Bias Reduction for Nonparametric and Semiparametric Regression Models

Ming-Yen Cheng, Tao Huang, Peng Liu, and Heng Peng

Hong Kong Baptist University and National Taiwan University,

Shanghai University of Finance and Economics,

University of Washington and Fred Hutchinson Cancer Research Center,

Hong Kong Baptist University

Supplementary Material

Technical conditions and proofs of the main theoretical results

We need the following technical conditions for theoretical investigation for our methods.

(a) For an $s > 2$, $\mathrm{E}|Y|^{2s} < \infty$ and $\mathrm{E}|X|^{2s} < \infty$.

(b) The density function of X, $f(x)$, is continuous and positive on its compact support.

(c) The second derivatives of $f(x)$ and $\sigma^2(x)$ are continuous and bounded.

(d) The fourth derivative of $m(x)$, $m^{(4)}(x)$, is continuous.

(e) The kernel function $K(t)$ is a asymmetric density function and is absolutely continuous on its support set $[-A, A]$.

(e1) $K(A) \neq 0$ or

(e2) $K(A) = 0, K(t)$ is a absolutely continuous, and $K^2(t)$ and $(K'(t))^2$ are integrable on the $(-\infty, +\infty)$.

Lemma 1. *Under conditions (a)-(e), for $\widehat{m}_h(x_0)$, we have the following higher-order expansion of its bias:*

$$Bias(\widehat{m}_h(x_0)|\,\mathbb{X}) = \frac{1}{2}\mu_2 m^{(2)}(x_0)h^2 + \boldsymbol{a}(x_0)h^4 + o_p(h^4) \tag{S.1}$$

where

$$\begin{aligned}\boldsymbol{a}(x_0) &= \frac{1}{24}\mu_4 m^{(4)}(x_0) - \frac{m^{(2)}(x_0)}{2}\boldsymbol{b}(x_0)\mu_4,\\ \boldsymbol{b}(x_0) &= \left(\frac{f^{(1)}(x_0)}{f(x_0)}\right)^2.\end{aligned}$$

Proof of Lemma 1

From Ruppert and Wand (1994), we know that:

$$\begin{aligned}\mathrm{E}[\widehat{m}_h(x_0) - m(x_0)|\,\mathbb{X}] &= e_1^T(\boldsymbol{X}^T\boldsymbol{W}_{h_n}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_{h_n}\boldsymbol{r}\\ &= e_1^T\left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{W}_{h_n}\boldsymbol{X}\right)^{-1}(S+R),\end{aligned} \tag{S.2}$$

where

$$e_1 = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix}, S = \frac{1}{n}\boldsymbol{X}^T\boldsymbol{W}_{h_n}\left\{\frac{m^{(2)}(x_0)}{2!}\begin{bmatrix} (X_1 - x_0)^2 \\ \vdots \\ (X_n - x_0)^2 \end{bmatrix} + \cdots + \frac{m^{(4)}(x_0)}{4!}\begin{bmatrix} (X_1 - x_0)^4 \\ \vdots \\ (X_n - x_0)^4 \end{bmatrix}\right\},$$

and R is the remainder term in the Taylor expansion. We denote

$$A = \begin{bmatrix} 1 & 0 \\ \\ 0 & h \end{bmatrix}, Q_1 = \begin{bmatrix} \mu_1 & \mu_2 \\ \\ \mu_2 & \mu_3 \end{bmatrix}, N_1 = \begin{bmatrix} 1 & \mu_1 \\ \\ \mu_1 & \mu_2 \end{bmatrix}.$$

Then for any $k = 0, 1, \cdots$, we have

$$\begin{aligned} e_1^T(n^{-1}\boldsymbol{X}^T\boldsymbol{W}_{h_n}\boldsymbol{X})^{-1} &= \frac{1}{f(x_0)}\left\{e_1^T N^{-1} - h\frac{f'(x_0)}{f(x_0)}e_1^T N^{-1}Q_1N^{-1}\right\}A^{-1} \\ &\quad + o_p\left(\begin{bmatrix} h & 1 \end{bmatrix}\right) \end{aligned} \tag{S.3}$$

$$A^{-1}\frac{1}{n}\boldsymbol{X}^T\boldsymbol{W}_{h_n}\begin{bmatrix} (X_1 - x_0)^k \\ \vdots \\ (X_n - x_0)^k \end{bmatrix} = h^k f(x_0)\begin{bmatrix} \mu_k \\ \\ \mu_{k+1} \end{bmatrix} + h^{k+1}f'(x_0)\begin{bmatrix} \mu_{k+1} \\ \\ \mu_{k+2} \end{bmatrix} \tag{S.4}$$

Substituting (S.3) and (S.4) into (S.2) and after some calculation, we complete our proof of the lemma. □

Proof of Theorem 1.

Combining (S.1) and (2.4), we have

$$\begin{aligned}
& E[\widehat{\alpha}_B - m(x_0)|\,\mathbb{X}] \\
= & \sum_{i=1}^{B} \boldsymbol{g}_i \times \left[\frac{1}{2}\mu_2 m^{(2)}(x_0)h_i^2 + \boldsymbol{a}(x_0)h_i^4\right] + o_p(h^4) \\
= & \frac{1}{2}\mu_2 m^{(2)}(x_0) \times \frac{\sum_{k=1}^{B} h_k^4 \sum_{i=1}^{B} h_i^2 - \sum_{k=1}^{B} h_k^2 \sum_{i=1}^{B} h_i^4}{B\sum_{k=1}^{B} h_k^4 - \left(\sum_{k=1}^{B} h_k^2\right)^2} \\
& + \boldsymbol{a}(x_0) \times \frac{\sum_{k=1}^{B} h_k^4 \sum_{i=1}^{B} h_i^4 - \sum_{k=1}^{B} h_k^2 \sum_{i=1}^{B} h_i^6}{B\sum_{k=1}^{B} h_k^4 - \left(\sum_{k=1}^{B} h_k^2\right)^2} + o_p(h^4) \\
= & \boldsymbol{C}(x_0)h^4 + o_p(h^4).
\end{aligned}$$

In the following we calculate the variance of $\widehat{\alpha}_B$. First, we have

$$\mathrm{Var}\left[\widehat{\alpha}_B|\,\mathbb{X}\right] = \sum_{i=1}^{B} \boldsymbol{g}_i^2 \nu_0 \frac{\sigma^2(x_0)}{nf(x_0)}\frac{1}{h_i} + 2\sum_{i<j}^{B} \boldsymbol{g}_i \boldsymbol{g}_j \mathrm{Cov}(V_i, V_j).$$

Through some calculation, we know that

$$\mathrm{Cov}(V_i, V_j) \;=\; \frac{\sigma^2(x_0)}{nf(x_0)}\left[\psi_{ij}^{(0)} - 2\boldsymbol{b}(x_0)h_i h_j \psi_{ij}^{(1)} + \left(\boldsymbol{b}(x_0)h_i h_j\right)^2 \psi_{ij}^{(2)}\right]. \quad \text{(S.5)}$$

From the expression (S.5), when $h_i = h_j = h$, then

$$\begin{aligned}
\mathrm{Cov}(V_i, V_i) &= \frac{\sigma^2(x_0)}{nf(x_0)}\left[\int K(hu)^2 du - 2\boldsymbol{b}(x_0)h^2 \int K(hu)^2 u du + \left(\boldsymbol{b}(x_0)h^2\right)^2 \int K(hu)^2 u^2 du\right] \\
&= \frac{\sigma^2(x_0)}{nhf(x_0)}\left[\nu_0 - 2h\boldsymbol{b}(x_0)\nu_1 + h^2\left(\boldsymbol{b}(x_0)\right)^2 \nu_2\right] \\
&= \frac{\sigma^2(x_0)}{nhf(x_0)}(\nu_0 + o_p(1)).
\end{aligned}$$

Under some conditions, we have

$$\mathrm{Cov}(V_i, V_j) \;=\; \frac{\sigma^2(x_0)}{nf(x_0)}\left(\psi_{ij}^{(0)} + o_p(1)\right).$$

So we have

$$\mathrm{Var}\left[\widehat{\alpha}_B |\, \mathbb{X}\right] = \frac{\sigma^2(x_0)}{nf(x_0)} \sum_{i=1}^{B}\sum_{j=1}^{B} \boldsymbol{g}_i \boldsymbol{g}_j \left(\psi_{ij}^{(0)} + o_p(1)\right).$$

□

Proof of Theorem 2.

From the discussion given in Section 2, we know that

$$\mathrm{Bias}(\widehat{m}_B(x)) = O(h^4 + h^2 h_0^2),$$

and

$$\begin{aligned}
\widehat{m}_B(x) - m(x) &= \widehat{m}_h(x) - \widehat{\beta}_B h^2 - m(x) \\
&= \widehat{m}_h(x) - \mathrm{E}(\widehat{m}_h(x)) + \mathrm{E}(m(x)) - \widehat{\beta}_B h^2 - m(x) \\
&= \widehat{m}_h(x) - \mathrm{E}(\widehat{m}_h(x)) + O(h^4 + h^2 h_0^2).
\end{aligned}$$

By results of the local linear estimator shown by Fan and Gijbels (1996), we have

$$\begin{aligned}
\widehat{m}_h(x) - \mathrm{E}(\widehat{m}_h(x)) &= \frac{T_{n,0}S_{n,2} - T_{n,1}S_{n,1}}{S_{n,2}S_{n,0} - S_{n,1}S_{n,1}} - \mathrm{E}(\widehat{m}_h(x)) \\
&= \frac{1}{nhf(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) e_i (1 + o_p(1)).
\end{aligned}$$

So we obtain that

$$\widehat{m}_B(x) - m(x) = \frac{1}{nhf(x)} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) e_i + o_p(1/\sqrt{nh}) + O(h^4 + h^2 h_0^2).$$

Then based on the conditions on h and h_0, and following the steps in the proof of the uniform convergence lemma, Theorem 1 and 2 in Fan and Zhang (2000) for varying coefficient models, Theorem 2 is easily proved. $\square$

Proof of Theorem 3.

By the definition of $\widetilde{m}_B(x)$ and the fact that $\sum_{i=1}^{B} \boldsymbol{g}_i = 1$, we have

$$\begin{aligned}
\widetilde{m}_B(x) - m(x) &= \sum_{i=1}^{B} \boldsymbol{g}_i V_i - m(x) = \sum_{i=1}^{B} \boldsymbol{g}_i (V_i - m(x)) \\
&= \sum_{i=1}^{B} \left\{ \boldsymbol{g}_i \frac{1}{nh_i f(x)} \sum_{j=1}^{n} K\left(\frac{X_j - x}{h_i}\right) e_j (1 + o_p(1)) \right\} + O_p(h^4)
\end{aligned}$$

Then by the definition of $K_1(t)$ and the bandwidth series $h_i, i = 1, \ldots, B$,

$$\begin{aligned}
\widetilde{m}_B(x) - m(x) &= \sum_{i=1}^{B} \left\{ \frac{1}{nhC_i f(x)} \sum_{j=1}^{n} \boldsymbol{g}_i K\left(\frac{X_i - x}{C_i h}\right) e_j (1 + o_p(1)) \right\} + O_p(h^4) \\
&= \frac{1}{nhf(x)} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{B} \boldsymbol{g}_j K\left(\frac{X_i - x}{C_j h}\right) / C_j \right\} e_i (1 + o_p(1)) + O_p(h^4) \\
&= \frac{1}{nhf(x)} \sum_{i=1}^{n} K_1\left(\frac{X_i - x}{h}\right) (1 + o_p(1)) + O_p(h^4).
\end{aligned}$$

Regard $K_1(\cdot)$ as an equivalent kernel function. Then as in the proof of Theorem 2, following the steps in the proof of Lemma 1, Theorems 1 and 2 of Fan and Zhang (2000), and by some complicated calculation, Theorem 3 is proved. $\square$