Supplementary material to "Nonparametric identification and estimation of a generalized additive model with a flexible additive structure and unknown link"

Songnian Chen* Nianqing Liu† Jian Zhang‡ Yahong Zhou§

Abstract

The supplementary material includes two appendices. Appendix S.1 introduces some additional notation on local polynomial regression for the convenience of discussion in the text and proofs. Appendix S.2 states and proves some technical lemmas needed to show the main theorems in the text.

S.1 Additional Notation on Local Polynomial Regression

For r-th order local polynomial regression of Y_i on X_i, let $\mathbf{j} = (j_1, j_2, \cdots j_d)$ be an arbitrary d-tuple of integers, denote $|\mathbf{j}| = j_1 + j_2 + \cdots + j_d$, $\mathbf{j}! = j_1! \times j_2! \times \cdots \times j_d!$, $x^{\mathbf{j}} = (x^1)^{j_1} \times (x^2)^{j_2} \times \cdots (x^d)^{j_d}$, $D^{\mathbf{j}}H(x) = \frac{\partial^{|\mathbf{j}|}H(x)}{\partial(x^1)^{j_1}\partial(x^2)^{j_2}\cdots\partial(x^d)^{j_d}}$, and $\sum_{0\le|\mathbf{j}|\le r} = \sum_{k=0}^{r}\sum_{j_1+j_2+\cdots+j_d=k}$. The total number of d-tuples with $|\mathbf{j}| = s$ is $M_s = \begin{pmatrix} r+d-1 \\ d-1 \end{pmatrix}$. We arrange these tuples in an ascending lexicographical order style as in Masry (1996)[1]. The correspondent position of each tuple forms a one-to-one map which is called π_s, i.e. $\pi_s(1) = (s, 0, 0, \cdots, 0)$, ... $\pi_s(M_s) = (0, 0, \cdots, 0, s)$. Denote a vector-value function $\mu(\cdot)$ for an arbitrary entry $x \in \mathbb{R}^d$ such that $\mu_s(x)$ is an $M_s \times 1$ vector with l-entry given by $\left[\mu_s(x)\right]_l = x^{\pi_s(l)}$. and we stack these vectors and define an $N_r \times 1$ vector as $\mu(x) = [\mu_0(x), \mu_1(x), \cdots, \mu_r(x)]'$, where $N_r = M_0 + M_1 + \cdots + M_r$. Also, we denote $M_s \times 1$ vectors $H_s(x)$ ($s = 0, 1, \ldots, r+1$) to store $H(x)$ and its derivatives (up to $(r+1)$-th order) such that the l-entry of $\alpha_s(x)$ equals to $\left[H_s(x)\right]_l = \frac{1}{\pi_s(l)!}D^{\pi_s(l)}H(x)$, and $\alpha(x)$ stacks $\alpha_s(x)$ ($s = 0, 1, \ldots, r.$) as $\alpha(x) = [H_0(x), H_1(x), \cdots, H_r(x)]'$, then $\mu(y-x)'\alpha(x)$ is the r-th order Taylor expansion of $H(y)$ at x. Let $S_{n,p,q}(x)$ and $S_{p,q}$ be $M_p \times M_q$ matrices with (l,k)-element given by $\left[S_{n,p,q}(x)\right]_{l,k} = \int u^{\pi_p(l)+\pi_q(k)}K(u)p_X(x+h_H u)du$ and $\left[S_{p,q}\right]_{l,k} = \int u^{\pi_p(l)+\pi_q(k)}K(u)du$, where $u = (u_1, u_2, \cdots, u_d)$, $K(u) = K_1(u^1)K_2(u^2)$ with $u^1 = (u_1, \cdots, u_{d_1})$ and $u^2 = (u_{d_1+1}, \cdots, u_d)$, and $p_X(\cdot)$ is the probability density function of X. Define $N_r \times N_r$ matrices

*Tsingshan Institute for Advanced Business Studies and School of Economics, Zhejiang University, Hangzhou, China.

†Paula and Gregory Chow Institute for Studies in Economics, Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, China.

‡School of Economics, Nankai University, Tianjin, China.

§School of Economics and Dishui Lake Advanced Finance Institute, The Shanghai University of Finance and Economics, Shanghai, China.

[1]The highest priority of the order is based on j_1, second we order by j_2, so on and so forth, finally we order by j_d.

$S_{n,r}(x)$ and S_r as

$$S_{n,r}(x)=\begin{pmatrix} S_{n,0,0}(x) & S_{n,0,1}(x) & \cdots & S_{n,0,r}(x)\\ S_{n,1,0}(x) & S_{n,1,1}(x) & \cdots & S_{n,1,r}(x)\\ \vdots & \vdots & \ddots & \vdots\\ S_{n,r,0}(x) & S_{n,r,1}(x) & \cdots & S_{n,r,r}(x)\end{pmatrix}, S_r=\begin{pmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,r}\\ S_{1,0} & S_{1,1} & \cdots & S_{1,r}\\ \vdots & \vdots & \ddots & \vdots\\ S_{r,0} & S_{r,1} & \cdots & S_{r,r}\end{pmatrix},$$

and $N_r \times M_{r+1}$ matrices $S_{n,r}^{r+1}(x)$ and S_r^{r+1} as $S_{n,r}^{r+1}(x) = \left(S_{n,0,r+1}(x)', S_{n,1,r+1}(x)', \cdots, S_{n,r,r+1}(x)'\right)'$ and $S_r^{r+1} = \left(S'_{0,r+1}, S'_{1,r+1}, \cdots, S'_{r,r+1}\right)'$. For r-th order local polynomial regression of Y_i on $Z_i = \zeta(X_i)$, similarly, for each 2-tuple $\tilde{\mathbf{j}} = (j_1, j_2)$, we can define summation, factorial operation, multiplication and partial derivatives. In the same style as M_s, N_r, $\pi_s(\cdot)$, we can define $\tilde{M}_s$, $\tilde{N}_r$, and the lexicographical order map $\tau_s(\cdot)$. Similar to $\mu(\cdot)$, $S_{n,p,q}(x)$, $S_{p,q}$, $S_{n,r}(x)$, and S_r, we can define $\tilde{\mu}(\cdot)$, $\tilde{S}_{n,p,q}(z)$, $\tilde{S}_{p,q}$, $\tilde{S}_{n,r}(z)$, and $\tilde{S}_r$ with $z = (z^1, z^2)$. Let $\mathcal{S}_{n,p,q}(z,\zeta)$ be an $\tilde{M}_p \times \tilde{M}_q$ matrix with (l,k)-element defined by $\left[\mathcal{S}_{n,p,q}(z,\zeta)\right]_{l,k} = \frac{1}{nh_{\mathcal{H}}^2}\sum_{i=1}^n \left(\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}\right)^{\tau_p(l)+\tau_q(k)} \tilde{K}\left(\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}\right)$, and $\mathcal{Q}_{n,p,0}(z,\zeta)$ be an $\tilde{M}_p \times 1$ vector with k-th entry given by $\left[\mathcal{Q}_{n,p,0}(z,\zeta)\right]_k = \frac{1}{nh_{\mathcal{H}}^2}\sum_{i=1}^n Y_i\left(\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}\right)^{\tau_p(k)} \tilde{K}\left(\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}\right)$, where $z = (z^1, z^2)$, $\zeta = (\zeta^1, \zeta^2)$, $\zeta(X_i) = (\zeta_1(X_i^1), \zeta_2(X_i^2))$, and $\tilde{K}(u) = k_1(u^1)k_2(u^2)$. Also, we define the kernel derivatives $\partial_k \tilde{K}(u) = k'_k(u^k)k_{-k}(u^{-k})$ for $k = 1, 2$. By stacking $\mathcal{S}_{n,p,q}(z,\zeta)$ and $\mathcal{Q}_{n,p}(z,\zeta)$, we define an $\tilde{N}_r \times \tilde{N}_r$ matrix $\mathcal{S}_{n,r}(z,\zeta)$ and an $\tilde{N}_r \times 1$ vector $\mathcal{Q}_{n,r}(z,\zeta)$ as

$$\mathcal{S}_{n,r}(z,\zeta)=\begin{pmatrix} \mathcal{S}_{n,0,0}(z,\zeta) & \mathcal{S}_{n,0,1}(z,\zeta) & \cdots & \mathcal{S}_{n,0,r}(z,\zeta)\\ \mathcal{S}_{n,1,0}(z,\zeta) & \mathcal{S}_{n,1,1}(z,\zeta) & \cdots & \mathcal{S}_{n,1,r}(z,\zeta)\\ \vdots & \vdots & \ddots & \vdots\\ \mathcal{S}_{n,r,0}(z,\zeta) & \mathcal{S}_{n,r,1}(z,\zeta) & \cdots & \mathcal{S}_{n,r,r}(z,\zeta)\end{pmatrix}, \mathcal{Q}_{n,r}(z,\zeta)=\begin{bmatrix} \mathcal{Q}_{n,0,0}(z,\zeta)\\ \mathcal{Q}_{n,1,0}(z,\zeta)\\ \vdots\\ \mathcal{Q}_{n,r,0}(z,\zeta)\end{bmatrix}.$$

Then infeasible local polynomial estimator is $\tilde{\beta}(z) = B_{\mathcal{H}}^{-1}\mathcal{S}_{n,r}(z,\zeta)^{-1}\mathcal{Q}_{n,r}(z,\zeta)$ with unknown parameter $\zeta(\cdot)$, and correspondent feasible estimator is $\widehat{\beta}(z) = B_{\mathcal{H}}^{-1}\mathcal{S}_{n,r}(z,\widehat{\zeta})^{-1}\mathcal{Q}_{n,r}(z,\widehat{\zeta})$, where $B_{\mathcal{H}}$ is a $\tilde{N}_r \times \tilde{N}_r$ diagonal matrix with diagonal vector $D_h = [D_{h,0}, D_{h,1}, \cdots, D_{h,r}]'$ and $D_{h,s} = (h_{\mathcal{H}}^{|\tau(k)|})_{k=1,2,\ldots,\tilde{M}_s}$.

In order to represent the first-order derivatives of $\mathcal{H}(z)$ by $\beta(z)$, we introduce an $\tilde{N}_r \times 2$ vector given by $e_d = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0\\ 0 & 0 & 1 & 0 & \cdots & 0\end{pmatrix}'$, then $\left(\partial_1\mathcal{H}(z), \partial_2\mathcal{H}(z)\right)' = e_d'\beta(z)$. For r-th local polynomial regression of Y_i on $T_i = f_1(X_i^1) + f_2(X_i^2)$. Similar to $\mu(\cdot)$, $H_r(x)$, $S_{n,r}(x)$, S_r and S_r^{r+1}, we can also define $\mu_G(\cdot)$, $G_{r+1}(\tau)$, $S_{n,r}^G(\tau)$, S_r^G and $S_r^{G,r+1}$.

S.2 Technical Lemmas

We state and show in this section the lemmas used to prove the theorems in the text.

S.2.1 Lemma S.1

Lemma S.1 modifies Lemma 3.1 of Powell, Stock, and Stoker (1989) and Lemma 5 of Horowitz (1998). It provides sufficient conditions for approximation error of U-statistic projection other than $o_p(1/\sqrt{n})$. In particular, it degenerates to the case of Lemma 3.1 of Powell, Stock, and Stoker (1989) when $\lambda_n = n$. Denote $U_n = 2\cdot[n(n-1)]^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^n q_n(W_i, W_j)$ and $\widehat{U}_n = E[q_n(W_1, W_2)] + (2/n)\sum_{i=1}^n \left(E[q_n(W_i, W_j)|W_i] - E[q_n(W_1, W_2)]\right)$. As a matter of fact, Lemma S.1 can be further modified as $\widehat{U}_n - U_n = O_p\left[n^{-1} \cdot \sqrt{E\left[q_n(W_1, W_2)^2\right]}\right]$ by its proof.

Lemma S.1. *Suppose that $\{W_i\}_{i=1}^n$ is a sequence of independently and identically distributed random variables or vectors. Let $q_n(\cdot,\cdot)$ be a symmetric function, and λ_n be a sequence of positive scalars. If $E[q_n(W_1, W_2)^2] = o(\lambda_n)$, then $\widehat{U}_n - U_n = o_p\left[\sqrt{\lambda_n}/n\right]$.*

Proof. Follow the same idea as the proof of Lemma 3.1 of Powell, Stock, and Stoker (1989) to get $E(\widehat{U}_n - U_n)^2 = O\left[n^{-2} \cdot E\left[q_n(W_1, W_2)^2\right]\right]$. Thus $(n^2/\lambda_n) \cdot E(\widehat{U}_n - U_n)^2 = O\left[(n^2/\lambda_n) \cdot n^{-2} \cdot E\left[q_n(W_1, W_2)^2\right]\right] = o(1)$. The desired conclusion therefore holds by Markov's inequality. □

S.2.2 Lemma S.2

Lemma S.2 finds the uniform convergence rate and asymptotic representation of the nonparametric regression estimator $\widehat{H}(\cdot)$. We give a proof of Lemma S.2 for completeness.

Lemma S.2. *Let Assumptions 1-5 hold, and the bandwidth h_H satisfy (i) $h_H \to 0$ and (ii) $log(n)/(nh_H^d) \to 0$ as $n \to \infty$. Then*

$$\sup_{x \in \mathcal{S}_X} \left|\widehat{H}(x) - H(x)\right| = O(\xi_H)$$

in probability as $n \to \infty$. Moreover, the asymptotic representation of $\widehat{H}(x) - H(x)$ is given by

$$\begin{aligned}
&\widehat{H}(x) - H(x) \\
=&\frac{1}{nh_H^d}\sum_{i=1}^{n}\left(Y_i - \mu(X_i - x)'\alpha(x)\right)\left\{e_1' S_{n,r}(x)^{-1}\mu\left(\frac{X_i - x}{h_H}\right)\right\} \\
&\cdot K_1\left(\frac{x^1 - X_i^1}{h_H}\right)K_2\left(\frac{x^2 - X_i^2}{h_H}\right) + O(\xi_H^2)
\end{aligned}$$

as $n \to \infty$ in probability uniformly over $x \in \mathcal{S}_X$, where $e_1 = (1, 0, \cdots, 0)'$ is an $N_r \times 1$ vector, $\mu(X_i - x)'\alpha(x)$ represents the r-th order Taylor expansion of $H(X_i)$ at $X_i = x$. $S_{n,r}(x)$, $\mu(\cdot)$ and $\alpha(x)$ are defined in Appendix S.1.

Proof. The first part can be established by an argument similar to the proof of Theorem 6 of Masry (1996). Its proof is hence omitted here. According to the uniform bahadur representation in Remark 1 of Theorem 3.2 in Kong, Linton, and Xia (2010),

$$\begin{aligned}
\widehat{H}(x) - H(x) =&\frac{1}{nh_H^d} e_1' S_{n,r}(x)^{-1} B_H^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_H}\right)\left(Y_i - \mu(X_i - x)'\alpha(x)\right)\mu(X_i - x) \\
&+ O\left(\frac{\log n}{nh_H^d}\right) \\
=&\frac{1}{nh_H^d} e_1' S_{n,r}(x)^{-1} B_H^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_H}\right)\left(Y_i - \mu(X_i - x)'\alpha(x)\right)\mu(X_i - x) \\
&+ O(\xi_H^2)
\end{aligned}$$

as $n \to \infty$ in probability uniformly over $x \in \mathcal{S}_X$, where $e_1 = (1, 0, \cdots, 0)'(N_r - 1$ copies of 0), B_H is the diagonal matrix with diagonal vector $b_H = (b_{H,s}')'_{s=0,1,\cdots,r}$ and $b_{H,s} = (h_H^{|\pi_s(k)|})_{k=1,2,\ldots,M_s}$. By simplifying this equation, it establishes the second part and hence completes the whole proof.

□

S.2.3 Lemma S.3

Lemma S.3 shows the large sample properties of the estimators of partial integrations $\zeta_k(\cdot)$ for $k = 1, 2$. It establishes the uniform convergence rate and asymptotic representation of the estimators $\widehat{\zeta}_k(\cdot)$'s for $k = 1, 2$. In particular, the asymptotic representation decomposes the difference between the estimator and true value of $\zeta_k(\cdot)$ (i.e. $\widehat{\zeta}_k - \zeta_k$) into a weighted sum of i.i.d. quantities (with a mean of 0) and a bias term $h_H^r D_k(x^k)$ for $k = 1, 2$ up to some higher order error.

Lemma S.3. *Let Assumptions 1-6 hold. Then for any $k = 1, 2$, as $n \to \infty$, (i) $\sup_{x^k \in \mathcal{S}_{X^k}} \left|\widehat{\zeta}_k(x^k) - \zeta_k(x^k)\right| = O(\xi_{Hk})$ in probability with $\xi_{Hk} = h_H^{r+1} + \sqrt{log(n)/(nh_H^{d_k})}$. (ii) Moreover, for any $x^k \in \mathcal{S}_{X^k}$, $\widehat{\zeta}_k(x^k) - \zeta_k(x^k)$ can be written as*

$$\widehat{\zeta}_k(x^k) - \zeta_k(x^k) = J_{nk}(x^k) - E[J_{nk}(x^k)] + h_H^{r+1} \cdot D_k(x^k) + o_p(h_H^{r+1}),$$

where $J_{nk}(x^k)$ and $D_k(x^k)$ are defined respectively by (A.2) and (A.3).

Proof. Only $\zeta_1(\cdot)$ part is shown here. The $\zeta_2(\cdot)$ part can be shown similarly. Let $W = (Y, X^1, X^2)$. Apply Lemma 1 of Horowitz (1998) (or Theorem 2.37 of Pollard (1984)) and Lemma S.2 to obtain

$$\begin{aligned}
&\widehat{\zeta}_1(x^1) - \zeta_1(x^1) \\
=&\frac{1}{n}\sum_{j=1}^{n}\left[\widehat{H}_{-j}(x^1, X_j^2) - H(x^1, X_j^2)\right] + o\Big(\frac{log(n)}{\sqrt{n}}\Big) \\
=&\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}\frac{1}{h_H^d}\Big(Y_i - \mu(X_i^1 - x^1, X_i^2 - X_j^2)'\alpha(x^1, X_j^2)\Big)K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)K_2\Big(\frac{X_j^2 - X_i^2}{h_H}\Big) \\
&\cdot\Big(e_1' S_{n,r}(x^1, X_j^2)^{-1}\mu(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H})\Big) + O(\xi_H^2) + o\Big(\frac{log(n)}{\sqrt{n}}\Big) \\
=&\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}\psi_1(W_i, W_j) + \frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}\psi_2(W_i, W_j) + O(\xi_H^2) + o\Big(\frac{log(n)}{\sqrt{n}}\Big) \\
=&\frac{1}{n}\sum_{i=1}^{n}E\big[\psi_1(W_i, W_j)|W_i\big] + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\psi_1(W_i, W_j) - E\big[\psi_1(W_i, W_j)|W_i\big]\Big) \\
&+\frac{1}{n}\sum_{j=1}^{n}E\big[\psi_2(W_i, W_j)|W_j\big] + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Big(\psi_2(W_i, W_j) - E\big[\psi_2(W_i, W_j)|W_j\big]\Big) \\
&+O(\xi_H^2) + o\Big(\frac{log(n)}{\sqrt{n}}\Big) \\
=:&T_{1n} + T_{2n} + T_{3n} + T_{4n} + O(\xi_H^2) + o\Big(\frac{log(n)}{\sqrt{n}}\Big)
\end{aligned} \tag{1}$$

as $n \to \infty$ in probability uniformly over $x^1 \in \mathcal{S}_{X^1}$, where $\widehat{H}_{-j}(x^1, X_j^2)$ is a leave-one-out local polynomial estimator and

$$\begin{aligned}
\psi_1(W_i, W_j) =&\frac{1}{h_H^d}\Big(Y_i - H(X_i)\Big)K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)K_2\Big(\frac{X_j^2 - X_i^2}{h_H}\Big) \\
&\cdot\Big(e_1' S_{n,r}(x^1, X_j^2)^{-1}\mu(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H})\Big),
\end{aligned}$$

$$\psi_2(W_i,W_j) = \frac{1}{h_H^d}\Big(H(X_i) - \mu\big(X_i^1 - x^1, X_i^2 - X_j^2\big)'\alpha(x^1, X_j^2)\Big) K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big) K_2\Big(\frac{X_j^2 - X_i^2}{h_H}\Big)$$
$$\cdot\Big(e_1' S_{n,r}(x^1, X_j^2)^{-1}\mu\big(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H}\big)\Big).$$

The rest of the proof establishes the asymptotic representation of T_{1n}, T_{2n}, T_{3n}, and T_{4n}. It is accomplished in four steps. The asymptotic representation of T_{1n} characterizes the stochastic leading term, and T_{3n} characterizes the leading bias term.

Step 1. For T_{1n},

$$\begin{aligned}
&E[\psi_1(W_i,W_j)|W_i] \\
=&\frac{1}{h_H^{d_1}} K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)\big(Y_i - H(X_i)\big) \\
&\qquad \cdot \int \Big(e_1' S_{n,r}(x^1, X_j^2)^{-1}\mu\big(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H}\big)\Big)\frac{1}{h_H^{d_2}} K_2\Big(\frac{X_j^2 - X_i^2}{h_H}\Big) p_{X^2}(X_j^2) dX_j^2 \\
=&\frac{1}{h_H^{d_1}} K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)\big(Y_i - H(X_i)\big) e_1'\big(S_r^{-1} + O(h_H)\big) \\
&\qquad \cdot \int \mu\big(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H}\big)\frac{1}{h_H^{d_1}} K_2\Big(\frac{X_j^2 - X_i^2}{h_H}\Big)\frac{p_{X^2}(X_j^2)}{p_X(x^1, X_j^2)} dX_j^2 \\
=&\frac{1}{h_H^{d_1}} K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)\frac{Y_i - H(X_i)}{p_{X^1|X^2}(x^1|X_i^2)} e_1' S_r^{-1} V_1^{\mu}\Big(\frac{x^1 - X_i^1}{h_H}\Big)\{1 + O(h_H)\} \\
=&\frac{1}{h_H^{d_1}} K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)\frac{Y_i - H(X_i)}{p_{X^1|X^2}(x^1|X_i^2)} e_1' S_r^{-1} V_1^{\mu}\Big(\frac{x^1 - X_i^1}{h_H}\Big) + O(h_H \xi_{H1}),
\end{aligned}$$

where the last second equation is given by change of variable and first order Taylor expansion, and the last equation is based on the proof of Theorem 6 in Masry (1996). Thus, T_{1n} can be written as

$$T_{1n} = \frac{1}{n h_H^{d_1}} \sum_{i=1}^n K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)\frac{Y_i - H(X_i)}{p_{X^1|X^2}(x^1|X_i^2)} e_1' S_r^{-1} V_1^{\mu}\Big(\frac{x^1 - X_i^1}{h_H}\Big) + O(h_H \xi_{H1}).$$

Step 2. For T_{2n}, it can be decomposed as

$$\begin{aligned}
T_{2n} =&\frac{1}{n(n-1)}\sum_{i=1}^n \sum_{j\neq i}\Big(\psi_1(W_i,W_j) - E\big[\psi_1(W_i,W_j)|W_i\big]\Big) \\
=&\frac{1}{n h_H^d}\sum_{i=1}^n K_1\Big(\frac{x^1 - X_i^1}{h_H}\Big)\frac{1}{n-1}\sum_{j\neq i}\Big(\tilde{\psi}_1(W_i,W_j) - E\big[\tilde{\psi}_1(W_i,W_j)|W_i\big]\Big) \\
=&o\Big(\frac{log(n)^2}{n h_H^{d/2}}\Big),
\end{aligned}$$

where $\tilde{\psi}_1(W_i,W_j) = \frac{1}{h_H^{d_2}} K_2\Big(\frac{X_j^2 - X_i^2}{h_H}\Big)\big(Y_i - H(x_i)\big)\Big(e_1' S_{n,p}(x^1, X_j^2)^{-1}\mu\big(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H}\big)\Big)$, and the last equality is obtained by applying Lemma 1 of Horowitz (1998) (or Theorem 2.37 of Pollard (1984)).[2]

[2]A similar argument is used by Horowitz (1998) to establish its (C.5).

Step 3. For T_{3n}, the summand $E[\psi(W_i, W_j)|W_j]$ can be simplified as

$$
\begin{aligned}
&E[\psi_2(W_i, W_j)|W_j] \\
=&E\left[\frac{1}{h_H^d}\Big(H(X_i)-\mu(X_i^1 - x^1, X_i^2 - X_j^2)'\alpha(x^1, X_j^2)\Big)K_1\Big(\frac{x^1-X_i^1}{h_H}\Big)K_2\Big(\frac{X_j^2-X_i^2}{h_H}\Big)\right.\\
&\left.\cdot\Big(e_1'S_{n,r}(x^1, X_j^2)^{-1}\mu\big(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H}\big)\Big)\right| W_j\Big] \\
=&e_1'S_{n,r}(x^1, X_j^2)^{-1}\int\frac{1}{h_H^d}\Big(H(X_i)-\mu(X_i^1 - x^1, X_i^2 - X_j^2)'\alpha(x^1, X_j^2)\Big)\\
&\qquad\cdot K_1\Big(\frac{x^1-X_i^1}{h_H}\Big)K_2\Big(\frac{X_j^2-X_i^2}{h_H}\Big)\mu\big(\frac{X_i^1 - x^1}{h_H}, \frac{X_i^2 - X_j^2}{h_H}\big)p_X(X_i)dX_i \\
=&e_1'S_{n,r}(x^1, X_j^2)^{-1}\sum_{|\mathbf{s}|=r+1}\frac{1}{\mathbf{s}!}D^{\mathbf{s}}H(x^1, X_j^2)\int u^{\mathbf{s}}\mu(u)p_X(X_j+h_H u)du\cdot h_H^{r+1}+o(h_H^{r+1}) \\
=&e_1'S_{n,r}(x^1, X_j^2)^{-1}\Big(h_H^{r+1}S_{n,r}^{r+1}(x^1, X_j^2)H_{r+1}(x^1, X_j^2)+o_p(h_H^{r+1})\Big) \\
=&h_H^{r+1}e_1'S_r^{-1}S_r^{r+1}H_{r+1}(x^1, X_j^2)+o_p(h_H^{r+1}),
\end{aligned} \tag{2}
$$

where the last second equality is derived by change of variable in the integration and Taylor expansion, and the last equality is due to the approximations $S_{n,r}(x)^{-1} = \{p_X(x)\}^{-1}S_r^{-1} + O(h_H)$ and $S_{n,r}^{r+1}(x) = p_X(x)S_r^{r+1} + O(h_H)$ in the proof of Proposition 3.1 in Kong, Linton, and Xia (2010).[3] Thus, the weighted sum can be represented as

$$
\begin{aligned}
T_{3n} =&\frac{1}{n}\sum_{j=1}^{n}E[\psi_2(W_i, W_j)|W_j] \\
=&h_H^{r+1}e_1'S_r^{-1}S_r^{r+1}\Big(\frac{1}{n}\sum_{j=1}^{n}H_{r+1}(x^1, X_j^2)p_X(x^1, X_j^2)\Big)+o_p(h_H^{r+1}) \\
=&h_H^{r+1}e_1'S_r^{-1}S_r^{r+1}E\big[H_{r+1}(x^1, X^2)\big]+o_p(h_H^{r+1}).
\end{aligned}
$$

Step 4. For T_{4n}, note that

$$
\begin{aligned}
&E\left[\Big(\frac{1}{n-1}\sum_{j\neq i}\big[\psi_2(W_i, W_j) - E\big[\psi_2(W_i, W_j)|W_i\big]\Big)^2\right] \\
=&\frac{1}{n-1}E\Big[\Big(\psi_2(W_2, W_1) - E\big[\psi_2(W_2, W_1)|W_1\big]\Big)^2\Big] \\
\leq&\frac{1}{n-1}E\Big[\Big(\psi_2(W_2, W_1)\Big)^2\Big] \\
=&O\big(\frac{h_H^{2r+2}}{n}\big),
\end{aligned}
$$

where the last equality is obtained by Taylor expansion similar to Step 3. By applying Lemma 1 of

[3] When r is even, $e_1'S_r^{-1}S_r^{r+1} = 0$ and thus the first term on the right hand side of the last equality (2) vanishes. In this case, the bias term is of order $O(h^{r+2})$ if we further assume that all functions and densities are $(r+2)$ continuously differentiable.

Horowitz (1998) (or Theorem 2.37 of Pollard (1984)), we derive

$$
\begin{aligned}
T_{4n} =& \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \Big[\psi_2(W_i, W_j) - E\big[\psi_2(W_i, W_j)|W_j\big]\Big] \\
=& \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n-1} \sum_{j \neq i} \Big[\psi_2(W_i, W_j) - E\big[\psi_2(W_i, W_j)|W_i\big]\Big] \\
=& o\Big(h_H^{r+1} \frac{log(n)}{n}\Big).
\end{aligned}
$$

With the bandwidths satisfying Assumption 6, combining steps 1-4 yields

$$
\begin{aligned}
& \widehat{\zeta}_1(x^1) - \zeta_1(x^1) \\
=& T_{1n} + T_{2n} + T_{3n} + T_{4n} + O(\xi_H^2) + o\Big(\frac{log(n)}{\sqrt{n}}\Big) \\
=& \frac{1}{nh_H^{d_1}} \sum_{i=1}^{n} K_1\big(\frac{x^1 - X_i^1}{h_H}\big) \cdot \frac{Y_i - H(X_i)}{p_{X^1|X^2}(x^1|X_i^2)} e_1' S_r^{-1} V_1^{\mu}\Big(\frac{x^1 - X_i^1}{h_H}\Big) \\
& + e_1' S_r^{-1} S_r^{r+1} E\big[H_{r+1}(x^1, X^2)\big] h_H^{r+1} \\
& + O(\xi_H^2 + h_H \xi_{H1}) + o\Big(\frac{log(n)}{\sqrt{n}} + \frac{log(n)^2}{nh_H^{d/2}} + h_H^{r+1} + h_H^{r+1} \frac{log(n)}{n}\Big) \\
=& J_{n1}(x^1) + D_1(x^1) h_H^{r+1} + o(h_H^{r+1})
\end{aligned}
$$

in probability as $n \to \infty$ uniformly over $x^1 \in \mathcal{S}_{X^1}$, where the first term on the right hand side of the last equality is given by the definition of $J_{n1}(x^1)$ in (A.2), and $E[J_{n1}(x^1)] = 0$.[4] The asymptotic representation of $\widehat{\zeta}_1(x^1) - \zeta_1(x^1)$ is hence established.

Following an idea similar to the proof of Theorem 6 of Masry (1996), we have

$$
\sup_{x^1 \in \mathcal{S}_{X^1}} \left| \frac{1}{nh_H^{d_1}} \sum_{i=1}^{n} K_1\big(\frac{x^1 - X_i^1}{h_H}\big) \cdot \frac{Y_i - H(X_i)}{p_{X^1|X^2}(x^1|X_i^2)} e_1' S_p^{-1} V^{\mu}\Big(\frac{x^1 - X_i^1}{h_H}\Big) \right| = O\Big(\sqrt{\frac{log(n)}{nh_H^{d_1}}}\Big) \tag{3}
$$

in probability as $n \to \infty$. Based on the asymptotic representation, this implies that $\sup_{x^1 \in \mathcal{S}_{X^1}} |\widehat{\zeta}_1(x^1) - \zeta_1(x^1)| = O\big(h_H^{r+1} + \sqrt{\frac{log(n)}{nh_H^{d_1}}}\big) = O\big(\xi_{H1}\big)$ in probability as $n \to \infty$. This completes the proof. □

S.2.4 Lemma S.4

Lemma S.4 characterizes the uniform convergence rate and asymptotic representation of the Local linear estimators. There are three terms in the leading part (excluding all higher order remainders) of the difference $\widehat{\mathcal{H}}(z) - \mathcal{H}(z)$. The first term in the asymptotic representation is the oracle term with true $\zeta_1(\cdot)$ and $\zeta_2(\cdot)$. The second and third terms represent the error by estimating $\zeta_1(\cdot)$ and $\zeta_2(\cdot)$, respectively. Note that $\xi_{H1} \geq \xi_{H2} > 0$ due to $d_1 \geq d_2$. This implies that $O(\xi_{H1} + \xi_{H2}) = O(\xi_{H1})$.

[4]It is easy to obtain $E[J_{n1}(x^1)] = E\Big[\frac{1}{nh_H^{d_1}} K_1\big(\frac{x^1 - X_i^1}{h_H}\big) \cdot \frac{1}{p_{X^1|X^2}(x^1|X_i^2)} e_1' S_r^{-1} V_1^{\mu}\Big(\frac{x^1 - X_i^1}{h_H}\Big) E[Y_i - H(X_i)|X_i]\Big] = 0$ by the law of iterative expectations.

Lemma S.4. *Suppose that Assumptions 1-6 hold. Then*

$$\sup_{z\in\mathcal{S}_Z}\left|B_{\mathcal{H}}\big(\widehat{\beta}(z)-\beta(z)\big)\right| = O(\xi_{\mathcal{H}}+\xi_{H1})$$

in probability as $n\to\infty$. Moreover, the asymptotic representation of $\widehat{\beta}(z)-\beta(z)$ is given by

$$\begin{aligned}
&B_{\mathcal{H}}\big(\widehat{\beta}(z)-\beta(z)\big)\\
=&\frac{1}{nh_{\mathcal{H}}^2}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}k\Big(\frac{z^1-\zeta_1(X_i^1)}{h_{\mathcal{H}}}\Big)k\Big(\frac{z^2-\zeta_2(X_i^2)}{h_{\mathcal{H}}}\Big)\Big\{Y_i-\tilde{\mu}\big(\zeta(X_i)-z\big)'\beta(z)\Big\}\mu\Big(\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}\Big)\\
&+\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}\left[\Big(\frac{\partial}{\partial u^1}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_1\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\right]\big(\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)\big)\\
&+\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}\left[\Big(\frac{\partial}{\partial u^2}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_2\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\right]\big(\widehat{\zeta}_2(X_i^2)-\zeta_2(X_i^2)\big)\\
&+O(\xi_{\mathcal{H}}^2+\xi_{H1}^2)
\end{aligned}$$

as $n\to\infty$ in probability uniformly over $z\in\mathcal{S}_Z$, where $\widehat{\beta}(z)$ is the r-th order local polynomial estimator of true value $\beta(z)$, $u=(u^1,u^2)$, and $t(u,Y_i;z)=\tilde{\mu}(u)\big(Y_i-\tilde{\mu}(u)'B_{\mathcal{H}}\beta(z)\big)$. $B_{\mathcal{H}}$, $\tilde{S}_{n,r}(z)$ and $\tilde{\mu}(u)$ are defined in Appendix S.1.

Proof. Note that

$$B_{\mathcal{H}}\big(\widehat{\beta}(z)-\beta(z)\big)=B_{\mathcal{H}}\big(\widehat{\beta}(z)-\tilde{\beta}(z)\big)+B_{\mathcal{H}}\big(\tilde{\beta}(z)-\beta(z)\big).$$

First we consider $B_{\mathcal{H}}\big(\widehat{\beta}(z)-\tilde{\beta}(z)\big)$.

$$\begin{aligned}B_{\mathcal{H}}\big(\widehat{\beta}(z)-\tilde{\beta}(z)\big)=&[\mathcal{S}_{n,r}(z,\widehat{\zeta})^{-1}-\mathcal{S}_{n,r}(z,\zeta)^{-1}]\mathcal{Q}_{n,r}(z,\zeta)+\mathcal{S}_{n,r}(z,\zeta)^{-1}[\mathcal{Q}_{n,r}(z,\widehat{\zeta})-\mathcal{Q}_{n}(z,\zeta)]\\
&+[\mathcal{S}_{n,r}(z,\widehat{\zeta})^{-1}-\mathcal{S}_{n,r}(z,\zeta)^{-1}]\cdot[\mathcal{Q}_{n,r}(z,\widehat{\zeta})-\mathcal{Q}_{n,r}(z,\zeta)].\end{aligned}\tag{4}$$

$\mathcal{S}_{n,r}(\cdot)$ and $\mathcal{Q}_{n,r}(\cdot)$ are defined in Appendix S.1. As for $\mathcal{Q}_{n,r}(z,\widehat{\zeta})-\mathcal{Q}_{n,r}(z,\zeta)$, we apply Taylor expansion. By Lemma 1 of Horowitz (1998) (or Theorem 2.37 of Pollard (1984)) and our Lemma S.3,

$$\mathcal{Q}_{n,r}(z,\widehat{\zeta})-\mathcal{Q}_{n,r}(z,\zeta)=\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n\Big\{D_\zeta\mathcal{Q}_{in}^1(z,\zeta)\big(\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)\big)+D_\zeta\mathcal{Q}_{in}^2(z,\zeta)\big(\widehat{\zeta}_2(X_i^2)-\zeta_2(X_i^2)\big)\Big\}+O(\xi_{H1}^2)\tag{5}$$

and $\mathcal{Q}_{n,r}(z,\widehat{\zeta})-\mathcal{Q}_{n,r}(z,\zeta)=O(\xi_{H1})$ in probability as $n\to\infty$ uniformly over $z\in\mathcal{S}_Z$, where $D_\zeta\mathcal{Q}_{in}^1(z,\zeta)$ is an $\tilde{N}_r\times 1$ vector with

$$\begin{aligned}&\big[D_\zeta\mathcal{Q}_{in}^1(z,\zeta)\big]_l=\\
&\begin{cases}Y_i\big(\frac{\zeta_2(X_i^2)-z^2}{h_{\mathcal{H}}}\big)^{r_2}\partial_1\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big), & r_1=0\\
Y_i\big(\frac{\zeta_2(X_i^2)-z^2}{h_{\mathcal{H}}}\big)^{r_2}\Big[\partial_1\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\big(\frac{\zeta_1(X_i^1)-z^1}{h_{\mathcal{H}}}\big)^{r_1}+r_1\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\big(\frac{\zeta_1(X_i^1)-z^1}{h_{\mathcal{H}}}\big)^{r_1-1}\Big], & r_1\geq 1\end{cases},\end{aligned}$$

where $\tilde{r} = (r_1, r_2)$ is the correspondent power numbers of the l-th entry of $\mathcal{Q}_{n,r}(z,\zeta)$, i.e.

$$\big[\mathcal{Q}_{n,r}(z,\zeta)\big]_l = \frac{1}{nh_{\mathcal{H}}^2}\sum_{i=1}^{n}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)^{\tilde{r}}\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big).$$

Similarly, we can define $D_\zeta \mathcal{Q}_{in}^2(z,\zeta)$.
As for $\mathcal{S}_n(z,\widehat{\zeta})^{-1} - \mathcal{S}_n(z,\zeta)^{-1}$, Similarly, we can derive that

$$\mathcal{S}_n(z,\widehat{\zeta}) - \mathcal{S}_n(z,\zeta) = \frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^{n}\Big\{D_\zeta \mathcal{S}_{in}^1(z,\zeta)\big(\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)\big) + D_\zeta \mathcal{S}_{in}^2(z,\zeta)\big(\widehat{\zeta}_2(X_i)-\zeta_2(X_i)\big)\Big\} + O(\xi_{H1}^2),$$

and $\mathcal{S}_n(z,\widehat{\zeta}) - \mathcal{S}_n(z,\zeta) = O(\xi_{H1})$ as $n\to\infty$ in probability uniformly over $z\in\mathcal{S}_Z$, where matrix $D_\zeta \mathcal{S}_{in}^v(z,\zeta)$ ($v=1,2$) satisfies that its (l,k)-entry ($l,k=1,2,\ldots,N_r$) is

$$\begin{aligned}&\big[D_\zeta \mathcal{S}_{in}^1(z,\zeta)\big]_{lk} = \\ &\begin{cases}\big(\frac{\zeta_2(X_i^2)-z^2}{h_{\mathcal{H}}}\big)^{r_2}\partial_1\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big), & r_1 = 0\\ \big(\frac{\zeta_2(X_i^2)-z^2}{h_{\mathcal{H}}}\big)^{r_2}\Big[\partial_1\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\big(\frac{\zeta_1(X_i^1)-z^1}{h_{\mathcal{H}}}\big)^{r_1} + r_1\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\big(\frac{\zeta_1(X_i^1)-z^1}{h_{\mathcal{H}}}\big)^{r_1-1}\Big], & r_1\ge 1\end{cases},\end{aligned}$$

where $\tilde{r} = (r_1,r_2)$ is the power number of the (l,k)-element of $\mathcal{S}_n(z,\zeta)$. Similarly, we can define $D_\zeta \mathcal{S}_{in}^2(z,\zeta)$. Similar to the arguments in the proof of Theorem 3.2 in Kong, Linton, and Xia (2010), we have $\sup_{z\in\mathcal{S}_Z}|\mathcal{S}_n(z,\zeta) - \tilde{S}_{n,r}(z)| = O(\xi_{\mathcal{H}})$ as $n\to\infty$ in probability. Thus, the triangular inequality implies that

$$\mathcal{S}_n(z,\widehat{\zeta}) - \tilde{S}_{n,r}(z) = O(\xi_{H1}+\xi_{\mathcal{H}})$$

as $n\to\infty$ in probability uniformly over $z\in\mathcal{S}_Z$. Therefore, we can derive that

$$\begin{aligned}&\mathcal{S}_n(z,\widehat{\zeta})^{-1} - \mathcal{S}_n(z,\zeta)^{-1}\\ =& -\mathcal{S}_n(z,\widehat{\zeta})^{-1}\Big(\mathcal{S}_n(z,\widehat{\zeta}) - \mathcal{S}_n(z,\zeta)\Big)\mathcal{S}_n(z,\zeta)^{-1}\\ =& -\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^{n}\tilde{S}_{n,r}(z)^{-1}\Big\{D_\zeta \mathcal{S}_{in}^1(z,\zeta)\big(\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)\big) + D_\zeta \mathcal{S}_{in}^2(z,\zeta)\big(\widehat{\zeta}_2(X_i^2)-\zeta_2(X_i^2)\big)\Big\}\tilde{S}_{n,r}(z)^{-1}\\ &+O\big(\xi_{\mathcal{H}}\cdot\xi_{H1}+\xi_{H1}^2\big)\end{aligned} \tag{6}$$

and $\mathcal{S}_n(z,\widehat{\zeta})^{-1} - \mathcal{S}_n(z,\zeta)^{-1} = O(\xi_{H1})$ as $n\to\infty$ in probability uniformly over $z\in\mathcal{S}_Z$. Also, we have $\mathcal{Q}_n(z,\zeta) = \mathcal{S}_n(z,\zeta)B_{\mathcal{H}}\tilde{\beta}(z)$. By Theorem 6 in Masry (1996), $\sup_{z\in\mathcal{S}_Z}\big|B_{\mathcal{H}}\big(\tilde{\beta}(z)-\beta(z)\big)\big| = O(\xi_{\mathcal{H}})$ in probability as $n\to\infty$. Therefore, we have

$$\mathcal{Q}_n(z,\zeta) - \tilde{S}_{n,r}(z)B_{\mathcal{H}}\beta(z) = O(\xi_{\mathcal{H}}) \tag{7}$$

as $n\to\infty$ in probability uniformly over $z\in\mathcal{S}_Z$. According to (5), (6) and (7), (4) can be rewritten as

$$\begin{aligned}&B_{\mathcal{H}}\big(\widehat{\beta}(z)-\tilde{\beta}(z)\big)\\ =&[\mathcal{S}_n(z,\widehat{\zeta})^{-1} - \mathcal{S}_n(z,\zeta)^{-1}]\big(\tilde{S}_{n,r}(z)B_{\mathcal{H}}\beta(z)+O(\xi_{\mathcal{H}})\big) + \big(\tilde{S}_{n,r}(z)^{-1}+O(\xi_{\mathcal{H}})\big)[\mathcal{Q}_n(z,\widehat{\zeta}) - \mathcal{Q}_n(z,\zeta)]\\ &+O(\xi_{H1}^2)\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}\big[-D_\zeta \mathcal{S}_{in}^1(z,\zeta)B_{\mathcal{H}}\beta(z)+D_\zeta \mathcal{Q}_{in}^1(z,\zeta)\big]\big(\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)\big)\\
&+\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}\big[-D_\zeta \mathcal{S}_{in}^2(z,\zeta)B_{\mathcal{H}}\beta(z)+D_\zeta \mathcal{Q}_{in}^2(z,\zeta)\big]\big(\widehat{\zeta}_2(X_i^2)-\zeta_2(X_i^2)\big)\\
&+O\big(\xi_{\mathcal{H}}\cdot\xi_{H1}+\xi_{H1}^2\big)\\
=&\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}\bigg[\Big(\frac{\partial}{\partial u^1}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_1\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\bigg]\big(\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)\big)\\
&\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}\bigg[\Big(\frac{\partial}{\partial u^2}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_2\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\bigg]\big(\widehat{\zeta}_2(X_i^2)-\zeta_2(X_i^2)\big)\\
&+O\big(\xi_{\mathcal{H}}^2+\xi_{H1}^2\big) \qquad (8)
\end{aligned}
$$

and $B_{\mathcal{H}}\big(\widehat{\beta}(z)-\tilde{\beta}(z)\big)=O(\xi_{H1})$ in probability as $n\to\infty$ uniformly over $z\in\mathcal{S}_Z$.
Second we consider $B_{\mathcal{H}}\big(\tilde{\beta}(z)-\beta(z)\big)$. The asymptotic linear representation is a direct application of Theorem 3.2 in Kong, Linton, and Xia (2010), that is,

$$
\begin{aligned}
&B_{\mathcal{H}}\big(\tilde{\beta}(z)-\beta(z)\big)\\
=&\frac{1}{nh_{\mathcal{H}}^2}\sum_{i=1}^n \tilde{S}_{n,r}(z)^{-1}k\big(\frac{z^1-\zeta_1(X_i^1)}{h_{\mathcal{H}}}\big)k\big(\frac{z^2-\zeta_2(X_i^2)}{h_{\mathcal{H}}}\big)\Big\{Y_i-\tilde{\mu}\big(\zeta(X_i)-z\big)'\beta(z)\Big\}\mu\Big(\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}\Big)\\
&+O(\xi_{\mathcal{H}}^2) \qquad (9)
\end{aligned}
$$

and $B_{\mathcal{H}}\big(\tilde{\beta}(z)-\beta(z)\big)=O(\xi_{\mathcal{H}})$ in probability as $n\to\infty$ uniformly over $z\in\mathcal{S}_Z$. Finally, the desired representation of $B_{\mathcal{H}}\big(\widehat{\beta}(\cdot)-\beta(\cdot)\big)$ can then be established by (8) and (9). □

S.2.5 Lemma S.5

$\partial_k\widehat{\mathcal{H}}(z)$ is the r-th order local polynomial estimator of first derivatives $\partial_k\mathcal{H}(z)$ ($k=1,2$) based on data $\{Y_i,\widehat{\zeta}_1(X_i),\widehat{\zeta}_2(X_i)\}_{i=1}^n$, while $\partial_k\tilde{\mathcal{H}}(z)$ is the infeasible version with data $\{Y_i,\zeta_1(X_i),\zeta_2(X_i)\}_{i=1}^n$. Lemma S.5 studies the asymptotic properties of $\partial_k\widehat{\mathcal{H}}(z)$. It shows the uniform convergence and asymptotic representation of such statistics. Particularly, the first two terms in the asymptotic representation come from the (asymptotic) representation of infeasible estimator $\partial_k\tilde{\mathcal{H}}(z)$, while the third term is the additional bias appearing in the difference between feasible and infeasible estimators, namely $\partial_k\widehat{\mathcal{H}}(z)-\partial_k\tilde{\mathcal{H}}(z)$.

Lemma S.5. *Suppose that Assumptions 1-6 hold. Then for $k=1,2$, (i) $\sup_{z\in\mathcal{S}_Z}\big|\partial_k\widehat{\mathcal{H}}(z)-\partial_k\mathcal{H}(z)\big|=O\big(\xi'_{\mathcal{H}}+\xi_{H1}\big)$ in probability as $n\to\infty$; (ii) $\partial_k\widehat{\mathcal{H}}(z)-\partial_k\mathcal{H}(z)$ has an asymptotic representation as*

$$
\begin{aligned}
\begin{bmatrix}\widehat{\partial_1\mathcal{H}}(z)\\ \widehat{\partial_2\mathcal{H}}(z)\end{bmatrix}-\begin{bmatrix}\partial_1\mathcal{H}(z)\\ \partial_2\mathcal{H}(z)\end{bmatrix}=&\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n e_d'\tilde{S}_{n,r}(z)^{-1}\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\big(Y_i-H(X_i)\big)\mu\Big(\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}\Big)+\mathfrak{D}(z)h_{\mathcal{H}}^r+\mathcal{D}(z)h_H^{r+1}\\
&+o\big(h_{\mathcal{H}}^r+h_H^{r+1}\big)+O\Big(\frac{log(n)}{nh_{\mathcal{H}}^4}+\frac{log(n)}{nh_H^{d_1}}\Big)
\end{aligned}
$$

in probability as $n\to\infty$ uniformly over $z\in\mathcal{S}_Z$.

Proof. Note that

$$\begin{bmatrix}\widehat{\partial_1 \mathcal{H}}(z)\\ \widehat{\partial_2 \mathcal{H}}(z)\end{bmatrix} - \begin{bmatrix}\partial_1 \mathcal{H}(z)\\ \partial_2 \mathcal{H}(z)\end{bmatrix} = e_d'\big(\widehat{\beta}(z)-\beta(z)\big) = e_d' B_{\mathcal{H}}^{-1}\cdot B_{\mathcal{H}}\big(\widehat{\beta}(z)-\beta(z)\big)$$
$$=\frac{1}{h_{\mathcal{H}}} e_d' \cdot B_{\mathcal{H}}\big(\widehat{\beta}(z)-\beta(z)\big).$$

Thus, part (i) is trivial by Lemma S.5. To find the asymptotic representation, we just need to further decompose $\widehat{\beta}(z)-\beta(z)$. According to Lemma S.4, we just need to derive the asymptotic representation of the following three parts,

$$A_{1n}(z) = \frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n e_d'\tilde{S}_{n,r}(z)^{-1}\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\Big\{Y_i - \tilde{\mu}\big(z-\zeta(X_i)\big)'\beta(z)\Big\}\mu\Big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\Big);$$
$$A_{2n}(z) = \frac{1}{nh_{\mathcal{H}}^4}\sum_{i=1}^n e_d'\tilde{S}_{n,r}(z)^{-1}\bigg[\Big(\frac{\partial}{\partial u^1}t(u,Y_i;z)\tilde{K}(u) + t(u,Y_i;z)\partial_1\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\bigg]\big(\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)\big);$$
$$A_{3n}(z) = \frac{1}{nh_{\mathcal{H}}^4}\sum_{i=1}^n e_d'\tilde{S}_{n,r}(z)^{-1}\bigg[\Big(\frac{\partial}{\partial u^2}t(u,Y_i;z)\tilde{K}(u) + t(u,Y_i;z)\partial_2\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\bigg]\big(\widehat{\zeta}_2(X_i^2)-\zeta_2(X_i^2)\big).$$

First we consider $A_{1n}(z)$,

$$\begin{aligned}A_{1n}(z) =& \frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n e_d'\tilde{S}_{n,r}(z)^{-1}\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\Big\{\mathcal{H}\big(\zeta(X_i)\big) - \tilde{\mu}\big(z-\zeta(X_i)\big)'\beta(z)\Big\}\mu\Big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\Big)\\ &+\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n e_d'\tilde{S}_{n,r}(z)^{-1}\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\Big\{Y_i - \mathcal{H}\big(\zeta(X_i)\big)\Big\}\mu\Big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\Big)\\ =&:A_{11n}(z)+A_{12n}(z). \end{aligned}\tag{10}$$

As for $A_{11n}(z)$, note that

$$\begin{aligned}&E\big[A_{11n}(z)\big]\\ =&\frac{1}{h_{\mathcal{H}}^3}e_d'B_{\mathcal{H}}^{-1}\tilde{S}_{n,r}(z)^{-1}\int\tilde{K}\big(\frac{z-Z}{h_{\mathcal{H}}}\big)\Big\{\mathcal{H}(Z)-\tilde{\mu}(z-Z)'\beta(z)\Big\}\mu\Big(\frac{z-Z}{h_{\mathcal{H}}}\Big)p_Z(Z)dZ\\ =&e_d'\tilde{S}_{n,r}(z)^{-1}\sum_{|\mathbf{j}|=r+1}\frac{1}{\mathbf{j}!}D^{\mathbf{j}}\mathcal{H}(z)\int u^{\mathbf{j}}\tilde{\mu}(u)\tilde{K}(u)p_Z(z+h_{\mathcal{H}}u)du\cdot h_{\mathcal{H}}^r + o(h^r)\\ =&e_d'\tilde{S}_{n,r}(z)^{-1}\tilde{S}_{n,r}^{r+1}(z)\mathcal{H}_{r+1}(z)\cdot h_{\mathcal{H}}^r + o(h_{\mathcal{H}}^r)\\ =&e_d'\tilde{S}_r^{-1}\tilde{S}_r^{r+1}\mathcal{H}_{r+1}(z)\cdot h_{\mathcal{H}}^r + o(h_{\mathcal{H}}^r),\end{aligned}$$

where the second equality is derived by change of variables and Taylor expansion, and the last equality is due to the approximations $\tilde{S}_{n,r}(z)^{-1} = \tilde{S}_r^{-1}p_Z(z)^{-1}+O(h_{\mathcal{H}})$ and $\tilde{S}_{n,r}^{r+1}(z) = \tilde{S}_r^{r+1}p_Z(z)+O(h_{\mathcal{H}})$ in the proof of Proposition 3.1 in Kong, Linton, and Xia (2010). Also, by Lemma 1 of Horowitz (1998) (or Theorem 2.37 of Pollard (1984)), we derive

$$\begin{aligned}&A_{11n}(z)\\ =&E[A_{11n}(z)] + \big(A_{11n}(z)-E[A_{11n}(z)]\big)\end{aligned}$$

$$=e_d'\tilde{S}_r^{-1}\tilde{S}_r^{r+1}\mathcal{H}_{r+1}(z)\cdot h_{\mathcal{H}}^r+o(h_{\mathcal{H}}^r)+o\Big(h_{\mathcal{H}}^r\frac{log(n)}{n^{1/2}}\Big)$$
$$=e_d'\tilde{S}_r^{-1}\tilde{S}_r^{r+1}\mathcal{H}_{r+1}(z)\cdot h_{\mathcal{H}}^r+o(h_{\mathcal{H}}^r).$$

Therefore by (10), we derive

$$\begin{aligned}&A_{1n}(z)\\=&\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n e_d'\tilde{S}_{n,r}(z)^{-1}\tilde{K}\big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\big)\big(Y_i-H(X_i)\big)\mu\Big(\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}\Big)+e_d'\tilde{S}_r^{-1}\tilde{S}_r^{r+1}\mathcal{H}_{r+1}(z)\cdot h_{\mathcal{H}}^r\\&+o(h_{\mathcal{H}}^r).\end{aligned}\tag{11}$$

Second we consider $A_{2n}(z)$, plugging the asymptotic representation of $\widehat{\zeta}_1(X_i^1)-\zeta_1(X_i^1)$ given by Lemma S.3 into $A_{2n}(z)$, under Assumption 6,

$$\begin{aligned}&A_{2n}(z)\\=&e_d'\tilde{S}_{n,r}(z)^{-1}\Bigg(\frac{1}{nh_{\mathcal{H}}^4}\sum_{i=1}^n\Bigg[\Big(\frac{\partial}{\partial u^1}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_1\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\Bigg]D_1(X_i^1)\cdot h_H^{r+1}\\&+\frac{1}{nh_{\mathcal{H}}^4}\sum_{i=1}^n\Bigg[\Big(\frac{\partial}{\partial u^1}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_1\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\Bigg]J_{n1}(X_i^1)\Bigg)\\&+o(h_H^{r+1})\\=:&e_d'\tilde{S}_{n,r}(z)^{-1}\big(A_{21n}(z)+A_{22n}(z)\big)+o(h_H^{r+1}).\end{aligned}\tag{12}$$

As for $A_{21n}(z)$, note that by the product rule of derivatives

$$\begin{aligned}&\Big(\frac{\partial}{\partial u^1}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_1\tilde{K}(u)\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\\=&\frac{\partial}{\partial u^1}\Big\{t(u,Y_i;z)\tilde{K}(u)\Big\}\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\\=&\big(Y_i-\mathcal{H}(\zeta(X_i))\big)\Big(\frac{\partial}{\partial u^1}\{\tilde{\mu}(u)\tilde{K}(u)\}\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\\&+\frac{\partial}{\partial u^1}\Big\{t(u,\mathcal{H}(z+h_{\mathcal{H}}u);z)\tilde{K}(u)\Big\}\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\\&-\tilde{\mu}(u)\tilde{K}(u)\big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\cdot\frac{\partial}{\partial z^1}\mathcal{H}(z)\big|_{z=\zeta(X_i)}\cdot h_{\mathcal{H}},\end{aligned}$$

where $t(u,Y_i;z)\tilde{K}(u)=\tilde{\mu}(u)\big(Y_i-\tilde{\mu}(u)'B_{\mathcal{H}}\beta(z)\big)\tilde{K}(u)$. Thus, we can further decompose $A_{21n}(z)$ as

$$\begin{aligned}&A_{21n}(z)\\=&\frac{1}{nh_{\mathcal{H}}^4}\sum_{i=1}^n\Big(Y_i-\mathcal{H}\big(\zeta(X_i)\big)\Big)D_1(X_i^1)\Big(\frac{\partial}{\partial u^1}\{\tilde{\mu}(u)\tilde{K}(u)\}\Big)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\cdot h_H^{r+1}\\&+\frac{1}{nh_{\mathcal{H}}^4}\sum_{i=1}^n\frac{\partial}{\partial u^1}\Big\{t(u,\mathcal{H}(z+h_{\mathcal{H}}u);z)\tilde{K}(u)\Big\}\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}}\cdot D_1(X_i^1)\cdot h_H^{r+1}\end{aligned}$$

$$- \frac{1}{nh_{\mathcal{H}}^3} \sum_{i=1}^{n} \tilde{\mu}(u)\tilde{K}(u)\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}} \cdot \frac{\partial}{\partial z^1}\mathcal{H}(z)\Big|_{z=\zeta(X_i)} \cdot D_1(X_i^1) \cdot h_H^{r+1}$$
$$=: A_{211n}(z) + A_{212n}(z) + A_{213n}(z).$$

By Lemma 1 of Horowitz (1998) (or Theorem 2.37 of Pollard (1984)),

$$A_{211n}(z) = o\Big(\frac{h_H^{r+1} log(n)}{n^{1/2} h_{\mathcal{H}}^3}\Big),$$

and

$$\begin{aligned}
&A_{212n}(z)\\
=&E[A_{212n}(z)] + \Big(A_{212n}(z) - E[A_{212n}(z)]\Big)\\
=&\frac{h_H^{r+1}}{h_{\mathcal{H}}^4} \int \frac{\partial}{\partial u^1}\Big\{t(u, \mathcal{H}(z + h_{\mathcal{H}}u); z)\tilde{K}(u)\Big\}\Big|_{u=\frac{\zeta(X_i)-z}{h_{\mathcal{H}}}} E[D_1(X_i^1)|\zeta(X_i) = Z] p_Z(Z) dZ + o\Big(h_H^{r+1} \frac{h_{\mathcal{H}}^{(r-5)/2} log(n)}{n^{1/2}}\Big)\\
=&-\frac{h_H^{r+1}}{h_{\mathcal{H}}^2} \int \Big\{t(u, \mathcal{H}(z + h_{\mathcal{H}}u); z)\tilde{K}(u)\Big\} \frac{\partial}{\partial u^1} E[D_1(X_i^1)|\zeta(X_i) = z + h_{\mathcal{H}}u] p_Z(z + h_{\mathcal{H}}u) du + o\Big(\frac{h_H^{r+1} h_{\mathcal{H}}^{(r-5)/2} log(n)}{n^{1/2}}\Big)\\
=&O(h_H^{r+1} \cdot h_{\mathcal{H}}^r) + o\Big(\frac{h_H^{r+1} h_{\mathcal{H}}^{(r-5)/2} log(n)}{n^{1/2}}\Big),
\end{aligned}$$

where the last second equality is derived by change of variable and integration by parts, and the last equality is due to Taylor expansion. As for $A_{213n}(z)$, similar to $A_{212n}(z)$, by Lemma 1 of Horowitz (1998) (or Theorem 2.37 of Pollard (1984)), we have

$$\begin{aligned}
&A_{213n}(z)\\
=&E[A_{213n}(z)] + \big(A_{213n}(z) - E[A_{213n}(z)]\big)\\
=&-\frac{h_H^{r+1}}{h_{\mathcal{H}}^3} \int \tilde{\mu}(u)\tilde{K}(u)\big|_{u=\frac{Z-z}{h_{\mathcal{H}}}} \cdot \frac{\partial}{\partial z^1}\mathcal{H}(z)\Big|_{z=Z} E[D_1(X_i^1)|\zeta(X_i) = Z] p_Z(Z) dZ + o\Big(\frac{h_H^{r+1} log(n)}{n^{1/2} h_{\mathcal{H}}^2}\Big)\\
=&-V_r^{\tilde{\mu}} \frac{\partial}{\partial z^1}\mathcal{H}(z) E[D_1(X_i^1)|\zeta(X_i) = z] p_Z(z) \cdot \frac{h_H^{r+1}}{h_{\mathcal{H}}}\\
&-V_r^{\tilde{\mu}}(1) \frac{\partial}{\partial z^1}\Bigg\{\frac{\partial}{\partial z^1}\mathcal{H}(z) E[D_1(X_i^1)|\zeta(X_i) = z] p_Z(z)\Bigg\} \cdot h_H^{r+1}\\
&-V_r^{\tilde{\mu}}(2) \frac{\partial}{\partial z^2}\Bigg\{\frac{\partial}{\partial z^1}\mathcal{H}(z) E[D_1(X_i^1)|\zeta(X_i) = z] p_Z(z)\Bigg\} \cdot h_H^{r+1}\\
&+o\Big(h_H^{r+1} + \frac{h_H^{r+1} log(n)}{n^{1/2} h_{\mathcal{H}}^2}\Big),
\end{aligned}$$

where $V_r^{\tilde{\mu}} = \int \tilde{\mu}(u)\tilde{K}(u)du$, $V_r^{\tilde{\mu}}(1) = \int u^1 \tilde{\mu}(u)\tilde{K}(u)du$, and $V_r^{\tilde{\mu}}(2) = \int u^2 \tilde{\mu}(u)\tilde{K}(u)du$. Therefore, by adding up $A_{211n}(z)$, $A_{212n}(z)$, and $A_{213n}(z)$, we derive

$$e_d' \tilde{S}_{n,r}(z)^{-1} A_{21n}$$

$$
\begin{aligned}
=&e_d'\big(\tilde{S}_r^{-1}p_Z(z)^{-1}+O(h_{\mathcal{H}})\big)\cdot\Bigg(-V_r^{\tilde{\mu}}\frac{\partial}{\partial z^1}\mathcal{H}(z)E[D_1(X_i^1)|\zeta(X_i)=z]p_Z(z)\cdot h_H^{r+1}\\
&-V_r^{\tilde{\mu}}(1)p_Z(z)^{-1}\frac{\partial}{\partial z^1}\Big\{\frac{\partial}{\partial z^1}\mathcal{H}(z)E[D_1(X_i^1)|\zeta(X_i)=z]p_Z(z)\Big\}\cdot h_H^{r+1}\\
&-V_r^{\tilde{\mu}}(2)p_Z(z)^{-1}\frac{\partial}{\partial z^2}\Big\{\frac{\partial}{\partial z^1}\mathcal{H}(z)E[D_1(X_i^1)|\zeta(X_i)=z]p_Z(z)\Big\}\cdot h_H^{r+1}\Bigg)\\
&+o\Big(h_H^{r+1}+\frac{h_H^{r+1}log(n)}{n^{1/2}h_{\mathcal{H}}^3}\Big)\\
=&-\tilde{e}_1p_Z(z)^{-1}\frac{\partial}{\partial z^1}\Big\{\frac{\partial}{\partial z^1}\mathcal{H}(z)E[D_1(X_i^1)|\zeta(X_i)=z]p_Z(z)\Big\}\cdot h_H^{r+1}\\
&-\tilde{e}_2p_Z(z)^{-1}\frac{\partial}{\partial z^2}\Big\{\frac{\partial}{\partial z^1}\mathcal{H}(z)E[D_1(X_i^1)|\zeta(X_i)=z]p_Z(z)\Big\}\cdot h_H^{r+1}\\
&+o\Big(h_H^{r+1}+\frac{h_H^{r+1}log(n)}{n^{1/2}h_{\mathcal{H}}^3}\Big),
\end{aligned}
\tag{13}
$$

where the last equality is due to the facts that $\tilde{S}_r^{-1}V_r^{\tilde{\mu}}=e_1$, $\tilde{S}_r^{-1}V_r^{\tilde{\mu}}(1)=(0,1,0,\cdots,0)'$, and $\tilde{S}_r^{-1}V_r^{\tilde{\mu}}(1)=(0,0,1,\cdots,0)'$.
As for $A_{22n}(z)$, note that by (3) in Lemma S.3,

$$
\sup_{x^1\in\mathcal{S}_{X^1}}\big|J_{n1}(x^1)\big|=O\Big(\sqrt{\frac{log(n)}{nh_H^{d_1}}}\Big)
$$

in probability as $n\to\infty$, and $E[J_{n1}(X_i^1)|X_i^1=x^1]=0$. Thus by Lemma 1 of Horowitz (1998) (or Theorem 2.37 of Pollard (1984)),

$$
\begin{aligned}
&A_{22n}(z)\\
=&\frac{1}{nh_{\mathcal{H}}^4}\sum_{i=1}^n\Bigg[\Big(\frac{\partial}{\partial u^1}t(u,Y_i;z)\tilde{K}(u)+t(u,Y_i;z)\partial_1\tilde{K}(u)\Big)\Big|_{u=\frac{z-\zeta(X_i)}{h_{\mathcal{H}}}}\Bigg]J_{n1}(X_i^1)\Bigg)\\
=&o\Big(\frac{log(n)^{3/2}}{nh_{\mathcal{H}}^3h_H^{d_1/2}}\Big).
\end{aligned}
\tag{14}
$$

Plugging (13) and (14) into (12), we get

$$
\begin{aligned}
&A_{2n}(z)\\
=&-\tilde{e}_1p_Z(z)^{-1}\frac{\partial}{\partial z^1}\Big\{\frac{\partial}{\partial z^1}\mathcal{H}(z)E[D_1(X_i^1)|\zeta(X_i)=z]p_Z(z)\Big\}\cdot h_H^{r+1}\\
&-\tilde{e}_2p_Z(z)^{-1}\frac{\partial}{\partial z^2}\Big\{\frac{\partial}{\partial z^1}\mathcal{H}(z)E[D_1(X_i^1)|\zeta(X_i)=z]p_Z(z)\Big\}\cdot h_H^{r+1}\\
&+o\Big(h_H^{r+1}+\frac{h_H^{r+1}log(n)}{n^{1/2}h_{\mathcal{H}}^3}+\frac{h_H^{r+1}h_{\mathcal{H}}^{(r-5)/2}log(n)}{n^{1/2}}+\frac{log(n)^{3/2}}{nh_{\mathcal{H}}^3h_H^{d_1/2}}\Big).
\end{aligned}
\tag{15}
$$

Similarly, we have the decomposition for $A_{3n}(z)$. By adding up the representations of $A_{1n}(z)$, $A_{2n}(z)$, and $A_{3n}(z)$, The desired conclusion therefore follows from Assumption 6. □

S.2.6 Lemma S.6

Let $\check{\tilde{f}}_k(\cdot)$ be an infeasible estimator of $\tilde{f}_k(\cdot)$ with an (infeasible) data of $\{Y_i, \zeta_1(X_i^1), \zeta_2(X_i^2)\}_{i=1}^n$, while $\widehat{\tilde{f}}_k(\cdot)$ be a feasible estimator of $\tilde{f}_k(\cdot)$ with a data of $\{Y_i, \widehat{\zeta}_1(X_i^1), \widehat{\zeta}_2(X_i^2)\}_{i=1}^n$. Lemma S.6 establishes the uniform convergence rate and asymptotic representation of the feasible estimator of transformed component function $\tilde{f}_k(\cdot)$ for $k = 1, 2$. In particular, the first three terms in the asymptotic representation come from the (asymptotic) representation of infeasible estimator $\check{\tilde{f}}_k(z^k)$, while the fourth term is the additional bias appearing in the difference between feasible and infeasible estimators, namely $\widehat{\tilde{f}}_k(z^k) - \check{\tilde{f}}_k(z^k)$.

Lemma S.6. *If Assumptions 1-6 hold, then for $k = 1, 2$, (i) $\widehat{\tilde{f}}_k(z^k) - \tilde{f}_k(z^k) = \tilde{\mathfrak{I}}_{nk}(z^k) - E[\tilde{\mathfrak{I}}_{nk}(z^k)] + h_{\mathcal{H}}^r \mathfrak{B}_k(z^k) + h_H^{r+1} \tilde{\mathcal{B}}_k(z^k) + o_p(h_{\mathcal{H}}^r + h_H^{r+1})$, and (ii) $\widehat{\tilde{f}}_k(z^k) - \tilde{f}_k(z^k) = O_p(h_{\mathcal{H}}^r + \sqrt{\frac{log(n)}{nh_{\mathcal{H}}}} + h_H^{r+1})$ as $n \to \infty$ uniformly over $z^k \in \mathcal{S}_{Z^k}$.*

Proof. Only the case for $k = 2$ is proved. The proof for $k = 1$ is similar. The definition of $\widehat{\tilde{f}}_2(\cdot)$ yields

$$\widehat{\tilde{f}}_2(z^2) - \tilde{f}_2(z^2) = \int_{z_0^2}^{z^2} \int [\frac{\partial_2 \widehat{\mathcal{H}}(v)}{\partial_1 \widehat{\mathcal{H}}(v)} - \frac{\partial_2 \mathcal{H}(v)}{\partial_1 \mathcal{H}(v)}] \omega_3(v^1) dv^1 dv^2. \tag{16}$$

By applying Taylor expansion to the integrand,

$$\begin{aligned} & \frac{\partial_2 \widehat{\mathcal{H}}(v)}{\partial_1 \widehat{\mathcal{H}}(v)} - \frac{\partial_2 \mathcal{H}(v)}{\partial_1 \mathcal{H}(v)} \\ = & \frac{\partial_2 \widehat{\mathcal{H}}(v) - \partial_2 \mathcal{H}(v)}{\partial_1 \mathcal{H}(v)} - \frac{\partial_2 \mathcal{H}(v)}{[\partial_1 \mathcal{H}(v)]^2} [\partial_1 \widehat{\mathcal{H}}(v) - \partial_1 \mathcal{H}(v)] + O(\xi_{H1}^2 + [\xi_{\mathcal{H}}']^2) \\ = & q_2(v)' \cdot \left(\begin{bmatrix} \widehat{\partial_1 \mathcal{H}}(v) \\ \widehat{\partial_2 \mathcal{H}}(v) \end{bmatrix} - \begin{bmatrix} \partial_1 \mathcal{H}(v) \\ \partial_2 \mathcal{H}(v) \end{bmatrix} \right) + O(\xi_{H1}^2 + [\xi_{\mathcal{H}}']^2) \end{aligned} \tag{17}$$

in probability as $n \to \infty$ uniformly over $v \in \mathcal{S}_Z$, where $q_2(v) = \left[-\frac{\partial_2 \mathcal{H}(v)}{[\partial_1 \mathcal{H}(v)]^2}, \frac{1}{\partial_1 \mathcal{H}(v)} \right]'$. By Lemma S.5 and plugging the representations of $\widehat{\partial_k \mathcal{H}}(v) - \partial_k \mathcal{H}(v)$ into (17), we derive

$$\begin{aligned} \frac{\partial_2 \widehat{\mathcal{H}}(v)}{\partial_1 \widehat{\mathcal{H}}(v)} - \frac{\partial_2 \mathcal{H}(v)}{\partial_1 \mathcal{H}(v)} = & \frac{1}{nh_{\mathcal{H}}^3} q_2(v)' \sum_{i=1}^n e_d' \tilde{S}_{n,r}(v)^{-1} \tilde{K}\left(\frac{v - \zeta(X_i)}{h_{\mathcal{H}}}\right)(Y_i - H(X_i)) \mu\left(\frac{\zeta(X_i) - v}{h_{\mathcal{H}}}\right) \\ & + q_2(v)' \mathfrak{D}(v) h_{\mathcal{H}}^r + q_2(v)' \mathcal{D}(v) h^{r+1} \\ & + o\left(h_{\mathcal{H}}^r + h_H^{r+1}\right) + O\left(\frac{log(n)}{nh_{\mathcal{H}}^4} + \frac{log(n)}{nh_H^{d_1}}\right) \end{aligned} \tag{18}$$

in probability as $n \to \infty$ uniformly over $z \in \mathcal{S}_Z$. Therefore by integrating (18) and Assumption 6,

$$\begin{aligned} & \widehat{\tilde{f}}_2(z^2) - \tilde{f}_2(z^2) \\ = & \int_{z_0^2}^{z^2} \int [\frac{\partial_2 \widehat{\mathcal{H}}(v)}{\partial_1 \widehat{\mathcal{H}}(v)} - \frac{\partial_2 \mathcal{H}(v)}{\partial_1 \mathcal{H}(v)}] \omega_3(v^1) dv^1 dv^2 \end{aligned}$$

$$=\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n\int_{z_0^2}^{z^2}\int q_2(\nu)'e_d'\tilde{S}_{n,r}(\nu)^{-1}\tilde{K}\Big(\frac{\nu-\zeta(X_i)}{h_{\mathcal{H}}}\Big)\big(Y_i-H(X_i)\big)\mu\Big(\frac{\zeta(X_i)-\nu}{h_{\mathcal{H}}}\Big)\omega_3(\nu^1)d\nu^1d\nu^2$$
$$+\mathfrak{B}_2(z^2)h_{\mathcal{H}}^r+\tilde{\mathcal{B}}_2(z^2)h^{r+1}+o\big(h_{\mathcal{H}}^r+h_H^{r+1}\big). \tag{19}$$

The rest of the proof is to analyse the first term of the right hand side in (19). A change of variables and a Taylor expansion show that

$$\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n\int_{z_0^2}^{z^2}\int q_2(\nu)'e_d'\tilde{S}_{n,r}(\nu)^{-1}\tilde{K}\Big(\frac{\nu-\zeta(X_i)}{h_{\mathcal{H}}}\Big)\big(Y_i-H(X_i)\big)\mu\Big(\frac{\zeta(X_i)-\nu}{h_{\mathcal{H}}}\Big)\omega_3(\nu^1)d\nu^1d\nu^2$$
$$=\frac{1}{nh_{\mathcal{H}}^2}\sum_{i=1}^n\int_{z_0^2}^{z^2}q_2(\zeta_1(X_i^1),\nu^2)'e_d'\tilde{S}_r^{-1}k_2\Big(\frac{\nu^2-\zeta_2(X_i^2)}{h_{\mathcal{H}}}\Big)\frac{Y_i-H(X_i)}{p_Z(\zeta_1(X_i^1),\nu^2)}V_2^{\tilde{\mu}}\Big(\frac{\zeta_2(X_i^2)-\nu^2}{h_{\mathcal{H}}}\Big)\omega_3(\zeta_1(X_i^1))d\nu^2$$
$$+\frac{1}{nh_{\mathcal{H}}^3}\sum_{i=1}^n\int_{z_0^2}^{z^2}\int q_2(\nu)'e_d'\big(\tilde{S}_{n,r}(\nu)^{-1}-\{p_Z(\nu)\}^{-1}\tilde{S}_r^{-1}\big)$$
$$\cdot\tilde{K}\Big(\frac{\nu-\zeta(X_i)}{h_{\mathcal{H}}}\Big)\big(Y_i-H(X_i)\big)\mu\Big(\frac{\zeta(X_i)-\nu}{h_{\mathcal{H}}}\Big)\omega_3(\nu^1)d\nu^1d\nu^2$$
$$+\frac{1}{n}\sum_{i=1}^n\int_{(z_0^2-\zeta_2(X_i))/h_{\mathcal{H}}}^{(z^2-\zeta_2(X_i))/h_{\mathcal{H}}}\frac{\partial}{\partial z^1}\Bigg[q_2(z^1,\zeta_2(X_i^2)+h_{\mathcal{H}}u)'e_d'\tilde{S}_r^{-1}$$
$$\cdot\frac{(Y_i-H(X_i))\omega_3(z^1)}{p_Z(z^1,\zeta_2(X_i^2)+h_{\mathcal{H}}u)}\Bigg]\Bigg|_{z^1=\zeta_1(X_i^1)}\Big\{\int u^1\tilde{\mu}(u)k_1(u^1)du^1\Big\}k_2(u^2)du^2+o\Big(\frac{1}{n}\Big)$$
$$=:Q_{1n}(z^2)+Q_{2n}(z^2)+Q_{3n}(z^2)+o(h_H^{r+1}), \tag{20}$$

where $o(1/n)=o(h_H^{r+1})$ is due to Assumption 6. By Lemma 1 in Horowitz (1998) (or Theorem 2.37 in Pollard (1984)) and Assumption 6, $Q_{2n}(z^2)=o(h_H^{r+1})$ and $Q_{3n}(z^2)=o(h_H^{r+1})$. As for $Q_{1n}(z^2)$, an integration by parts implies that

$$Q_{1n}(z^2)$$
$$=\frac{1}{nh_{\mathcal{H}}}\sum_{i=1}^n\Bigg[q_2(\zeta_1(X_i^1),z^2)'e_d'\tilde{S}_r^{-1}\frac{Y_i-H(X_i)}{p_Z(\zeta_1(X_i^1),z^2)}V_2^{\tilde{\mu}}\Big(\frac{\zeta_2(X_i^2)-z^2}{h_{\mathcal{H}}}\Big)\omega_3(\zeta_1(X_i^1))\mathcal{K}_2\Big(\frac{z^2-\zeta_2(X_i^2)}{h_{\mathcal{H}}}\Big)$$
$$-q_2(\zeta_1(X_i^1),z_0^2)'e_d'\tilde{S}_r^{-1}\frac{Y_i-H(X_i)}{p_Z(\zeta_1(X_i^1),z_0^2)}V_2^{\tilde{\mu}}\Big(\frac{\zeta_2(X_i^2)-z_0^2}{h_{\mathcal{H}}}\Big)\omega_3(\zeta_1(X_i^1))\mathcal{K}_2\Big(\frac{z_0^2-\zeta_2(X_i^2)}{h_{\mathcal{H}}}\Big)\Bigg]$$
$$-\frac{1}{n}\sum_{i=1}^n\int_{(z_0^2-\zeta_2(X_i))/h_{\mathcal{H}}}^{(z^2-\zeta_2(X_i))/h_{\mathcal{H}}}\frac{\partial}{\partial z^2}\Bigg[q_2(\zeta_1(X_i^1),z^2)'e_d'\tilde{S}_r^{-1}$$
$$\cdot\frac{(Y_i-H(X_i))\omega_3(\zeta_1(X_i^1))}{p_Z(\zeta_1(X_i^1),z^2)}V_2^{\tilde{\mu}}(u)\Bigg]\Bigg|_{z^2=\zeta_2(X_i^2)+h_{\mathcal{H}}u}\mathcal{K}_2(u^2)du^2$$
$$=\tilde{\mathfrak{J}}_{n2}(z^2)+Q_{12n}(z^2), \tag{21}$$

where $\mathcal{K}_2(u)=\int_{-\infty}^u k_2(t)dt$. By Lemma 1 in Horowitz (1998) (or Theorem 2.37 in Pollard (1984)) and Assumption 6, $Q_{12n}(z^2)=o(h_H^{r+1})$ in probability as $n\to\infty$ uniformly over $z^2\in\mathcal{S}_{Z^2}$. Rearranging (19), (20), and (21), then part (i) is proved. Also, an argument similar to the proof of Theorem 6 in Masry (1996) shows that $\sup_{z^2\in\mathcal{S}_{Z^2}}\big|\tilde{\mathfrak{J}}_{n2}(z^2)\big|=O_p\Big(\sqrt{\frac{log(n)}{nh_{\mathcal{H}}}}\Big)$. Thus, part (ii) follows from Assumption 6. □

References

HOROWITZ, J. (1998): "Nonparametric estimation of a generalized additive model with an unknown link function," Working paper, University of Iowa.

KONG, E., O. LINTON, AND Y. XIA (2010): "Uniform Bahadur representation for local polynomial estimates of m-regression and its application to the additive model," *Econometric Theory*, 26(5), 1529–1564.

MASRY, E. (1996): "Multivariate local polynomial regression for time series: uniform strong consistency and rates," *Journal of Time Series Analysis*, 17(6), 571–599.

POLLARD, D. (1984): *Convergence of stochastic processes*. Springer-Verlag.

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric estimation of index coefficients," *Econometrica*, pp. 1403–1430.