Spring 1-1-2015

Essays in Nonparametric Estimation in Besov Spaces

Na Kyeong Lee
University of Colorado at Boulder, nakyolee@gmail.com

Essays in Nonparametric Estimation in Besov Spaces

by

Na Kyeong Lee

B.S., University of Minnesota, 2007

M.A., University of Colorado, 2012

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

2015

This thesis entitled:
Essays in Nonparametric Estimation in Besov Spaces
written by Na Kyeong Lee
has been approved for the Department of Economics

Carlos Martins-Filho, Chair

Professor Kairat Mynbaev

Date ________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Lee, Na Kyeong (Ph.D., Economics)

Essays in Nonparametric Estimation in Besov Spaces

Thesis directed by Professor Carlos Martins-Filho, Chair

In the first chapter of this thesis, a class of local constant kernel estimators for a regression in Besov spaces is developed based on a novel set of kernels provided by Mynbaev and Martins-Filho (2010). The proposed class of local constant estimators includes the Nadaraya-Watson estimator. I show that bias reduction for the estimators in the class can be achieved without the potential negativity of the underlying estimated densities. Our estimators have faster uniform convergence rates than the Nadaraya-Watson estimator. I establish consistency and asymptotic normality of the estimators in the class. These results have been established without using higher-order kernels and imposing less restrictive conditions on the true density and the regression. A Monte Carlo study is provided to illustrate the finite sample performance of the estimators.

In the second chapter of this thesis, I propose a family of estimators for a measure of polarization via a kernel-based density estimator provided by Mynbaev and Martins-Filho (2010), as well as a distribution function estimator based on integration of the estimated density. The existing estimator for polarization measure proposed by Duclos et al. (2004) is based on the empirical distribution that suffers from lack of smoothness. I modified their work by using both the density estimators proposed by Mynbaev and Martins-Filho (2010) and the integration of the estimated density. I show that a class of estimators for the distribution function is asymptotically unbiased and consistent. In addition, I establish that a class of estimators for polarization measure is asymptotically unbiased. Finally, I study the behavior of the estimators using a Monte Carlo simulation.

Dedication

This thesis is dedicated to my parents.

For their endless love, support and encouragement.

Acknowledgements

I would have never been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family.

I am deeply indebted to my advisor Professor Carlos Martins-Filho for his fundamental role in my doctoral work. I am grateful for his guidance and the opportunities he has afforded me. He is incredibly organized and a great problem solver, both of these qualities were immensely helpful in moving my project forward. Under his mentorship I have learned nonparametric econometrics, which is an invaluable skill to have as my career moves forward. I could not have imagined having a better advisor and mentor in my life.

Besides my advisor, I would like to express my gratitude to Professor Kairat Mynbaev for his encouragement, insightful comments and hard questions.

My sincere thanks also go to Dr. Xiaodong Liu, Dr. Donald Waldman, Dr. Scott Savage and Patricia C Holcomb for their support.

Last, but not the least, I take this opportunity to express my profound thanks from the depth of my heart, to my beloved parents, Jongeon Lee and Jaemook Lee for their love and continuous support-both spiritually and materially.

Contents

Tables

Table

Figures

Figure

Chapter 1

Local Constant Regression Estimation in Besov Spaces

1.1 Introduction

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a random sample from a population having density $f_{X,Y}(x, y)$. Let $f(x)$ be the marginal density of X. Consider the following nonparametric regression model

$$Y = m(X) + u, \tag{1.1}$$

where m is a real valued function, $E[u|X = x] = 0$ and $Var[u\ |X = x] = \sigma^2$. We call a kernel any function K on $\mathbb{R}$ such that $\int_{-\infty}^{\infty} K(t)dt = 1$. [20] and [26] introduced an estimator for a regression m evaluated at $x \in \mathbb{R}$ based on the Rosenblatt-Parzen estimator $\hat{f}$ for the density f which is denoted by $\hat{m}(x)$ and is given by

$$\hat{m}(x) = \frac{\frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) Y_t}{\frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)} \quad \text{where} \quad \hat{f}(x) = \frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right). \tag{1.2}$$

Here, h_n is a bandwidth sequence tending to zero as n goes to infinity. It is well known that if m has its s^{th} derivative bounded and continuous at x an interior point in the support of m and the kernel is of order s, that is, K satisfies $\int_{-\infty}^{+\infty} K(t)t^j dt = 0$ for $j = 1, 2, \cdots, s-1$ and $\int_{-\infty}^{\infty} K(t)t^s dt < \infty$, then the bias of $\hat{m}$ depends on the order s. In order to attain bias reduction, higher-order kernels ($s > 2$) have been suggested ([14], [23]). However, this approach is inconvenient since the condition that the kernel density estimator $\hat{f}$ should be a true density must be relaxed. That is, higher order kernels assign negative weights which can result in negative density estimates. There exist other approaches for bias reduction such as design-adaptive regression ([13]), data sharpening methods

([6]), iterative method ([22]) and parametrically guided nonparametric estimation ([15], [17]) but for all these methods $m(x) \in \mathcal{C}^s(\mathbb{R})$ where $\mathcal{C}^s(\mathbb{R})$ indicates the space of s-times differentiable, continuous and bounded functions in $\mathbb{R}$ for $s \in \mathbb{Z}_+$. In this paper, this assumption is substantially weakened.

[18] propose a new density estimator that achieves bias reduction relative to the Rosenblatt-Parzen estimator by introducing a family of kernels $\{M_k(x)\}_{k=1,2,\cdots}$. For a seed kernel K, natural number k and for any $x \in \mathbb{R}$,

$$M_k(x) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{x}{s}\right) \tag{1.3}$$

where the binomial coefficients $C_{2k}^N = \frac{(2k)!}{(2k-N)!N!}$, $N = 0, \cdots, 2k$, $k \in \{1, 2, \cdots\}$ and $c_{k,s} = (-1)^{s+k} C_{2k}^{s+k}$, $s = -k, \cdots, k$. [19] obtained new results on nonparametric prediction by relaxing the conditions in [5] and allowing fractional smoothness of the density.[1] In this paper, by extending the approaches of [18] and [19] we propose a new family of local constant estimators. Based on the kernels M_k in (1.3) we define a class of local constant estimators indexed by k such that

$$\hat{m}_k(x) = \frac{\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right) Y_t}{\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)} \quad . \tag{1.4}$$

The estimators $\hat{m}_k(x)$ form a general class of local constant estimators. When $k = 1$ and the seed kernel K is symmetric, our estimator $\hat{m}_1(x)$ coincides with $\hat{m}(x)$ which is given by (1.2). That is, the Nadaraya-Watson estimator $\hat{m}$ is a special case of our estimators $\hat{m}_k$.

Throughout this paper, we assume that the true regression m belongs to a Besov space $\mathcal{B}^r_{\infty,q}$ where $1 \le q \le \infty$ and $r > 0$. This assumption is desirable for the following reasons: (i) l-times continuous differentiability and uniform boundedness of m is stronger than $m \in \mathcal{B}^r_{\infty,q}$ when $l < r$, that is, $\mathcal{C}^l(\mathbb{R}) \subseteq \mathcal{B}^r_{\infty,q}$ where $\mathcal{C}^l(\mathbb{R})$ denotes the space of l times differentiable, continuous and bounded functions in $\mathbb{R}$; (ii) the space of higher order differentiable, continuous and bounded functions in $\mathbb{R}$ is a subset of the space of lower order differentiable, continuous and bounded functions, that is, $\mathcal{C}^s(\mathbb{R}) \subseteq \mathcal{C}^l(\mathbb{R})$ where $l \le s$.

[1] [19] replaced conditions (4.2) and (4.3) from [5] with their lighter assumptions 2.1 and 2.2.

The class of local constant regression estimators we propose can be used more generally than the Nadaraya-Watson estimator, since our estimators are appropriate even when a true density f or a true regression m are not differentiable. At any point at which a function has a kink or a sharp bend, the use of our estimators $\hat{m}_k$ is desirable. In many real world scenarios kinks or sharp bends are relevant. For example, suppose that m maps a general underlying trend. Then a kink can be denoted to identify a sharp change in trend from an upward trend to a downward trend. In this example, the use of Nadaraya-Watson estimator is inappropriate since m is not be differentiable at the kink point.

The first contribution of this paper is to show that the estimators $\hat{m}_k(x)$ attain a reduction in the order of the bias relative to the Nadaraya-Watson estimator while maintaining the same order for the variance. We obtain bias reduction without using higher-order kernels and potentially bypassing the disadvantage of negativity of the estimated density. The second contribution of this paper is to show that the estimators $\hat{m}_k$ are uniformly consistent. We improve the rate of uniform consistency relative to the existing literature ([10], [7], [16]) under less restrictive assumptions. The third contribution of this paper is to establish the asymptotic normality of $\hat{m}_k(x)$. The expression for the variance of the asymptotic distribution is similar to that of the Nadaraya-Watson estimator. Lastly, we conduct a Monte Carlo study to investigate the finite sample performance of the local constant estimators we propose and compare it to that of the Nadaraya-Watson estimator using a Gaussian kernel. The simulation results indicate improved performance, measured by the absolute average bias and the average root mean squared error when the kernels proposed in [18] are used.

The remainder of the paper is organized as follows. Section 1.2 provides a brief discussion of Besov spaces and discusses properties of the density estimator. In section 1.3, we provide the main asymptotic properties of local constant estimators. Section 1.4 contains a small Monte Carlo study that gives some evidence on the finite sample performance of our estimators. Section 1.5 summarizes the findings. The appendices contain all proofs, tables and figures that summarize the Monte Carlo simulation.

1.2 A Nonparametric density estimator

1.2.1 Finite differences and Besov Spaces

In this section, we define a class of density estimators $\{\hat{f}_k\}_{k=1,2,\cdots}$ using the family of kernels $\{M_k\}_{k=1,2,\cdots}$ introduced by [18]. We need a series of definitions that support the construction of the class. The properties of nonparametric density estimators are traditionally obtained by assumptions on the smoothness of the underlying density. Smoothness can be regulated by finite differences, which can be defined as forward, backward, or centered. Let $C_s^l = \frac{s!}{(s-l)!l!}$ for $l = 1, 2, \cdots, s$ and $s \in \mathbb{Z}_+$ be the binomial coefficients. An s-th order forward difference is defined by

$$\tilde{\Delta}_h^s f(x) = \sum_{j=0}^{s} (-1)^{s-j} C_s^j f(x + jh) \text{ where } s = 1, 2, \cdots. \text{ and for } h \in \mathbb{R}. \tag{1.5}$$

Lemma 1 relates forward differences to differentiability by means of a recursion. Before giving the lemma, first we define $\mathcal{D}^l f(u_1, \cdots, u_l) = \frac{\partial^l f}{\partial u_1 \cdots \partial u_l}$.

Lemma 1.1. *Let* $\tilde{\Delta}_h^0 f(x) = f(x)$, $(\tilde{\Delta}_h^s f)(x) = \tilde{\Delta}_h^1(\tilde{\Delta}_h^{s-1} f)(x)$ *where* $x \in \mathbb{R}$, $h \in \mathbb{R}_+$, $s \in \mathbb{N}$ *be the iterated differences in* $\mathbb{R}$. *For* $x \in \mathbb{R}$ *and* $s \in \mathbb{Z}_+$, *we have*

$$\tilde{\Delta}_h^s f(x) = \int_0^h \cdots \int_0^h \tilde{\Delta}_h^{s-l} \mathcal{D}^l f\left(x + \sum_{i=1}^{l} u_i\right) \prod_{i=1}^{l} du_i \qquad \textit{where } l = 1, 2, \cdots, s. \tag{1.6}$$

When we consider forward even-order difference, (1.5) can be written as

$$\tilde{\Delta}_h^{2k} f(x) = \sum_{|s|=0}^{k} c_{k,s} f(x + kh + sh) \tag{1.7}$$

where $c_{k,s} = (-1)^{s+k} C_{2k}^{s+k}$ for $s = -k, \cdots, k$ and $k \in \{1, 2, \cdots\}$. It is easy to verify that for $s = 2k$, $\tilde{\Delta}_h^{2k} f(x) = \sum_{j=0}^{2k} (-1)^{2k-j} C_{2k}^j f(x + jh) = \sum_{|s|=0}^{k} (-1)^{s+k} C_{2k}^{s+k} f(x + kh + sh)$.

Next, we introduce Besov spaces $\mathcal{B}_{p,q}^r(\mathbb{R})$ where $1 \le p \le \infty$, $1 \le q \le \infty$, $r > 0$, and the norm in $\mathcal{B}_{p,q}^r(\mathbb{R})$ is defined by $||f||_{\mathcal{B}_{p,q}^r} = ||f||_{b_{p,q}^r} + ||f||_p$ where the first part $||f||_{b_{p,q}^r}$ characterizes smoothness of f and is given by

$$||f||_{b_{p,q}^r} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} \left| \tilde{\Delta}_h^{2k} f(x) \right|^p dx \right)^{1/p}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{1/q}$$

for $k \in \mathbb{Z}_+$ satisfying $2k > r$ ([24], [19]). When $p = \infty$ and/or $q = \infty$, the integral(s) is (are) replaced by supremum and we write, for example, $\sup_{x\in\mathbb{R}} |\Delta_h^{2k} f(x)| = ||\Delta_h^{2k} f(x)||_\infty$. $\mathcal{C}^0(\mathbb{R})$ is defined as the collection of all real-valued, bounded and uniformly continuous functions in $\mathbb{R}$, equipped with the norm $||f|\mathcal{C}^0(\mathbb{R})||_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. [2] The following lemma shows that the class $\mathcal{C}^l(\mathbb{R})$ is a subset of $\mathcal{B}^r_{\infty,q}$ whenever $l < r$.

Lemma 1.2. *If* $l = 1, 2, 3, \cdots$, *we define* $\mathcal{C}^l(\mathbb{R}) = \{f|\mathcal{D}^l f \in \mathcal{C}^{l-1}(\mathbb{R})\}$. *Let* $0 \le q \le \infty$. *For* $r > l$, *we have*

$$\mathcal{C}^l(\mathbb{R}) \subseteq \mathcal{B}^r_{\infty,q}(\mathbb{R}). \tag{1.8}$$

A full description of the relationships between $\mathcal{C}^l(\mathbb{R})$ and a Besov space $\mathcal{B}^r_{p,q}$ can be found in [4]. Since,

$$M_k(x) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{x}{s}\right) \tag{1.9}$$

we can express the bias of our proposed estimators $\hat{m}_k$ in terms of higher order finite differences. Let $\lambda_{k,s} = \frac{(-1)^{s+1}(k!)^2}{(k+s)!(k-s)!}$ where $s = 1, 2, \cdots, k$ and since $-\frac{c_{k,s}}{c_{k,0}} = -\frac{c_{k,-s}}{c_{k,0}} = \lambda_{k,s}$, $s = 1, \cdots, k$, we can write $M_k(x) = \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s} \left(K\left(\frac{x}{s}\right) + K\left(-\frac{x}{s}\right)\right)$. Consequently, $M_k(x) = M_k(-x)$ for $x \in \mathbb{R}$, that is, M_k is symmetric. Since the coefficients $c_{k,s}$ satisfy $\sum_{|s|=0}^{k} c_{k,s} = (1-1)^{2k} = 0$, the following equations are true:

$$-\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} = 1 \quad \text{and} \quad \sum_{s=1}^{k} \lambda_{k,s} = \frac{1}{2}. \tag{1.10}$$

Equation (1.10) and $\int K(\psi)d\psi = 1$ imply that

$$\int M_k(\psi)d\psi = \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s} \left[\int K\left(\frac{\psi}{s}\right) d\psi + \int K\left(-\frac{\psi}{s}\right) d\psi\right] = 1, \quad \text{for all } \mathbb{R}.$$

The kernel M_k defines a new family of density estimators indexed by k as follows,

$$\hat{f}_k(x) = \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right) \tag{1.11}$$

[2] See [25].

where h_n is a bandwidth sequence tending to zero as $n \to \infty$. When $k = 1$ and K is symmetric, the density estimator in (1.11) coincides with the Rosenblatt-Parzen density estimator. Since the kernel $M_k(x)$ is symmetric, by using forward even-order differences (1.7), for any function f we have

$$\Delta_h^{2k} f(x) = \sum_{s=-k}^{k} c_{k,s} f(x + sh) \quad \text{for } h \in \mathbb{R}.$$

It is easy to verify that $\tilde{\Delta}_h^{2k} f(x) = \Delta_h^{2k}[f(x + kh)]$ ([19]). Hence, we use centered even-order difference for a smoothness characteristic, and we have

$$||f||_{b_{p,q}^r} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} \left|\Delta_h^{2k} f(x)\right|^p dx\right)^{1/p}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{1/q}.$$

1.2.2 Density Estimation

We now list assumptions that will be used throughout the paper.

Assumption 1 : $\{Y_t, X_t\}_{t=1}^n$ is an IID sequence.

Assumption 2 : (1) $f \in \mathcal{B}_{\infty,q}^r$ with $r > 0$ and $1 \leq q \leq \infty$; (2) $f \in \mathcal{C}^0(\mathbb{R})$; (3) f is bounded away from 0.

Assumption 3 : $h_n > 0$ for all n, $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

Assumption 4 : For all $x \in \mathbb{R}$,

(1) $K(x) : \mathbb{R} \to \mathbb{R}$ is a measurable function. (2) $\int K(x)dx = 1$; (3)$\int |K(x)|dx < \infty$; (4) $\sup_{x\in\mathbb{R}} |K(x)| < C < \infty$; (5) $|K(x) - K(x')| < c|x - x'|$ for some $c < \infty$ and $x \neq x'$, $x, x' \in \mathbb{R}$.

The following theorem gives the bias and its order for the density estimator $\hat{f}_k$.

Theorem 1.1. *Suppose* Assumption 1, Assumption 2(1) *and* Assumption 4(1)-(2) *hold. In addition, suppose that* $\left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q'} d\psi\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for* $1 \leq q \leq \infty$. *For* *all* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots$, *we have*

$$(a)\ Bias(\hat{f}_k(x)) = \int -\frac{1}{c_{k,0}} K(\psi) \Delta_{h\psi}^{2k} f(x) d\psi$$

$$(b)\ |Bias(\hat{f}_k(x))| \leq ch_n^r \left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q'} d\psi \right]^{1/q'} ||f||_{b_{\infty,q}^r} \quad where\ \ 2k > r.$$

We note that the order of the bias for our estimator is similar to that attained by the Rosenblatt density estimator constructed with a kernel of order r and $f \in \mathcal{C}^r(\mathbb{R})$. Given Assumption 3 we have that $Bias(\hat{f}_k(x)) \to 0$ as $n \to \infty$, which implies that $\hat{f}_k$ is asymptotically unbiased. The following theorem deals with the consistency of $\hat{f}_k$.

Theorem 1.2. *Suppose* Assumption 1, Assumption 2(1)-(2), Assumption 3 *and* Assumption 4(1)-(4) *hold. In addition, suppose that* $\left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q'} d\psi\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for* $1 \leq q \leq \infty$. *Then, for all* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots$,

$$\hat{f}_k(x) - f(x) = o_p(1).$$

It is of interest to establish the uniform consistency of $\hat{f}_k$. The following theorem provides conditions under which $\hat{f}_k(x)$ converges to $f(x)$ uniformly in probability.

Theorem 1.3. *Suppose* Assumption 1, Assumption 2(1)-(2), Assumption 3 *and* Assumption 4(1)-(5) *hold. In addition, suppose that* $\left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q'} d\psi\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for* $1 \leq q \leq \infty$. *Let* $\mathcal{G}$ *be a compact subset of* $\mathbb{R}$. *For all* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots$, *we have*

$$\sup_{x \in \mathcal{G}} |\hat{f}_k(x) - f(x)| = O_p\left(\left(\frac{\log n}{nh_n} \right)^{1/2} + h_n^r \right). \tag{1.12}$$

Uniform consistency of the density estimator requires $\left(\frac{\log n}{nh_n}\right) \to 0$ as $n \to \infty$. From (1.12), the order of $\hat{f}_k$ is similar to that attained by Rosenblatt density estimator with a kernel of order r and $f \in \mathcal{C}^r(\mathbb{R})$. We achieve faster uniform convergence rate by imposing the less restrictive assumption $f \in \mathcal{B}_{\infty,q}^r$. The next theorem gives the asymptotic normality of the density estimator $\hat{f}_k(x)$ for all

$x \in \mathbb{R}$ under suitable normalization.

Theorem 1.4. *Suppose* Assumption 1, Assumption 2(1)-(2), Assumption 3 *and* Assumption 4(1)-(4) *hold. Then, for all $x \in \mathbb{R}$ and $k = 1, 2, \cdots$, we have*

$$\sqrt{nh_n}\left(\hat{f}_k(x) - f(x) + O(h_n^r)\right) \xrightarrow{d} \mathcal{N}\left(0, f(x)\int M_k^2(\psi)d\psi\right).$$

Suppose, additionally, that $nh_n^{1+2r} \to 0$ as $n \to \infty$. Then

$$\sqrt{nh_n}\left(\hat{f}_k(x) - f(x)\right) \xrightarrow{d} \mathcal{N}\left(0, f(x)\int M_k^2(\psi)d\psi\right). \tag{1.13}$$

This result is similar to that attained for a Rosenblatt density estimator with the exception that K is replaced by the M_k kernel in the expression for the variance of the asymptotic distribution. In order to obtain asymptotic normality of $\sqrt{nh_n}(\hat{f}_k - f)$, we write $\sqrt{nh_n}(\hat{f}_k(x) - f(x)) = \sqrt{nh_n}(\hat{f}_k(x) - E[\hat{f}_k(x)]) + \sqrt{nh_n}(E[\hat{f}_k(x) - f(x)])$. From Theorem 1.1, we know the second term in the decomposition is of order $\sqrt{nh_n}\ O(h_n^r)$. The quantity $\sqrt{nh_n}(\hat{f}_k(x) - f(x))$ will be asymptotically normally distributed with mean zero if the second term in the decomposition tends to zero as $n \to \infty$. Thus, we need $nh_n^{1+2r} \to 0$ as $n \to \infty$. In this case we obtain equation (1.13).

1.3 Local Constant Estimator

In this section, we investigate the asymptotic properties of the estimators $\hat{m}_k$ for $k = 1, 2, \cdots$. We assume that the conditional density of Y given $X = x$ exists and is denoted by $f_{Y|X}(y) = \frac{f_{Y,X}(y,x)}{f(x)}$ where $f_{Y,X}$ is the density of (Y, X) and $f(x)$ denotes the marginal density of X with $f(x) \neq 0$. If the conditional expectation $E[Y_t|X_t = x]$ exists, we write

$$m(x) = E[Y_t|X_t = x] = \int y f_{Y|X}(y)dy = \int y \frac{f_{Y,X}(y,x)}{f(x)}dy.$$

Assumption 5 : (1) $m \in \mathcal{B}^{\rho}_{\infty,\infty}$ with $\rho > r$ where r is as in Assumption 2 (1) and $\mathcal{B}^{\rho}_{\infty,\infty}(\mathbb{R})$ is a Zygmund space $\mathcal{Z}^{\rho}(\mathbb{R})$; (2) $m \in \mathcal{C}^0(\mathbb{R})$.

By Corollary 2.8.2 (i) in [24], multiplication by a function $m \in \mathcal{Z}^\rho(\mathbb{R})$ is bounded in $\mathcal{B}^r_{p,q}$ if $\rho > r$, that is

$$||mf||_{\mathcal{B}^r_{p,q}} \leq c||m||_{\mathcal{Z}^\rho}||f||_{\mathcal{B}^r_{p,q}}. \tag{1.14}$$

In the existing literature, for the Nadaraya-Watson estimator it is assumed that the regression function $m(\cdot)$ is continuous, uniformly bounded and differentiable. From Lemma 1.2, Assumption 5 seems desirable since $\mathcal{B}^r_{\infty,q}$ is wider than $\mathcal{C}^l(\mathbb{R})$ where $l \leq r$. That is, we impose less restrictive assumptions than the existing literature for (1.1). We make the following additional assumption.
Assumption 6: $E[|Y - m(X)|^{2+\delta}|X] < \infty$ for $\delta > 0$ and denote $Var(Y|X = x) = \sigma^2$.

The estimators $\hat{m}_k$ are similar to the Nadarya-Watson estimator with the exception that K is replaced by M_k kernel. When $k = 1$ and a seed kernel K is symmetric, the estimator $\hat{m}_1(x)$ concides with the Nadaraya-Watson estimator (henceforth NW). Thus, the NW estimator is an element of the class defined in (1.4). To investigate the asymptotic properties of $\hat{m}_k$, we write

$$\hat{m}_k(x) = \frac{\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)Y_t}{\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)} = \frac{\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)Y_t}{\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)} = \frac{\hat{g}_k(x)}{\hat{f}_k(x)}$$

where $\hat{g}_k(x) \equiv \hat{m}_k(x)\hat{f}_k(x) = \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)Y_t$ for $x \in \mathbb{R}$. We put $g(x) \equiv m(x)f(x)$. From (1.14), Assumption 2(1) and Assumption 5(1), we know $g \in \mathcal{B}^r_{\infty,q}$ since $||g||_{\mathcal{B}^r_{p,q}} \leq c||m||_{\mathcal{Z}^\rho}||f||_{\mathcal{B}^r_{p,q}}$ for $r < \rho$ ([24]). In the previous section, we considered the properties of $\hat{f}_k$, so the only step needed to investigate the properties of $\hat{m}_k$ is to consider the properties of $\hat{g}_k(x)$.

Theorem 1.5. *Suppose* Assumption 1-3*,* Assumption 4(1)-(4)*,* Assumption 5*, and* Assumption 6 *hold. In addition, suppose that* $\left[\int |K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for*

$1 \leq q \leq \infty$. *For* $x \in \mathbb{R}$. *Then for* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots$, *we have*

$$(a)\ Bias(\hat{g}_k(x)) = \left(-\frac{1}{c_{k,0}}\right) \int K(\psi) \Delta^{2k}_{h_n \psi} g(x) d\psi,$$

$$(b)\ |Bias(\hat{g}_k(x))| \leq C h_n^r \left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q'} d\psi\right]^{1/q'} ||g||_{b^r_{\infty,q}},$$

$$(c)\ \hat{g}_k(x) - g(x) = o_p(1) \quad \textit{for } g(x) = m(x)f(x).$$

Avoiding higher-order restrictions and using fractional smoothness on m and f, we obtain the order of the bias of $\hat{m}_k(x)\hat{f}_k(x)$ to be $O(h_n^r)$ where $2k > r$. The following theorem provides conditions under which $\hat{g}(x)$ converges to $E[\hat{g}(x)]$ uniformly in probability.

Theorem 1.6. *Suppose* Assumption 1, Assumption 2(2), Assumption 3, Assumption 4(1),(4) *and* (5), Assumption 5(2) *and* Assumption 6 *hold. In addition, suppose that* $\frac{nh_n}{\log n} \to \infty$ *as* $n \to \infty$. *For* $k = 1, 2, \cdots$,

$$\sup_{x \in \mathcal{G}} \left| \hat{g}_k(x) - E[\hat{g}_k(x)] \right| = O_p\left(\left(\frac{\log n}{nh_n}\right)^{1/2}\right) \tag{1.15}$$

where $\mathcal{G}$ *is a compact set in* $\mathbb{R}$.

We now establish asymptotic normality of $\hat{g}_k(x)$ under a suitable normalization.

Theorem 1.7. *Suppose* Assumption 1-3, Assumption 4(1)-(4), Assumption 6 *hold. For* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots$, *we have*

$$\sqrt{nh}\left[\hat{g}_k(x) - E(\hat{g}_k(x)|X_t)\right] \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 f(x) \int M_k^2(\psi) d\psi\right).$$

Given $\hat{f}_k(x)$ such that $\hat{f}_k(x) = f(x) + o_p(1)$ in Theorem 1.2, we have

$$E[\hat{m}_k(x) - m(x)] = \frac{1}{f(x)} \left(-\frac{1}{c_{k,0}}\right) \int K(\psi) \Delta^{2k}_{h_n \psi} m(x) f(x) d\psi.$$

Given the results on $\hat{g}_k$ and $\hat{f}_k$, we obtain following properties for $\hat{m}_k(x)$. First, Theorem 1.8 gives the order of the bias for $\hat{m}_k$.

Theorem 1.8. *Suppose* Assumption 1-2*,* Assumption 4(1)-(4) *and* Assumption 5 *hold. In addition, suppose that* $\left[\int |K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'} < \infty$ *where* $1/q+1/q'=1$ *for* $1\le q\le\infty$*. For* $x\in\mathbb{R}$ *and* $k=1,2,\cdots,$ *we have* $|Bias(\hat{m}_k(x))| = O(h_n^r)$*.*

Note that the order of the bias for our estimators are similar to that attained by the NW estimator constructed with a kernel of order r. It is interesting to compare the order of the bias for the estimator $\hat{m}_k$ to that of the NW estimator. It is worth noting that in Theorem 1.8 symmetry of K is not required, nor is compactness of its support. The advantage of our estimator $\hat{m}_k$ for $k=1,2,\cdots$ is that we achieve bias reduction and avoid negative density estimators by imposing less restrictive conditions i.e., $f\in B^r_{\infty,q}$ and $m\in B^\rho_{\infty,\infty}$ where $\rho > r$.

Next theorem states that $\hat{m}_k(x)$ converges to $m(x)$ uniformly in probability.

Theorem 1.9. *Suppose* Assumption 1-6 *hold.*
In addition, suppose that $\left[\int |K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'} < \infty$ *where* $1/q+1/q'=1$ *for* $1\le q\le\infty$*.*
For $x\in\mathbb{R}$*,* $k=1,2,\cdots,$

$$\sup_{x\in\mathcal{G}}|\hat{m}_k(x)-m(x)| = O_p\left(h_n^r + \left(\frac{\log n}{nh_n}\right)^{1/2}\right).$$

Uniform consistency of $\hat{m}_k$ requires $\left(\frac{\log n}{nh_n}\right)\to 0$ as $n\to\infty$. We improve the rate of uniform consistency relative to the existing literatures ([10], [7], [16]) by avoiding higher-order conditions on the kernel and imposing less restrictive conditions. Consistency follows from Theorem 1.9. We now give sufficient condition for asymptotic normality of $\hat{m}_k(x)$ under suitable centering and normalization.

Theorem 1.10. *Suppose* Assumption 1-6 *hold.*
In addition, suppose that $\left[\int |K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'} < \infty$ *where* $1/q+1/q'=1$ *for* $1\le q\le\infty$*.*
For $x\in\mathbb{R}$ *and* $k=1,2,\cdots,$ *we have*

$$\sqrt{nh_n}\left(\hat{m}_k(x)-m(x)+O_p(h_n^r)\right)\xrightarrow{d}\mathcal{N}\left(0,\sigma^2 f(x)^{-1}\int M_k^2(\psi)d\psi\right).$$

Suppose, additionally, that $nh_n^{1+2r} \to 0$ *as* $n \to \infty$*, then*

$$\sqrt{nh_n}\left(\hat{m}_k(x) - m(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 f(x)^{-1} \int M_k^2(\psi)d\psi\right).$$

For the local constant estimator the normalizing factor is $(nh_n)^{1/2}$ and we use the decomposition $(nh_n)^{1/2}[\hat{m}_k(x) - m(x)] = (nh_n)^{1/2}\left[\hat{m}_k(x) - E(\hat{m}_k(x)|X_t)\right] + (nh_n)^{1/2}\left[E(\hat{m}_k(x)|X_t) - m(x)\right]$. The first term in the decomposition is asymptotically normally distributed and the second term is the conditional bias $[E(\hat{m}_k(x)|X_t) - m(x)] = O(h_n^r)$. To eliminate the asymptotic bias in the limiting distribution of the estimator, we need an additional assumption such as $nh_n^{1+2r} \to 0$ as $n \to \infty$. The expression for the variance term of the asymptotic distribution is similar to that the NW estimator with exception that K is replaced by M_k kernel.

1.4 Monte Carlo Study

In this section we perform a small Monte Carlo study to investigate the finite sample performance of our proposed local constant estimator. For comparison purpose, we also implement the Nadaraya-Watson estimator, which is given by $\hat{m}_{NW}(x) \equiv \hat{m}_1(x) \equiv \frac{(nh_n)^{-1}\sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right)Y_j}{(nh_n)^{-1}\sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right)}$ with $K(\cdot)$ is a Gaussian kernel. We consider following data generating processes (DGPs),

$$DGP1: \ y = m_1(x) + \epsilon, \ m_1(x) = 3x + \frac{20}{\sqrt{2\pi}} exp\left\{-100(x-0.5)^2\right\}$$

where $X \sim N(\mu_X, \sigma_X^2), \epsilon \sim N(0, \sigma_\epsilon^2), \ \mu_X = 0.5, \ \sigma_X^2 = 1/3.92^2, \sigma_\epsilon^2 = 0.673$

$$DGP2: \ y = m_2(x) + \epsilon, \ m_2(x) = exp\{x\}sin(5x^2),$$

where $X \sim N(\mu_x, \sigma_X^2), \epsilon \sim N(0, \sigma_\epsilon^2), \ \mu_X = 0, \ \sigma_X^2 = 1, \sigma_\epsilon^2 = 2$

$$DGP3: \ y = binornd(1, m_3(x)), \ m_3(x) = 0.5sin(10\pi x) + 0.5, \ X \sim U[0, n]$$

$$DGP4: \ y = binornd(1, m_4(x)), \ m_4(x) = 0.5sin(2\pi x) + 0.5, \ X \sim U[0, n]$$

where $y = binornd(1, m)$ generates random numbers from the binomial distribution with parameters specified by the number of trials 1, and probability of success for each trial m. In each DGP, we evaluate the regression at 101 points from 0 to 1 with increments 0.01. At each point, we compute absolute bias, variance and root mean square error. Then, we average the absolute bias, variance and root mean squared error across all 101 evaluation points. A Gaussian seed kernel is used to construct the estimators.

In our simulations, for each of these DGPs, 1000 samples of size $n = 400$ and 1000 were considered and four estimators $\hat{m}_{NW}$, $\hat{m}_2$, $\hat{m}_3$ and $\hat{m}_4$ were studied. For values of x where the denominator of the local constant estimator, i.e.,$\hat{f}_k(x)$ is close to zero, the local constant estimator at x is ill defined. To avoid this situation, we introduce a trimming parameter $\delta > 0$. That is, we only consider the observations where the density estimate $\hat{f}_k$ is above δ. We select both bandwidth h and the trimming parameter δ by minimizing a cross validation criterion. $CV(h, \delta) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{m}_{-i}(X_i)\right)^2$ where

$$\hat{m}_{-i}(X_i) = \frac{\frac{1}{nh_n}\sum_{l\neq i}^{n} Y_l M_k\left(\frac{X_i - X_l}{h_n}\right)}{\left[\frac{1}{nh_n}\sum_{l\neq i}^{n} M_k\left(\frac{X_i-X_l}{h_n}\right)\right]\mathbf{I}\left(\left[\frac{1}{nh_n}\sum_{l\neq i}^{n} M_k\left(\frac{X_i-X_l}{h_n}\right)\right]\geq\delta\right) + \delta\mathbf{I}\left(\left[\frac{1}{nh_n}\sum_{l\neq i}^{n} M_k\left(\frac{X_i-X_l}{h_n}\right)\right]<\delta\right)}$$

and $\mathbf{I}(\cdot)$ is an indicator function. For each estimator at point x where the denominator was smaller than δ, it was replaced by δ. Table A.1 provides average absolute bias (B), average variance (V) and average root MSE (R) for each estimator considered for $n = 400, 1000$, respectively.

Table A.1, Figure A.1 and A.2 reveal the following general regularities. First, for all four DGPs the average absolute bias (B), average variance (V) and average root MSE (R) of our estimators $\hat{m}_k$ decrease as the sample size increases from 400 to 1000. Box plots also show that the root mean squared error falls as the sample size increases. Second, as expected from the theoretical results, an increase in the value of k reduces average absolute bias (B). Third, for $k = 2$ the case where the smallest bias reductions are attained, (B) can be reduced by as much as 58% relative to $\hat{m}_{NW}$. Fourth, reduction in root mean square error (R) due to the increase in k is much less pronounced. When we observe the true regression function m_4, root mean squared error (R) tends to increase as k rises but the largest difference of the root mean squared errors between $\hat{m}_{NW}$ and $\hat{m}_k$ is negligible magnitude. Finally, we observe that for DGP1, DGP3 and DGP4, our proposed estimators $\hat{m}_2$, $\hat{m}_3$ and $\hat{m}_4$ outperform the Nadaraya-Watson estimator $\hat{m}_{NW}$ in terms of both the average absolute bias (B) and root mean squared error (R) and among all estimators, $\hat{m}_2$ achieves the smallest root mean square error (R).

1.5 Summary

The use of higher order kernels is a well-known method for bias reduction in density and regression estimation. This method of bias reduction has the disadvantage of potential negativity of the underlying estimated density. To avoid this, [18] pioneered a new set of nonparametric kernel based estimators for a density that achieves bias reduction by using a new family of kernels. In addition, [19] obtained much faster convergence of nonparametric prediction by allowing fractional smoothness for the relevant densities. By extending both approaches, in this paper, we propose local constant estimators for regression which are more general than the Nadaraya-Watson (NW) estimator. The main contribution in this paper is that bias reduction may be achieved relative to the NW estimator, and our proposed estimators attain faster uniform convergence without using higher-order kernels and allowing for fractional smoothness for the relevant densities and regressions. We also provide consistency and asymptotic normality of the estimators in the class we propose. A small Monte Carlo study reveals that our estimator performs well relative to the NW estimator and the promised bias reduction is obtained, experimentally in finite samples.

Chapter 2

Kernel Density Estimation for a Polarization Measure

2.1 Introduction

According to the OECD ([21]), over the past two decades, income inequality as measured by the Gini coefficient, has widened in most OECD countries. For example, in the U.S., more than 40 percent of total income is owned by the wealthiest 10 percent of the population. Thus, there has been growing interest in the area of measuring inequality and the consequences of unequal economic performance during last few decades ([12],[27], [28], [11], [2], [1]).

The most well known and widely used measures of inequality are Gini, Atkinson and General entropy measures. The existing standard measures, however, are not able to explain all characteristics of inequality. Since these measures focus on deviations from a global mean and assessing the expected divergence, clustering around local mean or disappearing middle class can not be explained by the existing inequality measures ([27]). Therefore, polarization measures have been suggested to explain these characteristics that inequality measures fail to capture.

[12] measured polarization in one dimension designed for discrete random variables. Later, [11] proposed a polarization measure called the DER index, [11] also provided an estimator for the case of continuous random variables. The proposed estimator derived by [11], however, is based on a Rosenblatt-Parzen density estimator and the empirical distribution function that suffers from lack of smoothness.

In this paper, for the estimation of the DER index, we use alternative nonparametric kernel based density estimators, $\hat{f}_k$ that were introduced by [18]. These new density estimators $\hat{f}_k$

achieve bias reduction relative to the Rosenblatt-Parzen density estimator $\hat{f}$. In addition, instead of using the empirical distribution function, we obtain the estimator of the DER index based on integrating $\hat{f}_k$. Let F be the distribution with the density f such that $F(x) = \int_{-\infty}^{x} f(v)dv$. We denote a class of distribution estimators by integrating $\hat{f}_k$ as $\hat{F}_k(x) = \int_{-\infty}^{x} \hat{f}_k(v)dv$. Integrating $\hat{f}_k$ seems desirable for the following reasons: (i) $\hat{F}_k$ forms a general class of nonparametric distribution estimators. When $k = 1$, the newly developed distribution estimators $\hat{F}_k$ coincide with integrating a Rosenblatt-Parzen density estimator. (ii) mean squared error of the distribution estimators obtained by integrating density estimator $\hat{f}_k$ can be smaller than that of the empirical distribution function.

Throughout this paper, we assume that the true density f belongs to a Besov space $\mathcal{B}^r_{\infty,q}$ where $1 \leq q \leq \infty$ and $r > 0$. This assumption is desirable since l-times continuous differentiability and uniform boundedness of f is stronger than $f \in \mathcal{B}^r_{\infty,q}$ where $l < r$, that is $\mathcal{C}^l(\mathbb{R}) \subseteq \mathcal{B}^r_{\infty,q}$ where $\mathcal{C}^l(\mathbb{R})$ denotes the space of l-times differentiable and continuous and bounded functions in $\mathbb{R}$. [1] We also assume that $F \in \mathcal{B}^{r+1}_{\infty,q}$. This assumption is reasonable in the following sense. The Besov space $\mathcal{B}^r_{\infty,q}$, $r > 0$ and $0 < q \leq \infty$ is a smooth subspace of $\mathcal{L}_\infty(\mathbb{R})$ [9].[2] The parameter r gives the order of smoothness in $\mathcal{L}_\infty(\mathbb{R})$. Since F is once-differentiable, the order of smoothness for F in the Besov space, $r + 1$ is greater than that for f in the Besov space, r.

The first contribution of this paper is that we achieve bias reduction for $\hat{F}_k$ relative to integrating the Rosenblatt-Parzen density estimator $\hat{f}$ without using higher-order kernels. The second contribution of this paper shows that our class of estimators for the DER index $P_\alpha(\hat{F}_k)$ is asymptotically unbiased. Lastly, we conduct a Monte Carlo study to investigate the finite sample performance of our estimators $P_\alpha(\hat{F}_k)$ we propose and compare it to that of the existing estimator $P_\alpha(\tilde{F})$ where $\tilde{F}$ is the empirical distribution function. The simulation results indicate improved performance measured by bias and root mean squared error.

The remainder of the paper is organized as follows. Section 2.2 provides the estimator for

[1] see Lemma 1.2 in Chapter 1

[2] $\mathcal{L}_\infty$ generalizes the $\mathcal{L}_p$ spaces to $p = \infty$. The norm on $\mathcal{L}_\infty$ is given by the essential supremum.

the polarization measure. Section 2.3 contains a Monte Carlo study that implements our proposed estimator and compares its performance with that of the estimator suggested by [11]. Lastly, section 2.4 provides a summary.

2.2 Estimator for a Polarization Measure

2.2.1 Finite difference and Besov Spaces

In this section, we provide a class of density estimators $\{\hat{f}_k\}_{k=1,2,\cdots}$ and an associate class of distribution estimators $\{\hat{F}_k\}_{k=1,2,\cdots}$ using the family of kernels $\{M_k\}_{k=1,2,\cdots}$ introduced by [18]. We need a series of definitions that support the construction of the class. Properties of nonparametric density estimators and the smoothed estimators of the distribution functions are traditionally obtained by using assumptions on the smoothness of the underlying density and cumulative distribution function. Smoothness can be regulated by finite differences which can be defined as forward, backward or centered. Let $C_s^l = \frac{s!}{(s-l)!l!}$ for $l = 1, 2, \cdots, s$ and $s \in \mathbb{Z}_+$ be the binomial coefficients. A s-th order forward difference is defined by

$$\tilde{\Delta}_h^s f(x) = \sum_{j=0}^{s} (-1)^{s-j} C_s^j f(x + jh) \quad \text{where } s = 1, 2, \cdots, \text{ for } x \text{ and } h \in \mathbb{R}. \tag{2.1}$$

When we consider forward even-order differences, (2.1) can be written as

$$\tilde{\Delta}_h^{2k} f(x) = \sum_{|s|=0}^{k} c_{k,s} f(x + kh + sh) \quad \text{for } k = 0, 1, 2 \cdots \tag{2.2}$$

where $c_{k,s} = (-1)^{s+k} \mathrm{C}_{2k}^{s+k}$ for $s = -k, \cdots, k$ and $k \in \{1, 2, \cdots\}$. From chapter 1, for $s = 2k$, $\tilde{\Delta}_h^{2k} f(x) = \sum_{j=0}^{2k} (-1)^{2k-j} C_{2k}^j f(x + jh) = \sum_{|s|=0}^{k} (-1)^{s+k} C_{2k}^{s+k} f(x + kh + sh)$. Again from chapter 1, we introduce Besov spaces $\mathcal{B}_{p,q}^r(\mathbb{R})$ where $1 \le p \le \infty$, $1 \le q \le \infty$, $r > 0$, with norm defined by $||f||_{B_{p,q}^r} = ||f||_{b_{p,q}^r} + ||f||_p$ where the first part $||f||_{b_{p,q}^r}$ characterizes smoothness of f and is given by

$$||f||_{b_{p,q}^r} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} \left| \tilde{\Delta}_h^{2k} f(x) \right|^p dx \right)^{1/p}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{1/q}$$

for $k \in \mathbb{Z}_+$ satisfying $2k > r$. When $p = \infty$ and/or $q = \infty$, the integral(s) is (are) replaced by supremum. As before, $\mathcal{C}^0(\mathbb{R})$ is defined as the collection of all real-valued, bounded and uniformly continuous functions in $\mathbb{R}$, equipped with the norm $||f||_{\mathcal{C}^0} = \sup_{x\in\mathbb{R}} |f(x)|$([25]).

Since,

$$M_k(x) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{x}{s}\right) \tag{2.3}$$

from chapter 1, we have

$$\hat{f}_k(x) = \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right) \tag{2.4}$$

where h_n is a bandwidth sequence tending to zero as $n \to \infty$. When $k = 1$ and K is symmetric, the density estimator in (2.4) coincides with the Rosenblatt-Parzen density estimator.

$$\Delta_h^{2k} f(x) = \sum_{s=-k}^{k} c_{k,s} f(x + sh) \quad \text{for } h \in \mathbb{R}.$$

It is easy to verify that $\tilde{\Delta}_h^{2k} f(x) = \Delta_h^{2k}[f(x + kh)]$ ([19]). Hence, we use centered even-order difference for a smoothness characteristic, and we have

$$||f||_{b^r_{p,q}} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} |\Delta_h^{2k} f(x)|^p \, dx\right)^{1/p}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{1/q}.$$

2.2.2 Distribution Function Estimation

Let F be the distribution function with density f. We now list assumption that will be used throughout this chapter.

Assumption 1 : $\{X_t\}_{t=1}^n$ is an IID sequence.

Assumption 2 : (1) $F \in \mathcal{B}^{r+1}_{\infty,q}$ and $f \in \mathcal{B}^{r}_{\infty,q}$with $r > 0$ and $1 \le q \le \infty$; (2) $\sup_{x\in\mathbb{R}} |f(x)| < \infty$.

Assumption 3 : $h_n > 0$ for all n , $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

Assumption 4 : For all $x \in \mathbb{R}$,

(1) $K(x) : \mathbb{R} \to \mathbb{R}$ is a measurable function; (2) $\int K(x)dx = 1$; (3)$\int |K(x)|dx < \infty$; (4) $\int xK(x)dx = 0$; (5) $\int |x||K(x)|dx < \infty$; (6) $\sup_{x\in\mathbb{R}} |K(x)| < C < \infty$ where $C \in \mathbb{R}$; (7) K has a compact support.

Let's define a family of estimators for the distribution F by integrating $\hat{f}_k$. Thus, let

$$\hat{F}_k(x) = \int_{-\infty}^{x} \hat{f}_k(v)dv = \int_{-\infty}^{x} \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{X_t - v}{h_n}\right) dv$$

When $k = 1$ and K is symmetric, the distribution function estimator coincides with integrating the Rosenblatt-Parzen density estimator. The following theorem gives the bias for the distribution function estimators $\hat{F}_k$ and its order.

Theorem 2.1. *Let* Assumption $1-4$ *hold. For any* $h_n > 0$, $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots$,

(a) $Bias(\hat{F}_k(x)) = -\frac{1}{c_{k.0}} \int K(\psi)\Delta^{2k}_{-h_n\psi} F(x)d\psi$

(b) $|Bias(\hat{F}_k(x))| \leq O(h_n^{r+1})$ *where* $r + 1 < 2k$.

(c) $\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |Bias(\hat{F}_k(x))| = 0$.

(d) $\hat{F}_k(x) - F(x) \xrightarrow{p} 0$ *as* $n \to \infty$.

Theorem 2.1 (c) implies that $\hat{F}_k$ is asymptotically unbiased and (d) gives consistency of $\hat{F}_k$. According to the traditional asymptotic theory regarding the properties of a nonparametric estimator for a distribution, F is assumed to be continuous and differentiable in $\mathbb{R}$ in order to obtain bias of $\hat{F}$ and its order. Then, the order of the bias for $\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(v)dv$, where $\hat{f}$ is the Rosenblatt-Parzen density estimator, is $O(h_n^2)$ if a second-order kernel is used. Hence, we obtain bias reduction for the estimators of a distribution in the class by imposing less conditions on F.

[11] constructs a class of polarization measures based on the alienation-identification framework. When an individual is located, for example, at income x, the individual's sense of identification depends on the density at x, i.e., $f(x)$. Alienation is measured by means of a distance $|x - y|$. Given the alienation and identification, the DER index is given by

$$P_\alpha(F) = \int_{\mathbb{R}} f(y)^\alpha a(y)dF(y)$$

where f and F are density and distribution functions for income and α is a parameter related to the importance of the identification factor and is defined by the user.[3] The values of α can be described by the degree of polarization sensitivity and the greater is its value, the greater is the

[3] [11] defines $\alpha \in [0.25, 1]$.

difference from an inequality measurement. If $\alpha = 0$, the polarization measure resembles the Gini coefficient. The identification effect is denoted by $f(y)^\alpha$ and $a(y)$ represents the alienation effect, with $a(y) = \int |x - y| dF(x)$ for all x and $y \in \mathbb{R}$.

An estimator of $P_\alpha(F)$ is given by substituting $F(y)$ by $\hat{F}_k$ and $a(y)$ by $\hat{a}_k(y)$ where $\hat{a}_k(y) = \int |x - y| d\hat{F}_k(y)$. Hence, a family of estimators for the DER index can be denoted as

$$P_\alpha(\hat{F}_k) = \int \hat{f}_k(y)^\alpha \hat{a}_k(y) d\hat{F}_k(y)$$

First, we consider a family of estimators for the alienation $a(y)$.

$$\hat{a}_k(y) = \int |x - y| d\hat{F}_k(x) = \int x\hat{f}_k(x)dx - y + 2\hat{F}_k(y) - 2y \int_{-\infty}^{y} x\hat{f}_k(x)dx$$

We now add the extra assumption.

Assumption 5 : $E[|X_t|^{2p}] < \infty$ where $p > 1$.

Theorem 2.2. *Let* Assumption *1-5 hold. For every $y \in \mathbb{R}$ and $k = 1, 2, \cdots$,*

(a) $\lim_{n\to\infty} |Bias(\hat{a}_k(y))| = 0$ *and* $Var(\hat{a}_k(y)) \to 0$ *as* $n \to \infty$.

(b) $\hat{a}_k(x) - a(x) = o_p(1)$.

Consequently, $\hat{a}_k(y)$ is a class of consistent estimators for $a(y)$ for $y \in \mathbb{R}$. Our estimator for alienation is provided by integrating $\hat{f}_k$ which belongs to a Besov space and is continuous. In contrast, the existing estimator for alienation provided by [11], was constructed based on the empirical distribution function. In some cases it may be more interesting to consider a smooth estimator of alienation. For example, income distribution is generally to be a continuous function. Thus, the use of our estimator is more appropriate than the estimator proposed by [11].

The following theorem states that a class of estimators for the DER index is asymptotically unbiased.

Theorem 2.3. *Let* Assumption 1-5 *hold. For $k = 1, 2, \cdots$ we have*

$$Bias[P_\alpha(\hat{F}_k)] \longrightarrow 0 \quad as\ n \to \infty$$

2.3 Monte Carlo Study

In this section, we perform a small Monte Carlo study to investigate the finite sample performance of our proposed estimators for a polarization measure. For comparison purpose, we implement the existing estimator provided by [11], which is given by $\tilde{P}_\alpha(\tilde{F}) = \int \hat{f}(y)^\alpha \tilde{a}(y) d\tilde{F}(y) = n^{-1}\sum_{i=1}^n \hat{f}(y_i)^\alpha \tilde{a}(y_i)$ for $y_1 \le y_2 \le \cdots \le y_n$. Note that y_i is the empirical quantile for percentiles between $(i-1)/n$ and i/n. $\tilde{a}(y_i)$ is defined as $\tilde{a}(y_i) = \tilde{\mu} + y_i(n^{-1}(2i-1)-1) - n^{-1}\left(2\sum_{j=1}^{i-1} y_j + y_i\right)$, where $\tilde{\mu}$ is the sample mean. $\hat{f}(y_i)^\alpha$ is the Rosenblatt-Parzen density estimator. We consider simulated data from two different densities. They are: 1) Bimodal ($f_1(x) = \frac{3}{4}N(-1, 1.8) + \frac{1}{4}N(4, 0.4)$) and 2)Trimodal ($f_2(x) = \frac{1}{5}N(0, 0.5) + \frac{3}{5}N(5, 1) + \frac{2}{5}N(7, 2.5)$). For each these densities 1000 samples of size and $n = 200, 800$ were generated. We use bandwidth $h = 4.7n^{-1/2}\sigma\alpha^{0.1}$ where σ is variance provided by [11].

Results are reported in Table A.2. We observe the following general regularities. First, as predicted by our asymptotic results, for all densities considered bias of our estimators, $P_\alpha(\hat{F}_k)$ for $k = 1, 2, 3, 4$ fall as the sample size increases from 200 to 800. Second, the increase in the values of k reduces bias and root MSE, but this is not verified for all experiments. When we observe the case of f_2 for $k = 1$ in Table A.2, the existing estimators outperform $P_\alpha(\hat{F}_1)$. However, $P_\alpha(\hat{F}_2), P_\alpha(\hat{F}_3)$ and $P_\alpha(\hat{F}_4)$ show better performance than the existing estimators for a polarization measure in terms of bias and root MSE. For $k = 4$ the case where the smallest bias reductions are attained, bias can be reduced by as much as 80 percent relative to the estimator $P_\alpha(\tilde{F})$. Additionally, the magnitude of bias reduction produced by our estimator increases with the sample size. Third, density functions with large curvature (in increasing order of curvature f_1 and f_2) are more difficult to estimate both in terms of bias and root mean MSE for our estimators.

2.4 Summary

[11] introduced polarization index and its estimators. Their estimator was derived by using a Rosenblatt-Parzen density estimator and the empirical distribution function. Since the empirical

distribution function jumps by $1/n$ at each of the n data points, the empirical distribution suffers from lack of smoothness. For a new estimator of polarization measure, I used a class of new kernels $M_k(\cdot)$ and density estimators $\hat{f}_k$ introduced by [18] and integrate the elements of the class to produce density estimators $\hat{F}_k$. In this paper, I obtained the bias and variance for $\hat{F}_k$ and established its consistency. In addition, I show that my estimators for the DER index are asymptotically unbiased. Finally, a Monte Carlo simulation results show that our estimators can outperform the existing estimator in terms of bias and root mean squared error.

Bibliography

[1] Gordon Anderson. Polarization measurement and inference in many dimensions when subgroups can not be identified. Economics: The Open-Access, Open-Assessment E-Journal, 5:11, 2011.

[2] Gordon Anderson, Oliver B Linton, and Yoon-Jae Whang. Nonparametric estimation of a polarization measure. LSE STICERD Research Paper No. EM534, 2009.

[3] George Bennett. Probability inequalities for the sum of independent random variables. Journal of the American Statistical Association, 57:33–45, 1962.

[4] Oleg V Besov, P Valentin, and Sergey M Nikol'skii. Integral Representations of Functions and Imbedding: Theorems.: Vol.: 1. John Wiley & Sons, 1978.

[5] Raymond J Carroll, Aurore Delaigle, and Peter Hall. Nonparametric prediction in measurement error models. Journal of the American Statistical Association, 104:993–1014, 2009.

[6] Edwin Choi, Peter Hall, and Valentin Rousson. Data sharpening methods for bias reduction in nonparametric regression. Annals of Statistics, 28:1339–1355, 2000.

[7] Gérard Collomb. Estimation non-paramétrique de la régression: Revue bibliographique. International Statistical Review/Revue Internationale de Statistique, 49:75–93, 1981.

[8] James Davidson. Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians. Oxford University Press, 1994.

[9] Ronald A DeVore and Bradley J Lucier. Wavelets. Acta Numerica, 1:1–56, 1992.

[10] Luc P Devroye. The uniform convergence of the nadaraya-watson regression function estimate. Canadian Journal of Statistics, 6:179–191, 1978.

[11] Jean-Yves Duclos, Joan Esteban, and Debraj Ray. Polarization: concepts, measurement, estimation. Econometrica, 72:1737–1772, 2004.

[12] Joan-Maria Esteban and Debraj Ray. On the measurement of polarization. Econometrica: Journal of the Econometric Society, 62:819–851, 1994.

[13] Jianqing Fan. Design-adaptive nonparametric regression. Journal of the American Statistical Association, 87:998–1004, 1992.

[14] Theo Gasser, Hans-Georg Muller, and Volker Mammitzsch. Kernels for nonparametric curve estimation. Journal of the Royal Statistical Society. Series B (Methodological), 47:238–252, 1985.

[15] Ingrid K Glad. Parametrically guided non-parametric regression. Scandinavian Journal of Statistics, 25:649–668, 1998.

[16] Yue-Pok Mack and Bernard W Silverman. Weak and strong uniform consistency of kernel regression estimates. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 61:405–415, 1982.

[17] Carlos Martins-Filho, Santosh Mishra, and Aman Ullah. A class of improved parametrically guided nonparametric regression estimators. Econometric Reviews, 27:542–573, 2008.

[18] Kairat Mynbaev and Carlos Martins-Filho. Bias reduction in kernel density estimation via lipschitz condition. Journal of Nonparametric Statistics, 22:219–235, 2010.

[19] Kairat Mynbaev and Carlos Martins-Filho. Consistency and asymptotic normality for a nonparametric prediction under measurement errors. Journal of Multivariate Analysis, 139:166–188, 2015.

[20] Elizbar A Nadaraya. On estimating regression. Theory of Probability & Its Applications, 9:141–142, 1964.

[21] Divided We Stand OECD. Why inequality keeps rising, 2011.

[22] Jeffrey Racine. Bias-corrected kernel regression. Journal of Quantitative Economics, 17:25–42, 2001.

[23] William R Schucany. On nonparametric regression with higher-order kernels. Journal of Statistical Planning and Inference, 23:145–151, 1989.

[24] Hans Triebel. Theory of Function Spaces. Birkhäuser, 1985.

[25] Hans Triebel. Theory of Function Spaces II. Birkhäuser, 2010.

[26] Geoffrey S Watson. Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A, 24:359–372, 1964.

[27] Michael C Wolfson. When inequalities diverge. The American Economic Review, 84:353–358, 1994.

[28] Xiaobo Zhang and Ravi Kanbur. What difference do polarisation measures make? an application to china. Journal of Development Studies, 37:85–98, 2001.

Appendix A

Appendix

A.1 Appendix for Chapter 1: Proofs, tables and figures

Proof of Lemma 1.1

Proof. Let $\tilde{\Delta}_h^0 f(x) = f(x)$, $(\tilde{\Delta}_h^{s+1} f)(x) = \tilde{\Delta}_h^1(\tilde{\Delta}_h^s f)(x)$ where $x \in \mathbb{R}$, $h \in \mathbb{R}_+$, $s \in \mathbb{N}$ be the iterated differences in $\mathbb{R}$. From the definition of $\tilde{\Delta}_h^s f(x)$ (2.1), we have

$$\begin{aligned}\tilde{\Delta}_h^s f(x) &= \sum_{j=0}^{s}(-1)^{s-j}C_s^j f(x+jh)\\ &= \sum_{j=0}^{s-1}(-1)^{s-1-j}C_{s-1}^j[f(x+(j+1)h)-f(x+jh)].\end{aligned}$$

Let $\mathcal{D}^l f(u_1, \cdots, u_l) = \frac{\partial^l f}{\partial u_1 \cdots \partial u_l}$. First, we need to verify the case : (i) $s = 1$.

$$\tilde{\Delta}_h^1 f(x) = f(x+h) - f(x) = \int_0^h \mathcal{D}f(x+u_1)du_1$$

Consider (ii) $s = 2$.

$$\begin{aligned}\tilde{\Delta}_h^2 f(x) = \tilde{\Delta}_h^1[\tilde{\Delta}_h^1 f(x)] = \tilde{\Delta}_h^1 \int_0^h \mathcal{D}f(x+u_1)du_1 &= \int_0^h [\mathcal{D}f(x+u_1+h) - \mathcal{D}f(x+u_1)]du_1\\ &= \int_0^h \int_0^h \mathcal{D}^2 f(x+u_1+u_2)du_1 du_2\end{aligned}$$

Assume that $s = k$ is true such that

$$\tilde{\Delta}_h^k f(x) = \int_0^h \cdots \int_0^h \tilde{\Delta}_h^{k-l}\mathcal{D}^l f\left(x + \sum_{i=1}^{l} u_i\right)\prod_{i=1}^{l} du_i \quad \text{where } l = 1, 2, \cdots, k.$$

Now, we must prove the case (iii) $s = k+1$.

$$\begin{aligned}
\tilde{\Delta}_h^{k+1} f(x) &= \tilde{\Delta}_h^1[\tilde{\Delta}_h^k f(x)] = \tilde{\Delta}_h^1\left(\int_0^h \tilde{\Delta}_h^{k-1}\mathcal{D}f(x+u_1)du_1\right)\\
&= \int_0^h \left(\tilde{\Delta}_h^{k-1}\mathcal{D}f(x+u_1+h) - \tilde{\Delta}_h^{k-1}\mathcal{D}f(x+u_1)\right)du_1\\
&= \int_0^h\int_0^h \tilde{\Delta}_h^{k-1}\mathcal{D}^2 f(x+u_1+u_2)du_1du_2 = \int_0^h\int_0^h \tilde{\Delta}_h^1[\tilde{\Delta}_h^{k-2}\mathcal{D}^2 f(x+u_1+u_2)]du_1du_2\\
&= \cdots = \int_0^h\cdots\int_0^h \mathcal{D}^{k+1} f\left(x+\sum_{i=1}^{k+1} u_i\right)\prod_{i=1}^{k+1} du_i
\end{aligned}$$

□

Proof of Lemma 1.2

Proof. Let s and $l \in \mathbb{Z}_+$ such that $l < r < s$.

$$||f||_{b^r_{\infty,q}} = \int_0^\infty h^{-rq}||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h} + \int_{-\infty}^0 (-h)^{-rq}||\tilde{\Delta}_{-h}^s f(x)||_\infty^q \frac{dh}{(-h)} = 2\int_0^\infty h^{-rq}||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h}$$

Now,

$$\begin{aligned}
&\int_0^\infty h^{-rq}||\tilde{\Delta}_h^s f||_\infty^q \frac{dh}{h}\\
&= \int_0^1 h^{-rq}||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h} + \int_1^\infty h^{-rq}||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h}\\
&+ \int_0^1 \gamma^{-rq}||\tilde{\Delta}_{-\gamma}^s f(x)||_\infty^q \frac{d\gamma}{\gamma} + \int_1^\infty \gamma^{-rq}||\tilde{\Delta}_{-\gamma}^s f(x)||_\infty^q \frac{d\gamma}{\gamma}\\
&= \int_0^1 h^{-rq}\left[||\tilde{\Delta}_h^s f(x)||_\infty^q + ||\tilde{\Delta}_{-h}^s f(x)||_\infty^q\right]\frac{dh}{h} + \int_1^\infty h^{-rq}\left[||\tilde{\Delta}_h^s f(x)||_\infty^q + ||\tilde{\Delta}_{-h}^s f(x)||_\infty^q\right]\frac{dh}{h}\\
&= \int_0^1 h^{-rq}\left[\left|\left|\int_0^h\cdots\int_0^h \mathcal{D}^s f(x+\sum_{i=1}^s u_i)\prod_{i=1}^s du_i\right|\right|_\infty^q + \left|\left|\int_0^{-h}\cdots\int_0^{-h} \mathcal{D}^s f(x+\sum_{i=1}^s u_i)\prod_{i=1}^s du_i\right|\right|_\infty^q\right]\frac{dh}{h}\\
&+ \int_1^\infty h^{-rq}\left[\left|\left|\int_0^h\cdots\int_0^h \tilde{\Delta}_h^{s-l}\mathcal{D}^l f(x+\sum_{i=1}^l u_i)\prod_{i=1}^l du_i\right|\right|_\infty^q\right]\frac{dh}{h}\\
&+ \int_1^\infty h^{-rq}\left[\left|\left|\int_0^{-h}\cdots\int_0^{-h} \tilde{\Delta}_{-h}^{s-l}\mathcal{D}^l f(x+\sum_{i=1}^l u_i)\prod_{i=1}^l du_i\right|\right|_\infty^q\right]\frac{dh}{h}
\end{aligned}$$

$$
\begin{aligned}
&\le \int_0^1 h^{-rq}\left\{\sup_{x\in\mathbb{R}}\left[\int_0^h\cdots\int_0^h\left|\mathcal{D}^s f(x+\sum_{i=1}^s u_i)\right|\prod_{i=1}^s du_i\right]^q\right\}\frac{dh}{h}\\
&+\int_0^1 h^{-rq}\left\{\sup_{x\in\mathbb{R}}\left[\int_0^{-h}\cdots\int_0^{-h}\left|\mathcal{D}^s f(x+\sum_{i=1}^s u_i)\right|\prod_{i=1}^s du_i\right]^q\right\}\frac{dh}{h}\\
&+\int_1^\infty h^{-rq}\left\{\sup_{x\in\mathbb{R}}\left[\int_0^h\cdots\int_0^h\left|\tilde{\Delta}_h^{s-l}\mathcal{D}^l f(x+\sum_{i=1}^l u_i)\right|\prod_{i=1}^l du_i\right]^q\right\}\frac{dh}{h}\\
&+\int_1^\infty h^{-rq}\left\{\sup_{x\in\mathbb{R}}\left[\int_0^{-h}\cdots\int_0^{-h}\left|\tilde{\Delta}_{-h}^{s-l}\mathcal{D}^l f(x+\sum_{i=1}^l u_i)\right|\prod_{i=1}^l du_i\right]^q\right\}\frac{dh}{h}\\
&\le \sup_{x\in\mathbb{R}}|\mathcal{D}^s f(x)|^q\int_0^1 h^{-rq}\left\{\left[\int_0^h\cdots\int_0^h\prod_{i=1}^s du_i\right]^q+\left[\int_0^{-h}\cdots\int_0^{-h}\prod_{i=1}^s du_i\right]^q\right\}\frac{dh}{h}\\
&+c_1\sup_{x\in\mathbb{R}}|\mathcal{D}^l f(x)|^q\int_1^\infty h^{-rq}\left[\int_0^h\cdots\int_0^h\prod_{i=1}^l du_i\right]^q\frac{dh}{h}\\
&+c_2\sup_{x\in\mathbb{R}}|\mathcal{D}^l f(x)|^q\int_1^\infty h^{-rq}\left[\int_0^{-h}\cdots\int_0^{-h}\prod_{i=1}^l du_i\right]^q\frac{dh}{h}\\
&=\sup_{x\in\mathbb{R}}|\mathcal{D}^s f(x)|^q\int_0^1\left[h^{-rq+sq-1}+h^{-rq-1}(-h)^{sq}\right]dh\\
&+\sup_{x\in\mathbb{R}}|\mathcal{D}^l f(x)|^q\int_1^\infty\left[c_1h^{-rq+lq-1}+c_2h^{-rq-1}(-h)^{lq}\right]dh\\
&=\sup_{x\in\mathbb{R}}|\mathcal{D}^s f(x)|^q\left[\frac{1}{(s-r)q}\right](1+(-1)^{sq})+\sup_{x\in\mathbb{R}}|\mathcal{D}^l f(x)|^q\left[\frac{1}{(r-l)q}\right](c_1+c_2(-1)^{lq})
\end{aligned}
$$

where $s>r>l$ and some $c_1,c_2<\infty$. Therefore, for some c_3,c_4 and $c<\infty$, we have

$$\left[\int|h|^{-rq}||\tilde{\Delta}_h^s f(x)||_\infty^q\frac{dh}{|h|}\right]^{1/q}\le c_3\sup_{x\in\mathbb{R}}|\mathcal{D}^s f(x)|+c_4\sup_{x\in\mathbb{R}}|\mathcal{D}^l f(x)|\le c\sup_{x\in\mathbb{R}}|\mathcal{D}^l f(x)|$$

The last inequality follows from the fact that $\mathcal{C}^s(\mathbb{R})\subseteq\mathcal{C}^l(\mathbb{R})$ for $s>l$. Hence, we have $||f||_{\mathcal{B}^r_{\infty,q}}\le C||f||_{\mathcal{C}^l}+||f||_p$. That is, $\mathcal{C}^l(\mathbb{R})\subseteq\mathcal{B}^r_{\infty,q}(\mathbb{R})$ where $l<r$. □

Proof of Theorem 1.1

Proof. (a)

$$E(\hat{f}_k(x)) \;=\; \int\frac{1}{h_n}\Big[-\frac{1}{c_{k,0}}\sum_{|s|=1}^k\frac{c_{k,s}}{|s|}K\left(\frac{y-x}{sh_n}\right)\Big]f(y)dy=\int\Big[-\frac{1}{c_{k,0}}\sum_{|s|=1}^k c_{k,s}K(\psi)\Big]f(x+sh_n\psi)d\psi$$

Therefore, $Bias(\hat{f}_k(x))$ can be denoted as follows,

$$Bias(\hat{f}_k(x))=E(\hat{f}_k(x))-f(x)=-\frac{1}{c_{k,0}}\int K(\psi)\Delta^{2k}_{h_n\psi}f(x)d\psi$$

by $-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}c_{k,s}=1$ and Assumption 3(2).

(*b*) We proceed to investigate the order of $Bias(\hat{f}_k(x))$ using (*a*).
Given that $\left[\int|K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'}<\infty$, we have

$$
\begin{aligned}
&|Bias(\hat{f}_k(x))| = |E(\hat{f}_k(x)) - f(x)| = \left|\int -\frac{1}{c_{k,0}}K(\psi)\Delta^{2k}_{h_n\psi}f(x)d\psi\right| \\
&= \left|\frac{1}{c_{k,0}}\int K(\psi)|h_n\psi|^{r+1/q}\frac{1}{|h_n\psi|^{r+1/q}}\Delta^{2k}_{h\psi}f(x)d\psi\right| \\
&\leq \left|\frac{1}{c_{k,0}}\right|\left[\int\left\{|K(\psi)||h_n\psi|^{r+1/q}\right\}^{q'}d\psi\right]^{1/q'}\left[\int\left\{\frac{sup_{x\in\mathbb{R}}|\Delta^{2k}_{h_n\psi}f(x)|}{|h_n\psi|^{r+1/q}}\right\}^{q}d\psi\right]^{1/q} \text{ by Hölder's Inequality} \\
&= \left|\frac{1}{c_{k,0}}\right|\left[\int\left\{|K(\psi)||h_n\psi|^{r+1/q}\right\}^{q'}d\psi\right]^{1/q'}\left[\int\left\{\frac{sup_{x\in\mathbb{R}}|\Delta^{2k}_{h_n\psi}f(x)|}{|h_n\psi|^{r}}\right\}^{q}\frac{1}{|h_n\psi|}d\psi\right]^{1/q} \\
&= \left|\frac{1}{c_{k,0}}\right|\left[\int|K(\psi)|^{q'}|h_n\psi|^{(r+1/q)q'}d\psi\right]^{1/q'}\left[\int\left\{\frac{sup_{x\in\mathbb{R}}|\Delta^{2k}_{t}f(x)|}{|t|^{r}}\right\}^{q}\frac{1}{|t|}h_n^{-1}dt\right]^{1/q} \\
&\leq h_n^r\left|\frac{1}{c_{k,0}}\right|\left[\int|K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'}||f||_{b^r_{\infty,q}} = O(h_n^r)
\end{aligned}
$$

where $1/q+1/q'=1$ and $1\leq q\leq\infty$. □

Proof of Theorem 1.2

Proof.

$$
\begin{aligned}
Var(\hat{f}_k(x)) &= E[\hat{f}_k(x)^2] - (E[\hat{f}_k(x)])^2 \\
&= \int\left\{\frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{y-x}{h_n}\right)\right\}^2 f(y)dy - \left\{\int\frac{1}{nh}\sum_{t=1}^{n}M_k\left(\frac{y-x}{h_n}\right)f(y)dy\right\}^2 \\
&= \int\frac{1}{nh_n}M_k^2(\psi)f(x+h_n\psi)d\psi - \frac{1}{n}\left\{\int M_k(\psi)f(x+h_n\psi)d\psi\right\}^2 \\
&\leq \frac{1}{nh_n}\int M_k^2(\psi)f(x+h_n\psi)d\psi
\end{aligned}
$$

Now provided that Assumption 2(2), Assumption 3 and Assumption 4(1),(3),(4) , we have

$$\int M_k^2(\psi) f(x+h_n\psi)d\psi = \int M_k^2(\psi)\left[f(x+h_n\psi)-f(x)\right]d\psi + f(x)\int M_k^2(\psi)d\psi$$
$$\leq \int_{|\psi|\leq\delta/h_n} M_k^2(\psi)|f(x+h_n\psi)-f(x)|d\psi + \int_{|\psi|>\delta/h_n} M_k^2(\psi)|f(x+h_n\psi)-f(x)|d\psi + f(x)\int M_k^2(\psi)d\psi$$
$$\leq \sup_{|y|\leq\delta, x\in\mathbb{R}} |f(x+y)-f(x)| \int M_k^2(\psi)d\psi + 2\sup_{x\in\mathbb{R}}|f(x)| \int_{|\psi|>\delta/h_n} M_k^2(\psi)d\psi + \sup_{x\in\mathbb{R}}|f(x)| \int M_k^2(\psi)d\psi$$
$$\text{since } f\in\mathcal{C}^0(\mathbb{R}) < \infty \tag{A.1}$$

The inequality follows from $\int M_k^2(\psi)d\psi \leq C\int |K(\psi)|d\psi < \infty$ by Assumption 4(3)-(4) for some $C < \infty$ and $\sup_{x\in\mathbb{R}}|f(x)| < \infty$ by Assumption 2(2). If $h_n \to 0$ and $nh_n \to \infty$ as $n\to\infty$ (Assumption 3), from Theorem 1 and equation (A.1), $\hat{f}_k(x) - f(x) = o_p(1)$ for all $x\in\mathbb{R}$. □

Proof of Theorem 1.3

Proof. Let $\{X_t\}_{t=1,2,\cdots,n}$ be a sequence of IID random variables in $\mathbb{R}$ (Assumption 1). For $x\in\mathbb{R}$, $\hat{f}_k(x) = \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t-x}{h_n}\right)$ where $h_n > 0$. Let $\mathcal{G}$ be a compact subset of $\mathbb{R}$. The collection $F = \{B(x,r) : x\in\mathcal{G}, r>0\}$ where $B(x,r) = \{x_0\in\mathbb{R} : |x-x_0| < r\}$ is an open covering of $\mathcal{G}$. By the Heine-Borel Theorem, the open covering has a finite subcovering. That is, there exists a collection $F' = \{B(x_\tau, r) : x_\tau\in\mathcal{G}, r>0, \tau = 1,2,\cdots,m, \text{ where } m \text{ is finite}\}$ such that $\mathcal{G}\subseteq F'$. Given that K satisfies a Lipschitz condition of order 1 Assumption 4(5), for $x\in\mathcal{G}$, we have

$$|\hat{f}_k(x)-\hat{f}_k(x_\tau)| = \left|\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t-x}{h_n}\right) - \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t-x_\tau}{h_n}\right)\right|$$
$$\leq \frac{1}{nh_n}\sum_{t=1}^n \left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^k \frac{|c_{k,s}|}{|s|}\left|K\left(\frac{X_t-x}{sh_n}\right) - K\left(\frac{X_t-x_\tau}{sh_n}\right)\right|$$
$$\leq \frac{c}{nh_n}\sum_{t=1}^n\sum_{|s|=1}^k \left|\frac{c_{k,s}}{|s|}\right| \frac{|x_\tau - x|}{|sh_n|} \leq \frac{c}{h_n^2}|x-x_\tau| \quad \text{for some contant } c<\infty.$$

Also, $|E[\hat{f}_k(x)] - \hat{f}_k(x_\tau)| \leq c\frac{1}{h_n^2}|x-x_\tau|$. If $x\in B(x_\tau, r)$, then $|x-x_\tau| < r$. Then, by the Triangle

Inequality

$$\begin{aligned}|\hat{f}_k(x) - E[\hat{f}_k(x)]| &\leq |\hat{f}_k(x) - \hat{f}_k(x_\tau)| + |\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| + |E[\hat{f}_k(x_\tau)] - E[\hat{f}_k(x)]| \\ &\leq |\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| + 2c\frac{r}{h_n^2}.\end{aligned}$$

Since for each $x \in \mathcal{G}$, there exists $B(x_\tau, r)$ that contains x

$$d_n = \sup_{x\in\mathcal{G}} |\hat{f}_k(x) - E[\hat{f}_k(x)]| \leq 2c\frac{r}{h_n^2} + \max_{1\leq\tau\leq m} |\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]|.$$

$$\frac{d_n}{a_n} = \frac{|d_n|}{a_n} \leq \frac{2cr}{a_n h_n^2} + \frac{1}{a_n}\max_{1\leq\tau\leq m} |\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| = \frac{2cr}{a_n h_n^2} + \frac{1}{a_n} d_{2,n}$$

where $d_{2,n} \equiv \max_{1\leq\tau\leq m} |\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]|$.

$$P\left[\frac{d_n}{a_n} > M_\epsilon\right] \leq P\left[\frac{2cr}{a_n h_n^2} + \frac{d_{2,n}}{a_n} > M_\epsilon\right] = P\left[\frac{d_{2,n}}{a_n} > M_\epsilon - \frac{2cr}{a_n h_n^2}\right] = P\left[\frac{d_{2,n}}{a_n} > M_{n,\epsilon}\right]$$

where $M_{n,\epsilon} = M_\epsilon - \frac{2cr}{a_n h_n^2}$.

Then, we have

$$\begin{aligned}P\left[\frac{d_{2,n}}{a_n} > M_{n,\epsilon}\right] &= P\left[\frac{1}{a_n}\max_{1\leq\tau\leq m} |\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > M_{n,\epsilon}\right] \\ &\leq \sum_{\tau=1}^{m} P\left[\frac{1}{a_n}|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > M_{n,\epsilon}\right] \\ &= \sum_{\tau=1}^{m} P\left[|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > M_{n,\epsilon} a_n\right].\end{aligned}$$

Hence, $P\left[\frac{d_n}{a_n} > M_\epsilon\right] \leq \sum_{\tau=1}^{m} P\left[|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > M_{n,\epsilon} a_n\right]$.

$$\begin{aligned}|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| &= \left|\frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{h_n}M_k\left(\frac{X_t - x_\tau}{h_n}\right) - \frac{1}{h_n}E\left[M_k\left(\frac{X_t - x_\tau}{h_n}\right)\right]\right]\right| \\ &= \left|\frac{1}{n}\sum_{t=1}^{n} W_{tn}\right| = \left|\frac{1}{n}\sum_{t=1}^{n}(X_{tn} - E[X_{tn}])\right|\end{aligned}$$

where $W_{tn} \equiv \left[\frac{1}{h_n}M_k\left(\frac{X_t - x_\tau}{h_n}\right) - \frac{1}{h_n}E\left[M_k\left(\frac{X_t - x_\tau}{h_n}\right)\right]\right] = X_{tn} - E[X_{tn}]$ and $X_{tn} \equiv \frac{1}{h_n}M_k\left(\frac{X_t - x_\tau}{h_n}\right)$.

Given that $|K(x)| \leq C$ for all $x \in \mathbb{R}$ (Assumption 4(4)), note that

$$|X_{tn} - E[X_{tn}]| = \left| \frac{1}{h_n} M_k\left(\frac{X_t - x_\tau}{h_n}\right) - \frac{1}{h_n}\int M_k\left(\frac{\alpha - x}{h_n}\right) f(\alpha)d\alpha \right|$$
$$= \left| \frac{1}{h_n}\left[-\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{X_t - x}{sh_n}\right)\right] - \frac{1}{h_n}\int \left[-\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{\alpha - x}{sh_n}\right)\right] f(\alpha)d\alpha \right|$$
$$\leq \frac{1}{h_n}\left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^{k} \left|\frac{c_{k,s}}{|s|}\right| C + \frac{1}{h_n}\left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^{k} \left|\frac{c_{k,s}}{|s|}\right| C \int |f(\alpha)|d\alpha \leq 2\frac{1}{h_n}\left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^{k} \left|\frac{c_{k,s}}{|s|}\right| C$$

since $\int |f(\alpha)|d\alpha \leq 1$.

Then, we have $|W_{tn}| \leq \frac{1}{h_n} C C_1$ where $C_1 = 2\left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^{k} \left|\frac{c_{k,s}}{|s|}\right|$. Next we consider $Var(W_{tn})$. Let $\sigma_{tn}^2 = Var(W_{tn})$. Given that $E[W_{tn}] = 0$, we have

$$\sigma_{tn}^2 = Var(W_{tn}) = E[W_{tn}^2] = \frac{1}{h_n^2}\int M_k^2\left(\frac{\alpha - x_\tau}{h_n}\right) f(\alpha)d\alpha - \frac{1}{h_n^2}\left[\int M_k\left(\frac{\alpha - x_\tau}{h_n}\right) f(\alpha)d\alpha\right]^2$$
$$= \frac{1}{h_n}\int M_k^2(\psi) f(x_\tau + h_n\psi)d\psi - \left[\int M_k(\psi) f(x_\tau + h_n\psi)d\psi\right]^2.$$

Given that $\{X_t\}_{t=1,2,\cdots,n}$ is an IID sequence,

$$\sum_{t=1}^{n} \sigma_{tn}^2 = \frac{n}{h_n}\int M_k^2(\psi) f(x_\tau + h_n\psi)d\psi - n\left[\int M_k(\psi) f(x_\tau + h_n\psi)d\psi\right]^2.$$

Note that $h_n\sigma_{tn}^2 = \int M_k^2(\psi) f(x_\tau + h_n\psi)d\psi - h_n\left[\int M_k(\psi) f(x_\tau + h_n\psi)d\psi\right]^2 = g_n(x_\tau)$. By Bernstein's Inequality ([3]), we have

$$P\left[|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > a_n M_{n,\epsilon}\right] = P\left[\left|\frac{1}{n}\sum_{t=1}^{n} W_{tn}\right| > a_n M_{n,\epsilon}\right]$$
$$\leq 2\exp\left\{-\frac{n a_n^2 M_{n,\epsilon}^2}{2\frac{1}{n}\sum_{t=1}^{n} Var(W_{tn}) + \frac{2}{3}\frac{C}{h_n} C_1 a_n M_{n,\epsilon}}\right\} \quad \text{where } C_1 = 2\left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^{k} \left|\frac{c_{k,s}}{|s|}\right|$$
$$= 2\exp\left\{-\frac{n^2 a_n^2 M_{n,\epsilon}^2 / nE(W_{tn}^2)}{2 + \frac{2}{3}\frac{C}{h_n} C_1 \frac{n a_n M_{n,\epsilon}}{nE(W_{tn}^2)}}\right\} \quad \text{by Assumption 1 and } Var(W_{tn}) = E(W_{tn}^2)$$

$$= 2\exp\left\{-\frac{na_n^2 M_{n,\epsilon}^2/E(W_{tn}^2)}{2\frac{E[W_{tn}^2]}{E[W_{tn}^2]} + \frac{2}{3}\frac{C}{h_n}C_1\frac{a_n M_{n,\epsilon}}{E[W_{tn}^2]}}\right\}$$

$$= 2\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 nh_n}{2h_n E[W_{tn}^2] + \frac{2}{3}CC_1 a_n M_{n,\epsilon}}\right\}$$

$$P\left[\frac{1}{a_n}\max_{1\le\tau\le m}|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > M_{n,\epsilon}\right] \le \sum_{\tau=1}^{m} 2\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 nh_n}{2h_n E[W_{tn}^2] + \frac{2}{3}CC_1 a_n M_{n,\epsilon}}\right\}$$

$$\le 2m\max_{1\le\tau\le m}\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 nh_n}{2h_n E[W_{tn}^2] + \frac{2}{3}CC_1 a_n M_{n,\epsilon}}\right\}$$

$$\le 2m\max_{1\le\tau\le m}\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 nh_n}{2g_n(x_\tau) + \frac{2}{3}CC_1 a_n M_{n,\epsilon}}\right\} \quad \text{where } g_n(x_\tau) = h_n E[W_{tn}^2]$$

$$\le 2m\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 nh_n}{2g_n(x^m) + \frac{2}{3}CC_1 a_n M_{n,\epsilon}}\right\}$$

where x^m corresponds to the point of the given function such that $\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 nh_n}{2h_n E[W_{tn}^2]+\frac{2}{3}CC_1 a_n M_{n,\epsilon}}\right\}$ attains its maximum value and $g_n(x^m) = \int M_k^2(\psi)f(x^m + h_n\psi)d\psi - h_n\left[\int M_k(\psi)f(x^m + h_n\psi)d\psi\right]^2$. Let $a_n = \left(\frac{\log n}{nh_n}\right)^{1/2}$ and $r = \left(\frac{h_n^3}{n}\right)^{1/2}$. Note that $a_n M_{n,\epsilon} = a_n\left(M_\epsilon - \frac{2cr}{a_n h_n^2}\right) = a_n M_\epsilon - \frac{2cr}{h_n^2}$.

Hence,

$$\begin{aligned}(a_n M_{n,\epsilon})^2 &= \left(a_n M_\epsilon - \frac{2cr}{h_n^2}\right)^2 = a_n^2 M_\epsilon^2 + \frac{4c^2 r^2}{h_n^4} - 4a_n M_\epsilon \frac{cr}{h_n^2} \\ &= \left(\frac{\log n}{nh_n}\right)M_\epsilon^2 + \frac{4c^2}{h_n^4}\left(\frac{h_n^3}{n}\right) - 4M_\epsilon c\frac{1}{h_n^2}\left(\frac{h_n^3}{n}\right)^{1/2}\left(\frac{\log n}{nh_n}\right)^{1/2} \\ &= \left(\frac{\log n}{nh_n}\right)M_\epsilon^2 + \frac{4c^2}{nh_n} - 4M_\epsilon c\frac{(\log n)^{1/2}}{nh_n}\end{aligned}$$

$$\begin{aligned}-nh_n(a_n M_{n,\epsilon})^2 = -(\log n)M_\epsilon^2 - 4c^2 + 4M_\epsilon c(\log n)^{1/2} &= -\log n\left[M_\epsilon^2 - \frac{4M_\epsilon c}{(\log n)^{1/2}} + \frac{4c^2}{\log n}\right] \\ &= -\triangle_n \log n\end{aligned}$$

where $\triangle_n = M_\epsilon^2 - \frac{4M_\epsilon c}{(\log n)^{1/2}} + \frac{4c^2}{\log n}$.

$$\begin{aligned}a_n M_{n,\epsilon} &= a_n M_\epsilon - \frac{2cr}{h_n^2} = \left(\frac{\log n}{nh_n}\right)^{1/2} M_\epsilon - \frac{2c}{h_n^2}\left(\frac{h_n^3}{n}\right)^{1/2} = \left(\frac{\log n}{nh_n}\right)^{1/2} M_\epsilon - \frac{2c}{(nh_n)^{1/2}} \\ &= \frac{1}{(nh_n)^{1/2}}\left[(\log n)^{1/2} M_\epsilon - 2c\right]\end{aligned}$$

Hence, if $\left(\frac{\log n}{nh_n}\right) \to 0$ as $n \to \infty$, then $a_n M_{n,\epsilon} \to 0$. From the above inequality, we have

$$2m \exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 nh_n}{2g_n(x^m) + \frac{2}{3}CC_1 a_n M_{n,\epsilon}}\right\} = 2m \exp\left\{-\frac{\triangle_n \log n}{v_n}\right\}$$
$$= 2m \exp\left\{\log n^{-\triangle_n/v_n}\right\} = 2mn^{-\triangle_n/v_n}$$

where $v_n = 2g_n(x^m) + \frac{2}{3}CC_1 a_n M_{n,\epsilon}$. Hence, $P\left[\frac{1}{a_n}\max_{1\le\tau\le m}|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > M_{n,\epsilon}\right] \le 2mn^{-\triangle_n/v_n}$. The $B(x_\tau, r)$ for $x \in \mathbb{R}$ is $2r = 2\left(\frac{h_n^3}{n}\right)^{1/2} = 2r_n$. Since $F' = \{B(x_\tau, r) : x_\tau \in \mathcal{G}, r > 0, \tau = 1, \cdots, m$ where m is finite $\}$ such that $\mathcal{G} \subseteq F'$ is a covering for $\mathcal{G}$, it must be that $r \to 0$ which implies $m \to \infty$ and since $\mathcal{G}$ is bounded, there exists $x_0 \in \mathbb{R}$ and $r_0 < \infty$ such that $\mathcal{G} \subseteq B(x_0, r_0)$.

Hence for every $x \in \mathbb{R}$, $2mr_n = 2m\left(\frac{h_n^3}{n}\right)^{1/2} \le 2r_0$ which implies that $m \le r_0 r_n^{-1} = r_0\left(\frac{h_n^3}{n}\right)^{-1/2}$.

Then,

$$\begin{aligned}
2mn^{-\triangle_n/v_n} &\le 2r_0\left(\frac{h_n^3}{n}\right)^{-1/2}\frac{1}{n^{\triangle_n/v_n}} = 2r_0\left(\frac{n}{h_n^3 n^{2\triangle_n/v_n}}\right)^{1/2} = 2r_0\left(\frac{1}{h_n^3 n^{2\triangle_n/v_n - 1}}\right)^{1/2} \\
&= 2r_0\left(\frac{1}{nh_n^3}\right)^{1/2}\left(\frac{1}{n^{2(\triangle_n/v_n - 1)}}\right)^{1/2} = 2r_0\left(\frac{1}{nh_n}\right)^{1/2}\frac{1}{h_n}\frac{1}{n^{\triangle_n/v_n - 1}}
\end{aligned}$$

Since $nh_n \to \infty$ it suffices to have $n^{\triangle_n/v_n - 1}h_n$ bounded away from 0 as $n \to \infty$.

Given that $\triangle_n = M_\epsilon^2 - \frac{4M_\epsilon c}{(\log n)^{1/2}} + \frac{4c^2}{\log n}$ and $v_n = 2g_n(x^m) + \frac{2}{3}CC_1 a_n M_{n,\epsilon}$, $\triangle_n \to M_\epsilon^2$, $g_n(x^m) \to f(x^m)\int M_k^2(\psi)d\psi$ as $n \to \infty$ and $v_n \to 2f(x^m)\int M_k^2(\psi)d\psi$. Then, $\frac{\triangle_n}{v_n} - 1 = \frac{M_\epsilon^2}{2f(x^m)\int M_k^2(\psi)d\psi} - 1$. Since $nh_n \to \infty$ it suffices to choose M_ϵ large enough to have $\frac{M_\epsilon^2}{2f(x^m)\int M_k^2(\psi)d\psi} - 1 \ge 1$ or $\frac{M_\epsilon^2}{2f(x^m)\int M_k^2(\psi)d\psi} \ge 2$ to obtain $n^{\frac{\triangle_n}{v_n} - 1}h_n \to \infty$.

Now,

$$\begin{aligned}
\sup_{x\in\mathcal{G}}|\hat{f}_k(x) - f(x)| &\le \sup_{x\in\mathcal{G}}|\hat{f}_k(x) - E[\hat{f}_k(x)]| + \sup_{x\in\mathcal{G}}|E[\hat{f}_k(x)] - f(x)| \\
&= \left(\frac{\log n}{nh_n}\right)^{1/2} O_p(1) + \sup_{x\in\mathcal{G}}|E[\hat{f}_k(x)] - f(x)| = \left(\frac{\log n}{nh_n}\right)^{1/2} O_p(1) + h_n^r O(1)
\end{aligned}$$

□

Proof of Theorem 1.4

Proof. We have for $x \in \mathbb{R}$, $\hat{f}_k(x) - E[\hat{f}_k(x)] = \sum_{t=1}^n \left[\frac{1}{nh_n} M_k\left(\frac{X_t - x}{h_n}\right) - \frac{1}{nh_n} E\left[M_k\left(\frac{X_t-x}{h_n}\right)\right]\right]$. Let $Z_{nt} = \frac{1}{nh_n} M_k\left(\frac{X_t-x}{h_n}\right)$, then $E[Z_{nt}] = \mu_n$ and $S_n^2 = \sum_{t=1}^n E[Z_{nt} - \mu_n]^2$. We have

$$\begin{aligned} S_n^2 &= \sum_{t=1}^n E\left[\frac{1}{nh} M_k\left(\frac{X_t - x}{h_n}\right) - E\left(\frac{1}{nh} M_k\left(\frac{X_t - x}{h_n}\right)\right)\right]^2 = \frac{1}{nh_n^2} Var\left(M_k\left(\frac{X_t - x}{h_n}\right)\right) \\ &= Var(\hat{f}_k(x)). \end{aligned}$$

Hence, $\frac{\hat{f}_k(x) - E[\hat{f}_k(x)]}{\sqrt{Var(\hat{f}_k(x))}} = \sum_{t=1}^n \left(\frac{Z_{nt} - \mu_n}{S_n}\right) = \sum_{t=1}^n X_{nt}$ with $E[X_{nt}] = 0$, $E[X_{nt}^2] = \frac{1}{S_n^2} E\left[(Z_{nt} - \mu_n)^2\right]$ and $\sum_{t=1}^n E[X_{nt}^2] = 1$.

In order to use Liapounov's CLT , we need to verify $\lim_{n\to\infty} \sum_{t=1}^n E|X_{nt}|^{2+\delta} = 0$.

$$\begin{aligned} \sum_{t=1}^n E|X_{nt}|^{2+\delta} &= \sum_{t=1}^n E\left[\left|\frac{Z_{nt} - \mu_n}{S_n}\right|^{2+\delta}\right] = \sum_{t=1}^n Var(\hat{f}_k(x))^{-1-\delta/2} E\left[|Z_{nt} - \mu_n|^{2+\delta}\right] \\ &= Var(\hat{f}_k(x))^{-1-\delta/2} n E\left[|Z_{nt} - \mu_n|^{2+\delta}\right] \end{aligned}$$

We need to show that $|\mu_n|^{2+\delta} \leq C$.

$$\mu_n = E\left[\frac{1}{nh_n} M_k\left(\frac{X_t - x}{h_n}\right)\right] = \frac{1}{n} \int M_k(\psi) f(x + h_n\psi) d\psi = O(n^{-1})$$

Therefore, $|\mu|^{2+\delta} \leq M^{2+\delta} \left(\frac{1}{n}\right)^{2+\delta}$. By the C_r inequality and the fact that $\mu_n = O(n^{-1})$, we have

$$E\left[|Z_{nt} - \mu_n|^{2+\delta}\right] \leq 2^{1+\delta} E\left[|Z_{nt}|^{2+\delta}\right] + 2^{1+\delta} E[|\mu_n|^{2+\delta}].$$

Now, we show that $\sum_{t=1}^n E|X_{nt}|^{2+\delta} \to 0$ as $n \to \infty$.

$$\begin{aligned} \sum_{t=1}^n E|X_{nt}|^{2+\delta} &\leq Var(\hat{f}_k(x))^{-(1+\delta/2)} n \left[2^{1+\delta} E(|Z_{nt}|^{2+\delta}) + o(1)\right] \\ &= \frac{1}{nh_n Var(\hat{f}_k(x))^{1+\delta/2}} \frac{2^{1+\delta}}{(nh_n)^{\delta/2}} \frac{1}{h_n} \left\{E\left[\left|M_k\left(\frac{X_t - x}{h_n}\right)\right|^{2+\delta}\right] + o(1)\right\}. \end{aligned}$$

The following result provides that $nh_n Var(\hat{f}_k(x)) = O(1)$.

Note that

$$\begin{aligned}
&\left|nh_nVar(\hat{f}_k(x)) - \int M_k^2(\psi)f(x)d\psi\right| \\
&\leq \int_{|h_n\psi|\leq\delta} M_k^2(\psi)\left|f(x+h_n\psi)-f(x)\right|d\psi + \int_{|h_n\psi|>\delta} M_k^2(\psi)\left|f(x+h_n\psi)-f(x)\right|d\psi \\
&+ h_n\left\{\int M_k^2(\psi)f(x+h_n\psi)d\psi\right\}^2 \\
&\leq \sup_{|t|\leq\delta, x\in\mathbb{R}}|f(x+t)-f(x)|\int M_k^2(\psi)d\psi + 2\sup_{x\in\mathbb{R}}[f(x)]\int_{|\psi|>\delta/h_n} M_k^2(\psi)d\psi \\
&+ h_n\left\{\sup_{x\in\mathbb{R}}[f(x)]\int M_k^2(\psi)d\psi\right\}^2.
\end{aligned}$$

Consequently, $nh_nVar(\hat{f}_k(x)) \longrightarrow \int M_k^2(\psi)f(x)d\psi$ as $n\to\infty$ by Assumption 1, Assumption 2(2) Assumption 3 and Assumption 4(3)-(4).

Similarly,

$$\frac{1}{h_n}E\left[\left|M_k\left(\frac{X_t-x}{h_n}\right)\right|^{2+\delta}\right] = \int |M_k(\psi)|^{2+\delta}f(x+h_n\psi)hd\psi \longrightarrow f(x)\int |M_k(\psi)|^{2+\delta}d\psi < \infty$$

as $n\to\infty$ since $\int|K(\psi)|^{2+\delta}d\psi<\infty$. Therefore, $\sum_{t=1}^n E|X_{nt}|^{2+\delta}\longrightarrow 0$ as $n\to\infty$ provided that $\frac{1}{(nh_n)^{\delta/2}}\to 0$. Hence, $\sqrt{nh}\left(\hat{f}_k(x)-E[\hat{f}_k(x)]\right)\xrightarrow{d}\mathcal{N}\left(0, f(x)\int M_k^2(\psi)d\psi\right)$.

Now,

$$\begin{aligned}
\sqrt{nh_n}\left(\hat{f}_k(x)-f(x)\right) &= \sqrt{nh_n}\left(\hat{f}_k(x)-E[\hat{f}_k(x)]\right)+\sqrt{nh_n}\left(E[\hat{f}_k(x)]-f(x)\right) \\
&= \sqrt{nh_n}O(h_n^r)+\sqrt{nh_n}\left(\hat{f}_k(x)-E[\hat{f}_k(x)]\right) \quad \text{from Theorem 1.1}(a)
\end{aligned}$$

Therefore, $\sqrt{nh_n}\left(\hat{f}_k(x)-f(x)+O(h_n^r)\right)\xrightarrow{d}\mathcal{N}\left(0,f(x)\int M_k^2(\psi)d\psi\right)$. If $nh_n^{1+2r}\to 0$ as $n\to\infty$, $\hat{f}_k(x)$ has an asymptotic normal distribution as $\sqrt{nh_n}\left(\hat{f}_k(x)-f(x)\right)\xrightarrow{d}\mathcal{N}\left(0,f(x)\int M_k^2(\psi)d\psi\right)$.

□

Proof of Theorem 1.5

Proof. First, consider the proof of (a). Let $\hat{g}_k(x)=\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t-x}{h_n}\right)Y_t$.

$$E[\hat{g}_k(x)|X_t] = \frac{1}{nh}\sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right)E[Y_t|X_t]$$
$$= \frac{1}{nh}\sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right)m(X_t) \quad \text{where } E[Y_t|X_t] = m(X_t)$$

Then, given Assumption 1 we have

$$E[\hat{g}_k(x)] = \left(-\frac{1}{c_{k,0}}\right)\int \sum_{|s|=1}^{k} c_{k,s}K(\psi)m(x + sh_n\psi)f(x + sh_n\psi)d\psi.$$

Let $g(x) = f(x)m(x)$. $Bias(\hat{g}_k(x))$ is denoted by

$$Bias(\hat{g}_k(x)) = E[\hat{g}_k(x)] - g(x)$$
$$= \left(-\frac{1}{c_{k,0}}\right)\int \sum_{|s|=1}^{k} c_{k,s}\ K(\psi)m(x + sh_n\psi)f(x + sh_n\psi)d\psi - g(x)$$
$$= \left(-\frac{1}{c_{k,0}}\right)\int K(\psi)\Delta^{2k}_{h_n\psi}g(x)d\psi$$

by Assumption 4(2) and $-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k} c_{k,s} = 1$.

Next, we prove (b); the order of $Bias(\hat{g}_k(x))$.

$$|Bias(\hat{g}_k(x))|$$
$$\leq C\left|\int K(\psi)\Delta^{2k}_{h_n\psi}\ g(x)d\psi\right| = C\left|\int K(\psi)|h_n\psi|^{r+1/q}\ \frac{\Delta^{2k}_{h_n\psi}\ g(x)}{|h_n\psi|^{r+1/q}}\ d\psi\right|$$
$$\leq C\left\{\int\left[|K(\psi)||h_n\psi|^{(r+1/q)}\right]^{q'} d\psi\right\}^{1/q'}\left\{\int\left[\frac{\sup_{x\in\mathbb{R}}\ |\Delta^{2k}_{h_n\psi}\ g(x)|}{|h_n\psi|^{r+1/q}}\right]^{q} d\psi\right\}^{1/q}$$
$$= C\left[\int |K(\psi)|^{q'}|h_n\psi|^{(r+1/q)q'}d\psi\right]^{1/q'}\left\{\int\left[\frac{\sup|\Delta^{2k}_t g(x)|}{|t|^r}\right]^{q}\frac{1}{|t|}h_n^{-1}dt\right\}^{1/q} \quad \text{by Hölder's Inequality}$$
$$= h_n^r\left[\int |K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'} ||g||_{b^r_{\infty,q}} = O(h_n^r)$$

where the order follows from Assumption 2(1), Assumption 5(1) and the fact that $\left[\int |K(\psi)|^{q'}|\psi|^{(r+1/q)q'}d\psi\right]^{1/q'} < \infty$ where $1/q + 1/q' = 1$ for $1 \leq q \leq \infty$. From the result of ([24]), we know that $||g||_{\mathcal{B}^r_{\infty,q}} \leq C||m||_{\mathcal{Z}^\rho}||f||_{\mathcal{B}^r_{\infty,q}}$ where $\rho > r$.

For (c) it is sufficient to show that $Var[\hat{g}_k(x)] \longrightarrow 0$ as $n \to \infty$. Since, $Var[\hat{g}_k(x)] = E[Var_X(\hat{g}_k(x))] + Var[E_X(\hat{g}_k(x))]$ and $Var[Y_t|X_t] = \sigma^2$, we have

$$\begin{aligned}
Var[\hat{g}_k(x)] &= \sigma^2 E\left[\frac{1}{n^2h_n^2}\sum_{t=1}^{n} M_k^2\left(\frac{X_t-x}{h_n}\right)\right] + Var\left[\frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)m(X_t)\right] \\
&= \frac{\sigma^2}{n^2h_n^2}E\left[\sum_{t=1}^{n} M_k^2\left(\frac{X_t-x}{h_n}\right)\right] + \frac{1}{n^2h_n^2}E\left\{\left[\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)m(X_t)\right]^2\right\} \\
&- \left\{E\left[\frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)m(X_t)\right]\right\}^2 \\
&= \frac{\sigma^2}{nh_n^2}\int M_k^2\left(\frac{y-x}{h_n}\right)f(y)dy + \frac{1}{nh_n}\int M_k^2(\psi)m^2(x+h_n\psi)f(x+h_n\psi)d\psi \\
&+ \frac{n(n-1)}{n^2h_n^2}\left[\int h_n^2 M_k^2(\psi)m(x+h_n\psi)f(x+h_n\psi)d\psi\right]^2 - \left\{\int M_k(\psi)m(x+h_n\psi)f(x+h_n\psi)d\psi\right\}^2 \\
&= \frac{\sigma^2}{nh_n}\int M_k^2(\psi)f(x+h_n\psi)d\psi + \frac{1}{nh_n}\int M_k^2(\psi)m^2(x+h_n\psi)f(x+h_n\psi)d\psi \\
&- \frac{1}{n}\left\{\int M_k(\psi)m(x+h_n\psi)f(x+h_n\psi)d\psi\right\}^2.
\end{aligned}$$

Since f and $m \in \mathcal{C}^0(\mathbb{R})$ and ASSUMPTION 4(3)-(4), we have that $Var[\hat{g}_k(x)] \to 0$ as $n \to \infty$. Hence $\hat{g}_k(x) \xrightarrow{p} g(x)$. □

Proof of Theorem 1.6

Proof. Let $\{X_t\}_{t=1,2,\cdot}$ be a sequence of IID random variable in $\mathbb{R}$.

$$\begin{aligned}
\hat{g}_k(x) &= \frac{1}{nh}\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)[m(X_t)+u_t] \\
&= \frac{1}{nh}\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)m(X_t) + \frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)u_t
\end{aligned}$$

Let $s_1(x) = \frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)m(X_t)$ and $s_2(x) = \frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t-x}{h_n}\right)u_t$.

Let $\mathcal{G}$ be a compact set in $\mathbb{R}$. For every $x \in \mathcal{G}$, define $B(x,r) = \{y : |x-y| < r\}$. The collection $F' = \{B(x,r) : x \in \mathcal{G}, r > 0\}$ is an open covering of $\mathcal{G}$. By the Heine-Borel Theorem, there exists a collection $F' = \{B(x_\tau, r) : x_\tau \in \mathcal{G}, r > 0, \tau = 1, 2, \cdots, m, m \text{ finite}\}$ such that $\mathcal{G} \subseteq F'$. For $x \in \mathcal{G}$ and $x_\tau \in \mathcal{G}$ where $\tau = 1, 2, \cdots, m$,

$$|s_1(x) - E[s_1(x)]| \leq |s_1(x) - s_1(x_\tau)| + |s_1(x_\tau) - E[s_1(x_\tau)]| + |E[s_1(x_\tau)] - E[s_1(x)]|. \quad \text{(A.2)}$$

Given that $c = \left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^{k} \left|\frac{c_{k,s}}{|s|}\right|$, we have

$$\begin{aligned}
|s_1(x) - s_1(x_\tau)| &= \left| \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right) m(X_t) - \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{X_t - x_\tau}{h_n}\right) m(X_t) \right| \\
&\leq \frac{1}{nh_n} \sum_{t=1}^{n} \left| M_k\left(\frac{X_t - x}{h_n}\right) - M_k\left(\frac{X_t - x_\tau}{h_n}\right) \right| |m(X_t)| \\
&\leq \frac{1}{nh_n} \sum_{t=1}^{n} \left[-\frac{1}{c_{k,0}} \sum_{t=1}^{n} \frac{c_{k,s}}{|s|} \right] \left| K\left(\frac{X_t - x}{sh_n}\right) - K\left(\frac{X_t - x_\tau}{sh_n}\right) \right| |m(X_t)|
\end{aligned}$$

by Lipschitz condition on K (Assumption 4(5)) and $m \in \mathcal{C}^0(\mathbb{R})$(Assumption 5(2)).

$$\leq c \sup_{x \in \mathbb{R}} |m(x)| \frac{|x_\tau - x|}{h_n^2} \leq c \sup_{x \in \mathbb{R}} |m(x)| \frac{r}{h_n^2} \quad \text{since } x \in B(x_\tau, r) \text{ which implies } |x - x_\tau| < r$$

and $|E[s_1(x)] - E[s_1(x_\tau)]| \leq c \sup_{x \in \mathbb{R}} |m(x)| \frac{r}{h_n^2}$.

Thus, from (A.2) we have

$$|s_1(x) - E[s_1(x)]| \leq \frac{2cr}{h_n^2} + |s_1(x_\tau) - E[s_1(x_\tau)]|.$$

Since for each $x \in \mathcal{G}$, there exists $B(x_\tau, r)$ that contains x,

$$d_n = \sup_{x \in \mathbb{R}} |s_1(x) - E[s_1(x)]| \leq \frac{2cr}{h_n^2} + \max_{1 \leq \tau \leq m} |s_1(x_\tau) - E[s_1(x_\tau)]|$$

Let $d_{2,n} = \max_{1 \leq \tau \leq m} |s_1(x_\tau) - E[s_1(x_\tau)]|$.

$$\begin{aligned}
P\left[\frac{d_n}{a_n} > M_\epsilon\right] &\leq P\left[\frac{2cr}{a_n h_n^2} + \frac{d_{2,n}}{a_n} > M_\epsilon\right] = P\left[\frac{d_{2,n}}{a_n} > M_\epsilon - \frac{2rc}{a_n h_n^2}\right] \\
&= P\left[\frac{1}{a_n} \max_{1 \leq \tau \leq m} |s_1(x_\tau) - E[s_1(x_\tau)]| > M_{n,\epsilon}\right] \\
&\leq \sum_{\tau=1}^{m} P\left[\frac{1}{a_n} |s_1(x_\tau) - E[s_1(x_\tau)]| > M_{n,\epsilon}\right] = \sum_{\tau=1}^{m} P\left[|s_1(x_\tau) - E[s_1(x_\tau)]| > a_n M_{n,\epsilon}\right]
\end{aligned}$$

$$|s_1(x_\tau) - E[s_1(x_\tau)]| = \left| \frac{1}{n} \sum_{t=1}^{n} W_{tn} \right| \quad \text{(A.3)}$$

where $W_{tn} = \frac{1}{h_n} M_k\left(\frac{X_t - x_\tau}{h_n}\right) m(X_t) - \frac{1}{h_n} E\Big[M_k\left(\frac{X_t - x_\tau}{h_n}\right) m(X_t)\Big]$.

$$
\begin{aligned}
&|W_{tn}| \\
&= \left|\frac{1}{h_n}\Big(-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\Big) K\left(\frac{X_t - x_\tau}{sh_n}\right) m(X_t) - \frac{1}{h_n}\Big(-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\Big) E\Big[K\left(\frac{X_t - x_\tau}{sh_n}\right) m(X_t)\Big]\right| \\
&\le \frac{1}{h_n} Cc \sup_{x\in\mathbb{R}} |m(x)| \Big[1 + \int |f(\alpha)| d\alpha\Big] \le 2Cc\frac{1}{h_n}\sup_{x\in\mathbb{R}}|m(x)| \quad \text{where } c = \left|\frac{1}{c_{k,0}}\right| \textstyle\sum_{|s|=1}^{k}\left|\frac{c_{k,s}}{|s|}\right|.
\end{aligned}
$$

since $\int |f(\alpha)| \le 1$, $m \in \mathcal{C}^0(\mathbb{R})$ and $\sup_{x\in\mathbb{R}} |K(x)| \le C$ for all $x \in \mathbb{R}$.

$$
\begin{aligned}
&Var(W_{tn}) = E(W_{tn}^2) \\
&= \frac{1}{h_n^2}\int M_k^2\left(\frac{\alpha - x_\tau}{h_n}\right) m(\alpha) f(\alpha) d\alpha - \frac{1}{h_n^2}\Big[\int M_k\left(\frac{\alpha - x_\tau}{h_n}\right) m(\alpha) f(\alpha) d\alpha\Big]^2 \\
&= \frac{1}{h_n}\int M_k^2(\psi) m^2(x_\tau + h_n\psi) f(x_\tau + h_n\psi) d\psi - \Big[\int M_k(\psi) m(x_\tau + h_n\psi) f(x_\tau + h_n\psi) d\psi\Big]^2
\end{aligned}
$$

$$
\begin{aligned}
&h_n Var(W_{tn}) \\
&= \int M_k^2(\psi) m^2(x_\tau + h_n\psi) f(x_\tau + h_n\psi) d\psi - h_n\Big[\int M_k(\psi) m(x_\tau + h_n\psi) f(x_\tau + h_n\psi) d\psi\Big]^2
\end{aligned}
\tag{A.4}
$$

From (A.3), we have

$$
\begin{aligned}
P\left[|s_1(x_\tau) - E[s_1(x_\tau)]| > a_n M_{n,\epsilon}\right] &= P\left[\left|\frac{1}{n}\sum_{t=1}^{n} W_{tn}\right| > a_n M_{n,\epsilon}\right] = P\left[\left|\sum_{t=1}^{n} W_{tn}\right| > n a_n M_{n,\epsilon}\right] \\
&\le 2\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 n h_n}{2h_n Var(W_{tn}) + \frac{2}{3}Cc\sup_{x\in\mathbb{R}}|m(x)| a_n M_{n,\epsilon}}\right\} \\
&\qquad \text{by Bernstein's inequality.}
\end{aligned}
$$

Let $g_n(x_\tau) = h_n Var(W_{tn})$. Then,

$$
\begin{aligned}
&P\left[\frac{1}{a_n}\max_{1\le\tau\le m} |s_1(x_\tau) - E[s_1(x_\tau)]| > M_{n,\epsilon}\right] \\
&\le \sum_{\tau=1}^{m} 2\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 n h_n}{2h_n Var(W_{tn}) + \frac{2}{3}Cc\sup_{x\in\mathbb{R}}|m(x)| a_n M_{n,\epsilon}}\right\} \\
&\le 2m \max_{1\le\tau\le m} \exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 n h_n}{2g_n(x_\tau) + \frac{2}{3}Cc\sup_{x\in\mathbb{R}}|m(x)| a_n M_{n,\epsilon}}\right\} \\
&= 2m \exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 n h_n}{2g_n(x^m) + \frac{2}{3}Cc\sup_{x\in\mathbb{R}}|m(x)| a_n M_{n,\epsilon}}\right\}
\end{aligned}
\tag{A.5}
$$

where x^m corresponds to the point of the given function such that

$$\exp\left\{-\frac{a_n^2 M_{n,\epsilon}^2 n h_n}{2h_n E[W_{tn}^2] + \frac{2}{3}Cc\sup_{x\in\mathbb{R}}|m(x)|a_n M_{n,\epsilon}}\right\}$$

attains its maximum value. Thus we have

$$g_n(x^m) = \int M_k^2(\psi)m^2(x^m + h_n\psi)f(x^m + h_n\psi)d\psi - h_n\left[\int M_k(\psi)m(x^m + h_n\psi)f(x^m + h_n\psi)d\psi\right]^2 \tag{A.6}$$

Let $a_n = \left(\frac{\log n}{nh_n}\right)^{1/2}$ and $r = \left(\frac{h_n^3}{n}\right)^{1/2}$. We have

$$\begin{aligned} a_n M_{n,\epsilon} &= a_n M_\epsilon - \frac{2cr}{h_n^2} = \frac{1}{(nh_n)^2}\left[(\log n)^{1/2} M_\epsilon c - 2c\right] \\ (a_n M_{n,\epsilon})^2 &= \left(\frac{\log n}{nh_n}\right) M_\epsilon^2 + \frac{4c^2}{nh_n} - 4M_\epsilon c\frac{(\log n)^{1/2}}{nh_n} \end{aligned}$$

to obtain $a_n M_{n,\epsilon} \to 0$ we want $\frac{\log n}{nh_n} \to 0$ as $n \to \infty$. Hence,

$$\begin{aligned} -nh_n(a_n M_{n,\epsilon})^2 = -M_\epsilon^2 \log n - 4c^2 + 4M_\epsilon c(\log n)^{1/2} &= -\log n\left[M_\epsilon^2 + \frac{4c^2}{\log n} - \frac{4M_\epsilon c}{(\log n)^2}\right] \\ &= -\triangle_n(\log n) \end{aligned}$$

where $\triangle_n = M_\epsilon^2 + \frac{4c^2}{\log n} - \frac{4M_\epsilon c}{(\log n)^2}$. Let $v_n = 2g_n(x^m) + \frac{2}{3}Cc\sup_{x\in\mathbb{R}}|m(x)|a_n M_{n,\epsilon}$.

From (A.5)

$$P\left[\frac{1}{a_n}\max_{1\le\tau\le m}|s_1(x_\tau) - E[s_1(x_\tau)]| > M_{n,\epsilon}\right] \le 2m\exp\left\{-\frac{\triangle_n \log n}{v_n}\right\} = 2mn^{-\triangle_n/v_n} \tag{A.7}$$

From (A.4) and (A.6), $g_n(x^m) \to m^2(x^m)f(x^m)\int M_k^2(\psi)d\psi$ as $n \to \infty$ since $f \in \mathcal{C}^0(\mathbb{R})$, $m \in \mathcal{C}^0(\mathbb{R})$, $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. The volume of $B(x_\tau, r)$ for $x_\tau \in \mathbb{R}$ is $2r$. Since F' is a covering for $\mathcal{G}$, it must be that $m \to \infty$ since $r \to 0$. Since $\mathcal{G}$ is bounded, there exists $x_0 \in \mathbb{R}$ and $r_0 < \infty$ such that $\mathcal{G} \subseteq B(x_0, r_0)$. Hence, $2mr \le 2r_0$ which implies $m \le r_0\left(\frac{n}{h_n^3}\right)^{1/2}$. From (A.7),

$$2mn^{-\triangle_n/v_n} \le 2\left(\frac{n}{h_n^3}\right)^{1/2} r_0\frac{1}{n^{\triangle_n/v_n}} \le 2\left(\frac{1}{nh_n}\right)^{1/2}\frac{r_0}{h_n}\left[\frac{1}{n^{\triangle_n/v_n - 1}}\right]$$

Since $nh_n \to \infty$ as $n \to \infty$ it suffices to have $n^{\triangle_n/v_n - 1} h_n$ bounded away from 0 as $n \to \infty$. Note that $\triangle_n \to M_\epsilon^2$ and $v_n \to 2m^2(x^m)f(x^m)\int M_k^2(\psi)d\psi$. Since $nh_n \to \infty$ it suffices to choose M_ϵ large enough to have

$$\frac{\triangle_n}{v_n} - 1 \longrightarrow \frac{M_\epsilon^2}{2m^2(x^m)f(x^m)\int M_k^2(\psi)d\psi} \geq 2$$

to obtain $n^{\triangle_n/v_n - 1} h_n \to \infty$. Hence, we have $\sup_{x\in\mathbb{R}} |s_1(x) - E[s_1(x)]| = O_p\Big(\left(\frac{\log n}{nh_n}\right)^{1/2}\Big)$.

Now consider $s_2(x) = \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)u_t$. For $x, x_\tau \in \mathcal{G}$ $\tau = 1, 2, \cdots, m$, by the Triangle Inequality

$$|s_2(x)| \leq |s_2(x) - s_2(x_\tau)| + |s_2(x_\tau)|$$

Note that

$$|s_2(x) - s_2(x^\tau)| \leq \frac{1}{nh_n}\sum_{t=1}^n \left| M_k\left(\frac{X_t - x}{h_n}\right) - M_k\left(\frac{X_t - x_\tau}{h_n}\right)\right| |u_t| \quad \leq c\left(\frac{r}{h_n^2}\right) O_p(1) \text{ by } x \in B(x_\tau, r)$$

where $c = \left|\frac{1}{c_{k,0}}\right| \sum_{|s|=1}^k \left|\frac{c_{k,s}}{|s|}\right|$. We have that $\{|u_t|\}_{t=1,2,\cdots}$ is IID. By condition $E[|u_t|^a] < \infty$ for some $a \geq 2$ and $\frac{1}{n}\sum_{t=1}^n (|u_t| - E[|u_t|]) = o_p(1)$ by Kolmogorov's LLN we have $|s_2(x) - s_2(x_\tau)| \leq c\left(\frac{r}{h_n^2}\right) O_p(1)$.

Then,

$$|E(s_1(x)) - E(s_1(x^\tau))| \leq c\left(\frac{r}{h_n^2}\right) O_p(1) \text{ and } |E(s_2(x)) - E(s_2(x^\tau))| \leq c\left(\frac{r}{h_n^2}\right) O_p(1)$$

By the Triangle Inequality,

$$\begin{aligned} |s_2(x) - E[s_2(x)]| &\leq |s_2(x) - s_2(x_\tau)| + |s_2(x_\tau) - E[s_2(x_\tau)]| + |E(s_2(x_\tau)) - E[s_2(x)]| \\ &\leq |s_2(x_\tau) - E[s_2(x_\tau)]| + 2c\left(\frac{r}{h_n^2}\right) O_p(1) \end{aligned} \tag{A.8}$$

Let $\hat{s}_2(x) = \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)u_t\chi_{\{|u_t|\leq B_n\}}$ with $B_1 \leq B_2 \leq \cdots$ such that $\sum_{t=1}^\infty B_t^{-a} < \infty$ for some $a > 1$. Note that

$$|s_2(x_\tau) - E[s_2(x_\tau)]| \leq |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| + |s_2(x_\tau) - \hat{s}_2(x_\tau)| + |E[s_2(x_\tau)] - E[\hat{s}_2(x_\tau)]| \tag{A.9}$$

From (A.8) and (A.9) for each $x \in \mathcal{G}$, there exists $B(x_\tau, r)$ that contains x

$$\gamma_n = \sup_{x\in\mathcal{G}} |s_2(x) - E[s_2(x)]| \leq \frac{2cr}{h_n^2} O_p(1) + \sup_{x\in\mathcal{G}} |s_2(x_\tau) - E[s_2(x_\tau)]|$$
$$\leq \frac{2cr}{h_n^2} O_p(1) + \sup_{x\in\mathcal{G}} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| + \sup_{x\in\mathcal{G}} |s_2(x_\tau) - \hat{s}_2(x_\tau)| + \sup_{x\in\mathcal{G}} |E[s_2(x_\tau)] - E[\hat{s}_2(x_\tau)]|$$

Let $T_1 = \sup_{x\in\mathcal{G}} |s_2(x) - \hat{s}_2(x)|$ and $T_2 = \sup_{x\in\mathcal{G}} |E[s_2(x) - \hat{s}_2(x)]|$. We show that (1) $T_1 = o_{a.s}(1)$ and (2) $T_2 = O(B_n^{1-a})$ for $a > 1$. Note that $T_1 = \sup_{x\in\mathcal{G}} \left| \frac{1}{nh_n} \sum_{t=1}^n M_k \left(\frac{X_t - x}{h_n}\right) u_t \chi_{\{|u_t|>B_n\}} \right|$. Assume $E[|u_t|^a] < C$.

By Chebyshev's Inequality, for $a > 0$, $P[|u_t| > B_t] < \frac{E[|u_t|^a]}{B_t^a} < C\frac{1}{B_t{}^a}$ and $\sum_{t=1}^\infty P[|u_t| > B_t] < \sum_{t=1}^\infty \frac{C}{B_t^a} < \infty$.

By the Borel-Cantelli Lemma, $\sum_{t=1}^\infty P[|u_t| > B_t] < \infty$, which implies $P[|u_t| > B_t \ \ i.o.] = 0$ where $i.o.$ is an abbreviation for "infinitely often". Hence, for any $\epsilon > 0$ and for all m satisfying $m' < m$ we have $P[|u_m| \leq B_m] > 1 - u$ since $\{B_t\}_{t=1,2,\dots}$ is an increasing sequence, for $n > m > m'$ we have $P[|u_m| \leq B_n] > 1 - u$. Hence, there exists N such that for $n > \max\{m, N\}$ we have that for all $t \leq n$, $P[|u_t| \leq B_n] > 1 - \epsilon$ which implies $\chi_{\{|u_t|>B_n\}} = 0$ with probability 1. Therefore $T_1 = o_{a.s}(1)$.

For T_2,

$$\begin{aligned} E[s_2(x) - \hat{s}_2(x)] &= \frac{1}{nh_n} \sum_{t=1}^n \int \int_{|u_t|>B_n} M_k \left(\frac{\alpha - x}{h_n}\right) u_t \ f(\alpha) f_u(u_t) d\alpha du_t \\ &= \frac{1}{n} \sum_{t=1}^n \int M_k(\psi) f(x + h_n\psi) d\psi \int_{|u_t|>B_n} u_t f_u(u_t) du_t \\ &= \int M_k(\psi) f(x + h_n\psi) d\psi \int u \ f_u(u) \chi_{\{|u|>B_n\}} du \end{aligned}$$

By Hölder's Inequality,

$$\begin{aligned} \int |u_t| f(u_t) \chi_{\{|u_t|>B_n\}} du_t &\leq \left[\int |u_t|^a f(u_t) du_t\right]^{1/a} \left[\int \chi_{\{|u_t|>B_n\}} f_u(u_t) du_t\right]^{1-1/a} \\ &= \Big[E[|u_t|^a]\Big]^{1/a} \left[\int \chi_{\{|u_t|>B_n\}} f_u(u_t) du_t\right]^{1-1/a} \end{aligned}$$

where $\left[\int \chi_{\{|u_t|>B_n\}} f_u(u_t) du_t\right]^{1-1/a} = [P(|u_t| > B_n)]^{1-1/a} \leq C \left[\frac{E(|u_t|^a)}{B_n^a}\right]^{1-1/a} \leq CB_n^{1-a}$ by using Chebychev's Inequality.

Hence, $T_2 = O(B_n^{1-a})$ for $a > 1$. Given $T_1 = o_{a.s}(1)$ and $T_2 = O(B_n^{1-a})$ we have

$$\begin{aligned}\gamma_n = \sup_{x\in\mathcal{G}} |s_2(x) - E[s_2(x)]| &\leq \frac{2cr}{h_n^2} O_p(1) + T_1 + T_2 + \sup_{x_\tau\in\mathcal{G}} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| \\ &\leq \frac{2cr}{h_n^2} O_p(1) + O(B_n^{1-a}) + \max_{1\leq\tau\leq m} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]|\end{aligned}$$

Let γ_n be a sequence of stochastic variables and $\gamma_{2,n} = \max_{1\leq\tau\leq m} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]|$. For all $\epsilon > 0$, there exists $\tilde{M}_\epsilon > 0$ and a nonstochastic sequence $\{b_n\}$ such that $P\left[\frac{|\gamma_n|}{b_n} > \tilde{M}_\epsilon\right] < \epsilon$ for all n.

$$\begin{aligned}&P\left[\frac{\gamma_n}{b_n} > \tilde{M}_\epsilon\right] \leq P\left[\frac{2cr}{h_n^2} O_p(1) + O(B_n^{1-a}) + \frac{\gamma_{2,n}}{b_n} > \tilde{M}_\epsilon\right] \leq P\left[\frac{\gamma_{2,n}}{b_n} > \tilde{M}_\epsilon - \frac{2cr}{h_n^2} O_p(1) - O(B_n^{1-a})\right] \\ &\leq P\left[\frac{1}{b_n} \max_{1\leq\tau\leq m} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > \tilde{M}_{n,\epsilon}\right] \quad \text{where } \tilde{M}_{n,\epsilon} = \tilde{M}_\epsilon - \tfrac{2cr}{h_n^2} O_p(1) - O(B_n^{1-a}). \\ &\leq \sum_{\tau=1}^{m} P\left[\frac{1}{b_n} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > \tilde{M}_{n,\epsilon}\right] \leq \sum_{\tau=1}^{m} P\left[|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > b_n \tilde{M}_{n,\epsilon}\right] \qquad \text{(A.10)}\end{aligned}$$

Note that

$$\begin{aligned}&|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| \\ &= \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{1}{h_n} M_k\left(\frac{X_t - x_\tau}{h_n}\right) u_t \chi_{\{|u_t|\leq B_n\}} - \frac{1}{h_n} E\left[M_k\left(\frac{X_t - x_\tau}{h_n}\right) u_t \chi_{\{|u_t|\leq B_n\}}\right]\right\}\right| \\ &= \left| \frac{1}{n} \sum_{t=1}^{n} Z_{tn} \right|\end{aligned}$$

where $Z_{tn} = \left\{\frac{1}{h_n} M_k\left(\frac{X_t - x_\tau}{h_n}\right) u_t \chi_{\{|u_t|\leq B_n\}} - \frac{1}{h_n} E\left[M_k\left(\frac{X_t - x_\tau}{h_n}\right) u_t \chi_{\{|u_t|\leq B_n\}}\right]\right\}$. Note that

$$\begin{aligned}|Z_{tn}| &= \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{1}{h_n} M_k\left(\frac{X_t - x_\tau}{h_n}\right) u_t \chi_{\{|u_t|\leq B_n\}} - \frac{1}{h_n} E\left[M_k\left(\frac{X_t - x_\tau}{h_n}\right) u_t \chi_{\{|u_t|\leq B_n\}}\right]\right\}\right| \\ &= \left| \frac{1}{h_n}\left[- \frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{X_t - x_\tau}{s h_n}\right)\right] u_t \chi_{\{|u_t|\leq B_n\}} \right. \\ &\qquad \left. - \frac{1}{h_n} E\left[\left[- \frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{X_t - x_\tau}{s h_n}\right)\right] u_t \chi_{\{|u_t|\leq B_n\}}\right] \right| \\ &\leq \frac{1}{h_n} cC \left[\left|u_t \chi_{\{|u_t|\leq B_n\}}\right| + \int |f_X(\alpha)| d\alpha \int \left|u_t \chi_{\{|u_t|\leq B_n\}}\right| |f_{u|X}(u)| du \right] \\ &\leq \frac{1}{h_n} cC \left[B_n + B_n \int\int |f_{u|X}(u)| du \right] \leq 2cC B_n \frac{1}{h_n}\end{aligned}$$

since $|K(\cdot)| \leq C$ (Assumption 4(4)) and $|u_t \chi_{\{|u_t| \leq B_n\}}| \leq B_n$.

$$Var(Z_{tn}) = \frac{1}{h_n} \int\int M_k^2(\psi) u^2 \chi_{\{|u^2| \leq B_n\}} f(x_\tau + h_n \psi) f_{u|X}(u) d\psi du - \left[\int\int M_k(\psi) u \chi_{\{|u| \leq B_n\}} f(x_\tau + h_n\psi) f_{u|X}(u) d\psi du \right]^2$$

Let $l_n(x_\tau) = hVar(Z_{tn})$

$$\begin{aligned} P\left[|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > b_n \tilde{M}_{n,\epsilon} \right] &= P\left[\left| \frac{1}{n} \sum_{t=1}^{n} Z_{tn} \right| > b_n \tilde{M}_{n,\epsilon} \right] = P\left[\left| \sum_{t=1}^{n} Z_{tn} \right| > n b_n \tilde{M}_{n,\epsilon} \right] \\ &\leq 2\exp\left\{ -\frac{b_n^2 \tilde{M}_{n,\epsilon} n h_n}{2h_n Var[Z_{tn}] + \frac{2}{3} cCB_n b_n \tilde{M}_{n,\epsilon}} \right\} \end{aligned}$$

by Bernstein's inequality. Then,

$$\begin{aligned} P\left[\frac{1}{b_n} \max_{1\leq\tau\leq m} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > \tilde{M}_{n,\epsilon} \right] &\leq \sum_{\tau=1}^{m} 2\exp\left\{ -\frac{b_n^2 \tilde{M}_{n,\epsilon} n h_n}{2h_n Var[Z_{tn}] + \frac{2}{3} CcB_n b_n \tilde{M}_{n,\epsilon}} \right\} \\ &\leq 2m \max_{1\leq\tau\leq m} \exp\left\{ -\frac{b_n^2 \tilde{M}_{n,\epsilon} n h_n}{2h_n Var[Z_{tn}] + \frac{2}{3} CcB_n b_n \tilde{M}_{n,\epsilon}} \right\} \\ &= 2m \exp\left\{ -\frac{b_n^2 \tilde{M}_{n,\epsilon} n h_n}{2l_n(x^m) + \frac{2}{3} CcB_n b_n \tilde{M}_{n,\epsilon}} \right\} \qquad \text{(A.11)} \end{aligned}$$

where x^m corresponds to the point of the given function such that $\exp\left\{ -\frac{b_n^2 \tilde{M}_{n,\epsilon} n h_n}{2h_n Var[Z_{tn}] + \frac{2}{3} CcB_n b_n \tilde{M}_{n,\epsilon}} \right\}$ which the function $\exp\{\cdot\}$ attains its maximum value. Thus we have

$$l_n(x^m) = \int\int M_k^2(\psi) u^2 \chi_{\{|u^2| \leq B_n\}} f(x^m + h_n\psi) f_{u|X}(u) d\psi du - h_n \left[\int\int M_k(\psi) u \chi_{\{|u| \leq B_n\}} f(x^m + h_n\psi) f_{u|X}(u) d\psi du \right]^2 \qquad \text{(A.12)}$$

Let $b_n = \left(\frac{\log n}{nh_n} \right)^{1/2}$ and $r = \left(\frac{h_n^3}{n} \right)^{1/2}$. Given $\tilde{M}_{n,\epsilon} = \tilde{M}_\epsilon - \frac{2cr}{h_n^2} O_p(1) - O(B_n^{1-a})$ we have

$$b_n M_{n,\epsilon} = \left(\frac{\log n}{nh_n} \right)^{1/2} \tilde{M}_\epsilon - \frac{2c}{(nh_n)^{1/2}} O_p(1) - O(B_n^{1-a}) \qquad \text{(A.13)}$$

$$B_n b_n \tilde{M}_{n,\epsilon} = \left(\frac{\log n}{nh_n} \right)^{1/2} B_n \tilde{M}_\epsilon - \frac{2c}{(nh_n)^{1/2}} B_n O_p(1) - O(B_n^{2-a}) \quad \text{where } a > 2. \qquad \text{(A.14)}$$

We want $\left(\frac{\log n}{nh_n}\right) \longrightarrow 0$ as $n \to \infty$ that implies $b_n M_{n,\epsilon} \to 0$ for $a > 2$ as $n \to \infty$. From (A.13),

$$\begin{aligned}(b_n M_{n,\epsilon})^2 =& \left[\left(\frac{\log n}{nh_n}\right)^{1/2} \tilde{M}_\epsilon - \frac{2c}{(nh_n)^{1/2}} O_p(1) - O(B_n^{1-a})\right]^2 \\ =& \frac{\log n}{nh_n} \tilde{M}_\epsilon^2 + \frac{4c^2}{nh_n} O_p(1) + O(B_n^{2(1-a)}) - 2\left(\frac{\log n}{nh_n}\right)^{1/2} \tilde{M}_\epsilon \frac{2c}{(nh_n)^{1/2}} O_p(1) \\ & - 2\left(\frac{\log n}{nh_n}\right)^{1/2} \tilde{M}_\epsilon O(B_n^{1-a}) + 4c\left(\frac{1}{(nh_n)^{1/2}}\right) O_p(1) O(B_n^{1-a}).\end{aligned}$$

Note that

$$\begin{aligned}& - nh_n(b_n^2 \tilde{M}_{n,\epsilon}^2) \\ & = - \log n \Bigg[\tilde{M}_\epsilon^2 + \frac{4c^2}{\log n} O_p(1) + \frac{nh_n}{\log n} O(B_n^{2(1-a)}) - \frac{4c}{(\log n)^{1/2}} O_p(1) \\ & \qquad - 2\left(\frac{nh_n}{\log n}\right)^{1/2} \tilde{M}_\epsilon O(B_n^{1-a}) + 4c\left(\frac{nh_n}{\log n}\right)^{1/2} O(B_n^{1-a})\Bigg] = -\triangle_n \log n\end{aligned}$$

where $\triangle_n = \tilde{M}_\epsilon^2 + \frac{4c^2}{\log n} O_p(1) + \frac{nh_n}{\log n} O(B_n^{2(1-a)}) - \frac{4c}{(\log n)^{1/2}} O_p(1) - 2\left(\frac{nh_n}{\log n}\right)^{1/2} \tilde{M}_\epsilon O(B_n^{1-a})$ $+4c\left(\frac{nh_n}{\log n}\right)^{1/2} O(B_n^{1-a})$. Choose B_n such that $\left(\frac{\log n}{nh_n}\right)^{1/2} O(B_n^{1-a}) \to 0$ and $\left(\frac{nh_n}{\log n}\right)^{1/2} O(B_n^{1-a}) O(1) \to 0$ as $n \to \infty$ for $a > 2$. Let $B_n = O(\left(\frac{nh}{\log n}\right))$. Then, $\left(\frac{\log n}{nh_n}\right)^{1/2} O(B_n^{1-a}) = \left(\frac{\log n}{nh_n}\right)^{a-1/2} \to 0$ and $\left(\frac{nh_n}{\log n}\right)^{1/2} O(B_n^{1-a}) = \left(\frac{nh_n}{\log n}\right)^{2/3-a} O(1) \to 0$ as $n \to \infty$ for $a > 2$. In addition, $\left(\frac{\log n}{nh_n}\right)^{1/2} O(B_n^{1-a}) = o(1)$ implies that $\left(\frac{1}{nh_n}\right)^{1/2} O(B_n^{1-a}) = o(1)$. Let $v_n = 2l_n(x^m) + \frac{2}{3} C c B_n b_n \tilde{M}_{n,\epsilon}$. From (A.11), we have

$$P\left[\frac{1}{b_n} \max_{1 \le \tau \le m} |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > \tilde{M}_{n,\epsilon}\right] \le 2mn^{-\triangle_n/v_n} \le 2r_0 \left(\frac{1}{nh_n}\right)^{1/2} \frac{1}{h_n} \frac{1}{n^{\triangle_n/v_n - 1}}$$

The last inequality follows from that since F' is a covering for $\mathcal{G}$, it must be that $m \to \infty$ and since $\mathcal{G}$ is bounded there exists $x_0 \in \mathbb{R}$ and $r_0 < \infty$ such that $\mathcal{G} \subseteq B(x_0, r_0)$. That is, $2mr \le 2r_0$ which implies that $m \le r_0 \left(\frac{n}{h_n^3}\right)^{1/2}$.

From (A.12), $l_n(x^m) \to f(x^m) \int M_k^2(\psi) d\psi E[u^2|X] < \infty$ by Assumption 6. Since $nh_n \to \infty$ it suffices to have $n^{\triangle_n/v_n - 1} h_n$ bounded away from 0 as $n \to \infty$. We have $\triangle_n \to \tilde{M}_\epsilon^2$ and $v_n \to 2f(x^m) \int M_k^2(\psi) d\psi \sigma_u^2$. Choose $\tilde{M}_\epsilon$ large enough to have $\frac{\triangle_n}{v_n} - 1 \to \frac{\tilde{M}_\epsilon^2}{2f(x^m) \int M_k^2(\psi) d\psi \sigma_u^2} \ge 2$ to obtain

$n^{\triangle_n/v_n - 1}h_n \to \infty$ as $n \to \infty$. Then, we have $\sup_{x\in\mathcal{G}} |s_2(x) - E[s_2(x)]| = O_p\bigg(\left(\frac{\log n}{nh_n}\right)^{1/2} \bigg)$.

Hence, $\sup_{x\in\mathcal{G}} |\hat{g}_k(x) - E[\hat{s}_k(x)]| = O_p\bigg(\left(\frac{\log n}{nh_n}\right)^{1/2} \bigg)$. □

Proof of Theorem 1.7

Proof. Note that $\hat{g}_k(x) = \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right) Y_t$ and $E\left[\hat{g}_k(x)|X_t\right] = \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right) m(X_t)$. We have $\hat{g}_k(x) - E[\hat{g}_k(x)|X_t] = \frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t-x}{h_n}\right)[Y_t - m(X_t)]$. Let $Z_{tn} = \frac{1}{nh_n} M_k\left(\frac{X_t-x}{h_n}\right)[Y_t - m(X_t)]$ with $E[Z_{tn}] = 0$ where $m(X_t) = E[Y_t|X_t]$.

$$\begin{aligned} Var(Z_{tn}) &= E[Z_{tn}^2] = E\left[\left\{\frac{1}{nh_n}M_k\left(\frac{X_t - x}{h_n}\right)(Y_t - m(X_t))\right\}^2\right] = \frac{\sigma^2}{n^2h_n^2}E\left[M_k^2\left(\frac{X_t - x}{h_n}\right)\right] \\ &= \frac{\sigma^2}{n^2h_n^2}\int M_k^2\left(\frac{y-x}{h_n}\right)f(y)dy \end{aligned}$$

Let $S_n^2 = \sum_{t=1}^n E[Z_{tn}^2]$ and $X_{tn} = \frac{Z_{tn}}{S_n} = \frac{\frac{1}{nh_n}M_k\left(\frac{X_t-x}{h_n}\right)[Y_t - m(X_t)]}{\left[\frac{\sigma^2}{nh_n^2}\int M_k^2\left(\frac{X_t - x}{h_n}\right)f(X_t)dX_t\right]^{1/2}}$.

Then

$$S_n^2 = \frac{\sigma^2}{n^2h_n^2}\sum_{t=1}^n \int M_k^2\left(\frac{y-x}{h_n}\right)f(y)dy = \frac{\sigma^2}{nh_n^2}\int M_k^2\left(\frac{y-x}{h_n}\right)f(y)dy.$$

By Liapounov's CLT $\sum_{t=1}^n X_{tn} \xrightarrow{d} \mathcal{N}(0,1)$ provided that $\lim_{n\to\infty}\sum_{t=1}^n E\left[|X_{tn}|^{2+\delta}\right] = 0$ for some $\delta > 0$. Note that $|X_{tn}| = \frac{\left|M_k\left(\frac{X_t-x}{h_n}\right)\right||Y_t - m(X_t)|}{(nh_n)^{1/2}(c(n))^{1/2}}$ with $c(n) = \frac{\sigma^2}{h_n}\int M_k^2\left(\frac{y-x}{h_n}\right)f(y)dy$.

Therefore,

$$|X_{tn}|^{2+\delta} = \frac{\left|M_k\left(\frac{X_t-x}{h_n}\right)\right|^{2+\delta}|Y_t - m(X_t)|^{2+\delta}}{(nh_n)^{1+\delta/2}(c(n))^{1+\delta/2}} \text{ where } c(n) \text{ is non stochastic.}$$

$$E\left[|X_{tn}|^{2+\delta}\right] = (nh_nc(n))^{-1-\delta/2}E\left[\left|M_k\left(\frac{X_t-x}{h_n}\right)\right|^{2+\delta}|Y_t - m(X_t)|^{2+\delta}\right] \text{ and}$$
$$\sum_{t=1}^n E\left[|X_{tn}|^{2+\delta}\right] = (nh_nc(n))^{-1-\delta/2}\sum_{t=1}^n E\left[\left|M_k\left(\frac{X_t-x}{h_n}\right)\right|^{2+\delta}|Y_t - m(X_t)|^{2+\delta}\right].$$

Now given that $E\left[|Y_t - m(X_t)|^{2+\delta}|X_t\right] < \infty$, for some $C < \infty$,

$$E\left[\left|M_k\left(\frac{X_t-x}{h_n}\right)\right|^{2+\delta}|Y_t-m(X_t)|^{2+\delta}\right]=E\left[\left|M_k\left(\frac{X_t-x}{h_n}\right)\right|^{2+\delta}E\left(|Y_t-m(X_t)|^{2+\delta}|X_t\right)\right]$$
$$\leq CE\left[\left|M_k\left(\frac{X_t-x}{h_n}\right)\right|^{2+\delta}\right]=C\int\left|M_k\left(\frac{y-x}{h_n}\right)\right|^{2+\delta}f(y)dy.$$

Consequently,

$$\begin{aligned}\sum_{t=1}^{n}E\left[|X_{tn}|^{2+\delta}\right] &\leq (nh_n\ c(n))^{-1-\delta/2}n\ C\int\left|M_k\left(\frac{y-x}{h_n}\right)\right|^{2+\delta}f(y)dy\\ &= (nh_n)^{-\delta/2}(c(n))^{-1-\delta/2}C\int|M_k(\psi)|^{2+\delta}f(x+h_n\psi)d\psi.\end{aligned}$$

According to assumptions that $\sup_{x\in\mathbb{R}}|K(x)|<\infty$, $\int|K(x)|dx<\infty$ and $f\in B^r_{\infty,q}$, we have

$$\int|M_k(\psi)|^{2+\delta}f(x+h_n\psi)d\psi=\int\left|-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}K\left(\frac{\psi}{s}\right)\right|^{2+\delta}f(x+h_n\psi)d\psi$$
$$\leq C2^{1+\delta}\int\sum_{|s|=1}^{k}\left|\frac{c_{k,s}}{|s|}K\left(\frac{\psi}{s}\right)\right|^{2+\delta}|f(x+h_n\psi)|d\psi\quad\text{by } C_r \text{ inequality}$$
$$\leq C2^{1+\delta}\sum_{|s|=1}^{k}|c_{k,s}|^{2+\delta}\sup_{x\in\mathbb{R}}|f(x)|\int|K(t)|^{2+\delta}dt<\infty$$

since $f\in\mathcal{C}^0(\mathbb{R})$ (Assumption 2(2)) and Assumption 4(3)-(4).

Thus, $\lim_{n\to\infty}\sum_{t=1}^{n}E\left[|X_{tn}|^{2+\delta}\right]=0$. Then, $\sum_{t=1}^{n}X_{tn}\xrightarrow{d}\mathcal{N}(0,1)$ which implies
$\dfrac{\sum_{t=1}^{n}\frac{1}{nh_n}M_k\left(\frac{X_t-x}{h_n}\right)[Y_t-m(X_t)]}{\left[\frac{\sigma^2}{nh_n^2}\int M_k^2\left(\frac{X_t-x}{h_n}\right)f(X_t)dX_t\right]^{1/2}}\xrightarrow{d}\mathcal{N}(0,1)$.
Thus, $\sqrt{nh_n}\left[\hat{g}_k(x)-E\left(\hat{g}_k(x)|X_t\right)\right]\xrightarrow{d}\mathcal{N}\left(0,\sigma^2f(x)\int M_k^2(\psi)d\psi\right)$. □

Proof of Theorem 1.8

Proof. For $x\in\mathbb{R}$ such that $\hat{f}_k(x),f(x)\neq0$ we have

$$\frac{\hat{g}_k(x)}{\hat{f}_k(x)}-\frac{g(x)}{f(x)}=\frac{1}{f(x)}\left(\hat{g}_k(x)\left(1+\frac{\hat{f}_k(x)-f(x)}{f(x)}\right)^{-1}-g(x)\right)$$

For $z \neq -1$ we have by Taylor's Theorem that $(1+z)^{-1} = 1 - z + z^2 - \cdots$. Hence, since $\hat{f}_k(x) \neq 0$, we write

$$\frac{\hat{g}_k(x)}{f(x)}\left(1+\frac{\hat{f}_k(x)-f(x)}{f(x)}\right)^{-1} = \frac{\hat{g}_k(x)}{f(x)}\left(1-\left(\frac{\hat{f}_k(x)-f(x)}{f(x)}\right)+\left(\frac{\hat{f}_k(x)-f(x)}{f(x)}\right)^2-\cdots\right).$$

Consequently,

$$\frac{\hat{g}_k(x)}{f(x)}-\frac{g(x)}{f(x)} = \frac{\hat{g}_k(x)-g(x)}{f(x)}-\frac{\hat{g}_k(x)}{f(x)}\left(\frac{\hat{f}_k(x)-f(x)}{f(x)}\right)+\frac{\hat{g}_k(x)}{f(x)}\left(\frac{\hat{f}_k(x)-f(x)}{f(x)}\right)^2-\cdots.$$

Taking expectations and by the Triangle Inequality, we have

$$\begin{aligned}|E(\hat{m}_k(x))-m(x)| \leq{}& \frac{|E(\hat{g}_k(x)-g(x))|}{f(x)}+\frac{1}{f(x)^2}\Big|E\Big(\hat{g}_k(x)(\hat{f}_k(x)-f(x))\Big)\Big| \\ &+\frac{1}{f(x)^3}\Big|E\Big(\hat{g}_k(x)(\hat{f}_k(x)-f(x))^2\Big)\Big|-\cdots.\end{aligned}$$

By Hölder's Inequality and Theorem 1.1 and 1.5 we have,

$$|E(\hat{m}_k(x))-m(x)| \leq Ch_n^r + C\left(\frac{1}{(nh_n)^{1/2}}\left(h_n^r+\frac{1}{(nh_n)^{1/2}}\right)\right).$$

Finally, if $nh_n^{r+1} \to C$ we have $|E(\hat{m}_k(x))-m(x)| = O(h_n^r)$. □

Proof of Theorem 1.9

Proof. From Theorem 1.1, $E\Big((\hat{f}_k(x)-f(x))^2\Big) = O(h_n^{2r}+(nh_n)^{-1})$ which implies that $\hat{f}_k(x) - f(x) = O_p(h_n^r+(nh_n)^{-1/2})$. Similarly, from Theorem 1.5, $\hat{g}_k(x)-g(x) = O(h_n^r+(nh_n)^{-1})$. Now,

$$\begin{aligned}\hat{m}_k(x)-m(x) &= \frac{\hat{g}_k(x)}{\hat{f}_k(x)}-\frac{g(x)}{f(x)} \\ &= \frac{(\hat{g}_k(x)+O_p(h_n^r+(nh_n)^{-1/2}))f(x)-(f(x)+O_p(h_n^r+(nh_n)^{-1/2}))g(x)}{f(x)(f(x)+O_p(h_n^r+(nh_n)^{-1/2}))} \\ &= O_p(h_n^r+(nh_n)^{-1/2})\end{aligned}$$

□

Proof of Theorem 1.10

Proof. Note that $\hat{m}_k(x) - E[\hat{m}_k(x)|X_t] = \frac{\hat{g}_k(x) - E[\hat{g}_k(x)|X_t]}{\hat{f}_k(x)}$. From Theorem 1.2, we know that $\hat{f}_k(x) - f(x) = o_p(1)$ for all $x \in \mathbb{R}$. Consequently, we have the following result.

$$\sqrt{nh}\Big(\hat{g}_k(x) - E[\hat{g}_k(x)|X_t]\Big)/\hat{f}_k(x) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 f(x)^{-1} \int M_k^2(\psi)d\psi\right) \tag{A.15}$$

$$\sqrt{nh_n}\,(\hat{m}_k(x) - m(x)) = \sqrt{nh_n}\,(\hat{m}_k(x) - E[\hat{m}_k(x)|X_t]) + \sqrt{nh_n}\,[E(\hat{m}_k(x)|X_t) - m(x)]$$

From (A.15), we know $\sqrt{nh_n}\,(\hat{m}_k(x) - E[\hat{m}_k(x)|X_t]) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \frac{1}{f(x)} \int M_k^2(\psi)d\psi\right)$. Thus, we need to consider $\sqrt{nh_n}\,[E(\hat{m}_k(x)|X_t) - m(x)]$. Note that

$$\begin{aligned} E[\hat{m}_k(x)|X_t] - m(x) &= \frac{\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right) m(X_t)}{\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)} - m(x) \\ &= \frac{1}{\hat{f}_k(x)}\left[\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)\Big(m(X_t) - m(x)\Big)\right] \end{aligned}$$

Now, we show that $\left[\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)\Big(m(X_t) - m(x)\Big)\right] = O_p(h_n^r)$.

$$\begin{aligned} &E\left[\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)\Big(m(X_t) - m(x)\Big)\right] \\ &= \frac{1}{h_n}\left[-\frac{1}{c_{k,0}}\sum_{|s|=1}^k \frac{c_{k,s}}{|s|}\right]\int K\left(\frac{y - x}{sh_n}\right)\big(m(y) - m(x)\big)f(y)dy \\ &= -\frac{1}{c_{k,0}}\left[\int K(\psi)\sum_{|s|=0}^k c_{k,s} m(x + sh_n\psi) f(x + sh_n\psi)d\psi - \int K(\psi)\sum_{|s|=0}^k c_{k,s} m(x) f(x + sh_n\psi)d\psi\right] \\ &= -\frac{1}{c_{k,0}}\left[\int K(\psi)[\Delta_{h_n\psi}^{2k} m(x)f(x)]d\psi - m(x)\int K(\psi)[\Delta_{h_n\psi}^{2k} f(x)]d\psi\right] = O(h_n^r) \end{aligned}$$

Therefore $E\left[\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)\Big(m(X_t) - m(x)\Big)\right] = O(h_n^r)$ since $\hat{f}(x) = f(x) + O_p(h_n^r + (nh_n)^{-1/2})$ and $E(E[\hat{m}_k(x)|X_t] - m(x)) = E(\hat{m}_k(x)) - m(x)$, we have $Bias(\hat{m}_k(x)) = O(h_n^r)$. Hence $\frac{1}{nh_n}\sum_{t=1}^n M_k\left(\frac{X_t - x}{h_n}\right)\Big(m(X_t) - m(x)\Big) = O_p(h_n^r)$ Note that $E[\hat{m}_k(x)|X_t] - m(x) = \frac{1}{f(x)} O_p(h_n^r)$.

Consequently,

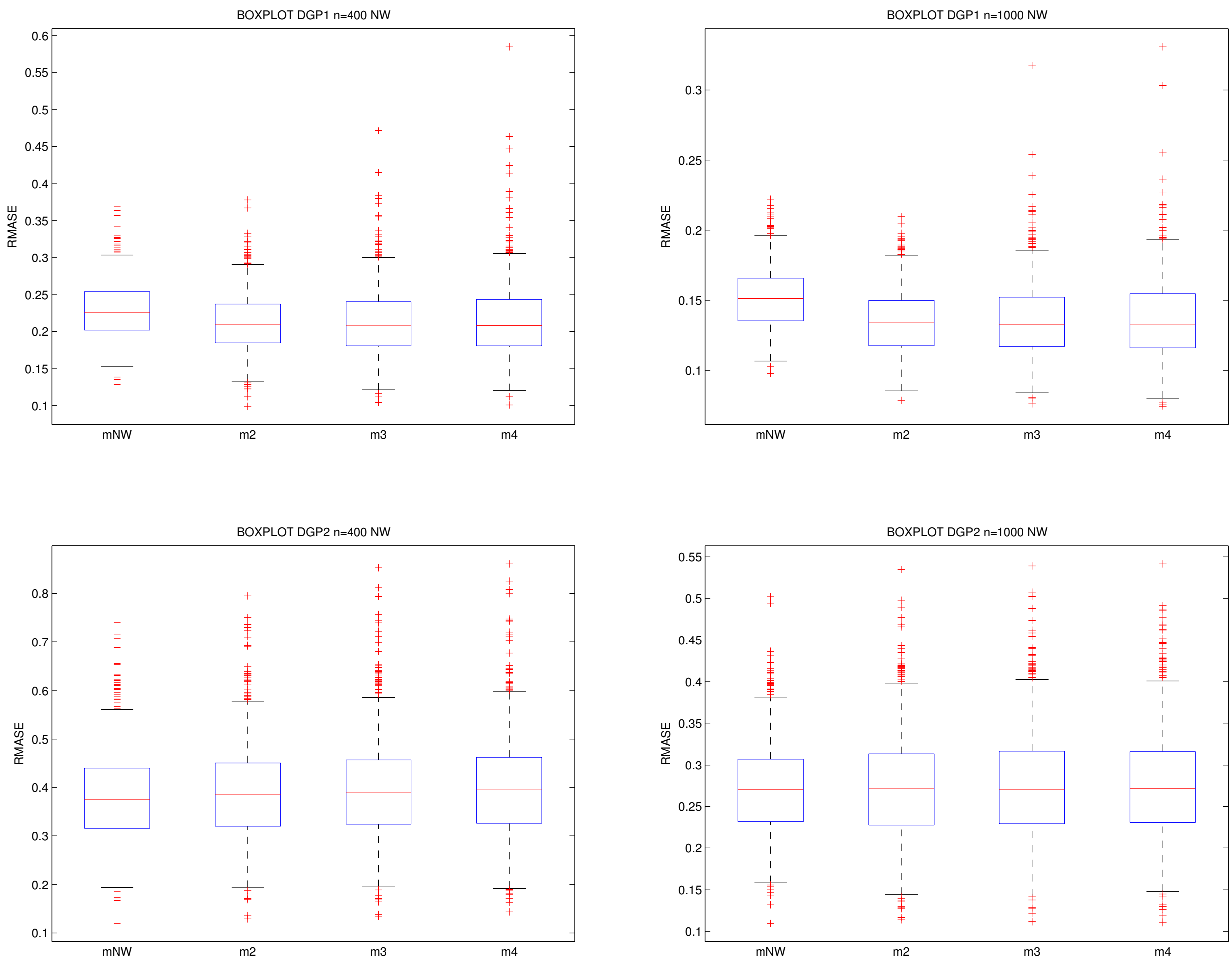
$$\sqrt{nh_n}(\hat{m}_k(x) - m(x)) = \sqrt{nh_n}\,(\hat{m}_k(x) - E[\hat{m}_k(x)|X_t]) + \sqrt{nh_n}\,(E[\hat{m}_k(x)|X_t] - m(x))$$
$$= \sqrt{nh_n}\,(\hat{m}_k(x) - E[\hat{m}_k(x)|X_t]) + \sqrt{nh_n}\; O_p(h_n^r) = \sqrt{nh_n}\,(\hat{m}_k(x) - E[\hat{m}_k(x)|X_t] + O_p(h^r))$$
$$\xrightarrow{d} \mathcal{N}\left(0, \sigma^2 f(x)^{-1} \int M_k^2(\psi) d\psi\right)$$

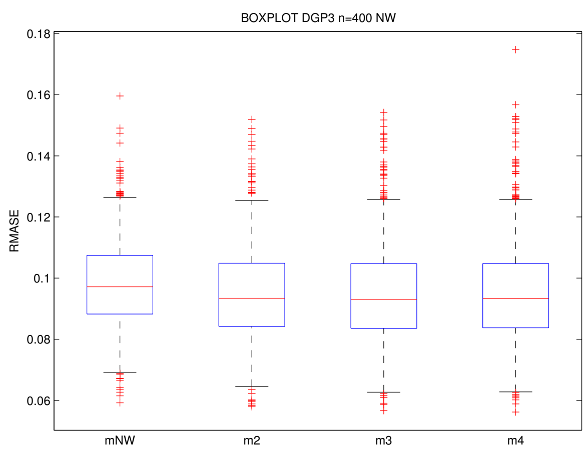
If $nh_n^{1+2r} \to 0$ as $n \to \infty$, we have $\sqrt{nh_n}(\hat{m}_k(x) - m(x)) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 f(x)^{-1} \int M_k^2(\psi) d\psi\right)$. □

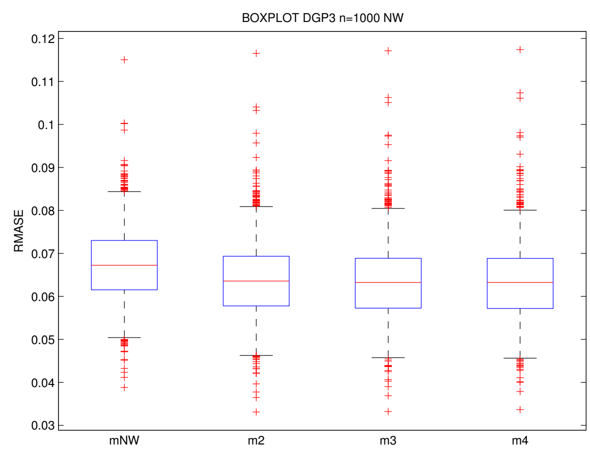
Table A.1: Local constant estimators with cross validation bandwidth h^{CV}; Average Absolute Bias (B); Average Variance (V); Average Root Mean Squared Error (R).

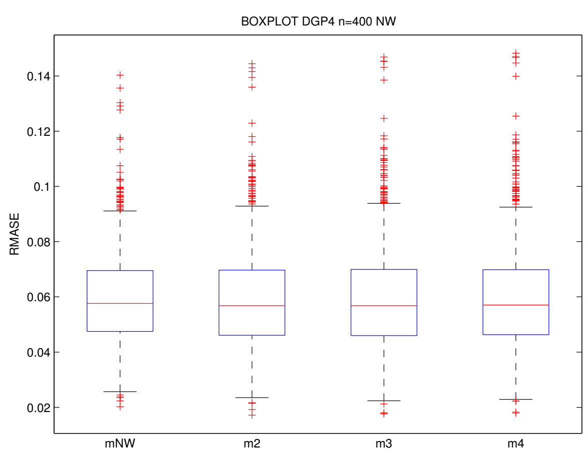
	$m_1(x)$			$m_2(x)$		
$n = 400$	B	V	R	B	V	R
$\hat{m}_{NW}$	0.0517	0.0470	0.2320	0.0384	0.1500	0.3909
$\hat{m}_2$	0.0395	0.0430	0.2164	0.0151	0.1622	0.4032
$\hat{m}_3$	0.0355	0.0453	0.2194	0.0123	0.1675	0.4095
$\hat{m}_4$	0.0334	0.0473	0.2227	0.0114	0.1703	0.4128
	$m_3(x)$			$m_4(x)$		
$n = 400$	B	V	R	B	V	R
$\hat{m}_{NW}$	0.0369	0.0078	0.0993	0.0171	0.0032	0.0618
$\hat{m}_2$	0.0268	0.0076	0.0960	0.0120	0.0034	0.0619
$\hat{m}_3$	0.0232	0.0077	0.0960	0.0110	0.0035	0.0624
$\hat{m}_4$	0.0213	0.0077	0.0963	0.0108	0.0035	0.0627
	$m_1(x)$			$m_2(x)$		
$n = 1000$	B	V	R	B	V	R
$\hat{m}_{NW}$	0.0360	0.0203	0.1534	0.0183	0.0761	0.2775
$\hat{m}_2$	0.0270	0.0172	0.1373	0.0076	0.0787	0.2807
$\hat{m}_3$	0.0225	0.0183	0.1392	0.0070	0.0795	0.2821
$\hat{m}_4$	0.0207	0.0188	0.1402	0.0071	0.0797	0.2825
	$m_3(x)$			$m_4(x)$		
$n = 1000$	B	V	R	B	V	R
$\hat{m}_{NW}$	0.0251	0.0036	0.0681	0.0127	0.0014	0.0426
$\hat{m}_2$	0.0168	0.0034	0.0646	0.0087	0.0015	0.0421
$\hat{m}_3$	0.0139	0.0033	0.0642	0.0082	0.0015	0.0422
$\hat{m}_4$	0.0124	0.0033	0.0643	0.0083	0.0015	0.0424

Figure A.1: Box plots of RMSE for estimators $\hat{m}_{NW}, \hat{m}_2, \hat{m}_3$ and $\hat{m}_4$ and four DGPs. DGP1, DGP2, DGP3 and DGP4 indicate $m_1(x)$, $m_2(x)$, $m_3(x)$ and $m_4(x)$ respectively. We consider the sample size $n = 400$ and $n = 1000$




BOXPLOT DGP3 n=400 NW
RMASE
0.18
0.16
0.14
0.12
0.1
0.08
0.06
mNW
m2
m3
m4


BOXPLOT DGP3 n=1000 NW
RMASE
0.12
0.11
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
mNW
m2
m3
m4


BOXPLOT DGP4 n=400 NW
RMASE
0.14
0.12
0.1
0.08
0.06
0.04
0.02
mNW
m2
m3
m4

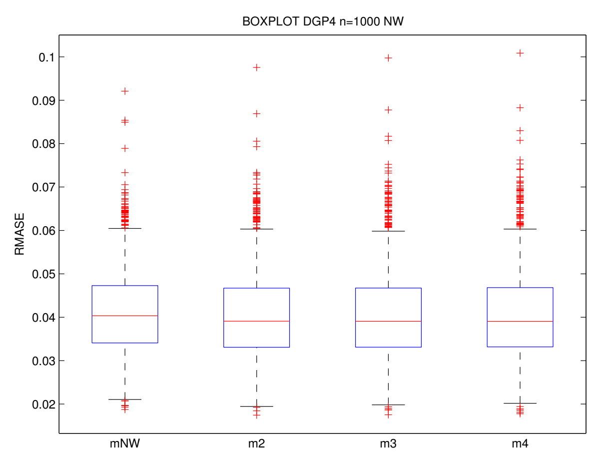
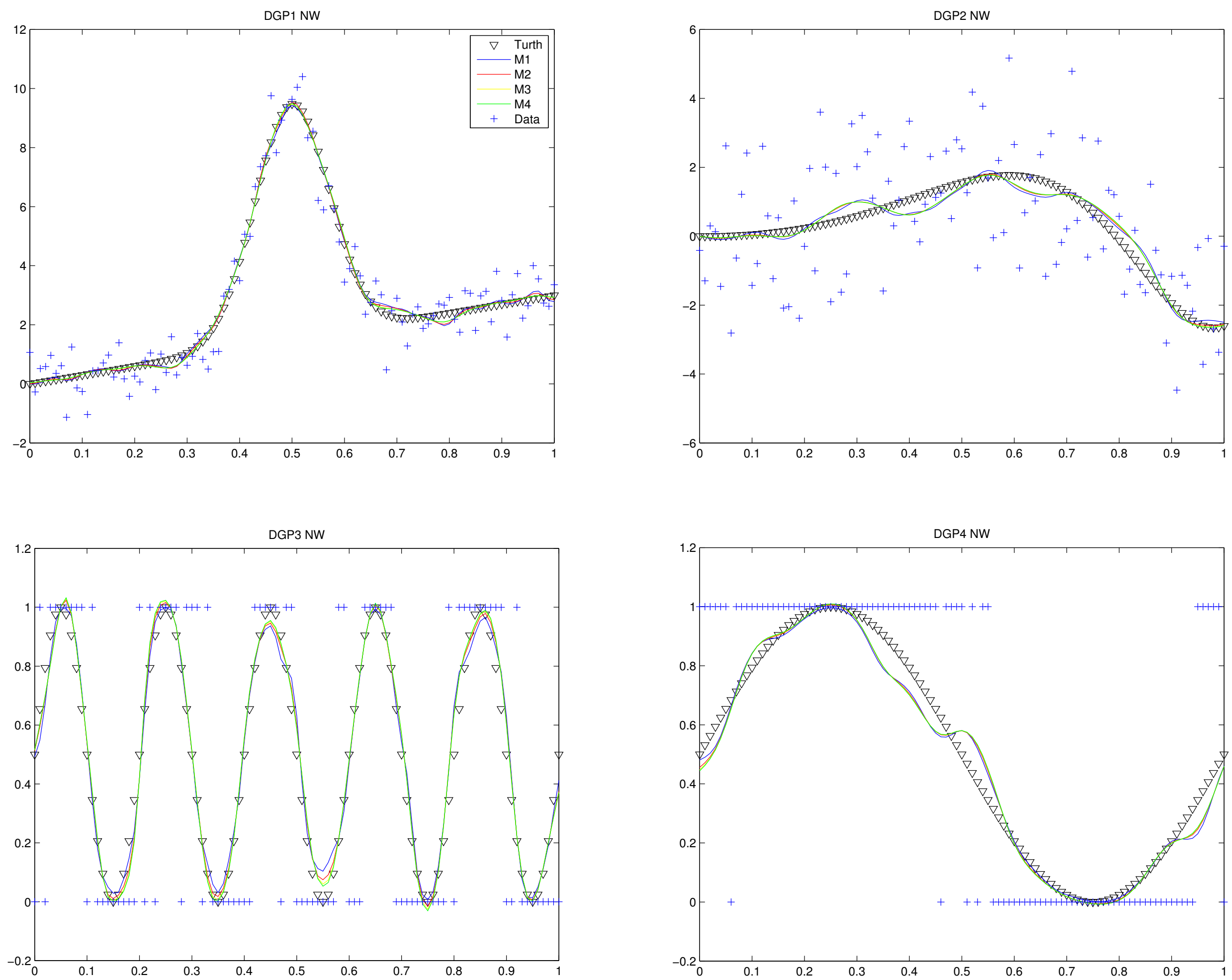

BOXPLOT DGP4 n=1000 NW
RMASE
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
mNW
m2
m3
m4

Figure A.2: These figures represent four data generating processes with four local constant regression estimators when $n = 400$. $\bigtriangledown$ is a true line, the blue line is NW regression estimator, the red line is a local constant estimator based on M_2 kernel. The yellow line indicates a local constant estimator baed on M_3 kernel. The green line represents a local constant estimator based on M_4. $+$ is an observed data points.



A.2 Appendix for Chapter 2: Proofs and tables

Proof of Theorem 2.1

Proof. (a) Given the independent and identically distributed (IID) assumption and $M_k(-x) = M_k(x)$, we have

$$\hat{F}_k(x) = \int_{-\infty}^{x} \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{v - X_t}{h_n}\right) dv = \frac{1}{n} \sum_{t=1}^{n} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \int_{-\infty}^{\frac{x-X_t}{sh_n}} K(\phi)d\phi$$
$$= \frac{1}{n} \sum_{t=1}^{n} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s}\mathcal{G}\left(\frac{x - X_t}{sh_n}\right)$$

where $\mathcal{G}(x) = \int_{-\infty}^{x} K(v)dv$.

$$E[\hat{F}_k(x)] = E\Bigg[-\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s}\mathcal{G}\left(\frac{x - X_1}{sh_n}\right)\Bigg] = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \int c_{k,s}\mathcal{G}(\psi)f(x - sh_n\psi)sh_n d\psi$$
$$= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \int c_{k,s}\mathcal{G}(\psi)dF(x - sh_n\psi)(-1)$$
$$= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \left[-\big[\mathcal{G}(\psi)F(x - sh_n\psi)\big]\Big|_{\psi=-\infty}^{\psi=+\infty} + \int K(\psi)F(x - sh_n\psi)d\psi\right]$$
$$= -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \int K(\psi)F(x - sh_n\psi)d\psi$$

Hence, the bias for $\hat{F}_k$ is,

$$Bias(\hat{F}_k(x)) = E[\hat{F}_k(x) - F(x)] = -\frac{1}{c_{k,0}} \sum_{|s|=0}^{k} c_{k,s} \int K(\psi)F(x - sh_n\psi)d\psi$$
$$= -\frac{1}{c_{k,0}} \int K(\psi)\Delta^{2k}_{-h_n\psi}F(x)d\psi \tag{A.16}$$

(b) From the result of (a), we have

$$|Bias(\hat{F}_k(x))| = \left|-\frac{1}{c_{k,0}} \int K(\psi)\Delta^{2k}_{-h_n\psi}F(x)d\psi\right|$$

$$= \left|\frac{1}{c_{k,0}}\right| \left[\int \left\{|K(\psi)||h_n\psi|^{r+1+1/q}\right\}^{q'} d\psi\right]^{1/q'} \left[\int \left\{\frac{\sup_{x\in\mathbb{R}} |\Delta^{2k}_{-h_n\psi}F(x)|}{|h_n\psi|^{r+1+1/q}}\right\}^{q} d\psi\right]^{1/q}$$

$$= h_n^{r+1} \left|\frac{1}{c_{k,0}}\right| \left[\int \left\{|K(\psi)||\psi|^{r+1+1/q}\right\}^{q'} d\psi\right]^{1/q'} ||F||_{b^{r+1}_{\infty,q}}$$

where $1/q + 1/q' = 1$ for $1 \leq q \leq \infty$ and $r + 1 < 2k$.

(c) The proof follows directly from (b) since $h_n \to 0$ as $n \to \infty$.

(d) Since $\mathcal{G}(\phi) = \int_{-\infty}^{\phi} K(x)dx$,

$$Var(\hat{F}_k(x)) = \frac{1}{n}\Big(-\frac{1}{c_{k,0}}\Big)^2 \sum_{|s|=1}^{k} c_{k,s}^2 Var\Big(\int_{-\infty}^{\frac{x-X_1}{sh_n}} K(\psi)d\psi\Big)$$

$$\leq \frac{1}{n}\Big(-\frac{1}{c_{k,0}}\Big)^2 \sum_{|s|=1}^{k} c_{k,s}^2 E\Big(\mathcal{G}^2\Big(\frac{x-X_1}{sh_n}\Big)\Big)$$

$$= \frac{1}{n}\Big(-\frac{1}{c_{k,0}}\Big)^2 \sum_{|s|=1}^{k} c_{k,s}^2 \int \mathcal{G}^2(\psi)dF(x - sh_n\psi) = O(n^{-1})$$

by the arguments in Theorem 2.1 and the fact that $\sup_{x\in\mathbb{R}} |F(x)| \leq 1$ and $|\mathcal{G}(\phi)| = |\int_{-\infty}^{\phi} K(v)dv| \leq C$ given that K has a compact support. Therefore, as $n \to \infty$, we have $Var(\hat{F}_k(x)) \to 0$ and $|Bias(\hat{F}_k(x))| \to 0$. □

Proof of Theorem 2.2

Proof. (a)

$$E[\hat{a}_k(y)] - a(y)$$

$$= E\left[\int y\hat{f}_k(y)dy - \mu\right] + 2E\left[\hat{F}_k(y) - F(y)\right] - 2yE\left[\int_{-\infty}^{y} x\hat{f}_k(x)dx - \int_{-\infty}^{y} xf(x)dx\right]$$

$$= E\left[\int y\hat{f}_k(y)dy - \mu\right] + 2E\left[\hat{F}_k(y) - F(y)\right] - 2y\left[\int_{-\infty}^{y} xE\left[\hat{f}_k(x) - f(x)\right]dx\right]$$

$$E\left[\int y\hat{f}_k(y)dy\right] = \frac{1}{n}\sum_{t=1}^{n} -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} \int (X_t + sh_n\psi)K(\psi)d\psi = \frac{1}{n}\sum_{t=1}^{n} X_t$$

Note that $E[\hat{f}_k(x) - f(x)] \leq Ch_n^r ||f||_{b^r_{\infty,q}}$ and $E[\hat{F}_k(x) - F(x)] \leq Ch_n^{r+1}||F||_{b^r_{\infty,q}}$ for $r+1 < 2k$. Assume that $\int_{-\infty}^y xdx = O(1)$ provided the existence of lower bound of x. Therefore, $\hat{a}_k(x)$ is asymptotically unbiased.

(b) Now we consider $Var(\hat{a}_k(y))$.

$$\begin{aligned}
Var(\hat{a}_k(y)) &= Var\big(\hat{\mu} - y + 2\hat{F}_k(y) - 2y\int_{-\infty}^y x\hat{f}_k(x)dx\big)\\
&= Var\big(\hat{\mu} + 2\hat{F}_k(y) - 2y\int_{-\infty}^y x\hat{f}_k(x)dx\big)\\
&= Var\big(\hat{\mu}\big) + 4Var\big(\hat{F}_k(y)) + 4y^2 Var\big(\int_{-\infty}^y x\hat{f}_k(x)dx\big) + 4Cov(\hat{\mu}, \hat{F}_k(y))\\
&- 4yCov(\hat{\mu}, \int_{-\infty}^y x\hat{f}_k(x)dx) - 8yCov(\hat{F}_k(y), 2y\int_{-\infty}^y x\hat{f}_k(x)dx)
\end{aligned}$$

where $\hat{\mu} = \int x\hat{f}_k(x)dx$.

First, we show that $Var(\hat{\mu}) = O(n^{-1})$. Consequently, we conclude that $Var(\hat{\mu}) \to 0$ as $n \to \infty$.

$$Var(\hat{\mu}) = Var\left(\frac{1}{n}\sum_{t=1}^n X_t\right) = \frac{1}{n^2}\sum_{t=1}^n Var(X_t) = \frac{\sigma^2}{n}$$

From Theorem 2.1 (d), we show that $Var(\hat{F}_k(x)) = O(n^{-1})$. Consequently, $Var(\hat{F}_k(x)) \to 0$ as $n \to \infty$. Next, we need to show that $Var(\int_{-\infty}^y x\hat{f}_k(x)dx) \to 0$ as $n \to \infty$. Let $\mathcal{M}_k(x) = \int_{-\infty}^x M_k(v)dv$. Suppose that $E[X_t^{2p}] < \infty$ for $p > 1$.

$$\begin{aligned}
&Var\bigg(\int_{-\infty}^y x\hat{f}_k(x)dx\bigg) \leq E\bigg[\Big\{\int_{-\infty}^y x\hat{f}_k(x)dx\Big\}^2\bigg]\\
&= \frac{1}{n}E\bigg[\Big\{\int_{-\infty}^{\frac{y-X_1}{h_n}} (X_1 + h_n\psi)M_k(\psi)d\psi\Big\}^2\bigg]\\
&\leq \frac{2}{n}E\bigg[\Big\{X_1\mathcal{M}_k\Big(\frac{y-X_1}{h_n}\Big)\Big\}^2\bigg] + \frac{2}{n}h_n^2 E\bigg[\Big(\int_{-\infty}^{\frac{y-X_1}{h_n}} \psi M_k(\psi)d\psi\Big)^2\bigg]\\
&\leq \frac{2}{n}\bigg\{\Big(E[X_1^{2p}]\Big)^{1/p}\Big(E\Big[\mathcal{M}_k^{2q}\Big(\frac{y-X_1}{h_n}\Big)\Big]\Big)^{1/q}\bigg\} + \frac{2}{n}h_n^2 E\bigg[\Big(\int_{-\infty}^{\frac{y-X_1}{h_n}} \psi M_k(\psi)d\psi\Big)^2\bigg]
\end{aligned}$$

where $p > 1$ and $1/p + 1/q = 1$.

Consider the part of $E\Big[\mathcal{M}_k^{2q}\Big(\frac{y-X_1}{h_n}\Big)\Big] = O(1)$.

$$\left|E\Big[\mathcal{M}_k^{2q}\Big(\frac{y-X_1}{h_n}\Big)\Big]\right| = \left|\Big\{-\Big[\mathcal{M}_k^{2q}(\phi)F(y-h_n\phi)\Big]\Big|_{\phi=-\infty}^{\phi=\infty} + 2q\int \mathcal{M}_k^{2q-1}(\phi)M_k(\phi)F(y-h_n\phi)d\phi\Big\}\right|$$
$$\leq 2q\int|\mathcal{M}_k^{2q-1}(\phi)||M_k(\phi)||F(y-h_n\phi)|d\phi \leq 2q\int|M_k(\phi)|d\phi<\infty$$

provided that $\int|M_k(\psi)|d\psi<\infty$, $\sup_{x\in\mathbb{R}}\mathcal{M}_k(x)\leq 1$ and $\sup_{x\in\mathbb{R}}F(x)\leq 1$.

$$h_n^2E\Big[\Big(\int_{-\infty}^{\frac{y-X_1}{h_n}}\psi M_k(\psi)d\psi\Big)^2\Big] = h_n^2\Big[\int\Big(\int_{-\infty}^{\phi}\psi M_k(\psi)d\psi\Big)^2 dF(y-h_n\phi)(-1)\Big]$$
$$= h_n^2\Big\{-\Big[\Big(\int_{-\infty}^{\phi}\psi M_k(\psi)d\psi\Big)^2F(y-h_n\phi)\Big]\Big|_{\phi=-\infty}^{\phi=\infty} + \int 2\Big(\int_{-\infty}^{\phi}\psi M_k(\psi)d\psi\Big)\phi M_k(\phi)F(y-h_n\phi)d\phi\Big\}$$
$$= 2h_n^2\int\Big(\int_{-\infty}^{\phi}\psi M_k(\psi)d\psi\Big)\phi M_k(\phi)F(y-h_n\phi)d\phi<\infty \text{ since } \int\psi M_k(\psi)d\psi\leq\int|\psi||M_k(\psi)|d\psi<\infty.$$

Therefore, $Var\Big(\int_{-\infty}^{y}x\hat{f}_k(x)dx\Big)\to 0$ as $n\to\infty$.

From now, we consider the part of covariance. First, we compute $Cov(\hat{\mu},\hat{F}_k(y))$.

$$Cov(\hat{\mu},\hat{F}_k(y)) = E[\hat{\mu}\hat{F}_k(y)] - E[\hat{\mu}]E[\hat{F}_k(y)]$$

$$= E\Big[\Big(\frac{1}{n}\sum_{t=1}^{n}X_t\Big)\Big(\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)d\psi\Big)\Big] - E\Big[\frac{1}{n}\sum_{t=1}^{n}X_t\Big]E\Big[\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\frac{y-X_t}{h_n}}M_k(\psi)d\psi\Big]$$
$$= E\Big[\Big(\frac{1}{n}\sum_{t=1}^{n}X_t\Big)\Big(\frac{1}{n}\sum_{t=1}^{n}\mathcal{M}_k\Big(\frac{y-X_t}{h_n}\Big)\Big)\Big] - E\Big[\Big(\frac{1}{n}\sum_{t=1}^{n}X_t\Big)\Big]E\Big[\frac{1}{n}\sum_{t=1}^{n}\mathcal{M}_k\Big(\frac{y-X_t}{h_n}\Big)\Big]$$
$$= \frac{1}{n^2}E\Big[\sum_{t=1}^{n}X_t\mathcal{M}_k\Big(\frac{y-X_t}{h_n}\Big)\Big] + \frac{n(n-1)}{n^2}E[X_1]E\Big[\mathcal{M}_k\Big(\frac{y-X_t}{h_n}\Big)\Big] - E[X_1]E\Big[\mathcal{M}_k\Big(\frac{y-X_1}{h_n}\Big)\Big]$$
$$= \frac{1}{n}E\Big[X_1\mathcal{M}_k\Big(\frac{y-X_1}{h_n}\Big)\Big] - \frac{1}{n}E[X_1]E\Big[\mathcal{M}_k\Big(\frac{y-X_1}{h_n}\Big)\Big]$$
$$= \frac{1}{n}\int X_1\mathcal{M}_k\Big(\frac{y-X_1}{h_n}\Big)f(X_1)dX_1 - \frac{1}{n}E[X_1]\int\mathcal{M}_k\Big(\frac{y-X_1}{h_n}\Big)f(X_1)dX_1 = O(n^{-1})$$

since $|\mathcal{M}_k(\psi)| = \left|\int_{-\infty}^{\psi}M_k(x)dx\right| \leq \left|-\frac{1}{c_{k,0}}\right|\left|\sum_{|s|=1}^{k}c_{k,s}\right|\int|K(\phi)|d\phi<\infty$ provided $\int|K(x)|dx<\infty$.

$$
\begin{aligned}
&Cov\Big(\hat{\mu}, \int_{-\infty}^{y} x\hat{f}_k(x)dx\Big)\\
&= E\Big[\hat{\mu}\int_{-\infty}^{y} x\hat{f}_k(x)dx\Big] - E[\hat{\mu}]E\Big[\int_{-\infty}^{y} x\hat{f}_k(x)dx\Big]\\
&= E\Big[\frac{1}{n}\sum_{t=1}^{n} X_t \int_{-\infty}^{y} x\frac{1}{nh_n}\sum_{t=1}^{n} M_k\Big(\frac{x-X_t}{h_n}\Big)dx\Big] - E\Big[\frac{1}{n}\sum_{t=1}^{n} X_t\Big]E\Big[\int_{-\infty}^{y} x\frac{1}{nh_n}\sum_{t=1}^{n} M_k\Big(\frac{x-X_t}{h_n}\Big)dx\Big]\\
&= \frac{1}{nh_n}E\Big[X_1\int_{-\infty}^{y} xM_k\Big(\frac{x-X_1}{h_n}\Big)dx\Big] - \frac{1}{nh_n}E[X_1]E\Big[\int_{-\infty}^{y} xM_k\Big(\frac{x-X_1}{h_n}\Big)dx\Big]\\
&\leq \frac{1}{nh_n}(E[|X_1|^2])^{1/2}\Big(E\Big[\Big\{\int_{-\infty}^{y} xM_k\Big(\frac{x-X_1}{h_n}\Big)dx\Big\}^2\Big]\Big)^{1/2} - \frac{1}{nh_n}E[X_1]E\Big[\int_{-\infty}^{y} xM_k\Big(\frac{x-X_1}{h_n}\Big)dx\Big]\\
&= \frac{1}{n}(E[|X_1|^2])^{1/2}E\Big[\Big(\int_{-\infty}^{\frac{y-X_1}{h_n}} (X_1+h_n\psi)M_k(\psi)d\psi\Big)^2\Big]^{1/2}\\
&- \frac{1}{n}E[X_1]E\Big[X_1\int_{-\infty}^{\frac{y-X_1}{h_n}} M_k(\psi)d\psi + h_n\int_{-\infty}^{\frac{y-X_1}{h_n}} \psi M_k(\psi)d\psi\Big]\\
&= O(n^{-1}) \text{ given that } \textstyle\int |M_k(\psi)|d\psi < \infty \text{ and } \int |\psi||M_k(\psi)|d\psi < \infty.
\end{aligned}
$$

Similarly, we have $Cov\Big(\hat{F}_k(y), \int_{-\infty}^{y} x\hat{f}_k(x)dx\Big) = O(n^{-1})$. Therefore, $Var(\hat{a}_k(y)) = O(n^{-1})$ and $Var(\hat{a}_k(y)) \to 0$ as $n \to \infty$. Consequently, $\hat{a}_k(y) - a(y) = o_p(1)$. □

Proof of Theorem 2.3

Proof.

$$
\begin{aligned}
[\hat{P}_\alpha(\hat{F}_k) - P_\alpha(F)] &= \int \hat{f}_k(y)^\alpha \hat{a}_k(y)d\hat{F}_k(y) - \int f(y)^\alpha a(y)dF(y)\\
&= \int \hat{p}_\alpha(y)d\hat{F}_(y) - \int p_\alpha(y)dF(y)\\
&= \int [\hat{p}_\alpha(y) - p_\alpha(y)]dF(y) + \int p_\alpha(y)d[\hat{F}_k(y) - F(y)] + \int [\hat{p}_\alpha(y) - p_\alpha(y)]d[\hat{F}_k - F](y)
\end{aligned}
$$

where $\hat{p}_\alpha(y) \equiv \hat{f}_k^\alpha(y)\hat{a}_k(y)$ and $p_\alpha(y) \equiv f(y)^\alpha a(y)$.

$$
\begin{aligned}
&\hat{p}_\alpha(y) - p_\alpha(y) = \hat{f}_k^\alpha(y)\hat{a}_k(y) - f(y)^\alpha a(y)\\
&= [\hat{f}_k(y)^\alpha - f(y)^\alpha][\hat{a}_k(y) - a(y)] + [\hat{f}_k(y)^\alpha - f(y)^\alpha]a(y) + f(y)^\alpha[\hat{a}_k(y) - a(y)]\\
&= \alpha f^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)]a(y) + \alpha f^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)][\hat{a}_k(y) - a(y)] + f(y)^\alpha[\hat{a}_k(y) - a(y)]
\end{aligned}
$$

According to the Mean Value Theorem, we have $\hat{f}_k(y)^\alpha - f(y)^\alpha = \alpha f^*(y)^{\alpha-1}(\hat{f}_k(y) - f(y))$ where $f^*(y)$ lies in between f and $\hat{f}_k$ for all $y \in \mathbb{R}$. That is, $f^*(y) = \lambda_y \hat{f}_k(y) + (1-\lambda_y)f(y)$ where $\lambda_y \in (0,1)$ for all $y \in \mathbb{R}$.

$$|f^*(y)| \leq (1-\lambda_y)|\hat{f}_k(y)| + (1-\lambda_y)|f(y)| \leq |\hat{f}_k(y)| + |f(y)| \text{ for } y \in \mathbb{R}.$$

Hence, given the assumptions of $\sup_{y\in\mathbb{R}} |K(y)| < \infty$ and $\sup_{y\in\mathbb{R}} |f(y)| < \infty$ for $y \in \mathbb{R}$, we have

$$\sup_{y\in\mathbb{R}} |f^*(y)| \leq \frac{1}{nh_n}\Big| - \frac{1}{c_{k,0}}\Big| \sum_{t=1}^{n} \sum_{|s|=1}^{k} \Big|\frac{c_{k,s}}{|s|}\Big|\Big|K\Big(\frac{y - X_t}{sh_n}\Big)\Big| + |f(y)| < \infty \text{ for each } n\ .$$

Hence, $\sup_{y\in\mathbb{R}} |f^*(y)|^{\alpha-1} < \infty$ where $\alpha \in [0.25, 1]$.

$$\begin{aligned} E\Big[|f^*(y)^{\alpha-1}(\hat{f}_k(y) - f(y))|\Big] &\leq \Big(E\Big[|f^*(y)|^{2\alpha-2}\Big]\Big)^{1/2}\Big(E\Big[|\hat{f}_k(y) - f(y)|^2\Big]\Big)^{1/2} \\ &\leq C\Big(E\Big[|\hat{f}_k(y) - f(y)|^2\Big]\Big)^{1/2} \text{ where } C < \infty \end{aligned}$$

For some $y \in \mathbb{R}$ and each n, we have

$$\begin{aligned} |\hat{a}_k(y)| = \Big|\int |x-y|\hat{f}_k(x)dx\Big| &\leq \frac{1}{nh_n}\sum_{t=1}^{n}\int |x-y|\Big|M_k\Big(\frac{x-X_t}{h_n}\Big)\Big|dx \\ &\leq \frac{1}{n}\sum_{t=1}^{n}\int |X_t + h_n\psi - y||M_k(\psi)|d\psi \\ &\leq \frac{1}{n}\sum_{t=1}^{n}|X_t|\int |M_k(\psi)|d\psi + \frac{h_n}{n}\sum_{t=1}^{n}\int |\psi||M_k(\psi)|d\psi + |y|\int |M_k(\psi)|d\psi < C_1 < \infty. \end{aligned}$$

For all $y \in \mathbb{R}$, note that

$$\begin{aligned} &E[|\hat{p}_\alpha(y) - p_\alpha(y)|] \\ &\leq E\Big[|\alpha f^*(y)^{\alpha-1}||[\hat{f}_k(y) - f(y)]||a(y)|\Big] + E\Big[|\alpha f^*(y)^{\alpha-1}||[\hat{f}_k(y) - f(y)]||[\hat{a}_k(y) - a(y)]|\Big] \\ &+ E\Big[|f(y)^\alpha||[\hat{a}_k(y) - a(y)]|\Big] \\ &\leq |a(y)|E\Big[|\alpha f^*(y)^{2\alpha-2}|\Big]^{1/2}E\Big[|\hat{f}_k(y) - f(y)|^2\Big]^{1/2} \\ &+ CE\Big[|\hat{f}_k(y) - f(y)|^2\Big]^{1/2}E\Big[|\hat{a}_k(y) - a(y)|^2\Big]^{1/2} + |f(y)|^\alpha E\Big[|\hat{a}_k(y) - a(y)|^2\Big]^{1/2}. \end{aligned}$$

Therefore, $E[|\hat{p}_\alpha(y) - p_\alpha(y)|] \to 0$ as $n \to \infty$.

$$\begin{aligned} E[\hat{P}_\alpha(\hat{F}_k) - P_\alpha(F)] = E\Big[\int [\hat{p}_\alpha(y) - p_\alpha(y)]dF(y)\Big] + E\Big[\int p_\alpha(y)d[\hat{F}_k(y) - F(y)]\Big] \\ + E\Big[\int [\hat{p}_\alpha(y) - p_\alpha(y)]d[\hat{F}_k(y) - F(y)]\Big] \end{aligned}$$

$E\left[\int p_\alpha(y)d[\hat{F}_k(y) - F(y)]\right] = \int p_\alpha(y)E[\hat{f}_k(y) - f(y)]dy \le O(h_n^r)$. Note that $E\left[(\hat{f}_k(y) - f(y))^2\right] = Bias(\hat{f}_k(y))^2 + Var(\hat{f}_k(y)) = O(h_n^{2r}) + O((nh_n)^{-1})$. Hence, $\left|E\left[\int[\hat{p}_\alpha(y) - p_\alpha(y)]dF(y)\right]\right| \to 0$ as $n \to \infty$.

$$
\begin{aligned}
&E\left[|\hat{p}_\alpha(y) - p_\alpha(y)|^2\right] \\
&= E\left[\left|\alpha f^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)]a(y) + af^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)][\hat{a}_k(y) - a(y)] + f(y)^\alpha[\hat{a}_k(y) - a(y)]\right|^2\right] \\
&\le 3E\left[\left|\alpha f^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)]a(y)\right|^2\right] + 3E\left[\left|\alpha f^*(y)^{\alpha-1}[\hat{f}_k(y) - f(y)][\hat{a}_k(y) - a(y)]\right|^2\right] \\
&+ 3E\left[\left|f(y)^\alpha[\hat{a}_k(y) - a(y)]\right|^2\right] \text{ by using } Cr \text{ inequality ([8] page 141)}
\end{aligned}
$$

Given the result of $\hat{f}_k(x) - f(x) = o_p(1)$ for all $x \in \mathbb{R}$(Theorem 1.2 in Chapter 1) and $\sup_{y\in\mathbb{R}} |f^*(y)| < C < \infty$, we have $E\left[|\hat{p}_\alpha(y) - p_\alpha(y)|^2\right] \to 0$ as $n \to \infty$. Similarly, we have

$$
E\left[\int[\hat{p}_\alpha(y) - p_\alpha(y)]d[\hat{F}_k - F](y)\right] = \int \left\{E\left[(\hat{p}_\alpha(y) - p_\alpha(y))^2\right]\right\}^{1/2}\left\{E\left[(\hat{f}_k(y) - f(y))^2\right]\right\}^{1/2} dy \to 0
$$

as $n \to \infty$. Hence, $Bias(\hat{P}_\alpha(\hat{F}_k)) \to 0$ as $n \to \infty$. □

Table A.2: Bias (B) and Root Mean Squared Error (R) for densities $f_1(x)$ and $f_2(x)$ and $P_\alpha(\tilde{F})$, $\{P(\tilde{F}_k)\}_{k=1}^4$.

f_1	$\alpha = 0.25$		$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 1.0$	
$n = 200$	B	R	B	R	B	R	B	R
$P_\alpha(\tilde{F})$	0.0366	0.0679	0.0535	0.0623	0.0928	0.0945	0.0594	0.0605
$P_\alpha(\hat{F}_1)$	0.0338	0.0735	0.0503	0.0632	0.0911	0.0937	0.0580	0.0596
$P_\alpha(\hat{F}_2)$	0.0202	0.0685	0.0298	0.0503	0.0707	0.0747	0.0411	0.0441
$P_\alpha(\hat{F}_3)$	0.0198	0.0689	0.0259	0.0490	0.0657	0.0704	0.0364	0.0402
$P_\alpha(\hat{F}_4)$	0.0203	0.0693	0.0245	0.0487	0.0636	0.0686	0.0343	0.0385

f_1	$\alpha = 0.25$		$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 1.0$	
$n = 800$	B	R	B	R	B	R	B	R
$P_\alpha(\tilde{F})$	0.0178	0.0363	0.0438	0.0487	0.0382	0.0411	0.0353	0.0370
$P_\alpha(\hat{F}_1)$	0.0097	0.0341	0.0381	0.0441	0.0347	0.0381	0.0328	0.0348
$P_\alpha(\hat{F}_2)$	0.0063	0.0341	0.0265	0.0355	0.0204	0.0269	0.0190	0.0233
$P_\alpha(\hat{F}_3)$	0.0036	0.0340	0.0220	0.0326	0.0153	0.0237	0.0143	0.0201
$P_\alpha(\hat{F}_4)$	0.0032	0.0341	0.0205	0.0318	0.0134	0.0228	0.0123	0.0190

f_2	$\alpha = 0.25$		$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 1.0$	
$n = 200$	B	R	B	R	B	R	B	R
$P_\alpha(\tilde{F})$	0.0462	0.0917	0.0420	0.0576	0.0643	0.0671	0.0671	0.0679
$P_\alpha(\hat{F}_1)$	0.1282	0.1556	0.0942	0.1065	0.0989	0.1028	0.0899	0.0971
$P_\alpha(\hat{F}_2)$	0.0743	0.1249	0.0503	0.0774	0.0637	0.0725	0.0625	0.0663
$P_\alpha(\hat{F}_3)$	0.0513	0.1168	0.0319	0.0701	0.0493	0.0619	0.0515	0.0569
$P_\alpha(\hat{F}_4)$	0.0390	0.1140	0.0218	0.0680	0.0414	0.0568	0.0455	0.0520

f_2	$\alpha = 0.25$		$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 1.0$	
$n = 800$	B	R	B	R	B	R	B	R
$P_\alpha(\hat{F}_1)$	0.0191	0.0525	0.0208	0.0354	0.0266	0.0318	0.0217	0.0247
$P_\alpha(\hat{F}_2)$	-0.0034	0.0501	-0.0032	0.0300	0.0041	0.0192	0.0024	0.0131
$P_\alpha(\hat{F}_3)$	-0.0077	0.0508	-0.0087	0.0314	-0.0014	0.0192	-0.0026	0.0135
$P_\alpha(\hat{F}_4)$	-0.0076	0.0509	-0.0097	0.0319	-0.0029	0.0195	-0.0041	0.0140