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 [y]

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.

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

where the binomial coefficients $C_{2k}^N = \frac{(2k)!}{(2k-N)!N!}$, $N = 0, \ldots, 2k$, $k \in \{1, 2, \ldots\}$ and $c_{k,s} = (-1)^{s+k} C_{2k}^{s+k}$, $s = -k, \ldots, k$. 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

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

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 $B^r_{\infty,q}$ where $1 \le q \le \infty$ and $r > 0$. This assumption is desirable for the following reasons: (i) l-times continuous differentiability and uniform boundedness of m is stronger than $m \in B^r_{\infty,q}$ where $l < r$, that is, $C^l(\mathbb{R}) \subset B^r_{\infty,q}$ where $C^l(\mathbb{R})$ denotes the space of l

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 where $c_{k;s} = (-1)^{s+k} C_{2k}^{s+k}$ for $s = -k, \ldots, k$ and $k \in \{1, 2, \ldots\}$. 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 B

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

$$\hat{f}_k($$

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

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

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

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* $\left[\int |K(\ )|^{q'} | |^{(r+1/q)q'} d\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for* $1 \le q \le \infty$. *Then, for all* $x \in \mathbb{R}$ *and* $k = 1, 2, \ldots$,

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

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* $\left[\int |K(\ )|^{q'} | |^{(r+1/q)q'} d\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for* $1 \le q \le \infty$. *Let* G *be a compact subset of* $\mathbb{R}$. *For all* $x \in \mathbb{R}$ *and* $k = 1, 2, \ldots$, *we have*

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

Uniform consistency of the density estimator requires $\frac{\log n}{nh_n} \to 0$ 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* $k = 1, 2, \ldots$, *we have*

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

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

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

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

where 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*

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

Given $\hat{f}_k(x)$ such that $\hat{f}_k(x) = f(x) + o_p(1)$

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* $\left[\int |K(\cdot)|^{q'} |\cdot|^{(r+1/q)q'} d\cdot\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for* $1 \le q \le \infty$. *For* $x \in \mathbb{R}$, $k = 1, 2, \cdots,$

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

Uniform consistency of $\hat{m}_k$ requires $\frac{\log n}{nh_n} \to 0$ 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* $\left[\int |K(\cdot)|^{q'} |\cdot|^{(r+1/q)q} d\cdot\right]^{1/q'} < \infty$ *where* $1/q + 1/q' = 1$ *for* $1 \le q \le \infty$. *For* $x \in \mathbb{R}$ *and* $k = 1, 2, \cdots$, *we have*

$$\sqrt{nh_n}\left(\hat{m}_k(x) - m(x) + O_p(h_n^r)\right) \xrightarrow{d} N$$

estimator, which is given by $\hat{m}_{NW}(x) \quad \hat{m}_1(x) \quad (nh_n)^{1} \sum_{j=1}^{n} K \left(\frac{X_j \quad x}{h_n} \right) Y$

absolute bias (B), average variance (V) and average root MSE (R) of our estimators $\hat{m}$

Lemma 2

Proof. Let s and l 2 $\mathbb{Z}_+$ such that $l < r < s$.

Z

jhj

$$= \sup_{x \in \mathbb{R}} |D^s f(x)|^q \int_0^1 h^{-rq+sq-1} + h^{-rq-1}(-h)^{sq}\, dh$$

$$+ \sup_{x \in \mathbb{R}} |D^l f(x)|^q \int_1^{\infty} c_1 h^{-rq+lq-1} + c_2 h^{-rq-1}(-h)^{lq}\, dh$$

$$= \sup_{x \in \mathbb{R}} |D^s f(x)|^q \frac{1}{(s-r)q}(1 + (-1)^{sq}) + \sup_{x \in \mathbb{R}} |D^l f(x)| \frac{1}{(r-l)q}(c_1 + c_2(-1)^{lq})$$

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

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

The last inequality follows from the fact that $C^s(\mathbb{R}) \subset C^l(\mathbb{R})$ for $s > l$. Hence, we have $||f||_{B^r_{\infty,q}}$

That is, $C^l(\mathbb{R}) \subset B^r_{\infty,q}(\mathbb{R})$ where $l < r$.

Theorem 1

Proof. (a)

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

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

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

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

$$|Bias(\hat{f}_k(x))| = |E(\hat{f}_k(x)) - f(x)| = \int \frac{1}{c_{k,0}} K(\ )$$

(3962 0 Td [( )]T 9.962

Given that

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

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

Theorem 2

Proof.

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

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

$$
\begin{aligned}
\int M_k^2(\psi)f(x+h\psi)d\psi &= \int M_k^2(\psi)[f(x+h\psi) - f(x)]d\psi + \int M_k^2(\psi)f(x)d\psi \\
\le \int_{|h\psi|\le\delta} M_k^2(\psi)|f(x+h\psi) - f(x)|d\psi &+ \int_{|h\psi|>\delta} M_k^2(\psi)|f(x+h\psi) - f(x)|d\psi + f(x)\int M_k^2(\psi)d\psi \\
\le \sup_{|y|\le\delta;x\in\mathbb{R}}|f(x+y) - f(x)|\int M_k^2(\psi)d\psi &+ 2\sup_{x\in\mathbb{R}}|f(x)|\int_{|h\psi|>\delta} M_k^2(\psi)d\psi + \sup_{x\in\mathbb{R}}|f(x)|\int M_k^2(\psi)d\psi \\
\text{since } f \in C^0(\mathbb{R}) < \infty &
\end{aligned} \tag{16}
$$

The inequality follows from $\int M_k^2(\psi)d\psi \le C\int|K(\psi)|d\psi < \infty$ 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 $\{X_t\}_{t=1,2,\ldots,n}$ 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 G be a compact subset of $\mathbb{R}$ that is, $G \subset \mathbb{R}$. The collection $F = \{B(x; r) : x \in G; r > 0\}$ is an open covering of G. By the Heine-Borel theorem, the open covering has a finite subcovering. That is, there exists a collection $F' = \{B(x_i; r) : x_i \in G; r > 0; i = 1, 2, \cdots, m;$ where m is finite$\}$ such that $G \subset F'$. Given that K satisfies a Lipschitz condition of order 1 Assumption 4(5), for $x \in G$, we have

$$
\begin{aligned}
|\hat{f}_k(x) - \hat{f}_k(x_i)| &= \left| \frac{1}{nh_n} \sum_{t=1}^{n} M_k\left(\frac{X_t - x}{h_n}\right) - \frac{1}{nh} \sum_{t=1}^{n} M_k\left(\frac{X_t - x_i}{h_n}\right) \right| \\
&\leq \frac{1}{nh_n} \sum_{t=1}^{n} \frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{|c_{k,s}|}{|s|} \left| K\left(\frac{X_t - x}{sh_n}\right) \cdots \right.
\end{aligned}
$$

Hence, $P\left[\frac{d_n}{a_n} > M\right]$

Hence, P h $\frac{1}{a_n}$ max$_1$ $_m$ $j\hat{f}_k$ [redacted] ume of $B(x\ ;r)$ for x 2 $\mathbb{R}$ is

$2r = 2$ $\frac{h_n^3}{n}$ $^{1=2}$ $= 2r_n$. Sin [redacted] where m is finite$g$ such that

G F^0 is a covering for G, i [redacted] ce G is bounded, there exists

x_0 2 $\mathbb{R}$ and $r_0 < 1$ such t [redacted] $= 2m$ $\frac{h_n^3}{n}$ $^{1=2}$ $2r_0$ which

implies that m $r_0 r_n{}^1 = r_0$ [redacted]

$$2mn^{\ 4_n = v_n} \quad 2r \text{[redacted]} \quad \frac{1}{h_n^3 n^{2\ 4_n=v_n\ \ 1}}^{\ 1=2}$$

$$= \quad 2r \text{[redacted]} \quad \frac{1}{n^{4_n=v_n\ \ 1}}$$

Since nh_n ! 1 it su□ces to [redacted] .

Given that $4_n = M^2$ [redacted] , 4_n ! M^2, $g_n(x^m)$!

$f(x^m)$ R $M_k^2(\ )d$ as n ! [redacted] $\frac{M^2}{2f(x^m)\ \ \text{R}\ M_k^2(\ )d}$ 1. Since

nh_n ! 1 it su□ces to choo [redacted] 1 or $\frac{M^2}{2f(x^m)\ \ \text{R}\ M_k^2(\ )d}$ 2 to

obtain $n^{\frac{4_n}{v_n}\ 1}h_n$! 1 .

Now,

$$\sup_{x2G} j\hat{f}_k(x) \quad f(x)j \quad \text{[redacted]}$$

$$= \text{[redacted]} \ ^{1=2} O_p(1) + h_n^r O(1)$$

□

Theorem 4 [redacted]

Proof. We have for x 2 $\mathbb{R}$, f [redacted] $\frac{X_t\ \ x}{h_n}$ i i

Let $Z_{nt} = \frac{1}{nh_n} M_k$ $\frac{X_t\ \ x}{h_n}$ [redacted] e

$$S_n^2 = \sum_{t=1} E \ \frac{1}{nh} M \text{[redacted]} \frac{X_t\ \ x}{h_n}$$

$$= \ \mathcal{V}ar \text{[redacted]}$$

h

1

=1

since $\int jK(\;)j$

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

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

$$|Bias(\hat{g}_k(x))| = \frac{1}{c_{k;0}} \int \sum_{|s|=0}^{k} c_{k;s} K(\;) m(x + sh_n$$

/F11 9.9626 Tf 3.874 0 Td [(x)]TJ/F8 9.9626 Tf

Then,

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

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

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

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

Theorem 6

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

;

1 2

$\hat{g}_k(x)$ ∫

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

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

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

$$C\sup_{x\in\mathbb{R}}|m(x)|\frac{|x-x^*|}{h_n^2} \le C\sup_{x\in\mathbb{R}}|m(x)|\frac{r}{h_n^2} \quad \text{since } x \in B(x^*;r) \text{ which implies } |x-x^*|<r$$

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

Thus, from (22) we have

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

Since for each $x \in G$, there exists $B(x^*;r)$ that contains x,

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

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 $d_{2;n} = \max_{1 \le \ell \le m} |s_1(x$

$$
\begin{aligned}
|W_{tn}| &= \left| \frac{1}{h_n} M_k\left(\frac{X_t - x}{h_n}\right) m(X_t) - \frac{1}{h_n} E\left[M_k\left(\frac{X_t - x}{h_n}\right) m(X_t)\right] \right| \\
&= \left| \frac{1}{h_n}\left(\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} K\left(\frac{X_t - x}{s h_n}\right) m(X_t)\right) - \frac{1}{h_n}\left(\frac{1}{c_{k,0}} \sum_{|s|=1}^{k} \frac{c_{k,s}}{|s|} E\left[K\left(\frac{X_t - x}{s h_n}\right) m(X_t)\right]\right)\right| \\
&\le \frac{1}{h_n} c B_1 \sup_{x \in \mathbb{R}} |m(x)| \left(1 + \int |f(\theta)| d\theta\right) \le 2 c B_1 \frac{1}{h_n} \sup_{x \in \mathbb{R}} |m(x)| \quad \text{where } c = \left| \tfrac{1}{c_{k,0}} \textstyle\sum_{|s|=1}^{k} \tfrac{c_{k,s}}{|s|} \right|.
\end{aligned}
$$

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

$$
\begin{aligned}
Var(W_{tn}) &= E(W_{tn}^2) \\
&= \frac{1}{h_n^2} \int M_k^2\left(\frac{\theta - x}{h_n}\right) m(\theta) f(\theta) d\theta - \left(\frac{1}{h_n^2} \int M_k\left(\frac{\theta - x}{h_n}\right) m(\theta) f(\theta) d\theta\right)^2 \\
&= \frac{1}{h_n} \int M_k^2(\psi) m^2(x + h_n \psi) f(x + h_n \psi) d\psi - \left(\int M_k(\psi) m(x + h_n \psi) f(x + h_n \psi) d\psi\right)^2
\end{aligned}
$$

$$
\begin{aligned}
& h_n Var(W_{tn}) \\
&= \int M_k^2(\psi) m^2(x + h_n \psi) f(x + h_n \psi) d\psi - h_n \left(\int M_k(\psi) m(x + h_n \psi) f(x + h_n \psi) d\psi\right)^2 \qquad (24)
\end{aligned}
$$

From (23), we have

$$
\begin{aligned}
P\left[|s_1(x_\ell) - E[s_1(x_\ell)]| > a_n M_{n,\ell}\right] &= P\left[\left|\frac{1}{n}\sum_{t=1}^{n} W_{tn}\right| > a_n M_{n,\ell}\right] = P\left[\left|\sum_{t=1}^{n} W_{tn}\right| > n a_n M_{n,\ell}\right] \\
&\le 2\exp\left\{ - \frac{a_n^2 M_{n,\ell}^2 n h_n}{2 h_n Var(W_{tn}) + \frac{2}{3} B_1 c \sup_{x \in \mathbb{R}} |m(x)| a_n M_{n,\ell}} \right\} \\
&\quad \text{by Bernstein's inequality.}
\end{aligned}
$$

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

$$
\begin{aligned}
P\left(\frac{1}{a_n} \max_{1 \le \ell \le m} |s_1(x_\ell) - E[s_1(x_\ell)]| > M_{n,\ell}\right) &\le \sum_{\ell=1}^{m} 2\exp\left\{ - \frac{a_n^2 M_{n,\ell}^2 n h_n}{2 h_n Var(W_{tn}) + \frac{2}{3} B_1 c \sup_{x \in \mathbb{R}} |m(x)| a_n M_{n,\ell}} \right\} \\
&\le 2m \max_{1 \le \ell \le m} \exp\left\{ - \frac{a_n^2 M_{n,\ell}^2 n h_n}{2 g_n(x_\ell) + \frac{2}{3} B_1 c \sup_{x \in \mathbb{R}} |m(x)| a_n M_{n,\ell}} \right\} \\
&= 2m \exp\left\{ - \frac{a_n^2 M_{n,\ell}^2 n h_n}{2 g_n(x^m) + \frac{2}{3} B_1 c \sup_{x \in \mathbb{R}} |m(x)| a_n M_{n,\ell}} \right\} \qquad (25)
\end{aligned}
$$

where x^m corresponds to the point of the given function such that $\exp\left\{ - \frac{a_n^2 M_{n,\ell}^2 n h_n}{2 h_n E[W_{tn}^2] + \frac{2}{3} B_1 c \sup_{x \in \mathbb{R}} |m(x)| a_n M_{n,\ell}} \right\}$ which the function $\exp\{\cdot\}$ attains its maximum value. Thus we have

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

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

a

exists $\bar{M} > 0$ and a nonstochastic sequence $\{b_n\}$ such that $P\left[\frac{|\ _n|}{}\right.$

$$P\left[|\hat{s}_2(x) - E[\hat{s}_2(x)]| > b_n \tilde{M}_{n;} \right] = P\left[\frac{1}{n} \sum_{t=1} Z_{tn} > b_n \tilde{M}_{n;} \right] = P\left[\sum_{t=1} Z_{tn} > n b_n \tilde{M}_{n;} \right]$$

$$\leq 2 \exp\left\{ - \frac{b_n^2 \tilde{M}_{n;}\, n h_n}{2 h_n Var[Z_{tn}] + \frac{2}{3} c B_1 B_n b_n \tilde{M}_{n;}} \right\}$$

by Bernstein's inequality. Then,

$$P\left(\frac{1}{b_n} \max_{1 \leq \ \leq m} |\hat{s}_2(x) - E[\hat{s}_2(x)]| > \tilde{M}_{n;} \right) \leq \sum_{=1}^{m} 2 \exp\left\{ - \frac{b_n^2 \tilde{M}_{n;}\, n h_n}{2 h_n Var[Z_{tn}] + \frac{2}{3} c B_1 B_n b_n \tilde{M}_{n;}} \right\}$$

$$\leq 2m \max_{1 \leq \ \leq m} \exp\left\{ - \frac{b_n^2 \tilde{M}_{n;}\, n h_n}{2 h_n Var[Z_{tn}] + \frac{2}{3} c B_1 B_n b_n \tilde{M}_{n;}} \right\}$$

$$= 2m \exp\left\{ - \frac{b_n^2 \tilde{M}_{n;}\, n h_n}{2 l_n(x^m) + \frac{2}{3} c B_1 B_n b_n \tilde{M}_{n;}} \right\} \qquad (31)$$

where x^m corresponds to the point of the given function such that $\exp\left\{ - \frac{b_n^2 \tilde{M}_{n;}\, n h_n}{2 h_n Var[Z_{tn}] + \frac{2}{3} \tilde{M}_{n;}} \right.$

Note that

$$
\begin{aligned}
& nh_n(b_n^2 \tilde{M}^2_{n;} \;) \\
= \quad & \log n \;\; \tilde{M}^2 + \frac{4c^2}{\log n} O_p(1) + \frac{nh_n}{\log n} O(B^{2(1 \;\; a)}) \quad \frac{4c}{(\log n)^{1=2}} O_p(1) \\
& \qquad 2 \;\; \frac{nh_n}{\log n}^{1=2} \tilde{M} \;\; O(B_n^{1 \;\; a}) + 4c \;\; \frac{nh_n}{\log n}^{1=2} O(B_n^{1 \;\; a} \;\; \mathrm{n} \quad \mathrm{n}
\end{aligned}
$$

$E[Z_{tn}] = 0$ where $m(X_t) = E[Y_t|X_t]$.

$$Var(Z_{tn}) \quad = \quad E[Z_{tn}^2] = E\left[\left(\frac{1}{nh_n}M_k\left(\frac{X_t \quad x}{h_n}\right)(Y_t \quad m(X_t))\right)^2\right] = \frac{2}{n^2h_n^2}E\left(M_k^2\left(\frac{X_t \quad x}{h_n}\right)\right)$$

$$= \quad \frac{2}{n^2h_n^2}\int M_k^2\left(\frac{y \quad x}{h_n}\right) f(y)dy$$

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

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

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

Therefore,

$$|X_{tn}|^{2+} \quad = \quad \left|M_k\left(\frac{X_t \quad x}{h_n}\right)\right|^{2+} |Y_t \quad m(X_t$$

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

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

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

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

since $f\in C^0(\mathbb{R})$ (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, $\sum_{t=1}^n X_{tn} \xrightarrow{d} N(0;1)$ which implies $\frac{\sum_{t=1}^n \frac{1}{nh_n}M_k(\frac{X_t-x}{h_n})[Y_t - m(X_t)]}{\left(\frac{\sigma^2}{nh_n^2}\int M_k^2(\frac{X_t-x}{h_n})f(X_t)dX_t\right)^{1/2}} \xrightarrow{d} N(0;1)$. Thus, $\sqrt{nh_n}[\hat g_k(x) - E(\hat g_k(x)|X_t)] \xrightarrow{d} N\left(0;\ \sigma^2 f(x)\int M_k^2(\nu)d\nu\right)$.

□

Theorem 8

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

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

since $E[|$

Theorem 9

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

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

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

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

$\sqrt{nh_n}[E(\hat{m}_k(x)|X_t) - m(x)]$. Note that

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

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

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

$$= \int \frac{1}{nh_n}\sum_{t=1}^{n}\left[\frac{1}{c_{k;0}}\sum_{|s|=1}^{k}\frac{c_{k;s}}{|s|}K\left(\frac{y - x}{sh_n}\right)\right]\left(m(y) - m(x)\right) f(y)dy$$

$$= \int \left[\frac{1}{c_{k;0}}\sum_{|s|=1}^{k} c_{k;s}K(\ )\right][m(x\ ) \ \sum_{|s|=1}^{k} c_{k;s}K(\ )\Big] [m($$

m 0 Td [(x)]TJ/F8 9.9626 Tf 5.694 0 Td [( ))]TJ/F1 9.9626 Tf

Consequently,

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

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

$$\stackrel{d}{\to} N\left(0; \sigma^2 f(x)^{-1} \int M_k^2(\psi)d\psi\right)$$

If [illegible] $\to 0$ as n [illegible] have $\sqrt{nh_n}(\hat{m}_k(x) - m(x)) \stackrel{d}{\to} N\left(0; \sigma^2 f(x)^{-1} \int M_k^2(\psi)d\psi\right)$. □

Appendix 2 - Tables and figures

Table 1

Local constant estimator with cross validation bandwidth h^{CV}; Trimmed average absolute Bias (B); Trimmed average Variance (V); Trimmed average Root Mean Squared Error (R).

	$m_1(x)$			$m_2(x)$		
[illegible]	BV	VR	RB	BV	VR	R
$\hat{m}_{NW}$	0.0517	0.0470	0.2320	0.0386175315069400.3909 [())]TJ/29 408.416 Td [(n)]TJ/F8 9.		

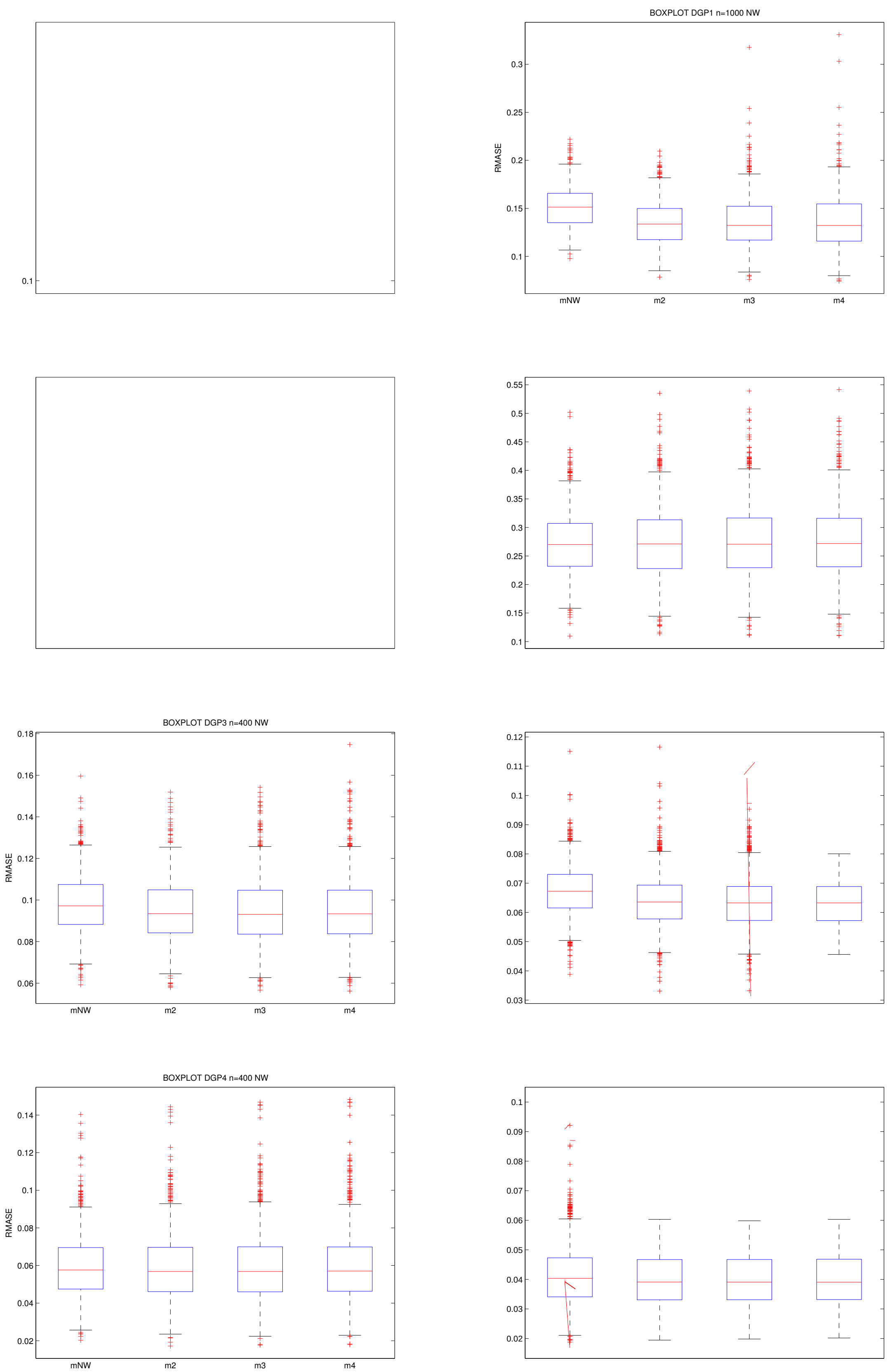
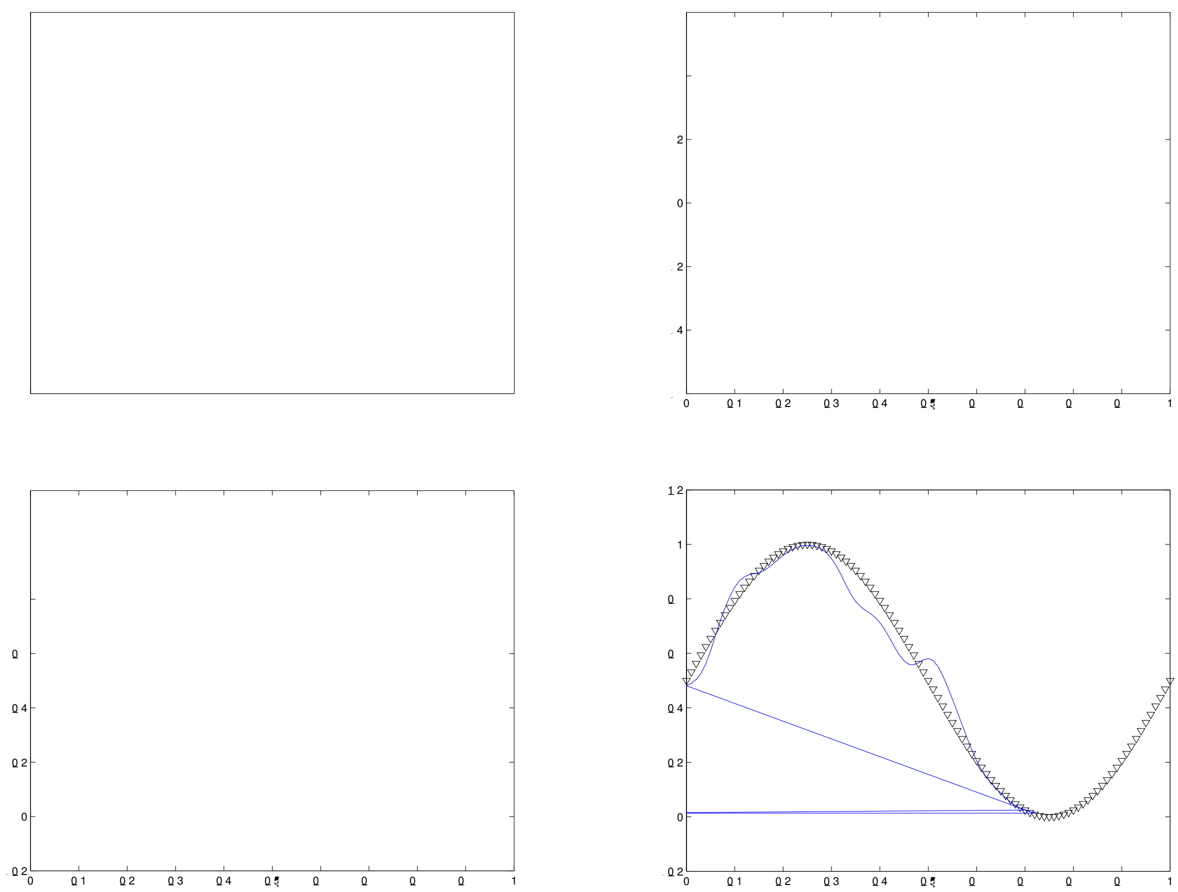


Figure 1: These figures are box plots of trimmed RMSE from estimators $\hat{m}_{NW}$; $\hat{m}_2$; $\hat{m}_3$ and $\hat{m}_4$ and four DGPs. DGP1, DGP2, DGP3 and DGP4 indicate $m_1(x)$, $m_2(x)$, $m_3(x)$ and $m_4(x)$ respeah8501



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.