(E^*[g(Y^*)|X = x] \right) = O\left(\frac{1}{nh}\right)$$

which implies

$$\begin{aligned}
E_{X,Y} \left(E^*[g(Y^*)|X = x] \right)^2 &= \text{Var}_{X,Y} \left(E^*[g(Y^*)|X = x] \right) \\
&\quad + \left[E_{X,Y} \left(E^*[g(Y^*)|X = x] \right) \right]^2
\end{aligned}$$

is bounded. Then let us see the first term:

$$\begin{aligned}
& \mathbb{E}_{X,Y} \left\{ \mathbb{E}^* [g(Y^*) | X = x]^2 \right\} \\
&= \mathbb{E}_{X,Y} \left\{ \frac{1}{nh} \sum_{i=1}^n g^2(Y_i) \tilde{K} \left(\frac{X_i - x}{h} \right) \right\} \\
&= \frac{1}{nh} \sum_{i=1}^n \mathbb{E}_{X,Y} \left\{ g^2(Y_i) \tilde{K} \left(\frac{X_i - x}{h} \right) \right\} \\
&= \frac{1}{nh} \sum_{i=1}^n h \left\{ f_X(x) \int \tilde{K}(u) du \cdot \mathbb{E}[g^2(Y) | X = x] + O(h^2) \right\} \quad \text{by lemma 5.1} \\
&= \tilde{A}(x) \mathbb{E}[g^2(Y) | X = x] + O(h^2) \\
&< \infty
\end{aligned}$$

Thus $\text{Var}_{X,Y,U} \{ \mathbb{E}^+ [g(Y^+) | X = x] - \mathbb{E}^* [g(Y^*) | X = x] \} = O(1/n)$. This and (A.2), Chebyshev inequality, and the definition of big O in probability notation leads to:

$$\mathbb{E}^+ [g(Y^+) | X = x] - \mathbb{E}^* [g(Y^*) | X = x] = O_p \left(\frac{1}{\sqrt{n}} \right)$$

Recall (A.1), we have:

$$\begin{aligned}
|\mathbb{E}^+ [g(Y^+) | X = x] - \mathbb{E}[g(Y) | X = x]| &= O_p \left(\frac{1}{\sqrt{n}} \right) + O_p \left(h^2 + \frac{1}{\sqrt{nh}} \right) \\
&= O_p \left(h^2 + \frac{1}{\sqrt{nh}} \right) \quad \square
\end{aligned}$$

Proof of lemma 4.6. Let $\hat{d}(x) = \frac{1}{nh} \sum_{i=1}^n \Lambda \left(\frac{Y_i - y}{h_0} \right) K \left(\frac{X_i - x}{h} \right)$, then we write:

$$\begin{aligned}
& \sup_{x \in \mathcal{S}} |\tilde{D}_x(y) - D_x(y)| \\
&= \sup_{x \in \mathcal{S}} \left| \frac{\hat{d}(x)}{\hat{f}(x)} - D_x(y) \right| \\
&\leq \frac{\sup_{x \in \mathcal{S}} |\hat{d}(x) - \mathbb{E}(\hat{d}(x)) + \mathbb{E}(\hat{d}(x)) - D_x(y) \hat{f}(x)|}{\inf_{x \in \mathcal{S}} |\hat{f}(x)|} \\
&\leq \frac{\sup_{x \in \mathcal{S}} |\hat{d}(x) - \mathbb{E}(\hat{d}(x))|}{\inf_{x \in \mathcal{S}} |\hat{f}(x)|} + \frac{\sup_{x \in \mathcal{S}} |\mathbb{E}(\hat{d}(x)) - D_x(y) \hat{f}(x)|}{\inf_{x \in \mathcal{S}} |\hat{f}(x)|}
\end{aligned}$$

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

By Assumption A (v) and thm 1.4 from [8], i.e. $\sup_{x \in \mathcal{S}} |\hat{f}(x) - f(x)| \xrightarrow{a.s.} 0$, for large sufficiently enough n we can find another $\delta' > 0$ s.t. $\inf_{x \in \mathcal{S}} \hat{f}(x) \geq \delta' > 0$. And for the second term, by theorem 6.2 in Li and Racine [8], we have:

$$\left\{ \mathbb{E}[\hat{d}(x)] - D_x(y) \hat{f}(x) \right\}$$

$$\begin{aligned}
&= \frac{h_0^2}{2} \int u^2 K(u) du \cdot \frac{\partial^2 D_x(y)}{\partial y^2} f(x) + \frac{h^2}{2} \int u^2 K(u) du \left[\frac{\partial^2 D_x(y)}{\partial x^2} f(x) \right. \\
&\quad \left. + 2 \frac{\partial f(x)}{\partial x} \cdot \frac{\partial D_x(y)}{\partial x} \right] + o(h_0^2 + h^2) \\
&= O(h_0^2 + h^2)
\end{aligned}$$

By Assumption A (ii) and Assumption B (v), $\frac{\partial^2 D_x(y)}{\partial y^2}$ and $\sup_{x \in \mathcal{S}} \left[\frac{\partial D_x(y)}{\partial x} f(x) + 2 \frac{\partial f(x)}{\partial x} \cdot \frac{\partial D_x(y)}{\partial x} \right]$ is bounded, which implies

$$\sup_{x \in \mathcal{S}} \left| \mathbb{E}(\widehat{d}(x)) - D_x(y) \widehat{f}(x) \right| = O(h_0^2 + h^2)$$

Then just need to show

$$\sup_{x \in \mathcal{S}} \left| \widehat{d}(x) - \mathbb{E}(\widehat{d}(x)) \right| = O\left(\left[\frac{\ln(n)}{nh}\right]^{\frac{1}{2}}\right) \quad \text{a.s.} \quad (\text{A.3})$$

As \mathcal{S} is compact, it can be covered by a finite number L_n of interval $\{I_k\}_1^{L_n}$ with length l_n , and $L_n = \text{constant}/l_n$. We write:

$$\begin{aligned}
\sup_{x \in \mathcal{S}} \left| \widehat{d}(x) - \mathbb{E}(\widehat{d}(x)) \right| &= \max_{1 \leq k \leq L_n} \sup_{x \in \mathcal{S} \cap I_k} \left| \widehat{d}(x) - \mathbb{E}(\widehat{d}(x)) \right| \\
&\leq \max_{1 \leq k \leq L_n} \sup_{x \in \mathcal{S} \cap I_k} \left| \widehat{d}(x) - \widehat{d}(x_{k,n}) \right| \\
&\quad + \max_{1 \leq k \leq L_n} \left| \widehat{d}(x_{k,n}) - \mathbb{E}[\widehat{d}(x_{k,n})] \right| \\
&\quad + \max_{1 \leq k \leq L_n} \sup_{x \in \mathcal{S} \cap I_k} \left| \mathbb{E}[\widehat{d}(x_{k,n})] - \mathbb{E}[\widehat{d}(x)] \right| \\
&= Q_1 + Q_2 + Q_3
\end{aligned}$$

For Q_2 , the sup can be ignored because it only concern $x_{k,n}$ but not x , and $x_{k,n}$ is the central point of the interval I_k . We will show Q_1 and Q_3 in the last part, just consider Q_2 first. To show $Q_2 = O(\eta_n)$ a.s., by Borel-Cantelli Lemma, we just need to show

$$\sum_{n=1}^{\infty} P[Q_2 > \eta_n] < \infty$$

Let $W_n(x) = \widehat{d}(x) - \mathbb{E}[\widehat{d}(x)] = \sum_i Z_{n,i}$, where

$$Z_{n,i} = \frac{1}{nh} \left\{ \Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right) - \mathbb{E}\left[\Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right)\right] \right\}$$

For any $\eta_n > 0$, we have

$$P[Q_2 > \eta_n] = P\left[\max_{1 \leq k \leq L_n} |W_n(x_{k,n})| > \eta_n\right]$$

$$\begin{aligned} &\leq \sum_{k=1}^{L_n} P[|W_n(x_{k,n})| > \eta_n] \\ &\leq L_n \sup_{x \in \mathcal{S}} P[|W_n(x)| > \eta_n] \end{aligned}$$

Since $K(\cdot)$ is bounded, let its supremum is A_1 , moreover, $\Lambda(\cdot)$ is a CDF and so $|\Lambda(\cdot) \leq 1|$, thus we have $Z_{n,i} \leq \frac{2A_1}{nh}$. Define $\lambda_n = (nh \ln(n))^{\frac{1}{2}}$, then $\lambda_n |Z_{n,i}| \leq 2A_1 \left[\frac{\ln(n)}{nh} \right]^{\frac{1}{2}}$. Then we can choose sufficiently large n such that $\lambda_n |Z_{n,i}| \leq \frac{1}{2}$ for all $i = 1, 2, \dots, n$. Now use inequality

$$\exp(x) \leq 1 + x + x^2 \quad \text{as } |x| \leq \frac{1}{2}$$

and

$$1 + x \leq \exp(x) \quad \text{as } x \geq 0$$

We have:

$$E[\exp(\pm \lambda_n Z_{n,i})] \leq 1 + E[\pm \lambda_n Z_{n,i}] + E[\lambda_n^2 |Z_{n,i}^2|] \leq \exp(E[\lambda_n^2 |Z_{n,i}^2|]) \quad \text{a.s.}$$

By the Markov inequality, ($a > 0$)

$$P[X > c] \leq \frac{E[\exp(Xa)]}{\exp(ac)} \quad (\text{A.4})$$

Thus,

$$\begin{aligned} P[|W_n(x)| > \eta] &= P\left[\left|\sum_{i=1}^n Z_{n,i}\right| > \eta\right] \\ &= P\left[\sum_{i=1}^n Z_{n,i} > \eta\right] + P\left[-\sum_{i=1}^n Z_{n,i} > \eta\right] \\ &\leq \frac{E[\exp(\lambda_n \sum_{i=1}^n Z_{n,i})] + E[\exp(-\lambda_n \sum_{i=1}^n Z_{n,i})]}{\exp(\lambda_n \eta)} \\ &\leq 2 \exp(-\lambda_n \eta) \prod_{i=1}^n \left[\exp\left(\lambda_n^2 E Z_{n,i}^2\right)\right] \\ &\leq 2 \exp(-\lambda_n \eta) \left[\exp\left(\frac{A_2 \lambda_n^2}{nh}\right)\right] \end{aligned}$$

Where we use

$$\begin{aligned} E Z_{n,i}^2 &= \frac{1}{(nh)^2} \text{Var}\left[\Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right)\right] \\ &\leq \frac{1}{n^2 h^2} E\left[\Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right)\right]^2 \\ &\leq \frac{1}{n^2 h^2} \int K^2\left(\frac{x_i - x}{h}\right) f_X(x_i) dx_i \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^2 h} \int K^2(u) f_X(x + uh) du \\
&= \frac{1}{n^2 h} \int K^2(u) \left[f_X(x) + \frac{\partial f_X(x) uh}{\partial x} + O(h^2) \right] du \\
&= \frac{1}{n^2 h} \left\{ f_X(x) \int K^2(u) du + h \frac{\partial f_X(x)}{\partial x} \int u K^2(u) du + O(h^2) \right\} \\
&= \frac{f(x) \int K^2(u) du}{n^2 h} [1 + O(h^2)] \\
&\leq \frac{A_2}{n^2 h}
\end{aligned}$$

As $f(x)$ is bounded, and $\int u K^2(u) du = 0$ since K is symmetric. Because A_2 is independent of x , we get

$$\sup_{x \in \mathcal{S}} P[|W_n(x)| > \eta_n] \leq 2 \exp\left(-\lambda_n \eta_n + \frac{A_2 \lambda_n^2}{nh}\right)$$

And let $\lambda_n = [(nh) \ln(n)]^{\frac{1}{2}}$ and $\eta_n = C_4 \ln(n) / \lambda_n = C_4 [\ln(n) / (nh)]^{\frac{1}{2}}$, where C_4 is positive. Choosing large enough C_4 , we get

$$\left(-\lambda_n \eta_n + \frac{A_2 \lambda_n^2}{nh}\right) = (-C_4 + A_2) \ln n = \alpha \ln n$$

where $\alpha = (-C_4 + A_2)$ is negative, and it can be small enough (by the choice of C_4) so that:

$$\sum_{n=1}^{\infty} P[Q_2 > \eta_n] \leq \sum_{n=1}^{\infty} \sup_{x \in \mathcal{S}} P[|W_n(x)| > \eta_n] \leq \sum_{n=1}^{\infty} \frac{L_n}{n^\alpha} < \infty$$

by choosing $L_n = \sqrt{\frac{n}{h^3 \ln n}}$. Thus, based on all the above and Borel-Cantelli Lemma, we know that,

$$Q_2 = O\left(\left[\frac{\ln(n)}{nh}\right]^{\frac{1}{2}}\right) \quad \text{a.s.}$$

Now consider Q_1 and Q_3 . By Assumption C (iv), $K(\cdot)$ is Lipschitz continuous and Lipschitz constant is C_1 , we know that

$$\begin{aligned}
&\sup_{\mathcal{S} \cap I_k} \left| \Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right) - \Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x_{k,n}}{h}\right) \right| \\
&\leq \frac{C_1}{h} \sup_{\mathcal{S} \cap I_k} \|x - x_{k,n}\| \\
&\leq \frac{C_1 l_n}{h}
\end{aligned}$$

By using the same choice of L_n above, we have $l_n = \text{constant} \cdot \sqrt{\frac{h^3 \ln n}{n}}$, and

$$|Q_1| \leq \frac{C_1 l_n}{h^2} = O\left(\left[\frac{\ln(n)}{nh}\right]^{\frac{1}{2}}\right) \quad \text{a.s.}$$

And by exactly same argument, we can also show

$$|Q_3| = O\left(\left[\frac{\ln(n)}{nh}\right]^{\frac{1}{2}}\right) \quad \text{a.s.}$$

Thus, finally, we get

$$\sup_{x \in \mathcal{S}} |\tilde{D}_x(y) - D_x(y)| = O\left(h_0^2 + h^2 + \left[\frac{\ln(n)}{nh}\right]^{\frac{1}{2}}\right) \quad \text{a.s.} \quad \square$$

Proof of Lemma 4.9. Recall (4.3), the definition of $\tilde{D}_x(\cdot)$ and (4.5), the definition of $\tilde{D}_x^{-1}(\cdot)$, and both are sequence functions with n . Denote $D_x(\cdot)$ the function $\tilde{D}_x(\cdot)$ converge to, then $D_x(\cdot)$ and $\tilde{D}_x(\cdot)$ are CDF and 1 to 1 increasing functions and have inverse function $D_x^{-1}(\cdot)$ and $\tilde{D}_x^{-1}(\cdot)$. For a fixed $q_0 \in (0, 1)$, let $y_0 = D_x^{-1}(q_0)$. Pick $M_1 > y_0$, then $\forall \Delta$, that $0 < \Delta < D_x(M_1) - D_x(y_0)$, by Lemma 4.6, $\exists N_1 = N_1(q_0) > 0, \forall n > N_1, |\tilde{D}_x(M_1) - D_x(M_1)| < \Delta$ a.s.. It implies

$$q_0 = D_x(y_0) < D_x(M_1) - \Delta < \tilde{D}_x(M_1) \quad \text{a.s.}$$

Apparently $\tilde{D}_x^{-1}(\cdot)$ is 1-1 increasing function as well, so

$$\tilde{D}_x^{-1}(q_0) < M_1 \quad \text{a.s.}$$

Similarly, we could show that, pick $M_2 < y_0, \exists N_2 = N_2(q_0), \forall n > N_2, \tilde{D}_x^{-1}(q_0) > M_2$ a.s.. Now consider that $\forall \epsilon > 0$, by supremum triangle inequality,

$$\begin{aligned} & \sup_{M_2 < u < M_1} \left[\tilde{D}_x(u - \epsilon) - \tilde{D}_x(u) \right] \\ &= \sup_{M_2 < u < M_1} \left[\tilde{D}_x(u - \epsilon) - D_x(u - \epsilon) + D_x(u - \epsilon) - D_x(u) + D_x(u) - \tilde{D}_x(u) \right] \\ &\leq \sup_{M_2 < u < M_1} \left[\tilde{D}_x(u - \epsilon) - D_x(u - \epsilon) \right] + \sup_{M_2 < u < M_1} \left[D_x(u - \epsilon) - D_x(u) \right] \\ &\quad + \sup_{M_2 < u < M_1} \left[D_x(u) - \tilde{D}_x(u) \right] \\ &\leq 2 \sup_u \left[D_x(u) - \tilde{D}_x(u) \right] + \sup_{M_2 < u < M_1} \left[D_x(u - \epsilon) - D_x(u) \right] \end{aligned} \quad (\text{A.5})$$

$D_x(\cdot)$ is continuous 1 to 1 increasing function, so $\exists \delta = \delta(\epsilon, q_0) > 0$ such that

$$\sup_{M_2 < u < M_1} \left[D_x(u - \epsilon) - D_x(u) \right] < -2\delta < 0$$

By Lemma 4.6 again, $\exists N_3 = N_3(\epsilon, q_0) > 0, \forall n > N_3$,

$$2 \sup_u \left[D_x(u) - \tilde{D}_x(u) \right] \leq 2 \sup_u \left| D_x(u) - \tilde{D}_x(u) \right| < \delta \quad \text{a.s.}$$

Back to (A.5), we have:

$$\sup_{M_2 < u < M_1} \left[\tilde{D}_x(u - \epsilon) - \tilde{D}_x(u) \right] < -\delta \quad \text{a.s.}$$

Similarly, we could show

$$\inf_{M_2 < u < M_1} \left[\tilde{D}_x(u + \epsilon) - \tilde{D}_x(u) \right] > \delta \text{ a.s.}$$

Let $N = \max(N_1, N_2, N_3)$, $\forall n > N$, we found δ , if $|q - q_0| < \delta$,

$$\tilde{D}_x(\tilde{D}_x^{-1}(q_0) - \epsilon) - \tilde{D}_x(\tilde{D}_x^{-1}(q_0)) \leq \sup_{M_2 < u < M_1} \left[\tilde{D}_x(u - \epsilon) - \tilde{D}_x(u) \right] < -\delta$$

Thus

$$\tilde{D}_x(\tilde{D}_x^{-1}(q_0) - \epsilon) \leq q_0 - \delta < q$$

Analogically,

$$\tilde{D}_x(\tilde{D}_x^{-1}(q_0) + \epsilon) \geq q_0 + \delta > q$$

Above all, we can show for $\forall \epsilon > 0$, $\exists N = N(\epsilon, q_0) > 0$, and $\delta = \delta(\epsilon, q_0)$, $\forall n > N$, if $|q - q_0| < \delta$, then

$$\begin{aligned} \tilde{D}_x(\tilde{D}_x^{-1}(q_0) - \epsilon) &< q < \tilde{D}_x(\tilde{D}_x^{-1}(q_0) + \epsilon) \\ \tilde{D}_x^{-1}(q_0) - \epsilon &< \tilde{D}_x^{-1}(q) < \tilde{D}_x^{-1}(q_0) + \epsilon \\ \left| \tilde{D}_x^{-1}(q) - \tilde{D}_x^{-1}(q_0) \right| &< \epsilon \text{ a.s.} \end{aligned}$$

Now by the same proof of Heine-Cantor theorem, we can find finite open covers for \mathcal{C} , thus finite different N and δ work for all $q \in \mathcal{C}$. Then obviously for $\forall q_0 \in \mathcal{C}$, $\exists N = N(\epsilon) > 0$ and $\delta = \delta(\epsilon)$, such that $\forall n > N$, $|q - q_0| < \delta$, we have

$$\left| \tilde{D}_x^{-1}(q) - \tilde{D}_x^{-1}(q_0) \right| < \epsilon \text{ a.s.} \quad \square$$

Proof of Theorem 4.10. First, by Probability Integral Transform theorem, we have

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}\{g[D_x^{-1}(U)]\}$$

The first expectation is with respect to $Y|x$, the expectation on the right side is with respect to r.v. U , that follows $\text{Unif}(0,1)$. By Law of Large Number, we have

$$\mathbb{E}\{g[D_x^{-1}(U)]\} = \frac{1}{n} \sum_{i=1}^n g[D_x^{-1}(U_i)] + o_p\left(\frac{1}{\sqrt{n}}\right)$$

Where U_i follows $\text{Unif}[0,1]$ i.i.d. On the other hand, by the algorithm of model free

$$\mathbb{E}^+[g(Y^+)|X = x] = \frac{1}{n} \sum_{i=1}^n g\left(\tilde{D}_x^{-1}[\tilde{D}_{x_i}(Y_i)]\right)$$

Then, to show the theorem, we just need to prove

$$\left| \frac{1}{n} \sum_{i=1}^n g\left(\tilde{D}_x^{-1}[\tilde{D}_{x_i}(Y_i)]\right) - \frac{1}{n} \sum_{i=1}^n g[D_x^{-1}(U_i)] \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

Because we are eventually showing the convergence in probability, according to assumption (i), (v) and (vi), both X and Y have finite expectation and variance, by Chebyshev's inequality, both X and Y are bounded in probability. Furthermore, it is obvious that $D_x(Y)$ is also bounded in probability, i.e. For $\forall \epsilon > 0, \exists M_\epsilon > 0$, such that

$$\mathbb{P}(D_x(Y) < 1 - M_\epsilon \text{ and } D_x(Y) > M_\epsilon) < \epsilon$$

By Lemma 4.6, with a large enough n , $\tilde{D}_x(Y)$ is also bounded in probability: for the same $\epsilon, \exists \tilde{M}_\epsilon > 0$, such that

$$\mathbb{P}(\tilde{D}_x(Y) < 1 - \tilde{M}_\epsilon \text{ and } \tilde{D}_x(Y) > \tilde{M}_\epsilon) < \epsilon$$

Then let $M'_\epsilon = \max(M_\epsilon, \tilde{M}_\epsilon)$. For X_i, Y_i such that $\tilde{D}_x(Y_i) < 1 - \tilde{M}_\epsilon$ and $\tilde{D}_x(Y_i) > \tilde{M}_\epsilon$

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n g\left(\tilde{D}_x^{-1}[\tilde{D}_{x_i}(Y_i)]\right) - \frac{1}{n} \sum_{i=1}^n g[D_x^{-1}(U_i)]\right| \geq O\left(h^2 + \frac{1}{\sqrt{nh}}\right)\right) \\ & \leq \mathbb{P}\left(\tilde{D}_x(Y) < 1 - M'_\epsilon \text{ and } \tilde{D}_x(Y) > M'_\epsilon\right) \\ & < \epsilon \end{aligned}$$

Thus in the following proof, we only need to consider X_i, Y_i such that $1 - M'_\epsilon \leq \tilde{D}_x(Y_i) \leq M'_\epsilon$. By Lemma 4.9, we only need to prove

$$\left|\frac{1}{n} \sum_{i=1}^n g\left(\tilde{D}_x^{-1}[D_{x_i}(Y_i)]\right) - \frac{1}{n} \sum_{i=1}^n g[D_x^{-1}(U_i)]\right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

Were U_i follows $\text{Unif}[0,1]$ i.i.d.. Notice that

$$\begin{aligned} & \left|\frac{1}{n} \sum_{i=1}^n g\left(\tilde{D}_x^{-1}[D_{x_i}(Y_i)]\right) - \frac{1}{n} \sum_{i=1}^n g[D_x^{-1}(U_i)]\right| \\ & \leq \left|\frac{1}{n} \sum_{i=1}^n g\left(\tilde{D}_x^{-1}[D_{x_i}(Y_i)]\right) - \frac{1}{n} \sum_{i=1}^n g(D_x^{-1}[D_{x_i}(Y_i)])\right| \\ & + \left|\frac{1}{n} \sum_{i=1}^n g(D_x^{-1}[D_{x_i}(Y_i)]) - \frac{1}{n} \sum_{i=1}^n g[D_x^{-1}(U_i)]\right| \end{aligned} \quad (\text{A.6})$$

$D_{x_i}(Y_i)$ essentially follows $\text{Unif}(0,1)$ by Probability Integral Transform Theorem. Thus the second term is $o_p\left(\frac{1}{\sqrt{n}}\right)$ by Law of Large Number. The first term is however, complicated since $\tilde{D}_x^{-1}(\cdot)$ includes X_i, Y_i . But it is straightforward if we can show

$$\sup_{x_i, y_i} \left|g\left(\tilde{D}_x^{-1, i}[D_{x_i}(y_i)]\right) - g(D_x^{-1}[D_{x_i}(y_i)])\right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

where $\tilde{D}_x^{-1,i}(\cdot)$ denote the same function $\tilde{D}_x^{-1}(\cdot)$ but with variables x_i, y_i instead of random variables X_i, Y_i . However, the equality above might fail because $D_{x_i}(y_i)$ can reach 0 or 1, which leads $\tilde{D}_x^{-1}(\cdot)$ and $D_x^{-1}(\cdot)$ to infinity. Fortunately, we just study these functions in probability, which implies $Y|X = x$, and $D_x(Y)$ are both bounded in probability. $\forall \epsilon > 0, \exists M_\epsilon$, s.t.

$$\mathbb{P}(|Y|X = x| > M_\epsilon) < \epsilon$$

For the same ϵ , there also exists $\tilde{M}_\epsilon > 0$, such that

$$\mathbb{P}\left(D_x(Y) < 1 - \tilde{M}_\epsilon \text{ and } D_x(Y) > \tilde{M}_\epsilon\right) < \epsilon$$

Here notice that $M_\epsilon, \tilde{M}_\epsilon$ does not depend on i . Now we consider to prove

$$\sup_{x_i; y_i \in [1 - \tilde{M}_\epsilon, \tilde{M}_\epsilon]} \left| g\left(\tilde{D}_x^{-1,i}[D_{x_i}(y_i)]\right) - g\left(D_x^{-1}[D_{x_i}(y_i)]\right) \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

Since the function inside is continuous, we can attain the maximum. Denote the maximizer by $x_0^{(n)}, y_0^{(n)}$, and let $u_0^{(n)} = D_{x_0^{(n)}}(y_0^{(n)})$, then

$$\begin{aligned} \sup_{x_i; y_i \in [1 - \tilde{M}_\epsilon, \tilde{M}_\epsilon]} & \left| g\left(\tilde{D}_x^{-1,i}[D_{x_i}(y_i)]\right) - g\left(D_x^{-1}[D_{x_i}(y_i)]\right) \right| \\ & = \left| g\left(\tilde{D}_x^{-1,i,0}(u_0^{(n)})\right) - g\left(D_x^{-1}(u_0^{(n)})\right) \right| \end{aligned} \quad (\text{A.7})$$

Let $s_n = D_x^{-1}(u_0^{(n)})$, $\tilde{s}_n = \tilde{D}_x^{-1,i,0}(u_0^{(n)})$, $\tilde{t}_n = \tilde{D}_x^{i,0}(s_n)$, where $\tilde{D}_x^{-1,i,0}(\cdot)$ and $\tilde{D}_x^{i,0}(\cdot)$ denote the same function $\tilde{D}_x^{-1}(\cdot)$ and $\tilde{D}_x(\cdot)$ but with maximizer $x_0^{(n)}, y_0^{(n)}$ instead of random variables X_i, Y_i . By the definition of $s_n, \tilde{s}_n, \tilde{t}_n$, we have

$$\begin{aligned} u_0^{(n)} &= D_x(s_n) = \tilde{D}_x^{i,0}(\tilde{s}_n) \\ s_n &= D_x^{-1}(u_0^{(n)}) = \tilde{D}_x^{-1,i,0}(\tilde{t}_n) \end{aligned}$$

Then we go back the right side of equation (A.7). In the following procedures, we are using the tricks $D_x^{-1}(u_0^{(n)}) = \tilde{D}_x^{-1,i,0}(\tilde{t}_n)$, $u_0^{(n)} = D_x(s_n)$, $\tilde{t}_n = \tilde{D}_x^{i,0}(s_n)$,

$$\left| g\left(\tilde{D}_x^{-1,i,0}(u_0^{(n)})\right) - g\left(D_x^{-1}(u_0^{(n)})\right) \right| = \left| g\left(\tilde{D}_x^{-1,i,0}(u_0^{(n)})\right) - g\left(\tilde{D}_x^{-1,i,0}(\tilde{t}_n)\right) \right|$$

Since $g(x) = x^k$ or $g(x) = |x|^k$, $k = 1, 2, 3$, let $h(x) = x^k$ only, then

$$\begin{aligned} \left| g\left(\tilde{D}_x^{-1,i,0}(u_0^{(n)})\right) - g\left(D_x^{-1}(u_0^{(n)})\right) \right| &= \left| g\left(\tilde{D}_x^{-1,i,0}(u_0^{(n)})\right) - g\left(\tilde{D}_x^{-1,i,0}(\tilde{t}_n)\right) \right| \\ &\leq \left| h\left(\tilde{D}_x^{-1,i,0}(u_0^{(n)})\right) - h\left(\tilde{D}_x^{-1,i,0}(\tilde{t}_n)\right) \right| \\ &= \frac{h'\left(\tilde{D}_x^{-1,i,0}(c_n)\right)}{\tilde{D}_x'^{i,0}[\tilde{D}_x^{-1,i,0}(c_n)]} \left| u_0^{(n)} - \tilde{t}_n \right| \\ &= \frac{h'\left(\tilde{D}_x^{-1,i,0}(c_n)\right)}{\tilde{D}_x'^{i,0}[\tilde{D}_x^{-1,i,0}(c_n)]} \left| \tilde{D}_x^{i,0}(s_n) - D_x(s_n) \right| \end{aligned}$$

The third line is derived by mean value theorem. First we consider the equation $\left| \tilde{D}_x^{i,0}(s_n) - D_x(s_n) \right|$

$$\left| \tilde{D}_x^{i,0}(s_n) - D_x(s_n) \right| \leq \left| \tilde{D}_x^{i,0}(s_n) - \tilde{D}_x(s_n) \right| + \left| \tilde{D}_x(s_n) - D_x(s_n) \right|$$

On the right side, the second term is shown by Q. Li and J.S. Racine [8] in theorem 6.2, that

$$\left| \tilde{D}_x(s_n) - D_x(s_n) \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

Thus, we only need to concern the first term $\left| \tilde{D}_x^{i,0}(s_n) - \tilde{D}_x(s_n) \right|$. Actually their difference is very minor

$$\begin{aligned} \tilde{D}_x^{i,0}(x) &= \frac{\sum_{j \neq i}^n \Lambda\left(\frac{Y_j - y}{h_0}\right) K\left(\frac{X_j - x}{h}\right)}{\sum_{j \neq i}^n K\left(\frac{X_j - x}{h}\right)} \\ &= \frac{\sum_{j=1}^n \Lambda\left(\frac{Y_j - y}{h_0}\right) K\left(\frac{X_j - x}{h}\right) + \Lambda\left(\frac{y_i - y}{h_0}\right) K\left(\frac{y_i - x}{h}\right) - \Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h}\right) + K\left(\frac{x_i - x}{h}\right) - K\left(\frac{X_i - x}{h}\right)} \end{aligned}$$

Let

$$\begin{aligned} a_n &= \sum_{j=1}^n \Lambda\left(\frac{Y_j - y}{h_0}\right) K\left(\frac{X_j - x}{h}\right), \delta_{a,i} = \Lambda\left(\frac{y_i - y}{h_0}\right) K\left(\frac{y_i - x}{h}\right), \\ \Delta_{a,i} &= \Lambda\left(\frac{Y_i - y}{h_0}\right) K\left(\frac{X_i - x}{h}\right), b_n = \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right), \delta_{b,i} = K\left(\frac{x_i - x}{h}\right), \\ \Delta_{b,i} &= K\left(\frac{X_i - x}{h}\right) \end{aligned}$$

Then

$$\begin{aligned} \tilde{D}_x^{i,0}(y) &= \frac{a_n + \delta_{a,i} - \Delta_{a,i}}{b_n + \delta_{b,i} - \Delta_{b,i}} \\ &= \frac{a_n}{b_n} \left(\frac{1 - \frac{\Delta_{a,i} - \delta_{a,i}}{a_n}}{1 - \frac{\Delta_{b,i} - \delta_{b,i}}{b_n}} \right) \\ &= \frac{a_n}{b_n} \left(1 - \frac{\Delta_{a,i} - \delta_{a,i}}{a_n} \right) \\ &\quad \times \left(1 + \frac{\Delta_{b,i} - \delta_{b,i}}{b_n} + \left[\frac{\Delta_{b,i} - \delta_{b,i}}{b_n} \right]^2 + o\left[\frac{\Delta_{b,i} - \delta_{b,i}}{b_n} \right]^2 \right) \\ &= \frac{a_n}{b_n} + O_p\left(\frac{1}{n}\right) \\ &= \tilde{D}_x(y) + O_p\left(\frac{1}{n}\right) \end{aligned}$$

Above all, we have that

$$\left| \tilde{D}_x^{i,0}(s_n) - D_x(s_n) \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

It also implies

$$\left| u_0^{(n)} - \tilde{t}_n \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

We know that $u_0^{(n)}$ is bounded in probability. With a large enough n , \tilde{t}_n is also bounded in probability. And c_n is a value between $u_0^{(n)}$ and \tilde{t}_n , and also bounded in probability. This leads to $\frac{h'(\tilde{D}_x^{-1,i,0}(c_n))}{\tilde{D}_x^{i,0}[\tilde{D}_x^{-1,i,0}(c_n)]}$ is bounded in probability. Thus

$$\sup_{x_i; y_i \in [1-\tilde{M}_\epsilon, \tilde{M}_\epsilon]} \left| g\left(\tilde{D}_x^{-1,i}[D_{x_i}(y_i)]\right) - g\left(D_x^{-1}[D_{x_i}(y_i)]\right) \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

Now we go back to (A.6), it is very straightforward to see our target equation holds

$$\left| \mathbb{E}^+[g(Y^+)|X=x] - \mathbb{E}[g(Y)|X=x] \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right) \quad \square$$

Proof of Theorem 5.2. Recall lemma 5.1, we have:

$$\begin{aligned} \mathbb{E}^* \hat{m}_{n,h}^*(x) &= \frac{1}{nh} \sum_{i=1}^n \mathbb{E}^* \left[Y_i^* K\left(\frac{X_i - x}{h}\right) \right] \\ &= \frac{1}{nh} \sum_{i=1}^n \left[h \int K(u) du \mathbb{E}^*(Y^*|X=x) + O(h^2) \right] \\ &= \mathbb{E}^*(Y^*|X=x) + O(h^2) \end{aligned}$$

By theorem 3.2 & 4.10, local bootstrap method, LMF, MF and PMF all satisfy

$$\left| \mathbb{E}^*(Y^*|X=x) - \mathbb{E}(Y|X=x) \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

If denote E^* as local bootstrap, E^+ as one of model free method, then

$$\left| \mathbb{E}^*(Y^*|X=x) - \mathbb{E}^+(Y^+|X=x) \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)$$

Notice, by definition of resampling method of local bootstrap,

$$\mathbb{E}^*(Y^*|X=x) = \hat{m}_{n,h}(x)$$

As a result,

$$\mathbb{E}^* \hat{m}_{n,h}^*(x) = \frac{1}{nh} \sum_{i=1}^n \mathbb{E}^* \left[Y_i^* K\left(\frac{X_i - x}{h}\right) \right]$$

$$\begin{aligned}
&= \mathbf{E}^*(Y^*|X=x) + O(h^2) \\
&= \widehat{m}_{n,h}(x) + O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

which implies $|\mathbf{E}^*\widehat{m}_{n,h}^*(x) - \widehat{m}_{n,h}(x)| = O_p(h^2)$ \square

Proof of Theorem 5.3.

$$\begin{aligned}
&nh \left| \text{Var}^*[\widehat{m}_{n,h}^*(x)] - \text{Var}[\widehat{m}_{n,h}(x)] \right| \\
&= nh \left| \frac{1}{(nh)^2} \sum_{i=1}^n \text{Var}^* \left[Y_i^* \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] - \frac{1}{(nh)^2} \sum_{i=1}^n \text{Var} \left[Y_i \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right| \\
&\leq \frac{1}{nh} \sum_{i=1}^n \left| \mathbf{E}^* \left[Y_i^* \widetilde{K} \left(\frac{X_i - x}{h} \right) \right]^2 - \left(\mathbf{E}^* \left[Y_i^* \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right)^2 \right. \\
&\quad \left. - \mathbf{E} \left[Y_i \widetilde{K} \left(\frac{X_i - x}{h} \right) \right]^2 + \left(\mathbf{E} \left[Y_i \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right)^2 \right| \\
&\leq \frac{1}{nh} \sum_{i=1}^n \left| \mathbf{E}^* \left[Y_i^* \widetilde{K} \left(\frac{X_i - x}{h} \right) \right]^2 - \mathbf{E} \left[Y_i \widetilde{K} \left(\frac{X_i - x}{h} \right) \right]^2 \right| \\
&\quad + \frac{1}{nh} \sum_{i=1}^n \left| \left(\mathbf{E}^* \left[Y_i^* \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right)^2 - \left(\mathbf{E} \left[Y_i \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right)^2 \right| \\
&= \frac{1}{nh} \sum_{i=1}^n h \left| f_X(x) \int \widetilde{K}^2(u) du \cdot (\mathbf{E}^*(Y^{*2}|X=x) - \mathbf{E}(Y^2|X=x)) + O(h^2) \right| + \\
&\quad \frac{1}{nh} \sum_{i=1}^n \left| \left(\mathbf{E}^* \left[Y_i^* \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] + \mathbf{E} \left[Y_i \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right) \left(\mathbf{E}^* \left[Y_i^* \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right. \right. \\
&\quad \left. \left. - \mathbf{E} \left[Y_i \widetilde{K} \left(\frac{X_i - x}{h} \right) \right] \right) \right| \\
&\leq f_X(x) \int \widetilde{K}^2(u) du \cdot |\mathbf{E}^*(Y^{*2}|X=x) - \mathbf{E}(Y^2|X=x)| + O(h^2) \\
&\quad + \frac{1}{nh} \sum_{i=1}^n \left| h \left[f_X(x) \int \widetilde{K}(u) du \cdot (\mathbf{E}^*(Y^*|x) + \mathbf{E}(Y|x)) + O(h^2) \right] \right. \\
&\quad \left. \cdot h \left[f_X(x) \int \widetilde{K}(u) du \cdot (\mathbf{E}^*(Y^*|x) - \mathbf{E}(Y|x)) + O(h^2) \right] \right| \\
&\leq f_X(x) \int \widetilde{K}^2(u) du \cdot |\mathbf{E}^*(Y^{*2}|x) - \mathbf{E}(Y^2|x)| \\
&\quad + h \left(f_X(x) \int \widetilde{K}(u) du \right)^2 |\mathbf{E}^*(Y^*|x) + \mathbf{E}(Y|x)| \cdot |\mathbf{E}^*(Y^*|x) - \mathbf{E}(Y|x)|
\end{aligned}$$

Here notice that

$$\begin{aligned}
|\mathbf{E}^*(Y^*|x) + \mathbf{E}(Y|x)| &\leq |\mathbf{E}^*(Y^*|x)| + |\mathbf{E}(Y|x)| \\
&\leq |\mathbf{E}(Y|x)| + |\mathbf{E}^*(Y^*|x) - \mathbf{E}(Y|x)| + |\mathbf{E}(Y|x)| \\
&< \infty
\end{aligned}$$

Thus, recall theorem 3.2 & 4.10, we can easily get:

$$nh \left| \text{Var}^*[\widehat{m}_{n,h}^*(x)] - \text{Var}[\widehat{m}_{n,h}(x)] \right| = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right) \quad \square$$

Proof of Theorem 5.4. Define $u_i^* = \frac{1}{nh} Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right)$ and $u_i = \frac{1}{nh} Y_i \widetilde{K}\left(\frac{X_i-x}{h}\right)$. Then we have $\widehat{m}_{n,h}^*(x) = \sum_{i=1}^n u_i^*$, $\widehat{m}_{n,h}(x) = \sum_{i=1}^n u_i$. Also, let

$$S_n^{*2} = \sum_{i=1}^n \text{Var}^*(u_i^*) = \text{Var}^*[\widehat{m}_{n,h}^*(x)], \quad S_n^2 = \sum_{i=1}^n \text{Var}(u_i) = \text{Var}[\widehat{m}_{n,h}(x)]$$

Our goal is to show

$$\sqrt{nh}(\widehat{m}_{n,h}^*(x) - \mathbb{E}^*[\widehat{m}_{n,h}^*(x)]) \xrightarrow{d} N\left(0, \frac{\kappa\sigma^2(x)}{f(x)}\right)$$

which is equivalent to

$$\sqrt{nh} \sum_{i=1}^n (u_i^* - \mathbb{E}^* u_i^*) \xrightarrow{d} N\left(0, \frac{\kappa\sigma^2(x)}{f(x)}\right)$$

It is easy to see that $\{u_i^*\}$ are independent but not identical, thus we can use Lyapunov Central Limit Theorem to show the above equation. Actually, we only need to show Lyapunov Condition

$$\frac{\sum_{i=1}^n \mathbb{E}^* |u_i^* - \mathbb{E}^* u_i^*|^3}{S_n^{*3}} \xrightarrow{p} 0$$

First, recall (2.2) and theorem 5.3, we have:

$$S_n^{*3} \leq |S_n^{*3} - S_n^3| + S_n^3 = o_p(1) + S_n^3 \xrightarrow{p} \frac{1}{(nh)^{\frac{3}{2}}} \frac{\kappa\sigma^2(x)}{f_X(x)}$$

Simply, we just need

$$S_n^{*3} = O_p\left(\frac{1}{(nh)^{\frac{3}{2}}}\right) \quad (\text{A.8})$$

$$\begin{aligned} & \mathbb{E}^* |u_i^* - \mathbb{E}^* u_i^*|^3 \\ &= \frac{1}{(nh)^3} \mathbb{E}^* \left| Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) - \mathbb{E}^* \left[Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) \right] \right|^3 \\ &\leq \frac{1}{(nh)^3} \left\{ \mathbb{E}^* \left| Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) \right|^3 + 3 \left| \mathbb{E}^* \left[Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) \right] \right| \cdot \mathbb{E}^* \left[Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) \right]^2 \right. \\ &\quad \left. + 3 \left(\mathbb{E}^* \left[Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) \right] \right)^2 \cdot \mathbb{E}^* \left| Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) \right| + \left| \mathbb{E}^* \left[Y_i^* \widetilde{K}\left(\frac{X_i-x}{h}\right) \right] \right|^3 \right\} \\ &\leq \frac{1}{(nh)^3} \left\{ h \widetilde{A}_3(x) \mathbb{E}^* |Y^*|^3 + 3h^2 \widetilde{A}_1(x) \widetilde{A}_2(x) |\mathbb{E}^*(Y|x)^*| \cdot \mathbb{E}^*(Y^{*2}|x) + \right. \end{aligned}$$

$$\begin{aligned}
& \left. 3h^3 \tilde{A}_1^3(x) (E^*(Y^*|x)^2 \cdot E^*|Y^*|x| + h^3 \tilde{A}_1^3 |E^*(Y^*|x)|^3 + O(h^2)) \right\} \\
&= O\left(\frac{1}{n^3 h^2}\right) \quad \text{in probability} \\
&= O_p\left(\frac{1}{n^3 h^2}\right)
\end{aligned}$$

Here we use $E^*|Y^*|x|^k \leq |E^*|Y^*|x|^k - E|Y|x|^k| + E|Y|x|^k < \infty$.

$$\sum_{i=1}^n E^*|u_i^* - E^*u_i^*|^3 = O_p\left(\frac{1}{(nh)^2}\right)$$

Recall (A.8), we have

$$\frac{\sum_{i=1}^n E^*|u_i^* - E^*u_i^*|^3}{S_n^{*3}} = O_p\left(\frac{1}{\sqrt{nh}}\right)$$

By Lyapunov Central Limit Theorem, we have

$$\frac{\hat{m}_{n,h}^*(x) - E^*[\hat{m}_{n,h}^*(x)]}{\sqrt{\text{Var}^*[\hat{m}_{n,h}^*(x)]}} = \frac{\sum_{i=1}^n (u_i^* - E^*u_i^*)}{S_n^*} \xrightarrow{d} N(0, 1)$$

With the result of theorem 5.3, $\text{Var}^*[\hat{m}_{n,h}^*(x)] \xrightarrow{p} \text{Var}[\hat{m}_{n,h}(x)] \rightarrow \frac{1}{nh} \frac{\kappa\sigma^2(x)}{f_X(x)}$, we can finally get

$$\sqrt{nh}(\hat{m}_{n,h}^*(x) - E[\hat{m}_{n,h}^*(x)]) \xrightarrow{d} N\left(0, \frac{\kappa\sigma^2(x)}{f(x)}\right) \quad \square$$

References

- [1] J. E. Angus. The probability integral transform and related results. *SIAM Review*, 36(4):652–654, 1994. [MR1306928](#)
- [2] L. Devroye. The uniform convergence of the nadaraya-watson regression function estimate. *The Canadian Journal of Statistics*, 6(2):179–191, 1978. [MR0532857](#)
- [3] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. [MR0515681](#)
- [4] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer, New York, 1992. [MR1145237](#)
- [5] W. Hardle and A. Bowman. Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83(401):102–110, 1988. [MR0941002](#)
- [6] O. Hossjer and D. Ruppert. Asymptotics for the transformation kernel density estimator. *The Annals of Statistics*, 23(4):1198–1222, 1995. [MR1353502](#)
- [7] E. Kong, O. Linton, and Y. Xia. Uniform bahadur representation for local polynomial estimates of m-regression and its application to the additive model. *Econometric Theory*, 26(5):1529–1564, 2010. [MR2684794](#)

- [8] Q. Li and J. Racine. *Nonparametric Econometrics*. Princeton University Press, 2007. [MR2283034](#)
- [9] Q. Li and J. Racine. Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics*, 26:423–434, 2008. [MR2459343](#)
- [10] E. Masry. Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Series Analysis*, 17(6):571–599, 1996. [MR1424907](#)
- [11] E. Nadaraya. On estimating regression. *Teor. Veroyatnost. i Primenen.*, 9(1):157–159, 1964. [MR0166874](#)
- [12] L. Pan and D. N. Politis. Model-free bootstrap for markov processes. In *Proceedings of the 60th World Statistics Congress-ISI2015*, pages 26–31, Rio de Janeiro, Brazil, 2015.
- [13] D. Politis. Model-free model-fitting and predictive distributions. *TEST*, 22(2):183–221, 2013. [MR3062250](#)
- [14] D. Politis. Bootstrap confidence intervals in nonparametric regression without an additive model. In M. G. Akritas, S. Lahiri, and D. N. Politis, editors, *Topics in NonParametric Statistics: Proceedings of the First Conference of the International Society for NonParametric Statistics*, pages 271–282, New York, USA, 2014. Springer. [MR3333354](#)
- [15] D. Politis and J. Romano. On subsampling estimators with unknown rate of convergence. *Journal of the American Statistical Association*, 94(446):569–579, 1999. [MR1702326](#)
- [16] D. N. Politis. Model-free vs. model-based volatility prediction. *J. Financial Econometrics*, 5(3):358–389, 2007. [MR2708017](#)
- [17] D. N. Politis. Model-free prediction with application to functional data analysis. In E. G. Bongiorno, A. Goia, and E. Salinelli, editors, *Contributions in Infinite-Dimensional Statistics and Related Topics*, pages 221–226, Bologna, 2014. Societa Editrice Esculapio. [MR3496480](#)
- [18] D. N. Politis. *Model-Free Prediction and Regression: A Transformation-based Approach to Inference*. Springer, New York, 2015. [MR3442999](#)
- [19] M. Priestley and M. Chao. Non-parametric function fitting. *Journal of the Royal Statistical Society*, 34(3):385–392, 1972. [MR0331616](#)
- [20] Z. Qu and J. Yoon. Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, 185(1):1–19, 2015. [MR3300333](#)
- [21] D. Ruppert and D. Cline. Bias reduction in kernel density estimation by smoothed empirical transformations. *The Annals of Statistics*, 22(1):185–210, 1994. [MR1272080](#)
- [22] J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer, New York, 1995. [MR1351010](#)
- [23] S. Shi. Local bootstrap. *Ann. Inst. Statist. Math.*, 43(4):667–676, 1991. [MR1149150](#)
- [24] S. Sperlich. Comments on: Model-free model-fitting and predictive distributions. *TEST*, 22(2):227–233, 2013. [MR3062253](#)