DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 14-11

A Class of Local Constant Kernel Estimators for a Regression in a Besov Space

Na Kyeong Lee
University of Colorado Boulder

November 6, 2014
Revised November 10, 2014

Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

A CLASS OF LOCAL CONSTANT KERNEL ESTIMATORS FOR A REGRESSION IN A BESOV SPACE *

NA KYEONG LEE †

University of Colorado, Boulder

November 2014

Abstract. The use of higher order kernels is a well-known method for bias reduction of density and regression estimators. This method of bias reduction has the disadvantage of potential negativity of the underlying estimated density. To avoid this, Mynbaev and Martins-Filho (2010) 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, Mynbaev and Martins-Filho (2014) 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.

Keywords and phrases. bias reduction, local polynomial estimation, asymptotic normality

JEL classifications. C13; C14.

AMS-MS classifications. 62G07, 62G08, 62G20.

---

*I am deeply grateful to my advisor Carlos Martins-Filho. I would like to thank Kairat Mynbaev and Xiaodong Liu for helpful comments and suggestions. All errors are my own.

†Department of Economics, University of Colorado, Boulder, CO 80309-0256 (Email:na.lee@colorado.edu; website: https://sites.google.com/site/econakyo)

1 Introduction

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

$$Y = m(X) + u \tag{1}$$

where m is a real valued function with $E[u|X = x] = 0$ and $Var[u\ |X = x] = \sigma^2$. We call a kernel any function K on $\mathbb{R}$ such that $\int_{-\infty}^{\infty} K(t)dt = 1$. Nadaraya (1964) and Watson (1964) 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

$$\hat{m}(x) = \frac{\frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) Y_t}{\frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)} \quad \text{where} \quad \hat{f}(x) = \frac{1}{nh_n}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right), \tag{2}$$

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 $j = 1, 2, \cdots, s-1$ 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 (Gasser et al. (1985), Schucany (1989)). However, this approach is inconvenient since the condition that the kernel density estimate $\hat{f}$ 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 the design-adaptive regression (Fan (1992)), data sharpening methods (Choi et al. (2000)), iterative method (Racine (2001)) and parametrically guided nonparametric estimation (Glad (1998), Martins-Filho et al. (2008)) 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.

Mynbaev and Martins-Filho (2010) 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 $x \in \mathbb{R}$,

$$M_k(x) = -\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{x}{s}\right) \tag{3}$$

where the binomial coefficients $C_{2k}^{N}=\frac{(2k)!}{(2k-N)!N!}$, $N=0,\cdots,2k$, $k\in\{1,2,\cdots\}$ and $c_{k,s}=(-1)^{s+k}C_{2k}^{s+k}$, $s=-k,\cdots,k$. Mynbaev and Martins-Filho (2014) obtain new results on nonparametric prediction by relaxing the conditions in Carroll et al. (2009)[1] and allowing fractional smoothness of the density. In this paper, by extending the approaches of Mynbaev and Martins-Filho (2010) and Mynbaev and Martins-Filho (2014) we propose a new family of local constant estimators. Based on the kernels M_k in (3) we define a class of local constant estimators indexed by k such that

$$\hat{m}_k(x)=\frac{\sum_{t=1}^{n}M_k\left(\frac{X_t-x}{h_n}\right)Y_t}{\sum_{t=1}^{n}M_k\left(\frac{X_t-x}{h_n}\right)}\quad. \tag{4}$$

The estimators $\hat{m}_k(x)$ form a general class of local constant estimators. When $k=1$ and a seed kernel K is symmetric, our estimator $\hat{m}_1(x)$ coincides with $\hat{m}(x)$ which is given by (2). That is, the Nadaraya-Watson estimator $\hat{m}$ can be considered as a special case of our estimators $\hat{m}_k$.

Throughout this paper, we assume that the true regression m belongs to a Besov space $\mathcal{B}_{\infty,q}^{r}$ 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 $m\in\mathcal{B}_{\infty,q}^{r}$ where $l<r$, that is, $\mathcal{C}^{l}(\mathbb{R})\subseteq\mathcal{B}_{\infty,q}^{r}$ 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 $l\le s$.

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 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 literatures (Devroye (1978), Collomb (1981), Mack and Silverman (1982)) by imposing 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

[1] Mynbaev and Martins-Filho (2014) replaced conditions (4.2) and (4.3) from Carroll et al. (2009) with their lighter assumptions 2.1 and 2.2.

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 the absolute average root mean squared error when the kernels proposed in Mynbaev and Martins-Filho (2010) are used.

The remainder of the paper is organized as follows. Section 2 provides a brief discussion of Besov spaces and discusses properties of the density estimator. In section 3, we provide the main asymptotic properties of local constant estimators. Section 4 contains a small Monte Carlo study that gives some evidence on the finite sample performance of our estimators. Section 5 summarizes the findings. The appendices contain all proofs, tables and figures that summarize the Monte Carlo simulation.

2 A Nonparametric density estimator

2.1 Finite differences and Besov Spaces

In this section, we define the class of density estimators $\{\hat{f}_k\}_{k=1,2,\cdots}$ using the family of kernels $\{M_k\}_{k=1,2,\cdots}$ introduced by Mynbaev and Martins-Filho (2010). We need a series of definitions that support the construction of the class. The properties of nonparametric density estimators are traditionally obtained by assumption 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. A s-th order forward difference is defined by

$$\tilde{\Delta}_h^s f(x) = \sum_{j=0}^{s} (-1)^{s-j} C_s^j f(x+jh) \;\; \text{where } s = 1, 2, \cdots. \text{ and for } h \in \mathbb{R}. \tag{5}$$

Lemma 1 relates forward differences to differentiability by means of a recursion.

Lemma 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*

$$\tilde{\Delta}_h^s f(x) = \int_0^h \cdots \int_0^h \tilde{\Delta}_h^{s-l} \mathcal{D}^l f\left(x + \sum_{i=1}^{l} u_i\right) \prod_{i=1}^{l} du_i \qquad \text{where } l = 1, 2, \cdots, s. \tag{6}$$

When we consider forward even-order difference, (5) can be written as

$$\tilde{\Delta}_h^{2k} f(x) = \sum_{|s|=1}^{k} c_{k,s} f(x + kh + sh) \tag{7}$$

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 $s = 2k$, $\tilde{\Delta}_h^{2k} f(x) = \sum_{j=0}^{2k} (-1)^{2k-j} C_{2k}^j f(x + jh) = \sum_{|s|=1}^{k} (-1)^{s+k} C_{2k}^{s+k} f(x + kh + sh)$.

Next, we introduce Besov spaces $\mathcal{B}_{p,q}^r(\mathbb{R})$ where $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $r > 0$, and the norm in $\mathcal{B}_{p,q}^r(\mathbb{R})$ is 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 $2k > r$ (Triebel (1985), Mynbaev and Martins-Filho (2014)). When $p = \infty$ and/or $q = \infty$, the integral(s) is (are) replaced by supremum. $\mathcal{C}^0(\mathbb{R})$ is defined as the collection of all real-valued, bounded and uniformly continuous functions in $\mathbb{R}$, equipped with the norm $||f|\mathcal{C}^0(\mathbb{R})|| = \sup_{x \in \mathbb{R}} |f(x)|$. [2] The following lemma shows that the class $\mathcal{C}^l(\mathbb{R})$ is a subset of $B_{\infty,q}^r$ whenever $l < r$.

Lemma 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* $0 \leq q \leq \infty$. *For* $r > l$, *we have*

$$\mathcal{C}^l(\mathbb{R}) \subseteq \mathcal{B}_{\infty,q}^r(\mathbb{R}). \tag{8}$$

A full description of the relationships between $\mathcal{C}^l(\mathbb{R})$ and a Besov space $\mathcal{B}_{p,q}^r$ can be found in Besov et al. (1978). Since,

$$M_k(x) = -\frac{1}{c_{k,s}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{x}{s}\right) \tag{9}$$

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 equation is true.

$$-\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} c_{k,s} = 1 \quad \text{or} \quad \sum_{s=1}^{k} \lambda_{k,s} = \frac{1}{2} \tag{10}$$

Equation (10) and $\int K(\psi) d\psi = 1$ imply that

$$\int M_k(\psi) d\psi = \sum_{s=1}^{k} \frac{\lambda_{k,s}}{s} \left[\int K\left(\frac{\psi}{s}\right) d\psi + \int K\left(-\frac{\psi}{s}\right) d\psi \right] = 1,$$

[2] See Triebel (2010).

which establishes that every M_k. The kernel M_k defines a new family of density estimators indexed by k as follows,

$$\hat{f}_k(x) = \frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right) \tag{11}$$

where h_n is a bandwidth sequence tending to zero as $n \to \infty$.[3] When $k = 1$ and K is symmetric, the density estimator in (11) coincides with the Rosenblatt-Parzen density estimator. Since the kernel $M_k(x)$ is symmetric, by using forward even-order differences (7), for a function f we have

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

It is easy to verify that $\tilde{\Delta}_h^{2k} f(x) = \Delta_h^{2k}[f(x + kh)]$ (Mynbaev and Martins-Filho (2014)). Hence, we use centered even-order difference for a smoothness characteristic, and we have

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

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) $f \in \mathcal{B}^r_{\infty,q}$ 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) $\sup_{x\in\mathbb{R}} |K(x)| < M < \infty$; (5) $|K(x) - K(x')| < c|x - x'|$ for some $c < \infty$ and $x \neq x'$, $x, x' \in \mathbb{R}$.

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

Theorem 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*

[3] Mynbaev and Martins-Filho (2010) defined this alternative family of density estimator.

$k = 1, 2, \cdots,$ *we have*

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

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

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. Given ASSUMPTION 3 we have $Bias(\hat{f}_k(x)) \to 0$ 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 2 *Suppose* ASSUMPTIONS 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 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*

$$\sup_{x \in \mathcal{G}} |\hat{f}_k(x) - f(x)| = O_p\left(\left(\frac{\log n}{n h_n}\right)^{1/2} + h_n^r\right). \tag{12}$$

Uniform consistency of the density estimator requires $\left(\frac{\log n}{n h_n}\right) \to 0$ as $n \to \infty$. From (12), the order of $\hat{f}_k$ is similar to that attained by Rosenblatt density estimator with a kernel of order r. We achieve much faster uniform convergence rate by imposing the less restrictive assumption $f \in B^r_{\infty,q}$. 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 4 *Suppose* ASSUMPTION 1*,* ASSUMPTION 2(1)-(2)*,* ASSUMPTION 3 *and* ASSUMPTION 4(1)-(4)*. Then for all* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots,$ *we have*

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

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

$$\sqrt{nh_n}\left(\hat{f}_k(x) - f(x)\right) \xrightarrow{d} \mathcal{N}\left(0, f(x)\int M_k^2(\psi)d\psi\right). \tag{13}$$

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 attain the asymptotic normality of $\hat{f}_k$, 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, 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 only 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 (13).

3 Local Constant Estimator

In this section, we establish the asymptotic normality of the estimators $\hat{m}_k$ for $k = 1, 2, \cdots$. We assume that the conditional density of Y_t given $X_t = x$ exists and is denoted by $f_{Y|X}(y) = \frac{f_{Y,X}(y,x)}{f(x)}$ where $f_{Y,X}$ denotes the density of $Z = (Y, X)$ 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); (2) $m \in \mathcal{C}^0(\mathbb{R})$.

$\mathcal{B}^{\rho}_{\infty,\infty}(\mathbb{R})$ is a Zygmund space $\mathcal{Z}^{\rho}(\mathbb{R})$[4]. By Corollary 2.8.2 (i) in Triebel (1985), the multiplication by a function $m \in \mathcal{Z}^{\rho}(\mathbb{R})$ is bounded in $\mathcal{B}^r_{p,q}$ if $\rho > r$, that is

$$||mf||_{\mathcal{B}^r_{p,q}} \leq c||m||_{\mathcal{Z}^{\rho}}||f||_{\mathcal{B}^r_{p,q}}. \tag{14}$$

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 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

[4]For a more detailed explanation, see Theorem on page 90 in Triebel (1985).

literature for (2). We make the following additional assumption.

ASSUMPTION 6: $E[|Y - m(X)|^{2+\delta}|X] < \infty$ for $\delta > 0$ and $Var(Y|X = x) = \sigma^2 < \infty$.

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 denoted by (9) 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 (4). To obtain an approximation to the finite sample properties of $\hat{m}_k$, we rewrite

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

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 (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 $r < \rho$ (Triebel (1985)). The class of local constant estimators in equation (4) derives from a nonparametric density estimator. In the previous section, we already considered density estimation, so the only step needed to investigate the properties of $\hat{m}_k$ is to consider the properties of $\hat{g}_k(x)$.

Theorem 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*

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

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

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

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 $2k > r$. Since $f \in \mathcal{C}^0(\mathbb{R})$, $m \in \mathcal{C}^0(\mathbb{R})$, $\int |K(\psi)|d\psi < \infty$ and $\sup_{x\in\mathbb{R}} |K(\psi)| < \infty$, $\int M_k^2(\psi)f(x + h_n\psi)d\psi = O(1)$, $\int M_k^2(\psi)m^2(x + h_n\psi)f(x + h_n\psi)d\psi = O(1)$ and $\int M_k(\psi)m(x + h\psi)f(x + h_n\psi)d\psi = O(1)$. Given that $nh_n \to \infty$ as $n \to \infty$ and from $Var[\hat{g}_k(x)] = \frac{\sigma^2}{nh_n}\int M_k^2(\psi)f(x + h_n\psi)d\psi - \frac{1}{n}\left\{\int M_k(\psi)m(x + h_n\psi)f(x + h_n\psi)d\psi\right\}^2 + \frac{1}{nh_n}\int M_k^2(\psi)m^2(x + h_n\psi)f(x + h_n\psi)d\psi$, we have $Var[\hat{g}_k(x)] \longrightarrow 0$ as $n \to \infty$. Hence $\hat{g}_k(x) - g(x) = o_p(1)$.

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

$$\sup_{x \in \mathcal{G}} |\hat{g}_k(x) - E[\hat{g}_k(x)] = O_p\left(\left(\frac{\log n}{nh_n}\right)^{1/2}\right) \tag{15}$$

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

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

Theorem 7 *Suppose* Assumption 1-3*,* Assumption 4(1)-(4)*,* Assumption 6 *hold. For* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots,$ *we have*

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

Given $\hat{f}_k(x)$ such that $\hat{f}_k(x) = f(x) + o_p(1)$ in Theorem 2, we have

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

Given the results on $\hat{g}_k$ and $\hat{f}_k$, we obtain following properties for $\hat{m}_k(x)$. First, Theorem 8 gives the order of the bias for $\hat{m}_k$.

Theorem 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* $1 \leq q \leq \infty$*. 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 bias for our estimator is 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 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 by avoiding the nonnegative density estimator and by imposing less restrictive conditions such that $f \in B_{\infty,q}^r$ and $m \in B_{\infty,\infty}^\rho$ where $\rho > r$.

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

Theorem 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 (Devroye (1978), Collomb (1981), Mack and Silverman (1982)) by avoiding higher-order conditions on the kernel and imposing less restrictive conditions.

We now give sufficient condition for asymptotic normality of $\hat{m}_k(x)$ under suitable centering and normalization.

Theorem 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* $n \to \infty$.

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

For the local constant estimator the normalizing factor will be $(nh_n)^{1/2}$ and we will work with the decomposition $(nh_n)^{1/2}[\hat{m}_k(x) - m(x)] = (nh_n)^{1/2}[\hat{m}_k(x) - E(\hat{m}_k(x)|X_t)] + (nh_n)^{1/2}[E(\hat{m}_k(x)|X_t) - m(x)]$. The first term in the decomposition is asymptotically normal 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 $n \to \infty$. Consistency follows from the fact that $(nh_n)^{1/2}(\hat{m}_k(x) - m(x))$ has a limiting distribution. 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.

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 kernel 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 Gaussian kernel. We consider following data generating processes (DGPs),

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

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

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

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

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

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

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, evaluate the regression of 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 obtained by using a Gaussian seed kernel. For values of x where the denominator of the local constant estimator $\hat{f}_k(x)$ closes to zero, the local constant estimator at x might not be 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 1 provides average absolute bias (B), average variance (V) and average root MSE (R) for each estimator considered for $n = 400, 1000$, respectively.

Table 1, Figure 1 and 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 value. Finally, we observe that for DGP1, DGP2 and DGP3, 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).

5 Summary

The use of higher order kernels is a well-known method for bias reduction of density and regression estimators. This method of bias reduction has the disadvantage of potential negativity of the underlying estimated density. To avoid this, Mynbaev and Martins-Filho (2010) 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, Mynbaev and Martins-Filho (2014) 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.

Appendix 1 - Proofs

Lemma 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 $\tilde{\Delta}_h^s f(x)$ (5), 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 $\mathcal{D}^s$ are classical derivatives. First, we need to verify the case : $(i)\quad s=1$.

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

Consider $(ii)\quad s=2$.

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

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) $s = k+1$.

$$\begin{aligned}&\tilde{\Delta}_h^{k+1} f(x) = \tilde{\Delta}_h^1[\tilde{\Delta}_h^k f(x)] = \tilde{\Delta}_h^1 \int_0^h \tilde{\Delta}_h^{k-1} \mathcal{D} f(x+u_1) du_1 = \int_0^h \tilde{\Delta}_h^{k-1} \mathcal{D} f(x+u_1+h) - \tilde{\Delta}_h^{k-1} \mathcal{D} f(x+u_1) du_1 \\ &= \int_0^h \int_0^h \tilde{\Delta}_h^{k-1} \mathcal{D}^2 f(x+u_1+u_2) du_1 du_2 = \int_0^h \int_0^h \tilde{\Delta}_h^1[\tilde{\Delta}_h^{k-2} \mathcal{D}^2 f(x+u_1+u_2)] du_1 du_2 \\ &= \int_0^h \int_0^h [\tilde{\Delta}_h^{k-2} \mathcal{D}^2 f(x+u_1+u_2+h) - \tilde{\Delta}_h^{k-2} \mathcal{D}^2 f(x+u_1+u_2+h)] du_1 du_2 \\ &= \int_0^h \int_0^h \int_0^h [\tilde{\Delta}_h^{k-2} \mathcal{D}^3 f(x+u_1+u_2+u_3)] du_1 du_2 du_3 = \cdots = \int_0^h \cdots \int_0^h \mathcal{D}^{k+1} f\left(x + \sum_{i=1}^{k+1} u_i\right) \prod_{i=1}^{k+1} du_i\end{aligned}$$

□

Lemma 2

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

$$\int |h|^{-rq} ||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{|h|} = \int_0^\infty h^{-rq} ||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h} + \int_{-\infty}^0 (-h)^{-rq} ||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{(-h)}$$
$$= \int_0^\infty h^{-rq} ||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h} + \int_0^\infty \gamma^{-rq} ||\tilde{\Delta}_\gamma^s f(x)||_\infty^q \frac{d\gamma}{\gamma} \quad \text{by change of variable by letting } -h = \gamma$$
$$= \int_0^1 h^{-rq} ||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h} + \int_1^\infty h^{-rq} ||\tilde{\Delta}_h^s f(x)||_\infty^q \frac{dh}{h}$$
$$+ \int_0^1 \gamma^{-rq} ||\tilde{\Delta}_{-\gamma}^s f(x)||_\infty^q \frac{d\gamma}{\gamma} + \int_1^\infty \gamma^{-rq} ||\tilde{\Delta}_{-\gamma}^s f(x)||_\infty^q \frac{d\gamma}{\gamma}$$
$$= \int_0^1 h^{-rq} \left[||\tilde{\Delta}_h^s f(x)||_\infty^q + ||\tilde{\Delta}_{-h}^s f(x)||_\infty^q \right] \frac{dh}{h} + \int_1^\infty h^{-rq} \left[||\tilde{\Delta}_h^s f(x)||_\infty^q + ||\tilde{\Delta}_{-h}^s f(x)||_\infty^q \right] \frac{dh}{h}$$
$$= \int_0^1 h^{-rq} \left[\left|\left| \int_0^h \cdots \int_0^h \mathcal{D}^s f(x + \sum_{i=1}^s u_i) \prod_{i=1}^s du_i \right|\right|_\infty^q + \left|\left| \int_0^{-h} \cdots \int_0^{-h} \mathcal{D}^s f(x + \sum_{i=1}^s u_i) \prod_{i=1}^s du_i \right|\right|_\infty^q \right] \frac{dh}{h}$$
$$+ \int_1^\infty h^{-rq} \left[\left|\left| \int_0^h \cdots \int_0^h \tilde{\Delta}_h^{s-l} \mathcal{D}^l f(x + \sum_{i=1}^l u_i) \prod_{i=1}^l du_i \right|\right|_\infty^q \right] \frac{dh}{h}$$
$$+ \int_1^\infty h^{-rq} \left[\left|\left| \int_0^{-h} \cdots \int_0^{-h} \tilde{\Delta}_{-h}^{s-l} \mathcal{D}^l f(x + \sum_{i=1}^l u_i) \prod_{i=1}^l du_i \right|\right|_\infty^q \right] \frac{dh}{h}$$
$$\leq \int_0^1 h^{-rq} \left\{ \sup_{x\in\mathbb{R}} \left[\int_0^h \cdots \int_0^h \left| \mathcal{D}^s f(x + \sum_{i=1}^s u_i) \right| \prod_{i=1}^s du_i \right]^q \right\} \frac{dh}{h}$$
$$+ \int_0^1 h^{-rq} \left\{ \sup_{x\in\mathbb{R}} \left[\int_0^{-h} \cdots \int_0^{-h} \left| \mathcal{D}^s f(x + \sum_{i=1}^s u_i) \right| \prod_{i=1}^s du_i \right]^q \right\} \frac{dh}{h}$$
$$+ \int_1^\infty h^{-rq} \left\{ \sup_{x\in\mathbb{R}} \left[\int_0^h \cdots \int_0^h \left| \tilde{\Delta}_h^{s-l} \mathcal{D}^l f(x + \sum_{i=1}^l u_i) \right| \prod_{i=1}^l du_i \right]^q \right\} \frac{dh}{h}$$
$$+ \int_1^\infty h^{-rq} \left\{ \sup_{x\in\mathbb{R}} \left[\int_0^{-h} \cdots \int_0^{-h} \left| \tilde{\Delta}_{-h}^{s-l} \mathcal{D}^l f(x + \sum_{i=1}^l u_i) \right| \prod_{i=1}^l du_i \right]^q \right\} \frac{dh}{h}$$
$$\leq \sup_{x\in\mathbb{R}} |\mathcal{D}^s f(x)|^q \int_0^1 h^{-rq} \left\{ \left[\int_0^h \cdots \int_0^h \prod_{i=1}^s du_i \right]^q + \left[\int_0^{-h} \cdots \int_0^{-h} \prod_{i=1}^s du_i \right]^q \right\} \frac{dh}{h}$$
$$+ c_1 \sup_{x\in\mathbb{R}} |\mathcal{D}^l f(x)|^q \int_1^\infty h^{-rq} \left[\int_0^h \cdots \int_0^h \prod_{i=1}^l du_i \right]^q \frac{dh}{h}$$
$$+ c_2 \sup_{x\in\mathbb{R}} |\mathcal{D}^l f(x)|^q \int_1^\infty h^{-rq} \left[\int_0^{-h} \cdots \int_0^{-h} \prod_{i=1}^l du_i \right]^q \frac{dh}{h}$$

$$= \sup_{x\in\mathbb{R}} |\mathcal{D}^s f(x)|^q \int_0^1 \left[h^{-rq+sq-1} + h^{-rq-1}(-h)^{sq} \right] dh$$
$$+ \sup_{x\in\mathbb{R}} |\mathcal{D}^l f(x)|^q \int_1^\infty \left[c_1 h^{-rq+lq-1} + c_2 h^{-rq-1}(-h)^{lq} \right] dh$$
$$= \sup_{x\in\mathbb{R}} |\mathcal{D}^s f(x)|^q \left[\frac{1}{(s-r)q} \right] (1+(-1)^{sq}) + \sup_{x\in\mathbb{R}} |\mathcal{D}^l f(x)| \left[\frac{1}{(r-l)q} \right] (c_1 + c_2(-1)^{lq})$$

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

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

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 $||f||_{\mathcal{B}^r_{\infty,q}} \leq C||f||_{\mathcal{C}^l}$. That is, $\mathcal{C}^l(\mathbb{R}) \subseteq \mathcal{B}^r_{\infty,q}(\mathbb{R})$ where $l < r$. □

Theorem 1

Proof. (a)

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

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

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

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

(b) We can proceed the order of $Bias(\hat{f}_k(x))$ using the result of (a).

Given that $\left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q'} d\psi \right]^{1/q'} < \infty$, we have

$$|Bias(\hat{f}_k(x))| = |E(\hat{f}_k(x)) - f(x)| = \left| \int -\frac{1}{c_{k,0}} K(\psi)\Delta^{2k}_{h_n\psi} f(x)d\psi \right|$$
$$\leq \left| -\frac{1}{c_{k,0}} \right| \left| \int K(\psi)|h_n\psi|^{r+1/q} \frac{1}{|h_n\psi|^{r+1/q}} \Delta^{2k}_{h\psi} f(x)d\psi \right|$$
$$\leq \left| -\frac{1}{c_{k,0}} \right| \left[\int \left\{ |K(\psi)||h_n\psi|^{r+1/q} \right\}^{q'} d\psi \right]^{1/q'} \left[\int \left\{ \frac{sup_{x\in\mathbb{R}} |\Delta^{2k}_{h_n\psi} f(x)|}{|h_n\psi|^{r+1/q}} \right\}^q d\psi \right]^{1/q} \quad \text{by Holder's inequality}$$
$$= \left| -\frac{1}{c_{k,0}} \right| \left[\int \left\{ |K(\psi)||h_n\psi|^{r+1/q} \right\}^{q'} d\psi \right]^{1/q'} \left[\int \left\{ \frac{sup_{x\in\mathbb{R}} |\Delta^{2k}_{h_n\psi} f(x)|}{|h_n\psi|^{r}} \right\}^q \frac{1}{|h_n\psi|} d\psi \right]^{1/q} \quad \text{by letting } h_n\psi \equiv t$$

$$= \left|-\frac{1}{c_{k,0}}\right| \left[\int |K(\psi)|^{q'} |h_n\psi|^{(r+1/q)q'} d\psi\right]^{1/q'} \left[\int \left\{\frac{sup_{x\in\mathbb{R}}|\Delta_t^{2k} f(x)|}{|t|^r}\right\}^q \frac{1}{|t|} h_n^{-1} dt\right]^{1/q}$$
$$\leq h_n^r \left|-\frac{1}{c_{k,0}}\right| \left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q'} d\psi\right]^{1/q'} ||f||_{\mathcal{B}^r_{\infty,q}} = O(h_n^r)$$

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

Theorem 2

Proof.

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

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

$$\int M_k^2(\psi)f(x+h\psi)d\psi = \int M_k^2(\psi)\left[f(x+h\psi)-f(x)\right]d\psi + \int M_k^2(\psi)f(x)d\psi$$

$$\leq \int_{|h\psi|\leq\delta} M_k^2(\psi)|f(x+h\psi)-f(x)|d\psi + \int_{|h\psi|>\delta} M_k^2(\psi)|f(x+h\psi)-f(x)|d\psi + f(x)\int M_k^2(\psi)d\psi$$
$$\leq \sup_{|y|\leq\delta, x\in\mathbb{R}} |f(x+y)-f(x)| \int M_k^2(\psi)d\psi + 2\sup_{x\in\mathbb{R}}|f(x)| \int_{|h\psi|>\delta} M_k^2(\psi)d\psi + \sup_{x\in\mathbb{R}}|f(x)|\int M_k^2(\psi)d\psi$$
$$\text{since } f \in \mathcal{C}^0(\mathbb{R}) < \infty \tag{16}$$

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 (16), $\hat{f}_k(x) - f(x) = o_p(1)$ for all $x \in \mathbb{R}$. □

Theorem 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 $\mathbb{R}$ that is, $\mathcal{G} \subseteq \mathbb{R}$. The collection $F = \{B(x,r) : x \in \mathcal{G}, r > 0\}$ is an open covering of $\mathcal{G}$. By the Heine-Borel theorem, 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

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

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

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

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

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

Let d_n be a sequence of stochastic variables. If $\forall \epsilon > 0$ there exists $M_\epsilon > 0$ and a non stochastic sequence $\{a_n\}$ such that $P\left[\frac{|d_n|}{a_n} > M_\epsilon\right] < \epsilon$ for all n. We write $d_n = O_p(a_n)$(Mann and Wald (1943) and Davidson (1994)).

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

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

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

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

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

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

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

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

Then, we have $|W_{tn}| \leq \frac{1}{h_n}B_kC_1$ where $C_1 = 2\left|-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\right|$. Next we consider $Var(W_{tn})$. Let $\sigma_{tn}^2 = Var(W_{tn})$. Given that $E[W_{tn}] = 0$, we have

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

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

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

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 (Bennett (1962)), we have

$$\begin{aligned}P\left[|\hat{f}_k(x_\tau) - E[\hat{f}_k(x_\tau)]| > a_nM_{n,\epsilon}\right] &= P\left[\left|\frac{1}{n}\sum_{t=1}^{n}W_{tn}\right| > a_nM_{n,\epsilon}\right] \\ &\leq 2\exp\left\{-\frac{na_n^2M_{n,\epsilon}^2}{2\frac{1}{n}\sum_{t=1}^{n}Var(W_{tn}) + \frac{2}{3}\frac{B_k}{h_n}C_1a_nM_{n,\epsilon}}\right\} \quad \text{where } C_1 = 2\left|-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\right|\end{aligned}$$

$$= 2\exp\left\{-\frac{n^2a_n^2M_{n,\epsilon}^2/nE(W_{tn}^2)}{2+\frac{2}{3}\frac{B_k}{h_n}C_1\frac{na_nM_{n,\epsilon}}{nE(W_{tn}^2)}}\right\} \quad \text{by Assumption 1 and } Var(W_{tn}) = E(W_{tn}^2) \tag{17}$$

$$= 2\exp\left\{-\frac{na_n^2M_{n,\epsilon}^2/E(W_{tn}^2)}{2\frac{E[W_{tn}^2]}{E[W_{tn}^2]}+\frac{2}{3}\frac{B_k}{h_n}C_1\frac{a_nM_{n,\epsilon}}{E[W_{tn}^2]}}\right\} \tag{18}$$

$$= 2\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2h_nE[W_{tn}^2]+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}}\right\} \tag{19}$$

$$P\left[\frac{1}{a_n}\max_{1\le\tau\le m}|\hat{f}_k(x_\tau)-E[\hat{f}_k(x_\tau)]|>M_{n,\epsilon}\right] \le \sum_{\tau=1}^{m}2\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2h_nE[W_{tn}^2]+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}}\right\}$$

$$\le 2m\max_{1\le\tau\le m}\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2h_nE[W_{tn}^2]+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}}\right\}$$

$$\le 2m\max_{1\le\tau\le m}\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2g_n(x_\tau)+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}}\right\} \quad \text{where } g_n(x_\tau)=h_nE[W_{tn}^2]$$

$$\le 2m\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2g_n(x^m)+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}}\right\} \tag{20}$$

where x^m corresponds to the point of the given function such that $\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2h_nE[W_{tn}^2]+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}}\right\}$ which the function $\exp\{\cdot\}$ 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 $a_nM_{n,\epsilon}=a_n\left(M_\epsilon-\frac{2cr}{a_nh_n^2}\right)=a_nM_\epsilon-\frac{2cr}{h_n^2}$. Hence,

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

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

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

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

Hence, if $\left(\frac{\log n}{nh_n}\right)\to 0$ as $n\to\infty$, then $a_nM_{n,\epsilon}\to 0$. From equation (20),

$$2m\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2g_n(x^m)+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}}\right\}=2m\exp\left\{-\frac{\triangle_n\log n}{v_n}\right\}=2m\exp\left\{\log n^{-\triangle_n/v_n}\right\}=2mn^{-\triangle_n/v_n}$$

Hence, $P\left[\frac{1}{a_n}\max_{1\leq\tau\leq m}|\hat{f}_k(x_\tau)-E[\hat{f}_k(x_\tau)]|>M_{n,\epsilon}\right]\leq 2mn^{-\triangle_n/v_n}$. The volume of $B(x_\tau,r)$ for $x\in\mathbb{R}$ is $2r=2\left(\frac{h_n^3}{n}\right)^{1/2}=2r_n$. Since $F'=\{B(x_\tau,r):x_\tau\in\mathcal{G},r>0,\tau=1,2,\cdots,m,\quad \text{where } m \text{ is finite}\}$ 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 $x\in\mathbb{R}$, $2mr_n=2m\left(\frac{h_n^3}{n}\right)^{1/2}\leq 2r_0$ which implies that $m\leq r_0r_n^{-1}=r_0\left(\frac{h_n^3}{n}\right)^{-1/2}$. Hence,

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

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 $v_n=2g_n(x^m)+\frac{2}{3}B_kC_1a_nM_{n,\epsilon}$, $\triangle_n\to M_\epsilon^2$, $g_n(x^m)\to f(x^m)\int M_k^2(\psi)d\psi$ 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 $\frac{M_\epsilon^2}{2f(x^m)\int M_k^2(\psi)d\psi}-1\geq 1$ or $\frac{M_\epsilon^2}{2f(x^m)\int M_k^2(\psi)d\psi}\geq 2$ to obtain $n^{\frac{\triangle_n}{v_n}-1}h_n\to\infty$.

Now,

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

□

Theorem 4

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

Let $Z_{nt}=\frac{1}{nh_n}M_k\left(\frac{X_t-x}{h_n}\right)$, $E[Z_{nt}]=\mu_n$ and $S_n^2=\sum_{t=1}^nE[Z_{nt}-\mu_n]^2$. We have

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

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

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

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

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

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

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

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

Therefore,

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

We have

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

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

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

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 $\frac{1}{(nh)^{\delta/2}} \to 0$. Hence, $\sqrt{nh}\left(\hat{f}_k(x) - E[\hat{f}_k(x)]\right) \xrightarrow{d} \mathcal{N}\left(0, f(x)\int M_k^2(\psi)d\psi\right)$. In practice we are interested in the distribution of $\sqrt{nh_n}(\hat{f}_k(x) - f(x))$ instead of $\sqrt{nh}\left(\hat{f}_k(x) - E[\hat{f}_k(x)]\right)$.

$$\begin{aligned}
\sqrt{nh_n}\left(\hat{f}_k(x) - f(x)\right) &= \sqrt{nh_n}\left(\hat{f}_k(x) - E[\hat{f}_k(x)]\right) + \sqrt{nh_n}\left(E[\hat{f}_k(x)] - f(x)\right) \\
&= \sqrt{nh_n}O(h_n^r) + \sqrt{nh_n}\left(\hat{f}_k(x) - E[\hat{f}_k(x)]\right) \quad \text{from Theorem 1(a)}
\end{aligned}$$

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

□

Theorem 5

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

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

Then, given Assumption 1 we have

$$\begin{aligned}
E[\hat{g}_k(x)] &= \int \frac{1}{nh_n}\sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right) m(X_t) f(X_t) dX_t \\
&= \left(-\frac{1}{c_{k,0}}\right)\int \sum_{|s|=1}^{k} c_{k,s} K(\psi) m(x + sh_n\psi) f(x + sh_n\psi) d\psi.
\end{aligned}$$

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

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

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 $Bias(\hat{g}_k(x))$.

$$
\begin{aligned}
|Bias(\hat{g}_k(x))| &= \left| \left(-\frac{1}{c_{k,0}}\right) \int \sum_{|s|=0}^{k} c_{k,s}\ K(\psi) m(x+sh_n\psi) f(x+sh_n\psi) d\psi \right| \\
&\leq C \left| \int K(\psi) \Delta_{h_n\psi}^{2k}\ g(x) d\psi \right| = C \left| \int K(\psi) |h\psi|^{r+1/q}\ \frac{\Delta_{h_n\psi}^{2k}\ g(x)}{|h_n\psi|^{r+1/q}}\ d\psi \right| \\
&\leq C \left\{ \int \left[|K(\psi)||h_n\psi|^{(r+1/q)} \right]^{q'} d\psi \right\}^{1/q'} \left\{ \int \left[\frac{\sup_{x\in\mathbb{R}}\ |\Delta_{h_n\psi}^{2k}\ g(x)|}{|h_n\psi|^{r+1/q}} \right]^q d\psi \right\}^{1/q} \\
&= C \left[\int |K(\psi)|^{q'} |h_n\psi|^{(r+1/q)q} d\psi \right]^{1/q'} \left\{ \int \left[\frac{\sup |\Delta_t^{2k} g(x)|}{|t|^r} \right]^q \frac{1}{|t|} h_n^{-1} d\psi \right\}^{1/q} \\
&= h_n^r \left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q} d\psi \right]^{1/q'} ||g||_{\mathcal{B}_{\infty,q}^r} = O(h_n^r)
\end{aligned}
$$

by ASSUMPTION 2(1), ASSUMPTION 5(1) and $\left[\int |K(\psi)|^{q'} |\psi|^{(r+1/q)q} d\psi\right]^{1/q'} < \infty$ where $1/q + 1/q' = 1$ for $1 \leq q \leq \infty$. From the result of (Triebel (1985)), we know that $||g||_{\mathcal{B}_{\infty,q}^r} \leq C||m||_{\mathcal{Z}^\rho}||f||_{\mathcal{B}_{p,q}^r}$ where $\rho > r$.

For (c) it is sufficient to show $Var[\hat{g}_k(x)] \longrightarrow 0$ as $n \to \infty$ to prove that $\hat{g}_k(x)$ is consistent. 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

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

Then,

$$
\begin{aligned}
Var[\hat{g}_k(x)] &= \frac{\sigma^2}{nh_n}\int M_k^2(\psi)f(x+h_n\psi)d\psi - \frac{1}{n}\left\{\int M_k(\psi)m(x+h_n\psi)f(x+h_n\psi)d\psi\right\}^2 \\
&+ \frac{1}{nh_n}\int M_k^2(\psi)m^2(x+h_n\psi)f(x+h_n\psi)d\psi \qquad (21)
\end{aligned}
$$

Since f and $m \in C^0(\mathbb{R})$ and Assumption 4(3)-(4) , we have

$$
\begin{aligned}
&\int M_k^2(\psi)f(x+h\psi)d\psi \leq \sup_{x\in\mathcal{G}}|f(x)|\int M_k^2(\psi)d\psi \leq \sup_{x\in\mathcal{G}}|f(x)|C\int|K(\psi)|d\psi = O(1) \\
&\int M_k^2(\psi)m^2(x+h\psi)f(x+h\psi)d\psi \leq \sup_{x\in\mathcal{G}}|m(x)|^2\sup_{x\in\mathcal{G}}|f(x)|\int M_k^2(\psi)d\psi = O(1) \\
&\int M_k(\psi)m(x+h\psi)f(x+h\psi)d\psi \leq \sup_{x\in\mathcal{G}}|m(x)|\sup_{x\in\mathcal{G}}|f(x)|C\int|K(\psi)|d\psi = O(1)
\end{aligned}
$$

Given that $nh_n \to \infty$ as $n \to \infty$, we have $Var[\hat{g}_k(x)] \longrightarrow 0$. Hence $\hat{g}_k(x)\xrightarrow{p}g(x)$. □

Theorem 6

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

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

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 $\tau = 1,2,\cdots,m$,

$$
|s_1(x) - E[s_1(x)]| \leq |s_1(x) - s_1(x_\tau)| + |s_1(x_\tau) - E[s_1(x_\tau)]| + |E[s_1(x_\tau)] - E[s_1(x)]|. \qquad (22)
$$

Note that

$$
\begin{aligned}
&|s_1(x) - s_1(x_\tau)| \\
&= \left|\frac{1}{nh}\sum_{t=1}^{n}M_k\left(\frac{X_t-x}{h}\right)m(X_t) - \frac{1}{nh_n}\sum_{t=1}^{n}M_k\left(\frac{X_t-x_\tau}{h_n}\right)m(X_t)\right|
\end{aligned}
$$

$$\leq \frac{1}{nh_n}\sum_{t=1}^{n}\left|M_k\left(\frac{X_t-x}{h_n}\right)-M_k\left(\frac{X_t-x^\tau}{h_n}\right)\right||m(X_t)|$$

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

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 (22) 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,

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

where d_n is a sequence of stochastic variables. If every $\epsilon > 0$ there exists $M_\epsilon > 0$ and a non stochastic sequence $\{a_n\}$ such that $P\left[\frac{|d_n|}{a_n} > M_\epsilon\right] < \epsilon$ for all n. We write $d_n = O_p(a_n)$.

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

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

$$\begin{aligned}
|s_1(x_\tau) - E[s_1(x_\tau)]| &= \left|\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{h_n}M_k\left(\frac{X_t-x_\tau}{h_n}\right)m(X_t) - \frac{1}{h_n}E\left[M_k\left(\frac{X_t-x_\tau}{h_n}\right)m(X_t)\right]\right\}\right| \\
&= \left|\frac{1}{n}\sum_{t=1}^{n}W_{tn}\right| \qquad (23)
\end{aligned}$$

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

$$
\begin{aligned}
|W_{tn}| &= \left|\frac{1}{h_n}M_k\left(\frac{X_t-x_\tau}{h_n}\right)m(X_t)-\frac{1}{h_n}E\Big[M_k\left(\frac{X_t-x_\tau}{h_n}\right)m(X_t)\Big]\right| \\
&= \left|\frac{1}{h_n}\Big(-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\Big)K\left(\frac{X_t-x_\tau}{sh_n}\right)m(X_t)-\frac{1}{h_n}\Big(-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\Big)E\Big[K\left(\frac{X_t-x_\tau}{sh_n}\right)m(X_t)\Big]\right| \\
&\leq \frac{1}{h_n}cB_1\sup_{x\in\mathbb{R}}|m(x)|\Big[1+\int|f(\alpha)|d\alpha\Big]\leq 2cB_1\frac{1}{h_n}\sup_{x\in\mathbb{R}}|m(x)| \quad \text{where } c=\left|-\tfrac{1}{c_{k,0}}\textstyle\sum_{|s|=1}^{k}\tfrac{c_{k,s}}{|s|}\right|.
\end{aligned}
$$

since $\int|f(\alpha)|\leq 1$, $m\in\mathcal{C}^0(\mathbb{R})$ and $\sup_{x\in\mathbb{R}}|K(x)|\leq B_1$ for all $x\in\mathbb{R}$.

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

$$
\begin{aligned}
&h_nVar(W_{tn}) \\
&=\int M_k^2(\psi)m^2(x_\tau+h_n\psi)f(x_\tau+h_n\psi)d\psi-h_n\Big[\int M_k(\psi)m(x_\tau+h_n\psi)f(x_\tau+h_n\psi)d\psi\Big]^2 \qquad (24)
\end{aligned}
$$

From (23), we have

$$
\begin{aligned}
P\left[|s_1(x_\tau)-E[s_1(x_\tau)]|>a_nM_{n,\epsilon}\right] &= P\left[\left|\frac{1}{n}\sum_{t=1}^{n}W_{tn}\right|>a_nM_{n,\epsilon}\right]=P\left[\left|\sum_{t=1}^{n}W_{tn}\right|>na_nM_{n,\epsilon}\right] \\
&\leq 2\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2h_nVar(W_{tn})+\frac{2}{3}B_1c\sup_{x\in\mathbb{R}}|m(x)|a_nM_{n,\epsilon}}\right\} \\
&\qquad \text{by Bernstein's inequality.}
\end{aligned}
$$

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

$$
\begin{aligned}
P\left[\frac{1}{a_n}\max_{1\leq\tau\leq m}|s_1(x_\tau)-E[s_1(x_\tau)]|>M_{n,\epsilon}\right] &\leq \sum_{\tau=1}^{m}2\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2h_nVar(W_{tn})+\frac{2}{3}B_1c\sup_{x\in\mathbb{R}}|m(x)|a_nM_{n,\epsilon}}\right\} \\
&\leq 2m\max_{1\leq\tau\leq m}\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2g_n(x_\tau)+\frac{2}{3}B_1c\sup_{x\in\mathbb{R}}|m(x)|a_nM_{n,\epsilon}}\right\} \\
&= 2m\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2g_n(x^m)+\frac{2}{3}B_1c\sup_{x\in\mathbb{R}}|m(x)|a_nM_{n,\epsilon}}\right\} \qquad (25)
\end{aligned}
$$

where x^m corresponds to the point of the given function such that $\exp\left\{-\frac{a_n^2M_{n,\epsilon}^2nh_n}{2h_nE[W_{tn}^2]+\frac{2}{3}B_1c\sup_{x\in\mathbb{R}}|m(x)|a_nM_{n,\epsilon}}\right\}$ which the function $\exp\{\cdot\}$ attains its maximum value. Thus we have

$$
g_n(x^m)=\int M_k^2(\psi)m^2(x^m+h_n\psi)f(x^m+h_n\psi)d\psi-h_n\Big[\int M_k(\psi)m(x^m+h_n\psi)f(x^m+h_n\psi)d\psi\Big]^2 \qquad (26)
$$

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

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

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

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

where $\triangle_n = M_\epsilon^2 + \frac{4c^2}{\log n} - \frac{4M_\epsilon c}{(\log n)^2}$. Let $v_n = 2g_n(x^m) + \frac{2}{3}B_1 c \sup_{x\in\mathbb{R}} |m(x)| a_n M_{n,\epsilon}$.

From (25),

$$P\left[\frac{1}{a_n}\max_{1\le\tau\le m} |s_1(x_\tau) - E[s_1(x_\tau)]| > M_{n,\epsilon}\right] \le 2m \exp\left\{-\frac{\triangle_n \log n}{v_n}\right\} = 2mn^{-\triangle_n/v_n} \tag{27}$$

From (24) and (26), $g_n(x^m) \to m^2(x^m)f(x^m)\int M_k^2(\psi)d\psi$ 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 (27),

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

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

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

to obtain $n^{\triangle_n/v_n - 1}h_n \to \infty$.

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

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 triangle inequality

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

Note that

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

where $c = \left|-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}\right|$. We have that $\{|u_t|\}_{t=1,2,\ldots}$ 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,

$$\begin{aligned} |s_2(x) - E[s_2(x)]| &\leq |s_2(x) - s_2(x_\tau)| + |s_2(x_\tau) - E[s_2(x_\tau)]| + |E(s_2(x_\tau)) - E[s_2(x)]| \\ &\leq |s_2(x_\tau) - E[s_2(x_\tau)]| + 2c\left(\frac{r}{h_n^2}\right)O_p(1) \end{aligned} \quad (28)$$

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

$$|s_2(x_\tau) - E[s_2(x_\tau)]| \leq |\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| + |s_2(x_\tau) - \hat{s}_2(x_\tau)| + |E[s_2(x_\tau)] - E[\hat{s}_2(x_\tau)]| \quad (29)$$

From (28) and (29) for each $x \in \mathcal{G}$, there exists $B(x_\tau, r)$ that contains x

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

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 $a > 0$, $P\left[|u_t| > B_t\right] < \frac{E[|u_t|^a]}{B_t^a} < C\frac{1}{B_t{}^a}$ and $\sum_{t=1}^\infty P\left[|u_t| > B_t\right] < \sum_{t=1}^\infty \frac{C}{B_t^a} < \infty$.

By the Borel-Cantelli Lemma, $\sum_{t=1}^\infty P\left[|u_t| > B_t\right] < \infty$, which implies $P\left[|u_t| > B_t \ i.o\right] = 0$. Hence, for any $\epsilon > 0$ and for all m satisfying $m' < m$ we have $P\left[|u_m| \le B_m\right] > 1 - u$ since $\{B_t\}_{t=1,2,\cdots}$ is an increasing sequence, for $n > m > m'$ we have $P\left[|u_m| \le B_n\right] > 1-u$. Hence, there exists N such that for $n > \max\{m, N\}$ we have that for all $t \le n$, $P\left[|u_t| \le B_n\right] > 1 - \epsilon$ 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 Holder's inequality,

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

where $\left[\int\chi_{\{|u_t|>B_n\}}f_u(u_t)du_t\right]^{1-1/a} = [P(|u_t| > B_n)]^{1-1/a} \le C\left[\frac{E(|u_t|^a)}{B_n^a}\right]^{1-1/a} \le CB_n^{1-a}$ by using Chebychev's inequality. Hence, $T_2 = O(B_n^{1-a})$ for $a > 1$. Given the results of $T_1 = o_{a.s}(1)$ and $T_2 = O(B_n^{1-a})$ we have

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

Let γ_n be a sequence of stochastic variables and $\gamma_{2,n} = \max_{1\le\tau\le m}|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]|$. 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.

$$\begin{aligned}
&P\left[\frac{\gamma_n}{b_n} > \tilde{M}_\epsilon\right] \le P\left[\frac{2cr}{h_n^2}O_p(1) + O(B_n^{1-a}) + \frac{\gamma_{2,n}}{b_n} > \tilde{M}_\epsilon\right] \le P\left[\frac{\gamma_{2,n}}{b_n} > \tilde{M}_\epsilon - \frac{2cr}{h_n^2}O_p(1) - O(B_n^{1-a})\right] \\
&\le P\left[\frac{1}{b_n}\max_{1\le\tau\le m}|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > \tilde{M}_{n,\epsilon}\right] \quad \text{where } \tilde{M}_{n,\epsilon} = \tilde{M}_\epsilon - \tfrac{2cr}{h_n^2}O_p(1) - O(B_n^{1-a}). \\
&\le \sum_{\tau=1}^m P\left[\frac{1}{b_n}|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > \tilde{M}_{n,\epsilon}\right] \le \sum_{\tau=1}^m P\left[|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > b_n\tilde{M}_{n,\epsilon}\right] \qquad (30)
\end{aligned}$$

Note that

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

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

$$\begin{aligned}
|Z_{tn}| &= \left|\frac{1}{n}\sum_{t=1}^n\left\{\frac{1}{h_n}M_k\left(\frac{X_t - x_\tau}{h_n}\right)u_t\chi_{\{|u_t|\le B_n\}} - \frac{1}{h_n}E\left[M_k\left(\frac{X_t - x_\tau}{h_n}\right)u_t\chi_{\{|u_t|\le B_n\}}\right]\right\}\right| \\
&= \left|\frac{1}{h_n}\left[-\frac{1}{c_{k,0}}\sum_{|s|=1}^k\frac{c_{k,s}}{|s|}K\left(\frac{X_t - x_\tau}{sh_n}\right)\right]u_t\chi_{\{|u_t|\le B_n\}}\right. \\
&\qquad\left. - \frac{1}{h_n}E\left[\left[-\frac{1}{c_{k,0}}\sum_{|s|=1}^k\frac{c_{k,s}}{|s|}K\left(\frac{X_t - x_\tau}{sh_n}\right)\right]u_t\chi_{\{|u_t|\le B_n\}}\right]\right| \\
&\le \frac{1}{h_n}cB_1\left[\left|u_t\chi_{\{|u_t|\le B_n\}}\right| + \int|f_X(\alpha)|d\alpha\int\left|u_t\chi_{\{|u_t|\le B_n\}}\right||f_{u|X}(u)|du\right] \\
&\le \frac{1}{h_n}cB_1\left[B_n + B_n\int\int|f_{u|X}(u)|du\right] \le 2cB_1B_n\frac{1}{h_n}
\end{aligned}$$

since $c = \left[-\frac{1}{c_{k,0}}\sum_{|s|=1}^k\frac{c_{k,s}}{|s|}\right]$,$|K(\cdot)| \le B_1$ (Assumption 4(4)) and $|u_t\chi_{\{|u_t|\le B_n\}}| \le B_n$.

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

$-\left[\frac{1}{h_n}\int\int M_k\left(\frac{\alpha - x_\tau}{h_n}\right)u\chi_{\{|u|\le B_n\}}f(\alpha)f_{u|X}(u)d\alpha du\right]^2$

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

Letting $l_n(x_\tau) = hVar(Z_{tn})$, we have

$$\begin{aligned}
l_n(x_\tau) &= \int\int M_k^2(\psi)u^2\chi_{\{|u^2|\le B_n\}}f(x_\tau + h_n\psi)f_{u|X}(u)d\psi du \\
&\quad - h_n\left[\int\int M_k(\psi)u\chi_{\{|u|\le B_n\}}f(x_\tau + h_n\psi)f_{u|X}(u)d\psi du\right]^2
\end{aligned}$$

$$P\left[|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > b_n\tilde{M}_{n,\epsilon}\right] = P\left[\left|\frac{1}{n}\sum_{t=1}^{n} Z_{tn}\right| > b_n\tilde{M}_{n,\epsilon}\right] = P\left[\left|\sum_{t=1}^{n} Z_{tn}\right| > nb_n\tilde{M}_{n,\epsilon}\right]$$
$$\leq 2\exp\left\{-\frac{b_n^2\tilde{M}_{n,\epsilon}nh_n}{2h_nVar[Z_{tn}] + \frac{2}{3}cB_1B_nb_n\tilde{M}_{n,\epsilon}}\right\}$$

by Bernstein's inequality. Then,

$$\begin{aligned} P\left[\frac{1}{b_n}\max_{1\leq\tau\leq m}|\hat{s}_2(x_\tau) - E[\hat{s}_2(x_\tau)]| > \tilde{M}_{n,\epsilon}\right] &\leq \sum_{\tau=1}^{m} 2\exp\left\{-\frac{b_n^2\tilde{M}_{n,\epsilon}nh_n}{2h_nVar[Z_{tn}] + \frac{2}{3}cB_1B_nb_n\tilde{M}_{n,\epsilon}}\right\} \\ &\leq 2m\max_{1\leq\tau\leq m}\exp\left\{-\frac{b_n^2\tilde{M}_{n,\epsilon}nh_n}{2h_nVar[Z_{tn}] + \frac{2}{3}cB_1B_nb_n\tilde{M}_{n,\epsilon}}\right\} \\ &= 2m\exp\left\{-\frac{b_n^2\tilde{M}_{n,\epsilon}nh_n}{2l_n(x^m) + \frac{2}{3}cB_1B_nb_n\tilde{M}_{n,\epsilon}}\right\} \end{aligned} \tag{31}$$

where x^m corresponds to the point of the given function such that $\exp\left\{-\frac{b_n^2\tilde{M}_{n,\epsilon}nh_n}{2h_nVar[Z_{tn}]+\frac{2}{3}cB_1B_nb_n\tilde{M}_{n,\epsilon}}\right\}$ which the function $\exp\{\cdot\}$ attains its maximum value. Thus we have

$$\begin{aligned} l_n(x^m) = &\int\int M_k^2(\psi)u^2\chi_{\{|u^2|\leq B_n\}}f(x^m + h_n\psi)f_{u|X}(u)d\psi du \\ &- h_n\left[\int\int M_k(\psi)u\chi_{\{|u|\leq B_n\}}f(x^m + h_n\psi)f_{u|X}(u)d\psi du\right]^2 \end{aligned} \tag{32}$$

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

$$b_nM_{n,\epsilon} = \left(\frac{\log n}{nh_n}\right)^{1/2}\tilde{M}_\epsilon - \frac{2c}{(nh_n)^{1/2}}O_p(1) - O(B_n^{1-a}) \tag{33}$$

$$B_nb_n\tilde{M}_{n,\epsilon} = \left(\frac{\log n}{nh_n}\right)^{1/2}B_n\tilde{M}_\epsilon - \frac{2c}{(nh_n)^{1/2}}B_nO_p(1) - O(B_n^{2-a}) \quad \text{where } a > 2. \tag{34}$$

We want $\left(\frac{\log n}{nh_n}\right) \longrightarrow 0$ as $n \to \infty$ that implies $b_nM_{n,\epsilon} \to 0$ for $a > 2$ as $n \to \infty$. From (33),

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

Note that

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

where $\triangle_n=\tilde{M}_\epsilon^2+\frac{4c^2}{\log n}O_p(1)+\frac{nh_n}{\log n}O(B^{2(1-a)})-\frac{4c}{(\log n)^{1/2}}O_p(1)-2\left(\frac{nh_n}{\log n}\right)^{1/2}\tilde{M}_\epsilon O(B_n^{1-a})$ $+\,4c\left(\frac{nh_n}{\log n}\right)^{1/2}O(B_n^{1-a})$. 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 $v_n=2l_n(x^m)+\frac{2}{3}cB_1B_nb_n\tilde{M}_{n,\epsilon}$. From (31), 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 (32), $l_n(x^m)\to f(x^m)\int M_k^2(\psi)d\psi E[u^2|X]<\infty$ 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 $\sup_{x\in\mathcal{G}}|s_2(x)-E[s_2(x)]|=O_p\left(\left(\frac{\log n}{nh_n}\right)^{1/2}\right)$.

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

□

Theorem 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,

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

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

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

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

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

$$\int|M_k(\psi)|^{2+\delta}f(x+h_n\psi)d\psi=\int\left|-\frac{1}{c_{k,0}}\sum_{|s|=1}^{k}\frac{c_{k,s}}{|s|}K\left(\frac{\psi}{s}\right)\right|^{2+\delta}f(x+h_n\psi)d\psi$$

$$\leq C2^{1+\delta}\int\sum_{|s|=1}^{k}\left|\frac{c_{k,s}}{|s|}K\left(\frac{\psi}{s}\right)\right|^{2+\delta}|f(x+h_n\psi)|d\psi\quad\text{by } C_r \text{ inequality}$$

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

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 $\frac{\sum_{t=1}^n\frac{1}{nh_n}M_k\left(\frac{X_t-x}{h_n}\right)[Y_t-m(X_t)]}{\left[\frac{\sigma^2}{nh_n^2}\int M_k^2\left(\frac{X_t-x}{h_n}\right)f(X_t)dX_t\right]^{1/2}}\xrightarrow{d}\mathcal{N}(0,1)$. Thus, $\sqrt{nh_n}\left[\hat{g}_k(x)-E\left(\hat{g}_k(x)|X_t\right)\right]\xrightarrow{d}\mathcal{N}\left(0,\sigma^2 f(x)\int M_k^2(\psi)d\psi\right)$.

□

Theorem 8

Proof. For $x\in\mathbb{R}$, we have

$$E[\hat{m}_k(x)]-m(x)=E[\hat{m}_k(x)-m(x)]=E\left[\frac{\hat{g}_k(x)}{\hat{f}_k(x)}-\frac{g(x)}{f(x)}\right]=\frac{1}{f(x)}E\left[\frac{\hat{g}_k(x)}{1+\frac{1}{f(x)}O_p\left(h_n^r+(nh_n)^{-1/2}\right)}\right]-\frac{g(x)}{f(x)}$$

since $E[|\hat{f}_k(x)-f(x)|^2]=O(h_n^{2r}+(nh_n)^{-1})$ which implies $|\hat{f}_k(x)-f(x)|=O_p(h_n^r+(nh_n)^{-1/2})$.

$$=\frac{1}{f(x)}E\left[\hat{g}_k(x)\left(1+\frac{1}{f(x)}O_p\left(h_n^r+(nh_n)^{-1/2}\right)\right)\right]-\frac{g(x)}{f(x)}\tag{35}$$

$$=\frac{1}{f(x)}E[\hat{g}_k(x)-g(x)]+\frac{1}{f(x)^2}E[\hat{g}_k(x)O_p(h^r+(nh)^{-1/2})]$$

since $E[|\hat{f}_k(x)-f(x)|]\leq\left(E[|\hat{f}_k(x)-f(x)|^2]\right)^{1/2}=O(h^r+(nh_n)^{-1/2})$.

$$\leq O(h_n^r)+\frac{1}{f(x)^2}(E[\hat{g}_k(x)^2])^{1/2}\left(E[|\hat{f}_k(x)-f(x)|^2]\right)^{1/2}\tag{36}$$

$$=O(h_n^r)+O(h_n^r+(nh_n)^{-1/2})\ \text{ since } Bias(\hat{g}_k(x))=O(h_n^r) \text{ and } E[\hat{g}_k(x)^2]<\infty.$$

$$=O(h_n^r)$$

For the equation(35), we use the fact that sum of the infinite series $\frac{1}{1+a}=1+(-a)+(-a)^2+\cdots$ such that $\frac{1}{1+\frac{1}{f(x)}O_p(h_n^r+(nh_n)^{-1/2})}=1-\frac{1}{f(x)}O_p(h^r+(nh_n)^{-1/2})+\left(\frac{1}{f(x)}O_p(h^r+(nh_n)^{-1/2})\right)^2+\cdots$. For the equation (36), we use the inequality as follows, for $1\leq p\leq q\leq\infty$, $(E|X|^p)^{1/p}\leq(E|X|^q)^{1/q}$ where X is a random variable.

□

Theorem 9

Proof. For $x \in \mathbb{R}$ and $k = 1, 2, \cdots$, we have

$$E[\hat{m}_k(x)] - \hat{m}_k(x) = E\left[\frac{\hat{g}_k(x)}{\hat{f}_k(x)}\right] - \frac{\hat{g}_k(x)}{\hat{f}_k(x)} = E\left[\frac{\hat{g}_k(x)}{f(x) + O_p(h_n^r + (nh_n)^{-1/2})}\right] - \frac{\hat{g}_k(x)}{f(x) + O_p(h_n^r + (nh_n)^{-1/2})}$$
$$= \frac{1}{f(x)}E\left[\hat{g}_k(x)\left(1 + \frac{1}{f(x)}O_p(h_n^r + (nh_n)^{-1/2})\right)\right] - \frac{1}{f(x)}\hat{g}_k(x)\left(1 + \frac{1}{f(x)}O_p(h_n^r + (nh_n)^{-1/2})\right)$$
$$= \frac{1}{f(x)}\Big(E[\hat{g}_k(x)] - \hat{g}_k(x)\Big) - \frac{1}{f(x)^2}\Big(E\left[\hat{g}_k(x)O_p(h_n^r + (nh_n)^{-1/2})\right] - \hat{g}_k(x)O_p(h_n^r + (nh_n)^{-1/2})\Big)$$
$$= \frac{1}{f(x)}\Big(E[\hat{g}_k(x)] - \hat{g}_k(x)\Big) + o(1) \quad \text{since } h_n \to 0 \text{ and } nh_n \to \infty \text{ as } n \to \infty.$$

since $E\left[|\hat{g}_k(x)O_p(h_n^r + (nh_n)^{-1/2})|\right] \le E\left[|\hat{g}_k(x)||\hat{f}_k(x) - f(x)|\right] \le (E[|\hat{g}_k(x)|^2])^{1/2}(E[|\hat{f}_k(x) - f(x)|]^2)^{1/2} = o(1)$ since $(E[|\hat{f}_k(x)] - f(x)|^2)^{1/2} = O(h_n^r + (nh_n)^{-1/2})$. From equation (15) $\sup_{x\in\mathcal{G}} |E[\hat{g}_k(x)] - \hat{g}_k(x)| = O_p\left(\left(\frac{\log n}{nh_n}\right)^{1/2}\right)$, we conclude as follows,

$$\sup_{x\in\mathcal{G}} |E[\hat{m}_k(x)] - \hat{m}_k(x)| = O_p\left(\left(\frac{\log n}{nh_n}\right)^{1/2}\right) \quad \text{where } k = 1, 2, \cdots.$$

Hence, $\sup_{x\in\mathcal{G}} |\hat{m}_k(x) - m(x)| \le \sup_{x\in\mathcal{G}} |\hat{m}_k(x) - E[\hat{m}_k(x)]| + \sup_{x\in\mathcal{G}} |E[\hat{m}_k(x)] - m(x)|$.

From Theorem 8, we have $\sup_{x\in\mathcal{G}} |\hat{m}_k(x) - m(x)| = O_p\left(h_n^r + \left(\frac{\log n}{nh_n}\right)^{1/2}\right)$. □

Theorem 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 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.

$$\sqrt{nh}\Big(\hat{g}_k(x) - E[\hat{g}_k(x)|X_t]\Big)/\hat{f}_k(x) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 f(x)^{-1}\int M_k^2(\psi)d\psi\right) \tag{37}$$

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

From (37), 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

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

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

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

Appendix 2 - Tables and figures

Table 1

Local constant estimators with cross validation bandwidth h^{CV}; Trimmed average absolute Bias (B);

Trimmed average Variance (V); Trimmed 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

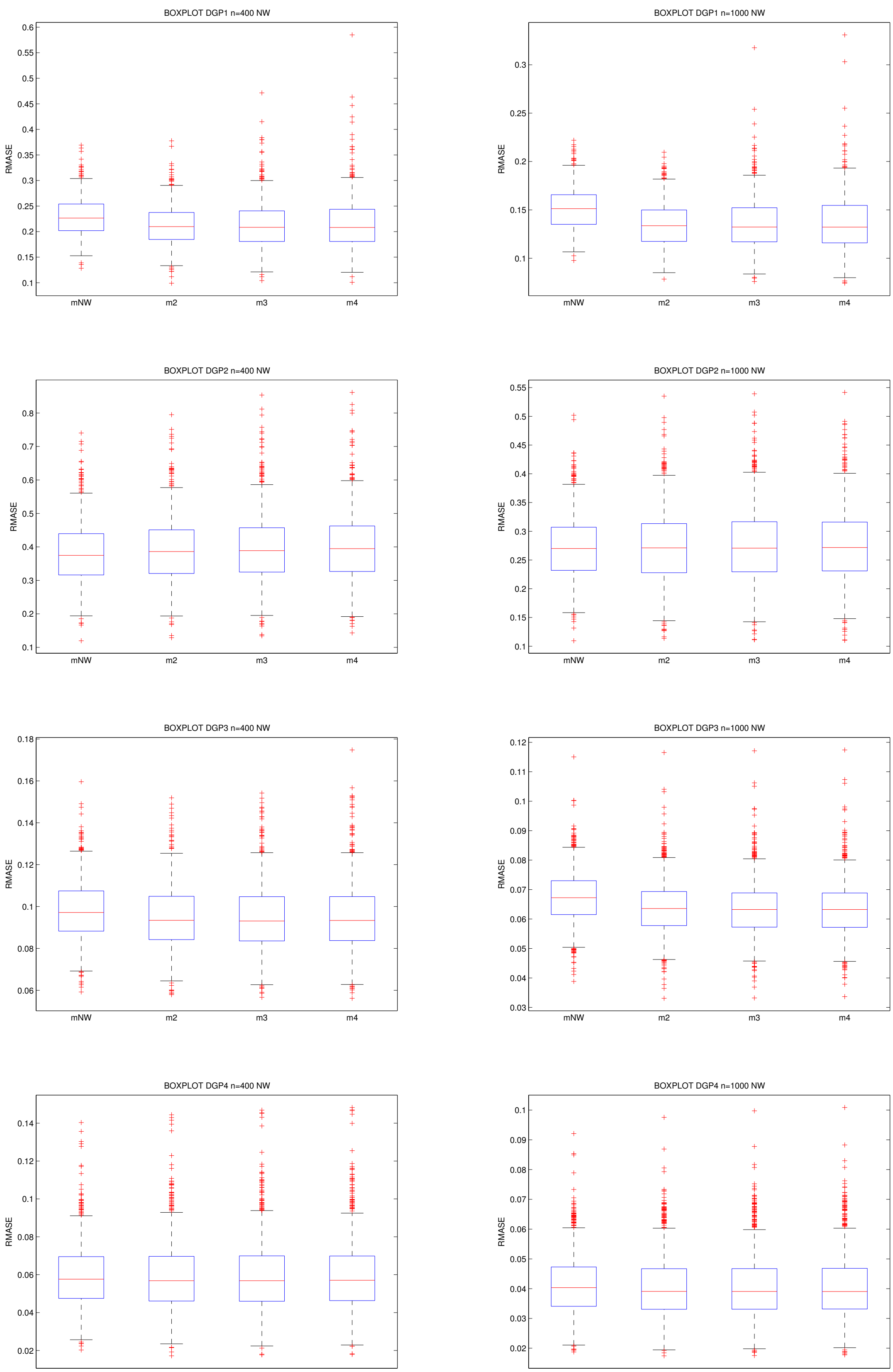

Figure 1: These figures are box plots of trimmed RMSE from estimators $\hat{m}_{NW}, \hat{m}_2, \hat{m}_3$ and $\hat{m}_4$ and four DGPs. DGP1, DGP2, DGP3 and DGP4 indicate $m_1(x)$, $m_2(x)$, $m_3(x)$ and $m_4(x)$ respectively. We consider the sample size $n = 400$ and $n = 1000$

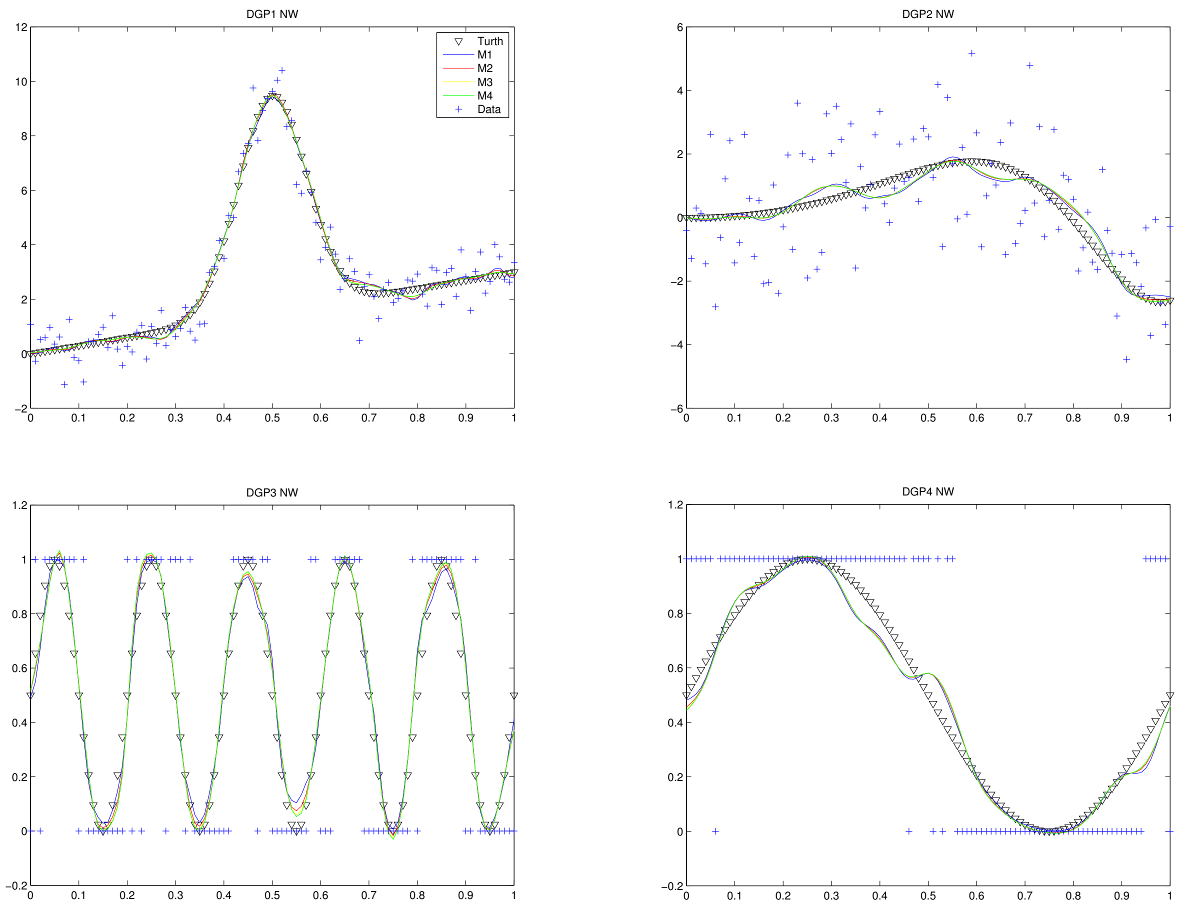

Figure 2: These figures represent four data generating processes with four local constant regression estimators. $\bigtriangledown$ is a true line, the blue line is NW regression estimator, the red line is a local constant estimator based on M_2 kernel. The yellow line indicates a local constant estimator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

References

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45.

Besov, O. V., Valentin, P., and Nikol'skii, S. M. (1978). *Integral Representations of Functions and Imbedding: Theorems.: Vol.: 1.* John Wiley & Sons.

Carroll, R. J., Delaigle, A., and Hall, P. (2009). Nonparametric prediction in measurement error models. *Journal of the American Statistical Association*, 104(487).

Choi, E., Hall, P., and Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *Annals of statistics*, pages 1339–1355.

Collomb, G. (1981). Estimation non-paramétrique de la régression: Revue bibliographique. *International Statistical Review/Revue Internationale de Statistique*, pages 75–93.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians.* Oxford university press.

Devroye, L. P. (1978). The uniform convergence of the nadaraya-watson regression function estimate. *Canadian Journal of Statistics*, 6(2):179–191.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American statistical Association*, 87(420):998–1004.

Gasser, T., Muller, H., and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 238–252.

Glad, I. K. (1998). Parametrically guided non-parametric regression. *Scandinavian Journal of Statistics*, 25(4):649–668.

Mack, Y. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61(3):405–415.

Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. *The Annals of Mathematical Statistics*, 14(3):217–226.

Martins-Filho, C., Mishra, S., and Ullah, A. (2008). A class of improved parametrically guided nonparametric regression estimators. *Econometric Reviews*, 27(4-6):542–573.

Mynbaev, K. and Martins-Filho, C. (2010). Bias reduction in kernel density estimation via lipschitz condition. *Journal of Nonparametric Statistics*, 22(2):219–235.

Mynbaev, K. and Martins-Filho, C. (2014). Consistency and asymptotic normality for a nonparametric prediction under measurement errors.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Racine, J. (2001). Bias-corrected kernel regression. *Journal of Quantitative Economics*, 17(1):25–42.

Schucany, W. R. (1989). On nonparametric regression with higher-order kernels. *Journal of Statistical Planning and Inference*, 23(2):145–151.

Triebel, H. (1985). *Theory of function spaces.* Birkhäuser.

Triebel, H. (2010). *Theory of function spaces II.* Birkhäuser.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.