Supplement for "Empirical Likelihood Covariate Adjustment for Regression Discontinuity Designs"

Jun Ma* Zhengfei Yu†

Contents

This version: April 22, 2024

*School of Economics, Renmin University of China

†Faculty of Humanities and Social Sciences, University of Tsukuba

S1 Proof of Theorem 1

$\max_i$ is understood as $\max_{1\leq i\leq n}$. For a square matrix A, $\|A\|$ is understood as the spectral norm of A. "With probability approaching one" is abbreviated as "wpa1". "Law of iterated expectations" is abbreviated as "LIE". "$a \lesssim b$" is understood as $a \leq C \cdot b$, for some universal constant $C > 0$ that does not depend on the distribution of the variables or the sample size but may depend on the kernel function.

S1.1 Proof of Part 1

Denote

$$\begin{aligned}\Pi_{p,+} &\coloneqq \mathrm{E}\left[\frac{1}{h} r_p\left(\frac{X}{h}\right) r_p^\top\left(\frac{X}{h}\right) K\left(\frac{X}{h}\right) \mathbb{1}\left(X>0\right)\right]\\ \Pi_{p,-} &\coloneqq \mathrm{E}\left[\frac{1}{h} r_p\left(\frac{X}{h}\right) r_p^\top\left(\frac{X}{h}\right) K\left(\frac{X}{h}\right) \mathbb{1}\left(X<0\right)\right]\end{aligned}$$

and

$$\begin{aligned}\widetilde{W}_{p;+,i} &\coloneqq \mathrm{e}_{p+1,1}^\top \Pi_{p,+}^{-1} r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}\left(X_i>0\right)\\ \widetilde{W}_{p;-,i} &\coloneqq \mathrm{e}_{p+1,1}^\top \Pi_{p,-}^{-1} r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}\left(X_i<0\right)\\ \widetilde{W}_{p,i} &\coloneqq \widetilde{W}_{p;+,i} - \widetilde{W}_{p;-,i}.\end{aligned}$$

We can easily show

$$\begin{aligned}\widehat{\Pi}_{p,\mathrm{s}} - \Pi_{p,\mathrm{s}} &= O_p\left((nh)^{-1/2}\right)\\ \Pi_{p,\mathrm{s}} &= \varphi \cdot \mathrm{V}_{p;\mathrm{s}} + O\left(h\right),\end{aligned} \tag{S1}$$

where the first equality follows from Chebyshev's inequality, change of variables, Taylor expansion and continuity of f_X, and the second equality follows from change of variables and continuity of f_X. It follows that $\forall s \in \{-,+\}$,

$$\left|\mathrm{mineig}\left(\Pi_{p,\mathrm{s}}\right) - \mathrm{mineig}\left(\varphi \cdot \mathrm{V}_{p;\mathrm{s}}\right)\right| \leq \left\|\Pi_{p,\mathrm{s}} - \varphi \cdot \mathrm{V}_{p;\mathrm{s}}\right\| = o\left(1\right) \tag{S2}$$

and

$$\left|\text{mineig}\left(\widehat{\Pi}_{p,\mathrm{s}}\right)-\text{mineig}\left(\Pi_{p,\mathrm{s}}\right)\right|\leq\left\|\widehat{\Pi}_{p,\mathrm{s}}-\Pi_{p,\mathrm{s}}\right\|=o_p\left(1\right). \tag{S3}$$

Then, if follows from this result, the equality

$$\mathrm{A}^{-1}-\mathrm{B}^{-1}=-\mathrm{B}^{-1}\left(\mathrm{A}-\mathrm{B}\right)\mathrm{B}^{-1}+\mathrm{B}^{-1}\left(\mathrm{A}-\mathrm{B}\right)\mathrm{A}^{-1}\left(\mathrm{A}-\mathrm{B}\right)\mathrm{B}^{-1} \tag{S4}$$

for positive definite matrices A and B, and $\text{mineig}\left(\mathrm{V}_{p;\mathrm{s}}\right)>0$ that $\forall s\in\{-,+\}$

$$\begin{aligned}\left\|\Pi_{p,\mathrm{s}}^{-1}-\varphi^{-1}\cdot\mathrm{V}_{p;\mathrm{s}}^{-1}\right\| &= O\left(h\right)\\ \left\|\widehat{\Pi}_{p,\mathrm{s}}^{-1}-\Pi_{p,\mathrm{s}}^{-1}\right\| &= O_p\left(\left(nh\right)^{-1/2}\right).\end{aligned} \tag{S5}$$

Lemma 1. *Let V denote a random variable and $\{V_1,...,V_n\}$ are i.i.d. copies of V. Assume that $nh\to\infty$. Suppose that K is a symmetric continuous PDF supported on $[-1,1]$. Let $\mathbb{B}\subseteq[\underline{x},\overline{x}]$ denote an open neighborhood of 0. The following results hold for all $(\mathrm{s},k)\in\{-,+\}\times\mathbb{N}$: (a) if g_V is uniformly continuous on $\mathbb{B}\setminus\{0\}$,*

$$\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;\mathrm{s}}^{k}V\right]=\frac{\mu_{V,\mathrm{s}}\omega_p^{0,k}}{\varphi^{k-1}}+o\left(1\right);$$

(b) if g_V is $(p+1)$-times continuously differentiable with uniformly continuous $g_V^{(p+1)}$ on $\mathbb{B}\setminus\{0\}$,

$$\frac{1}{nh}\sum_i\widehat{W}_{p;\mathrm{s},i}g_V\left(X_i\right)=\mu_{V,\mathrm{s}}+\frac{\mu_{V,\mathrm{s}}^{(p+1)}}{(p+1)!}\omega_{p;\mathrm{s}}^{p+1,1}h^{p+1}+o_p\left(h^{p+1}\right);$$

(c) if g_{V^2} is bounded on $\mathbb{B}\setminus\{0\}$,

$$\frac{1}{nh}\sum_i\widehat{W}_{p;\mathrm{s},i}^{k}V_i-\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;\mathrm{s}}^{k}V\right]=O_p\left(\left(nh\right)^{-1/2}\right);$$

(d) if $g_{|V|^r}$ is bounded on $\mathbb{B}\setminus\{0\}$ for some $r>2$, $\max_i\left|\widehat{W}_{p;\mathrm{s},i}V_i\right|=O_p\left(\left(nh\right)^{1/r}\right)$.

Proof of Lemma 1. We take $\mathrm{s}=+$ without loss of generality. For Part (a), we have

$$\begin{aligned}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^{k}V\right] &= \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^{k}g_V\left(X\right)\right]\\ &= \int_0^{\overline{x}}\frac{1}{h}\left(\mathrm{e}_{p+1,1}^{\top}\Pi_{p,+}^{-1}r_p\left(\frac{x}{h}\right)\right)^k K^k\left(\frac{x}{h}\right)g_V\left(x\right)f_X\left(x\right)\mathrm{d}x\\ &= \int_0^{\frac{\overline{x}}{h}}\left(\mathrm{e}_{p+1,1}^{\top}\Pi_{p,+}^{-1}r_p\left(y\right)\right)^k K^k\left(y\right)g_V\left(hy\right)f_X\left(hy\right)\mathrm{d}y\end{aligned}$$

$$= \frac{\mu_{V,\mathrm{s}}\omega_p^{0,k}}{\varphi^{k-1}} + o(1)$$

where the first equality follows from LIE, the third equality follows from change of variables, and the fourth equality follows from (S5), continuity of g_V and f_X and applying the equality

$$a^k - b^k = (a-b)\left(a^{k-1} + a^{k-2}b + \cdots + ab^{k-2} + b^{k-1}\right) \tag{S6}$$

to $\left(\mathrm{e}_{p+1,1}^\top \Pi_{p,+}^{-1} r_p(x/h)\right)^k - \left(\mathrm{e}_{p+1,1}^\top \left(\varphi^{-1} \cdot \mathrm{V}_{p;+}^{-1}\right) r_p(x/h)\right)^k$.

By Taylor's theorem, for $X_i > 0$,

$$g_V(X_i) = \mu_{V,+} + \mu_{V,+}^{(1)} X_i + \left(\frac{\mu_{V,+}^{(2)}}{2!}\right) X_i^2 + \cdots + \left(\frac{\mu_{V,+}^{(p)}}{p!}\right) X_i^p + \frac{g_V^{(p+1)}\left(\tilde{X}_i\right)}{(p+1)!} X_i^{p+1},$$

for some $\tilde{X}_i$ between 0 and X_i. Denote $\mu_+ := \left(\mu_{V,+}, \mu_{V,+}^{(1)}, \mu_{V,+}^{(2)}/2, ..., \mu_{V,+}^{(p)}/p!\right)^\top$. Then, we write

$$\frac{1}{nh}\sum_i \widehat{W}_{p;+,i} g_V(X_i) = \frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\left(r_p^\top(X_i)\mu_+\right) + \frac{1}{nh}\sum_i \widehat{W}_{p;+,i} \frac{g_V^{(p+1)}\left(\tilde{X}_i\right)}{(p+1)!} X_i^{p+1}. \tag{S7}$$

Clearly, by the definition of $\widehat{W}_{p;+,i}$,

$$\begin{aligned} \frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\left(r_p^\top(X_i)\mu_+\right) &= \frac{1}{nh}\sum_i \mathrm{e}_{p+1,1}^\top \widehat{\Pi}_{p,+}^{-1} r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}(X_i > 0)\, r_p^\top\left(\frac{X_i}{h}\right) \mathrm{H}\mu_+ \\ &= \mathrm{e}_{p+1,1}^\top \mathrm{H}\mu_+ \\ &= \mu_{V,+}. \end{aligned} \tag{S8}$$

Write

$$\frac{1}{nh}\sum_i \widehat{W}_{p;+,i} \frac{g_V^{(p+1)}\left(\tilde{X}_i\right)}{(p+1)!} X_i^{p+1} = \frac{1}{nh}\sum_i \widehat{W}_{p;+,i} \frac{\left(g_V^{(p+1)}\left(\tilde{X}_i\right) - \mu_{V,+}^{(p+1)}\right)}{(p+1)!} X_i^{p+1} + \frac{1}{nh}\sum_i \widehat{W}_{p;+,i} \frac{\mu_{V,+}^{(p+1)}}{(p+1)!} X_i^{p+1}. \tag{S9}$$

By change of variables and continuity of f_X,

$$\begin{aligned} \mathrm{E}\left[\frac{1}{nh}\sum_i \left\|r_p\left(\frac{X_i}{h}\right)\right\| \left|\left(\frac{X_i}{h}\right)^{p+1}\right| K\left(\frac{X_i}{h}\right) \mathbb{1}(X_i > 0)\right] &= \mathrm{E}\left[\frac{1}{h}\left\|r_p\left(\frac{X}{h}\right)\right\| \left|\left(\frac{X_i}{h}\right)^{p+1}\right| K\left(\frac{X}{h}\right) \mathbb{1}(X > 0)\right] \\ &= \int_0^{\overline{x}} \frac{1}{h}\left\|r_p\left(\frac{x}{h}\right)\right\| \left|\left(\frac{x}{h}\right)^{p+1}\right| \left|K\left(\frac{x}{h}\right)\right| f_X(x)\,\mathrm{d}x = O(1). \end{aligned}$$

Then, by Markov's inequality,

$$\frac{1}{nh}\sum_i \left\| r_p\left(\frac{X_i}{h}\right)\right\| \left|\left(\frac{X_i}{h}\right)^{p+1}\right| \left|K\left(\frac{X_i}{h}\right)\right| \mathbb{1}\left(X_i>0\right) = O_p\left(1\right). \tag{S10}$$

By this result and (S5),

$$\begin{aligned}\frac{1}{nh}\sum_i \left|\widehat{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1}\right| &\leq \left\|\widehat{\Pi}_{p,+}^{-1}\right\| \left(\frac{1}{nh}\sum_i \left\| r_p\left(\frac{X_i}{h}\right)\right\| \left|\left(\frac{X_i}{h}\right)^{p+1}\right| K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\right) \\ &= O_p\left(1\right). \end{aligned} \tag{S11}$$

By this result, $|K(X_i/h)| \lesssim \mathbb{1}(|X_i| \leq h)$ and continuity of $g_V^{(p+1)}$, we have

$$\begin{aligned}&\left|\frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\frac{\left(g_V^{(p+1)}\left(\tilde{X}_i\right)-\mu_{V,+}^{(p+1)}\right)}{(p+1)!}X_i^{p+1}\right| \\ &\qquad \leq \left(\frac{1}{nh}\sum_i\left|\widehat{W}_{p;+,i}\frac{(X_i/h)^{p+1}}{(p+1)!}\right|\right)\cdot\left(\sup_{0<x<h}\left|g_V^{(p+1)}(x)-\mu_{V,+}^{(p+1)}\right|\right)h^{p+1} = o_p\left(h^{p+1}\right).\end{aligned}$$

It now follows from this result, (S7), (S8) and (S9) that

$$\frac{1}{nh}\sum_i \widehat{W}_{p;+,i}g_V\left(X_i\right) = \mu_{V,+} + \frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\frac{\mu_{V,+}^{(p+1)}}{(p+1)!}X_i^{p+1} + o_p\left(h^{p+1}\right). \tag{S12}$$

By triangle inequality, (S5) and (S10),

$$\begin{aligned}&\left|\frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1} - \frac{1}{nh}\sum_i \widetilde{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1}\right| \leq \frac{1}{nh}\sum_i\left|\left(\widehat{W}_{p;+,i}-\widetilde{W}_{p;+,i}\right)\left(\frac{X_i}{h}\right)^{p+1}\right| \\ &\quad \leq \left\|\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right\|\left\{\frac{1}{nh}\sum_i\left\|r_p\left(\frac{X_i}{h}\right)\right\|\left|K\left(\frac{X_i}{h}\right)\right|\left|\left(\frac{X_i}{h}\right)^{p+1}\right|\mathbb{1}\left(X_i>0\right)\right\} \\ &\quad = O_p\left((nh)^{-1/2}\right).\end{aligned}$$

By (S5), Chebyshev's inequality, change of variables and continuity of f_X,

$$\frac{1}{nh}\sum_i \widetilde{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1} - \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}\left(\frac{X_i}{h}\right)^{p+1}\right] = O_p\left((nh)^{-1/2}\right).$$

By (S5), change of variables and continuity of f_X,

$$\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}\left(\frac{X_i}{h}\right)^{p+1}\right] = \int_0^{\overline{x}}\frac{1}{h}\left(\mathrm{e}_{p+1,1}^{\top}\Pi_{p,+}^{-1}r_p\left(\frac{x}{h}\right)\right)\left(\frac{x}{h}\right)^{p+1}K\left(\frac{x}{h}\right)f_X\left(x\right)\mathrm{d}x$$

$$= \omega_{p;+}^{p+1,1} + o\left(1\right).$$

It now follows from these results that

$$\begin{aligned}\frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\frac{\mu_{V,+}^{(p+1)}}{(p+1)!}X_i^{p+1} &= \left(\frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1}\right)\frac{\mu_{V,+}^{(p+1)}}{(p+1)!}h^{p+1}\\ &= \frac{\mu_{V,+}^{(p+1)}}{(p+1)!}\omega_{p;+}^{p+1,1}h^{p+1} + o_p\left(h^{p+1}\right).\end{aligned}$$

Part (b) follows from this result and (S12).

For Part (c), write

$$\frac{1}{nh}\sum_i \widehat{W}_{p;+,i}^k V_i - \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^k V\right] = \frac{1}{nh}\sum_i\left(\widehat{W}_{p;+,i}^k - \widetilde{W}_{p;+,i}^k\right)V_i + \left\{\frac{1}{nh}\sum_i \widetilde{W}_{p;+,i}^k V_i - \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^k V\right]\right\}.$$

Then, by triangle inequality, applying (S6) to $\left(\mathrm{e}_{p+1,1}^\top \widehat{\Pi}_{p,+}^{-1} r_p\left(X_i/h\right)\right)^k - \left(\mathrm{e}_{p+1,1}^\top \Pi_{p,+}^{-1} r_p\left(X_i/h\right)\right)^k$, LIE, change of variables, boundedness of $g_{|V|}$, Markov's inequality, and using (S5), we have

$$\left|\frac{1}{nh}\sum_i\left(\widehat{W}_{p;+,i}^k - \widetilde{W}_{p;+,i}^k\right)V_i\right| \leq \frac{1}{nh}\sum_i\left|\widehat{W}_{p;+,i}^k - \widetilde{W}_{p;+,i}^k\right|\left|V_i\right| = O_p\left((nh)^{-1/2}\right).$$

It follows from (S5), LIE, change of variables, Chebyshev's inequality and boundedness of g_{V^2} that

$$\frac{1}{nh}\sum_i \widetilde{W}_{p;+,i}^k V_i - \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^k V\right] = O_p\left((nh)^{-1/2}\right).$$

Part (c) follows from these results.

For Part (d),

$$\max_i\left|\widehat{W}_{p;+,i}V_i\right| \leq \left(\max_i\left|\widehat{W}_{p;+,i}V_i\right|^r\right)^{1/r} \leq \left(\sum_i\left|\widehat{W}_{p;+,i}V_i\right|^r\right)^{1/r} = \left(\frac{1}{nh}\sum_i\left|\widehat{W}_{p;+,i}V_i\right|^r\right)^{1/r}(nh)^{1/r}. \quad \text{(S13)}$$

By LIE, change of variables, Markov's inequality and boundedness of $g_{|V|^r}$,

$$\frac{1}{nh}\sum_i\left\|r_p\left(\frac{X_i}{h}\right)\right\|^r K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i > 0\right)\left|V_i\right|^r = O_p\left(1\right).$$

By this result, triangle inequality and (S5),

$$\frac{1}{nh}\sum_i\left|\widehat{W}_{p;+,i}V_i\right|^r \leq \frac{1}{nh}\sum_i\left|\widehat{W}_{p;+,i}\right|^r\left|V_i\right|^r$$

$$
\begin{aligned}
&= \left\|\widehat{\Pi}_{p,+}^{-1}\right\|^{r}\left(\frac{1}{nh}\sum_{i}\left\|r_{p}\left(\frac{X_{i}}{h}\right)\right\|^{r}K\left(\frac{X_{i}}{h}\right)\mathbb{1}\left(X_{i}>0\right)\left|V_{i}\right|^{r}\right) \\
&= O_{p}\left(1\right). \qquad \text{(S14)}
\end{aligned}
$$

The conclusion follows from this result and (S13). ■

Let A be a random variable/vector/matrix. $\{A_1, ..., A_n\}$ are i.i.d. copies of A. Denote $\widetilde{\Delta}_{A,k} \coloneqq \mathrm{E}\left[h^{-1}\widetilde{W}_p^k A\right]$, $\widehat{\Psi}_{A,k} \coloneqq (nh)^{-1}\sum_i \widehat{W}_{p,i}^k A_i$, $\widetilde{\Delta}_k \coloneqq \mathrm{E}\left[h^{-1}\widetilde{W}_p^k\right]$, $\widehat{\Psi}_k \coloneqq (nh)^{-1}\sum_i \widehat{W}_{p,i}^k$, $\widetilde{\Delta}_A \coloneqq \mathrm{E}\left[h^{-1}\widetilde{W}_p A\right]$ and $\widehat{\Psi}_A \coloneqq (nh)^{-1}\sum_i \widehat{W}_{p,i}A_i$. Also denote $\widetilde{\Psi}_{A,k} \coloneqq (nh)^{-1}\sum_i \widetilde{W}_{p,i}^k A_i$, $\widetilde{\Psi}_k \coloneqq (nh)^{-1}\sum_i \widetilde{W}_{p,i}^k$ and $\widetilde{\Psi}_A \coloneqq (nh)^{-1}\sum_i \widetilde{W}_{p,i}A_i$.

Lemma 2. *Suppose that Assumptions in the statement of Theorem 1 hold. Then,* $\lambda_p^{\mathsf{eb}} = O_p\left((nh)^{-1/2}\right)$ *and* $\lambda_p^{\mathsf{eb}} = \widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widehat{\Psi}_{\bar{Z}} + O_p\left((nh)^{-1}\right)$.

Proof. Since λ_p^{eb} is defined to be the optimizer of a convex optimization problem:

$$
\lambda_p^{\mathsf{eb}} \coloneqq \underset{\lambda}{\operatorname{argmax}} \sum_i \log\left(1 + \lambda^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)\right),
$$

therefore, we have the first-order conditions:

$$
\sum_i \frac{\widehat{W}_{p,i}\bar{Z}_i}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)} = 0_{d_z+1}.
$$

Then, by

$$
\frac{1}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)} = 1 - \frac{\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)} \qquad \text{(S15)}
$$

and the first-order conditions, we have

$$
\frac{1}{nh}\sum_i \widehat{W}_{p,i}\bar{Z}_i = \left(\frac{1}{nh}\sum_i \frac{\widehat{W}_{p,i}^2\bar{Z}_i\bar{Z}_i^\top}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}\right)\lambda_p^{\mathsf{eb}} \qquad \text{(S16)}
$$

and

$$
\frac{1}{nh}\sum_i \widehat{W}_{p,i}\left(\bar{Z}_i^\top\lambda_p^{\mathsf{eb}}\right) = \frac{1}{nh}\sum_i \frac{\widehat{W}_{p,i}^2\left(\bar{Z}_i^\top\lambda_p^{\mathsf{eb}}\right)^2}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}. \qquad \text{(S17)}
$$

Note that for all i,

$$
0 < 1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right) \leq 1 + \max_i \left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right) \leq 1 + \left\|\lambda_p^{\mathsf{eb}}\right\| \max_i \left\|\widehat{W}_{p,i}\bar{Z}_i\right\|,
$$

where the first inequality follows from construction of λ_p^{eb}. Then, by this result and (S17),

$$\left\|\lambda_p^{\mathsf{eb}}\right\|\left\|\widehat{\Psi}_{\bar{Z}}\right\| \geq \frac{1}{nh}\sum_i \widehat{W}_{p,i}\left(\bar{Z}_i^\top \lambda_p^{\mathsf{eb}}\right) = \frac{1}{nh}\sum_i \frac{\widehat{W}_{p,i}^2\left(\bar{Z}_i^\top \lambda_p^{\mathsf{eb}}\right)^2}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)} \geq \frac{1}{nh}\sum_i \frac{\widehat{W}_{p,i}^2\left(\bar{Z}_i^\top \lambda_p^{\mathsf{eb}}\right)^2}{1+\left\|\lambda_p^{\mathsf{eb}}\right\|\max_i\left\|\widehat{W}_{p,i}\bar{Z}_i\right\|}.$$

Then by rearranging, we have

$$\left(\lambda_p^{\mathsf{eb}}\right)^\top \widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2}\lambda_p^{\mathsf{eb}} \quad\leq\quad \left\|\lambda_p^{\mathsf{eb}}\right\|\left\|\widehat{\Psi}_{\bar{Z}}\right\|\left\{1+\left\|\lambda_p^{\mathsf{eb}}\right\|\max_i\left\|\widehat{W}_{p,i}\bar{Z}_i\right\|\right\}$$

and

$$\left(\lambda_p^{\mathsf{eb}}\right)^\top \widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\lambda_p^{\mathsf{eb}} \leq \left\|\lambda_p^{\mathsf{eb}}\right\|\left\|\widehat{\Psi}_{\bar{Z}}\right\|\left\{1+\left\|\lambda_p^{\mathsf{eb}}\right\|\max_i\left\|\widehat{W}_{p,i}\bar{Z}_i\right\|\right\}+\left\|\lambda_p^{\mathsf{eb}}\right\|^2\left\|\widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2}-\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right\|.$$

Then it is clear that

$$\left\|\lambda_p^{\mathsf{eb}}\right\|^2 \operatorname{mineig}\left(\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right) \leq \left\|\lambda_p^{\mathsf{eb}}\right\|\left\|\widehat{\Psi}_{\bar{Z}}\right\|\left\{1+\left\|\lambda_p^{\mathsf{eb}}\right\|\max_i\left\|\widehat{W}_{p,i}\bar{Z}_i\right\|\right\}+\left\|\lambda_p^{\mathsf{eb}}\right\|^2\left\|\widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2}-\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right\|$$

and by rearranging,

$$\left(\operatorname{mineig}\left(\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right)-\left\|\widehat{\Psi}_{\bar{Z}}\right\|\max_i\left\|\widehat{W}_{p,i}\bar{Z}_i\right\|-\left\|\widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2}-\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right\|\right)\left\|\lambda_p^{\mathsf{eb}}\right\| \leq \left\|\widehat{\Psi}_{\bar{Z}}\right\|. \tag{S18}$$

Note that by Lemma 1(a), $\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2} = \left(\omega_p^{0,2}/\varphi\right)\mu_{\bar{Z}\bar{Z}^\top,\pm}+o(1)$ and under our assumptions, $\mu_{\bar{Z}\bar{Z}^\top,\pm}$ is positive definite. Then we have $\operatorname{mineig}\left(\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right) = \operatorname{mineig}\left(\left(\omega_p^{0,2}/\varphi\right)\mu_{\bar{Z}\bar{Z}^\top,\pm}\right)+o(1)$ and $\operatorname{mineig}\left(\left(\omega_p^{0,2}/\varphi\right)\mu_{\bar{Z}\bar{Z}^\top,\pm}\right) > 0$. Therefore, $\operatorname{mineig}\left(\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right)$ is bounded away from zero when n is sufficiently large. We have $\widehat{\Psi}_{\bar{Z}} = \left(0,\widehat{\Psi}_Z^\top\right)^\top$ and by Lemma 1(b),

$$\begin{aligned}\widehat{\Psi}_Z &= \frac{1}{nh}\sum_i \widehat{W}_{p,i}g_Z(X_i)+\frac{1}{nh}\sum_i \widehat{W}_{p,i}\left(Z_i-g_Z(X_i)\right)\\ &= \frac{1}{nh}\sum_i \widehat{W}_{p,i}\left(Z_i-g_Z(X_i)\right)+\frac{\left(\mu_{Z,+}^{(p+1)}\omega_{p;+}^{p+1,1}-\mu_{Z,-}^{(p+1)}\omega_{p;-}^{p+1,1}\right)}{(p+1)!}h^{p+1}+o_p\left(h^{p+1}\right).\end{aligned} \tag{S19}$$

By Lemma 1(c), $(nh)^{-1}\sum_i \widehat{W}_{p,i}\left(Z_i-g_Z(X_i)\right) = O_p\left((nh)^{-1/2}\right)$. Then it follows that $\left\|\widehat{\Psi}_{\bar{Z}}\right\| = O_p\left((nh)^{-1/2}\right)$. By Lemma 1(d), $\left\|\widehat{\Psi}_{\bar{Z}}\right\|\max_i\left\|\widehat{W}_{p,i}\bar{Z}_i\right\| = o_p(1)$. By Lemma 1(c), $\left\|\widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2}-\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}\right\| = O_p\left((nh)^{-1/2}\right)$. Therefore, we have the coefficient of $\left\|\lambda_p^{\mathsf{eb}}\right\|$ in the parentheses on the left hand side of (S18) is bounded away from zero wpa1. This result, $\left\|\widehat{\Psi}_{\bar{Z}}\right\| = O_p\left((nh)^{-1/2}\right)$ and (S18) implies that $\left\|\lambda_p^{\mathsf{eb}}\right\| = O_p\left((nh)^{-1/2}\right)$.

Now (S15) and (S16) imply that

$$\widehat{\Psi}_{\bar{Z}} = \left\{ \frac{1}{nh} \sum_i \widehat{W}_{p,i}^2 \bar{Z}_i \bar{Z}_i^\top \cdot \left(1 - \frac{\left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)} \right) \right\} \lambda_p^{\mathsf{eb}}$$

and

$$\widehat{\Psi}_{\bar{Z}} = \widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2} \lambda_p^{\mathsf{eb}} - \frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i}^2 \bar{Z}_i \left(\bar{Z}_i^\top \lambda_p^{\mathsf{eb}}\right) \left(\lambda_p^{\mathsf{eb}} \left(\widehat{W}_{p,i} \bar{Z}_i\right)\right)}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)}. \tag{S20}$$

By arguments similar to those used to prove (S11), $(nh)^{-1} \sum_i \left\| \widehat{W}_{p,i} \bar{Z}_i \right\|^3 = O_p(1)$. Therefore, by this result, $\|\lambda_p^{\mathsf{eb}}\| = O_p\left((nh)^{-1/2}\right)$ and Lemma 1(d),

$$\begin{aligned} \left\| \frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i}^2 \bar{Z}_i \left(\bar{Z}_i^\top \lambda_p^{\mathsf{eb}}\right) \left(\left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)\right)}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)} \right\| &\leq \frac{\left(\frac{1}{nh} \sum_i \left\| \widehat{W}_{p,i} \bar{Z}_i \right\|^3\right) \|\lambda_p^{\mathsf{eb}}\|^2}{1 - \|\lambda_p^{\mathsf{eb}}\| \max_i \left\| \widehat{W}_{p,i} \bar{Z}_i \right\|} \\ &= O_p\left((nh)^{-1}\right). \end{aligned} \tag{S21}$$

The second conclusion follows from this result, (S20), $\left\| \widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2} - \widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2} \right\| = O_p\left((nh)^{-1/2}\right)$ and $\|\lambda_p^{\mathsf{eb}}\| = O_p\left((nh)^{-1/2}\right)$. ■

Proof of Theorem 1 Part 1. Denote $\widehat{\vartheta}_Y^{\mathsf{eb}} := \sum_i w_i^{\mathsf{eb}} \widehat{W}_{p,i} Y_i / h$ and $\widehat{\vartheta}_D^{\mathsf{eb}} := \sum_i w_i^{\mathsf{eb}} \widehat{W}_{p,i} D_i / h$ for notational simplicity. By using

$$\frac{1}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)} = 1 - \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right) + \frac{\left(\left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)\right)^2}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)},$$

we have

$$\begin{aligned} \widehat{\vartheta}_Y^{\mathsf{eb}} &= \frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i} Y_i}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)} \\ &= \frac{1}{nh} \sum_i \widehat{W}_{p,i} Y_i \left\{ 1 - \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right) + \frac{\left(\left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)\right)^2}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)} \right\} \\ &= \widehat{\Psi}_Y - \widehat{\Psi}_{Y\bar{Z},2}^\top \lambda_p^{\mathsf{eb}} + \frac{1}{nh} \sum_i \frac{\left(\widehat{W}_{p,i} Y_i\right) \left(\left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)\right)^2}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} \bar{Z}_i\right)}. \end{aligned} \tag{S22}$$

By Lemma 1(c), $\widehat{\Psi}_{Y\bar{Z},2} - \widetilde{\Delta}_{Y\bar{Z},2} = O_p\left((nh)^{-1/2}\right)$. By arguments similar to those used in the proof of (S21),

$$\frac{1}{nh}\sum_i \frac{\left(\widehat{W}_{p,i}Y_i\right)\left(\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)\right)^2}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)} = O_p\left((nh)^{-1}\right).$$

By these results, we have $\widehat{\vartheta}_Y^{\mathsf{eb}} = \widehat{\Psi}_Y - \widetilde{\Delta}_{Y\bar{Z},2}^\top\lambda_p^{\mathsf{eb}} + O_p\left((nh)^{-1}\right)$. And similarly, $\widehat{\vartheta}_D^{\mathsf{eb}} = \widehat{\Psi}_D - \widetilde{\Delta}_{D\bar{Z},2}^\top\lambda_p^{\mathsf{eb}} + O_p\left((nh)^{-1}\right)$. By these results and Lemma 1(b,c), $\widehat{\vartheta}_D^{\mathsf{eb}} - \mu_{D,\dagger} = O_p\left((nh)^{-1/2}\right)$ and $\widehat{\vartheta}_Y^{\mathsf{eb}} - \mu_{Y,\dagger} = O_p\left((nh)^{-1/2}\right)$. By these results and simple algebra, we have

$$\begin{aligned}\frac{\widehat{\vartheta}_Y^{\mathsf{eb}}}{\widehat{\vartheta}_D^{\mathsf{eb}}} - \frac{\mu_{Y,\dagger}}{\mu_{D,\dagger}} &= \left(\frac{\widehat{\vartheta}_Y^{\mathsf{eb}}}{\mu_{D,\dagger}} - \frac{\mu_{Y,\dagger}}{\mu_{D,\dagger}}\right) + \frac{\widehat{\vartheta}_Y^{\mathsf{eb}}}{\mu_{D,\dagger}}\cdot\left(\frac{\mu_{D,\dagger}}{\widehat{\vartheta}_D^{\mathsf{eb}}} - 1\right)\\ &= \frac{\widehat{\Psi}_Y - \widetilde{\Delta}_{Y\bar{Z},2}^\top\lambda_p^{\mathsf{eb}} - \mu_{Y,\dagger}}{\mu_{D,\dagger}} - \frac{\vartheta\left(\widehat{\Psi}_D - \widetilde{\Delta}_{D\bar{Z},2}^\top\lambda_p^{\mathsf{eb}} - \mu_{D,\dagger}\right)}{\mu_{D,\dagger}} + O_p\left((nh)^{-1}\right)\\ &= \frac{\widehat{\Psi}_M - \widetilde{\Delta}_{M\bar{Z},2}^\top\lambda_p^{\mathsf{eb}}}{\mu_{D,\dagger}} + O_p\left((nh)^{-1}\right). \qquad \text{(S23)}\end{aligned}$$

By writing $\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}$ as a block matrix and inverting,

$$\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1} = \begin{bmatrix}\widetilde{\Delta}_2^{-1} & 0_{d_z}^\top\\ 0_{d_z} & 0_{d_z\times d_z}\end{bmatrix} + \begin{bmatrix}-\frac{\widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\\ \mathrm{I}_{d_z}\end{bmatrix}\left(\widetilde{\Delta}_{ZZ^\top,2} - \frac{\widetilde{\Delta}_{Z,2}\widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\right)^{-1}\begin{bmatrix}-\frac{\widetilde{\Delta}_{Z,2}}{\widetilde{\Delta}_2} & \mathrm{I}_{d_z}\end{bmatrix}. \qquad \text{(S24)}$$

Then, by Lemma 2, we have

$$\begin{aligned}\widetilde{\Delta}_{M\bar{Z},2}^\top\lambda_p^{\mathsf{eb}} &= \begin{bmatrix}\widetilde{\Delta}_{M,2} & \widetilde{\Delta}_{MZ,2}^\top\end{bmatrix}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\begin{bmatrix}0\\ \widehat{\Psi}_Z\end{bmatrix} + O_p\left((nh)^{-1}\right)\\ &= \widetilde{\gamma}_M^\top\widehat{\Psi}_Z + O_p\left((nh)^{-1}\right), \qquad \text{(S25)}\end{aligned}$$

where

$$\widetilde{\gamma}_M \coloneqq \left(\widetilde{\Delta}_{ZZ^\top,2} - \frac{\widetilde{\Delta}_{Z,2}\widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\right)^{-1}\left(\widetilde{\Delta}_{MZ,2} - \frac{\widetilde{\Delta}_{M,2}\widetilde{\Delta}_{Z,2}}{\widetilde{\Delta}_2}\right).$$

By Lemma 1(a), $\widetilde{\Delta}_{ZZ^\top,2} = \left(\omega_p^{0,2}/\varphi\right)\mu_{ZZ^\top,\pm} + o(1)$, $\widetilde{\Delta}_{Z,2} = 2\left(\omega_p^{0,2}/\varphi\right)\mu_Z + o(1)$, $\widetilde{\Delta}_2 = 2\left(\omega_p^{0,2}/\varphi\right) + o(1)$, $\widetilde{\Delta}_{MZ,2} = 2\left(\omega_p^{0,2}/\varphi\right)\mu_{MZ,\pm} + o(1)$ and $\widetilde{\Delta}_{M,2} = 2\left(\omega_p^{0,2}/\varphi\right)\mu_M + o(1)$. Therefore, it follows from these results, (S4) and the fact that for $\mathrm{s}\in\{-,+\}$, $\mu_{ZZ^\top,\mathrm{s}} - \mu_Z\mu_{Z^\top}$ is positive definite that $\widetilde{\gamma}_M = \gamma_M + o(1)$. Therefore, by this result, $\widehat{\Psi}_Z = O_p\left((nh)^{-1/2}\right)$, (S23) and (S25), we have $\widehat{\vartheta}_p^{\mathsf{eb}} - \vartheta = \widehat{\Psi}_\epsilon/\mu_{D,\dagger} + o_p\left((nh)^{-1/2}\right)$. By lemma

1(b),

$$\begin{aligned}\widehat{\Psi}_{\epsilon} &= \frac{1}{nh}\sum_{i}\widehat{W}_{p,i}g_{\epsilon}\left(X_{i}\right)+\frac{1}{nh}\sum_{i}\widehat{W}_{p,i}\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)\\ &= \frac{\left(\mu_{\epsilon,+}^{(p+1)}\omega_{p;+}^{p+1,1}-\mu_{\epsilon,-}^{(p+1)}\omega_{p;-}^{p+1,1}\right)}{(p+1)!}h^{p+1}+\frac{1}{nh}\sum_{i}\widehat{W}_{p,i}\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)+o\left(h^{p+1}\right).\end{aligned}$$

Then, we write

$$\begin{aligned}&\frac{1}{nh}\sum_{i}\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)\\ &\qquad=\mathrm{e}_{p+1,1}^{\top}\left(\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right)\left(\frac{1}{nh}\sum_{i}r_{p}\left(\frac{X_{i}}{h}\right)K\left(\frac{X_{i}}{h}\right)\mathbb{1}\left(X_{i}>0\right)\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)\right).\end{aligned}$$

By change of variables and Chebyshev's inequality,

$$\frac{1}{nh}\sum_{i}r_{p}\left(\frac{X_{i}}{h}\right)K\left(\frac{X_{i}}{h}\right)\mathbb{1}\left(X_{i}>0\right)\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)=O_{p}\left((nh)^{-1/2}\right).$$

Therefore, by this result and (S5), $(nh)^{-1}\sum_{i}\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)=O_{p}\left((nh)^{-1}\right)$. It then follows that

$$\widehat{\Psi}_{\epsilon}=\frac{\left(\mu_{\epsilon,+}^{(p+1)}\omega_{p;+}^{p+1,1}-\mu_{\epsilon,-}^{(p+1)}\omega_{p;-}^{p+1,1}\right)}{(p+1)!}h^{p+1}+\frac{1}{nh}\sum_{i}\widetilde{W}_{p,i}\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)+o_{p}\left((nh)^{-1/2}\right). \tag{S26}$$

By Lemma 1(a),

$$\begin{aligned}\operatorname{Var}\left[\frac{1}{\sqrt{nh}}\sum_{i}\widetilde{W}_{p,i}\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)\right] &= \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p}^{2}\left(\epsilon-g_{\epsilon}\left(X\right)\right)^{2}\right]\\ &= \frac{\omega_{p}^{0,2}\sigma^{2}}{\varphi}+o\left(1\right).\end{aligned} \tag{S27}$$

Let $\varsigma\in(0,1)$. We have

$$\begin{aligned}\sum_{i}\mathrm{E}\left[\left|\frac{1}{\sqrt{nh}}\widetilde{W}_{p,i}\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)\right|^{2+\varsigma}\right] &= \frac{1}{(nh)^{1+\varsigma/2}}\sum_{i}\mathrm{E}\left[\left|\widetilde{W}_{p,i}\left(\epsilon_{i}-g_{\epsilon}\left(X_{i}\right)\right)\right|^{2+\varsigma}\right]\\ &\lesssim \frac{\mathrm{E}\left[h^{-1}\left|\widetilde{W}_{p}\right|^{2+\varsigma}\left(|\epsilon|^{2+\varsigma}+\left|g_{\epsilon}\left(X\right)\right|^{2+\varsigma}\right)\right]}{(nh)^{\varsigma/2}},\end{aligned}$$

where the inequality follows from Loève's c_r inequality. By (S5), change of variables and Markov's inequality, $\mathrm{E}\left[h^{-1}\left|\widetilde{W}_p\right|^{2+\varsigma}|\epsilon|^{2+\varsigma}\right] = O(1)$ and $\mathrm{E}\left[h^{-1}\left|\widetilde{W}_p\right|^{2+\varsigma}|g_\epsilon(X)|^{2+\varsigma}\right] = O(1)$. Therefore, we have verified Lyapunov's condition

$$\sum_i \mathrm{E}\left[\left|\frac{1}{\sqrt{nh}}\widetilde{W}_{p,i}\left(\epsilon_i - g_\epsilon(X_i)\right)\right|^{2+\varsigma}\right] \to 0.$$

By Lyapunov's central limit theorem,

$$\frac{\frac{1}{\sqrt{nh}}\sum_i \widetilde{W}_{p,i}\left(\epsilon_i - g_\epsilon(X_i)\right)}{\sqrt{\operatorname{Var}\left[\frac{1}{\sqrt{nh}}\sum_i \widetilde{W}_{p,i}\left(\epsilon_i - g_\epsilon(X_i)\right)\right]}} \to_d \mathrm{N}(0,1).$$

It follows from this result, (S26) and (S27) that

$$\sqrt{nh}\left(\widehat{\Psi}_\epsilon - \frac{\left(\mu_{\epsilon,+}^{(p+1)}\omega_{p;+}^{p+1,1} - \mu_{\epsilon,-}^{(p+1)}\omega_{p;-}^{p+1,1}\right)}{(p+1)!}h^{p+1}\right) \to_d \mathrm{N}\left(0, \frac{\omega_p^{0,2}\sigma^2}{\varphi}\right). \tag{S28}$$

The conclusion follows from this result, $\widehat{\vartheta}_p^{\mathsf{eb}} - \vartheta = \widehat{\Psi}_\epsilon/\mu_{D,\dagger} + o_p\left((nh)^{-1/2}\right)$ and Slutsky's lemma. ■

S1.2 Proof of Part 2

The following lemma is an analogue of Lemma 1.

Lemma 3. *Let V denote a random variable and $\{V_1, ..., V_n\}$ are i.i.d. copies of V. Assume that $nh \to \infty$. Suppose that K is a symmetric continuous PDF supported on $[-1,1]$. Let $\mathbb{B}$ denote a neighborhood of 0. The following results hold for all $(\mathrm{s}, k) \in \{-,+\} \times \mathbb{N}$. (a) If m_V is uniformly continuous on $\mathbb{B}\setminus\{0\}$, for $k \geq 2$,*

$$\mathrm{E}\left[\frac{1}{h}W_{p;\mathrm{s}}^k V\right] = \psi_{V,\mathrm{s}}\omega_p^{0,k} + o(1);$$

(b) If m_V is $(p+1)$-times continuously differentiable with uniformly continuous $m_V^{(p+1)}$ on $\mathbb{B}\setminus\{0\}$,

$$\mathrm{E}\left[\frac{1}{h}W_{p;\mathrm{s}}V\right] = \psi_{V,\mathrm{s}} + \frac{\psi_{V,\mathrm{s}}^{(p+1)}}{(p+1)!}\omega_{p;\mathrm{s}}^{p+1,1}h^{p+1} + o\left(h^{p+1}\right);$$

(c) If g_{V^2} is bounded on $\mathbb{B}\setminus\{0\}$,

$$\frac{1}{nh}\sum_i W_{p;\mathrm{s},i}^k V_i - \mathrm{E}\left[\frac{1}{h}W_{p;\mathrm{s}}^k V\right] = O_p\left((nh)^{-1/2}\right);$$

(d) If $g_{|V|^r}$ is bounded on $\mathbb{B}\setminus\{0\}$, $\max_i |W_{p;\mathrm{s},i}V_i| = O_p\left((nh)^{1/r}\right)$.

Proof of Lemma 3. We take s = + without loss of generality. By LIE, change of variables and continuity of g_V and f_X,

$$\begin{aligned}\mathrm{E}\left[\frac{1}{h}W_{p;+}^{k}V\right] &= \int_0^{\overline{x}}\frac{1}{h}\mathcal{K}_{p;+}^{k}\left(\frac{x}{h}\right)m_V\left(x\right)\mathrm{d}x \\ &= \int_0^{\frac{\overline{x}}{h}}\mathcal{K}_{p;+}^{k}\left(y\right)m_V\left(hy\right)\mathrm{d}y \\ &= \psi_{V,+}\omega_p^{0,k}+o\left(1\right). \end{aligned} \tag{S29}$$

Part (a) follows from the above calculation.

Part (b) is a straightforward extension of Bickel and Doksum (2015, Proposition 11.3.1), which follows from LIE and $(p+1)$-th order Taylor expansion. Denote $\psi_+ := \left(\psi_{V,+},\psi_{V,+}^{(1)},\psi_{V,+}^{(2)}/2,...,\psi_{V,+}^{(p)}/p!\right)^\top$. By Taylor's theorem,

$$\begin{aligned}\mathrm{E}\left[\frac{1}{h}W_{p;+}V\right] &= \int_0^{\frac{\overline{x}}{h}}\mathcal{K}_{p;+}\left(y\right)m_V\left(hy\right)\mathrm{d}y \\ &= \int_0^{\frac{\overline{x}}{h}}\mathcal{K}_{p;+}\left(y\right)\left(\psi_+^\top \mathrm{H}r_p\left(y\right)\right)\mathrm{d}y+\int_0^{\frac{\overline{x}}{h}}\mathcal{K}_{p;+}\left(y\right)\left(\frac{m_V^{(p+1)}\left(\tilde{t}\right)}{(p+1)!}\right)\left(hy\right)^{p+1}\mathrm{d}y, \end{aligned} \tag{S30}$$

where $0<\tilde{t}<hy$ denotes the mean value. By the definition of $\mathcal{K}_{p;+}$, when h is sufficiently small, we have

$$\int_0^{\frac{\overline{x}}{h}}\mathcal{K}_{p;+}\left(y\right)\left(\psi_+^\top \mathrm{H}r_p\left(y\right)\right)\mathrm{d}y=\mathrm{e}_{p+1,1}^\top\mathrm{H}\psi_+=\psi_{V,+}. \tag{S31}$$

By continuity of g_V and f_X,

$$\int_0^{\frac{\overline{x}}{h}}\mathcal{K}_{p;+}\left(y\right)\left(\frac{m_V^{(p+1)}\left(\tilde{t}\right)}{(p+1)!}\right)\left(hy\right)^{p+1}\mathrm{d}y=\frac{\psi_{V,+}^{(p+1)}\omega_{p;+}^{p+1,1}}{(p+1)!}h^{p+1}+o\left(h^{p+1}\right).$$

Part (b) follows from this result and (S30).

By LIE and change of variables,

$$\mathrm{Var}\left[\frac{1}{h}W_{p;+}^{k}V\right]=\frac{1}{h}\cdot\left\{\mathrm{E}\left[\frac{1}{h}W_{p;+}^{2k}V^2\right]-h\left(\mathrm{E}\left[\frac{1}{h}W_{p;+}^{k}V\right]\right)^2\right\}=O\left(h^{-1}\right)$$

Part (c) follows from this result and Chebyshev's inequality.

For Part (d), by arguments similar to those used to prove (S13)

$$\max_i\left|W_{p;+,i}V_i\right|\leq\left(\frac{1}{nh}\sum_i\left|W_{p;+,i}V_i\right|^r\right)^{1/r}\left(nh\right)^{1/r}. \tag{S32}$$

By LIE, change of variables, Markov's inequality and boundedness of $g_{|V|^r}$, $(nh)^{-1}\sum_i |W_{p;+,i}V_i|^r = O_p(1)$. The conclusion follows from this result and (S32). ■

Let A be a random variable/vector/matrix. $\{A_1,...,A_n\}$ are i.i.d. copies of A. Denote $\Delta_{A,k} \coloneqq \mathrm{E}\left[h^{-1}W_p^k A\right]$, $\Psi_{A,k} \coloneqq (nh)^{-1}\sum_i W_{p,i}^k A_i$, $\Delta_k \coloneqq \mathrm{E}\left[h^{-1}W_p^k\right]$, $\Psi_k \coloneqq (nh)^{-1}\sum_i W_{p,i}^k$, $\Delta_A \coloneqq \mathrm{E}\left[h^{-1}W_p A\right]$, $\Psi_A \coloneqq (nh)^{-1}\sum_i W_{p,i}A_i$ and $\Psi \coloneqq (nh)^{-1}\sum_i W_{p,i}$.

Lemma 4. *Suppose that Assumptions in the statement of Theorem 1 hold. Then,* $\lambda_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$ *and* $\lambda_p^{\mathsf{mc}} = \Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Psi_{\bar{Z}} + O_p\left((nh)^{-1}\right)$.

Proof of Lemma 4. By arguments similar to those used in the proof of Lemma 2,

$$\left(\lambda_p^{\mathsf{mc}}\right)^\top \Delta_{\bar{Z}\bar{Z}^\top,2}\lambda_p^{\mathsf{mc}} \leq \left\|\lambda_p^{\mathsf{mc}}\right\| \left\|\Psi_{\bar{Z}}\right\| \left\{1 + \left\|\lambda_p^{\mathsf{mc}}\right\| \max_i \left\|W_{p,i}\bar{Z}_i\right\|\right\} + \left\|\lambda_p^{\mathsf{mc}}\right\|^2 \left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\| \tag{S33}$$

and

$$\left(\text{mineig}\left(\Delta_{\bar{Z}\bar{Z}^\top,2}\right) - \left\|\Psi_{\bar{Z}}\right\| \max_i \left\|W_{p,i}\bar{Z}_i\right\| - \left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\|\right) \left\|\lambda_p^{\mathsf{mc}}\right\| \leq \left\|\Psi_{\bar{Z}}\right\|. \tag{S34}$$

Write $\Psi_{\bar{Z}} = \Psi_{\bar{Z}} - \Delta_{\bar{Z}} + \Delta_{\bar{Z}}$. It follows from Lemma 3(c) that $\Psi_{\bar{Z}} - \Delta_{\bar{Z}}$ is $O_p\left((nh)^{-1/2}\right)$. By Lemma 3(b), $\Delta_{\bar{Z}} = O\left(h^{p+1}\right)$. Therefore, $\Psi_{\bar{Z}} = O_p\left((nh)^{-1/2}\right)$. It also follows from this result, Lemma 3(d) that $\left\|\Psi_{\bar{Z}}\right\| \left(\max_i \left\|W_{p,i}\bar{Z}_i\right\|\right) = o_p(1)$. By Lemma 3(c), $\left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\| = O_p\left((nh)^{-1/2}\right)$. Since it follows from Lemma 3(a) that $\Delta_{\bar{Z}\bar{Z}^\top,2} = \psi_{\bar{Z}\bar{Z}^\top,\pm}\omega_p^{0,2} + o(1)$, we have $\text{mineig}\left(\Delta_{\bar{Z}\bar{Z}^\top,2}\right) = \text{mineig}\left(\psi_{\bar{Z}\bar{Z}^\top,\pm}\omega_p^{0,2}\right) + o(1)$. Therefore wpa1,

$$\text{mineig}\left(\Delta_{\bar{Z}\bar{Z}^\top,2}\right) - \left\|\Psi_{\bar{Z}}\right\| \max_i \left\|W_{p,i}\bar{Z}_i\right\| - \left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\| > \frac{1}{2} \cdot \text{mineig}\left(\psi_{\bar{Z}\bar{Z}^\top,\pm}\omega_p^{0,2}\right) > 0. \tag{S35}$$

By this result and (S34), $\lambda_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$.

By arguments similar to those used to show (S20), we have

$$\Psi_{\bar{Z}} = \Psi_{\bar{Z}\bar{Z}^\top,2}\lambda_p^{\mathsf{mc}} - \frac{1}{nh}\sum_i \frac{W_{p,i}^2 \bar{Z}_i \left(\bar{Z}_i^\top \lambda_p^{\mathsf{mc}}\right)\left(\left(\lambda_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\bar{Z}_i\right)\right)}{1 + \left(\lambda_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\bar{Z}_i\right)}. \tag{S36}$$

By Lemma 3(a) and Markov's inequality, $(nh)^{-1}\sum_i \left\|W_{p,i}\bar{Z}_i\right\|^3 = O_p(1)$. Then by using this result, Lemma 3(d) and $\lambda_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$, the second term on the right hand side of (S36) is $O_p\left((nh)^{-1}\right)$. The second conclusion follows from this result, $\left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\| = O_p\left((nh)^{-1/2}\right)$, $\lambda_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$ and (S36). ■

Proof of Theorem 1 Part 2. Denote $\widehat{\vartheta}_Y^{\mathsf{mc}} \coloneqq \sum_i w_i^{\mathsf{mc}} W_{p,i}Y_i/h$ and $\widehat{\vartheta}_D^{\mathsf{mc}} \coloneqq \sum_i w_i^{\mathsf{mc}} W_{p,i}D_i/h$ for notational simplicity. By writing $\Delta_{\bar{Z}\bar{Z}^\top,2}$ as a block matrix and inverting,

$$\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1} = \begin{bmatrix} \Delta_2^{-1} & 0_{d_z}^\top \\ 0_{d_z} & 0_{d_z \times d_z} \end{bmatrix} + \begin{bmatrix} -\frac{\Delta_{Z^\top,2}}{\Delta_2} \\ \mathrm{I}_{d_z} \end{bmatrix} \left(\Delta_{ZZ^\top,2} - \frac{\Delta_{Z,2}\Delta_{Z^\top,2}}{\Delta_2} \right)^{-1} \begin{bmatrix} -\frac{\Delta_{Z,2}}{\Delta_2} & \mathrm{I}_{d_z} \end{bmatrix}. \tag{S37}$$

Let

$$\bar{\gamma}_M := \left(\Delta_{ZZ^\top,2} - \frac{\Delta_{Z,2}\Delta_{Z^\top,2}}{\Delta_2} \right)^{-1} \left(\Delta_{MZ,2} - \frac{\Delta_{M,2}\Delta_{Z,2}}{\Delta_2} \right).$$

It follows from Lemma 3(a) that $\Delta_{Z,2} = 2\omega_p^{0,2}\psi_Z + o(1)$, $\Delta_2 = 2\omega_p^{0,2}\varphi$, $\Delta_{M,2} = 2\omega_p^{0,2}\psi_M + o(1)$, $\Delta_{ZZ^\top,2} = \omega_p^{0,2}\psi_{ZZ^\top,\pm} + o(1)$ and $\Delta_{MZ,2} = \omega_p^{0,2}\psi_{MZ,\pm} + o(1)$. It follows from these results, (S4) and the fact that for $\mathrm{s} \in \{-,+\}$, $\mu_{ZZ^\top,\mathrm{s}} - \mu_Z\mu_{Z^\top}$ is positive definite that $\bar{\gamma}_M = \gamma_M + o(1)$. It follows from Lemma 3(b,c) that $\Psi_{\bar{Z}} = O_p\left((nh)^{-1/2}\right)$. By these results,

$$\begin{aligned} \Delta_{M\bar{Z},2}^\top \Delta_{\bar{Z}\bar{Z}^\top,2}^{-1} \Psi_{\bar{Z}} &= \frac{\Delta_{M,2}}{\Delta_2} \cdot \Psi + \bar{\gamma}_M^\top \left(\Psi_Z - \frac{\Delta_{Z,2}}{\Delta_2} \cdot \Psi \right) \\ &= \mu_M \cdot \Psi + \gamma_M^\top (\Psi_Z - \mu_Z \cdot \Psi) + o_p\left((nh)^{-1/2}\right). \end{aligned} \tag{S38}$$

By arguments similar to those used to show (S22),

$$\widehat{\vartheta}_Y^{\mathsf{mc}} = \Psi_Y - \Psi_{Y\bar{Z},2}^\top \lambda_p^{\mathsf{mc}} + \frac{1}{nh} \sum_i \frac{(W_{p,i}Y_i)\left(\left(\lambda_p^{\mathsf{mc}}\right)^\top (W_{p,i}\bar{Z}_i) \right)^2}{1 + \left(\lambda_p^{\mathsf{mc}}\right)^\top (W_{p,i}\bar{Z}_i)}. \tag{S39}$$

Then,

$$\begin{aligned} \left| \frac{1}{nh} \sum_i \frac{(W_{p,i}Y_i)\left(\left(\lambda_p^{\mathsf{mc}}\right)^\top (W_{p,i}\bar{Z}_i) \right)^2}{1 + \left(\lambda_p^{\mathsf{mc}}\right)^\top (W_{p,i}\bar{Z}_i)} \right| &\leq \frac{\left(\frac{1}{nh} \sum_i \left\| W_{p,i}\bar{Z}_i \right\|^2 |W_{p,i}Y_i| \right) \left\| \lambda_p^{\mathsf{mc}} \right\|^2}{1 - \left\| \lambda_p^{\mathsf{mc}} \right\| \max_i \left\| W_{p,i}\bar{Z}_i \right\|} \\ &= O_p\left((nh)^{-1}\right). \end{aligned} \tag{S40}$$

It follows from Lemma 1(c) that $\Psi_{Y\bar{Z},2} - \Delta_{Y\bar{Z},2} = O_p\left((nh)^{-1/2}\right)$. Therefore, $\widehat{\vartheta}_Y^{\mathsf{mc}} = \Psi_Y - \Delta_{Y\bar{Z},2}^\top \lambda_p^{\mathsf{mc}} + O_p\left((nh)^{-1}\right)$. Similarly, $\widehat{\vartheta}_D^{\mathsf{mc}} = \Psi_D - \Delta_{D\bar{Z},2}^\top \lambda_p^{\mathsf{mc}} + O_p\left((nh)^{-1}\right)$. By Lemma 1(b,c), $\Psi_Y - \psi_{Y,\dagger} = O_p\left((nh)^{-1/2}\right)$ and $\Psi_D - \psi_{D,\dagger} = O_p\left((nh)^{-1/2}\right)$. By arguments similar to those used to show (S23),

$$\frac{\widehat{\vartheta}_Y^{\mathsf{mc}}}{\widehat{\vartheta}_D^{\mathsf{mc}}} - \frac{\psi_{Y,\dagger}}{\psi_{D,\dagger}} = \frac{\Psi_M - \Delta_{M\bar{Z},2}^\top \lambda_p^{\mathsf{mc}}}{\psi_{D,\dagger}} + O_p\left((nh)^{-1}\right).$$

Therefore, by this result and (S38),

$$\sqrt{nh}\left(\widehat{\vartheta}_{p}^{\mathsf{mc}}-\vartheta\right)=\frac{1}{\psi_{D,\dagger}}\cdot\frac{1}{\sqrt{nh}}\sum_{i}W_{p,i}\left(\epsilon_{i}-\mu_{\epsilon}\right)+o_{p}\left(1\right), \tag{S41}$$

where $\epsilon_i := M_i - Z_i^{\top}\gamma_M$. Let $\mathring{\epsilon}_i := \epsilon_i - \mu_\epsilon$. Then,

$$\sqrt{nh}\left(\widehat{\vartheta}_{p}^{\mathsf{mc}}-\vartheta-\frac{\Delta_{\mathring{\epsilon}}}{\psi_{D,\dagger}}\right)=\frac{1}{\psi_{D,\dagger}}\cdot\sum_{i}\left(\frac{W_{p,i}\mathring{\epsilon}_{i}}{\sqrt{nh}}-\mathrm{E}\left[\frac{W_{p}\mathring{\epsilon}}{\sqrt{nh}}\right]\right)+o_{p}\left(1\right) \tag{S42}$$

follows from subtracting both sides of (S41) by $\sqrt{nh}\cdot\Delta_{\mathring{\epsilon}}/\psi_{D,\dagger}$. By Lemma 3(b),

$$\Delta_{\mathring{\epsilon}}=\frac{\left(\psi_{\epsilon,+}^{(p+1)}-\mu_{\epsilon}\varphi^{(p+1)}\right)\omega_{p;+}^{p+1,1}-\left(\psi_{\epsilon,-}^{(p+1)}-\mu_{\epsilon}\varphi^{(p+1)}\right)\omega_{p;-}^{p+1,1}}{(p+1)!}\cdot h^{p+1}+o\left(h^{p+1}\right)$$

and $\Delta_{\mathring{\epsilon}^2,2}=\omega_p^{0,2}\psi_{\mathring{\epsilon}^2,\pm}+o\left(1\right)$. It follows from simple algebraic calculations that $\psi_{\mathring{\epsilon}^2,\pm}=\sigma^2\varphi$. Then, we have

$$\mathrm{Var}\left[\frac{W_{p}\mathring{\epsilon}}{\sqrt{h}}\right]=\Delta_{\mathring{\epsilon}^{2}}-h\Delta_{\mathring{\epsilon}}^{2}=\omega_{p}^{0,2}\sigma^{2}\varphi+o\left(1\right). \tag{S43}$$

Let $\varsigma\in(0,1)$. Then,

$$\begin{aligned}\sum_{i}\mathrm{E}\left[\left|\frac{W_{p,i}\mathring{\epsilon}_{i}}{\sqrt{nh}}-\mathrm{E}\left[\frac{W_{p}\mathring{\epsilon}}{\sqrt{nh}}\right]\right|^{2+\varsigma}\right] &= \frac{1}{(nh)^{1+\varsigma/2}}\sum_{i}\mathrm{E}\left[\left|W_{p,i}\mathring{\epsilon}_{i}-\mathrm{E}\left[W_{p}\mathring{\epsilon}\right]\right|^{2+\varsigma}\right]\\ &\lesssim \frac{\mathrm{E}\left[h^{-1}\left|W_{p}\mathring{\epsilon}\right|^{2+\varsigma}\right]+h^{-1}\left|\mathrm{E}\left[W_{p}\mathring{\epsilon}\right]\right|^{2+\varsigma}}{(nh)^{\varsigma/2}},\end{aligned} \tag{S44}$$

where the inequality follows from Loève's c_r inequality and the equality follows from (S43), $\mathrm{E}\left[h^{-1}\left|W_p\mathring{\epsilon}\right|^{2+\varsigma}\right]=O\left(1\right)$ and $\mathrm{E}\left[W_p\mathring{\epsilon}\right]=O\left(h^{p+2}\right)$. (S44) verifies Lyapunov's condition. By Lyapunov's central limit theorem,

$$\frac{\frac{1}{\sqrt{nh}}\sum_{i}\left(W_{p,i}\mathring{\epsilon}_{i}-\mathrm{E}\left[W_{p}\mathring{\epsilon}\right]\right)}{\sqrt{\mathrm{Var}\left[\frac{1}{\sqrt{nh}}\sum_{i}\left(W_{p,i}\mathring{\epsilon}_{i}-\mathrm{E}\left[W_{p}\mathring{\epsilon}\right]\right)\right]}}\to_{d}\mathrm{N}\left(0,1\right).$$

The conclusion follows from this result, (S42), (S43) and Slutsky's lemma. ■

S2 Proof of Theorem 2

S2.1 Calculating the nonlinearity bias of $\widehat{\vartheta}_p^{\mathsf{mc}}$

Let $G_i := \begin{bmatrix} D_i & 0_{d_z+1}^\top \end{bmatrix}^\top$. Denote $\Sigma := \left(\Delta_G^\top \Delta_{UU^\top,2}^{-1} \Delta_G\right)^{-1}$, $\mathrm{N} := \Delta_{UU^\top,2}^{-1}\Delta_G \Sigma$ and $\mathrm{Q} := \Delta_{UU^\top,2}^{-1} - \mathrm{N}\Delta_G^\top \Delta_{UU^\top,2}^{-1}$. By writing $\Delta_{UU^\top,2}$ as a block matrix and inverting,

$$\Delta_{UU^\top,2}^{-1} = \begin{bmatrix} \left(\Delta_{M^2,2} - \Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}\right)^{-1} & -\frac{\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}}{\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}} \\ -\frac{\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}}{\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}} & \Delta_{\bar{Z}\bar{Z}^\top,2}^{-1} - \Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\left(\frac{\Delta_{M\bar{Z},2}\Delta_{M\bar{Z}^\top,2}}{\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}}\right)\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1} \end{bmatrix}. \tag{S45}$$

By straightforward calculation,

$$\mathrm{Q} = \begin{bmatrix} 0 & 0_{d_z+1}^\top \\ 0_{d_z+1} & \Delta_{\bar{Z}\bar{Z}^\top,2}^{-1} \end{bmatrix}. \tag{S46}$$

Consider the singular value decomposition of $\Delta_{UU^\top,2}^{-1/2}(-\Delta_G)$:

$$\mathrm{S}^\top \Delta_{UU^\top,2}^{-1/2}(-\Delta_G)\,\mathrm{T} = \begin{bmatrix} \Lambda \\ 0_{d_z+1} \end{bmatrix},$$

where $\mathrm{S}^\top\mathrm{S} = \mathrm{I}_{d_z+2}$, $\mathrm{T} = 1$ and $\Lambda := \sqrt{\Delta_G^\top \Delta_{UU^\top,2}^{-1}\Delta_G}$. We follow Chen and Cui (2007) to rotate the moment conditions by $\mathrm{S}^\top \Delta_{UU^\top,2}^{-1/2}$ so that results from Chen and Cui (2007) can be applied. Let $V_i(\theta) := \mathrm{S}^\top \Delta_{UU^\top,2}^{-1/2} U_i(\theta)$, $V_i := V_i(\vartheta)$ and $H_i := \mathrm{S}^\top \Delta_{UU^\top,2}^{-1/2}(-G_i)$. It is easy to see that the criterion function is invariant to the rotation:

$$\ell_p^{\mathsf{mc}}(\theta \mid h) = \sup_\lambda \frac{1}{n}\sum_i \log\left(1 + \lambda^\top \left(W_{p,i} V_i(\theta)\right)\right).$$

Let $\widehat{\lambda}_p^{\mathsf{mc}}$ denote the Lagrange multipliers

$$\widehat{\lambda}_p^{\mathsf{mc}} := \operatorname*{argmax}_\lambda \frac{1}{n}\sum_i \log\left(1 + \lambda^\top\left(W_{p,i} V_i\left(\widehat{\vartheta}_p^{\mathsf{mc}}\right)\right)\right).$$

It is clear that we have $\widehat{\lambda}_p^{\mathsf{mc}} = \left(\mathrm{S}^\top \Delta_{UU^\top,2}^{1/2}\right)\left(0, \left(\lambda_p^{\mathsf{mc}}\right)^\top\right)^\top$. Lemma 4 implies that $\widehat{\lambda}_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$. Let $\Omega := \Lambda^{-1}$. Then, $\Delta_{VV^\top,2} = \mathrm{I}_{d_z+2}$ and $\Delta_H := \mathrm{S}^\top \Delta_{UU^\top,2}^{-1/2}(-\Delta_G) = \begin{bmatrix} \Lambda & 0_{d_z+1}^\top \end{bmatrix}^\top$. By inverting the block matrices and simple algebra,

$$
\begin{aligned}
&\begin{bmatrix} -\Delta_{VV^\top,2} & \Delta_H \\ \Delta_H^\top & 0 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} -\Delta_{VV^\top,2}^{-1} + \Delta_{VV^\top,2}^{-1}\Delta_H\left(\Delta_H^\top\Delta_{VV^\top,2}^{-1}\Delta_H\right)^{-1}\Delta_H^\top\Delta_{VV^\top,2}^{-1} & \Delta_{VV^\top,2}^{-1}\Delta_H\left(\Delta_H^\top\Delta_{VV^\top,2}^{-1}\Delta_H\right)^{-1} \\ \left(\Delta_H^\top\Delta_{VV^\top,2}^{-1}\Delta_H\right)^{-1}\Delta_H^\top\Delta_{VV^\top,2}^{-1} & \left(\Delta_H^\top\Delta_{VV^\top,2}^{-1}\Delta_H\right)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0_{d_z+1}^\top & \Omega \\ 0_{d_z+1} & -\mathrm{I}_{d_z+1} & 0_{d_z+1} \\ \Omega & 0_{d_z+1}^\top & \Omega^2 \end{bmatrix} = \begin{bmatrix} -\mathrm{S}^\top\Delta_{UU^\top,2}^{1/2}\mathrm{Q}\Delta_{UU^\top,2}^{1/2}\mathrm{S} & -\mathrm{S}^\top\Delta_{UU^\top,2}^{1/2}\mathrm{N} \\ -\mathrm{N}^\top\Delta_{UU^\top,2}^{1/2}\mathrm{S} & \Sigma \end{bmatrix}. \qquad \text{(S47)}
\end{aligned}
$$

The first order conditions are given by

$$
\begin{aligned}
0 &= \sum_i \frac{W_{p,i}V_i\left(\widehat{\vartheta}_p^{\mathsf{mc}}\right)}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}V_i\left(\widehat{\vartheta}_p^{\mathsf{mc}}\right)\right)} \\
0 &= \sum_i \frac{\left(W_{p,i}H_i\right)^\top\widehat{\lambda}_p^{\mathsf{mc}}}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}V_i\left(\widehat{\vartheta}_p^{\mathsf{mc}}\right)\right)}. \qquad \text{(S48)}
\end{aligned}
$$

Let $\mathring{\vartheta}_p^{\mathsf{mc}} := \widehat{\vartheta}_p^{\mathsf{mc}} - \vartheta$. Theorem 1 implies that $\mathring{\vartheta}_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$. Denote $\widehat{V}_i^{\mathsf{mc}} := V_i\left(\widehat{\vartheta}_p^{\mathsf{mc}}\right)$. By simple algebra, we expand the right hand sides of (S48) to get

$$
\begin{aligned}
0 &= \sum_i W_{p,i}\widehat{V}_i^{\mathsf{mc}}\left\{1-\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)+\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2-\frac{\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^3}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)}\right\} \\
0 &= \sum_i \left(W_{p,i}H_i\right)^\top\widehat{\lambda}_p^{\mathsf{mc}}\left\{1-\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)+\frac{\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)}\right\}. \qquad \text{(S49)}
\end{aligned}
$$

By $\widehat{V}_i^{\mathsf{mc}} = V_i + H_i\mathring{\vartheta}_p^{\mathsf{mc}}$, triangle inequality, Lemma 3(a,c,d), Markov's inequality, Lemma 4 and $\mathring{\vartheta}_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$, we have

$$
\begin{aligned}
\frac{1}{nh}\sum_i \frac{W_{p,i}\widehat{V}_i^{\mathsf{mc}}\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^3}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)} &= O_p\left((nh)^{-3/2}\right) \\
\frac{1}{nh}\sum_i \frac{\left(\left(W_{p,i}H_i\right)^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)} &= O_p\left((nh)^{-3/2}\right).
\end{aligned}
$$

By $\widehat{V}_i^{\mathsf{mc}} = V_i + H_i \mathring{\vartheta}_p^{\mathsf{mc}}$, $\mathring{\vartheta}_p^{\mathsf{mc}} = O_p\left((nh)^{-1/2}\right)$, Lemma 3(a), and Markov's inequality,

$$
\begin{aligned}
\frac{1}{nh}\sum_i W_{p,i}\widehat{V}_i^{\mathsf{mc}} &= \frac{1}{nh}\sum_i W_{p,i}V_i + \frac{1}{nh}\sum_i W_{p,i}H_i\mathring{\vartheta}_p^{\mathsf{mc}} \\
\frac{1}{nh}\sum_i W_{p,i}\widehat{V}_i^{\mathsf{mc}}\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right) &= \frac{1}{nh}\sum_i W_{p,i}^2 V_i V_i^\top \widehat{\lambda}_p^{\mathsf{mc}} + \frac{1}{nh}\sum_i W_{p,i}^2 V_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right) \\
&\quad + \frac{1}{nh}\sum_i W_{p,i}^2 H_i\mathring{\vartheta}_p^{\mathsf{mc}}\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right) + O_p\left((nh)^{-3/2}\right) \\
\frac{1}{nh}\sum_i W_{p,i}\widehat{V}_i^{\mathsf{mc}}\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2 &= \frac{1}{nh}\sum_i W_{p,i}^3 V_i\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2 + O_p\left((nh)^{-3/2}\right) \\
\frac{1}{nh}\sum_i \left(W_{p,i}H_i\right)^\top\widehat{\lambda}_p^{\mathsf{mc}}\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right) &= \frac{1}{nh}\sum_i W_{p,i}^2\left(H_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right) + O_p\left((nh)^{-3/2}\right)
\end{aligned}
$$

By plugging these results into the right hand side of (S49), we have

$$
\begin{aligned}
-\Delta_{VV^\top,2}\widehat{\lambda}_p^{\mathsf{mc}} + \Delta_H\mathring{\vartheta}_p^{\mathsf{mc}} &= -\frac{1}{nh}\sum_i W_{p,i}V_i + \frac{1}{nh}\sum_i W_{p,i}^2 V_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right) - \frac{1}{nh}\sum_i W_{p,i}^3 V_i\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2 \\
&\quad + \frac{1}{nh}\sum_i W_{p,i}^2 H_i\mathring{\vartheta}_p^{\mathsf{mc}}\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right) + \left(\Psi_{VV^\top,2} - \Delta_{VV^\top,2}\right)\widehat{\lambda}_p^{\mathsf{mc}} - \left(\Psi_H - \Delta_H\right)\mathring{\vartheta}_p^{\mathsf{mc}} \\
&\quad + O_p\left((nh)^{-3/2}\right) \\
\Delta_H\widehat{\lambda}_p^{\mathsf{mc}} &= \frac{1}{nh}\sum_i W_{p,i}^2\left(H_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right) - \left(\Psi_H - \Delta_H\right)^\top\widehat{\lambda}_p^{\mathsf{mc}} + O_p\left((nh)^{-3/2}\right). \quad \text{(S50)}
\end{aligned}
$$

We now switch to coordinate notations. Let $\alpha^{\mathsf{k}} := \Delta_{V^{(\mathsf{k})}}$, $\alpha^{\mathsf{kl}} := \Delta_{V^{(\mathsf{k})}V^{(\mathsf{l})},2}$, $\alpha^{\mathsf{klm}} := \Delta_{V^{(\mathsf{k})}V^{(\mathsf{l})}V^{(\mathsf{m})},3}$, and $\gamma^{\mathsf{k:l}} := \Delta_{V^{(\mathsf{k})}H^{(\mathsf{l})},2}$. Let $A^{\mathsf{k}} := \Psi_{V^{(\mathsf{k})}}$, $A^{\mathsf{kl}} := \Psi_{V^{(\mathsf{k})}V^{(\mathsf{l})},2} - \Delta_{V^{(\mathsf{k})}V^{(\mathsf{l})},2}$ and $C^{\mathsf{k}} := \Psi_{H^{(\mathsf{k})}} - \Delta_{H^{(\mathsf{k})}}$. By multiplying both sides of (S50) by (S47) and replacing $\left(\widehat{\lambda}_p^{\mathsf{mc}}, \mathring{\vartheta}_p^{\mathsf{mc}}\right)$ with their leading terms in the stochastic expansion, we can get a stochastic expansion for $\left(\widehat{\lambda}_p^{\mathsf{mc}}, \mathring{\vartheta}_p^{\mathsf{mc}}\right)$ in the form of a quadratic polynomial of $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, C^{\mathsf{k}}\right)$. The algebraic calculations have been done in Chen and Cui (2007) so that we use them directly here. In the following proofs, summation over repeated indices is taken implicitly with the "$\sum$" notation suppressed and ranges of indices fixed: $\mathsf{k},\mathsf{l},\mathsf{m},\mathsf{n},\mathsf{o},\mathsf{v} = 1,...,d_z+2$ and $\mathsf{a},\mathsf{b},\mathsf{c},\mathsf{d} = 1,...,d_z+1$, $\mathsf{e},\mathsf{f},\mathsf{g},\mathsf{h} = 1,...,p$. By these calculations and Lemma 3(c), we have

$$
\mathring{\vartheta}_p^{\mathsf{mc}} = -\Omega A^1 + \Omega A^{1\,1+\mathsf{a}}A^{1+\mathsf{a}} + \Omega^2 C^1 A^1 - \Omega\alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}}A^{1+\mathsf{a}}A^{1+\mathsf{b}} - \Omega^2\gamma^{1+\mathsf{a}:1}A^{1+\mathsf{a}}A^1 + O_p\left((nh)^{-3/2}\right).
$$

Let $\mathring{A}^{\mathsf{k}} := \Psi_{V^{(\mathsf{k})}} - \alpha^{\mathsf{k}}$, then we have

$$
\mathring{\vartheta}_p^{\mathsf{mc}} = -\Omega\left(\mathring{A}^1 + \alpha^1\right) + I_1^{\mathsf{mc}} + I_2^{\mathsf{mc}} + I_3^{\mathsf{mc}} + O_p\left((nh)^{-3/2}\right), \quad \text{(S51)}
$$

where

$$\begin{aligned}
I_1^{\mathsf{mc}} &\coloneqq \Omega A^{1\,1+\mathsf{a}} \mathring{A}^{1+\mathsf{a}} + \Omega^2 C^1 \mathring{A}^1 - \Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \mathring{A}^{1+\mathsf{a}} \mathring{A}^{1+\mathsf{b}} - \Omega^2 \gamma^{1+\mathsf{a}:1} \mathring{A}^{1+\mathsf{a}} \mathring{A}^1 \\
I_2^{\mathsf{mc}} &\coloneqq \Omega A^{1\,1+\mathsf{a}} \alpha^{1+\mathsf{a}} + \Omega^2 C^1 \alpha^1 - 2\Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \mathring{A}^{1+\mathsf{a}} \alpha^{1+\mathsf{b}} - \Omega^2 \gamma^{1+\mathsf{a}:1} \mathring{A}^{1+\mathsf{a}} \alpha^1 - \Omega^2 \gamma^{1+\mathsf{a}:1} \alpha^{1+\mathsf{a}} \mathring{A}^1 \\
I_3^{\mathsf{mc}} &\coloneqq -\Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \alpha^{1+\mathsf{a}} \alpha^{1+\mathsf{b}} - \Omega^2 \gamma^{1+\mathsf{a}:1} \alpha^{1+\mathsf{a}} \alpha^1.
\end{aligned}$$

I_2^{mc} is the interaction term of the smoothing bias and stochastic variability. I_3^{mc} is the second-order smoothing bias. The nonlinearity bias is defined to be the leading term of the expectation of I_1^{mc}. By straightforward calculation and Lemma 3(b), we have

$$\begin{aligned}
\mathrm{E}\left[A^{1\,1+\mathsf{a}} \mathring{A}^{1+\mathsf{a}}\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\left(\frac{1}{h} W_p^2 V^{(1)} V^{(1+\mathsf{a})}\right)\left(\frac{1}{h} W_p V^{(1+\mathsf{a})}\right)\right] - \mathrm{E}\left[\frac{1}{h} W_p^2 V^{(1)} V^{(1+\mathsf{a})}\right] \mathrm{E}\left[\frac{1}{h} W_p V^{(1+\mathsf{a})}\right]\right\} \\
&= \frac{1}{nh} \cdot \alpha^{1\,1+\mathsf{a}\,1+\mathsf{a}} \\
\mathrm{E}\left[C^1 \mathring{A}^1\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\left(\frac{1}{h} W_p H^{(1)}\right)\left(\frac{1}{h} W_p V^{(1)}\right)\right] - \mathrm{E}\left[\frac{1}{h} W_p H^{(1)}\right] \mathrm{E}\left[\frac{1}{h} W_p V^{(1)}\right]\right\} \\
&= \frac{1}{nh} \cdot \gamma^{1:1} + O\left(n^{-1} h^{1+p}\right) \\
\mathrm{E}\left[\mathring{A}^{1+\mathsf{a}} \mathring{A}^{1+\mathsf{b}}\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\left(\frac{1}{h} W_p V^{(1+\mathsf{a})}\right)\left(\frac{1}{h} W_p V^{(1+\mathsf{b})}\right)\right] - \mathrm{E}\left[\frac{1}{h} W_p V^{(1+\mathsf{a})}\right] \mathrm{E}\left[\frac{1}{h} W_p V^{(1+\mathsf{b})}\right]\right\} \\
&= \mathbb{1}\left(\mathsf{a} = \mathsf{b}\right) + O\left(n^{-1} h^{2(1+p)}\right) \\
\mathrm{E}\left[\mathring{A}^{1+\mathsf{a}} \mathring{A}^1\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\left(\frac{1}{h} W_p V^{(1+\mathsf{a})}\right)\left(\frac{1}{h} W_p V^{(1)}\right)\right] - \mathrm{E}\left[\frac{1}{h} W_p V^{(1+\mathsf{a})}\right] \mathrm{E}\left[\frac{1}{h} W_p V^{(1)}\right]\right\} \\
&= O\left(n^{-1} h^{2(1+p)}\right).
\end{aligned}$$

Therefore, we have $\mathrm{E}\left[I_1^{\mathsf{mc}}\right] = (nh)^{-1}\left(\Omega^2 \gamma^{1:1} + o(1)\right)$. Since

$$\frac{\mathrm{S}^\top \Delta_{UU^\top,2}^{-1/2}\left(-\Delta_G\right)}{\sqrt{\Delta_G^\top \Delta_{UU^\top,2}^{-1} \Delta_G}} = \mathrm{e}_{d_z+2,1}, \tag{S52}$$

by (S47), we have

$$\begin{aligned}
H^{(1)} &= \frac{\Delta_G^\top \Delta_{UU^\top,2}^{-1} G}{\sqrt{\Delta_G^\top \Delta_{UU^\top,2}^{-1} \Delta_G}} \\
V^{(1)} &= -\frac{\Delta_G^\top \Delta_{UU^\top,2}^{-1} U}{\sqrt{\Delta_G^\top \Delta_{UU^\top,2}^{-1} \Delta_G}}.
\end{aligned}$$

Therefore

$$\Omega^2\gamma^{1:1} = -\frac{\mathrm{E}\left[\frac{1}{h}W_p^2\left(\Delta_G^\top\Delta_{UU^\top,2}^{-1}G\right)\left(\Delta_G^\top\Delta_{UU^\top,2}^{-1}U\right)\right]}{\left(\Delta_G\Delta_{UU^\top,2}^{-1}\Delta_G\right)^2}. \tag{S53}$$

By (S45), we have

$$\begin{aligned}
\Delta_G^\top\Delta_{UU^\top,2}^{-1}G &= \frac{\Delta_D\cdot D}{\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}} \\
\Delta_G^\top\Delta_{UU^\top,2}^{-1}U &= \frac{\Delta_D\left(M-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z}\right)}{\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}} \\
\Delta_G^\top\Delta_{UU^\top,2}^{-1}\Delta_G &= \frac{\Delta_D^2}{\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}}.
\end{aligned} \tag{S54}$$

By (S37),

$$M-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z} = \left(M-\frac{\Delta_{M,2}}{\Delta_2}\right)-\bar{\gamma}_M^\top\left(Z-\frac{\Delta_{Z,2}}{\Delta_2}\right).$$

By these results, $\bar{\gamma}_M = \gamma_M + o(1)$, $\Delta_{M,2}/\Delta_2 = \mu_M + o(1)$, $\Delta_{Z,2}/\Delta_2 = \mu_Z + o(1)$, (S53) and Lemma 3(a,b),

$$\begin{aligned}
\Omega^2\gamma^{1:1} &= -\Delta_D^{-2}\cdot\mathrm{E}\left[\frac{1}{h}W_p^2\left(M-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z}\right)D\right] \\
&= \omega_p^{0,2}\cdot\frac{\mathrm{Cov}_{|0^\pm}[\epsilon,D]}{\varphi\mu_{D,\dagger}^2}+o(1).
\end{aligned} \tag{S55}$$

S2.2 Calculating the nonlinearity bias of $\widehat{\vartheta}_p^{\mathsf{eb}}$

Similarly, denote $\widetilde{\Sigma} := \left(\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G\right)^{-1}$, $\widetilde{\mathrm{N}} := \widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G\widetilde{\Sigma}$ and $\widetilde{\mathrm{Q}} := \widetilde{\Delta}_{UU^\top,2}^{-1} - \widetilde{\mathrm{N}}\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}$. By writing $\widetilde{\Delta}_{UU^\top,2}$ as a block matrix and inverting,

$$\widetilde{\Delta}_{UU^\top,2}^{-1} = \begin{bmatrix} \left(\widetilde{\Delta}_{M^2,2}-\widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}\right)^{-1} & -\frac{\widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}}{\widetilde{\Delta}_{M^2,2}-\widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}} \\ -\frac{\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}}{\widetilde{\Delta}_{M^2,2}-\widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}} & \widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}-\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\left(\frac{\widetilde{\Delta}_{M\bar{Z},2}\widetilde{\Delta}_{M\bar{Z}^\top,2}}{\widetilde{\Delta}_{M^2,2}-\widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}}\right)\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1} \end{bmatrix} \tag{S56}$$

and by straightforward calculation,

$$\widetilde{\mathrm{Q}} = \begin{bmatrix} 0 & 0_{d_z+1}^\top \\ 0_{d_z+1} & \widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1} \end{bmatrix}. \tag{S57}$$

Consider the singular value decomposition of $\widetilde{\Delta}_{UU^\top,2}^{-1/2}\left(-\widetilde{\Delta}_G\right)$:

$$\mathrm{S}^\top \widetilde{\Delta}_{UU^\top,2}^{-1/2}\left(-\widetilde{\Delta}_G\right)\mathrm{T} = \begin{bmatrix} \Lambda \\ 0_{d_z+1} \end{bmatrix},$$

where $\mathrm{S}^\top\mathrm{S} = \mathrm{I}_{d_z+2}$, $\mathrm{T} = 1$ and $\Lambda \coloneqq \sqrt{\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G}$. Let $V_i(\theta) \coloneqq \mathrm{S}^\top \widetilde{\Delta}_{UU^\top,2}^{-1/2} U_i(\theta)$, $V_i \coloneqq \mathrm{S}^\top \widetilde{\Delta}_{UU^\top,2}^{-1/2} U_i$ and $H_i \coloneqq \mathrm{S}^\top \widetilde{\Delta}_{UU^\top,2}^{-1/2}(-G_i)$. The estimator $\widehat{\vartheta}_p^{\mathsf{eb}} = \operatorname{argmin}_\theta \ell_p^{\mathsf{eb}}(\theta \mid h)$ is invariant the rotation, where

$$\ell_p^{\mathsf{eb}}(\theta \mid h) = \sup_\lambda \frac{1}{n}\sum_i \log\left(1 + \lambda^\top\left(\widehat{W}_{p,i}V_i(\theta)\right)\right).$$

Let $\widehat{\lambda}_p^{\mathsf{eb}}$ to denote the associated Lagrange multipliers:

$$\widehat{\lambda}_p^{\mathsf{eb}} \coloneqq \operatorname*{argmax}_\lambda \frac{1}{n}\sum_i \log\left(1 + \lambda^\top\left(\widehat{W}_{p,i}V_i\left(\widehat{\vartheta}_p^{\mathsf{eb}}\right)\right)\right).$$

Then $\widehat{\lambda}_p^{\mathsf{eb}} = \mathrm{S}^\top \widetilde{\Delta}_{UU^\top,2}^{1/2}\left(0, \left(\lambda_p^{\mathsf{eb}}\right)^\top\right)^\top$ and Lemma 2 implies that $\widehat{\lambda}_p^{\mathsf{eb}} = O_p\left((nh)^{-1/2}\right)$. Let $\Omega \coloneqq \Lambda^{-1}$. Then, $\widetilde{\Delta}_{VV^\top,2} = \mathrm{I}_{d_z+2}$ and $\widetilde{\Delta}_H \coloneqq \mathrm{S}^\top \widetilde{\Delta}_{UU^\top,2}^{-1/2}\left(-\widetilde{\Delta}_G\right) = \begin{bmatrix} \Lambda & 0_{d_z+1}^\top \end{bmatrix}^\top$. By inverting the block matrices,

$$\begin{aligned}
&\begin{bmatrix} -\widetilde{\Delta}_{VV^\top,2} & \widetilde{\Delta}_H \\ \widetilde{\Delta}_H^\top & 0 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} -\widetilde{\Delta}_{VV^\top,2}^{-1} + \widetilde{\Delta}_{VV^\top,2}^{-1}\widetilde{\Delta}_H\left(\widetilde{\Delta}_H^\top\widetilde{\Delta}_{VV^\top,2}^{-1}\widetilde{\Delta}_H\right)^{-1}\widetilde{\Delta}_H^\top\widetilde{\Delta}_{VV^\top,2}^{-1} & \widetilde{\Delta}_{VV^\top,2}^{-1}\widetilde{\Delta}_H\left(\widetilde{\Delta}_H^\top\widetilde{\Delta}_{VV^\top,2}^{-1}\widetilde{\Delta}_H\right)^{-1} \\ \left(\widetilde{\Delta}_H^\top\widetilde{\Delta}_{VV^\top,2}^{-1}\widetilde{\Delta}_H\right)^{-1}\widetilde{\Delta}_H^\top\widetilde{\Delta}_{VV^\top,2}^{-1} & \left(\widetilde{\Delta}_H^\top\widetilde{\Delta}_{VV^\top,2}^{-1}\widetilde{\Delta}_H\right)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0_{d_z+1}^\top & \Omega \\ 0_{d_z+1} & -\mathrm{I}_{d_z+1} & 0_{d_z+1} \\ \Omega & 0_{d_z+1}^\top & \Omega^2 \end{bmatrix} = \begin{bmatrix} -\mathrm{S}^\top\widetilde{\Delta}_{UU^\top,2}^{1/2}\widetilde{\mathrm{Q}}\widetilde{\Delta}_{UU^\top,2}^{1/2}\mathrm{S} & -\mathrm{S}^\top\widetilde{\Delta}_{UU^\top,2}^{1/2}\widetilde{\mathrm{N}} \\ -\widetilde{\mathrm{N}}^\top\widetilde{\Delta}_{UU^\top,2}^{1/2}\mathrm{S} & \widetilde{\Sigma} \end{bmatrix}. \qquad \text{(S58)}
\end{aligned}$$

Let $\mathring{\vartheta}_p^{\mathsf{eb}} \coloneqq \widehat{\vartheta}_p^{\mathsf{eb}} - \vartheta$. Theorem 1 implies that $\mathring{\vartheta}_p^{\mathsf{eb}} = O_p\left((nh)^{-1/2}\right)$. We expand the right hand sides of the first order conditions

$$\begin{aligned}
0 &= \sum_i \frac{\widehat{W}_{p,i}V_i\left(\widehat{\vartheta}_p^{\mathsf{eb}}\right)}{1 + \left(\widehat{\lambda}_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}V_i\left(\widehat{\vartheta}_p^{\mathsf{eb}}\right)\right)} \\
0 &= \sum_i \frac{\left(\widehat{W}_{p,i}H_i\right)^\top\widehat{\lambda}_p^{\mathsf{eb}}}{1 + \left(\widehat{\lambda}_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}V_i\left(\widehat{\vartheta}_p^{\mathsf{eb}}\right)\right)}
\end{aligned}$$

and use Lemma 1 and $V_i\left(\widehat{\vartheta}_p^{\mathsf{eb}}\right) = V_i + H_i \mathring{\vartheta}_p^{\mathsf{eb}}$ to get

$$\begin{aligned}
-\widetilde{\Delta}_{VV^\top,2}\widehat{\lambda}_p^{\mathsf{eb}} + \widetilde{\Delta}_H \mathring{\vartheta}_p^{\mathsf{eb}} &= -\frac{1}{nh}\sum_i \widehat{W}_{p,i}V_i + \frac{1}{nh}\sum_i \widehat{W}_{p,i}^2 V_i \left(\left(\widehat{\lambda}_p^{\mathsf{eb}}\right)^\top H_i \mathring{\vartheta}_p^{\mathsf{eb}}\right) - \frac{1}{nh}\sum_i \widehat{W}_{p,i}^3 V_i \left(V_i^\top \widehat{\lambda}_p^{\mathsf{eb}}\right)^2 \\
&\quad + \frac{1}{nh}\sum_i \widehat{W}_{p,i}^2 H_i \mathring{\vartheta}_p^{\mathsf{eb}} \left(V_i^\top \widehat{\lambda}_p^{\mathsf{eb}}\right) + \left(\widehat{\Psi}_{VV^\top,2} - \widetilde{\Delta}_{VV^\top,2}\right)\widehat{\lambda}_p^{\mathsf{eb}} - \left(\widehat{\Psi}_H - \widetilde{\Delta}_H\right)\mathring{\vartheta}_p^{\mathsf{eb}} \\
&\quad + O_p\left((nh)^{-3/2}\right) \\
\widetilde{\Delta}_H \widehat{\lambda}_p^{\mathsf{eb}} &= \frac{1}{nh}\sum_i \widehat{W}_{p,i}^2 \left(H_i^\top \widehat{\lambda}_p^{\mathsf{eb}}\right)\left(V_i^\top \widehat{\lambda}_p^{\mathsf{eb}}\right) - \left(\widehat{\Psi}_H - \widetilde{\Delta}_H\right)^\top \widehat{\lambda}_p^{\mathsf{eb}} + O_p\left((nh)^{-3/2}\right). \qquad \text{(S59)}
\end{aligned}$$

Denote $A^{\mathsf{k}} := \widehat{\Psi}_{V^{(\mathsf{k})}}$, $A^{\mathsf{kl}} := \widehat{\Psi}_{V^{(\mathsf{k})}V^{(\mathsf{l})},2} - \widetilde{\Delta}_{V^{(\mathsf{k})}V^{(\mathsf{l})},2}$ and $C^{\mathsf{k}} := \widehat{\Psi}_{H^{(\mathsf{k})}} - \widetilde{\Delta}_{H^{(\mathsf{k})}}$. Let $\alpha^{\mathsf{klm}} := \widetilde{\Delta}_{V^{(\mathsf{k})}V^{(\mathsf{l})}V^{(\mathsf{m})},3}$, and $\gamma^{\mathsf{k:l}} := \widetilde{\Delta}_{V^{(\mathsf{k})}H^{(\mathsf{l})},2}$. By (S58), (S59), calculations in Chen and Cui (2007) and Lemma 1(c),

$$\mathring{\vartheta}_p^{\mathsf{eb}} = -\Omega A^1 + \Omega A^{1\,1+\mathsf{a}} A^{1+\mathsf{a}} + \Omega^2 C^1 A^1 - \Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} A^{1+\mathsf{a}} A^{1+\mathsf{b}} - \Omega^2 \gamma^{1+\mathsf{a}:1} A^{1+\mathsf{a}} A^1 + O_p\left((nh)^{-3/2}\right). \quad \text{(S60)}$$

Denote $\Phi_{\mathrm{s}}^{\mathsf{ef}} := \left(\Pi_{p,\mathrm{s}}^{-1}\right)^{(\mathsf{ef})}$ and $S_{\mathrm{s}}^{\mathsf{ef}} := \widehat{\Pi}_{p,\mathrm{s}}^{(\mathsf{ef})} - \Pi_{p,\mathrm{s}}^{(\mathsf{ef})}$ for $\mathrm{s} \in \{-,+\}$. Let

$$\begin{aligned}
S_+^{\mathsf{e:k}} &:= \frac{1}{nh}\sum_i r_p^{(\mathsf{e})}\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}(X_i > 0)\left(V_i^{(\mathsf{k})} - g_{V^{(\mathsf{k})}}(X_i)\right) \\
S_-^{\mathsf{e:k}} &:= \frac{1}{nh}\sum_i r_p^{(\mathsf{e})}\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}(X_i < 0)\left(V_i^{(\mathsf{k})} - g_{V^{(\mathsf{k})}}(X_i)\right) \\
\Upsilon_+^{\mathsf{ek}} &:= \mathrm{E}\left[\frac{1}{h} r_p^{(\mathsf{e})}\left(\frac{X}{h}\right) K\left(\frac{X}{h}\right) \mathbb{1}(X > 0) H^{(\mathsf{k})}\right] \\
\Upsilon_-^{\mathsf{ek}} &:= \mathrm{E}\left[\frac{1}{h} r_p^{(\mathsf{e})}\left(\frac{X}{h}\right) K\left(\frac{X}{h}\right) \mathbb{1}(X < 0) H^{(\mathsf{k})}\right] \\
\Pi_+^{\mathsf{ef:kl}} &:= \mathrm{E}\left[\frac{1}{h} r_p^{(\mathsf{e})}\left(\frac{X}{h}\right) r_p^{(\mathsf{f})}\left(\frac{X}{h}\right) K\left(\frac{X}{h}\right) \mathbb{1}(X > 0) V^{(\mathsf{k})} V^{(\mathsf{l})}\right] \\
\Pi_-^{\mathsf{ef:kl}} &:= \mathrm{E}\left[\frac{1}{h} r_p^{(\mathsf{e})}\left(\frac{X}{h}\right) r_p^{(\mathsf{f})}\left(\frac{X}{h}\right) K\left(\frac{X}{h}\right) \mathbb{1}(X < 0) V^{(\mathsf{k})} V^{(\mathsf{l})}\right].
\end{aligned}$$

Let $\alpha^{\mathsf{k}} := (nh)^{-1}\sum_i \widehat{W}_{p,i} g_{V^{(\mathsf{k})}}(X_i)$, $\widetilde{A}^{\mathsf{k}} := (nh)^{-1}\sum_i \widetilde{W}_{p,i}\left(V_i^{(\mathsf{k})} - g_{V^{(\mathsf{k})}}(X_i)\right)$, $\widetilde{A}^{\mathsf{kl}} := \widetilde{\Psi}_{V^{(\mathsf{k})}V^{(\mathsf{l})},2} - \widetilde{\Delta}_{V^{(\mathsf{k})}V^{(\mathsf{l})},2}$ and $\widetilde{C}^{\mathsf{k}} := \widetilde{\Psi}_{H^{(\mathsf{k})}} - \widetilde{\Delta}_{H^{(\mathsf{k})}}$. Note that

$$\begin{aligned}
A^{\mathsf{k}} - \alpha^{\mathsf{k}} - \widetilde{A}^{\mathsf{k}} &= \frac{1}{nh}\sum_i \left(\widehat{W}_{p,i} - \widetilde{W}_{p,i}\right)\left(V_i^{(\mathsf{k})} - g_{V^{(\mathsf{k})}}(X_i)\right) \\
&= \frac{1}{nh}\sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,+}^{-1} - \Pi_{p,+}^{-1}\right) r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}(X_i > 0)\left(V_i^{(\mathsf{k})} - g_{V^{(\mathsf{k})}}(X_i)\right) \\
&\quad - \frac{1}{nh}\sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,-}^{-1} - \Pi_{p,-}^{-1}\right) r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}(X_i < 0)\left(V_i^{(\mathsf{k})} - g_{V^{(\mathsf{k})}}(X_i)\right),
\end{aligned}$$

$$\begin{aligned} A^{\mathsf{kl}} - \widetilde{A}^{\mathsf{kl}} &= \frac{1}{nh} \sum_i \left(\widehat{W}_{p,i} - \widetilde{W}_{p,i} \right) V_i^{(\mathsf{k})} V_i^{(\mathsf{l})} \\ &= \frac{1}{nh} \sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,+}^{-1} - \Pi_{p,+}^{-1} \right) r_p \left(\frac{X_i}{h} \right) K \left(\frac{X_i}{h} \right) \mathbb{1}\left(X_i > 0 \right) V_i^{(\mathsf{k})} V_i^{(\mathsf{l})} \\ &\quad - \frac{1}{nh} \sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,-}^{-1} - \Pi_{p,-}^{-1} \right) r_p \left(\frac{X_i}{h} \right) K \left(\frac{X_i}{h} \right) \mathbb{1}\left(X_i < 0 \right) V_i^{(\mathsf{k})} V_i^{(\mathsf{l})} \end{aligned}$$

and

$$\begin{aligned} C^{\mathsf{k}} - \widetilde{C}^{\mathsf{k}} &= \frac{1}{nh} \sum_i \left(\widehat{W}_{p,i} - \widetilde{W}_{p,i} \right) H_i^{(\mathsf{k})} \\ &= \frac{1}{nh} \sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,+}^{-1} - \Pi_{p,+}^{-1} \right) r_p \left(\frac{X_i}{h} \right) K \left(\frac{X_i}{h} \right) \mathbb{1}\left(X_i > 0 \right) H_i^{(\mathsf{k})} \\ &\quad - \frac{1}{nh} \sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,-}^{-1} - \Pi_{p,-}^{-1} \right) r_p \left(\frac{X_i}{h} \right) K \left(\frac{X_i}{h} \right) \mathbb{1}\left(X_i < 0 \right) H_i^{(\mathsf{k})}. \end{aligned}$$

By Chebyshev's inequality, change of variables and also (S1), (S4) and (S5),

$$\begin{aligned} A^{\mathsf{k}} &= \widetilde{A}^{\mathsf{k}} + \alpha^{\mathsf{k}} - \Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} S_+^{\mathsf{g:k}} + \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} S_-^{\mathsf{g:k}} + O_p\left((nh)^{-3/2} \right) \\ A^{\mathsf{kl}} &= \widetilde{A}^{\mathsf{kl}} - 2 \cdot \Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} \Pi_+^{\mathsf{gh:kl}} \Phi_+^{\mathsf{h}1} + 2 \cdot \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} \Pi_-^{\mathsf{gh:kl}} \Phi_-^{\mathsf{h}1} + O_p\left((nh)^{-1} \right) \\ C^{\mathsf{k}} &= \widetilde{C}^{\mathsf{k}} - \Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} \Upsilon_+^{\mathsf{gk}} + \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} \Upsilon_-^{\mathsf{gk}} + O_p\left((nh)^{-1} \right). \end{aligned}$$

Then, by these results and (S60),

$$\mathring{\vartheta}_p^{\mathsf{eb}} = -\Omega \left(\widetilde{A}^1 + \alpha^{\mathsf{k}} \right) + I_1^{\mathsf{eb}} + I_2^{\mathsf{eb}} + I_3^{\mathsf{eb}} + O_p\left((nh)^{-3/2} \right),$$

where

$$\begin{aligned} I_1^{\mathsf{eb}} &:= -\Omega \left(-\Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} S_+^{\mathsf{g:1}} + \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} S_-^{\mathsf{g:1}} \right) \\ &\quad + \Omega \left(\widetilde{A}^{1\,1+\mathsf{a}} - 2 \cdot \Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} \Pi_+^{\mathsf{gh:1\,1+a}} \Phi_+^{\mathsf{h}1} - 2 \cdot \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} \Pi_-^{\mathsf{gh:1\,1+a}} \Phi_-^{\mathsf{h}1} \right) \widetilde{A}^{1+\mathsf{a}} \\ &\quad + \Omega^2 \left(\widetilde{C}^1 - \Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} \Upsilon_+^{\mathsf{g}1} + \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} \Upsilon_-^{\mathsf{g}1} \right) \widetilde{A}^1 - \Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \widetilde{A}^{1+\mathsf{a}} \widetilde{A}^{1+\mathsf{b}} - \Omega^2 \gamma^{1+\mathsf{a}:1} \widetilde{A}^{1+\mathsf{a}} \widetilde{A}^1 \\ I_2^{\mathsf{eb}} &:= \Omega \left(\widetilde{A}^{1\,1+\mathsf{a}} - 2 \cdot \Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} \Pi_+^{\mathsf{gh:1\,1+a}} \Phi_+^{\mathsf{h}1} - 2 \cdot \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} \Pi_-^{\mathsf{gh:1\,1+a}} \Phi_-^{\mathsf{h}1} \right) \alpha^{1+\mathsf{a}} \\ &\quad + \Omega^2 \left(\widetilde{C}^1 - \Phi_+^{1\mathsf{e}} S_+^{\mathsf{ef}} \Phi_+^{\mathsf{fg}} \Upsilon_+^{\mathsf{g}1} + \Phi_-^{1\mathsf{e}} S_-^{\mathsf{ef}} \Phi_-^{\mathsf{fg}} \Upsilon_-^{\mathsf{g}1} \right) \alpha^1 - \Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \widetilde{A}^{1+\mathsf{a}} \alpha^{1+\mathsf{b}} \\ &\quad - \Omega^2 \gamma^{1+\mathsf{a}:1} \widetilde{A}^{1+\mathsf{a}} \alpha^1 - \Omega^2 \gamma^{1+\mathsf{a}:1} \alpha^{1+\mathsf{a}} \widetilde{A}^1 \\ I_3^{\mathsf{eb}} &:= -\Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \alpha^{1+\mathsf{a}} \alpha^{1+\mathsf{b}} - \Omega^2 \gamma^{1+\mathsf{a}:1} \alpha^{1+\mathsf{a}} \alpha^1. \end{aligned}$$

Similarly, I_2^{eb} is the interaction term of the smoothing bias and stochastic variability. I_3^{eb} is the second-order smoothing bias. The nonlinearity bias is defined to be the leading term of the expectation of I_1^{eb}. By straightforward calculation and LIE,

$$\mathrm{E}\left[S_+^{\mathsf{ef}} S_+^{\mathsf{g}:1}\right] = \mathrm{E}\left[S_-^{\mathsf{ef}} S_-^{\mathsf{g}:1}\right] = \mathrm{E}\left[S_+^{\mathsf{ef}} \widetilde{A}^{1+\mathsf{a}}\right] = \mathrm{E}\left[S_-^{\mathsf{ef}} \widetilde{A}^{1+\mathsf{a}}\right] = \mathrm{E}\left[S_+^{\mathsf{ef}} \widetilde{A}^{1}\right] = \mathrm{E}\left[S_-^{\mathsf{ef}} \widetilde{A}^{1}\right] = 0.$$

Therefore,

$$\mathrm{E}\left[I_1^{\mathsf{eb}}\right] = \Omega \cdot \mathrm{E}\left[\widetilde{A}^{1\,1+\mathsf{a}} \widetilde{A}^{1+\mathsf{a}}\right] + \Omega^2 \cdot \mathrm{E}\left[\widetilde{C}^1 \widetilde{A}^1\right] - \Omega \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \cdot \mathrm{E}\left[\widetilde{A}^{1+\mathsf{a}} \widetilde{A}^{1+\mathsf{b}}\right] - \Omega^2 \gamma^{1+\mathsf{a}:1} \cdot \mathrm{E}\left[\widetilde{A}^{1+\mathsf{a}} \widetilde{A}^{1}\right]. \tag{S61}$$

Similarly, by straightforward calculation,

$$\begin{aligned}
\mathrm{E}\left[\widetilde{A}^{1\,1+\mathsf{a}} \widetilde{A}^{1+\mathsf{a}}\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\frac{1}{h^2} \widetilde{W}_p^3 V^{(1)} V^{(1+\mathsf{a})} \left(V^{(1+\mathsf{a})} - g_{V^{(1+\mathsf{a})}}(X)\right)\right]\right\} \\
\mathrm{E}\left[\widetilde{C}^1 \widetilde{A}^1\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\frac{1}{h^2} \widetilde{W}_p^2 H^{(1)} \left(V^{(1)} - g_{V^{(1)}}(X)\right)\right]\right\} \\
\mathrm{E}\left[\widetilde{A}^{1+\mathsf{a}} \widetilde{A}^{1+\mathsf{b}}\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\frac{1}{h^2} \widetilde{W}_p^2 \left(V^{(1+\mathsf{a})} - g_{V^{(1+\mathsf{a})}}(X)\right)\left(V^{(1+\mathsf{b})} - g_{V^{(1+\mathsf{b})}}(X)\right)\right]\right\} \\
\mathrm{E}\left[\widetilde{A}^{1+\mathsf{a}} \widetilde{A}^{1}\right] &= \frac{1}{n}\left\{\mathrm{E}\left[\frac{1}{h^2} \widetilde{W}_p^2 \left(V^{(1+\mathsf{a})} - g_{V^{(1+\mathsf{a})}}(X)\right)\left(V^{(1)} - g_{V^{(1)}}(X)\right)\right]\right\}.
\end{aligned} \tag{S62}$$

Then, by

$$\frac{\mathrm{S}^\top \widetilde{\Delta}_{UU^\top,2}^{-1/2}\left(-\widetilde{\Delta}_G\right)}{\sqrt{\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} \widetilde{\Delta}_G}} = \mathrm{e}_{d_z+2,1},$$

we have

$$\begin{aligned}
H^{(1)} &= \frac{\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} G}{\sqrt{\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} \widetilde{\Delta}_G}} \\
V^{(1)} &= -\frac{\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} U}{\sqrt{\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} \widetilde{\Delta}_G}}.
\end{aligned} \tag{S63}$$

By (S24), (S57) and (S58),

$$\begin{aligned}
V^{(1+\mathsf{a})} g_{V^{(1+\mathsf{a})}}(X) &= \bar{Z}^\top \widetilde{\Delta}_{\bar{Z}\bar{Z}^\top}^{-1} g_{\bar{Z}}(X) \\
&= \widetilde{\Delta}_2^{-1} + \left(Z^\top - \frac{\widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\right)\left(\widetilde{\Delta}_{ZZ^\top,2} - \frac{\widetilde{\Delta}_{Z,2} \widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\right)^{-1}\left(g_Z(X) - \frac{\widetilde{\Delta}_{Z,2}}{\widetilde{\Delta}_2}\right).
\end{aligned}$$

By (S56),

$$\begin{aligned}
\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} G &= \frac{\widetilde{\Delta}_D \cdot D}{\widetilde{\Delta}_{M^2,2} - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}} \\
\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} \widetilde{\Delta}_G &= \frac{\widetilde{\Delta}_D^2}{\widetilde{\Delta}_{M^2,2} - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}} \\
\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} U &= \widetilde{\Delta}_D \cdot \frac{M - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z}}{\widetilde{\Delta}_{M^2,2} - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}} \\
\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} g_U(X) &= \widetilde{\Delta}_D \cdot \frac{g_M(X) - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1} g_{\bar{Z}}(X)}{\widetilde{\Delta}_{M^2,2} - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\widetilde{\Delta}_{M\bar{Z},2}}.
\end{aligned} \tag{S64}$$

By these results, the fact that $\widetilde{\Delta}_D = \mu_{D,\dagger} + o(1)$, $\widetilde{\Delta}_{Z,2} = 2\left(\omega_p^{0,2}/\varphi\right)\mu_Z + o(1)$ and $\widetilde{\Delta}_2 = 2\left(\omega_p^{0,2}/\varphi\right) + o(1)$, (S62) and Lemma 1(a),

$$\begin{aligned}
\Omega \cdot \mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^3 V^{(1)} V^{(1+\mathsf{a})} g_{V^{(1+\mathsf{a})}}(X)\right] &= -\frac{\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^3 \left(\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} U\right)\left(\bar{Z}^\top \widetilde{\Delta}_{\bar{Z}\bar{Z}^\top}^{-1} g_{\bar{Z}}(X)\right)\right]}{\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G} \\
&= -\widetilde{\Delta}_2^{-1}\widetilde{\Delta}_D^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^3\left(M - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z}\right)\right] \\
-\widetilde{\Delta}_D^{-1}\cdot \mathrm{E}\Bigg[\frac{1}{h}\widetilde{W}_p^3\left(M - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z}\right)&\left(Z^\top - \frac{\widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\right)\left(\widetilde{\Delta}_{ZZ^\top,2} - \frac{\widetilde{\Delta}_{Z,2}\widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\right)^{-1}\left(g_Z(X) - \frac{\widetilde{\Delta}_{Z,2}}{\widetilde{\Delta}_2}\right)\Bigg] \\
&= o(1).
\end{aligned}$$

Then by (S64), $\widetilde{\Delta}_D = \mu_{D,\dagger} + o(1)$, $\widetilde{\Delta}_{Z,2} = 2\left(\omega_p^{0,2}/\varphi\right)\mu_Z + o(1)$, $\widetilde{\Delta}_2 = 2\left(\omega_p^{0,2}/\varphi\right) + o(1)$, $\widetilde{\Delta}_{M,2} = 2\left(\omega_p^{0,2}/\varphi\right)\mu_M + o(1)$, (S24) and Lemma 1(a), we have

$$\begin{aligned}
\Omega^2 \cdot \mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 H^{(1)} g_{V^{(1)}}(X)\right] &= -\frac{\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2\left(\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} G\right)\left(\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} g_U(X)\right)\right]}{\left(\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G\right)^2} \\
&= -\widetilde{\Delta}_D^{-2}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 D\left(g_M(X) - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1} g_{\bar{Z}}(X)\right)\right] \\
&= -\widetilde{\Delta}_D^{-2}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 D\left(g_M(X) - \frac{\widetilde{\Delta}_{M,2}}{\widetilde{\Delta}_2} - \widetilde{\gamma}_M^\top\left(g_Z(X) - \frac{\widetilde{\Delta}_{Z,2}}{\widetilde{\Delta}_2}\right)\right)\right] \\
&= o(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Omega \cdot \mathrm{E}\left[\widetilde{A}^{1\,1+\mathsf{a}}\widetilde{A}^{1+\mathsf{a}}\right] &= \frac{1}{nh}\cdot\left(\Omega\cdot\alpha^{1\,1+\mathsf{a}\,1+\mathsf{a}} + o(1)\right) \\
\Omega^2 \cdot \mathrm{E}\left[\widetilde{C}^1\widetilde{A}^1\right] &= \frac{1}{nh}\cdot\left(\Omega^2\gamma^{1:1} + o(1)\right).
\end{aligned} \tag{S65}$$

By (S57), (S58) and (S63),

$$\Omega\alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{V^{(1+\mathsf{a})}}(X)\, g_{V^{(1+\mathsf{b})}}(X)\right] = -\frac{\operatorname{tr}\left(\left(\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^3\left(\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}U\right)\bar{Z}\bar{Z}^\top\right]\right)\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{\bar{Z}}(X)\, g_{\bar{Z}^\top}(X)\right]\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\right)}{\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G}. \tag{S66}$$

By Lemma 1(a),

$$\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{\bar{Z}}(X)\, g_{\bar{Z}^\top}(X)\right] = 2\cdot\frac{\omega_p^{0,2}}{\varphi}\cdot\begin{bmatrix} 1 & \mu_{Z^\top} \\ \mu_Z & \mu_Z\mu_{Z^\top}\end{bmatrix} + o(1).$$

by (S24),

$$\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{\bar{Z}}(X)\, g_{\bar{Z}^\top}(X)\right]\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1} = \left(2\cdot\frac{\omega_p^{0,2}}{\varphi}\right)^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 0\end{bmatrix} + o(1).$$

By this result, (S66) and Lemma 1(a),

$$\begin{aligned}\Omega\alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{V^{(1+\mathsf{a})}}(X)\, g_{V^{(1+\mathsf{b})}}(X)\right] &= \frac{\left(2\cdot\frac{\omega_p^{0,2}}{\varphi}\right)^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^3\left(\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}U\right)\right]}{\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G} + o(1) \\ &= \left(2\cdot\frac{\omega_p^{0,2}}{\varphi}\right)^{-1}\widetilde{\Delta}_D^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^3\left(M-\widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z}\right)\right] + o(1) \\ &= o(1).\end{aligned}$$

And, by (S24), (S58), (S63), (S64) and Lemma 1(a),

$$\begin{aligned}&\Omega^2\gamma^{1+\mathsf{a}:1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{V^{(1+\mathsf{a})}}(X)\, g_{V^{(1)}}(X)\right] \\ &\quad = \frac{\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2\left(\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}G\right)\bar{Z}^\top\right]\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{\bar{Z}}(X)\left(\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}g_U(X)\right)\right]}{\left(\widetilde{\Delta}_G^\top\widetilde{\Delta}_{UU^\top,2}^{-1}\widetilde{\Delta}_G\right)^2} \\ &= \widetilde{\Delta}_D^{-2}\cdot\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 D\bar{Z}^\top\right]\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 g_{\bar{Z}}(X)\left(g_M(X)-\frac{\widetilde{\Delta}_{M,2}}{\widetilde{\Delta}_2}-\widetilde{\gamma}_M^\top\left(g_Z(X)-\frac{\widetilde{\Delta}_{Z,2}}{\widetilde{\Delta}_2}\right)\right)\right] = o(1).\end{aligned}$$

Then,

$$\begin{aligned}\Omega\alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}}\cdot\mathrm{E}\left[\widetilde{A}^{1+\mathsf{a}}\widetilde{A}^{1+\mathsf{b}}\right] &= \frac{1}{nh}\cdot\left(\Omega\cdot\alpha^{1\,1+\mathsf{a}\,1+\mathsf{a}} + o(1)\right) \\ \Omega^2\gamma^{1+\mathsf{a}:1}\cdot\mathrm{E}\left[\widetilde{A}^{1+\mathsf{a}}\widetilde{A}^{1}\right] &= o\left((nh)^{-1}\right).\end{aligned} \tag{S67}$$

And, by $\widetilde{\Delta}_D = \mu_{D,\dagger} + o(1)$, (S24), (S63), (S64) and Lemma 1(a),

$$\begin{aligned}
\Omega^2 \gamma^{1:1} &= \frac{\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2\left(\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} G\right)\left(\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} U\right)\right]}{\left(\widetilde{\Delta}_G^\top \widetilde{\Delta}_{UU^\top,2}^{-1} \widetilde{\Delta}_G\right)^2} \\
&= \widetilde{\Delta}_D^{-2} \mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 D\left(M - \widetilde{\Delta}_{M\bar{Z}^\top,2}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\bar{Z}\right)\right] \\
&= \widetilde{\Delta}_D^{-2} \mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 D\left(M - \frac{\widetilde{\Delta}_{M,2}}{\widetilde{\Delta}_2} - \widetilde{\gamma}_M^\top\left(Z - \frac{\widetilde{\Delta}_{Z,2}}{\widetilde{\Delta}_2}\right)\right)\right] \\
&= \omega_p^{0,2} \cdot \frac{\mathrm{Cov}_{|0^\pm}[\epsilon, D]}{\varphi \mu_{D,\dagger}^2} + o(1).
\end{aligned}$$

The conclusion follows from this result, (S61), (S65) and (S67).

S3 Proof of Theorem 3

The following lemma provides simple bounds for projection coefficients.

Lemma 5. *Let V be a random variable and ω be a nonnegative random variable. Assume that $\mathrm{E}\left[\omega P P^\top\right]$ is invertible. Then,*

$$\mathrm{mineig}\left(\left(\mathrm{E}\left[\omega P P^\top\right]\right)^{-1}\right)\|\mathrm{E}[\omega P V]\|^2 \leq \mathrm{E}\left[\omega V^2\right].$$

Proof of Lemma 5. Denote $\beta := \left(\mathrm{E}\left[\omega P P^\top\right]\right)^{-1}\mathrm{E}[\omega P V]$. Then, we have

$$\begin{aligned}
0 &\leq \mathrm{E}\left[\omega\left(V - P^\top\beta\right)^2\right] \\
&= \mathrm{E}\left[\omega V^2\right] + \beta^\top \mathrm{E}\left[\omega P P^\top\right]\beta - 2\cdot\beta^\top \mathrm{E}[\omega P V] \\
&= \mathrm{E}\left[\omega V^2\right] - \mathrm{E}\left[\omega P^\top V\right]\left(\mathrm{E}\left[\omega P P^\top\right]\right)^{-1}\mathrm{E}[\omega P V] \\
&\leq \mathrm{E}\left[\omega V^2\right] - \mathrm{mineig}\left(\left(\mathrm{E}\left[\omega P P^\top\right]\right)^{-1}\right)\|\mathrm{E}[\omega P V]\|^2.
\end{aligned}$$

■

Lemma 6. *Suppose that the assumptions in the statement of Theorem 3 are satisfied. Then, $\left\|\eta^* - \rho^\top\gamma_P\right\|_\infty = O\left((\alpha_k + h)\beta_k\right)$, where $\gamma_P := \widetilde{\Delta}_{PP^\top,2}^{-1}\widetilde{\Delta}_{MP,2}$.*

Proof of Lemma 6. Let $\gamma^* := \mathrm{argmin}_\gamma \left\|\eta^* - \rho^\top\gamma\right\|_\infty$. Then, by triangle inequality,

$$\left\|\eta^* - \rho^\top\gamma_P\right\|_\infty \leq \left\|\eta^* - \rho^\top\gamma^*\right\|_\infty + \left\|\rho^\top\gamma^* - \rho^\top\gamma_P\right\|_\infty, \tag{S68}$$

where $\alpha_k \coloneqq \left\|\eta^* - \rho^\top \gamma^*\right\|_\infty$. Write

$$\begin{aligned}\left|\rho^\top(z)\gamma^* - \rho^\top(z)\gamma_P\right| &= \rho^\top(z)\widetilde{\Delta}^{-1}_{PP^\top,2}\widetilde{\Delta}_{PP^\top,2}\gamma^* - \rho^\top(z)\widetilde{\Delta}^{-1}_{PP^\top,2}\widetilde{\Delta}_{PM,2} \\ &\le \|\rho(z)\|\left\|\widetilde{\Delta}^{-1}_{PP^\top,2}\right\|\left\|\widetilde{\Delta}_{PM,2} - \widetilde{\Delta}_{PP^\top,2}\gamma^*\right\|.\end{aligned}$$

Then by this result and writing $\widetilde{\Delta}_{PM,2} - \widetilde{\Delta}_{PP^\top,2}\gamma^*$ as $\mathrm{E}\left[h^{-1}\widetilde{W}_p^2 P\left(M - P^\top\gamma^*\right)\right]$, we have

$$\begin{aligned}\left\|\rho^\top\gamma^* - \rho^\top\gamma_P\right\|_\infty &\le \beta_k \cdot \operatorname{maxeig}\left(\widetilde{\Delta}^{-1}_{PP^\top,2}\right)\left\|\widetilde{\Delta}_{PM,2} - \widetilde{\Delta}_{PP^\top,2}\gamma^*\right\| \\ &\le \beta_k \cdot \operatorname{maxeig}\left(\widetilde{\Delta}^{-1}_{PP^\top,2}\right)\left\{\left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 P\left(M - \eta^*(Z)\right)\right]\right\| + \left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 P\left(\eta^*(Z) - P^\top\gamma^*\right)\right]\right\|\right\}\end{aligned} \tag{S69}$$

It follows from $\operatorname{mineig}\left(\widetilde{\Delta}_{PP^\top,2}\right) \ge \widetilde{\Delta}_2\underline{\sigma}$ if h is sufficiently small and Lemma 1(a) that $\operatorname{maxeig}\left(\widetilde{\Delta}^{-1}_{PP^\top,2}\right) = \left(\operatorname{mineig}\left(\widetilde{\Delta}_{PP^\top,2}\right)\right)^{-1} = O(1)$. Similarly, it follows from the fact that

$$\left(\operatorname{mineig}\left(\widetilde{\Delta}^{-1}_{PP^\top,2}\right)\right)^{-1} = \operatorname{maxeig}\left(\widetilde{\Delta}_{PP^\top,2}\right) \le \widetilde{\Delta}_2\overline{\sigma}$$

if h is sufficiently small and Lemma 1(a) that $\left(\operatorname{mineig}\left(\widetilde{\Delta}^{-1}_{PP^\top,2}\right)\right)^{-1} = O(1)$. Then, by this result, Lemma 5 and Lemma 1(a),

$$\begin{aligned}\left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 P\left(\eta^*(Z) - P^\top\gamma^*\right)\right]\right\|^2 &\le \left(\operatorname{mineig}\left(\widetilde{\Delta}^{-1}_{PP^\top,2}\right)\right)^{-1}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2\left(\eta^*(Z) - P^\top\gamma^*\right)^2\right] \\ &\le \left(\operatorname{mineig}\left(\widetilde{\Delta}^{-1}_{PP^\top,2}\right)\right)^{-1}\widetilde{\Delta}_2\alpha_k^2 \\ &= O\left(\alpha_k^2\right).\end{aligned} \tag{S70}$$

By LIE and triangle inequality,

$$\begin{aligned}\left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2 P\left(M - \eta^*(Z)\right)\right]\right\| &\le \left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 P\left(g_{M|ZX}(Z,X) - \mu_+^*(Z)\right)\right]\right\| \\ &\quad + \left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;-}^2 P\left(g_{M|ZX}(Z,X) - \mu_-^*(Z)\right)\right]\right\| \\ &\quad + \left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 P\left(\mu_+^*(Z) - \eta^*(Z)\right)\right] + \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;-}^2 P\left(\mu_-^*(Z) - \eta^*(Z)\right)\right]\right\|\end{aligned} \tag{S71}$$

By Lemma 5 with $\omega = \widetilde{W}_{p;+}^2/h$, Lipschitz continuity (Assumption 4(e)) and Lemma 1(a),

$$\left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 P\left(g_{M|ZX}(Z,X) - \mu_+^*(Z)\right)\right]\right\|^2$$

$$\leq \left(\text{mineig}\left(\left(\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 PP^\top\right]\right)^{-1}\right)\right)^{-1} \left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 \left(g_{M|ZX}(Z,X) - \mu_+^*(Z)\right)^2\right]\right\|$$

$$\leq \text{maxeig}\left(\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 PP^\top\right]\right)\widetilde{\Delta}_2 L_g^2 h^2 \leq \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2\right]\overline{\sigma}\widetilde{\Delta}_2 L_g^2 h^2 = O\left(h^2\right). \quad \text{(S72)}$$

A similar bound can be shown for the second term on the right hand side of (S71). Denote $\eta^\dagger(Z) \coloneqq \left(\mu_+^*(Z) - \mu_-^*(Z)\right)/2$. Then,

$$\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 P\left(\mu_+^*(Z) - \eta^*(Z)\right)\right] + \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;-}^2 P\left(\mu_-^*(Z) - \eta^*(Z)\right)\right] = \mathrm{E}\left[\frac{1}{h}\left(\widetilde{W}_{p;+}^2 - \widetilde{W}_{p;-}^2\right)P\eta^\dagger(Z)\right].$$

Denote $\widetilde{\mathcal{K}}_{p;\mathrm{s}} \coloneqq \mathrm{e}_{p+1,1}^\top \Pi_{p;\mathrm{s}}^{-1} r_p(t) K(t)$ for $\mathrm{s} \in \{-,+\}$. Then, $\widetilde{W}_{p;+} = \widetilde{\mathcal{K}}_{p;+}(X/h)\mathbb{1}(X>0)$ and $\widetilde{W}_{p;-} = \widetilde{\mathcal{K}}_{p;-}(X/h)\mathbb{1}(X<0)$. Now we can write

$$\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 P\eta^\dagger(Z)\right] = \int_0^\infty \widetilde{\mathcal{K}}_{p;+}^2\left(\frac{y}{h}\right)\mathrm{E}\left[\rho(Z)\eta^\dagger(Z) \mid X=y\right] f_X(y)\,\mathrm{d}y.$$

And, for x, y in the right neighborhood of 0, by Lemma 5 with $\omega = 1$,

$$\left\|\mathrm{E}\left[\rho(Z)\eta^\dagger(Z) \mid X=y\right] f_X(y) - \mathrm{E}\left[\rho(Z)\eta^\dagger(Z) \mid X=x\right] f_X(x)\right\|^2$$
$$= \left\|\mathrm{E}\left[\rho(Z)\eta^\dagger(Z)\left(f_{X|Z}(y \mid Z) - f_{X|Z}(x \mid Z)\right)\right]\right\|^2$$
$$\leq \text{maxeig}\left(\mathrm{E}\left[PP^\top\right]\right)\mathrm{E}\left[\left(\eta^\dagger\right)^2(Z)\left(f_{X|Z}(y \mid Z) - f_{X|Z}(x \mid Z)\right)^2\right].$$

By this result, change of variables and Lipschitz continuity (Assumption 4(e)),

$$\left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}^2 P\eta^\dagger(Z)\right] - \left(\int_0^1 \widetilde{\mathcal{K}}_{p;+}^2(u)\,\mathrm{d}u\right)\psi_{P\eta^\dagger(Z),+}\right\|$$
$$\leq \int_0^1 \widetilde{\mathcal{K}}_{p;+}^2(u)\left\|\mathrm{E}\left[\rho(Z)\eta^\dagger(Z) \mid X=hu\right] f_X(hu) - \psi_{P\eta^\dagger(Z),+}\right\|\mathrm{d}u$$
$$\leq \left(\int_0^1 \widetilde{\mathcal{K}}_{p;+}^2(u)\,u\,\mathrm{d}u\right)\sqrt{\text{maxeig}\left(\mathrm{E}\left[PP^\top\right]\right)\mathrm{E}\left[\left(\eta^\dagger\right)^2(Z)\right]L_f^2}\,h^2.$$

Similarly,

$$\left\|\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;-}^2 P\eta^\dagger(Z)\right] - \left(\int_0^1 \widetilde{\mathcal{K}}_{p;-}^2(u)\,\mathrm{d}u\right)\psi_{P\eta^\dagger(Z),-}\right\| = O(h).$$

By these results,

$$\left\|\mathrm{E}\left[\frac{1}{h}\left(\widetilde{W}_{p;+}^2 - \widetilde{W}_{p;-}^2\right)P\eta^\dagger(Z)\right]\right\| = \left\|\left(\int_0^1 \widetilde{\mathcal{K}}_{p;+}^2(u)\,\mathrm{d}u\right)\psi_{P\eta^\dagger(Z),+} - \left(\int_0^1 \widetilde{\mathcal{K}}_{p;-}^2(u)\,\mathrm{d}u\right)\psi_{P\eta^\dagger(Z),-}\right\| + O(h).$$

By Lemma 5 with $\omega = 1$ and the expectation being replaced by the conditional expectation given $X = x$,

$$\left\| \mathrm{E}\left[P\eta^{\dagger}(Z) \mid X = x \right] \right\|^2 \leq \operatorname{maxeig}\left(\mathrm{E}\left[PP^{\top} \mid X = x \right] \right) \mathrm{E}\left[\left(\eta^{\dagger}(Z) \right)^2 \mid X = x \right].$$

By this result and Assumption 5(b), $\left\| \psi_{P\eta^{\dagger}(Z),\mathrm{s}} \right\| = O(1)$ for s $\in \{-,+\}$. By Assumption 5(a), $\psi_{P\eta^{\dagger}(Z),+} = \psi_{P\eta^{\dagger}(Z),-}$. And by (S5), $\int_0^1 \widetilde{\mathcal{K}}_{p;+}^2(u)\,\mathrm{d}u = \varphi^{-2}\omega_p^{0,2} + O(h)$, $\int_0^1 \widetilde{\mathcal{K}}_{p;-}^2(u)\,\mathrm{d}u = \varphi^{-2}\omega_p^{0,2} + O(h)$, it follows that

$$\begin{aligned} &\left\| \mathrm{E}\left[\frac{1}{h} \widetilde{W}_{p;+}^2 P\left(\mu_+^*(Z) - \eta^*(Z) \right) \right] + \mathrm{E}\left[\frac{1}{h} \widetilde{W}_{p;-}^2 P\left(\mu_-^*(Z) - \eta^*(Z) \right) \right] \right\| \\ &\qquad = \left\| \mathrm{E}\left[\frac{1}{h} \left(\widetilde{W}_{p;+}^2 - \widetilde{W}_{p;-}^2 \right) P\eta^{\dagger}(Z) \right] \right\| = O(h). \end{aligned}$$

It follows from this result, (S71) and (S72) that $\left\| \mathrm{E}\left[h^{-1} \widetilde{W}_p^2 P\left(M - \eta^*(Z) \right) \right] \right\| = O(h)$. By this result, (S69), (S70) and $\operatorname{maxeig}\left(\widetilde{\Delta}_{PP^{\top},2}^{-1} \right) = O(1)$, we have $\left\| \rho^{\top}\gamma^* - \rho^{\top}\gamma_P \right\|_{\infty} = O\left(\beta_k \left(\alpha_k + h \right) \right)$. The conclusion follows from this result and (S68). ■

Lemma 7. *Under the assumptions in the statement of Theorem 3, (a)* $\left\| \widehat{\Psi}_P \right\| = O_p\left(\sqrt{k/(nh)} \right)$*; (b)* $\left\| \widehat{\Psi}_{PP^{\top},2} - \widetilde{\Delta}_{PP^{\top},2} \right\| = O_p\left(\beta_k \sqrt{\log(k)/(nh)} \right)$*; (c)* $\max_i \left\| \widehat{W}_{p,i} P_i \right\| = O_p(\beta_k)$.

Proof of Lemma 7. For (a), write

$$\widehat{\Psi}_P = \frac{1}{nh} \sum_i \widehat{W}_{p,i} P_i = \frac{1}{nh} \sum_i \widehat{W}_{p,i} g_P(X_i) + \frac{1}{nh} \sum_i \widetilde{W}_{p,i} \left(P_i - g_P(X_i) \right) + \frac{1}{nh} \sum_i \left(\widehat{W}_{p,i} - \widetilde{W}_{p,i} \right) \left(P_i - g_P(X_i) \right). \tag{S73}$$

Then, by Taylor expansion and $\mu_{P,+} = \mu_{P,-}$,

$$\frac{1}{nh} \sum_i \widehat{W}_{p,i} g_P(X_i) = \left(\frac{1}{nh} \sum_i \widehat{W}_{p,i} \frac{g_P^{(p+1)}\left(\tilde{X}_i \right)}{(p+1)!} \left(\frac{X_i}{h} \right)^{p+1} \right) h^{p+1}, \tag{S74}$$

where $\tilde{X}_i$ denotes the mean value that lies between 0 and X_i. Then by Assumption 5(d) and (S11), $\left\| (nh)^{-1} \sum_i \widehat{W}_{p,i} g_P(X_i) \right\| = O_p\left(\sqrt{k} \cdot h^{p+1} \right)$. Then, for the second term on the right hand side of (S73),

$$\mathrm{E}\left[\left\| \frac{1}{nh} \sum_i \widetilde{W}_{p,i} \left(P_i - g_P(X_i) \right) \right\|^2 \right] = \frac{1}{nh^2} \cdot \mathrm{E}\left[\widetilde{W}_p^2 \left\| P - g_P(X) \right\|^2 \right] \leq \frac{1}{nh^2} \cdot \mathrm{E}\left[\widetilde{W}_p^2 P^{\top} P \right] = \frac{1}{nh} \cdot \operatorname{tr}\left(\widetilde{\Delta}_{PP^{\top},2} \right). \tag{S75}$$

Since $\operatorname{tr}\left(\widetilde{\Delta}_{PP^{\top},2} \right) \leq \mathrm{E}\left[h^{-1} \widetilde{W}_p^2 \operatorname{tr}\left(\mathrm{E}\left[PP^{\top} \mid X \right] \right) \right] \leq \widetilde{\Delta}_2 \overline{\sigma} k$, when h is sufficiently small, by Lemma 1(a), we have $\operatorname{tr}\left(\widetilde{\Delta}_{PP^\top,2}\right)=O(k)$. By this result, Markov's inequality and (S75), we have

$$\frac{1}{nh}\sum_i \widetilde{W}_{p,i}\left(P_i-g_P\left(X_i\right)\right)=O_p\left(\sqrt{\frac{k}{nh}}\right). \tag{S76}$$

Write

$$\begin{aligned}&\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\left(P_i^\top-g_{P^\top}\left(X_i\right)\right)\\&\quad=\mathrm{e}_{p+1,1}^\top\left(\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right)\left\{\frac{1}{nh}\sum_i r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\left(P_i^\top-g_{P^\top}\left(X_i\right)\right)\right\}\\&\qquad-\mathrm{e}_{p+1,1}^\top\left(\widehat{\Pi}_{p,-}^{-1}-\Pi_{p,-}^{-1}\right)\left\{\frac{1}{nh}\sum_i r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i<0\right)\left(P_i^\top-g_{P^\top}\left(X_i\right)\right)\right\}.\end{aligned} \tag{S77}$$

Then, since the operator norm is dominated by the Frobenius norm,

$$\begin{aligned}&\left\|\frac{1}{nh}\sum_i r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\left(P_i^\top-g_{P^\top}\left(X_i\right)\right)\right\|\\&\qquad\leq\sum_{j=0}^p\left\|\frac{1}{nh}\sum_i r_p^{(j)}\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\left(P_i^\top-g_{P^\top}\left(X_i\right)\right)\right\|.\end{aligned}$$

For all $j=0,...,p$,

$$\left\|\frac{1}{nh}\sum_i r_p^{(j)}\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\left(P_i^\top-g_{P^\top}\left(X_i\right)\right)\right\|=O_p\left(\sqrt{\frac{k}{nh}}\right)$$

follows from arguments similar to those used to show $(nh)^{-1}\sum_i\widetilde{W}_{p,i}\left(P_i-g_P\left(X_i\right)\right)=O_p\left(\sqrt{k/(nh)}\right)$. Therefore, it follows from this result and (S5) that

$$\mathrm{e}_{p+1,1}^\top\left(\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right)\left\{\frac{1}{nh}\sum_i r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\left(P_i^\top-g_{P^\top}\left(X_i\right)\right)\right\}=O_p\left(\frac{\sqrt{k}}{nh}\right).$$

Similarly, the second term of (S77) is also $O_p\left(\sqrt{k}/(nh)\right)$. The conclusions follows from these results, (S73), (S74), (S76) and (S77).

For (b), write

$$\widehat{\Psi}_{PP^\top,2}-\widetilde{\Delta}_{PP^\top,2}=\widehat{\Psi}_{PP^\top,2}-\widetilde{\Psi}_{PP^\top,2}+\widetilde{\Psi}_{PP^\top,2}-\widetilde{\Delta}_{PP^\top,2}. \tag{S78}$$

Write

$$\widehat{\Psi}_{PP^\top,2}-\widetilde{\Psi}_{PP^\top,2}=\frac{2}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\widetilde{W}_{p,i}P_iP_i^\top+\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)^2P_iP_i^\top. \tag{S79}$$

First, by triangle inequality, Loève's c_r inequality and (S5),

$$\begin{aligned}\left\|\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)^2 P_iP_i^\top\right\| &\leq \frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)^2\|P_i\|^2\\ &\lesssim \left\|\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right\|^2\left(\frac{1}{nh}\sum_i\mathbb{1}\left(X_i>0\right)\left\|r_p\left(\frac{X_i}{h}\right)\right\|^2K^2\left(\frac{X_i}{h}\right)\|P_i\|^2\right)\\ &\quad+\left\|\widehat{\Pi}_{p,-}^{-1}-\Pi_{p,-}^{-1}\right\|^2\left(\frac{1}{nh}\sum_i\mathbb{1}\left(X_i<0\right)\left\|r_p\left(\frac{X_i}{h}\right)\right\|^2K^2\left(\frac{X_i}{h}\right)\|P_i\|^2\right). \end{aligned}\tag{S80}$$

By LIE, $\operatorname{tr}\left(\mathrm{E}\left[PP^\top\mid X\right]\right)\leq k\overline{\sigma}$ if $X\in(-h,0)\cup(0,h)$ and h is sufficiently small, and change of variables,

$$\mathrm{E}\left[\frac{1}{h}\left\|r_p\left(\frac{X}{h}\right)\right\|^2K^2\left(\frac{X}{h}\right)\|P\|^2\mathbb{1}\left(X>0\right)\right]=\mathrm{E}\left[\frac{1}{h}\left\|r_p\left(\frac{X}{h}\right)\right\|^2K^2\left(\frac{X}{h}\right)\mathbb{1}\left(X>0\right)\operatorname{tr}\left(\mathrm{E}\left[PP^\top\mid X\right]\right)\right]=O\left(k\right).$$

By this result, Markov's inequality and (S5), the first term on the right hand side of the second inequality is $O_p\left(k/\left(nh\right)\right)$. A similar result holds for the second term. Therefore,

$$\left\|\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)^2P_iP_i^\top\right\|=O_p\left(\frac{k}{nh}\right).\tag{S81}$$

By triangle inequality, (S5),

$$\begin{aligned}\left\|\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\widetilde{W}_{p,i}\mathrm{E}\left[P_iP_i^\top\mid X_i\right]\right\| &\leq\frac{1}{nh}\sum_i\left|\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\widetilde{W}_{p,i}\right|\left\|\mathrm{E}\left[P_iP_i^\top\mid X_i\right]\right\|\\ &\leq\left\|\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right\|\left\|\Pi_{p,+}^{-1}\right\|\left(\frac{1}{nh}\sum_i\left\|r_p\left(\frac{X_i}{h}\right)\right\|^2K^2\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\left\|\mathrm{E}\left[P_iP_i^\top\mid X_i\right]\right\|\right)\\ &\quad+\left\|\widehat{\Pi}_{p,-}^{-1}-\Pi_{p,-}^{-1}\right\|\left\|\Pi_{p,-}^{-1}\right\|\left(\frac{1}{nh}\sum_i\left\|r_p\left(\frac{X_i}{h}\right)\right\|^2K^2\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i<0\right)\left\|\mathrm{E}\left[P_iP_i^\top\mid X_i\right]\right\|\right)\\ &=O_p\left(\left(nh\right)^{-1/2}\right),\end{aligned}\tag{S82}$$

since $\left\|\mathrm{E}\left[P_iP_i^\top\mid X_i\right]\right\|\leq\overline{\sigma}$ if $X_i\in(-h,0)\cup(0,h)$ and h is sufficiently small. Since the operator norm is dominated by the Frobenius norm, by triangle inequality and Loève's c_r inequality,

$$\begin{aligned}&\left\|\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\widetilde{W}_{p,i}\left(P_iP_i^\top-\mathrm{E}\left[P_iP_i^\top\mid X_i\right]\right)\right\|^2\\ &\leq\sum_{j=1}^k\sum_{j'=1}^k\left(\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\widetilde{W}_{p,i}\left(P_i^{(j)}P_i^{(j')}-\mathrm{E}\left[P_i^{(j)}P_i^{(j')}\mid X_i\right]\right)\right)^2\\ &\lesssim\left\|\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right\|^2\sum_{j=1}^k\sum_{j'=1}^k\left\|\frac{1}{nh}\sum_i r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\widetilde{W}_{p;+,i}\left(P_i^{(j)}P_i^{(j')}-\mathrm{E}\left[P_i^{(j)}P_i^{(j')}\mid X_i\right]\right)\right\|^2\end{aligned}$$

$$+\left\|\widehat{\Pi}_{p,-}^{-1}-\Pi_{p,-}^{-1}\right\|^{2}\sum_{j=1}^{k}\sum_{j'=1}^{k}\left\|\frac{1}{nh}\sum_{i}r_{p}\left(\frac{X_{i}}{h}\right)K\left(\frac{X_{i}}{h}\right)\mathbb{1}\left(X_{i}<0\right)\widetilde{W}_{p;-,i}\left(P_{i}^{(j)}P_{i}^{(j')}-\mathrm{E}\left[P_{i}^{(j)}P_{i}^{(j')}\mid X_{i}\right]\right)\right\|^{2}. \tag{S83}$$

Then, by simple calculation,

$$\begin{aligned}\mathrm{E}&\left[\sum_{j=1}^{k}\sum_{j'=1}^{k}\left\|\frac{1}{nh}\sum_{i}r_{p}\left(\frac{X_{i}}{h}\right)K\left(\frac{X_{i}}{h}\right)\mathbb{1}\left(X_{i}>0\right)\widetilde{W}_{p;+,i}\left(P_{i}^{(j)}P_{i}^{(j')}-\mathrm{E}\left[P_{i}^{(j)}P_{i}^{(j')}\mid X_{i}\right]\right)\right\|^{2}\right]\\&\leq\sum_{j=1}^{k}\sum_{j'=1}^{k}\frac{1}{nh^{2}}\mathrm{E}\left[\left\|r_{p}\left(\frac{X}{h}\right)\right\|^{2}K^{2}\left(\frac{X}{h}\right)\mathbb{1}\left(X>0\right)\widetilde{W}_{p;+}^{2}\left(P^{(j)}P^{(j')}\right)^{2}\right]\\&=\frac{1}{nh^{2}}\mathrm{E}\left[\left\|r_{p}\left(\frac{X}{h}\right)\right\|^{2}K^{2}\left(\frac{X}{h}\right)\mathbb{1}\left(X>0\right)\widetilde{W}_{p;+}^{2}\left\|P\right\|^{4}\right].\end{aligned}$$

And, by LIE, $\mathrm{tr}\left(\mathrm{E}\left[PP^{\top}\mid X\right]\right)\leq k\overline{\sigma}$ if $X\in(-h,0)\cup(0,h)$ and h is sufficiently small, and change of variables,

$$\begin{aligned}\mathrm{E}&\left[\frac{1}{h}\left\|r_{p}\left(\frac{X}{h}\right)\right\|^{2}K^{2}\left(\frac{X}{h}\right)\mathbb{1}\left(X>0\right)\widetilde{W}_{p;+}^{2}\left\|P\right\|^{4}\right]\\&\leq\beta_{k}^{2}\cdot\mathrm{E}\left[\frac{1}{h}\left\|r_{p}\left(\frac{X}{h}\right)\right\|^{2}K^{2}\left(\frac{X}{h}\right)\mathbb{1}\left(X>0\right)\widetilde{W}_{p;+}^{2}\mathrm{tr}\left(\mathrm{E}\left[PP^{\top}\mid X\right]\right)\right]=O\left(\beta_{k}^{2}k\right).\end{aligned}$$

Therefore,

$$\mathrm{E}\left[\sum_{j=1}^{k}\sum_{j'=1}^{k}\left\|\frac{1}{nh}\sum_{i}r_{p}\left(\frac{X_{i}}{h}\right)K\left(\frac{X_{i}}{h}\right)\mathbb{1}\left(X_{i}>0\right)\widetilde{W}_{p;+,i}\left(P_{i}^{(j)}P_{i}^{(j')}-\mathrm{E}\left[P_{i}^{(j)}P_{i}^{(j')}\mid X_{i}\right]\right)\right\|^{2}\right]=O\left(\frac{\beta_{k}^{2}k}{nh}\right).$$

By Markov's inequality and (S5), the first term on the right hand side of (S83) is $O_p\left(\left(\beta_k^2 k\right)/\left(nh\right)^2\right)$. A similar result holds for the second term. Therefore,

$$\left\|\frac{1}{nh}\sum_{i}\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\widetilde{W}_{p,i}\left(P_{i}P_{i}^{\top}-\mathrm{E}\left[P_{i}P_{i}^{\top}\mid X_{i}\right]\right)\right\|=O_{p}\left(\frac{\beta_{k}\sqrt{k}}{nh}\right).$$

It follows from this result and (S82) that

$$\left\|\frac{1}{nh}\sum_{i}\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\widetilde{W}_{p,i}P_{i}P_{i}^{\top}\right\|=O_{p}\left(\left(nh\right)^{-1/2}+\frac{\beta_{k}\sqrt{k}}{nh}\right).$$

By (S79), (S81) and this result,

$$\left\|\widehat{\Psi}_{PP^{\top},2}-\widetilde{\Psi}_{PP^{\top},2}\right\|=O_{p}\left(\left(nh\right)^{-1/2}+\frac{\beta_{k}\sqrt{k}}{nh}+\frac{k}{nh}\right). \tag{S84}$$

By Belloni et al. (2015, Lemma 6.2) with $p_i = \widetilde{W}_{p,i} P_i / \sqrt{h}$ and $\|p_i\| \lesssim \beta_k / \sqrt{h}$,

$$\mathrm{E}\left[\left\|\widetilde{\Psi}_{PP^\top,2} - \widetilde{\Delta}_{PP^\top,2}\right\|\right] \lesssim \frac{\beta_k^2 \log(k)}{nh} + \beta_k \sqrt{\frac{\log(k)\left\|\widetilde{\Delta}_{PP^\top,2}\right\|}{nh}}.$$

It follows from this result and $\left\|\widetilde{\Delta}_{PP^\top,2}\right\| \leq \widetilde{\Delta}_2 \overline{\sigma}$ that

$$\left\|\widetilde{\Psi}_{PP^\top,2} - \widetilde{\Delta}_{PP^\top,2}\right\| = O_p\left(\beta_k \sqrt{\frac{\log(k)}{nh}}\right)$$

Part (b) follows from this result, (S78) and (S84).

Part (c) follows from $\max_i \left\|\widehat{W}_{p,i} P_i\right\| \lesssim \left(\sum_{s \in \{-,+\}} \left\|\widehat{\Pi}_{p,s}^{-1}\right\|\right) \left(\sup_{z \in \mathcal{Z}} \|\rho(z)\|\right)$, (S1) and (S5). ■

By Lagrangian multiplier method, we get

$$w_i^{\mathsf{sieve}} = \frac{1}{n} \cdot \frac{1}{1 + \left(\lambda_p^{\mathsf{eb}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)},$$

where

$$\lambda_p^{\mathsf{sieve}} := \underset{\lambda}{\operatorname{argmax}} \sum_i \log\left(1 + \lambda^\top \left(\widehat{W}_{p,i} P_i\right)\right).$$

We have following result on the rate of convergence and linearization of $\lambda_p^{\mathsf{sieve}}$.

Lemma 8. *Under the assumptions in the statement of Theorem 3, we have* $\left\|\lambda_p^{\mathsf{sieve}}\right\| = O_p\left(\sqrt{k/(nh)}\right)$ *and* $\left\|\lambda_p^{\mathsf{sieve}} - \widetilde{\Delta}_{PP^\top,2}^{-1} \widehat{\Psi}_P\right\| = O_p\left((\beta_k k)/(nh)\right)$.

Proof of Lemma 8. By similar arguments as in the proof of Lemma 2,

$$\frac{1}{nh} \sum_i \widehat{W}_{p,i} P_i = \left(\frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i}^2 P_i P_i^\top}{1 + \left(\lambda_p^{\mathsf{sieve}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)}\right) \lambda_p^{\mathsf{sieve}}$$

and

$$\left(\lambda_p^{\mathsf{sieve}}\right)^\top \widetilde{\Delta}_{PP^\top,2} \lambda_p^{\mathsf{sieve}} \leq \left\|\lambda_p^{\mathsf{sieve}}\right\| \left\|\widehat{\Psi}_P\right\| \left\{1 + \left\|\lambda_p^{\mathsf{sieve}}\right\| \max_i \left\|\widehat{W}_{p,i} P_i\right\|\right\} + \left\|\lambda_p^{\mathsf{sieve}}\right\|^2 \left\|\widehat{\Psi}_{PP^\top,2} - \widetilde{\Delta}_{PP^\top,2}\right\|.$$

When h is sufficiently small, $\left(\lambda_p^{\mathsf{sieve}}\right)^\top \widetilde{\Delta}_{PP^\top,2} \lambda_p^{\mathsf{sieve}} \geq \underline{\sigma} \widetilde{\Delta}_2 \left\|\lambda_p^{\mathsf{sieve}}\right\|^2$. Therefore,

$$\left(\underline{\sigma} \widetilde{\Delta}_2 - \left\|\widehat{\Psi}_P\right\| \max_i \left\|\widehat{W}_{p,i} P_i\right\| - \left\|\widehat{\Psi}_{PP^\top,2} - \widetilde{\Delta}_{PP^\top,2}\right\|\right) \left\|\lambda_p^{\mathsf{sieve}}\right\| \leq \left\|\widehat{\Psi}_P\right\|.$$

The first conclusion follows from Lemma 7 and $\widetilde{\Delta}_2 = 2\omega_p^{0,2}/\varphi + o(1)$, which follows from Lemma 1(a).

By using the first order conditions, we have

$$\widehat{\Psi}_P = \left\{ \frac{1}{nh} \sum_i \widehat{W}_{p,i}^2 P_i P_i^\top \cdot \left(1 - \frac{\left(\lambda_p^{\mathsf{sieve}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)}{1 + \left(\lambda_p^{\mathsf{sieve}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)} \right) \right\} \lambda_p^{\mathsf{sieve}}$$

and

$$\widehat{\Psi}_P = \widehat{\Psi}_{PP^\top,2} \lambda_p^{\mathsf{sieve}} - \frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i}^3 P_i \left(P_i^\top \lambda_p^{\mathsf{sieve}}\right)^2}{1 + \left(\lambda_p^{\mathsf{sieve}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)}. \tag{S85}$$

Then, by triangle inequality, Lemma 7(b,c), $\left\|\lambda_p^{\mathsf{sieve}}\right\| = O_p\left(\sqrt{k/(nh)}\right)$ and $\left\|\widetilde{\Delta}_{PP^\top,2}\right\| = \operatorname{maxeig}\left(\widetilde{\Delta}_{PP^\top,2}\right) = O(1)$,

$$\left\| \frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i}^3 P_i \left(P_i^\top \lambda_p^{\mathsf{sieve}}\right)^2}{1 + \left(\lambda_p^{\mathsf{sieve}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)} \right\| \le \frac{\left(\max_i \left\|\widehat{W}_{p,i} P_i\right\|\right) \left\|\widehat{\Psi}_{PP^\top,2}\right\| \left\|\lambda_p^{\mathsf{sieve}}\right\|^2}{1 - \left\|\lambda_p^{\mathsf{sieve}}\right\| \max_i \left\|\widehat{W}_{p,i} P_i\right\|} = O_p\left(\frac{\beta_k k}{nh}\right). \tag{S86}$$

By Lemma 7(b) and $\left\|\lambda_p^{\mathsf{sieve}}\right\| = O_p\left(\sqrt{k/(nh)}\right)$,

$$\left\| \left(\widehat{\Psi}_{PP^\top,2} - \widetilde{\Delta}_{PP^\top,2}\right) \lambda_p^{\mathsf{sieve}} \right\| = O_p\left(\frac{\beta_k \sqrt{\log(k)\, k}}{nh}\right).$$

The second conclusion follows from this result, (S85), (S86) and $\left\|\widetilde{\Delta}_{PP^\top,2}^{-1}\right\| = \operatorname{maxeig}\left(\widetilde{\Delta}_{PP^\top,2}^{-1}\right) = O(1)$. ■

Proof of Theorem 3. Denote $\widehat{\vartheta}_Y^{\mathsf{sieve}} := \sum_i w_i^{\mathsf{sieve}} \widehat{W}_{p,i} Y_i / h$ and $\widehat{\vartheta}_D^{\mathsf{sieve}} := \sum_i w_i^{\mathsf{sieve}} \widehat{W}_{p,i} D_i / h$ for notational simplicity. By similar arguments as in the proof of (S22),

$$\widehat{\vartheta}_Y^{\mathsf{sieve}} = \widehat{\Psi}_Y - \widehat{\Psi}_{YP,2}^\top \lambda_p^{\mathsf{sieve}} + \frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i}^3 Y_i \left(\left(\lambda_p^{\mathsf{sieve}}\right)^\top P_i\right)^2}{1 + \left(\lambda_p^{\mathsf{sieve}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)}. \tag{S87}$$

Then, by triangle inequality, Lemma 7(b,c), $\left\|\widetilde{\Delta}_{PP^\top,2}\right\| = O(1)$ and Lemma 8,

$$\left| \frac{1}{nh} \sum_i \frac{\widehat{W}_{p,i}^3 Y_i \left(\left(\lambda_p^{\mathsf{sieve}}\right)^\top P_i\right)^2}{1 + \left(\lambda_p^{\mathsf{sieve}}\right)^\top \left(\widehat{W}_{p,i} P_i\right)} \right| \le \frac{\left(\max_i \left|\widehat{W}_{p,i} Y_i\right|\right) \left(\left\|\widehat{\Psi}_{PP^\top,2}\right\| \left\|\lambda_p^{\mathsf{sieve}}\right\|^2\right)}{1 - \left\|\lambda_p^{\mathsf{sieve}}\right\| \max_i \left\|\widehat{W}_{p,i} P_i\right\|} = O_p\left((nh)^{1/r} \cdot \frac{k}{nh}\right). \tag{S88}$$

Write $\widehat{\Psi}_{YP,2} - \widetilde{\Delta}_{YP,2} = \widehat{\Psi}_{YP,2} - \widetilde{\Psi}_{YP,2} + \widetilde{\Psi}_{YP,2} - \widetilde{\Delta}_{YP,2}$. We have

$$\widehat{\Psi}_{YP,2} - \widetilde{\Psi}_{YP,2} = \frac{1}{nh} \sum_i \left(\widehat{W}_{p,i} - \widetilde{W}_{p,i}\right) \left(\widehat{W}_{p,i} + \widetilde{W}_{p,i}\right) P_i Y_i$$

$$= \frac{1}{nh}\sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,+}^{-1} - \Pi_{p,+}^{-1}\right) r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}\left(X_i > 0\right)\left(\widehat{W}_{p,i} + \widetilde{W}_{p,i}\right) P_i Y_i$$
$$-\frac{1}{nh}\sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,-}^{-1} - \Pi_{p,-}^{-1}\right) r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}\left(X_i < 0\right)\left(\widehat{W}_{p,i} + \widetilde{W}_{p,i}\right) P_i Y_i \tag{S89}$$

By triangle inequality,

$$\left\|\frac{1}{nh}\sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,+}^{-1} - \Pi_{p,+}^{-1}\right) r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}\left(X_i > 0\right)\widetilde{W}_{p,i} P_i Y_i\right\|$$
$$\leq \left\|\widehat{\Pi}_{p,+}^{-1} - \Pi_{p,+}^{-1}\right\| \left(\frac{1}{nh}\sum_i \left\|r_p\left(\frac{X_i}{h}\right)\right\| \left|K\left(\frac{X_i}{h}\right)\right| \left|\mathbb{1}\left(X_i > 0\right)\right| \left|\widetilde{W}_{p,i}\right| \left\|P_i Y_i\right\|\right). \tag{S90}$$

By LIE, change of variables and $\mathrm{tr}\left(\mathrm{E}\left[PP^\top \mid X\right]\right) \leq k\overline{\sigma}$,

$$\mathrm{E}\left[\frac{1}{h}\left\|r_p\left(\frac{X}{h}\right)\right\| \left|K\left(\frac{X}{h}\right)\right| \left|\mathbb{1}\left(X > 0\right)\right| \left|\widetilde{W}_p\right| \left\|PY\right\|\right]$$
$$= \mathrm{E}\left[\frac{1}{h}\left\|r_p\left(\frac{X}{h}\right)\right\| \left|K\left(\frac{X}{h}\right)\right| \left|\mathbb{1}\left(X > 0\right)\right| \left|\widetilde{W}_p\right| \mathrm{E}\left[\left\|PY\right\| \mid X\right]\right]$$
$$\leq \mathrm{E}\left[\frac{1}{h}\left\|r_p\left(\frac{X}{h}\right)\right\| \left|K\left(\frac{X}{h}\right)\right| \left|\mathbb{1}\left(X > 0\right)\right| \left|\widetilde{W}_p\right| \sqrt{\mathrm{tr}\left(\mathrm{E}\left[PP^\top \mid X\right]\right)} \cdot \sqrt{\mathrm{E}\left[Y^2 \mid X\right]}\right] = O\left(\sqrt{k}\right).$$

It now follows from this result, (S90), Markov's inequality and (S5) that

$$\left\|\frac{1}{nh}\sum_i \mathrm{e}_{p+1,1}^\top \left(\widehat{\Pi}_{p,+}^{-1} - \Pi_{p,+}^{-1}\right) r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right) \mathbb{1}\left(X_i > 0\right)\widetilde{W}_{p,i} P_i Y_i\right\| = O_p\left(\sqrt{\frac{k}{nh}}\right).$$

A similar result holds for the second term on the right hand side of the second inequality in (S89). By these results, $\left\|\widehat{\Psi}_{YP,2} - \widetilde{\Psi}_{YP,2}\right\| = \sqrt{k/(nh)}$. And by LIE, change of variables and Lemma 1(a),

$$\mathrm{E}\left[\left\|\widetilde{\Psi}_{YP,2} - \widetilde{\Delta}_{YP,2}\right\|^2\right] \leq \frac{1}{nh^2}\mathrm{E}\left[\widetilde{W}_p^4 \left\|YP\right\|^2\right] \leq \frac{\beta_k^2}{nh}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^4 \left\|Y\right\|^2\right] = O\left(\frac{\beta_k^2}{nh}\right).$$

By Markov's inequality, we have $\Psi_{YP,2} - \widetilde{\Delta}_{YP,2} = O_p\left(\beta_k/\sqrt{nh}\right)$. By Lemma 5 with $\omega = \widetilde{W}_p^2/h$, the fact that $\mathrm{maxeig}\left(\widetilde{\Delta}_{PP^\top,2}\right) = O(1)$ and Lemma 1(a), we have $\left\|\widetilde{\Delta}_{YP,2}\right\| \leq \mathrm{maxeig}\left(\widetilde{\Delta}_{PP^\top,2}\right)\widetilde{\Delta}_{Y^2,2} = O(1)$. Similarly, $\left\|\widetilde{\Delta}_{DP,2}\right\| = O(1)$. Therefore, by this result, (S87), (S88) and Lemma 8,

$$\widehat{\vartheta}_Y^{\mathsf{sieve}} = \widehat{\Psi}_Y - \widetilde{\Delta}_{YP,2}^\top \lambda_p^{\mathsf{sieve}} + O_p\left((nh)^{1/r} \cdot \frac{k}{nh}\right)$$
$$= \widehat{\Psi}_Y - \widetilde{\Delta}_{YP,2}^\top \widetilde{\Delta}_{PP^\top,2}^{-1} \widehat{\Psi}_P + O_p\left(\frac{\beta_k k}{nh} + (nh)^{1/r} \cdot \frac{k}{nh}\right). \tag{S91}$$

By similar arguments,

$$\widehat{\vartheta}_D^{\mathsf{sieve}} = \widehat{\Psi}_D - \widetilde{\Delta}_{DP,2}^{\top}\widetilde{\Delta}_{PP^{\top},2}^{-1}\widehat{\Psi}_P + O_p\left(\frac{\beta_k k}{nh} + \frac{k}{nh}\right). \tag{S92}$$

By Lemma 1(b,c), $\widehat{\Psi}_Y = \mu_{Y,\dagger} + O\left((nh)^{-1/2}\right)$ and $\widehat{\Psi}_D = \mu_{D,\dagger} + O\left((nh)^{-1/2}\right)$. Denote $\epsilon_i^* := M_i - \eta^*(Z_i)$. Now it follows from Lemma 7(a), similar arguments as in the proof of (S23), (S91) and (S92) that

$$\begin{aligned}\widehat{\vartheta}_p^{\mathsf{sieve}} - \vartheta &= \frac{\widehat{\Psi}_M - \gamma_P^{\top}\widehat{\Psi}_P}{\mu_{D,\dagger}} + O_p\left(\frac{\beta_k k}{nh} + (nh)^{1/r}\cdot\frac{k}{nh}\right) \\ &= \frac{\widehat{\Psi}_M - \widehat{\Psi}_{\eta^*(Z)} + \widehat{\Psi}_{\eta^*(Z)} - \gamma_P^{\top}\widehat{\Psi}_P}{\mu_{D,\dagger}} + O_p\left(\frac{\beta_k k}{nh} + (nh)^{1/r}\cdot\frac{k}{nh}\right),\end{aligned} \tag{S93}$$

where we have

$$\begin{aligned}\widehat{\Psi}_M - \widehat{\Psi}_{\eta^*(Z)} &= \widehat{\Psi}_{\epsilon^*} \\ \widehat{\Psi}_{\eta^*(Z)} - \gamma_P^{\top}\widehat{\Psi}_P &= \frac{1}{nh}\sum_i \widehat{W}_{p,i}\left(\eta^*(Z_i) - \gamma_P^{\top}\rho(Z_i)\right).\end{aligned} \tag{S94}$$

Denote $\mathring{\eta} := \eta^* - \gamma_P^{\top}\rho$. By Lemma 6, $\|\mathring{\eta}\|_\infty \downarrow 0$ as $n \uparrow \infty$. Write

$$\frac{1}{nh}\sum_i \widehat{W}_{p,i}\mathring{\eta}(Z_i) = \frac{1}{nh}\sum_i \widehat{W}_{p,i}\left(\mathring{\eta}(Z_i) - g_{\mathring{\eta}(Z)}(X_i)\right) + \frac{1}{nh}\sum_i \widehat{W}_{p,i}g_{\mathring{\eta}(Z)}(X_i). \tag{S95}$$

Then, by Taylor expansion, triangle inequality, (S11), $\mu_{\mathring{\eta}(Z),+} = \mu_{\mathring{\eta}(Z),-}$ and Assumption 4(b),

$$\begin{aligned}\left|\frac{1}{nh}\sum_i \widehat{W}_{p,i}g_{\mathring{\eta}(Z)}(X_i)\right| &= \left|\left(\frac{1}{nh}\sum_i \widehat{W}_{p,i}\frac{g_{\mathring{\eta}(Z)}^{(p+1)}\left(\tilde{X}_i\right)}{(p+1)!}\left(\frac{X_i}{h}\right)^{p+1}\right)h^{p+1}\right| \\ &\le \left\{\frac{1}{nh}\sum_i \left|\widehat{W}_{p,i}\left(\frac{X_i}{h}\right)^{p+1}\right|\right\}\left(\sup_{x\in(-h,0)\cup(0,h)}\left|g_{\mathring{\eta}(Z)}^{(p+1)}(x)\right|\right)\frac{h^{p+1}}{(p+1)!} \\ &= o\left(h^{p+1}\right),\end{aligned}$$

where $\tilde{X}_i$ denotes the mean value that lies between 0 and X_i. And by simple calculation, $\|\mathring{\eta}\|_\infty = o(1)$ and Lemma 1(a),

$$\mathrm{E}\left[\left(\frac{1}{nh}\sum_i \widetilde{W}_{p,i}\left(\mathring{\eta}(Z_i) - g_{\mathring{\eta}(Z)}(X_i)\right)\right)^2\right] = \frac{1}{nh}\cdot\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2\left(\mathring{\eta}(Z) - g_{\mathring{\eta}(Z)}(X)\right)^2\right] \le \frac{1}{nh}\cdot\widetilde{\Delta}_2\cdot\|\mathring{\eta}\|_\infty^2 = o\left((nh)^{-1}\right).$$

By this result and Markov's inequality,

$$\frac{1}{nh}\sum_i \widetilde{W}_{p,i}\left(\mathring{\eta}(Z_i) - g_{\mathring{\eta}(Z)}(X_i)\right) = o_p\left((nh)^{-1/2}\right).$$

By (S5), similar calculation and Markov's inequality,

$$\begin{aligned}
&\left|\frac{1}{nh}\sum_i\left(\widehat{W}_{p,i}-\widetilde{W}_{p,i}\right)\left(\mathring{\eta}\left(Z_i\right)-g_{\mathring{\eta}(Z)}\left(X_i\right)\right)\right| \\
&\qquad\leq\left\|\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right\|\left\|\frac{1}{nh}\sum_i r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\left(\mathring{\eta}\left(Z_i\right)-g_{\mathring{\eta}(Z)}\left(X_i\right)\right)\right\| \\
&\qquad+\left\|\widehat{\Pi}_{p,-}^{-1}-\Pi_{p,-}^{-1}\right\|\left\|\frac{1}{nh}\sum_i r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i<0\right)\left(\mathring{\eta}\left(Z_i\right)-g_{\mathring{\eta}(Z)}\left(X_i\right)\right)\right\|=o_p\left(\left(nh\right)^{-1}\right).
\end{aligned}$$

It follows from these results and (S95) that $(nh)^{-1}\sum_i\widehat{W}_{p,i}\mathring{\eta}\left(Z_i\right)=o_p\left(\left(nh\right)^{-1/2}\right)$. By this result, (S93) and (S94), we have

$$\widehat{\vartheta}_p^{\mathsf{sieve}}-\vartheta=\frac{\widehat{\Psi}_{\epsilon^*}}{\mu_{D,\dagger}}+o_p\left(\left(nh\right)^{-1/2}\right). \tag{S96}$$

It follows from similar arguments as in the proof of (S28) that

$$\sqrt{nh}\left(\widehat{\Psi}_{\epsilon^*}-\frac{\mu_{\epsilon^*,+}^{(p+1)}\omega_{p;+}^{p+1,1}-\mu_{\epsilon^*,-}^{(p+1)}\omega_{p;-}^{p+1,1}}{(p+1)!}h^{p+1}\right)\to_d\mathrm{N}\left(0,\frac{\omega_p^{0,2}\sigma_{\mathsf{opt}}^2}{\varphi}\right).$$

The conclusion follows from this result, (S96) and Slutsky's lemma. ■

S4 Proof of Theorem 4

Let $f:\mathbb{R}^2\to\mathbb{R}$ be any function. Denote $\bar{Y}_i:=\left(Y_i,D_i\right)^\top$. Let

$$\begin{aligned}
R_f\left(\theta\right) &:= \max_{w_1,\ldots,w_n} 2\sum_i\log\left(n\cdot w_i\right) \\
&\quad\text{subject to } f\left(\sum_i w_i\frac{1}{h}W_{p,i}\bar{Y}_i\right)=\theta,\ \sum_i w_iW_{p,i}\bar{Z}_i=0_{d_z+1},\ \sum_i w_i=1
\end{aligned}$$

and for $\mu\in\mathbb{R}^2$, let

$$\begin{aligned}
R\left(\mu\right) &:= \max_{w_1,\ldots,w_n} 2\sum_i\log\left(n\cdot w_i\right) \\
&\quad\text{subject to } \sum_i w_i\frac{1}{h}W_{p,i}\bar{Y}_i=\mu,\ \sum_i w_iW_{p,i}\bar{Z}_i=0_{d_z+1},\ \sum_i w_i=1.
\end{aligned} \tag{S97}$$

Note that $R_f\left(\theta\right)\leq\max\left\{2\sum_i\log\left(n\cdot w_i\right):\sum_i w_i=1\right\}=0$ and similarly, $R\left(\mu\right)\leq 0$. Then it is easy to see that by arguments in Section 3 of the main text,

$$\max_{\theta} R_f(\theta) = \max_{\mu} R(\mu) = \max\left\{2\sum_i \log(n\cdot w_i) : \sum_i w_i W_{p,i}\bar{Z}_i = 0_{d_z+1},\ \sum_i w_i = 1\right\}$$
$$= 2\sum_i \log\left(n\cdot w_i^{\mathsf{mc}}\right) =: R^*.$$

By similar arguments as those in Owen (2001, Section 2.4), we can write

$$\begin{aligned}\{\mu : R(\mu) - R^* \geq -c_\tau\} &= \Bigg\{\mu : \exists w_1, ..., w_n \text{ such that } 2\sum_i \log(n\cdot w_i) - R^* \geq -c_\tau,\\ &\quad \sum_i w_i \frac{1}{h} W_{p,i}\bar{Y}_i = \mu,\ \sum_i w_i W_{p,i}\bar{Z}_i = 0_{d_z+1},\ \sum_i w_i = 1\Bigg\}\\ &= \left\{\sum_i w_i \frac{1}{h} W_{p,i}\bar{Y}_i : 2\sum_i \log(n\cdot w_i) - R^* \geq -c_\tau,\ \sum_i w_i W_{p,i}\bar{Z}_i = 0_{d_z+1},\ \sum_i w_i = 1\right\}\end{aligned}$$

and

$$\begin{aligned}\{\theta : R_f(\theta) - R^* \geq -c_\tau\} &= \Bigg\{\theta : \exists w_1, ..., w_n \text{ such that } 2\sum_i \log(n\cdot w_i) - R^* \geq -c_\tau,\\ &\quad f\left(\sum_i w_i \frac{1}{h} W_{p,i}\bar{Y}_i\right) = \theta,\ \sum_i w_i W_{p,i}\bar{Z}_i = 0_{d_z+1},\ \sum_i w_i = 1\Bigg\}\\ &= \left\{f\left(\sum_i w_i \frac{1}{h} W_{p,i}\bar{Y}_i\right) : 2\sum_i \log(n\cdot w_i) - R^* \geq -c_\tau,\ \sum_i w_i W_{p,i}\bar{Z}_i = 0_{d_z+1},\ \sum_i w_i = 1\right\}\\ &= f\left(\left\{\sum_i w_i \frac{1}{h} W_{p,i}\bar{Y}_i : 2\sum_i \log(n\cdot w_i) - R^* \geq -c_\tau,\ \sum_i w_i W_{p,i}\bar{Z}_i = 0_{d_z+1},\ \sum_i w_i = 1\right\}\right)\\ &= f\left(\{\mu : R(\mu) - R^* \geq -c_\tau\}\right). \end{aligned} \tag{S98}$$

For any $\alpha \in [0,1]$ and $\mu_1, \mu_2 \in \{\mu : R(\mu) - R^* \geq -c_\tau\}$, denote $\bar{\mu} := \alpha\mu_1 + (1-\alpha)\mu_2$. Let $(\bar{w}_1, ..., \bar{w}_n)$, $(w_{1,1}, ..., w_{1,n})$ and $(w_{2,1}, ..., w_{2,n})$ be the optimal weights associated with $R(\bar{\mu})$, $R(\mu_1)$ and $R(\mu_2)$ in (S97). Then, since $\sum_i w_{1,i} W_{p,i}\bar{Y}_i/h = \mu_1$ and $\sum_i w_{2,i} W_{p,i}\bar{Y}_i/h = \mu_2$, we have $\sum_i (\alpha w_{1,i} + (1-\alpha) w_{2,i}) W_{p,i}\bar{Y}_i/h = \bar{\mu}$. By definition of $(\bar{w}_1, ..., \bar{w}_n)$, $(w_{1,1}, ..., w_{1,n})$ and $(w_{2,1}, ..., w_{2,n})$, concavity of $\log(\cdot)$ and Jensen's inequality,

$$\sum_i \log(n\cdot \bar{w}_i) \geq \sum_i \log(n\cdot(\alpha w_{1,i} + (1-\alpha) w_{2,i})) \geq \alpha \sum_i \log(n\cdot w_{1,i}) + (1-\alpha)\sum_i \log(n\cdot w_{2,i}) \geq \frac{-c_\tau + R^*}{2}.$$

It follows from these inequalities that $R(\bar{\mu}) - R^* \geq -c_\tau$. Therefore, $\{\mu : R(\mu) - R^* \geq -c_\tau\} \subseteq \mathbb{R}^2$ is convex.

Write

$$\tilde{R}(\theta) := \max_{w_1,...,w_n} 2\sum_i \log(n\cdot w_i)$$

$$\text{subject to } \sum_i w_i \; W_{p,i} \begin{pmatrix} Y_i - \theta D_i \\ \bar{Z}_i \end{pmatrix} = 0_{d_z+2}, \; \sum_i w_i = 1.$$

Then, $\tilde{R}(\theta) = -2n \cdot \ell_p^{\mathsf{mc}}(\theta \mid h)$. Note that we have $2n \cdot \ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}} \mid h\right) = -R^*$. Then, $LR_p(\theta \mid h) \leq c_\tau$ if and only if $\tilde{R}(\theta) - R^* \geq -c_\tau$. And, it is clear that $CS_{p,\tau}(h) = \left\{\theta : \tilde{R}(\theta) - R^* \geq -c_\tau\right\}$.

Consider the coordinate projection $f_D(\{\mu : R(\mu) - R^* \geq -c_\tau\})$ of $\{\mu : R(\mu) - R^* \geq -c_\tau\}$, where $f_D(x, y) \coloneqq y$. Then, by (S98) with $f = f_D$, we have $f_D(\{\mu : R(\mu) - R^* \geq -c_\tau\}) = \{\theta : R_{f_D}(\theta) - R^* \geq -c_\tau\}$, where

$$\begin{aligned} R_{f_D}(\theta) \;\; &\coloneqq \;\; \max_{w_1,\ldots,w_n} 2\sum_i \log(n \cdot w_i) \\ &\qquad \text{subject to } \sum_i w_i \frac{1}{h} W_{p,i} D_i = \theta, \; \sum_i w_i W_{p,i} \bar{Z}_i = 0_{d_z+1}, \; \sum_i w_i = 1. \end{aligned}$$

Note that $CS_\tau^D(h) \coloneqq \{\theta : R_{f_D}(\theta) - R^* \geq -c_\tau\}$ is an EL confidence set for $\psi_{D,\dagger}$. And it can be shown by similar arguments as in the proof of Theorem 5 that $R^* - R_{f_D}(\psi_{D,\dagger}) \to_d \chi_1^2$ and $\Pr\left[\psi_{D,\dagger} \in CS_\tau^D(h)\right] \to 1 - \tau$. Let $\widehat{\theta}_D \coloneqq \mathrm{argmax}_\theta R_{f_D}(\theta)$. By similar arguments, $\widehat{\theta}_D = \sum_i w_i^{\mathsf{mc}} W_{p,i} D_i / h$ (i.e., $\widehat{\theta}_D$ is a covariate-adjusted estimator of $\psi_{D,\dagger}$) and $R_{f_D}\left(\widehat{\theta}_D\right) = R^*$. Assume $\mu_{D,\dagger} > 0$ without loss of generality. It is clear that $\widehat{\theta}_D \in CS_\tau^D(h)$. By convexity of $\{\mu : R(\mu) - R^* \geq -c_\tau\}$, $CS_\tau^D(h)$ is also convex. Let $\sigma_D^2 \coloneqq \Delta_{D,2} - \Delta_{D\bar{Z}^\top,2} \Delta_{\bar{Z}\bar{Z}^\top,2}^{-1} \Delta_{D\bar{Z},2}$. By Lemma 3(a), $\sigma_D^2 > 0$ is bounded away from zero when h is sufficiently small. By similar arguments as in the proofs of Theorems 1 and 5, we can show $\widehat{\theta}_D \to_p \psi_{D,\dagger}$ and a result similar to Owen (1988, Corollary 1): for any $\delta \in \mathbb{R}$, $R_{f_D}\left(\widehat{\theta}_D\right) - R_{f_D}\left(\widehat{\theta}_D + \sigma_D \delta / \sqrt{nh}\right) \to_p \delta^2$. Fix $\delta_0 > 0$ such that $\delta_0^2 > c_\tau$. Then wpa1, $R^* - R_{f_D}\left(\widehat{\theta}_D - \sigma_D \delta_0 / \sqrt{nh}\right) > c_\tau$ and by convexity of $CS_\tau^D(h)$ and the fact that $\widehat{\theta}_D \in CS_\tau^D(h)$,

$$\frac{\psi_{D,\dagger}}{2} < \widehat{\theta}_D - \frac{\sigma_D \delta_0}{\sqrt{nh}} \leq \inf CS_\tau^D(h) = \inf f_D(\{\mu : R(\mu) - R^* \geq -c_\tau\}) \leq \widehat{\theta}_D.$$

Therefore, wpa1, $\{\mu : R(\mu) - R^* \geq -c_\tau\} \subseteq \mathbb{R} \times [\psi_{D,\dagger}/2, \infty)$. Let $\bar{f}(x, y) \coloneqq x/y$. $\bar{f}$ is a continuous function on $\mathbb{R} \times [\psi_{D,\dagger}/2, \infty)$. Then it is clear that $\tilde{R}(\theta) = R_{\bar{f}}(\theta)$ and by (S98), $CS_\tau(h) = \left\{\tilde{R}(\theta) - R^* \geq -c_\tau\right\} = \bar{f}(\{\mu : R(\mu) - R^* \geq -c_\tau\})$. By similar arguments, wpa1,

$$\sup f_D(\{\mu : R(\mu) - R^* \geq -c_\tau\}) = \sup CS_\tau^D(h) \leq \widehat{\theta}_D + \frac{\sigma_D \delta_0}{\sqrt{nh}} < \frac{3}{2} \psi_{D,\dagger}$$

and therefore, $f_D(\{\mu : R(\mu) - R^* \geq -c_\tau\})$ is bounded. By using the same arguments to the other coordinate projection, we can show that wpa1, $\{\mu : R(\mu) - R^* \geq -c_\tau\}$ is bounded. The conclusion follows from convexity and boundedness of $\{\mu : R(\mu) - R^* \geq -c_\tau\}$ and continuity of $\bar{f}$.

S5 Proof of Theorem 5

For a sequence of classes of $\mathbb{R}$-valued functions $\mathfrak{F}_n$ defined on $\mathcal{S}$ (a compact set in a finite-dimensional Euclidean space), let $\|f\|_{Q,2} \coloneqq \left(\int f^2 \mathrm{d}Q\right)^{1/2}$ and $N\left(\varepsilon, \mathfrak{F}_n, \|\cdot\|_{Q,2}\right)$ denote the ε-covering number, i.e., the smallest integer m such that there are m balls of radius $\varepsilon > 0$ (with respect to $\|\cdot\|_{Q,2}$) centered at points in $\mathfrak{F}_n$ whose union covers $\mathfrak{F}_n$. A function $F_{\mathfrak{F}_n} : \mathcal{S} \to \mathbb{R}_+$ is an envelope of $\mathfrak{F}_n$ if $\sup_{f\in\mathfrak{F}_n} |f| \leq F_{\mathfrak{F}_n}$. We say that $\mathfrak{F}_n$ is a (uniform) Vapnik–Chervonenkis-type (VC-type) class with respect to the envelope $F_{\mathfrak{F}_n}$ (see, e.g., Chernozhukov et al., 2014b, Definition 2.1) if there exist some positive constants (VC characteristics) $A_{\mathfrak{F}_n} \geq e$ and $V_{\mathfrak{F}_n} > 1$ that are independent of the sample size n such that $\sup_{Q\in\mathcal{Q}_{\mathcal{S}}^{\mathtt{fd}}} N\left(\varepsilon \|F_{\mathfrak{F}_n}\|_{Q,2}, \mathfrak{F}_n, \|\cdot\|_{Q,2}\right) \leq (A_{\mathfrak{F}_n}/\varepsilon)^{V_{\mathfrak{F}_n}}$ $\forall \varepsilon \in (0,1]$ where $\mathcal{Q}_{\mathcal{S}}^{\mathtt{fd}}$ denotes the collection of all finitely discrete probability measures on $\mathcal{S}$. In the proofs in this section, whenever applied to quantities that depend on h, the $O_p(\cdot)$ and $o_p(\cdot)$ ($O(\cdot)$ and $o(\cdot)$) notations are understood as being uniform in $h \in \mathbb{H} \coloneqq \left[\underline{h}, \overline{h}\right]$. For notational simplicity, let $\underline{n} \coloneqq n\underline{h}$ and $\overline{n} \coloneqq n\overline{h}$.

Lemma 9. *Let V denote a random variable and $\{V_1, ..., V_n\}$ are i.i.d. copies of V. Let $\mathbb{B}$ denote an open neighborhood of 0. Suppose that $\underline{h}$ and $\overline{h}$ satisfy $\overline{h} = o(1)$ and K is a symmetric continuous PDF supported on $[-1,1]$. The following results hold for all $(\mathrm{s}, k) \in \{-,+\} \times \mathbb{N}$, uniformly in $h \in \mathbb{H}$: (a) if g_V is Lipschitz continuous on $\mathbb{B} \setminus \{0\}$, for $k \geq 2$,*

$$\mathrm{E}\left[\frac{1}{h} W_{p;\mathrm{s}}^{k} V\right] = \psi_{V,\mathrm{s}} \omega_p^{0,k} + O\left(\overline{h}\right);$$

(b) if g_V is $(p+1)$-times continuously differentiable with uniformly continuous $g_V^{(p+1)}$ on $\mathbb{B} \setminus \{0\}$,

$$\mathrm{E}\left[\frac{1}{h} W_{p;\mathrm{s}} V\right] = \psi_{V,\mathrm{s}} + \frac{\psi_{V,\mathrm{s}}^{(p+1)}}{(p+1)!} \omega_{p;\mathrm{s}}^{p+1,1} h^{p+1} + o\left(\overline{h}^{p+1}\right);$$

(c) if $g_{|V|^r}$ is bounded on $\mathbb{B} \setminus \{0\}$ for some integer $r > 2$,

$$\frac{1}{\sqrt{nh}} \sum_i \left(W_{p;\mathrm{s},i}^{k} V_i - \mathrm{E}\left[W_{p;\mathrm{s}}^{k} V\right]\right) = O_p\left(\sqrt{\log(n)} + \log(n) \cdot \frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right);$$

(d) if $g_{|V|^r}$ is bounded on $\mathbb{B} \setminus \{0\}$ for some integer $r > 2$, $\max_i |W_{p;\mathrm{s},i} V_i| / \sqrt{nh} = O_p\left(\overline{n}^{1/r} / \underline{n}^{1/2}\right)$.

Proof of Lemma 9. We take $\mathrm{s} = +$ without loss of generality. Part (a) and Part (b) are straightforward extensions of Lemma 3(a,b) and therefore follow from similar arguments. For (c), denote $\bar{q}(V, X \mid h) \coloneqq h^{-1/2} W_{p;+}^{k} V$ and $\bar{\mathfrak{Q}} \coloneqq \{\bar{q}(\cdot \mid h) : h \in \mathbb{H}\}$. Denote $\mathbb{P}_n^V f \coloneqq n^{-1} \sum_i f(V_i, X_i)$, $\mathbb{P}^V f \coloneqq \mathrm{E}[f(V, X)]$ and $\mathbb{G}_n^V \coloneqq$

$\sqrt{n}\left(\mathbb{P}_n^V-\mathbb{P}^V\right)$. Then we have

$$\left\|\mathbb{G}_n^V\right\|_{\bar{\mathfrak{Q}}}=\sup_{h\in\mathbb{H}}\left|\frac{1}{\sqrt{nh}}\sum_i\left(W_{p;+,i}^kV_i-\mathrm{E}\left[W_{p;+}^kV\right]\right)\right|.$$

Let $\sigma_{\bar{\mathfrak{Q}}}^2:=\sup_{f\in\bar{\mathfrak{Q}}}\mathbb{P}^Vf^2$. It follows from LIE and change of variables that $\sigma_{\bar{\mathfrak{Q}}}^2=\sup_{h\in\mathbb{H}}\mathrm{E}\left[h^{-1}W_{p;+}^{2k}g_{V^2}(X)\right]=O(1)$. Since K is supported on $[-1,1]$, we can write $\bar{q}(v,x\mid h)=\mathcal{K}_{p;+}^k(x/h)\,h^{-1/2}\mathbb{1}(0<x<h)\,v$. Since Assumption 3(b) also implies that $\mathcal{K}_{p;+}^k$ has bounded variation $\forall k\in\mathbb{N}$. By Giné and Nickl (2015, Proposition 3.6.12), $\left\{x\mapsto\mathcal{K}_{p;+}^k(x/h):h\in\mathbb{H}\right\}$ is VC-type with respect to a constant envelope and its VC characteristics are independent of n. By Kosorok (2007, Lemma 9.6), $\left\{(x,v)\mapsto h^{-1/2}\mathbb{1}(0<x<h)\,v:h\in\mathbb{H}\right\}$ is VC-subgraph with an envelope $(x,v)\mapsto\underline{h}^{-1/2}\mathbb{1}\left(0<x<\overline{h}\right)|v|$ and VC index being at most 3. By Kosorok (2007, Theorem 9.3) and Chernozhukov et al. (2014b, Corollary A.1), $\bar{\mathfrak{Q}}$ is VC-type with respect to an envelope $F_{\bar{\mathfrak{Q}}}$ that is proportional to $(x,v)\mapsto\underline{h}^{-1/2}\mathbb{1}\left(0<x<\overline{h}\right)|v|$. By Chen and Kato (2020, Corollary 5.5), $\mathrm{E}\left[\left\|\mathbb{G}_n^V\right\|_{\bar{\mathfrak{Q}}}\right]\lesssim\sigma_{\bar{\mathfrak{Q}}}\sqrt{\log(n)}+\log(n)\left(\mathbb{P}^V\left|F_{\bar{\mathfrak{Q}}}\right|^r\right)^{1/r}n^{1/r}/\sqrt{n}$, where $\mathbb{P}^V\left|F_{\bar{\mathfrak{Q}}}\right|^r=O\left(\overline{h}/\underline{h}^{r/2}\right)$. (c) follows from Markov's inequality.

For Part (d), since $|K(X_i/h)|\lesssim\mathbb{1}(|X_i|\leq h)$,

$$\frac{\max_i\left|W_{p;+,i}V_i\right|}{\sqrt{nh}}\lesssim\frac{\max_i\mathbb{1}\left(0<X_i<\overline{h}\right)|V_i|}{\sqrt{\underline{n}}}\leq\frac{\left(\sum_i\mathbb{1}\left(0<X_i<\overline{h}\right)|V_i|^r\right)^{1/r}}{\sqrt{\underline{n}}}.\tag{S99}$$

Then,

$$\mathrm{E}\left[\sum_i\mathbb{1}\left(0<X_i<\overline{h}\right)|V_i|^r\right]=n\int_0^{\overline{h}}g_{|V|^r}(x)\,f_X(x)\,\mathrm{d}x=O(\overline{n}),$$

where the second equality follows from boundedness of $g_{|V|^r}$ and continuity of f_X. By this result, (S99) and Markov's inequality, we have $\max_i\left|W_{p;+,i}V_i\right|/\sqrt{nh}=O_p\left(\overline{n}^{1/r}/\underline{n}^{1/2}\right)$. ■

We define

$$\widetilde{\lambda}_p^{\mathsf{mc}}:=\operatorname*{argmax}_{\lambda}\sum_i\log\left(1+\lambda^\top\left(W_{p,i}U_i\right)\right).$$

Then we have

$$\begin{aligned}LR_p(\vartheta\mid h)&=2n\left(\ell_p^{\mathsf{mc}}(\vartheta\mid h)-\ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}}\mid h\right)\right)\\&=2\left\{\sum_i\log\left(1+\left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}U_i\right)\right)-\sum_i\log\left(1+\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right)\right\}.\end{aligned}$$

The following lemma provides uniform-in-bandwidth rate of convergence and linearization for λ_p^{mc} and $\widetilde{\lambda}_p^{\mathsf{mc}}$

Lemma 10. *Suppose that the assumptions in the statement of Theorem 5 hold. Assume that* $n\overline{h}^{2p+3}=$

$O(1)$, *and* $\log(n)\left(\overline{n}^{1/4}/\underline{n}^{1/2}\right) = o(1)$. *Then, the following results hold uniformly in* $h \in \mathbb{H}$: *(a)* $\sqrt{nh} \cdot \lambda_p^{\mathsf{mc}} = O_p\left(\sqrt{\log(n)}\right)$ *and* $\sqrt{nh} \cdot \lambda_p^{\mathsf{mc}} = \Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\left(\sqrt{nh} \cdot \Psi_{\bar{Z}}\right) + O_p\left(\log(n)\left(\overline{n}^{1/r}/\underline{n}^{1/2}\right)\right)$; *(b)* $\sqrt{nh} \cdot \widetilde{\lambda}_p^{\mathsf{mc}} = O_p\left(\sqrt{\log(n)}\right)$ *and* $\sqrt{nh} \cdot \widetilde{\lambda}_p^{\mathsf{mc}} = \Delta_{UU^\top,2}^{-1}\left(\sqrt{nh} \cdot \Psi_U\right) + O_p\left(\log(n)\left(\overline{n}^{1/r}/\underline{n}^{1/2}\right)\right)$.

Proof of Lemma 10. By arguments similar to those used in the proof of Lemma 2,

$$\left(\lambda_p^{\mathsf{mc}}\right)^\top \Delta_{\bar{Z}\bar{Z}^\top,2}\lambda_p^{\mathsf{mc}} \leq \left\|\lambda_p^{\mathsf{mc}}\right\| \left\|\Psi_{\bar{Z}}\right\| \left\{1 + \left\|\lambda_p^{\mathsf{mc}}\right\| \max_i \left\|W_{p,i}\bar{Z}_i\right\|\right\} + \left\|\lambda_p^{\mathsf{mc}}\right\|^2 \left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\|$$

and

$$\left(\operatorname{mineig}\left(\Delta_{\bar{Z}\bar{Z}^\top,2}\right) - \left\|\Psi_{\bar{Z}}\right\| \max_i \left\|W_{p,i}\bar{Z}_i\right\| - \left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\|\right) \left\|\lambda_p^{\mathsf{mc}}\right\| \leq \left\|\Psi_{\bar{Z}}\right\|. \tag{S100}$$

Write

$$\sqrt{nh} \cdot \Psi_{\bar{Z}} = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \left(W_{p,i}\bar{Z}_i - \mathrm{E}\left[W_p\bar{Z}\right]\right) + \sqrt{nh} \cdot \Delta_{\bar{Z}}.$$

It follows from Lemma 9(c) that the first term on the right hand side is $O_p\left(\sqrt{\log(n)}\right)$. By Lemma 9(b), $\sqrt{nh} \cdot \Delta_{\bar{Z}} = O\left(\sqrt{n\overline{h}^{2p+3}}\right)$. Therefore, $\sqrt{nh} \cdot \Psi_{\bar{Z}} = O_p\left(\sqrt{\log(n)}\right)$. It also follows from this result, Lemma 9(d) that $\left\|\Psi_{\bar{Z}}\right\| \left(\max_i \left\|W_{p,i}\bar{Z}_i\right\|\right) = o_p(1)$. By Lemma 9(c), $\left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\| = O_p\left(\sqrt{\log(n)/\underline{n}}\right)$. Since it follows from Lemma 9(a) that $\Delta_{\bar{Z}\bar{Z}^\top,2} = \psi_{\bar{Z}\bar{Z}^\top,\pm}\omega_p^{0,2} + o(1)$, we have

$$\inf_{h\in\mathbb{H}} \operatorname{mineig}\left(\Delta_{\bar{Z}\bar{Z}^\top,2}\right) = \operatorname{mineig}\left(\psi_{\bar{Z}\bar{Z}^\top,\pm}\omega_p^{0,2}\right) + O\left(\overline{h}\right).$$

Therefore wpa1, for all $h \in \mathbb{H}$,

$$\operatorname{mineig}\left(\Delta_{\bar{Z}\bar{Z}^\top,2}\right) - \left\|\Psi_{\bar{Z}}\right\| \max_i \left\|W_{p,i}\bar{Z}_i\right\| - \left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\| > \frac{1}{2} \cdot \operatorname{mineig}\left(\psi_{\bar{Z}\bar{Z}^\top,\pm}\omega_p^{0,2}\right) > 0.$$

Therefore, by (S100) and $\sqrt{nh} \cdot \Psi_{\bar{Z}} = O_p\left(\sqrt{\log(n)}\right)$, $\left\|\sqrt{nh} \cdot \lambda_p^{\mathsf{mc}}\right\| = O_p\left(\sqrt{\log(n)}\right)$. By arguments similar to those used to show (S20), we have

$$\Psi_{\bar{Z}} = \Psi_{\bar{Z}\bar{Z}^\top,2}\lambda_p^{\mathsf{mc}} - \frac{1}{nh} \sum_i \frac{W_{p,i}^2\bar{Z}_i\left(\bar{Z}_i^\top\lambda_p^{\mathsf{mc}}\right)\left(\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right)}{1 + \left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)}. \tag{S101}$$

By writing $(nh)^{-1}\sum_i \left\|W_{p,i}\bar{Z}_i\right\|^2 = \Psi_{\|\bar{Z}\|^2,2}$ as $\left(\Psi_{\|\bar{Z}\|^2,2} - \Delta_{\|\bar{Z}\|^2,2}\right) + \Delta_{\|\bar{Z}\|^2,2}$ and Lemma 9(a,c), we have $\Psi_{\|\bar{Z}\|^2,2} = O_p(1)$. Then by using this result, Lemma 9(d) and $\left\|\sqrt{nh} \cdot \lambda_p^{\mathsf{mc}}\right\| = O_p\left(\sqrt{\log(n)}\right)$, we have

$$\left\|\frac{1}{\sqrt{nh}} \sum_i \frac{W_{p,i}^2\bar{Z}_i\left(\bar{Z}_i^\top\lambda_p^{\mathsf{mc}}\right)\left(\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right)}{1 + \left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)}\right\| \leq \frac{\left(\frac{1}{nh}\sum_i \left\|W_{p,i}\bar{Z}_i\right\|^2\right)\left\|\sqrt{nh} \cdot \lambda_p^{\mathsf{mc}}\right\| \left\|\lambda_p^{\mathsf{mc}}\right\| \left(\max_i \left\|W_{p,i}\bar{Z}_i\right\|\right)}{1 - \left\|\lambda_p^{\mathsf{mc}}\right\| \max_i \left\|W_{p,i}\bar{Z}_i\right\|}$$

$$= \quad O_p\left(\log(n)\cdot\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right). \tag{S102}$$

The second conclusion follows from this result, $\left\|\Psi_{\bar{Z}\bar{Z}^\top,2}-\Delta_{\bar{Z}\bar{Z}^\top,2}\right\| = O_p\left(\sqrt{\log(n)/\underline{n}}\right)$, $\left\|\sqrt{nh}\cdot\lambda_p^{\mathsf{mc}}\right\| = O_p\left(\sqrt{\log(n)}\right)$ and (S101). Part (b) follows from the same arguments. ∎

The following lemma shows that $\{LR_p(\vartheta \mid h) : h\in\mathbb{H}\}$ can be approximated by the square of an empirical process indexed by $h\in\mathbb{H}$. Denote $\mathbb{P}_n^T f := n^{-1}\sum_i f(T_i,X_i)$, $\mathbb{P}^T f := \mathrm{E}[f(T,X)]$ and $\mathbb{G}_n^T := \sqrt{n}\left(\mathbb{P}_n^T-\mathbb{P}^T\right)$, where $T_i := \left(Y_i,D_i,Z_i^\top\right)^\top$ (similarly, $T := \left(Y,D,Z^\top\right)^\top$). Denote $\|F\|_{\mathbb{P}^T,r} := \left(\mathbb{P}^T|F|^r\right)^{1/r}$. Let $\xi(x) := \mathrm{E}\left[\mathring{\epsilon}^2 \mid |X|=x\right]$ and $q(\cdot\mid h)$ be defined by $q(T_i,X_i\mid h) := h^{-1/2}W_{p,i}\mathring{\epsilon}_i/\sqrt{\xi(|X_i|)f_{|X|}(|X_i|)\,\omega_p^{0,2}}$, where $f_{|X|}$ denotes the PDF of $|X|$.

Lemma 11. *Suppose that the assumptions in the statement of Theorem 5 hold. Then,*

$$LR_p(\vartheta\mid h) = \left\{\mathbb{G}_n^T q(\cdot\mid h)\right\}^2 + O_p\left(\log(n)\,\overline{h} + \log(n)^{3/2}\left(\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right)\right).$$

Proof of Lemma 11. By Taylor expansion, $2n\cdot\ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}}\mid h\right)$ is equal to the sum of $2\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(\sum_i W_{p,i}\bar{Z}_i\right)$, $-\sum_i\left(\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right)^2$ and a remainder term bounded by $\sum_i\left|\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right|^3/\left(1-\left|\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right|\right)^3$ up to a constant. By Lemma 9(d), Lemma 10 and $(nh)^{-1}\sum_i\left\|W_{p,i}\bar{Z}_i\right\|^2 = O_p(1)$,

$$\begin{aligned}\sum_i\frac{\left|\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right|^3}{\left(1-\left|\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right|\right)^3} &\leq \frac{\left(\sum_i\left\|W_{p,i}\bar{Z}_i\right\|^2\right)\left\|\lambda_p^{\mathsf{mc}}\right\|^3\left(\max_i\left\|W_{p,i}\bar{Z}_i\right\|\right)}{\left(1-\left\|\lambda_p^{\mathsf{mc}}\right\|\max_i\left\|W_{p,i}\bar{Z}_i\right\|\right)^3}\\ &= O_p\left(\log(n)^{3/2}\cdot\left(\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right)\right).\end{aligned}$$

Therefore,

$$\sum_i\log\left(1+\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right) = 2\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(\sum_i W_{p,i}\bar{Z}_i\right) - \sum_i\left(\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right)^2 + O_p\left(\log(n)^{3/2}\cdot\left(\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right)\right).$$

It follows from these results, Lemma 10(a) and $\Psi_{\bar{Z}\bar{Z}^\top,2}-\Delta_{\bar{Z}\bar{Z}^\top,2} = O_p\left(\sqrt{\log(n)/\underline{n}}\right)$ that

$$\sum_i\log\left(1+\left(\lambda_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\bar{Z}_i\right)\right) = \left(\sqrt{nh}\cdot\Psi_{\bar{Z}}\right)^\top\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\left(\sqrt{nh}\cdot\Psi_{\bar{Z}}\right) + O_p\left(\log(n)^{3/2}\cdot\left(\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right)\right).$$

Similarly, we have

$$\sum_i \log\left(1+\left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top (W_{p,i}U_i)\right) = \left(\sqrt{nh}\cdot\Psi_U\right)^\top \Delta_{UU^\top,2}^{-1}\left(\sqrt{nh}\cdot\Psi_U\right) + O_p\left(\log(n)^{3/2}\cdot\left(\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right)\right).$$

By (S46),

$$LR_p(\vartheta\mid h) = \left(\sqrt{nh}\cdot\Psi_U\right)^\top\left(\Delta_{UU^\top,2}^{-1}-\mathrm{Q}\right)\left(\sqrt{nh}\cdot\Psi_U\right) + O_p\left(\log(n)^{3/2}\cdot\left(\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right)\right).$$

Then by (S54),

$$\begin{aligned}\left(\sqrt{nh}\cdot\Psi_U\right)^\top\left(\Delta_{UU^\top,2}^{-1}-\mathrm{Q}\right)\left(\sqrt{nh}\cdot\Psi_U\right) &= \left(\sqrt{nh}\cdot\Psi_U\right)^\top\Delta_{UU^\top,2}^{-1}\Delta_G\left(\Delta_G^\top\Delta_{UU^\top,2}^{-1}\Delta_G\right)^{-1}\Delta_G^\top\Delta_{UU^\top,2}^{-1}\left(\sqrt{nh}\cdot\Psi_U\right)\\ &= \frac{(nh)\left(\Psi_M-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Psi_{\bar{Z}}\right)^2}{\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2}}.\end{aligned} \tag{S103}$$

By $\bar{\gamma}_M=\gamma_M+O\left(\overline{h}\right)$ which follows from Lemma 9(a) and

$$\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2} = \frac{\Delta_{M,2}^2}{\Delta_2}+\bar{\gamma}_M^\top\left(\Delta_{MZ,2}-\frac{\Delta_{M,2}\Delta_{Z,2}}{\Delta_2}\right),$$

which follows from (S37), we have

$$\begin{aligned}\Delta_{M^2,2}-\Delta_{M\bar{Z}^\top,2}\Delta_{\bar{Z}\bar{Z}^\top,2}^{-1}\Delta_{M\bar{Z},2} &= \mathrm{E}\left[\frac{1}{h}W_p^2\left(M-\Delta_{M,2}\Delta_2^{-1}-Z^\top\bar{\gamma}_M+\Delta_2^{-1}\Delta_{Z^\top,2}\bar{\gamma}_M\right)^2\right]\\ &= \Delta_{\mathring{\epsilon}^2,2}+O\left(\overline{h}\right).\end{aligned}$$

By $\left\|\sqrt{nh}\cdot\Psi_U\right\|=O_p\left(\sqrt{\log(n)}\right)$ and $\bar{\gamma}_M=\gamma_M+O\left(\overline{h}\right)$, the numerator on the right hand side of the second equality in (S103) is $\left\{(nh)^{-1/2}\sum_i W_{p,i}\mathring{\epsilon}_i\right\}^2+O_p\left(\log(n)\,\overline{h}\right)$. Let $\tilde{q}(T_i,X_i\mid h):=h^{-1/2}W_{p,i}\mathring{\epsilon}_i/\sqrt{\Delta_{\mathring{\epsilon}^2,2}}$ and $\tilde{\mathfrak{Q}}:=\{\tilde{q}(\cdot\mid h):h\in\mathbb{H}\}$. Then it is clear that $\left\{(nh)^{-1/2}\sum_i W_{p,i}\mathring{\epsilon}_i\right\}^2/\Delta_{\mathring{\epsilon}^2,2}=\left\{\mathbb{G}_n^T\tilde{q}(\cdot\mid h)\right\}^2$ and therefore,

$$LR_p(\vartheta\mid h)=\left\{\mathbb{G}_n^T\tilde{q}(\cdot\mid h)\right\}^2+O_p\left(\log(n)\,\overline{h}+\log(n)^{3/2}\cdot\left(\frac{\overline{n}^{1/r}}{\underline{n}^{1/2}}\right)\right).$$

Denote $\mathfrak{Q}:=\{q(\cdot\mid h):h\in\mathbb{H}\}$ and $\mathfrak{D}:=\{q(\cdot\mid h)-\tilde{q}(\cdot\mid h):h\in\mathbb{H}\}$. Then it follows arguments similar to those used in the proof of Lemma 9(c) that $\tilde{\mathfrak{Q}}$ and $\mathfrak{Q}$ are both VC-type with respect to envelopes $F_{\tilde{\mathfrak{Q}}}$ and $F_{\mathfrak{Q}}$, where $F_{\tilde{\mathfrak{Q}}}(T_i,X_i)$ is proportional to $\underline{h}^{-1/2}\mathbb{1}\left(|X_i|\le\overline{h}\right)|\mathring{\epsilon}_i|/\sqrt{\inf_{h\in\mathbb{H}}\Delta_{\mathring{\epsilon}^2,2}}$ and $F_{\mathfrak{Q}}(T_i,X_i)$ is proportional to $\underline{h}^{-1/2}\mathbb{1}\left(|X_i|\le\overline{h}\right)|\mathring{\epsilon}_i|/\sqrt{\xi(|X_i|)\,f_{|X|}(|X_i|)}$, respectively. By change of variables, $\mathbb{P}^T F_{\tilde{\mathfrak{Q}}}^r\asymp\mathbb{P}^T F_{\mathfrak{Q}}^r=O\left(\overline{h}/\underline{h}^{r/2}\right)$. By Chernozhukov et al. (2014b, Lemma A.6), $\mathfrak{D}$ is VC-type with respect to the envelope

$F_{\mathfrak{D}} = F_{\tilde{\mathfrak{Q}}} + F_{\mathfrak{Q}}$. Let

$$\sigma_{\mathfrak{D}}^2 \coloneqq \sup_{f \in \mathfrak{D}} \mathbb{P}^T f^2 = \sup_{h \in \mathbb{H}} \mathrm{E}\left[\left(q\left(T, X \mid h\right) - \tilde{q}\left(T, X \mid h\right)\right)^2\right].$$

By LIE and the fact that $\left(W_{p;+} + W_{p;-}\right)^2 = \mathcal{K}_{p;+}^2\left(|X|/h\right)$,

$$\begin{aligned}
\mathrm{E}\left[\left(q\left(T, X \mid h\right) - \tilde{q}\left(T, X \mid h\right)\right)^2\right] &= \mathrm{E}\left[\frac{1}{h}\left(W_{p;+} + W_{p;-}\right)^2 \mathring{\epsilon}^2 \left(\frac{1}{\sqrt{\Delta_{\mathring{\epsilon}^2,2}}} - \frac{1}{\sqrt{\xi\left(|X|\right) f_{|X|}\left(|X|\right) \omega_p^{0,2}}}\right)^2\right] \\
&= \int_0^\infty \frac{1}{h}\mathcal{K}_{p;+}^2\left(\frac{z}{h}\right)\left(\sqrt{\frac{\xi\left(z\right) f_{|X|}\left(z\right)}{\Delta_{\mathring{\epsilon}^2,2}}} - \frac{1}{\sqrt{\omega_p^{0,2}}}\right)^2 \mathrm{d}z. \qquad \text{(S104)}
\end{aligned}$$

Note that $\Delta_{\mathring{\epsilon}^2,2} = \int_0^\infty h^{-1}\mathcal{K}_{p;+}^2\left(z/h\right)\xi\left(z\right) f_{|X|}\left(z\right)\mathrm{d}z$ and therefore, it follows from mean value expansion and (S104) that $\sigma_{\mathfrak{D}}^2 = O\left(\overline{h}^2\right)$. By Chen and Kato (2020, Corollary 5.5), $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\mathfrak{D}}\right] \lesssim \sigma_{\mathfrak{D}}\sqrt{\log\left(n\right)} + \log\left(n\right)\left\|F_{\mathfrak{D}}\right\|_{\mathbb{P}^T,r} n^{1/r}/\sqrt{n}$ and therefore, $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\mathfrak{D}}\right] = O\left(\sqrt{\log\left(n\right)}\cdot\overline{h} + \log\left(n\right)\left(\overline{n}^{1/r}/\underline{n}^{1/2}\right)\right)$. Let $\sigma_{\tilde{\mathfrak{Q}}}^2 \coloneqq \sup_{f\in\tilde{\mathfrak{Q}}}\mathbb{P}^T f^2$ and $\sigma_{\mathfrak{Q}}^2 \coloneqq \sup_{f\in\mathfrak{Q}}\mathbb{P}^T f^2$. It is easy to see that $\mathbb{P}^T f^2 = 1$, if $f \in \mathfrak{Q}$ or $f \in \tilde{\mathfrak{Q}}$ and therefore, $\sigma_{\tilde{\mathfrak{Q}}}^2 = \sigma_{\mathfrak{Q}}^2 = 1$. Similarly, $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\tilde{\mathfrak{Q}}}\right] \lesssim \sigma_{\tilde{\mathfrak{Q}}}\sqrt{\log\left(n\right)} + \log\left(n\right)\left\|F_{\tilde{\mathfrak{Q}}}\right\|_{\mathbb{P}^T,r} n^{1/r}/\sqrt{n}$ and a similar inequality with $\tilde{\mathfrak{Q}}$ replaced by $\mathfrak{Q}$ holds. Therefore, $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\tilde{\mathfrak{Q}}}\right] \asymp \mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\mathfrak{Q}}\right] = O\left(\sqrt{\log\left(n\right)}\right)$. Then it follows from Markov's inequality that $\left\{\mathbb{G}_n^T\tilde{q}\left(\cdot \mid h\right)\right\}^2 - \left\{\mathbb{G}_n^T q\left(\cdot \mid h\right)\right\}^2 = O_p\left(\log\left(n\right)\overline{h} + \log\left(n\right)^{3/2}\left(\overline{n}^{1/r}/\underline{n}^{1/2}\right)\right)$. The conclusion follows from this result and $LR_p\left(\vartheta \mid h\right) = \left\{\mathbb{G}_n^T\tilde{q}\left(\cdot \mid h\right)\right\}^2 + O_p\left(\log\left(n\right)\overline{h} + \log\left(n\right)^{3/2}\left(\overline{n}^{1/r}/\underline{n}^{1/2}\right)\right)$. ■

Proof of Theorem 5. Denote $Z_{\mathfrak{Q}_\pm} \coloneqq \sup_{f\in\mathfrak{Q}_\pm}\mathbb{G}_n^T f = \left\|\mathbb{G}_n^T\right\|_{\mathfrak{Q}}$. Since $F_{\mathfrak{Q}}$ is also an envelope of $\mathfrak{Q}_\pm \coloneqq \mathfrak{Q} \cup \left(-\mathfrak{Q}\right)$ ($-\mathfrak{Q} \coloneqq \left\{-f : f \in \mathfrak{Q}\right\}$) and the covering number of $\mathfrak{Q}_\pm$ is at most twice that of $\mathfrak{Q}$, $\mathfrak{Q}_\pm$ is also VC-type with respect to $F_{\mathfrak{Q}}$. By standard calculus calculations (see, e.g., the proof of Chernozhukov et al., 2014b, Corollary 5.1) and Chernozhukov et al. (2014b, Lemma 2.1), there exists a zero-mean Gaussian process $\left\{G^T\left(f\right) : f \in \mathfrak{Q}_\pm\right\}$ that is a tight random element in $\ell^\infty\left(\mathfrak{Q}_\pm\right)$ and also satisfies $\mathrm{E}\left[G^T\left(f\right)G^T\left(g\right)\right] = \mathrm{Cov}\left[f\left(T, X\right), g\left(T, X\right)\right]$, $\forall f, g \in \mathfrak{Q}_\pm$.[1] By Giné and Nickl (2015, Theorem 3.7.28), almost surely the sample paths $\mathfrak{Q}_\pm \ni f \mapsto G^T\left(f\right)$ are prelinear and therefore, almost surely, $\forall f \in \mathfrak{Q}$, $G^T\left(f\right) + G^T\left(-f\right) = 0$, and $\sup_{f\in\mathfrak{Q}_\pm} G^T\left(f\right) = \left\|G^T\right\|_{\mathfrak{Q}}$. Let $\bar{\Gamma}_G\left(h\right) \coloneqq G^T\left(q\left(\cdot \mid h\right)\right)$ and therefore, the zero-mean Gaussian process $\left\{\bar{\Gamma}_G\left(h\right) : h \in \mathbb{H}\right\}$ is a tight random element in $\ell^\infty\left(\mathbb{H}\right)$ and has the covariance structure $\mathrm{E}\left[\bar{\Gamma}_G\left(h\right)\bar{\Gamma}_G\left(h'\right)\right] = \mathrm{Cov}\left[q\left(T, X \mid h\right), q\left(T, X \mid h'\right)\right]$, $\forall \left(h, h'\right) \in \mathbb{H}^2$. By definition, $\left\|\bar{\Gamma}_G\right\|_{\mathbb{H}} = \left\|G^T\right\|_{\mathfrak{Q}}$. By change of variables and LIE, $\sup_{f\in\mathfrak{Q}}\mathbb{P}^T\left|f\right|^3 = \sup_{h\in\mathbb{H}}\mathrm{E}\left[\left|q\left(T, X \mid h\right)\right|^3\right] \lesssim \underline{h}^{-1/2}$ and similarly $\sup_{f\in\mathfrak{Q}}\mathbb{P}^T\left|f\right|^4 \lesssim \underline{h}^{-1}$. Also, $\mathbb{P}^T F_{\mathfrak{Q}}^r \lesssim \overline{h}/\underline{h}^{r/2}$. By Chernozhukov et al. (2016, Theorem 2.1) with $B\left(f\right) = 0$, $\mathcal{F} = \mathfrak{Q}_\pm$, $q = r$, $K_n = \log\left(n\right)$, $\sigma = 1$,

[1] Tightness of $\left\{G^T\left(f\right) : f \in \mathfrak{Q}_\pm\right\}$ is equivalent to the condition that $\mathfrak{Q}$ endowed with the intrinsic pseudo metric $\left(f, g\right) \mapsto \left\|f - g\right\|_{\mathbb{P}^T,2} \coloneqq \left(\mathbb{P}^T\left(f - g\right)^2\right)^{1/2}$ is totally bounded and almost surely the sample paths $f \mapsto G^T\left(f\right)$ are uniformly continuous with respect to the intrinsic pseudo metric. By Kosorok (2007, Lemmas 7.2 and 7.4), $\left\{G^T\left(f\right) : f \in \mathfrak{Q}_\pm\right\}$ is also separable as a stochastic process.

$b \lesssim \underline{h}^{-1/2}$ and $\gamma = \log(n)^{-1}$, there exists $\widetilde{Z}_{\mathfrak{Q}_\pm} =_d \sup_{f \in \mathfrak{Q}_\pm} G^T(f) = \left\|G^T\right\|_{\mathfrak{Q}}$ which satisfies $Z_{\mathfrak{Q}_\pm} - \widetilde{Z}_{\mathfrak{Q}_\pm} = O_p(\upsilon_n)$, where "$=_d$" is understood as being equal in distribution and $\upsilon_n := \left\{\log(n)(\log(n)\, n)^{1/r}\right\}/\underline{n}^{1/2} + \log(n)/\underline{n}^{1/6}$. By Dudley's entropy integral bound (Giné and Nickl, 2015, Theorem 2.3.7), Chen and Kato (2020, Lemma A.2) and standard calculus calculations (see, e.g., calculations in the proof of Chernozhukov et al., 2014b, Corollary 5.1),

$$\begin{aligned}\mathrm{E}\left[\left\|G^T\right\|_{\mathfrak{Q}}\right] &\lesssim \int_0^{\sigma_{\mathfrak{Q}} \vee n^{-1/2}\|F_{\mathfrak{Q}}\|_{\mathbb{P}^T,2}} \sqrt{1 + \log\left(N\left(\varepsilon, \mathfrak{Q}, \|\cdot\|_{\mathbb{P}^T,2}\right)\right)}\mathrm{d}\varepsilon \\ &\lesssim \left(\sigma_{\mathfrak{Q}} \vee n^{-1/2}\|F_{\mathfrak{Q}}\|_{\mathbb{P}^T,2}\right)\sqrt{\log(n)} = O\left(\sqrt{\log(n)}\right). \qquad \text{(S105)}\end{aligned}$$

By Lemma 11, $\sup_{h \in \mathbb{H}} LR_p(\vartheta \mid h) = \left\|\mathbb{G}_n^T\right\|_{\mathfrak{Q}}^2 + O_p\left(\log(n)\,\overline{h} + \log(n)^{3/2}\left(\overline{n}^{1/r}/\underline{n}^{1/2}\right)\right)$. By (S105) and the fact that $\mathrm{E}\left[\left\|\mathbb{G}_n^T\right\|_{\mathfrak{Q}}\right] = O\left(\sqrt{\log(n)}\right)$, we have $Z_{\mathfrak{Q}_\pm}^2 - \widetilde{Z}_{\mathfrak{Q}_\pm}^2 = O_p\left(\sqrt{\log(n)}\upsilon_n\right)$. Therefore, $\sup_{h \in \mathbb{H}} LR_p(\vartheta \mid h) = \widetilde{Z}_{\mathfrak{Q}_\pm}^2 + O_p\left(\sqrt{\log(n)}\upsilon_n + \log(n)\,\overline{h}\right)$. By Dudley (2002, Theorem 9.2.2) and $\sup_{h \in \mathbb{H}} LR_p(\vartheta \mid h) - \widetilde{Z}_{\mathfrak{Q}_\pm}^2 = o_p\left(\log(n)^{-1}\right)$, there exists a null sequence $\varepsilon_n \downarrow 0$ such that $\Pr\left[\left|\sup_{h \in \mathbb{H}} LR_p(\vartheta \mid h) - \widetilde{Z}_{\mathfrak{Q}_\pm}^2\right| > \varepsilon_n/\log(n)\right] \leq \varepsilon_n$ and by the fact that $(a-b)^2 \leq \left|a^2 - b^2\right|\ \forall a, b \geq 0$,

$$\Pr\left[\left|\sqrt{\sup_{h \in \mathbb{H}} LR_p(\vartheta \mid h)} - \widetilde{Z}_{\mathfrak{Q}_\pm}\right| > \sqrt{\varepsilon_n/\log(n)}\right] \leq \varepsilon_n. \qquad \text{(S106)}$$

It is easy to check that for random variables (V, W) and constants $r_1, r_2, t > 0$ such that $\Pr[|V - W| > r_1] \leq r_2$,

$$|\Pr[V \leq t] - \Pr[W \leq t]| \leq \Pr[|W - t| \leq r_1] + r_2. \qquad \text{(S107)}$$

Then, by (S106) and (S107),

$$\begin{aligned}&\left|\Pr\left[\sup_{h \in \mathbb{H}} LR_p(\vartheta \mid h) \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)^2\right] - \Pr\left[\widetilde{Z}_{\mathfrak{Q}_\pm}^2 \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)^2\right]\right| \\ &\qquad \leq \Pr\left[\left|\widetilde{Z}_{\mathfrak{Q}_\pm} - z_{1-\tau}\left(\overline{h}/\underline{h}\right)\right| \leq \sqrt{\varepsilon_n/\log(n)}\right] + \varepsilon_n. \qquad \text{(S108)}\end{aligned}$$

Since $\widetilde{Z}_{\mathfrak{Q}_\pm} =_d \left\|G^T\right\|_{\mathfrak{Q}}$ and $\left\{G^T(f) : f \in \mathfrak{Q}\right\}$ is a centered Gaussian process with $\mathrm{E}\left[G^T(f)^2\right] = 1$, $\forall f$, by using the Gaussian anti-concentration inequality (Chernozhukov et al., 2014a, Corollary 2.1) and (S105),

$$\Pr\left[\left|\widetilde{Z}_{\mathfrak{Q}_\pm} - z_{1-\tau}\left(\overline{h}/\underline{h}\right)\right| \leq \sqrt{\varepsilon_n/\log(n)}\right] \lesssim \sqrt{\varepsilon_n/\log(n)}\left(\mathrm{E}\left[\left\|G^T\right\|_{\mathfrak{Q}}\right] + 1\right) = O\left(\sqrt{\varepsilon_n}\right). \qquad \text{(S109)}$$

It then follows from (S108) and (S109) that

$$\Pr\left[LR_p(\vartheta \mid h) \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)^2, \forall h \in \mathbb{H}\right] = \Pr\left[\left\|\bar{\varGamma}_G\right\|_{\mathbb{H}} \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)\right] + o(1).$$

Let N be an $\mathrm{N}(0,1)$ random variable that is independent of $\left\{\bar{\mathit{\Gamma}}_G(h) : h \in \mathbb{H}\right\}$. Let $\tilde{\mathit{\Gamma}}_G(h) \coloneqq \bar{\mathit{\Gamma}}_G(h) + \mathrm{E}\left[q(T, X \mid h)\right] \cdot N$. By change of variables, $\sup_{h\in\mathbb{H}} |\mathrm{E}\left[q(T, X \mid h)\right]| = O\left(\overline{h}^{1/2}\right)$. $\left\{\tilde{\mathit{\Gamma}}_G(h) : h \in \mathbb{H}\right\}$ is a zero-mean Gaussian process which satisfies $\left\|\tilde{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} = \left\|\bar{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} + O_p\left(\overline{h}^{1/2}\right)$ and has the covariance structure $\mathrm{E}\left[\tilde{\mathit{\Gamma}}_G(h)\,\tilde{\mathit{\Gamma}}_G(h')\right] = \mathrm{E}\left[q(T, X \mid h)\, q(T, X \mid h')\right]$, $\forall (h, h') \in \mathbb{H}^2$. By LIE and change of variables, $\mathrm{E}\left[q(T, X \mid h)\, q(T, X \mid h')\right] = \sqrt{h/h'} \int_0^\infty \mathcal{K}_{p;+}(z)\, \mathcal{K}_{p;+}((h/h')\, z)\, \mathrm{d}z/\omega_p^{0,2}$. Let $\mathit{\Gamma}_G(s) \coloneqq \tilde{\mathit{\Gamma}}_G(s \cdot \underline{h})$, $s \in \left[1, \overline{h}/\underline{h}\right]$. Then it is easy to see that the zero-mean Gaussian process $\left\{\mathit{\Gamma}_G(s) : s \in \left[1, \overline{h}/\underline{h}\right]\right\}$ has a covariance structure given by (20) in the main text and $\|\mathit{\Gamma}_G\|_{[1,\overline{h}/\underline{h}]} = \left\|\tilde{\mathit{\Gamma}}_G\right\|_{\mathbb{H}}$. By Dudley (2002, Theorem 9.2.2) and $\left\|\tilde{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} - \left\|\bar{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} = o_p\left(\log(n)^{-1/2}\right)$, there exists a sequence $\tilde{\varepsilon}_n \downarrow 0$ such that $\Pr\left[\left|\left\|\tilde{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} - \left\|\bar{\mathit{\Gamma}}_G\right\|_{\mathbb{H}}\right| > \tilde{\varepsilon}_n/\sqrt{\log(n)}\right] \leq \tilde{\varepsilon}_n$. By similar arguments, we have $\Pr\left[\left\|\tilde{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)\right] - \Pr\left[\left\|\bar{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)\right] = o(1)$. By the definition of $z_{1-\tau}\left(\overline{h}/\underline{h}\right)$ and $\|\mathit{\Gamma}_G\|_{[1,\overline{h}/\underline{h}]} = \left\|\tilde{\mathit{\Gamma}}_G\right\|_{\mathbb{H}}$, $\Pr\left[\left\|\tilde{\mathit{\Gamma}}_G\right\|_{\mathbb{H}} \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)\right] = 1 - \tau$. It then follows that $\Pr\left[LR_p(\vartheta \mid h) \leq z_{1-\tau}\left(\overline{h}/\underline{h}\right)^2, \forall h \in \mathbb{H}\right] = 1 - \tau + o(1)$. ■

S6 Proof of Theorem 6

We write $\delta = O_p^\star(a_n)$ for some bounded sequence a_n if there exists some positive constants $c > 0$ such that $\Pr\left[|\delta| > c \cdot a_n\right] = O\left(\log(n)/(nh)^{3/2}\right)$. It is straightforward to check that if $\delta_1 = O_p^\star(a_n)$ and $\delta_2 = O_p^\star(b_n)$, then $\delta_1\delta_2 = O_p^\star(a_n b_n)$ and $\delta_1 + \delta_2 = O_p^\star(a_n + b_n)$, i.e., the algebra of the O_p notations carry over to $O_p^\star$ notations. We say that an event occurs wp* if its probability is $1 - O\left(\log(n)/(nh)^{3/2}\right)$.

Lemma 12. *Let V denote a random variable and $\{V_1, ..., V_n\}$ are i.i.d. copies of V. Assume that $nh \to \infty$. Suppose that K is a symmetric continuous PDF supported on $[-1, 1]$. Let $\mathbb{B}$ denote a neighborhood of 0. The following results hold for all $(\mathrm{s}, k) \in \{-, +\} \times \mathbb{N}$.*

(a) If $g_{|V|^3}$ is bounded on $\mathbb{B} \setminus \{0\}$, $(nh)^{-1} \sum_i \left|W_{p;\mathrm{s},i}^k V_i\right| = O_p^\star(1)$;

(b) If $g_{|V|^5}$ is bounded on $\mathbb{B} \setminus \{0\}$,

$$\frac{1}{n}\sum_i \frac{1}{h} W_{p;\mathrm{s},i}^k V_i - \mathrm{E}\left[\frac{1}{h} W_{p;\mathrm{s}}^k V\right] = O_p^\star\left(\sqrt{\frac{\log(n)}{nh}}\right);$$

(c) If if $g_{|V|^{3r}}$ is bounded on $\mathbb{B} \setminus \{0\}$, $\max_i \left|W_{p;\mathrm{s},i} V_i\right| = O_p^\star\left((nh)^{1/r}\right)$.

Proof of Lemma 12. For Part (a), let $c > 0$ be an arbitrary positive constant,

$$\Pr\left[\frac{1}{nh}\sum_i \left|W_{p;\mathrm{s},i}^k V_i\right| > \mathrm{E}\left[\frac{1}{h}\left|W_{p;\mathrm{s}}^k V\right|\right] + c\right] \leq \Pr\left[\left|\frac{1}{nh}\sum_i \left|W_{p;\mathrm{s},i}^k V_i\right| - \mathrm{E}\left[\frac{1}{h}\left|W_{p;\mathrm{s}}^k V\right|\right]\right|^3 > c^3\right]$$

$$\lesssim c^{-3}\left\{\left(\frac{1}{n}\cdot \mathrm{E}\left[\left(\frac{1}{h}\left|W_{p;\mathrm{s}}^{k}V\right|-\mathrm{E}\left[\frac{1}{h}\left|W_{p;\mathrm{s}}^{k}V\right|\right]\right)^{2}\right]\right)^{3/2}+\frac{1}{n^{2}}\cdot \mathrm{E}\left[\left|\frac{1}{h}\left|W_{p;\mathrm{s}}^{k}V\right|-\mathrm{E}\left[\frac{1}{h}\left|W_{p;\mathrm{s}}^{k}V\right|\right]\right|^{3}\right]\right\}$$
$$=O\left((nh)^{-3/2}\right),$$

where the second inequality follows from Markov's inequality and Rosenthal's inequality and the equality follows from change of variables and Loève's c_r inequality. The conclusion follows from the above results and $\mathrm{E}\left[h^{-1}\left|W_{p;\mathrm{s}}^{k}V\right|\right]=O(1)$, which follows from change of variables.

For Part (b), let $r_n \coloneqq \sqrt{(nh)/\log(n)}$, $\overline{V}_i \coloneqq V_i \mathbb{1}(|V_i|>r_n)$ and $\underline{V}_i \coloneqq V_i \mathbb{1}(|V_i|\leq r_n)$. Then we write

$$\begin{aligned}
\frac{1}{\sqrt{nh}}\sum_i\left(W_{p;\mathrm{s},i}^{k}V_i-\mathrm{E}\left[W_{p;\mathrm{s}}^{k}V\right]\right) &= \overline{W}+\underline{W}, \text{ where}\\
\overline{W} &\coloneqq \frac{1}{\sqrt{nh}}\sum_i\left(W_{p;\mathrm{s},i}^{k}\overline{V}_i-\mathrm{E}\left[W_{p;\mathrm{s}}^{k}\overline{V}\right]\right)\\
\underline{W} &\coloneqq \frac{1}{\sqrt{nh}}\sum_i\left(W_{p;\mathrm{s},i}^{k}\underline{V}_i-\mathrm{E}\left[W_{p;\mathrm{s}}^{k}\underline{V}\right]\right).
\end{aligned}$$

Let $\sigma_{\underline{W}}^{2} \coloneqq \mathrm{Var}\left[h^{-1/2}W_{p;\mathrm{s}}^{k}\underline{V}\right]$. By $\sigma_{\underline{W}}^{2}\leq \mathrm{E}\left[h^{-1}W_{p;\mathrm{s}}^{2k}\underline{V}^{2}\right]$, LIE and change of variables, $\sigma_{\underline{W}}^{2}=O(1)$. $\left|W_{p;\mathrm{s},i}^{k}\underline{V}_i-\mathrm{E}\left[W_{p;\mathrm{s}}^{k}\underline{V}\right]\right|$ is bounded by an upper bound that is proportional to r_n. Let $c_0>0$ denote an arbitrary positive constant. By Bernstein's inequality (Giné and Nickl, 2015, Theorem 3.1.7 with $u=\log(n^{c_0})$ and $c\lesssim r_n$), we have $\Pr\left[|\underline{W}|\geq\left(\sqrt{2c_0\sigma_{\underline{W}}^{2}}+c_0/3\right)\sqrt{\log(n)}\right]\leq 2n^{-c_0}$. By $\sigma_{\underline{W}}^{2}=O(1)$ and taking c_0 to be sufficiently large, $\underline{W}=O_p^{\star}\left(\sqrt{\log(n)}\right)$. By Markov's inequality, the fact that $\overline{V}^{2}\leq\overline{V}^{2}\left|V/r_n\right|^{3}$ and change of variables,

$$\Pr\left[\left|\overline{W}\right|\geq\sqrt{\log(n)}\right]\leq\frac{\mathrm{E}\left[h^{-1}W_{p;\mathrm{s}}^{2k}\overline{V}^{2}\right]}{\log(n)}\leq\frac{\mathrm{E}\left[h^{-1}W_{p;\mathrm{s}}^{2k}|V|^{5}\right]}{r_n^{3}\cdot\log(n)}=O\left(\frac{\sqrt{\log(n)}}{(nh)^{3/2}}\right)$$

and therefore, $\overline{W}=O_p^{\star}\left(\sqrt{\log(n)}\right)$.

Part (c) follows from $\max_i\left|W_{p;\mathrm{s},i}V_i\right|\leq\left((nh)^{-1}\sum_i\left|W_{p;\mathrm{s},i}V_i\right|^{r}\right)^{1/r}(nh)^{1/r}$ and Part (a). ■

We keep all notations defined in Section S2.1. We recycle some previous notation and let $\widetilde{\lambda}_p^{\mathsf{mc}}$ be redefined by

$$\widetilde{\lambda}_p^{\mathsf{mc}}\coloneqq\underset{\lambda}{\operatorname{argmax}}\sum_i\log\left(1+\lambda^{\top}\left(W_{p,i}V_i\right)\right).$$

Then we have

$$\begin{aligned}
LR_p(\vartheta\mid h) &= 2n\left(\ell_p^{\mathsf{mc}}(\vartheta\mid h)-\ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}}\mid h\right)\right)\\
&= 2\left\{\sum_i\log\left(1+\left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^{\top}\left(W_{p,i}V_i\right)\right)-\sum_i\log\left(1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^{\top}\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)\right\}.
\end{aligned}$$

The following result is an analogue of Lemma 4. Its proof essentially follows similar arguments.

Lemma 13. *Suppose that the assumptions in the statement of Theorem 6 hold. (a)* $\left\|\widehat{\lambda}_p^{\mathsf{mc}}\right\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$*; (b)* $\mathring{\vartheta}_p^{\mathsf{mc}} = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$*; (c)* $\left\|\widetilde{\lambda}_p^{\mathsf{mc}}\right\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$.

Proof of Lemma 13. It follows from Lemma 12(b,c) and Lemma 3(b) that $\|\Psi_{\bar{Z}}\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$, $\max_i \left\|W_{p,i}\bar{Z}_i\right\| = O_p^\star\left((nh)^{1/6}\right)$ and $\left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$. Therefore, $\operatorname{mineig}\left(\Delta_{\bar{Z}\bar{Z}^\top,2}\right) - \|\Psi_{\bar{Z}}\| \max_i \left\|W_{p,i}\bar{Z}_i\right\| - \left\|\Psi_{\bar{Z}\bar{Z}^\top,2} - \Delta_{\bar{Z}\bar{Z}^\top,2}\right\|$ is bounded away from zero wp*. By this result and (S34), $\left\|\lambda_p^{\mathsf{mc}}\right\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$ and Part (c) follows from similar arguments. Part (a) follows from $\left\|\lambda_p^{\mathsf{mc}}\right\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$ and the fact that $\widehat{\lambda}_p^{\mathsf{mc}} = \left(\mathrm{S}^\top \Delta_{UU^\top,2}^{1/2}\right)\left(0, \left(\lambda_p^{\mathsf{mc}}\right)^\top\right)^\top$.

By $\left\|\lambda_p^{\mathsf{mc}}\right\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$, (S39), the inequality in (S40) and Lemma 13(a,b,c), $\widehat{\vartheta}_Y^{\mathsf{mc}} = \Psi_Y - \Psi_{Y\bar{Z},2}^\top \lambda_p^{\mathsf{mc}} + O_p^\star\left(\log(n)/(nh)\right)$. By this result, $\left\|\lambda_p^{\mathsf{mc}}\right\| = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$, Lemma 3(b) and Lemma 12(b), $\widehat{\vartheta}_Y^{\mathsf{mc}} = \psi_{Y,\dagger} + O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$. Similarly, $\widehat{\vartheta}_D^{\mathsf{mc}} = \psi_{D,\dagger} + O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$. Part (b) follows from these results and the equality

$$\frac{\widehat{\vartheta}_Y^{\mathsf{mc}}}{\widehat{\vartheta}_D^{\mathsf{mc}}} - \frac{\psi_{Y,\dagger}}{\psi_{D,\dagger}} = \left(\frac{\widehat{\vartheta}_Y^{\mathsf{mc}}}{\psi_{D,\dagger}} - \frac{\psi_{Y,\dagger}}{\psi_{D,\dagger}}\right) + \frac{\widehat{\vartheta}_Y^{\mathsf{mc}}}{\psi_{D,\dagger}} \cdot \left(\frac{\psi_{D,\dagger}}{\widehat{\vartheta}_D^{\mathsf{mc}}} - 1\right).$$

■

Expanding the right hand sides of (S48) yields

$$\begin{aligned} 0 &= \sum_i W_{p,i}\widehat{V}_i^{\mathsf{mc}} \left\{1 - \left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right) + \left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2 \right. \\ &\quad \left. - \left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^3 + \frac{\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^4}{1 + \left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)}\right\} \\ 0 &= \sum_i \left(W_{p,i}H_i\right)^\top \widehat{\lambda}_p^{\mathsf{mc}} \left\{1 - \left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right) + \left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2 - \frac{\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^3}{1 + \left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)}\right\} \end{aligned} \tag{S119}$$

By Lemma 12(c), Lemma 13(b) and $\widehat{V}_i^{\mathsf{mc}} = V_i + H_i\mathring{\vartheta}_p^{\mathsf{mc}}$, $\max_i \left\|W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right\| = O_p^\star\left((nh)^{1/6}\right)$ and by this result, Lemma 13(a) and the Cauchy-Schwarz inequality, $\max_i \left|\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right| = O_p^\star\left(\sqrt{\log(n)}/(nh)^{1/3}\right)$. Therefore, $\max_i \left|\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top \left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right| < 1/2$ wp*. By $\widehat{V}_i^{\mathsf{mc}} = V_i + H_i\mathring{\vartheta}_p^{\mathsf{mc}}$, Lemma 12(a) and Lemma 13(b), we have $(nh)^{-1}\sum_i W_{p,i}^4 \left\|\widehat{V}_i^{\mathsf{mc}}\right\|^5 = O_p^\star(1)$ and $(nh)^{-1}\sum_i W_{p,i}^4 \left\|\widehat{V}_i^{\mathsf{mc}}\right\|^3 = O_p^\star(1)$. By these results and Lemma

13(a),

$$\begin{aligned}
\frac{1}{nh}\sum_i \frac{W_{p,i}\widehat{V}_i^{\mathsf{mc}}\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^4}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)} &= O_p^\star\left(\left(\frac{\log(n)}{nh}\right)^2\right)\\
\frac{1}{nh}\sum_i \frac{\left((W_{p,i}H_i)^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^3}{1+\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)} &= O_p^\star\left(\left(\frac{\log(n)}{nh}\right)^2\right).
\end{aligned}$$

By $\widehat{V}_i^{\mathsf{mc}} = V_i + H_i\mathring{\vartheta}_p^{\mathsf{mc}}$ and Lemma 12(a,b),

$$\begin{aligned}
\frac{1}{nh}\sum_i\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right) &= \sum_i W_{p,i}V_i^\top\widehat{\lambda}_p^{\mathsf{mc}} + \sum_i W_{p,i}\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\\
\frac{1}{nh}\sum_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2 &= \sum_i W_{p,i}^2\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2 + \sum_i W_{p,i}^2\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right)^2\\
&\quad +2\sum_i W_{p,i}^2\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right)\\
\frac{1}{nh}\sum_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^3 &= \sum_i W_{p,i}^3\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^3 + 3\sum_i W_{p,i}^3\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right)\\
&\quad +O_p^\star\left(\left(\frac{\log(n)}{nh}\right)^{5/2}\right).
\end{aligned}$$

By plugging these results into the right hand side of (S110),

$$\begin{aligned}
-\Delta_{VV^\top,2}\widehat{\lambda}_p^{\mathsf{mc}} + \Delta_H\mathring{\vartheta}_p^{\mathsf{mc}} &= -\frac{1}{nh}\sum_i W_{p,i}V_i + \frac{1}{nh}\sum_i W_{p,i}^2V_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right) - \frac{1}{nh}\sum_i W_{p,i}^3V_i\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2\\
&\quad -\frac{2}{nh}\sum_i W_{p,i}^3V_i\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right) + \frac{1}{nh}\sum_i W_{p,i}^3V_i\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^3\\
&\quad +\frac{1}{nh}\sum_i W_{p,i}^2H_i\mathring{\vartheta}_p^{\mathsf{mc}}\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right) + \frac{1}{nh}\sum_i W_{p,i}^2H_i\mathring{\vartheta}_p^{\mathsf{mc}}\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right)\\
&\quad -\frac{1}{nh}\sum_i W_{p,i}^3H_i\mathring{\vartheta}_p^{\mathsf{mc}}\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2 + \left(\Psi_{VV^\top,2}-\Delta_{VV^\top,2}\right)\widehat{\lambda}_p^{\mathsf{mc}} - \left(\Psi_H-\Delta_H\right)\mathring{\vartheta}_p^{\mathsf{mc}}\\
&\quad +O_p^\star\left(\left(\frac{\log(n)}{nh}\right)^2\right)\\
\Delta_H\widehat{\lambda}_p^{\mathsf{mc}} &= \frac{1}{nh}\sum_i W_{p,i}^2\left(H_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right) + \frac{1}{nh}\sum_i W_{p,i}^2\left(H_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right)\\
&\quad -\frac{1}{nh}\sum_i W_{p,i}^3\left(H_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2 - \left(\Psi_H-\Delta_H\right)^\top\widehat{\lambda}_p^{\mathsf{mc}} + O_p^\star\left(\left(\frac{\log(n)}{nh}\right)^2\right)
\end{aligned} \tag{S111}$$

A stochastic expansion is an approximation that is a polynomial of centered sample averages and has an approximation error of desired order of magnitude. We invert (S111) by using (S47) and get higher-order approximations for $\left(\widehat{\lambda}_p^{\mathsf{mc}}, \mathring{\vartheta}_p^{\mathsf{mc}}\right)$. Plugging these higher-order approximations back to the right hand side of (S111), replacing all sample averages except $(nh)^{-1}\sum_i W_{p,i}V_i$ which is approximately centered since $\|\Delta_V\| = O\left(h^{p+1}\right)$ with the sums of their population means and their centered versions and dropping terms that are $O_p^\star\left((\log(n)/nh)^2\right)$, we get a cubic stochastic expansion of $\left(\widehat{\lambda}_p^{\mathsf{mc}}, \mathring{\vartheta}_p^{\mathsf{mc}}\right)$. The same algebraic calculations have been done in Chen and Cui (2007) so that we use them directly here. Let $\alpha^{\mathsf{klmn}} := \Delta_{V^{(\mathsf{k})}V^{(\mathsf{l})}V^{(\mathsf{m})}V^{(\mathsf{n})}}$, $\gamma^{\mathsf{k}} := \Delta_{H^{(\mathsf{k})}}$, $\gamma^{\mathsf{kl:m}} := \Delta_{V^{(\mathsf{k})}V^{(\mathsf{l})}H^{(\mathsf{m})},3}$, $\gamma^{\mathsf{k;l}} := \Delta_{H^{(\mathsf{k})}H^{(\mathsf{l})},2}$, $A^{\mathsf{klm}} := \Psi_{V^{(\mathsf{k})}V^{(\mathsf{l})}V^{(\mathsf{m})},3} - \alpha^{\mathsf{klm}}$ and $C^{\mathsf{k:l}} := \Psi_{V^{(\mathsf{k})}H^{(\mathsf{l})},2} - \gamma^{\mathsf{k:l}}$. By Lemma 12(b), $\Psi_{V^{(\mathsf{k})}V^{(\mathsf{l})}V^{(\mathsf{m})}V^{(\mathsf{n})},4} - \alpha^{\mathsf{klmn}}$, $\Psi_{V^{(\mathsf{k})}V^{(\mathsf{l})}H^{(\mathsf{m})},3} - \gamma^{\mathsf{kl:m}}$ and $\Psi_{H^{(\mathsf{k})}H^{(\mathsf{l})},2} - \gamma^{\mathsf{k;l}}$ are all $O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$. We can show that $\left(\widehat{\lambda}_p^{\mathsf{mc}}, \mathring{\vartheta}_p^{\mathsf{mc}}\right)$ admit a cubic stochastic expansions with leading terms that are polynomials of $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, A^{\mathsf{klm}}, C^{\mathsf{k}}, C^{\mathsf{k:l}}\right)$ with coefficients given by $\left(\alpha^{\mathsf{kl}}, \alpha^{\mathsf{klm}}, \alpha^{\mathsf{klmn}}\right)$, $\left(\gamma^{\mathsf{k}}, \gamma^{\mathsf{k:l}}, \gamma^{\mathsf{kl:m}}, \gamma^{\mathsf{k;l}}\right)$ and Ω. Formally, their expressions are the same as those given in Chen and Cui (2007) (see Equation (2.6) therein).

By fifth-order Taylor expansion, $\widehat{V}_i^{\mathsf{mc}} = V_i + H_i\mathring{\vartheta}_p^{\mathsf{mc}}$, Lemma 12(a) and Lemma 13(a,b) and the fact that $\max_i\left|\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right| = O_p^\star\left(\sqrt{\log(n)}/(nh)^{1/3}\right)$, we have

$$
\begin{aligned}
2n\cdot\ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}} \mid h\right) \quad = \quad & 2\sum_i\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right) - \sum_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^2 + \frac{2}{3}\sum_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^3 \\
& -\frac{1}{2}\sum_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^4 + O_p^\star\left(\frac{(\log(n))^{5/2}}{(nh)^{3/2}}\right).
\end{aligned}
$$

By $\widehat{V}_i^{\mathsf{mc}} = V_i + H_i\mathring{\vartheta}_p^{\mathsf{mc}}$, Lemma 12(a) and Lemma 13(a,b),

$$
\sum_i\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top\left(W_{p,i}\widehat{V}_i^{\mathsf{mc}}\right)\right)^4 = \sum_i W_{p,i}^4\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^4 + O_p^\star\left(\frac{(\log(n))^{5/2}}{(nh)^{3/2}}\right).
$$

Therefore,

$$
\begin{aligned}
2n\cdot\ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}} \mid h\right) = & 2\sum_i W_{p,i}V_i^\top\widehat{\lambda}_p^{\mathsf{mc}} + 2\sum_i W_{p,i}\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right) - \sum_i W_{p,i}^2\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2 \\
& - \sum_i W_{p,i}^2\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right)^2 - 2\sum_i W_{p,i}^2\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right) + \frac{2}{3}\sum_i W_{p,i}^3\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^3 \\
& + 2\sum_i W_{p,i}^3\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^2\left(\left(\widehat{\lambda}_p^{\mathsf{mc}}\right)^\top H_i\mathring{\vartheta}_p^{\mathsf{mc}}\right) - \frac{1}{2}\sum_i W_{p,i}^4\left(V_i^\top\widehat{\lambda}_p^{\mathsf{mc}}\right)^4 + O_p^\star\left(\frac{(\log(n))^{5/2}}{(nh)^{3/2}}\right). \quad \text{(S112)}
\end{aligned}
$$

By the same steps and plugging stochastic expansions of $\left(\widehat{\lambda}_p^{\mathsf{mc}}, \mathring{\vartheta}_p^{\mathsf{mc}}\right)$ into the right hand side of (S112), we have a stochastic expansion of $2n\cdot\ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}} \mid h\right)$ so that $2n\cdot\ell_p^{\mathsf{mc}}\left(\widehat{\vartheta}_p^{\mathsf{mc}} \mid h\right) = \widehat{\ell} + O_p^\star\left(\log(n)^{5/2}/(nh)^{3/2}\right)$, where the leading term $\widehat{\ell}$ is a quartic polynomial of centered sample averages $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, A^{\mathsf{klm}}, C^{\mathsf{k}}, C^{\mathsf{k:l}}\right)$. The expression of $\widehat{\ell}$ is formally the same as those given in Chen and Cui (2007, Equation (2.8)).

Similarly, by expanding the right hand side of

$$0 = \sum_i \frac{W_{p,i}V_i}{1 + \left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top (W_{p,i}V_i)},$$

we get

$$0 = \frac{1}{nh}\sum_i W_{p,i}V_i \left\{ 1 - \left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top (W_{p,i}V_i) + \left(\left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top (W_{p,i}V_i)\right)^2 - \left(\left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top (W_{p,i}V_i)\right)^3 + \frac{\left(\left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top (W_{p,i}V_i)\right)^4}{1 + \left(\widetilde{\lambda}_p^{\mathsf{mc}}\right)^\top (W_{p,i}V_i)} \right\}.$$

Then by Lemma 12(a,c) and Lemma 13(c), the remainder on the right hand side of the above equation is $O_p^\star\left((\log(n)/(nh))^2\right)$. By using the same approach, we get a cubic stochastic expansion of $\widetilde{\lambda}_p^{\mathsf{mc}}$ and a quartic stochastic expansion of $2n\ell_p^{\mathsf{mc}}(\vartheta \mid h)$ so that $2n \cdot \ell_p^{\mathsf{mc}}(\vartheta \mid h) = \widetilde{\ell} + O_p^\star\left(\log(n)^{5/2}/(nh)^{3/2}\right)$. The stochastic expansions are polynomials of $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, A^{\mathsf{klm}}\right)$. The expression of $\widetilde{\ell}$ is formally the same as Chen and Cui (2007, Equation (2.5)). Let $LR \coloneqq \widetilde{\ell} - \widehat{\ell}$ so that $LR_p(\vartheta \mid h) = LR + O_p^\star\left(\log(n)^{5/2}/(nh)^{3/2}\right)$.

Lemma 14. *Suppose that the assumptions in the statement of Theorem 6 hold. Then,*

$$\Pr[LR \le x] = F_{\chi_1^2}(x) - \mathscr{C}_p^{\mathsf{pre}}(n,h)\, x f_{\chi_1^2}(x) + O\left(\upsilon_n^{\mathsf{pre}}\right),$$

where

$$\begin{aligned} \mathscr{C}_p^{\mathsf{pre}}(n,h) &\coloneqq (nh)\left(\alpha^1\right)^2 + (nh)^{-1}\left\{\frac{1}{2}\left(\alpha^{\mathsf{kkll}} - \alpha^{1+\mathsf{a}\,1+\mathsf{a}1+\mathsf{b}\,1+\mathsf{b}}\right) - \frac{1}{3}\left(\alpha^{\mathsf{klm}}\alpha^{\mathsf{klm}} - \alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}}\alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}}\right)\right\} \\ \upsilon_n^{\mathsf{pre}} &\coloneqq \frac{\log(n)\|\Delta_U\|}{\sqrt{nh}} + \frac{\log(n)^{5/2}}{(nh)^{3/2}} + h\|\Delta_U\| + n^{-1} + (nh)^2\|\Delta_U\|^4 + (nh)\|\Delta_U\|^3. \end{aligned}$$

Proof of Lemma 14. A decomposition $LR = (nh)\left(\tilde{R}_1^2 + 2\tilde{R}_1\tilde{R}_2 + 2\tilde{R}_1\tilde{R}_3 + \tilde{R}_2^2\right)$ can be derived. $\tilde{R}_k$ is a homogeneous k-th order polynomial of $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, A^{\mathsf{klm}}, C^{\mathsf{k}}, C^{\mathsf{k:l}}\right)$ so that $\tilde{R}_1 = O_p^\star\left(\sqrt{\log(n)/(nh)}\right)$, $\tilde{R}_2 = O_p^\star(\log(n)/(nh))$ and $\tilde{R}_3 = O_p^\star\left((\log(n)/(nh))^{3/2}\right)$. The expressions of $\left(\tilde{R}_1, \tilde{R}_2, \tilde{R}_3\right)$ are the same as those given in Chen and Cui (2007). Algebraic calculations in Chen and Cui (2007) show that by setting $\tilde{R}_1 \coloneqq A^1$,

$$\begin{aligned} \tilde{R}_2 \coloneqq\ & -\frac{1}{2}\cdot A^{11}A^1 - A^{1\,1+\mathsf{a}}A^{1+\mathsf{a}} + \frac{1}{3}\cdot\alpha^{111}A^1A^1 + \Omega\cdot C^{1+\mathsf{a}}A^{1+\mathsf{a}} \\ & + \alpha^{11\,1+\mathsf{a}}A^1A^{1+\mathsf{a}} + \left\{\alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} - \Omega\cdot\gamma^{1+\mathsf{a}:1+\mathsf{b}}\right\}A^{1+\mathsf{a}}A^{1+\mathsf{b}}, \end{aligned} \tag{S113}$$

and $\tilde{R}_3$ to be given by the formula in Chen and Cui (2007), we have $LR = nh\left(\tilde{R}_1^2 + 2\tilde{R}_1\tilde{R}_2 + 2\tilde{R}_1\tilde{R}_3 + \tilde{R}_2^2\right)$. In (S113), by (S46) and (S47), $\gamma^{1+\mathsf{a}:1+\mathsf{b}}A^{1+\mathsf{a}}A^{1+\mathsf{b}} = C^{1+\mathsf{a}}A^{1+\mathsf{a}} = 0$.

Let $\alpha^{\mathsf{k}} \coloneqq \Delta_{V^{(\mathsf{k})}}$ and $\mathring{A}^{\mathsf{k}} \coloneqq A^{\mathsf{k}} - \alpha^{\mathsf{k}}$. By replacing A^{k} with $\mathring{A}^{\mathsf{k}} + \alpha^{\mathsf{k}}$, we have $\tilde{R}_1 = \tilde{R}_{10} + \tilde{R}_{11}$, where $\tilde{R}_{10} \coloneqq \alpha^1$ and $\tilde{R}_{11} \coloneqq \mathring{A}^1$. Similarly, we replace A^{k} with $\mathring{A}^{\mathsf{k}} + \alpha^{\mathsf{k}}$ to decompose $\tilde{R}_2 = \tilde{R}_{22} + \tilde{R}_{21} + \tilde{R}_{20}$ so that $\tilde{R}_{2k}$ is a homogeneous $(2-k)$ −th order polynomial of $\alpha^1, ..., \alpha^{d_z+2}$:

$$\begin{aligned}\tilde{R}_{21} \quad \coloneqq \quad & -\frac{1}{2}\cdot A^{11}\alpha^1 - A^{1\,1+\mathsf{a}}\alpha^{1+\mathsf{a}} + \frac{2}{3}\cdot\alpha^{111}\mathring{A}^1\alpha^1 + \alpha^{1\,1\,1+\mathsf{a}}\left(\alpha^1\mathring{A}^{1+\mathsf{a}} + \alpha^{1+\mathsf{a}}\mathring{A}^1\right) \\ & + \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}}\left(\alpha^{1+\mathsf{a}}\mathring{A}^{1+\mathsf{b}} + \alpha^{1+\mathsf{b}}\mathring{A}^{1+\mathsf{a}}\right),\end{aligned} \tag{S114}$$

$\tilde{R}_{22}$ is defined by the right hand side of (S113) with A^{k} replaced by $\mathring{A}^{\mathsf{k}}$ and $\tilde{R}_{20} \coloneqq \tilde{R}_2 - \tilde{R}_{22} - \tilde{R}_{21} = O\left(\|\Delta_V\|^2\right)$. Let $R_0 \coloneqq \tilde{R}_{10} + \tilde{R}_{20}$, $R_1 \coloneqq \tilde{R}_{11} + \tilde{R}_{21}$ and $R_2 \coloneqq \tilde{R}_{22}$. We decompose $\tilde{R}_3 = \tilde{R}_{33} + \tilde{R}_{32} + \tilde{R}_{31} + \tilde{R}_{30}$ in a similar manner and let $R_3 \coloneqq \tilde{R}_{33}$. R_3 is given by the formula of $\tilde{R}_3$ with A^{k} replaced by $\mathring{A}^{\mathsf{k}}$. Then, let $R \coloneqq R_1 + R_2 + R_3$. By Lemma 12(b), $\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3 = R_0 + R + O_p^{\star}\left(\|\Delta_V\|\log(n)/(nh)\right)$ and therefore, $LR = (nh)(R_0 + R)^2 + O_p^{\star}\left(\upsilon_n^{\mathsf{pre}}\right)$.

Denote $\mathcal{K}_p(t) \coloneqq \mathbb{1}(t > 0)\mathcal{K}_{p;+}(t) - \mathbb{1}(t < 0)\mathcal{K}_{p;-}(t)$. Let $U_0 \coloneqq \left(M, Z^\top\right)^\top$, $l_1(U_0) \coloneqq \left(1, U_0^\top\right)^\top$, $l_2(U_0) \coloneqq \left(1, U_0^\top, \left(U_0^{\otimes 2}\right)^\top\right)^\top$, $l_3(U_0) \coloneqq \left(1, U_0^\top, \left(U_0^{\otimes 2}\right)^\top, \left(U_0^{\otimes 3}\right)^\top\right)^\top$ and

$$L \coloneqq \begin{pmatrix} \mathcal{K}_p\left(\frac{X}{h}\right) l_1(U_0) \\ \mathcal{K}_p\left(\frac{X}{h}\right) l_2(U_0) \\ \mathcal{K}_p^3\left(\frac{X}{h}\right) l_3(U_0) \end{pmatrix}.$$

Let d_l denote the dimension of L. It is clear that $\sqrt{nh}\cdot R \coloneqq h_n(\Upsilon)$, where $\Upsilon \coloneqq (nh)^{-1/2}\sum_i (L_i - \mathrm{E}[L])$ and h_n is a cubic polynomial. E.g.,

$$\begin{aligned}\sqrt{nh}\cdot\mathring{A}^1 \quad &= \quad \mathrm{e}_{d_z+2,1}^\top \mathrm{S}^\top \Delta_{UU^\top,2}^{-1/2}\left((nh)^{-1/2}\sum_i (W_{p,i}U_i - \mathrm{E}[W_pU])\right) \\ &= \quad -\frac{\Delta_G^\top \Delta_{UU^\top,2}^{-1}\left((nh)^{-1/2}\sum_i (W_{p,i}U_i - \mathrm{E}[W_pU])\right)}{\sqrt{\Delta_G^\top \Delta_{UU^\top,2}^{-1}\Delta_G}}.\end{aligned}$$

It can be shown that other terms on the right hand side of (S114) can also be written as linear functions of Υ. Similarly, $\sqrt{nh}\cdot R_2$ and $\sqrt{nh}\cdot R_3$ are homogenous quadratic and cubic polynomials of Υ.

Let $\kappa_j(V)$ denote the j-th cumulant of a random variable V. We follow arguments in the proof of Calonico et al. (2022, Theorem S.1) and apply Skovgaard (1986, Theorem 3.4) with $s = 4$ to $S_n \coloneqq \mathrm{B}^{-1/2}\Upsilon$ where $\mathrm{B} \coloneqq \mathrm{Var}[L]/h$. For any $t \in \mathbb{R}^{d_l}$ with $\|t\| = 1$, by change of variables and calculation of the moments (see, e.g., DiCiccio et al., 1988, Page 12), $\kappa_3\left(t^\top S_n\right) = \mathrm{E}\left[\left(t^\top S_n\right)^3\right] = O\left((nh)^{-1/2}\right)$, $\kappa_4\left(t^\top S_n\right) = \mathrm{E}\left[\left(t^\top S_n\right)^4\right] - 3\left(\mathrm{E}\left[\left(t^\top S_n\right)^2\right]\right)^2 = O\left((nh)^{-1}\right)$ and $\rho_{s,n}(t) := \max\left\{\left|\kappa_3\left(t^\top S_n\right)\right|/3!, \sqrt{\left|\kappa_4\left(t^\top S_n\right)\right|/4!}\right\} = O\left((nh)^{-1/2}\right)$, uniformly in t. Condition I and II of Skovgaard (1986, Theorem 3.4) are satisfied by taking $a_n(t) \propto \sqrt{nh}$ and $\epsilon_n = (nh)^{-3/2}$. Let $\Psi_V(t) := \mathrm{E}\left[\exp\left(\mathrm{i}t^\top V\right)\right]$ denote the characteristic function of a random vector V, where $\mathrm{i} := \sqrt{-1}$. Since $\mathcal{K}_p$ is supported on $[-1,1]$, by change of variables,

$$\begin{aligned}\Psi_L(t_1,t_2,t_3) &= 1 - \Pr[-h < X \le h]\\ &\quad + h\cdot\int_{-1}^{1}\int \exp\left(\mathrm{i}\left(t_1^\top \mathcal{K}_p(v)\, l_1(y) + t_2^\top \mathcal{K}_p^2(v)\, l_2(y) + t_3^\top \mathcal{K}_p^3(v)\, l_3(y)\right)\right) f_{U_0|X}(y \mid hv)\, f_X(hv)\,\mathrm{d}y\mathrm{d}v.\end{aligned}$$

Then, by triangle inequality and mean value expansion,

$$\begin{aligned}&\sup_{\|(t_1,t_2,t_3)\|>\varepsilon}\left|\Psi_L(t_1,t_2,t_3)\right| \le 1 - \Pr[-h < X \le h]\\ &+ h\cdot\varphi\left\{\sup_{\|(t_1,t_2,t_3)\|>\varepsilon}\left|\int_0^1\int \exp\left(\mathrm{i}\left(t_1^\top \mathcal{K}_p(v)\, l_1(y) + t_2^\top \mathcal{K}_p^2(v)\, l_2(y) + t_3^\top \mathcal{K}_p^3(v)\, l_3(y)\right)\right) f_{U_0|X}\left(y \mid 0^+\right)\mathrm{d}y\mathrm{d}v\right|\right.\\ &\left.\sup_{\|(t_1,t_2,t_3)\|>\varepsilon}\left|\int_{-1}^0\int \exp\left(\mathrm{i}\left(t_1^\top \mathcal{K}_p(v)\, l_1(y) + t_2^\top \mathcal{K}_p^2(v)\, l_2(y) + t_3^\top \mathcal{K}_p^3(v)\, l_3(y)\right)\right) f_{U_0|X}\left(y \mid 0^-\right)\mathrm{d}y\mathrm{d}v\right| + O(h)\right\},\end{aligned} \tag{S115}$$

where $f_{U_0|X}\left(y \mid 0^+\right) := \lim_{x\downarrow 0} f_{U_0|X}(y \mid x)$ and $f_{U_0|X}\left(y \mid 0^-\right) := \lim_{x\uparrow 0} f_{U_0|X}(y \mid x)$. Let $\mathcal{U}$ denote the support of U_0. Let $E_1 := \mathcal{K}_p(V)\, l_1(A)$, $E_2 := \mathcal{K}_p^2(V)\, l_2(A)$, and $E_3 := \mathcal{K}_p^3(V)\, l_3(A)$, where (V,A) has the joint density $(v,u) \mapsto 1(0 \le v \le 1)\, f_{U|X}\left(u \mid 0^+\right)$. Under Assumption 6, the functions $(v,u) \mapsto \left(1, \mathcal{K}_p(v)\, l_1^\top(u), \mathcal{K}_p^2(v)\, l_2^\top(u), \mathcal{K}_p^3(v)\, l_3^\top(u)\right)$ are linearly independent on $(0,1)\times\mathcal{U}$. By Bhattacharya (1977, Lemma 1.4), $\forall\varepsilon > 0$, $\exists c_\varepsilon > 0$ such that

$$\begin{aligned}&\sup_{\|(t_1,t_2,t_3)\|>\varepsilon}\left|\int_0^1\int \exp\left(\mathrm{i}\left(t_1^\top \mathcal{K}_p(v)\, l_1(y) + t_2^\top \mathcal{K}_p^2(v)\, l_2(y) + t_3^\top \mathcal{K}_p^3(v)\, l_3(y)\right)\right) f_{U_0|X}\left(y \mid 0^+\right)\mathrm{d}y\mathrm{d}v\right|\\ &\qquad = \sup_{\|(t_1,t_2,t_3)\|>\varepsilon}\left|\mathrm{E}\left[\exp\left(\mathrm{i}\left(t_1^\top E_1 + t_2^\top E_2 + t_3^\top E_3\right)\right)\right]\right| \le 1 - c_\varepsilon.\end{aligned}$$

A similar result holds for the term on the third line of (S115). Then by these results, $\Pr[-h < X \le h] = 2h(\varphi + O(h))$ and (S115), $\forall\varepsilon > 0$, $\exists c_\varepsilon > 0$ such that $\sup_{\|t\|>\varepsilon}\left|\Psi_L(t)\right| < 1 - c_\varepsilon h$, for all sufficiently small h. It follows from this result and arguments in the proof of Calonico et al. (2022, Theorem S.1) that $\forall\delta > 0$, $\exists c_\delta > 0$ such that $\sup_{\|t\|>\delta\sqrt{nh}}\left|\Psi_{S_n}(t)\right| \le (1 - c_\delta h)^n$ when n is sufficiently large. It is also easy to see that for $\forall\delta > 0$, $(1 - c_\delta h)^n \le \epsilon_n^{d_l/2+2}$, when n is sufficiently large. Therefore, Condition III''_α of Skovgaard (1986, Theorem 3.4 and Remark 3.5) is satisfied with $\alpha = 1$. Verification of Condition IV of Skovgaard (1986, Theorem 3.4) follows from essentially the same calculations and arguments in the proof of Calonico et al. (2022, Theorem S.1). Now all conditions for Skovgaard (1986, Theorem 3.4) are verified. It shows that S_n admits a valid Edgeworth expansion, i.e., conditions (3.1), (3.2) and (3.3) of Skovgaard (1981) are satisfied with $U_n = S_n$, $s = 4$, $\beta_{s,n} = (nh)^{-1}$ and the Edgeworth expansion holds uniformly over the class of all convex sets in $\mathbb{R}^{d_l}$. Note that we can write $\sqrt{nh} \cdot R = h_n\left(\mathrm{B}^{1/2} S_n\right)$. Then we apply Skovgaard (1981) to show that the Edgeworth expansion is preserved by smooth transformations. Condition (3.4) of Skovgaard (1981) is satisfied with g_n taken to be $x \mapsto h_n\left(\mathrm{B}^{1/2} x\right)$ whose the gradient at zero $\nabla g_n(0)$ is given by

$$\nabla g_n(0) = \mathrm{B}^{1/2}\left(-\frac{\Delta_G^\top \Delta_{UU^\top,2}^{-1}}{\sqrt{\Delta_G^\top \Delta_{UU^\top,2}^{-1}\Delta_G}}, 0_{d_l-(d_z+2)}^\top\right)^\top + O\left(\|\Delta_U\|\right)$$

by the chain rule. Then we apply Skovgaard (1981, Theorem 3.2) to $f_n(S_n) \coloneqq B_n^{-1} g_n(S_n)$, where $B_n^2 \coloneqq \nabla g_n(0)^\top \nabla g_n(0)$. Then,

$$B_n^2 = \frac{\Delta_G^\top \Delta_{UU^\top,2}^{-1}\left(\mathrm{Var}\left[W_p U\right]/h\right)\Delta_{UU^\top,2}^{-1}\Delta_G}{\Delta_G^\top \Delta_{UU^\top,2}^{-1}\Delta_G} + O\left(\|\Delta_U\|\right) = 1 + O\left(\|\Delta_U\|\right).$$

Condition I of Skovgaard (1981, Assumption 3.1) is satisfied with $p = 4$. Condition II of Skovgaard (1981, Assumption 3.1) is satisfied with $\lambda_n = O\left((nh)^{-1/2}\right)$ so that $\lambda_n^{p-1} = o\left((nh)^{-1}\right)$. Now all conditions for Skovgaard (1981, Theorem 3.2) are verified. It is left to compute the approximate cumulants.

Then we calculate the formal cumulants of $f_n(S_n) = B_n^{-1}\sqrt{nh} \cdot R$. In the calculations, we repeatedly use formulae for moments of products of sample averages (e.g., DiCiccio et al., 1988, Page 12) and Lemma 3(a,b). By definition, $\mathrm{E}[R_1] = 0$. We calculate $\mathrm{E}[R_2]$, let the remainder term absorb the terms that involve $\alpha^1, \ldots, \alpha^{d_z+2}$ and get $\mathrm{E}[R_2] = (nh)^{-1}\bar{\kappa}_1 + O\left(\|\Delta_U\|/n\right)$ where $\bar{\kappa}_1 \coloneqq -\alpha^{111}/6$. By formulae for third moments and Lemma 3(a), $\mathrm{E}[R_3] = O\left((nh)^{-2}\right)$. Therefore, $\kappa_1\left(\sqrt{nh} \cdot R\right) = \tilde{\kappa}_{1,n} + O\left((nh)^{-1/2}\|\Delta_U\| h + (nh)^{-3/2}\right)$ with $\tilde{\kappa}_{1,n} \coloneqq (nh)^{-1/2}\bar{\kappa}_1$. For the second cumulant, by definition, $\kappa_2(R) = \mathrm{E}\left[R^2\right] - (\mathrm{E}[R])^2$ and by formulae for fifth and sixth moments and Lemma 3(a),

$$\mathrm{E}\left[R^2\right] = \mathrm{E}\left[R_1^2\right] + 2 \cdot \mathrm{E}\left[R_1 R_2\right] + 2 \cdot \mathrm{E}\left[R_1 R_3\right] + \mathrm{E}\left[R_2^2\right] + O\left((nh)^{-3}\right).$$

By $R_1 = \tilde{R}_{11} + \tilde{R}_{21}$ and calculation,

$$\begin{aligned}
\mathrm{E}\left[R_1^2\right] &= \mathrm{E}\left[\tilde{R}_{11}^2\right] + 2 \cdot \mathrm{E}\left[\tilde{R}_{21}\tilde{R}_{11}\right] + O\left((nh)^{-1}\|\Delta_U\|^2\right) \\
\mathrm{E}\left[R_1 R_2\right] + \mathrm{E}\left[R_1 R_3\right] &= \mathrm{E}\left[\tilde{R}_{11} R_2\right] + \mathrm{E}\left[\tilde{R}_{11} R_3\right] + O\left((nh)^{-2}\|\Delta_U\|\right).
\end{aligned}$$

Then by calculation, $\mathrm{E}\left[\tilde{R}_{11}^2\right] = (nh)^{-1} + O\left(\|\Delta_U\|^2/n\right)$ and $2 \cdot \mathrm{E}\left[\tilde{R}_{21}\tilde{R}_{11}\right] = (nh)^{-1}\tilde{\kappa}_{21,n} + O\left(\|\Delta_U\|^2/n\right)$, where $\tilde{\kappa}_{21,n} := \alpha^{111}\alpha^1/3$. Then, $\mathrm{E}\left[R_1^2\right] = (nh)^{-1}(1+\tilde{\kappa}_{21,n}) + O\left((nh)^{-1}\|\Delta_U\|^2\right)$. Calculation of $2 \cdot \mathrm{E}\left[\tilde{R}_{11}R_2\right] + 2 \cdot \mathrm{E}\left[\tilde{R}_{11}R_3\right] + \mathrm{E}\left[R_2^2\right]$ follows from replication of calculations in Chen and Cui (2007) and we can directly use the results therein. By calculations in Chen and Cui (2007), we have

$$2 \cdot \mathrm{E}\left[\tilde{R}_{11}R_2\right] + 2 \cdot \mathrm{E}\left[\tilde{R}_{11}R_3\right] + \mathrm{E}\left[R_2^2\right] = (nh)^{-2}\bar{\kappa}_2 + O\left((nh)^{-2}\|\Delta_U\|h + (nh)^{-3}\right),$$

where

$$\begin{aligned} \bar{\kappa}_2 &:= \frac{1}{2}\alpha^{1111} + \alpha^{11\,1+\mathsf{a}\,1+\mathsf{a}} - \frac{1}{3}\alpha^{111}\alpha^{111} - \alpha^{11\,1+\mathsf{a}}\alpha^{11\,1+\mathsf{a}} - \alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}}\alpha^{1\,1+\mathsf{a}\,1+\mathsf{b}} \\ &= \frac{1}{2}\left(\alpha^{\mathsf{kkll}} - \alpha^{1+\mathsf{a}\,1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{b}}\right) - \frac{1}{3}\left(\alpha^{\mathsf{klm}}\alpha^{\mathsf{klm}} - \alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}}\alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}}\right) \end{aligned}$$

and the $O\left((nh)^{-2}\|\Delta_U\|h + (nh)^{-3}\right)$ remainder collects terms that depend on $\alpha^1, ..., \alpha^{d_z+2}$ and higher-order terms from the fourth moment calculation. Therefore,

$$\kappa_2\left(\sqrt{nh} \cdot R\right) = \tilde{\kappa}_{2,n} + O\left(\|\Delta_U\|^2 + (nh)^{-1}\|\Delta_U\| + (nh)^{-2}\right),$$

where $\tilde{\kappa}_{2,n} := 1 + \tilde{\kappa}_{21,n} + \tilde{\kappa}_{22,n}$ and $\tilde{\kappa}_{22,n} := (nh)^{-1}\left(\bar{\kappa}_2 - \bar{\kappa}_1^2\right)$.

By definition, $\kappa_3(R) = \mathrm{E}\left[R^3\right] - 3 \cdot \mathrm{E}[R]\,\mathrm{E}\left[R^2\right] + 2(\mathrm{E}[R])^3$ and by $\mathrm{E}[R] = \mathrm{E}[R_2] + O\left((nh)^{-2}\right)$, $\mathrm{E}[R_2] = O\left((nh)^{-1}\right)$, $\mathrm{E}\left[R^2\right] = \mathrm{E}\left[R_1^2\right] + O\left((nh)^{-2}\right)$ and $\mathrm{E}\left[R^3\right] = \mathrm{E}\left[R_1^3\right] + 3 \cdot \mathrm{E}\left[R_2R_1^2\right] + O\left((nh)^{-3}\right)$, which follows from formulae for higher moments, we have

$$\kappa_3(R) = \mathrm{E}\left[R_1^3\right] - 3\left(\mathrm{E}\left[R_2R_1^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[R_1^2\right]\right) + O\left((nh)^{-3}\right).$$

It is easy to check that $\mathrm{E}\left[R_1^3\right] = \mathrm{E}\left[\tilde{R}_{11}^3\right] + O\left((nh)^{-2}\|\Delta_U\|\right)$, $\mathrm{E}\left[R_2R_1^2\right] = \mathrm{E}\left[R_2\tilde{R}_{11}^2\right] + O\left((nh)^{-2}\|\Delta_U\|\right)$. By these results and $\mathrm{E}\left[R_1^2\right] = \mathrm{E}\left[\tilde{R}_{11}^2\right] + O\left((nh)^{-1}\|\Delta_U\|\right)$,

$$\kappa_3(R) = \mathrm{E}\left[\tilde{R}_{11}^3\right] - 3\left(\mathrm{E}\left[R_2\tilde{R}_{11}^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]\right) + O\left((nh)^{-3} + (nh)^{-2}\|\Delta_U\|\right).$$

Calculation and expansion of $\mathrm{E}\left[\tilde{R}_{11}^3\right] - 3\left(\mathrm{E}\left[R_2\tilde{R}_{11}^2\right] - \mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)$ follow from replication of calculations in Chen and Cui (2007). E.g., by calculation using formulae for moments (DiCiccio et al., 1988),

$$\mathrm{E}\left[\tilde{R}_{11}^3\right] = n^{-2}\left(\mathrm{E}\left[\left(h^{-1}W_pV^{(1)} - \alpha^1\right)^3\right]\right) = n^{-2}\mathrm{E}\left[\left(h^{-1}W_pV^{(1)}\right)^3\right] + O\left((nh)^{-2}\|\Delta_U\|h\right),$$

and the $O\left((nh)^{-2}\|\Delta_U\| h\right)$ remainder collects all terms in the expansion of the third moment which depend on $\alpha^1,...,\alpha^{d_z+2}$. Similarly, we calculate $\mathrm{E}\left[R_2\tilde{R}_{11}^2\right]-\mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]$. We note that coefficients of terms of order $(nh)^{-2}$ in $\mathrm{E}\left[\tilde{R}_{11}^3\right]-3\left(\mathrm{E}\left[R_2\tilde{R}_{11}^2\right]-\mathrm{E}[R_2]\,\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)$ are formally the same as those of the leading terms in the calculation of the formal third cumulant in Chen and Cui (2007). Calculations in Chen and Cui (2007) show that the sum of these coefficients are exactly zero and therefore, the leading term vanishes so that

$$\kappa_3\left(\sqrt{nh}\cdot R\right)=O\left(\frac{\|\Delta_U\|}{\sqrt{nh}}+(nh)^{-3/2}\right). \tag{S116}$$

By (S116), the fact that

$$\kappa_4(R)=\mathrm{E}\left[R^4\right]-3\left(\mathrm{E}\left[R^2\right]\right)^2-4\cdot\mathrm{E}[R]\,\kappa_3(R)+2\left(\mathrm{E}[R]\right)^4,$$

$\mathrm{E}[R]=O\left((nh)^{-1}\right)$, $R=\tilde{R}_{11}+\tilde{R}_{21}+R_2+R_3$ and standard calculations,

$$\begin{aligned}\kappa_4(R)=\mathrm{E}\left[R^4\right]-3\left(\mathrm{E}\left[R^2\right]\right)^2+O\left((nh)^{-3}\|\Delta_U\|+(nh)^{-4}\right)=\left\{\mathrm{E}\left[\tilde{R}_{11}^4\right]-3\left(\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)^2\right\}\\ +4\left\{\mathrm{E}\left[R_2\tilde{R}_{11}^3\right]-3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]\right\}+6\left\{\mathrm{E}\left[R_2^2\tilde{R}_{11}^2\right]-\mathrm{E}\left[R_2^2\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]\right\}\\ +4\left\{\mathrm{E}\left[R_2\tilde{R}_{11}^3\right]-3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]\right\}+O\left((nh)^{-3}\|\Delta_U\|+(nh)^{-4}\right).\end{aligned} \tag{S117}$$

And by standard calculations,

$$\begin{aligned}\mathrm{E}\left[\tilde{R}_{11}^4\right]-3\left(\mathrm{E}\left[\tilde{R}_{11}^2\right]\right)^2&=n^{-3}\left(\mathrm{E}\left[\left(h^{-1}W_pV^{(1)}-\alpha^1\right)^4\right]-3\left(\mathrm{E}\left[\left(h^{-1}W_pV^{(1)}-\alpha^1\right)^2\right]\right)^2\right)\\ &=n^{-3}\left(\mathrm{E}\left[\left(h^{-1}W_pV^{(1)}\right)^4\right]-3\left(\mathrm{E}\left[\left(h^{-1}W_pV^{(1)}\right)^2\right]\right)\right)+O\left((nh)^{-3}\|\Delta_U\| h\right),\end{aligned}$$

and the $O\left((nh)^{-3}\|\Delta_U\| h\right)$ remainder collects all terms that depend on $\alpha^1,...,\alpha^{d_z+2}$. Similarly, we also calculate $\mathrm{E}\left[R_2\tilde{R}_{11}^3\right]-3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]$, $\mathrm{E}\left[R_2^2\tilde{R}_{11}^2\right]-\mathrm{E}\left[R_2^2\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]$ and $\mathrm{E}\left[R_2\tilde{R}_{11}^3\right]-3\cdot\mathrm{E}\left[R_2\tilde{R}_{11}\right]\mathrm{E}\left[\tilde{R}_{11}^2\right]$ on the right hand side of the second equality in (S117), ignore small-order terms that depend on $\alpha^1,...,\alpha^{d_z+2}$ and take the sum of the leading terms. We do not need to rework on the calculations since they are formally the same as those done in Chen and Cui (2007). Calculations in Chen and Cui (2007) show that the sum of the leading terms on the right hand side of (S117) is exactly zero so that it follows from this result and (S117) that $\kappa_4\left(\sqrt{nh}\cdot R\right)=O\left((nh)^{-1}\|\Delta_U\|+(nh)^{-2}\right)$. By previous calculations and $B_n=1+O\left(\|\Delta_U\|\right)$, we get the approximate cumulants for $f_n(S_n)$: $\kappa_1(f_n(S_n))=B_n^{-1}\tilde{\kappa}_{1,n}+O\left((nh)^{-1/2}\|\Delta_U\| h+(nh)^{-3/2}\right)$, $\kappa_2(f_n(S_n))=B_n^{-2}\tilde{\kappa}_{2,n}+O\left(\|\Delta_U\|^2+(nh)^{-1}\|\Delta_U\|+(nh)^{-2}\right)$, $\kappa_3(f_n(S_n))=O\left(\|\Delta_U\|/\sqrt{nh}+(nh)^{-3/2}\right)$

and $\kappa_4(f_n(S_n)) = O\left((nh)^{-1}\|\Delta_U\| + (nh)^{-2}\right)$.

Let $\phi(\cdot \mid \mu, \sigma^2)$ denote the PDF of $\mathrm{N}(\mu, \sigma^2)$. By applying Skovgaard (1981, Theorem 3.2) to $f_n(S_n) = B_n^{-1}\sqrt{nh} \cdot R$,

$$\Pr\left[(nh)(R_0 + R)^2 \leq x\right] = \int_{\left|t+\left(\sqrt{nh}\cdot R_0\right)/B_n\right| \leq \sqrt{x}/B_n} \phi\left(t \mid B_n^{-1}\tilde{\kappa}_{1,n}, B_n^{-2}\tilde{\kappa}_{2,n}\right) \mathrm{d}t + O\left(\frac{\|\Delta_U\|}{\sqrt{nh}} + (nh)^{-3/2}\right), \tag{S118}$$

uniformly in $x > 0$. By using the recurrence properties of non-central χ^2 (Cohen, 1988) and mean value expansion, we have $(\partial/\partial\lambda) F(x \mid \lambda)|_{\lambda=\bar{\lambda}} = -x f_{\chi_1^2}(x) + O(\bar{\lambda})$. By this result, $B_n^2 = 1 + O(\|\Delta_U\|)$, change of variables and mean value expansion,

$$\begin{aligned}
&\int_{\left|t+\left(\sqrt{nh}\cdot R_0\right)/B_n\right| \leq \sqrt{x}/B_n} \phi\left(t \mid B_n^{-1}\tilde{\kappa}_{1,n}, B_n^{-2}\tilde{\kappa}_{2,n}\right) \mathrm{d}t = \int_{|t| \leq \sqrt{x/\tilde{\kappa}_{2,n}}} \phi\left(t \mid \frac{\sqrt{nh}\cdot R_0 + \tilde{\kappa}_{1,n}}{\sqrt{\tilde{\kappa}_{2,n}}}, 1\right) \mathrm{d}t \\
&= F\left(\frac{x}{\tilde{\kappa}_{2,n}} \mid \frac{\left(\sqrt{nh}\cdot R_0 + \tilde{\kappa}_{1,n}\right)^2}{\tilde{\kappa}_{2,n}}\right) = F_{\chi_1^2}(x) - x f_{\chi_1^2}(x)\left(\left(\sqrt{nh}\cdot \tilde{R}_{10} + \tilde{\kappa}_{1,n}\right)^2 + \tilde{\kappa}_{21,n} + \tilde{\kappa}_{22,n}\right) + O\left(\upsilon_n^{\mathsf{pre}}\right).
\end{aligned} \tag{S119}$$

By (S118) and (S119),

$$\Pr\left[(nh)(R_0 + R)^2 \leq x\right] = F_{\chi_1^2}(x) - \left\{(nh)\tilde{R}_{10}^2 + 2\sqrt{nh}\cdot\tilde{R}_{10}\tilde{\kappa}_{1,n} + \tilde{\kappa}_{21,n} + (nh)^{-1}\bar{\kappa}_2\right\} x f_{\chi_1^2}(x) + O\left(\upsilon_n^{\mathsf{pre}}\right). \tag{S120}$$

It is easy to see that $2\sqrt{nh}\cdot\tilde{R}_{10}\tilde{\kappa}_{1,n} + \tilde{\kappa}_{21,n} = 0$.

It is easily seen that the result (S118) with the weak inequality replaced by a strict inequality still holds (see Skovgaard, 1981, Theorem 3.2). By $LR = (nh)(R_0 + R)^2 + O_p^\star(\upsilon_n^{\mathsf{pre}})$ and the fact (S107),

$$\left|\Pr[LR \leq x] - \Pr\left[(nh)(R_0 + R)^2 \leq x\right]\right| \leq \Pr\left[\left|(nh)(R_0 + R)^2 - x\right| \leq c_1\upsilon_n^{\mathsf{pre}}\right] + c_2\left(\frac{\log(n)}{(nh)^{3/2}}\right) = O\left(\upsilon_n^{\mathsf{pre}}\right), \tag{S121}$$

where the equality follows from (S118) and boundedness of $\phi(\cdot \mid \tilde{\kappa}_{1,n}, \tilde{\kappa}_{2,n})$. The conclusion follows from (S120) and (S121). ■

Proof of Theorem 6. By (S46) and (S47),

$$\begin{aligned}
\alpha^{\mathsf{kkll}} &= \left(\Delta_{UU^\top,2}^{-1}\right)^{(\mathsf{kl})}\left(\Delta_{UU^\top,2}^{-1}\right)^{(\mathsf{mn})}\Delta_{U^{(\mathsf{k})}U^{(\mathsf{l})}U^{(\mathsf{m})}U^{(\mathsf{n})},4} \\
&= \left(\Delta_{UU^\top,2}^{-1}\right)^{(\mathsf{kl})}\mathrm{tr}\left(\Delta_{UU^\top,2}^{-1}\Delta_{U^{(\mathsf{k})}U^{(\mathsf{l})}UU^\top,4}\right)
\end{aligned}$$

$$\begin{aligned}
\alpha^{1+\mathsf{a}\,1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{b}} &= \left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\right)^{(\mathsf{ab})}\left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\right)^{(\mathsf{cd})}\Delta_{\bar{Z}^{(\mathsf{a})}\bar{Z}^{(\mathsf{b})}\bar{Z}^{(\mathsf{c})}\bar{Z}^{(\mathsf{d})},4}\\
&= \left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\right)^{(\mathsf{ab})}\operatorname{tr}\left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\Delta_{\bar{Z}^{(\mathsf{a})}\bar{Z}^{(\mathsf{b})}\bar{Z}\bar{Z}^\top,4}\right)
\end{aligned}$$

and

$$\begin{aligned}
\alpha^{\mathsf{klm}}\alpha^{\mathsf{klm}} &= \Delta_{U^{(\mathsf{k})}U^{(\mathsf{l})}U^{(\mathsf{m})},3}\left(\Delta^{-1}_{UU^\top,2}\right)^{(\mathsf{kk}')}\left(\Delta^{-1}_{UU^\top,2}\right)^{(\mathsf{ll}')}\left(\Delta^{-1}_{UU^\top,2}\right)^{(\mathsf{mm}')}\Delta_{U^{(\mathsf{k}')}U^{(\mathsf{l}')}U^{(\mathsf{m}')},3}\\
&= \left(\Delta^{-1}_{UU^\top,2}\right)^{(\mathsf{kl})}\operatorname{tr}\left(\Delta^{-1}_{UU^\top,2}\Delta_{U^{(\mathsf{k})}UU^\top,3}\Delta^{-1}_{UU^\top,2}\Delta_{U^{(\mathsf{l})}UU^\top,3}\right)\\
\alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}}\alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}} &= \Delta_{\bar{Z}^{(\mathsf{a})}\bar{Z}^{(\mathsf{b})}\bar{Z}^{(\mathsf{c})},3}\left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\right)^{(\mathsf{aa}')}\left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\right)^{(\mathsf{bb}')}\left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\right)^{(\mathsf{cc}')}\Delta_{\bar{Z}^{(\mathsf{a}')}\bar{Z}^{(\mathsf{b}')}\bar{Z}^{(\mathsf{c}')},3}\\
&\quad \left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\right)^{(\mathsf{ab})}\operatorname{tr}\left(\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\Delta_{\bar{Z}^{(\mathsf{a})}\bar{Z}\bar{Z}^\top,3}\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\Delta_{\bar{Z}^{(\mathsf{b})}\bar{Z}\bar{Z}^\top,3}\right).
\end{aligned}$$

It is easy to see that in Lemma 3(a), if g_V is Lipschitz continuous, the remainder term is $O(h)$. By this result, $\Delta_{U^{(\mathsf{k})}U^{(\mathsf{l})}UU^\top,4} = \omega_p^{0,4}\psi_{U^{(\mathsf{k})}U^{(\mathsf{l})}UU^\top,\pm} + O(h)$, $\Delta_{U^{(\mathsf{k})}UU^\top,3} = \omega_p^{0,3}\psi_{U^{(\mathsf{k})}UU^\top,\dagger} + O(h)$ and $\Delta^{-1}_{UU^\top,2} = \left(\omega_p^{0,2}\psi_{UU^\top,\pm}\right)^{-1} + O(h)$. Similarly, $\Delta_{\bar{Z}^{(\mathsf{a})}\bar{Z}^{(\mathsf{b})}\bar{Z}\bar{Z}^\top,4} = \omega_p^{0,4}\psi_{\bar{Z}^{(\mathsf{a})}\bar{Z}^{(\mathsf{b})}\bar{Z}\bar{Z}^\top,\pm} + O(h)$, $\Delta_{\bar{Z}^{(\mathsf{a})}\bar{Z}\bar{Z}^\top,3} = \omega_p^{0,3}\psi_{\bar{Z}^{(\mathsf{a})}\bar{Z}\bar{Z}^\top,\dagger} + O(h)$ and $\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2} = \left(\omega_p^{0,2}\psi_{\bar{Z}\bar{Z}^\top,\pm}\right)^{-1} + O(h)$. By these results, we have

$$\begin{aligned}
\frac{1}{2}\alpha^{\mathsf{kkll}} - \frac{1}{3}\alpha^{\mathsf{klm}}\alpha^{\mathsf{klm}} &= \mathscr{V}_p^{\dagger} + O(h)\\
\frac{1}{2}\alpha^{1+\mathsf{a}\,1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{b}} - \frac{1}{3}\alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}}\alpha^{1+\mathsf{a}\,1+\mathsf{b}\,1+\mathsf{c}} &= \mathscr{V}_p^{\ddagger} + O(h).
\end{aligned} \tag{S122}$$

By $\bar{\gamma}_M = \gamma_M + O(h)$ which follows from Lemma 3(a) with Lipschitz continuity, (S52) and (S54),

$$\left(\alpha^1\right)^2 = \frac{\left(\Delta_G\Delta^{-1}_{UU^\top,2}\Delta_U\right)^2}{\Delta_G^\top\Delta^{-1}_{UU^\top,2}\Delta_G} = \frac{\left(\Delta_M - \Delta_{M\bar{Z}^\top,2}\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\Delta_{\bar{Z}}\right)^2}{\Delta_{M^2,2} - \Delta_{M\bar{Z}^\top,2}\Delta^{-1}_{\bar{Z}\bar{Z}^\top,2}\Delta_{M\bar{Z},2}} = \frac{\Delta^2_{\mathring{\epsilon}}}{\Delta_{\mathring{\epsilon}^2,2}} + O\left(h^{2p+3}\right).$$

It is clear that in Lemma 3(b), the remainder term is $O\left(h^{p+1+\mathfrak{h}}\right)$ if $m_V^{(p+1)}$ is Hölder continuous with exponent $\mathfrak{h}$. Then, by this result, $\Delta_{\mathring{\epsilon}} = \psi_{D,\dagger}\mathscr{B}_p^{\mathsf{mc}}h^{p+1} + O\left(h^{p+1+\mathfrak{h}}\right)$ and $\Delta_{\mathring{\epsilon}^2,2} = \psi^2_{D,\dagger}\mathscr{V}_p + O(h)$. Then, by this result and Lemma 3(a,b),

$$(nh)\left(\alpha^1\right)^2 = (nh)\cdot\frac{\left(\mathscr{B}_p^{\mathsf{mc}}h^{p+1}\right)^2}{\mathscr{V}_p} + O\left(nh^{2p+3+\mathfrak{h}}\right).$$

The first conclusion follows from this result, (S122) and Lemma 14.

It follows from the same arguments used to show Lemma 3(b) that $\mathrm{E}\left[h^{-1}W_{p+1;\mathrm{s}}V\right] = \psi_{V,\mathrm{s}} + O\left(h^{p+1+\mathfrak{h}}\right)$ if $m_V^{(p+1)}$ is Hölder continuous with exponent $\mathfrak{h}$. Then, the second conclusion follows from Lemma 14 with p replaced by $p+1$ and similar arguments. In this case, we have $\left\|\mathrm{E}\left[h^{-1}W_{p+1}U\right]\right\| = O\left(h^{p+1+\mathfrak{h}}\right)$ and the bias part is now $O\left(nh^{2p+3+2\mathfrak{h}}\right)$. ■

S7 Proof of Theorem 7

Define

$$\tilde{U}_i(\theta_0,\theta_1) \coloneqq \begin{pmatrix} Y_i - \theta_0 D_i \\ 1 \\ Z_i - \theta_1 D_i \end{pmatrix}, \tilde{G}_{0,i} \coloneqq \begin{pmatrix} D_i \\ 0_{d_z+1} \end{pmatrix} \text{ and } \tilde{G}_{1,i} \coloneqq \begin{bmatrix} 0_{2\times d_z} \\ D_i \cdot \mathrm{I}_{d_z} \end{bmatrix}.$$

Then, we have $(\partial/\partial\theta_0)\tilde{U}_i(\theta_0,\theta_1) = -\tilde{G}_{0,i}$ and $\left(\partial/\partial\theta_1^\top\right)\tilde{U}_i(\theta_0,\theta_1) = -\tilde{G}_{1,i}$. Also let $\tilde{U}_i \coloneqq \tilde{U}_i(\vartheta, \delta\cdot l_n)$. We recycle some previous notations. $(\Delta_{A,k},\Psi_{A,k},\Delta_k,\Psi_k,\Delta_A,\Psi_A)$ are defined by the same formulae with p replaced by $p+1$. Consider the singular value decomposition

$$\tilde{\mathrm{S}}^\top \Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\left(-\Delta_{\tilde{G}_0}\right)\tilde{\mathrm{T}} = \begin{bmatrix} \tilde{\Lambda} \\ 0_{d_z+1} \end{bmatrix}$$

of $\Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\left(-\Delta_{\tilde{G}_0}\right)$, where $\tilde{\mathrm{S}}^\top\tilde{\mathrm{S}} = \mathrm{I}_{d_z+2}$, $\tilde{\mathrm{T}} = 1$ and $\tilde{\Lambda} \coloneqq \sqrt{\Delta_{\tilde{G}_0}^\top \Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\Delta_{\tilde{G}_0}}$. Then we have

$$\frac{\tilde{\mathrm{S}}^\top \Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\left(-\Delta_{\tilde{G}_0}\right)}{\sqrt{\Delta_{\tilde{G}_0}^\top \Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\Delta_{\tilde{G}_0}}} = \mathrm{e}_{d_z+2,1}. \tag{S123}$$

Let $\tilde{\Omega} \coloneqq \tilde{\Lambda}^{-1}$, $\tilde{V}_i(\theta_0,\theta_1) \coloneqq \tilde{\mathrm{S}}^\top \Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\tilde{U}_i(\theta_0,\theta_1)$, $\tilde{V}_i \coloneqq \tilde{V}_i(\vartheta,\delta\cdot l_n)$, $\tilde{H}_{0,i} \coloneqq \tilde{\mathrm{S}}^\top \Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\left(-\tilde{G}_{0,i}\right)$ and $\tilde{H}_{1,i} \coloneqq \tilde{\mathrm{S}}^\top \Delta_{\tilde{U}\tilde{U}^\top,2}^{-1/2}\left(-\tilde{G}_{1,i}\right)$. Then we have

$$\ell_{p+1}^{\mathsf{mc}}(\theta \mid h) = \sup_\lambda \frac{1}{n}\sum_i \log\left(1 + \lambda^\top\left(W_{p+1,i}\tilde{V}_i(\theta, 0_{d_z})\right)\right).$$

Let $\widehat{\lambda}_{p+1}^{\mathsf{mc}}$ be defined by

$$\widehat{\lambda}_{p+1}^{\mathsf{mc}} \coloneqq \operatorname*{argmax}_\lambda \sum_i \log\left(1 + \lambda^\top\left(W_{p+1,i}\tilde{V}_i\left(\widehat{\vartheta}_{p+1}^{\mathsf{mc}}, 0_{d_z}\right)\right)\right).$$

Let $\widetilde{\lambda}_{p+1}^{\mathsf{mc}}$ be redefined by

$$\widetilde{\lambda}_{p+1}^{\mathsf{mc}} \coloneqq \operatorname*{argmax}_\lambda \sum_i \log\left(1 + \lambda^\top\left(W_{p+1,i}\tilde{V}_i(\vartheta, 0_{d_z})\right)\right).$$

Then, the likelihood ratio is given by

$$LR_{p+1}(\vartheta \mid h) = 2\left\{\sum_i \log\left(1 + \left(\widetilde{\lambda}_{p+1}^{\mathsf{mc}}\right)^\top\left(W_{p+1,i}\tilde{V}_i(\vartheta, 0_{d_z})\right)\right) - \sum_i \log\left(1 + \left(\widehat{\lambda}_{p+1}^{\mathsf{mc}}\right)^\top\left(W_{p+1,i}\tilde{V}_i\left(\widehat{\vartheta}_{p+1}^{\mathsf{mc}}, 0_{d_z}\right)\right)\right)\right\}.$$

Let $\left(\tilde{A}^{\mathsf{k}}, \tilde{A}^{\mathsf{kl}}, \tilde{\alpha}^{\mathsf{k}}, \tilde{\alpha}^{\mathsf{kl}}, \tilde{\alpha}^{\mathsf{klm}}, \mathring{\tilde{A}}^{\mathsf{k}}\right)$ be defined by the same formulae as $\left(A^{\mathsf{k}}, A^{\mathsf{kl}}, \alpha^{\mathsf{k}}, \alpha^{\mathsf{kl}}, \alpha^{\mathsf{klm}}, \mathring{A}^{\mathsf{k}}\right)$ with V replaced by $\tilde{V}$. The ranges of the following indices are fixed: $a, b, c = 1, 2, ..., d_z$. Let $\tilde{\gamma}^{\mathsf{k},a} := \Delta_{\tilde{H}_1^{(\mathsf{k}a)}}$ and $\tilde{\gamma}^{\mathsf{k:l},a} := \Delta_{\tilde{V}^{(\mathsf{k})}\tilde{H}_1^{(\mathsf{l}a)},2}$. Let $\tilde{C}^{\mathsf{k},a} := \Psi_{\tilde{H}_1^{(\mathsf{k}a)}} - \Delta_{\tilde{H}_1^{(\mathsf{k}a)}}$ and $\tilde{C}^{\mathsf{k:l},a} := \Psi_{\tilde{V}^{(\mathsf{k})}\tilde{H}_1^{(\mathsf{l}a)},2} - \Delta_{\tilde{V}^{(\mathsf{k})}\tilde{H}_1^{(\mathsf{l}a)},2}$. $\left(\tilde{A}^{\mathsf{k}}, \tilde{A}^{\mathsf{kl}}, \mathring{\tilde{A}}^{\mathsf{k}}, \tilde{C}^{\mathsf{k},a}, \tilde{C}^{\mathsf{k:l},a}\right)$ are all $O_p^{\star}\left(\sqrt{\log(n)} \cdot l_n\right)$. Let $V_{0,i} := \tilde{V}_i\left(\vartheta, 0_{d_z}\right)$, $A_0^{\mathsf{k}} := \Psi_{V_0^{(\mathsf{k})}}$ and $A_0^{\mathsf{kl}} := \Psi_{V_0^{(\mathsf{k})}V_0^{(\mathsf{l})},2} - \tilde{\alpha}^{\mathsf{kl}}$. It is easy to see that

$$\begin{aligned} A_0^{\mathsf{k}} &= \tilde{A}^{\mathsf{k}} + l_n\left(\tilde{C}^{\mathsf{k},a} + \tilde{\gamma}^{\mathsf{k},a}\right)\delta^{(a)} \\ A_0^{\mathsf{kl}} &= \tilde{A}^{\mathsf{kl}} + l_n\left(\tilde{C}^{\mathsf{k:l},a} + \tilde{\gamma}^{\mathsf{k:l},a}\right)\delta^{(a)}\left[\mathsf{k},\mathsf{l}\right] + O_p^{\star}\left(l_n^2\right). \end{aligned} \tag{S124}$$

Denote $\bar{N}_i := \begin{pmatrix} 1 & N_i \end{pmatrix}^{\top}$. Let

$$\begin{aligned} \bar{\gamma}_N &:= \left(\Delta_{NN^{\top},2} - \frac{\Delta_{N,2}\Delta_{N^{\top},2}}{\Delta_2}\right)^{-1}\left(\Delta_{MN,2} - \frac{\Delta_{M,2}\Delta_{N,2}}{\Delta_2}\right) \\ \bar{\sigma}_N^2 &:= \Delta_{M^2,2} - \Delta_{M\bar{N}^{\top},2}\Delta_{\bar{N}\bar{N}^{\top},2}^{-1}\Delta_{M\bar{N},2} \end{aligned}$$

and $\mathscr{V}_N := \bar{\sigma}_N^2/\Delta_D^2$. By writing $\Delta_{\tilde{U}\tilde{U}^{\top},2}$ and $\Delta_{\bar{N}\bar{N}^{\top},2}$ as block matrices and inverting, (S45) and (S54) hold with $\left(G, U, \bar{Z}\right)$ replaced by $\left(\tilde{G}_0, \tilde{U}, \bar{N}\right)$ and (S37) holds with Z replaced by N. By these results,

$$M - \Delta_{M\bar{N}^{\top},2}\Delta_{\bar{N}\bar{N}^{\top},2}^{-1}\bar{N} = \left(M - \frac{\Delta_{M,2}}{\Delta_2}\right) - \bar{\gamma}_N^{\top}\left(N - \frac{\Delta_{N,2}}{\Delta_2}\right). \tag{S125}$$

By (S45) and (S37), $\Delta_{\tilde{G}_0}^{\top}\Delta_{\tilde{U}\tilde{U}^{\top},2}^{-1}\Delta_{\tilde{G}_1} = -\bar{\gamma}_N^{\top}/\mathscr{V}_N$. By this result, (S54) and (S123),

$$\tilde{\gamma}^{1,a}\delta^{(a)} = \frac{\Delta_{\tilde{G}_0}^{\top}\Delta_{\tilde{U}\tilde{U}^{\top},2}^{-1}\Delta_{\tilde{G}_1}\delta}{\sqrt{\Delta_{\tilde{G}_0}^{\top}\Delta_{\tilde{U}\tilde{U}^{\top},2}^{-1}\Delta_{\tilde{G}_0}}} = -\frac{\bar{\gamma}_N^{\top}\delta}{\sqrt{\mathscr{V}_N}}. \tag{S126}$$

Denote $\tilde{\mathrm{Q}} := \Delta_{\tilde{U}\tilde{U}^{\top},2}^{-1} - \Delta_{\tilde{U}\tilde{U}^{\top},2}^{-1}\Delta_{\tilde{G}_0}\left(\Delta_{\tilde{G}_0}^{\top}\Delta_{\tilde{U}\tilde{U}^{\top},2}^{-1}\Delta_{\tilde{G}_0}\right)^{-1}\Delta_{\tilde{G}_0}^{\top}\Delta_{\tilde{U}\tilde{U}^{\top},2}^{-1}$. Then (S46) holds with $\left(\mathrm{Q}, \bar{Z}\right)$ replaced by $\left(\tilde{\mathrm{Q}}, \bar{N}\right)$. (S47) holds with $\left(V, H, U, \Omega\right)$ replaced by $\left(\tilde{V}, \tilde{H}_0, \tilde{U}, \tilde{\Omega}\right)$.

It is easy to check that the conclusion of Lemma 13 still holds for $\left(\widehat{\lambda}_{p+1}^{\mathsf{mc}}, \widehat{\vartheta}_{p+1}^{\mathsf{mc}}, \widetilde{\lambda}_{p+1}^{\mathsf{mc}}\right)$. (S111) and (S112) hold for $\left(\widehat{\lambda}_{p+1}^{\mathsf{mc}}, \widehat{\vartheta}_{p+1}^{\mathsf{mc}}\right)$ with $\left(V_i, H_i\right)$ replaced by $\left(V_{0,i}, \tilde{H}_{0,i}\right)$. Let $\left(\tilde{R}_1^0, \tilde{R}_2^0\right)$ be defined by the formulae of $\left(\tilde{R}_1, \tilde{R}_2\right)$ in the proof of Lemma 14 with V_i replaced by $V_{0,i}$: $\tilde{R}_1^0 = A_0^1$ and

$$\tilde{R}_2^0 := -\frac{1}{2} \cdot A_0^{11}A_0^1 - A_0^{1\,1+\mathsf{a}}A_0^{1+\mathsf{a}} + \frac{1}{3} \cdot \tilde{\alpha}^{111}A_0^1A_0^1 + \tilde{\alpha}^{11\,1+\mathsf{a}}A_0^1A_0^{1+\mathsf{a}} + \tilde{\alpha}^{1\,1+\mathsf{a}\,1+\mathsf{b}}A_0^{1+\mathsf{a}}A_0^{1+\mathsf{b}}.$$

By arguments as in the proof of Lemma 14, $LR_{p+1}\left(\vartheta \mid h\right) = \left(nh\right)\left(\tilde{R}_1^0 + \tilde{R}_2^0\right)^2 + O_p^{\star}\left(\log\left(n\right)^2 l_n^2\right)$. By using these results and replacing $\tilde{A}^{\mathsf{k}}$ with $\mathring{\tilde{A}}^{\mathsf{k}} + \tilde{\alpha}^{\mathsf{k}}$, we decompose $\tilde{R}_1^0 = \tilde{R}_{11}^0 + \tilde{R}_{10}^0$, where $\tilde{R}_{11}^0 := \mathring{\tilde{A}}^1 + \delta^{(a)}l_n\tilde{C}^{1,a}$

and $\tilde{R}^0_{10} \coloneqq \tilde{\alpha}^1 + \tilde{\gamma}^{1,a}\delta^{(a)}l_n$. Similarly, by using (S124), we have $\tilde{R}^0_2 = \tilde{R}^0_{20} + \tilde{R}^0_{21} + \tilde{R}^0_{22} + O^\star_p\left(\log(n)\, l_n^3\right)$, where

$$\begin{aligned}\tilde{R}^0_{20} \quad = \quad & -\tilde{\gamma}^{1:1,a}\tilde{\gamma}^{1,b}\delta^{(a)}\delta^{(b)}l_n^2 - \left(\tilde{\gamma}^{1:1+\mathsf{a},a} + \tilde{\gamma}^{1+\mathsf{a}:1,a}\right)\tilde{\gamma}^{1+\mathsf{a},b}\delta^{(a)}\delta^{(b)}l_n^2 + \frac{1}{3}\cdot\tilde{\alpha}^{111}\tilde{\gamma}^{1,a}\tilde{\gamma}^{1,b}\delta^{(a)}\delta^{(b)}l_n^2 \\ & +\tilde{\alpha}^{11\,1+\mathsf{a}}\tilde{\gamma}^{1,a}\tilde{\gamma}^{1+\mathsf{a},b}\delta^{(a)}\delta^{(b)}l_n^2 + \tilde{\alpha}^{1\,1+\mathsf{a}\,1+\mathsf{b}}\tilde{\gamma}^{1+\mathsf{a},a}\tilde{\gamma}^{1+\mathsf{b},b}\delta^{(a)}\delta^{(b)}l_n^2 + O\left(l_n^3\right),\end{aligned}$$

$$\begin{aligned}\tilde{R}^0_{21} \quad = \quad & -\frac{1}{2}\cdot\tilde{A}^{11}\tilde{\gamma}^{1,a}\delta^{(a)}l_n - \tilde{\gamma}^{1:1,a}\mathring{\tilde{A}}^{1}\delta^{(a)}l_n - \tilde{A}^{1\,1+\mathsf{a}}\tilde{\gamma}^{1+\mathsf{a},a}\delta^{(a)}l_n \\ & - \left(\tilde{\gamma}^{1:1+\mathsf{a},a} + \tilde{\gamma}^{1+\mathsf{a}:1,a}\right)\mathring{\tilde{A}}^{1+\mathsf{a}}\delta^{(a)}l_n + \frac{2}{3}\cdot\tilde{\alpha}^{111}\mathring{\tilde{A}}^{1}\tilde{\gamma}^{1,a}\delta^{(a)}l_n + \tilde{\alpha}^{11\,1+\mathsf{a}}\mathring{\tilde{A}}^{1}\tilde{\gamma}^{1+\mathsf{a},a}\delta^{(a)}l_n \\ & +\tilde{\alpha}^{11\,1+\mathsf{a}}\tilde{\gamma}^{1,a}\mathring{\tilde{A}}^{1+\mathsf{a}}\delta^{(a)}l_n + 2\cdot\tilde{\alpha}^{1\,1+\mathsf{a}\,1+\mathsf{b}}\mathring{\tilde{A}}^{1+\mathsf{a}}\tilde{\gamma}^{1+\mathsf{b},a}\delta^{(a)}l_n + O^\star_p\left(\log(n)\, l_n^3\right)\end{aligned}$$

and

$$\tilde{R}^0_{22} \coloneqq -\frac{1}{2}\cdot\tilde{A}^{11}\mathring{\tilde{A}}^{1} - \tilde{A}^{1\,1+\mathsf{a}}\mathring{\tilde{A}}^{1+\mathsf{a}} + \frac{1}{3}\cdot\tilde{\alpha}^{111}\mathring{\tilde{A}}^{1}\mathring{\tilde{A}}^{1} + \tilde{\alpha}^{11\,1+\mathsf{a}}\mathring{\tilde{A}}^{1}\mathring{\tilde{A}}^{1+\mathsf{a}} + \tilde{\alpha}^{1\,1+\mathsf{a}\,1+\mathsf{b}}\mathring{\tilde{A}}^{1+\mathsf{a}}\mathring{\tilde{A}}^{1+\mathsf{b}}.$$

Let $R^0_0 \coloneqq \tilde{R}^0_{10} + \tilde{R}^0_{20}$, $R^0_1 \coloneqq \tilde{R}^0_{11} + \tilde{R}^0_{21}$, $R^0_2 \coloneqq \tilde{R}^0_{22}$ and $R^0 \coloneqq R^0_1 + R^0_2$ so that we have $LR_{p+1}(\vartheta \mid h) = (nh)\left(R^0_0 + R^0\right)^2 + O^\star_p(l_n h)$. Denote $\bar{\kappa}^{ab}_0 \coloneqq \tilde{\gamma}^{1,a}\tilde{\gamma}^{1,b}$. Then, by (S126), we have

$$\bar{\kappa}^{ab}_0\delta^{(a)}\delta^{(b)}l_n^2 = \frac{\left(\bar{\gamma}_N^\top\delta\right)^2 l_n^2}{\bar{\mathscr{V}}_N}. \tag{S127}$$

By tedious algebra and Lemma 3(b),

$$\left(R^0_0\right)^2 = \tilde{\alpha}^1\tilde{\gamma}^{1,a}\delta^{(a)}l_n + \bar{\kappa}^{ab}_0\delta^{(a)}\delta^{(b)}l_n^2 + \bar{\kappa}^{abc}_1\delta^{(a)}\delta^{(b)}\delta^{(c)}l_n^3 + o\left(l_n^3\right)$$

where

$$\begin{aligned}\bar{\kappa}^{abc}_1 \quad \coloneqq \quad & -\frac{2}{3}\cdot\tilde{\alpha}^{111}\tilde{\gamma}^{1,a}\tilde{\gamma}^{1,b}\tilde{\gamma}^{1,c} + 2\cdot\tilde{\alpha}^{11\,1+\mathsf{a}}\tilde{\gamma}^{1,a}\tilde{\gamma}^{1+\mathsf{a},b}\tilde{\gamma}^{1,c} + 2\cdot\tilde{\alpha}^{1\,1+\mathsf{a}\,1+\mathsf{b}}\tilde{\gamma}^{1+\mathsf{a},a}\tilde{\gamma}^{1+\mathsf{b},b}\tilde{\gamma}^{1,c} \\ & -2\cdot\tilde{\gamma}^{1:1,a}\tilde{\gamma}^{1,b}\tilde{\gamma}^{1,c} - 2\cdot\left(\tilde{\gamma}^{1:1+\mathsf{a},a} + \tilde{\gamma}^{1+\mathsf{a}:1,a}\right)\tilde{\gamma}^{1+\mathsf{a},b}\tilde{\gamma}^{1,c}.\end{aligned}$$

Let $\bar{\kappa}^a_2 \coloneqq \tilde{\alpha}^{111}\tilde{\gamma}^{1,a}/3$ and $\tilde{\kappa}_{2,n} \coloneqq 1 + \bar{\kappa}^a_2\delta^{(a)}l_n$. By calculation using arguments in the proof of Lemma 14, we have $\kappa_1\left(\sqrt{nh}\cdot R^0\right) = \tilde{\kappa}_{1,n} + o(l_n)$, $\kappa_2\left(\sqrt{nh}\cdot R^0\right) = \tilde{\kappa}_{2,n} + o(l_n)$ and $\kappa_3\left(\sqrt{nh}\cdot R^0\right) = o(l_n)$, where $\tilde{\kappa}_{1,n} \coloneqq -(nh)^{-1/2}\tilde{\alpha}^{111}/6$. Then, $2\sqrt{nh}\cdot R^0_0\tilde{\kappa}_{1,n} = -\bar{\kappa}^a_2\delta^{(a)}l_n + O\left(l_n^2\right)$. By arguments used to show (S118) and (S119) (i.e., Skovgaard, 1981 with $s = p = q = 3$, $\beta_{s,n} = l_n$ and $\lambda_n = O(l_n)$),

$$\Pr\left[(nh)\left(R^0_0 + R^0\right)^2 \le x\right] = F\left(\frac{x}{\tilde{\kappa}_{2,n}} \mid \frac{\left(\sqrt{nh}\cdot R^0_0 + \tilde{\kappa}_{1,n}\right)^2}{\tilde{\kappa}_{2,n}}\right) + o(l_n). \tag{S128}$$

Then by Taylor expansion,

$$F\left(\frac{x}{\tilde{\kappa}_{2,n}} \mid \frac{\left(\sqrt{nh}\cdot R_0^0+\tilde{\kappa}_{1,n}\right)^2}{\tilde{\kappa}_{2,n}}\right) = F\left(\frac{x}{\tilde{\kappa}_{2,n}} \mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right) + \Big\{(nh)\,\bar{\kappa}_1^{abc}\delta^{(a)}\delta^{(b)}\delta^{(c)}l_n^2$$
$$+(nh)\,\tilde{\alpha}^1\tilde{\gamma}^{1,a}\delta^{(a)} - (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\cdot\bar{\kappa}_2^a\delta^{(a)} - \bar{\kappa}_2^a\delta^{(a)}\Big\}\,F^{(1)}\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)l_n + o\left(l_n\right). \quad \text{(S129)}$$

Let $f(\cdot\mid\iota)$ denote the $\chi_1^2(\iota)$ PDF. By using the recurrence properties of non-central χ^2 (Cohen, 1988), $-xf(x\mid\iota) = 2\iota\cdot F^{(2)}(x\mid\iota) + (\iota+1)\,F^{(1)}(x\mid\iota)$. By these results, (S127) and Taylor expansion,

$$\begin{aligned}
F\left(\frac{x}{\tilde{\kappa}_{2,n}} \mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right) &= F\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right) + xf\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)\left(\frac{1}{\tilde{\kappa}_{2,n}}-1\right) + O\left(l_n^2\right) \\
&= F\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right) \\
&\quad +2\left((nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)F^{(2)}\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)\left(\bar{\kappa}_2^a\delta^{(a)}l_n\right) \\
&\quad +\left((nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2+1\right)F^{(1)}\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)\left(\bar{\kappa}_2^a\delta^{(a)}l_n\right) + o\left(l_n\right). \quad \text{(S130)}
\end{aligned}$$

It then follows from these results, (S129) and (S130) that

$$\begin{aligned}
\Pr\left[(nh)\left(R_0^0+R^0\right)^2\le x\right] &= F\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right) \\
&\quad +\left((nh)\,\bar{\kappa}_1^{abc}\delta^{(a)}\delta^{(b)}\delta^{(c)}l_n^2 + (nh)\,\tilde{\alpha}^1\tilde{\gamma}^{1,a}\delta^{(a)}\right)F^{(1)}\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)l_n \\
&\quad +2\left((nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)F^{(2)}\left(x\mid (nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}l_n^2\right)\left(\bar{\kappa}_2^a\delta^{(a)}l_n\right) + o\left(l_n\right). \quad \text{(S131)}
\end{aligned}$$

Let $\epsilon_N \coloneqq M - N^\top\gamma_N$, $\mathring{\epsilon}_N \coloneqq \epsilon_N - \mu_{\epsilon_N}$, $\mathring{N} \coloneqq N - \mu_N$ and

$$\begin{aligned}
\mathscr{P}_1(\delta) &\coloneqq H\Bigg\{-\frac{2}{3}\cdot\frac{\omega_{p+1}^{0,3}\mu_{\mathring{\epsilon}_N^3,\dagger}\varphi\mu_{D,\dagger}^3\left(\gamma_N^\top\delta\right)^3}{\left(\omega_{p+1}^{0,2}\mathrm{Var}_{|0^\pm}\left[\epsilon_N\right]\right)^3} + 2\cdot\frac{\omega_{p+1}^{0,3}\mu_{D,\dagger}^3\left(\mu_{\mathring{\epsilon}_N^2\mathring{N}^\top,\dagger}\right)\left(\mathrm{Var}_{|0^\pm}\left[N\right]\right)^{-1}\delta\left(\gamma_N^\top\delta\right)^2}{\left(\omega_{p+1}^{0,2}\right)^3\left(\mathrm{Var}_{|0^\pm}\left[\epsilon_N\right]\right)^2} \\
&\quad +2\cdot\frac{\omega_{p+1}^{0,3}\varphi^{1/2}\mu_{D,\dagger}^3\left(\delta^\top\left(\mathrm{Var}_{|0^\pm}\left[N\right]\right)^{-1}\left(\mu_{\mathring{\epsilon}_N\mathring{N}\mathring{N}^\top}\right)\left(\mathrm{Var}_{|0^\pm}\left[N\right]\right)^{-1}\delta\right)\left(\gamma_N^\top\delta\right)}{\left(\omega_{p+1}^{0,2}\right)^{5/2}\sqrt{\mathrm{Var}_{|0^\pm}\left[\epsilon_N\right]}} \\
&\quad +2\left(\frac{\varphi\mu_{D,\dagger}^2\mathrm{Cov}_{|0^\pm}\left[D,\epsilon_N\right]\left(\gamma_N^\top\delta\right)^3}{\omega_{p+1}^{0,2}\left(\mathrm{Var}_{|0^\pm}\left[\epsilon_N\right]\right)^2}\right) - 2\left(\frac{\left(\mu_{D,\dagger}\right)^2\varphi\mathrm{Cov}_{|0^\pm}\left[D,\epsilon_N\right]\delta^\top\left(\mathrm{Var}_{|0^\pm}\left[N\right]\right)^{-1}\delta\left(\gamma_N^\top\delta\right)}{\mathrm{Var}_{|0^\pm}\left[\epsilon_N\right]\omega_p^{0,2}}\right) \\
&\quad -2\cdot\frac{\left(\mu_{D,\dagger}\right)^2\left(\mathrm{Cov}_{|0^\pm}\left[D,N\right]\left(\mathrm{Var}_{|0^\pm}\left[N\right]\right)^{-1}\delta\right)\left(\gamma_N^\top\delta\right)^2}{\omega_{p+1}^{0,2}\mathrm{Var}_{|0^\pm}\left[\epsilon_N\right]}\Bigg\} \\
\mathscr{P}_2(\delta) &\coloneqq H\cdot\frac{2}{3}\cdot\frac{\omega_{p+1}^{0,3}\mu_{\mathring{\epsilon}_N^3,\dagger}\varphi\left(\gamma_N^\top\delta\right)^3\mu_{D,\dagger}^3}{\left(\omega_{p+1}^{0,2}\mathrm{Var}_{|0^\pm}\left[\epsilon_N\right]\right)^3}.
\end{aligned}$$

By tedious algebra and Lemma 3(a),

$$
\begin{aligned}
(nh)\,\bar{\kappa}_1^{abc}\delta^{(a)}\delta^{(b)}\delta^{(c)}l_n^2 &= \mathscr{P}_1(\delta)+o(1)\\
2(nh)\,\bar{\kappa}_0^{ab}\delta^{(a)}\delta^{(b)}\bar{\kappa}_2^{a}\delta^{(a)}l_n^2 &= \frac{2}{3}(nh)\,\tilde{\alpha}^{111}\tilde{\gamma}^{1,a}\tilde{\gamma}^{1,b}\tilde{\gamma}^{1,c}\delta^{(a)}\delta^{(b)}\delta^{(c)}l_n^2\\
&= \mathscr{P}_2(\delta)+o(1).
\end{aligned}
$$

By Lemma 3(a,b), (S54), (S126) and (S123), we have $(nh)\,\tilde{\alpha}^{1}\tilde{\gamma}^{1,a}\delta^{(a)}=O\left(h^{\mathfrak{h}}\right)$. The conclusion follows from these results, (S127), (S131), the fact that $LR_{p+1}(\vartheta\mid h)=(nh)\left(R_0^0+R^0\right)^2+O_p^{\star}(l_n h)$ and (S107).

S8 Proof of Theorem 8

Theorem 8 can be proven in a similar way as Theorem 1. Denote

$$
\begin{aligned}
\ddot{W}_{p;+,i} &\coloneqq \mathrm{e}_{p+1,2}^{\top}\Pi_{p,+}^{-1}r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}(X_i>0)\\
\ddot{W}_{p;-,i} &\coloneqq \mathrm{e}_{p+1,2}^{\top}\Pi_{p,-}^{-1}r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}(X_i<0)\\
\ddot{W}_{p,i} &\coloneqq \ddot{W}_{p;+,i}-\ddot{W}_{p;-,i}.
\end{aligned}
$$

Let $\varpi_{p,+}\coloneqq\int_0^1\mathcal{K}_{p;+}(t)\,\dot{\mathcal{K}}_{p;+}(t)\,\mathrm{d}t=\varpi_p$ and $\varpi_{p,-}\coloneqq\int_{-1}^0\mathcal{K}_{p;-}(t)\,\dot{\mathcal{K}}_{p;-}(t)\,\mathrm{d}t=-\varpi_p$.

Lemma 15. *Let V denote a random variable and $\{V_1,...,V_n\}$ are i.i.d. copies of V. Assume that $nh^3\to\infty$. Suppose that K is a symmetric continuous PDF supported on $[-1,1]$. Let $\mathbb{B}\subseteq[\underline{x},\overline{x}]\setminus\{0\}$ denote an open neighborhood of 0. The following results hold for all $(\mathrm{s},k)\in\{-,+\}\times\mathbb{N}$: (a) if g_V is uniformly continuous on $\mathbb{B}$,*

$$
\mathrm{E}\left[\frac{1}{h}\ddot{W}_{p;\mathrm{s}}^{k}V\right]=\frac{\mu_{V,\mathrm{s}}\dot{\omega}_p^{0,k}}{\varphi^{k-1}}+o(1)\ \ \textit{and}\ \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;\mathrm{s}}\ddot{W}_{p;\mathrm{s}}V\right]=\frac{\mu_{V,\mathrm{s}}\varpi_{p,\mathrm{s}}}{\varphi}+o(1);
$$

(b) if g_V is $(p+1)$-times continuously differentiable with uniformly continuous $g_V^{(p+1)}$ on $\mathbb{B}$,

$$
\frac{1}{nh^2}\sum_i\dot{W}_{p;\mathrm{s},i}g_V(X_i)=\mu_{V,\mathrm{s}}^{(1)}+\frac{\mu_{V,\mathrm{s}}^{(p+1)}}{(p+1)!}\dot{\omega}_{p;\mathrm{s}}^{p+1,1}h^p+o_p(h^p);
$$

(c) if g_{V^2} is bounded on $\mathbb{B}$,

$$
\frac{1}{nh}\sum_i\widehat{W}_{p;\mathrm{s},i}\dot{W}_{p;\mathrm{s},i}V_i-\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;\mathrm{s}}\ddot{W}_{p;\mathrm{s}}V\right]=O_p\left((nh)^{-1/2}\right).
$$

Proof of Lemma 15. We take $\mathrm{s} = +$ without loss of generality. For Part (a), we have

$$\begin{aligned}\mathrm{E}\left[\frac{1}{h}\ddot{W}^{k}_{p;+}V\right] &= \mathrm{E}\left[\frac{1}{h}\ddot{W}^{k}_{p;+}g_V(X)\right]\\ &= \int_0^{\overline{x}} \frac{1}{h}\left(\mathrm{e}^{\top}_{p+1,2}\Pi^{-1}_{p,+}r_p\left(\frac{x}{h}\right)\right)^k K^k\left(\frac{x}{h}\right) g_V(x) f_X(x)\,\mathrm{d}x\\ &= \int_0^{\frac{\overline{x}}{h}} \left(\mathrm{e}^{\top}_{p+1,2}\Pi^{-1}_{p,+}r_p(y)\right)^k K^k(y) g_V(hy) f_X(hy)\,\mathrm{d}y\\ &= \frac{\mu_{V,+}\dot{\omega}^{0,k}_p}{\varphi^{k-1}} + o(1).\end{aligned}$$

where the first equality follows from LIE, the third equality follows from change of variables, and the fourth equality follows from (S5), continuity of g_V and f_X and applying the equality (S6) to $\left(\mathrm{e}^{\top}_{p+1,2}\Pi^{-1}_{p,+}r_p(y)\right)^k - \left(\mathrm{e}^{\top}_{p+1,2}\left(\varphi^{-1}\cdot \mathrm{V}^{-1}_{p;+}\right)r_p(y)\right)^k$. Similarly, we have

$$\begin{aligned}\mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}\ddot{W}_{p;+}V\right] &= \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}\ddot{W}_{p;+}g_V(X)\right]\\ &= \int_0^{\overline{x}} \frac{1}{h}\left(\mathrm{e}^{\top}_{p+1,1}\Pi^{-1}_{p,+}r_p\left(\frac{x}{h}\right)K\left(\frac{x}{h}\right)\mathrm{e}^{\top}_{p+1,2}\Pi^{-1}_{p,+}r_p\left(\frac{x}{h}\right)K\left(\frac{x}{h}\right)\right) g_V(x) f_X(x)\,\mathrm{d}x\\ &= \int_0^{\frac{\overline{x}}{h}} \left(\mathrm{e}^{\top}_{p+1,1}\Pi^{-1}_{p,+}r_p(y)K(y)\mathrm{e}^{\top}_{p+1,2}\Pi^{-1}_{p,+}r_p(y)K(y)\right) g_V(hy) f_X(hy)\,\mathrm{d}y\\ &= \frac{\mu_{V,+}\varpi_{p,+}}{\varphi} + o(1).\end{aligned}$$

For Part (b), by Taylor's theorem, for $X_i > 0$,

$$g_V(X_i) = \mu_{V,+} + \mu^{(1)}_{V,+}X_i + \left(\frac{\mu^{(2)}_{V,+}}{2!}\right)X_i^2 + \cdots + \left(\frac{\mu^{(p)}_{V,+}}{p!}\right)X_i^p + \frac{g_V^{(p+1)}\left(\tilde{X}_i\right)}{(p+1)!}X_i^{p+1},$$

for some $\tilde{X}_i$ between 0 and X_i. Denote $\mu_+ \coloneqq \left(\mu_{V,+}, \mu^{(1)}_{V,+}, \mu^{(2)}_{V,+}/2, ..., \mu^{(p)}_{V,+}/p!\right)^{\top}$. Then, we write

$$\frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}g_V(X_i) = \frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\left(r_p^{\top}(X_i)\mu_+\right) + \frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\frac{g_V^{(p+1)}\left(\tilde{X}_i\right)}{(p+1)!}X_i^{p+1}. \tag{S132}$$

Clearly, by the definition of $\dot{W}_{p;+,i}$,

$$\begin{aligned}\frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\left(r_p^{\top}(X_i)\mu_+\right) &= \frac{1}{nh^2}\sum_i \mathrm{e}^{\top}_{p+1,2}\widehat{\Pi}^{-1}_{p,+}r_p\left(\frac{X_i}{h}\right)K\left(\frac{X_i}{h}\right)\mathbb{1}(X_i>0)\,r_p^{\top}\left(\frac{X_i}{h}\right)\mathrm{H}\mu_+\\ &= \frac{\mathrm{e}^{\top}_{p+1,2}\mathrm{H}\mu_+}{h}\\ &= \mu^{(1)}_{V,+}.\end{aligned} \tag{S133}$$

Write

$$\frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\frac{g_V^{(p+1)}\left(\tilde{X}_i\right)}{(p+1)!}X_i^{p+1} = \frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\frac{\left(g_V^{(p+1)}\left(\tilde{X}_i\right)-\mu_{V,+}^{(p+1)}\right)}{(p+1)!}X_i^{p+1} + \frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\frac{\mu_{V,+}^{(p+1)}}{(p+1)!}X_i^{p+1}. \tag{S134}$$

By (S10) and (S5),

$$\begin{aligned}\frac{1}{nh}\sum_i \left|\dot{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1}\right| &\leq \left\|\widehat{\Pi}_{p,+}^{-1}\right\|\left(\frac{1}{nh}\sum_i \left\|r_p\left(\frac{X_i}{h}\right)\right\|\left|\left(\frac{X_i}{h}\right)^{p+1}\right| K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i>0\right)\right)\\ &= O_p(1).\end{aligned} \tag{S135}$$

By this result, $|K(X_i/h)| \lesssim \mathbb{1}(|X_i|\leq h)$ and continuity of $g_V^{(p+1)}$, we have

$$\begin{aligned}&\left|\frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\frac{\left(g_V^{(p+1)}\left(\tilde{X}_i\right)-\mu_{V,+}^{(p+1)}\right)}{(p+1)!}X_i^{p+1}\right|\\ &\qquad\leq \left(\frac{1}{nh}\sum_i \left|\dot{W}_{p;+,i}\frac{(X_i/h)^{p+1}}{(p+1)!}\right|\right)\cdot\left(\sup_{0<x<h}\left|g_V^{(p+1)}(x)-\mu_{V,+}^{(p+1)}\right|\right)h^p = o_p\left(h^p\right).\end{aligned}$$

It now follows from this result, (S132), (S133) and (S134) that

$$\frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}g_V\left(X_i\right) = \mu_{V,+}^{(1)} + \frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\frac{\mu_{V,+}^{(p+1)}}{(p+1)!}X_i^{p+1} + o_p\left(h^p\right). \tag{S136}$$

By triangle inequality, (S5) and (S10),

$$\begin{aligned}&\left|\frac{1}{nh}\sum_i \dot{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1} - \frac{1}{nh}\sum_i \ddot{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1}\right| \leq \frac{1}{nh}\sum_i\left|\left(\dot{W}_{p;+,i}-\ddot{W}_{p;+,i}\right)\left(\frac{X_i}{h}\right)^{p+1}\right|\\ &\qquad\leq \left\|\widehat{\Pi}_{p,+}^{-1}-\Pi_{p,+}^{-1}\right\|\left\{\frac{1}{nh}\sum_i\left\|r_p\left(\frac{X_i}{h}\right)\right\|\left|K\left(\frac{X_i}{h}\right)\right|\left|\left(\frac{X_i}{h}\right)^{p+1}\right|\mathbb{1}\left(X_i>0\right)\right\}\\ &\qquad= O_p\left((nh)^{-1/2}\right).\end{aligned}$$

By (S5), Chebyshev's inequality, change of variables and continuity of f_X,

$$\frac{1}{nh}\sum_i \ddot{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1} - \mathrm{E}\left[\frac{1}{h}\ddot{W}_{p;+}\left(\frac{X_i}{h}\right)^{p+1}\right] = O_p\left((nh)^{-1/2}\right).$$

By (S5), change of variables and continuity of f_X,

$$\mathrm{E}\left[\frac{1}{h}\ddot{W}_{p;+}\left(\frac{X_i}{h}\right)^{p+1}\right] = \int_0^{\overline{x}} \frac{1}{h}\left(\mathrm{e}_{p+2,1}^{\top}\Pi_{p,+}^{-1}r_p\left(\frac{x}{h}\right)\right)\left(\frac{x}{h}\right)^{p+1} K\left(\frac{x}{h}\right) f_X(x)\,\mathrm{d}x$$
$$= \dot{\omega}_{p;+}^{p+1,1} + o(1).$$

It now follows from these results that

$$\frac{1}{nh^2}\sum_i \dot{W}_{p;+,i}\frac{\mu_{V,+}^{(p+1)}}{(p+1)!}X_i^{p+1} = \left(\frac{1}{nh}\sum_i \dot{W}_{p;+,i}\left(\frac{X_i}{h}\right)^{p+1}\right)\frac{\mu_{V,+}^{(p+1)}}{(p+1)!}h^p$$
$$= \frac{\mu_{V,+}^{(p+1)}}{(p+1)!}\dot{\omega}_{p;+}^{p+1,1}h^p + o_p(h^p).$$

Part (b) follows from this result and (S136).

For Part (c), write

$$\frac{1}{nh}\sum_i \widehat{W}_{p;+,i}\dot{W}_{p;+,i}V_i - \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}\ddot{W}_{p;+}V\right] = \frac{1}{nh}\sum_i\left(\widehat{W}_{p;+,i}\dot{W}_{p;+,i} - \widetilde{W}_{p;+,i}\ddot{W}_{p;+}\right)V_i$$
$$+\left\{\frac{1}{nh}\sum_i \widetilde{W}_{p;+,i}\ddot{W}_{p;+,i}V_i - \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}\ddot{W}_{p;+}V\right]\right\}$$

Then, by triangle inequality, LIE, change of variables, boundedness of $g_{|V|}$, Markov's inequality, and using (S5), we have

$$\left|\frac{1}{nh}\sum_i\left(\widehat{W}_{p;+,i}\dot{W}_{p;+,i} - \widetilde{W}_{p;+,i}\ddot{W}_{p;+}\right)V_i\right| \leq \frac{1}{nh}\sum_i\left|\widehat{W}_{p;+,i}\dot{W}_{p;+,i} - \widetilde{W}_{p;+,i}\ddot{W}_{p;+,i}\right||V_i| = O_p\left((nh)^{-1/2}\right).$$

It follows from (S5), LIE, change of variables, Chebyshev's inequality and boundedness of g_{V^2} that

$$\frac{1}{nh}\sum_i \widetilde{W}_{p;+,i}\ddot{W}_{p;+,i}V_i - \mathrm{E}\left[\frac{1}{h}\widetilde{W}_{p;+}\ddot{W}_{p;+}V\right] = O_p\left((nh)^{-1/2}\right).$$

Part (c) follows from these results. ■

Denote $\dot{\Psi}_A \coloneqq (nh)^{-1}\sum_i \widehat{W}_{p,i}\dot{W}_{p,i}A_i$ and $\ddot{\Delta}_A \coloneqq \mathrm{E}\left[h^{-1}\widetilde{W}_p\ddot{W}_pA\right]$.

Proof of Theorem 8. Using

$$w_i^{\mathsf{eb}} = \frac{1}{n}\cdot\frac{1}{1+\left(\lambda_p^{\mathsf{eb}}\right)^{\top}\left(\widehat{W}_{p,i}\bar{Z}_i\right)}$$

and

$$\frac{1}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}=1-\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)+\frac{\left(\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)\right)^2}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)},$$

we have

$$\begin{aligned}\widehat{\pi}_p^{\mathsf{eb}} &= \frac{1}{nh^2}\sum_i\frac{\dot{W}_{p,i}Y_i}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}\\ &= \frac{1}{nh^2}\sum_i\dot{W}_{p,i}Y_i\left\{1-\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)+\frac{\left(\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)\right)^2}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}\right\}\\ &= \frac{1}{nh^2}\sum_i\dot{W}_{p,i}Y_i-\left(\lambda_p^{\mathsf{eb}}\right)^\top\frac{1}{nh^2}\sum_i\widehat{W}_{p,i}\dot{W}_{p,i}Y_i\bar{Z}_i+\frac{1}{nh^2}\sum_i\frac{\left(\dot{W}_{p,i}Y_i\right)\left(\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)\right)^2}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}\end{aligned} \tag{S137}$$

By Lemma 15(c), $\dot{\Psi}_{Y\bar{Z}}-\ddot{\Delta}_{Y\bar{Z}}=O_p\left((nh)^{-1/2}\right)$. By arguments similar to those used in the proof of (S21),

$$\frac{1}{nh^2}\sum_i\frac{\left(\dot{W}_{p,i}Y_i\right)\left(\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)\right)^2}{1+\left(\lambda_p^{\mathsf{eb}}\right)^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right)}=O_p\left(\left(nh^2\right)^{-1}\right).$$

By these results, we have

$$\widehat{\pi}_p^{\mathsf{eb}}=\frac{1}{nh^2}\sum_i\dot{W}_{p,i}Y_i-\frac{\left(\lambda_p^{\mathsf{eb}}\right)^\top\ddot{\Delta}_{Y\bar{Z}}}{h}+O_p\left(\left(nh^2\right)^{-1}\right).$$

Using (S24) and Lemma 2, we have

$$\begin{aligned}\ddot{\Delta}_{Y\bar{Z}}^\top\lambda_p^{\mathsf{eb}} &= \begin{bmatrix}\ddot{\Delta}_Y & \ddot{\Delta}_{YZ}^\top\end{bmatrix}\widetilde{\Delta}_{\bar{Z}\bar{Z}^\top,2}^{-1}\begin{bmatrix}0\\ \widehat{\Psi}_Z\end{bmatrix}+O_p\left((nh)^{-1}\right)\\ &= \widetilde{\gamma}_{\mathsf{ted}}^\top\widehat{\Psi}_Z+O_p\left((nh)^{-1}\right),\end{aligned} \tag{S138}$$

where

$$\widetilde{\gamma}_{\mathsf{ted}}\coloneqq\left(\widetilde{\Delta}_{ZZ^\top,2}-\frac{\widetilde{\Delta}_{Z,2}\widetilde{\Delta}_{Z^\top,2}}{\widetilde{\Delta}_2}\right)^{-1}\left(\ddot{\Delta}_{YZ}-\frac{\widetilde{\Delta}_{Z,2}\ddot{\Delta}_Y}{\widetilde{\Delta}_2}\right).$$

By Lemmas 1(a) and 15(a), $\widetilde{\Delta}_{ZZ^\top,2}=\left(\omega_p^{0,2}/\varphi\right)\mu_{ZZ^\top,\pm}+o(1)$, $\widetilde{\Delta}_{Z,2}=2\left(\omega_p^{0,2}/\varphi\right)\mu_Z+o(1)$, $\widetilde{\Delta}_2=2\left(\omega_p^{0,2}/\varphi\right)+o(1)$, $\ddot{\Delta}_{YZ}=\left(\varpi_p/\varphi\right)\mu_{YZ,\dagger}+o(1)$ and $\ddot{\Delta}_Y=\left(\varpi_p/\varphi\right)\mu_{Y,\dagger}+o(1)$. Therefore, it follows from these results, (S4) and the fact that for $\mathsf{s}\in\{-,+\}$, $\mu_{ZZ^\top,\mathsf{s}}-\mu_Z\mu_{Z^\top}$ is positive definite that $\widetilde{\gamma}_{\mathsf{ted}}=$

$\left(\varpi_p/\omega_p^{0,2}\right)\gamma_{\mathsf{ted}}+o\left(1\right)$. Therefore, by this result, $\widehat{\Psi}_Z=O_p\left(\left(nh\right)^{-1/2}\right)$ and (S138), we have

$$\widehat{\pi}_p^{\mathsf{eb}}=\frac{1}{nh^2}\sum_i \dot{W}_{p,i}Y_i-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\frac{\gamma_{\mathsf{ted}}^\top\widehat{\Psi}_Z}{h}+o_p\left(\left(nh^3\right)^{-1/2}\right).$$

By lemma 15(b),

$$\begin{aligned}\frac{1}{nh^2}\sum_i \dot{W}_{p,i}Y_i &= \frac{1}{nh^2}\sum_i \dot{W}_{p,i}g_Y\left(X_i\right)+\frac{1}{nh^2}\sum_i \dot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)\\ &= \left(\mu_{Y,+}^{(1)}-\mu_{Y,-}^{(1)}\right)+\frac{\left(\mu_{Y,+}^{(p+1)}\dot{\omega}_{p;+}^{p+1,1}-\mu_{Y,-}^{(p+1)}\dot{\omega}_{p;-}^{p+1,1}\right)}{(p+1)!}h^p\\ &\quad +\frac{1}{nh^2}\sum_i \dot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)+o\left(h^p\right).\end{aligned}$$

Combining the above result and (S19), we have

$$\begin{aligned}\widehat{\pi}_p^{\mathsf{eb}}-\pi_{\mathsf{srd}} &= \frac{\left(\mu_{Y,+}^{(p+1)}\dot{\omega}_{p;+}^{p+1,1}-\mu_{Y,-}^{(p+1)}\dot{\omega}_{p;-}^{p+1,1}\right)}{(p+1)!}h^p+\frac{1}{nh^2}\sum_i \dot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)\\ &\quad -\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\gamma_{\mathsf{ted}}^\top\left\{\frac{\left(\mu_{Z,+}^{(p+1)}\omega_{p;+}^{p+1,1}-\mu_{Z,-}^{(p+1)}\omega_{p;-}^{p+1,1}\right)}{(p+1)!}h^p+\frac{1}{nh^2}\sum_i \widehat{W}_{p,i}\left(Z_i-g_Z\left(X_i\right)\right)\right\}\\ &\quad +o_p\left(\left(nh^3\right)^{-1/2}\right)\\ &= \left\{\frac{\left(\mu_{Y,+}^{(p+1)}\dot{\omega}_{p;+}^{p+1,1}-\mu_{Y,-}^{(p+1)}\dot{\omega}_{p;-}^{p+1,1}\right)}{(p+1)!}-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\gamma_{\mathsf{ted}}^\top\frac{\left(\mu_{Z,+}^{(p+1)}\omega_{p;+}^{p+1,1}-\mu_{Z,-}^{(p+1)}\omega_{p;-}^{p+1,1}\right)}{(p+1)!}\right\}h^p\\ &\quad +\frac{1}{nh^2}\sum_i \dot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\frac{1}{nh^2}\sum_i \widehat{W}_{p,i}\gamma_{\mathsf{ted}}^\top\left(Z_i-g_Z\left(X_i\right)\right)+o_p\left(\left(nh^3\right)^{-1/2}\right).\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\sqrt{nh^3}\left(\widehat{\pi}_p^{\mathsf{eb}}-\pi-\mathscr{B}_p^{\mathsf{ted}}h^p\right)\\ =\ &\frac{1}{\sqrt{nh}}\sum_i\left(\dot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\widehat{W}_{p,i}\gamma_{\mathsf{ted}}^\top\left(Z_i-g_Z\left(X_i\right)\right)\right)+o_p(1).\end{aligned}\tag{S139}$$

Following the same arguements for showing (S26), we obtain

$$\begin{aligned}&\frac{1}{\sqrt{nh}}\sum_i\left(\dot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\frac{\varpi_p}{\omega_p^{0,2}}\widehat{W}_{p,i}\gamma_{\mathsf{ted}}^\top\left(Z_i-g_Z\left(X_i\right)\right)\right)\\ =\ &\frac{1}{\sqrt{nh}}\sum_i\left(\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\frac{\varpi_p}{\omega_p^{0,2}}\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^\top\left(Z_i-g_Z\left(X_i\right)\right)\right)+o_p\left(\left(nh\right)^{-1/2}\right).\end{aligned}\tag{S140}$$

By Lemma 1(a) and the definition of $\gamma_{\mathsf{ted}}^{\top}$,

$$\begin{aligned}
&\operatorname{Var}\left[\frac{1}{\sqrt{nh}}\sum_{i}\left(\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\frac{\varpi_p}{\omega_p^{0,2}}\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right)\right]\\
=\;&\mathrm{E}\left[\frac{1}{h}\ddot{W}_p^2\left(Y_i-g_Y\left(X_i\right)\right)^2\right]+\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)^2\mathrm{E}\left[\frac{1}{h}\widetilde{W}_p^2\left(\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right)^2\right]\\
&-\frac{2\varpi_p}{\omega_p^{0,2}}\mathrm{E}\left[\frac{1}{h}\ddot{W}_p\widetilde{W}_p\left(Y_i-g_Y\left(X_i\right)\right)\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right]\\
=\;&\frac{\dot{\omega}_p^{0,2}\operatorname{Var}_{|0\pm}\left[Y\right]-\left(\frac{\varpi_p^2}{\omega_p^{0,2}}\right)\gamma_{\mathsf{ted}}^{\top}\operatorname{Var}_{|0\pm}\left[Z\right]\gamma_{\mathsf{ted}}}{\varphi}+o\left(1\right).
\end{aligned}\tag{S141}$$

Let $\varsigma\in(0,1)$. We have

$$\begin{aligned}
&\sum_i\mathrm{E}\left[\left|\frac{1}{\sqrt{nh}}\left(\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right)\right|^{2+\varsigma}\right]\\
=\;&\frac{1}{(nh)^{\frac{2+\varsigma}{2}}}\sum_i\mathrm{E}\left[\left|\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right|^{2+\varsigma}\right]\\
\lesssim\;&\frac{\mathrm{E}\left[h^{-1}\left|\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)\right|^{2+\varsigma}\right]+\mathrm{E}\left[h^{-1}\left|\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right|^{2+\varsigma}\right]}{(nh)^{\frac{\varsigma}{2}}}\\
\lesssim\;&\frac{\mathrm{E}\left[h^{-1}\left|\ddot{W}_{p,i}\right|^{2+\varsigma}\left(|Y|^{2+\varsigma}+|g_Y\left(X\right)|^{2+\varsigma}\right)\right]}{(nh)^{\frac{\varsigma}{2}}}\\
&+\frac{\mathrm{E}\left[h^{-1}\left|\widetilde{W}_p\right|^{2+\varsigma}\left(\left|\gamma_{\mathsf{ted}}^{\top}Z\right|^{2+\varsigma}+\left|\gamma_{\mathsf{ted}}^{\top}g_Z\left(X\right)\right|^{2+\varsigma}\right)\right]}{(nh)^{\frac{\varsigma}{2}}},
\end{aligned}$$

where the inequalities follow from Loève's c_r inequality. By (S5), change of variables and Markov's inequality, the numerators of the last two terms are $O\left(1\right)$. Therefore, we have verified Lyapunov's condition

$$\sum_i\mathrm{E}\left[\left|\frac{1}{\sqrt{nh}}\left(\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right)\right|^{2+\varsigma}\right]\to 0.$$

By Lyapunov's central limit theorem,

$$\frac{\frac{1}{\sqrt{nh}}\sum_i\left(\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right)}{\sqrt{\operatorname{Var}\left[\frac{1}{\sqrt{nh}}\sum_i\left(\ddot{W}_{p,i}\left(Y_i-g_Y\left(X_i\right)\right)-\left(\frac{\varpi_p}{\omega_p^{0,2}}\right)\widetilde{W}_{p,i}\gamma_{\mathsf{ted}}^{\top}\left(Z_i-g_Z\left(X_i\right)\right)\right)\right]}}\to_d \mathrm{N}\left(0,1\right).$$

The conclusion follows from this result, (S139), (S140), (S141) and Slutsky's lemma. ∎

S9 Generalized balancing and regression adjustment

Clearly, we have $D_{-2}(w_1,...,w_n \parallel 1/n,...,1/n) = (2n)^{-1}\sum_i \left\{(n\cdot w_i)^2 - 1\right\}$. Consider the problem

$$\begin{aligned} &\min_{w_1,\ldots,w_n} \frac{1}{2n}\sum_i \left\{(n\cdot w_i)^2 - 1\right\} \\ &\text{subject to } \sum_i w_i \ \widehat{W}_{p,i}\bar{Z}_i \ = 0, \sum_i w_i = 1. \end{aligned} \tag{S142}$$

Define the Lagrangian function

$$\mathcal{L}(w_1,...,w_n,\lambda,\mu) := \frac{1}{2n}\sum_i \left\{(n\cdot w_i)^2 - 1\right\} - \lambda^\top\left(\sum_i w_i \ \widehat{W}_{p,i}\bar{Z}_i \ \right) - \mu\left(\sum_i w_i - 1\right).$$

Then by the first-order conditions

$$\begin{aligned} n\cdot w_i - \lambda^\top\left(\widehat{W}_{p,i}\bar{Z}_i\right) - \mu &= 0 \\ \sum_i w_i \ \widehat{W}_{p,i}\bar{Z}_i &= 0 \\ \sum_i w_i - 1 &= 0 \end{aligned}$$

and partialing out μ, we find the dual characterization of the solution $\left(w^{\mathsf{gb}}_{-2,1},...,w^{\mathsf{gb}}_{-2,n}\right)$ to (S142):

$$w^{\mathsf{gb}}_{-2,i} = \frac{1}{n}\left(1 + \lambda_{\mathsf{gb}}^\top\left(\widehat{W}_{p,i}\bar{Z}_i - h\cdot\widehat{\Psi}_{\bar{Z}}\right)\right),$$

where

$$\lambda^{\mathsf{gb}}_p := -\left(\widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2} - h\cdot\widehat{\Psi}_{\bar{Z}}\widehat{\Psi}_{\bar{Z}}^\top\right)^{-1}\widehat{\Psi}_{\bar{Z}}.$$

Then by writing $\widehat{\Psi}_{ZZ^\top,2} - h\cdot\widehat{\Psi}_Z\widehat{\Psi}_Z^\top$ as a block matrix and inverting,

$$\begin{aligned} \frac{1}{h}\sum_i w^{\mathsf{gb}}_{-2,i}\widehat{W}_{p,i}Y_i &= \widehat{\Psi}_Y + \left(\widehat{\Psi}_{Y\bar{Z},2} - h\cdot\widehat{\Psi}_Y\widehat{\Psi}_{\bar{Z}}\right)^\top\lambda^{\mathsf{gb}}_p \\ &= \widehat{\Psi}_Y - \left(\widehat{\Psi}_{Y\bar{Z},2} - h\cdot\widehat{\Psi}_Y\widehat{\Psi}_{\bar{Z}}\right)^\top\left(\widehat{\Psi}_{\bar{Z}\bar{Z}^\top,2} - h\cdot\widehat{\Psi}_{\bar{Z}}\widehat{\Psi}_{\bar{Z}}^\top\right)^{-1}\widehat{\Psi}_{\bar{Z}} \\ &= \widehat{\Psi}_Y - \left(\widehat{\gamma}^{\mathsf{gb}}_Y\right)^\top\widehat{\Psi}_Z, \end{aligned}$$

where

$$\widehat{\gamma}^{\mathsf{gb}}_Y := \left(\widehat{\Psi}_{ZZ^\top,2} - \frac{\widehat{\Psi}_{Z,2}\widehat{\Psi}_{Z^\top,2}}{\widehat{\Psi}_2} - h\cdot\widehat{\Psi}_Z\widehat{\Psi}_Z^\top\right)^{-1}\left(\widehat{\Psi}_{YZ,2} - \frac{\widehat{\Psi}_{Y,2}\widehat{\Psi}_{Z,2}}{\widehat{\Psi}_2} - h\cdot\widehat{\Psi}_Y\widehat{\Psi}_Z\right).$$

Therefore,

$$\frac{1}{h}\sum_i w^{\mathsf{gb}}_{-2,i}\widehat{W}_{p,i}Y_i = \frac{1}{nh}\sum_i \widehat{W}_{p,i}\left(Y_i - Z_i^\top \widehat{\gamma}^{\mathsf{gb}}_Y\right).$$

By Lemma 1(a,b), $\widehat{\gamma}^{\mathsf{gb}}_Y \to_p \gamma_Y$.

S10 Covariate-adjusted inference on the treatment effect derivative

This section describes how to conduct inference using the EB TED estimator $\widehat{\pi}^{\mathsf{eb}}_p$ proposed in Section 6 of the main text. To calculate the standard error of $\widehat{\pi}^{\mathsf{eb}}_p$, we introduce the following notations:

$$\widehat{U}_{Y,i} \coloneqq \mathbb{1}\left(X_i > 0\right)\left(Y_i - r_p^\top\left(X_i\right)\widehat{\beta}_{Y,p,+}\right) + \mathbb{1}\left(X_i < 0\right)\left(Y_i - r_p^\top\left(X_i\right)\widehat{\beta}_{Y,p,-}\right) \tag{S143}$$

$$\widehat{U}^\top_{Z,i} \coloneqq \mathbb{1}\left(X_i > 0\right)\left(Z_i^\top - r_p^\top\left(X_i\right)\widehat{\beta}_{Z,p,+}\right) + \mathbb{1}\left(X_i < 0\right)\left(Z_i - r_p^\top\left(X_i\right)\widehat{\beta}_{Z,p,-}\right), \tag{S144}$$

where

$$\widehat{\beta}_{Y,p,+} \coloneqq \operatorname*{argmin}_{b\in\mathbb{R}^{p+1}} \sum_i \mathbb{1}\left(X_i > 0\right) K\left(\frac{X_i}{h}\right)\left\{Y_i - r_p^\top\left(\frac{X_i}{h}\right)\mathrm{H}b\right\}$$

is solved as

$$\widehat{\beta}_{Y,p,+} = \mathrm{H}^{-1}\left\{\frac{1}{nh}\sum_i \widehat{\Pi}^{-1}_{p,+} r_p\left(\frac{X_i}{h}\right) K\left(\frac{X_i}{h}\right)\mathbb{1}\left(X_i > 0\right)Y_i\right\}. \tag{S145}$$

Analogously, $\widehat{\beta}_{Y,p,-}$ is given by replacing $\mathbb{1}\left(X_i > 0\right)$ in (S145) with $\mathbb{1}\left(X_i < 0\right)$, $\widehat{\beta}_{Z,p,+}$ given by replacing Y_i with $Z_i^\top$, and $\widehat{\beta}_{Z,p,-}$ given by replacing both. Recalling the notations $\widehat{\Psi}_k \coloneqq (nh)^{-1}\sum_i \widehat{W}^k_{p,i}$, $\widehat{\Psi}_{A,k} \coloneqq (nh)^{-1}\sum_i \widehat{W}^k_{p,i}A_i$ and $\dot{\Psi}_A \coloneqq (nh)^{-1}\sum_i \widehat{W}_{p,i}\dot{W}_{p,i}A_i$ defined in Sections S1 and S8, we construct the following estimator $\widehat{\gamma}_{\mathsf{ted}}$ for $\varpi_p\gamma_{\mathsf{ted}}/\omega^{0,2}_p$:

$$\widehat{\gamma}_{\mathsf{ted}} = \left(\widehat{\Psi}_{ZZ^\top,2} - \frac{\widehat{\Psi}_{Z,2}\Psi_{Z^\top,2}}{\widehat{\Psi}_2}\right)^{-1}\left(\dot{\Psi}_{YZ} - \frac{\widehat{\Psi}_{Z,2}\dot{\Psi}_Y}{\widehat{\Psi}_2}\right).$$

Assume that the bandwidth satisfies $nh^{2p+3} = o\left(1\right)$ and $nh^3 \to \infty$, we have

$$\frac{\widehat{\pi}^{\mathsf{eb}}_p - \pi_{\mathsf{srd}}}{\sqrt{\left(nh^3\right)^{-1}\widehat{\mathscr{V}}_{\mathsf{ted}}}} \to_d \mathrm{N}\left(0,1\right),$$

where

$$
\begin{aligned}
\widehat{\mathscr{V}}_{+}^{\mathsf{ted}} &:= \frac{1}{nh}\sum_{i=1}^{n}\left\{\dot{W}_{p;+,i}^{2}\cdot\widehat{U}_{Y,i}^{2}+\widehat{W}_{p;+,i}^{2}\cdot\widehat{\gamma}_{\mathsf{ted}}^{\top}\left(\widehat{U}_{Z,i}\widehat{U}_{Z,i}^{\top}\right)\widehat{\gamma}_{\mathsf{ted}}-2\cdot\dot{W}_{p;+,i}\widehat{W}_{p;+,i}\cdot\widehat{U}_{Y,i}\left(\widehat{U}_{Z,i}^{\top}\widehat{\gamma}_{\mathsf{ted}}\right)\right\},\\
\widehat{\mathscr{V}}_{-}^{\mathsf{ted}} &:= \frac{1}{nh}\sum_{i=1}^{n}\left\{\dot{W}_{p;-,i}^{2}\cdot\widehat{U}_{Y,i}^{2}+\widehat{W}_{p;-,i}^{2}\cdot\widehat{\gamma}_{\mathsf{ted}}^{\top}\left(\widehat{U}_{Z,i}\widehat{U}_{Z,i}^{\top}\right)\widehat{\gamma}_{\mathsf{ted}}-2\cdot\dot{W}_{p;-,i}\widehat{W}_{p;-,i}\cdot\widehat{U}_{Y,i}\left(\widehat{U}_{Z,i}^{\top}\widehat{\gamma}_{\mathsf{ted}}\right)\right\},\\
\widehat{\mathscr{V}}^{\mathsf{ted}} &:= \widehat{\mathscr{V}}_{+}^{\mathsf{ted}}+\widehat{\mathscr{V}}_{-}^{\mathsf{ted}}.
\end{aligned}
$$

References

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics 186*(2), 345 366.

Bhattacharya, R. N. (1977). Refinements of the Multidimensional Central Limit Theorem and Applications. *The Annals of Probability 5*(1), 1–27.

Bickel, P. J. and K. A. Doksum (2015). *Mathematical statistics: basic ideas and selected topics*, Volume 2. CRC Press.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2022). Coverage error optimal confidence intervals for local polynomial regression. *Bernoulli*.

Chen, S. X. and H. Cui (2007). On the second-order properties of empirical likelihood with moment restrictions. *Journal of Econometrics 141*(2), 492–516.

Chen, X. and K. Kato (2020). Jackknife multiplier bootstrap: finite sample approximations to the u-process supremum with applications. *Probability Theory and Related Fields 176*(3-4), 1–67.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014a). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics 42*(5), 1787–1818.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014b). Gaussian approximation of suprema of empirical processes. *Annals of Statistics 42*(4), 1564–1597.

Chernozhukov, V., D. Chetverikov, and K. Kato (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Processes and their Applications 126*(12), 3632–3651.

Cohen, J. D. (1988). Noncentral chi-square: Some observations on recurrence. *The American Statistician 42*(2), 120–122.

DiCiccio, T., P. Hall, and J. Romano (1988). Bartlett adjustments for empirical likelihood. Technical report No. 298, Department of Statistics, Stanford University.

Dudley, R. M. (2002). *Real Analysis and Probability.* Cambridge University Press.

Giné, E. and R. Nickl (2015). *Mathematical foundations of infinite-dimensional statistical models*, Volume 40. Cambridge University Press.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference.* Springer Science & Business Media.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237 249.

Owen, A. B. (2001). *Empirical likelihood.* CRC press.

Skovgaard, I. M. (1981). Transformation of an edgeworth expansion by a sequence of smooth functions. *Scandinavian Journal of Statistics*, 207–217.

Skovgaard, I. M. (1986). On multivariate edgeworth expansions. *International Statistical Review/Revue Internationale de Statistique*, 169–186.