Supplement to "Complementarity and aggregate implications of assortative matching: A nonparametric analysis"

(*Quantitative Economics*, Vol. 5, No. 1, March 2014, 29–66)

Bryan S. Graham
Department of Economics, University of California—Berkeley and NBER

Guido W. Imbens
Graduate School of Business, Stanford University

Geert Ridder
Department of Economics, University of Southern California

This supplemental appendix gives proofs for the theorems in the main text. All notation is as defined in the main text unless noted otherwise.

Appendix A: Additional lemmas and theorems

In this appendix, we state a number of preliminary results that will be used in the proofs of the four theorems in the main text. Detailed proofs of the results listed below may be found in Appendix C below. The proofs of the main results are given in Appendix B.

The proof of Theorem 4.1 uses Lemmas A.1 and A.2 below. Theorems 5.1 and 5.2 use Lemmas A.3–A.12, A.14, A.15, Theorem A.2, and Lemmas A.16–A.18. Theorem 5.3 uses Lemmas A.9–A.12, A.14, Theorem A.3, and Lemmas A.24–A.28. Theorem 5.4 uses Lemmas A.9–A.14, Theorem A.1, and Lemmas A.19–A.23.

Definition A.1 (Sobolev Norm). The norm that we use for functions $g:\mathbb{Z}\subset\mathbb{R}^L\to\mathbb{R}$ that are at least j times continuously differentiable is the Sobolev norm

$$|g|_j=\sup_{z\in\mathbb{Z},|\lambda|\le j}\left|\frac{\partial g^{|\lambda|}}{\partial z^{\lambda}}(z)\right|.$$

Lemma A.1. *Let $f:\mathbb{X}\mapsto\mathbb{R}$, with $\mathbb{X}=[x_l,x_u]$ a compact subset of $\mathbb{R}$, be a twice continuously differentiable function, and let $g:\mathbb{R}\mapsto\mathbb{R}$ satisfy a Lipschitz condition, $|g(x+y)-g(x)|\le c\cdot|y|$. Then*

$$\left|f\bigl(g(\lambda)\bigr)-\left(f\bigl(g(0)\bigr)+\frac{\partial}{\partial x}f\bigl(g(0)\bigr)\cdot\bigl(g(\lambda)-g(0)\bigr)\right)\right|\le\frac{1}{2}\cdot\sup_{x\in\mathbb{X}}\left|\frac{\partial^2}{\partial x^2}f(x)\right|\cdot c^2\cdot\lambda^2.$$

Bryan S. Graham: bgraham@econ.berkeley.edu
Guido W. Imbens: imbens@stanford.edu
Geert Ridder: ridder@usc.edu

DOI: 10.3982/QE45

Lemma A.2. *Let X be a real-valued random variable with support $\mathbb{X} = [x_l, x_u]$, with density $f_X(x) > 0$ for all $x \in \mathbb{X}$, and let $h : \mathbb{X} \mapsto \mathbb{R}$ be a continuous function. Suppose that $\mathbb{E}[|h(X) \cdot X]$ is finite. Then*

$$\mathrm{Cov}\big(h(X), X\big) = \mathbb{E}\left[\frac{\partial}{\partial x} h(X) \cdot \gamma(X)\right],$$

where

$$\gamma(x) = \frac{F_X(x) \cdot (1 - F_X(x)}{f_X(x)} \cdot \big(\mathbb{E}[X|X > x] - \mathbb{E}[X|X \leq x]\big)$$

and $F_X(x)$ is the cumulative distribution function of X.

For completeness, we state a couple of results from Athey and Imbens (2006; AI hereafter).

Lemma A.3 (Lemma A.2 in AI). *Suppose Y is a real-valued, continuously distributed random variable with compact support $\mathbb{Y} = [y_l, y_u]$, with the probability density function $f_Y(y)$ continuous, bounded, and bounded away from zero, on $\mathbb{Y}$. Then, for any $\delta < 1/2$,*

$$\sup_{y \in \mathbb{Y}} N^{\delta} \cdot \big|\hat{F}_Y(y) - F_Y(y)\big| \xrightarrow{p} 0.$$

Lemma A.4 (Lemma A.3 in AI). *Suppose Y is a real-valued, continuously distributed random variable with compact support $\mathbb{Y} = [y_l, y_u]$, with the probability density function $f_Y(y)$ continuous, bounded, and bounded away from zero, on $\mathbb{Y}$. Then, for any $\delta < 1/2$,*

$$\sup_{q \in [0,1]} N^{\delta} \cdot \big|\hat{F}_Y^{-1}(q) - F_Y^{-1}(q)\big| \xrightarrow{p} 0.$$

Lemma A.5 (Lemma A.5 in AI). *Suppose Y is a real-valued random variable with compact support $\mathbb{Y} = [y_l, y_u]$, and suppose that the cumulative distribution function $F_Y(y)$ is twice continuously differentiable on $\mathbb{Y}$, with its first derivative $f_Y(y) = \frac{\partial F_Y}{\partial y}(y)$ bounded away from zero on $\mathbb{Y}$. Then, for $0 < \eta < 3/4$ and $\delta > \max(2\eta - 1, \eta/2)$,*

$$\sup_{y \in \mathbb{Y}, x \leq N^{-\delta}, x+y \in \mathbb{Y}} N^{\eta} \cdot \big|\hat{F}_Y(y + x) - \hat{F}_Y(y) - \big(F_Y(y + x) - F_Y(y)\big)\big| \xrightarrow{p} 0.$$

Lemma A.6 (Lemma A.6 in AI). *Suppose Y is a real-valued random variable with compact support $\mathbb{Y} = [y_l, y_u]$, and suppose that the cumulative distribution function $F_Y(y)$ is twice continuously differentiable on $\mathbb{Y}$, with its first derivative $f_Y(y) = \frac{\partial F_Y}{\partial y}(y)$ bounded away from zero on $\mathbb{Y}$. Then, for all $0 < \eta < 5/7$,*

$$\sup_{q \in [0,1]} N^{\eta} \cdot \left|\hat{F}_Y^{-1}(q) - F_Y^{-1}(q) + \frac{1}{f_Y(F_Y^{-1}(q))}\big(\hat{F}_Y\big(F_Y^{-1}(q)\big) - q\big)\right| \xrightarrow{p} 0.$$

LEMMA A.7. *Suppose X and Y are real-valued, continuously distributed, random variables with compact support $\mathbb{Y} = [y_l, y_u]$ and $\mathbb{X} = [x_l, x_u]$, with the probability density functions $f_Y(y)$ and $f_X(x)$ continuous, bounded, and bounded away from zero, on $\mathbb{Y}$ and $\mathbb{X}$. Then, for any $\delta < 1/2$,*

$$\sup_{x \in \mathbb{X}} N^{\delta} \cdot \left| \hat{F}_Y^{-1}\big(\hat{F}_X(x)\big) - F_Y^{-1}\big(F_X(x)\big) \right| \xrightarrow{p} 0.$$

LEMMA A.8. *Suppose Y is a real-valued random variable with compact support $\mathbb{Y} = [y_l, y_u]$, and the cumulative distribution function $F_Y(y)$ is twice continuously differentiable on $\mathbb{Y}$, with its first derivative $f_Y(y) = \frac{\partial F_Y}{\partial y}(y)$ bounded away from zero on $\mathbb{Y}$. Then, for $0 < \eta < 3/4$ and $\delta > \max(2\eta - 1, \eta/2)$,*

$$\sup_{y \in \mathbb{Y}, x \le N^{-\delta}, x+y \in \mathbb{Y}} N^{\eta} \cdot \left| \hat{F}_Y(y + x) - \hat{F}_Y(y) - f_Y(y) \cdot x \right| \xrightarrow{p} 0.$$

The next three lemmas are given without proof. Proofs can be found in Imbens and Ridder (2009). The first gives a bound on the bias of the NIP estimator.

LEMMA A.9 (Bias). *If for $m = 1, 2$ Assumptions 4.1–5.1 hold, and $q \ge 2s - 1$ and $r \ge s - 1$, then*

$$\sup_{z \in \mathbb{Z}} \left| \mathbb{E}\big[\hat{h}_{m,\text{nip},s}(z)\big] - h_m(z) \right| = O\big(b^s\big).$$

Note that by matching the order of the kernel and the degree of the polynomial in the NIP estimator, we obtain the same reduction in the bias on the full support as on the internal region, that is, the NIP estimator has a bias that is of the same order as that of the NW estimator on the internal region. The variance is bounded in the following lemma. We only use the following two results for the case with $L = 2$, but for convenience we give the general results.

LEMMA A.10 (Variance). *If Assumptions 4.1–5.1 hold and $q \ge s - 1$, $r \ge s - 1 + L$, then*

$$\sup_{z \in \mathbb{Z}} \left| \hat{h}_{m,\text{nip},s}(z) - \mathbb{E}\big[\hat{h}_{m,\text{nip},s}(z)\big] \right| = O_p\left(\left(\frac{\log N}{N b_N^L} \right)^{1/2} \right).$$

This is the same bound as for the NW estimator on the internal set.

The two lemmas imply a uniform rate for the NIP estimator.

LEMMA A.11 (Uniform Convergence). *If Assumptions 4.1–5.1 hold and $q \ge 2s - 1$, $r \ge s - 1 + L$, then*

$$\sup_{z \in \mathbb{Z}} \left| \hat{h}_{m,\text{nip},s}(z) - h_m(z) \right| = O_p\left(\left(\frac{\log N}{N \cdot b_N^L} \right)^{1/2} + b_N^s \right).$$

Lemma A.12. *If $\hat{h}(z)$ is a nonparametric estimator of $h(z)$, then*

$$\inf_{z\in\mathbb{Z}}\left|\hat{h}(z)\right| = \inf_{z\in\mathbb{Z}}\left|h(z)\right| + O_p\Big(\sup_{z\in\mathbb{Z}}\left|\hat{h}(z)-h(z)\right|\Big).$$

Therefore, if $\sup_{z\in\mathbb{Z}}|\hat{h}(z)-h(z)| = o_p(1)$ and $\inf_{z\in\mathbb{Z}}|h(z)|>0$, then $\inf_{z\in\mathbb{Z}}|\hat{h}(z)|$ converges in probability to a positive number. This lemma is useful if $\hat{h}(z)$ appears in the denominator. In this paper, $z=(w,x)$ or $z=w$.

Lemma A.13. *Suppose Assumptions* 4.1–5.2 *hold. Moreover, suppose that in these assumptions $q\geq 2s-1$, $r\geq s$. Then*

$$\sup_{w\in\mathbb{W}}\left|\hat{m}(w)-m(w)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N}\right)^{1/2} + b_N^s\right).$$

Lemma A.14. *Suppose Assumptions* 4.1–5.2 *hold. Moreover, suppose that $q\geq 2s+1$ and $r\geq s+3$. Then*

$$\text{(i)}\quad \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\hat{g}(w,x)-g(w,x)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^2}\right)^{1/2} + b_N^s\right),$$

$$\text{(ii)}\quad \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial\hat{g}}{\partial w}(w,x)-\frac{\partial g}{\partial w}(w,x)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^4}\right)^{1/2} + b_N^s\right),$$

and

$$\text{(iii)}\quad \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2\hat{g}}{\partial w^2}(w,x)-\frac{\partial^2 g}{\partial w^2}(w,x)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^6}\right)^{1/2} + b_N^s\right).$$

The next lemma shows that we can separate out the uncertainty in $\hat{\beta}^{\text{pam}}$ into five components: the uncertainty from estimating $g(\cdot)$, the uncertainty from estimating $\hat{F}_W^{-1}(\cdot)$, the uncertainty from estimating $\hat{F}_X(\cdot)$, the uncertainty from averaging $g(F_W^{-1}(F_X(X_i)),X_i)$ over the sample, and a remainder term that is $o_p(N^{-1/2})$. As defined in Section 6,

$$\hat{\beta}_g^{\text{pam}} = \frac{1}{N}\sum_{i=1}^N \hat{g}\big(F_W^{-1}\big(F_X(X_i)\big),X_i\big),$$

$$\hat{\beta}_W^{\text{pam}} = \frac{1}{N}\sum_{i=1}^N g\big(\hat{F}_W^{-1}\big(F_X(X_i)\big),X_i\big),$$

$$\hat{\beta}_X^{\text{pam}} = \frac{1}{N}\sum_{i=1}^N g\big(F_W^{-1}\big(\hat{F}_X(X_i)\big),X_i\big),$$

and

$$\overline{g}^{\text{pam}} = \frac{1}{N}\sum_{i=1}^N g\big(F_W^{-1}\big(F_X(X_i)\big),X_i\big).$$

Lemma A.15. *Suppose Assumptions* 4.1, 5.1, *and* 5.2 *hold with* $q \geq 2s+1$, $r \geq s+3$, *and* $0 \leq \delta < 1/6$. *Then*

$$\hat{\beta}^{\text{pam}} - \beta^{\text{pam}} = \big(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\big) + \big(\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}\big) + \big(\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}\big) + \big(\overline{g}^{\text{pam}} - \beta^{\text{pam}}\big) + o_p\big(N^{-1/2}\big). \tag{A.1}$$

The next two results are special cases of theorems in Imbens and Ridder (2009; IR hereafter). The first one refers to the full mean case, and focuses on the case where we take full means of regression functions and their first derivatives. The second result focuses on partial means of regression functions. The results in IR allow for more general dependence on higher order derivatives, even in the partial mean case. Here we also restrict the analysis to the case where the regressors are the pair (W_i, X_i). We also state the conditions that IR invoke.

Let $Z_i = (W_i, X_i)$, with $X_i \in \mathbb{X} \subset \mathbb{R}^{L_X}$, $W_i \in \mathbb{W} \subset \mathbb{R}^{L_W}$, $Z_i \in \mathbb{W} \times \mathbb{X} \subset \mathbb{R}^{L_Z}$, with $L_Z = L_X + L_W$. As before, $h(z) = (h_1(z), h_2(z))'$, with $h_1(z) = f_Z(z)$, and $h_2(z) = \mathbb{E}[Y|Z=z] \cdot f_Z(z)$. Let $n : \mathbb{R}^K \mapsto \mathbb{R}$, $tx : \mathbb{X} \mapsto \mathbb{W}$, and $\omega : \mathbb{X} \mapsto \mathbb{R}$, and define $\tilde{Y} = (\tilde{Y}_{i1}\ \tilde{Y}_{i2})'$, with $\tilde{Y}_{i1} = 1$ and $\tilde{Y}_{i2} = Y_i$. We are interested in full means (possibly depending on derivatives) of the regression function,

$$\theta^{\text{fm}} = \mathbb{E}\big[\omega(Z) n\big(h^{[\lambda]}(Z)\big)\big], \tag{A.2}$$

or partial means,

$$\theta^{\text{pm}} = \mathbb{E}\big[\omega(X) n\big(h\big(X, t(X)\big)\big)\big]. \tag{A.3}$$

Note that in the full mean case, $\omega : \mathbb{Z} \mapsto \mathbb{R}$, and in the partial mean case, $\omega : \mathbb{X} \mapsto \mathbb{R}$: the weight function depends only on the covariates that are being averaged over. In the full mean example, $h^{[\lambda]}$ denotes the vector with elements including all derivatives $h^{(\mu)}$ for $\mu \leq \lambda$. The estimators we focus on are

$$\hat{\theta}^{\text{fm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(Z_i) n\big(\hat{h}^{[\lambda]}_{\text{nip},s}(Z_i)\big) \quad \text{and} \quad \hat{\theta}^{\text{pm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(X_i) n\big(\hat{h}_{\text{nip},s}\big(t(X_i), X_i\big)\big).$$

It will also be useful to define the averages over the true regression functions and their derivatives,

$$\overline{\theta}^{\text{fm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(Z_i) n\big(h^{[\lambda]}(Z_i)\big) \quad \text{and} \quad \overline{\theta}^{\text{pm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(X_i) n\big(h\big(t(X_i), X_i\big)\big).$$

Assumption A.1 (Distribution).

(i) *The random vectors* $(Y_1, Z_1), (Y_2, Z_2), \ldots,$ *are independent and identically distributed.*

(ii) *The support of* Z *is* $\mathbb{Z} \subset \mathbb{R}^L$, $\mathbb{Z} = \bigotimes_{m=1}^{L} [z_{ml}, z_{mu}]$, $z_{ll} < z_{ul}$ *for all* $l = 1, \ldots, L$.

(iii) *We have* $\sup_{z \in \mathbb{Z}} \mathbb{E}[|Y|^p | Z = z] < \infty$.

(iv) *The function* $g(z)=\mathbb{E}[Y|Z=z]$ *is* q *times continuously differentiable on the interior of* $\mathbb{Z}$ *with the* q*th derivative bounded.*

(v) *The density* $f_Z(z)$ *is bounded and bounded away from zero on* $\mathbb{Z}$, *and is* q *times continuously differentiable on the interior of* $\mathbb{Z}$ *with the* q*th derivative bounded.*

Assumption A.2 (Kernel).

(i) *We have* $K:\mathbb{R}^L\to\mathbb{R}$, *with* $K(u)=\prod_{l=1}^{L}\mathcal{K}(u_l)$.

(ii) *We have* $K(u)=0$ *for* $u\notin\mathbb{U}$, *with* $\mathbb{U}=[-1,1]^L$, $\mathbb{U}_1=[-1,1]^{L_W}$, *and* $\mathbb{U}_2=[-1,1]^{L_X}$.

(iii) *The kernel* K *is* r *times continuously differentiable, with the* r*th derivative bounded on the interior of* $\mathbb{U}$.

(iv) *The kernel* K *is a kernel of order* s, *so that* $\int_{\mathbb{U}}K(u)\,\mathrm{d}u=1$ *and* $\int_{\mathbb{U}}u^{\lambda}K(u)\,\mathrm{d}u=0$ *for all* λ *such that* $0<|\lambda|<s$ *for some* $s\geq 1$.

(v) *The kernel* K *is a kernel of derivative order* d.

Assumption A.3. *The bandwidth* $b_N=N^{-\delta}$ *for some* $\delta>0$.

Assumption A.4 (Smoothness of n and ω).

(i) *The function* n *is* t *times continuously differentiable with its* t*th derivative bounded.*

(ii) *The function* ω *is* t *times differentiable on* $\mathbb{X}$ *with bounded* t*th derivative, and* $\frac{\partial^{\mu}\omega}{\partial z^{\mu}}(z)$ *is zero on the boundary of* $\mathbb{Z}$.

Assumption A.5 (Smoothness of t). *The function* $t:\mathbb{X}\mapsto\mathbb{W}$ *is twice continuously differentiable on* $\mathbb{X}$ *with its first derivative positive, bounded, and bounded away from zero.*

Theorem A.1 (Generalized Full Mean and Average Derivative (Theorem 4.2, IR)). *If Assumptions A.1, A.2, A.3, and A.4 hold with* $q\geq|\lambda|+2s-1$, $r\geq|\lambda|+s-1+L$, $t\geq|\lambda|+s$, $p\geq 3$, $d\geq\max\{\lambda_1,\dots,\lambda_L\}+s-1$, *all* $\mu\leq\lambda$, $0\leq|\mu|\leq|\lambda|-1$, *and*

$$\frac{1}{2s}<\delta<\min\left\{\frac{2-\dfrac{4}{p}}{2L+4\max\{1,|\lambda|\}},\frac{1}{2L+4|\lambda|}\right\},$$

then $\hat{\theta}_{\mathrm{fm}}$ *is asymptotically linear with*

$$\begin{aligned}\sqrt{N}\big(\hat{\theta}^{\mathrm{fm}}-\theta^{\mathrm{fm}}\big)&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\big(\omega(Z_i)n\big(h^{[\lambda]}(Z_i)\big)-\mathbb{E}\big[\omega(Z_i)n\big(h^{[\lambda]}(Z_i)\big)\big]\big)\\&\quad+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{\kappa\leq\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\big(\alpha_{\kappa m}^{(\kappa)}(X_i)\tilde{Y}_{im}-\mathbb{E}\big[\alpha_{\kappa m}^{(\kappa)}(X)\tilde{Y}_m\big]\big)\right)\\&\quad+o_p(1),\end{aligned}$$

with

$$\alpha^{(\kappa)}_{\kappa 1}(z) = f_X(z)\omega(z)\frac{\partial n}{\partial h^{(\kappa)}_1(z)}\big(h^{[\lambda]}(z)\big) \quad \textit{and}$$

$$\alpha^{(\kappa)}_{\kappa 2}(z) = f_X(z)\omega(z)\frac{\partial n}{\partial h^{(\kappa)}_2(z)}\big(h^{[\lambda]}(z)\big),$$

and $\tilde{Y} = (\tilde{Y}_{i1}\ \tilde{Y}_{i2})'$, *with* $\tilde{Y}_{i1} = 1$ *and* $\tilde{Y}_{i2} = Y_i$.

The second theorem from IR gives the asymptotic properties of the generalized partial mean (GPM) estimators

Theorem A.2 (Generalized Partial Mean (Theorem 4.3, IR)). *If Assumptions* A.1, A.2, A.3, A.4, *and* A.5 *hold with* $q \geq 2s-1$, $r \geq s-1+L$, $t \geq s$, $p \geq 4$, $d \geq s-1$, *and*

$$\frac{1}{2s} < \delta < \min\left\{\frac{2-\frac{4}{p}}{2L+4}, \frac{1}{2L}\right\},$$

then $\hat{\theta}^{\mathrm{pm}}$ *is asymptotically linear with*

$$\begin{aligned}
&\sqrt{N}\big(\hat{\theta}^{\mathrm{pm}} - \theta^{\mathrm{pm}}\big) \\
&\quad = \sqrt{N}\cdot\big(\overline{\theta}^{\mathrm{pm}} - \theta^{\mathrm{pm}}\big) + \frac{1}{b_N^{L_W}\sqrt{N}} \\
&\qquad \cdot \sum_{i=1}^{N}\sum_{m=1}^{2}\Bigg(\alpha_m(X_i)'\tilde{Y}_{im}\int_{\mathbb{U}_2} K\bigg(\frac{W_i - t(X_i)}{b_N} + \frac{\partial}{\partial x}t(X_i)\cdot u_2, u_2\bigg)\,\mathrm{d}u_2 \\
&\qquad - \mathbb{E}\bigg[\alpha_m(X)'\tilde{Y}_m \int_{\mathbb{U}_2} K\bigg(\frac{W - t(X)}{b_N} + \frac{\partial}{\partial x}t(X)\cdot u_2, u_2\bigg)\,\mathrm{d}u_2\bigg]\Bigg) + o_p(1),
\end{aligned}$$

with

$$\alpha_1(x) = f_Z\big(t(x), x\big)\omega(x)\frac{\partial n}{\partial h_1}\big(h\big(t(x), x\big)\big) \quad \textit{and}$$

$$\alpha_2(x) = f_Z\big(t(x), x\big)\omega(x)\frac{\partial n}{\partial h_2}\big(h\big(t(x), x\big)\big).$$

Moreover,

$$\left(\begin{pmatrix} \sqrt{N}\cdot\big(\overline{\theta}^{\mathrm{pm}} - \theta^{\mathrm{pm}}\big) \\ \sqrt{N}b_N^{L_W/2}\big(\hat{\theta}^{\mathrm{pm}} - \overline{\theta}^{\mathrm{pm}}\big)\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} V_1 & 0 \\ 0 & V_2\end{pmatrix}\right),$$

with

$$V_1 = \mathbb{E}\Big[\big(\omega(X)n\big(h\big(t(X), X\big)\big) - \theta_{\mathrm{pm}}\big)^2\Big]$$

and

$$V_2 = \sum_{m=1}^{2}\sum_{m'=1}^{2}\int_{\mathbb{X}} \mu_{mm'}\big(x, t(x)\big)\alpha_m(x)\alpha_{m'}(x)$$
$$\times \int_{\mathbb{U}_2}\left(\int_{\mathbb{U}_1} K\left(u_1, \frac{\partial t}{\partial x}(x)u_1 + u_2\right)\mathrm{d}u_1\right)^2 \mathrm{d}u_2 f_X\big(x, t(x)\big)\,\mathrm{d}x_1,$$

with $\mu_{mm'}(x) = \mathbb{E}[\tilde{Y}_{im}\tilde{Y}_{im'}|X = x]$ *for* $m, m' = 1, 2$.

LEMMA A.16. *Suppose Assumptions* 4.1, 5.1, *and* 5.2 *hold, with* $q \geq 2s-1$, $r \geq s+1$, $p \geq 4$, $d \geq s-1$, *and* $1/(2s) < \delta < 1/8$. *Then*

$$\begin{aligned}
&\sqrt{N}\big(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\big)\\
&\quad= \frac{1}{\sqrt{N}b_N}\sum_{i=1}^{N}\big(Y_i - g(W_i, X_i)\big)\\
&\qquad\cdot\int_{u_2} K\left(\frac{W_i - F_W^{-1}(F_X(X_i))}{b_N} + \frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u_2, u_2\right)\mathrm{d}u_2\\
&\qquad+ \frac{1}{\sqrt{N}b_N}\sum_{i=1}^{N}\Bigg\{\Big(g(W_i, X_i) - g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\Big)\\
&\qquad\cdot\int_{u_2} K\left(\frac{W_i - F_W^{-1}(F_X(X_i))}{b_N} + \frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u_2, u_2\right)\mathrm{d}u_2\\
&\qquad- \mathbb{E}\Bigg[\Big(g(W, X) - g\big(F_W^{-1}\big(F_X(X)\big), X\big)\Big)\\
&\qquad\cdot\int_{u_2} K\left(\frac{W_i - F_W^{-1}(F_X(X))}{b_N} + \frac{f_X(X)}{f_W(F_W^{-1}(F_X(X)))}\cdot u_2, u_2\right)\mathrm{d}u_2\Bigg]\Bigg\}\\
&\qquad+ o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
&\sqrt{N}b_N^{1/2}\big(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\big)\\
&\quad\xrightarrow{d} \mathcal{N}\Bigg(0, \mathbb{E}\Bigg[\sigma^2\big(F_W^{-1}\big(F_X(X)\big), X\big)\\
&\qquad\cdot\int_{u_1}\left(\int_{u_2} K\left(u_1 + \frac{f_X(X)}{f_W(F_W^{-1}(F_X(X)))}\cdot u_2, u_2\right)\mathrm{d}u_2\right)^2\mathrm{d}u_1\\
&\qquad\cdot f_{W|X}\big(F_W^{-1}\big(F_X(X)\big)|X\big)\Bigg]\Bigg).
\end{aligned}$$

LEMMA A.17. *Suppose Assumptions* 4.1, 5.1, *and* 5.2 *hold, with* $q \geq 2$. *Then*

$$\hat{\beta}_W^{\mathrm{pam}} - \overline{g}^{\mathrm{pam}} = \frac{1}{N}\sum_{i=1}^{N} \psi_W^{\mathrm{pam}}(W_i) + o_p\big(N^{-1/2}\big).$$

LEMMA A.18. *Suppose Assumptions* 4.1, 5.1, *and* 5.2 *hold, with* $q \geq 2$. *Then*

$$\hat{\beta}_X^{\mathrm{pam}} - \overline{g}^{\mathrm{pam}} = \frac{1}{N}\sum_{i=1}^{N} \psi_X^{\mathrm{pam}}(X_i) + o_p\big(N^{-1/2}\big).$$

Define

$$\hat{\beta}_g^{\mathrm{lc}} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big),$$

$$\hat{\beta}_m^{\mathrm{lc}} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - \hat{m}(W_i)\big),$$

and

$$\overline{g}^{\mathrm{lc}} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big).$$

LEMMA A.19. *Suppose Assumptions* 4.1, 5.1, *and* 5.2 *hold. Moreover, suppose that the estimators for* $g(w,x)$ *and* $m(w)$, *and* $\hat{g}(w,x)$ *and* $\hat{m}(w)$, *respectively, satisfy*

$$\sup_{w\in\mathbb{W}, x\in\mathbb{X}} \left| \frac{\partial \hat{g}}{\partial w}(w,x) - \frac{\partial g}{\partial w}(w,x) \right| = o_p\big(N^{-\eta}\big) \quad \textit{and}$$

$$\sup_{w\in\mathbb{W}} \big|\hat{m}(w) - m(w)\big| = o_p\big(N^{-\eta}\big)$$

for some $\eta > 1/4$. *Then*

$$\hat{\beta}^{\mathrm{lc}} - \beta^{\mathrm{lc}} = \big(\hat{\beta}_g^{\mathrm{lc}} - \overline{g}^{\mathrm{lc}}\big) + \big(\hat{\beta}_m^{\mathrm{lc}} - \overline{g}^{\mathrm{lc}}\big) + \big(\overline{g}^{\mathrm{lc}} - \beta^{\mathrm{lc}}\big) + o_p\big(N^{-1/2}\big). \tag{A.4}$$

LEMMA A.20. *Suppose Assumptions* 4.1, 5.1, *and* 5.2 *hold, with* $q \geq 2s$, $r \geq s$, $p \geq 3$, $d \geq s$, *and* $1/(2s) < \delta < 1/12$. *Then*

$$\hat{\beta}_g^{\mathrm{lc}} - \overline{g}^{\mathrm{lc}} = \frac{1}{N}\sum_{i=1}^{N} \psi_g^{\mathrm{lc}}(Y_i, W_i, X_i) + o_p\big(N^{-1/2}\big), \tag{A.5}$$

where

$$\begin{aligned}\psi_g^{\mathrm{lc}}(Y,W,X) = &-\frac{1}{f_{W,X}(W,X)}\frac{\partial f_{W,X}(W,X)}{\partial W}\big(Y - g(W,X)\big)d(W)\big(X - m(W)\big) \\ &- \frac{\partial m(W)}{\partial W} d(W)\big(Y - g(W,X)\big) \\ &+ \frac{\partial d}{\partial w}(W)\big(X - m(W)\big)\big(Y - g(W,X)\big).\end{aligned}$$

Lemma A.21. *Suppose Assumptions* 4.1–5.2 *hold, with* $q \geq 2s-1$, $r \geq s$, $p \geq 3$, $d \geq s-1$, *and*

$$\frac{1}{2s} < \delta < \frac{1}{3} - \frac{2}{3p}.$$

Then

$$\sup_{w \in \mathbb{W}} \left| \frac{1}{\hat{f}_W(w)} \big(\hat{f}_W(w) - f_W(w)\big)^2 \right| = o_p\big(N^{-1/2}\big).$$

Lemma A.22. *Let* $h(w) = (h_1(w), h_2(w))' = (\mathbb{E}[X|W=w] f_W(w), f_W(w))'$ *and suppose Assumptions* 4.1–5.2 *hold, with* $q \geq 2s-1$, $r \geq s$, $p \geq 3$, $d \geq s-1$, *and*

$$\frac{1}{2s} < \delta < \frac{1}{3} - \frac{2}{3p}.$$

Then

$$\sup_{w \in \mathbb{W}} \left| \frac{1}{\hat{h}_2(w)} \big(\hat{h}_1(w) - h_1(w)\big)\big(\hat{h}_2(w) - h_2(w)\big) \right| = o_p\big(N^{-1/2}\big).$$

Lemma A.23. *Suppose Assumptions* 4.1–5.2 *hold, with* $q \geq 2s-1$, $r \geq s$, $p \geq 3$, *and*

$$\frac{1}{2s} < \delta < \frac{1}{8}.$$

Then

$$\hat{\beta}_m^{\mathrm{lc}} - \overline{g}^{\mathrm{lc}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\big[g_W(W_i, X_i)|W_i\big] \cdot d(W_i) \cdot \big(X_i - m(W_i)\big) + o_p\big(N^{-1/2}\big). \tag{A.6}$$

Before the next theorem, we need some additional definitions. We split Z_i into $(Z_{i1}', Z_{i2}')'$, with the dimension of Z_{i1} equal to L_{Z1} and the dimension of Z_{i2} equal to L_{Z2}, so that $L = L_{Z1} + L_{Z2}$. We are interested in the distribution of

$$V = \sqrt{N} \cdot \left(\frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} n\big(\hat{h}_{\mathrm{nip},s}(Z_{1j}, Z_{2k})\big) - \frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} n\big(h(Z_{1j}, Z_{2k})\big) \right). \tag{A.7}$$

We show that this is, to first order, equivalent to a single normalized sum.

Theorem A.3. *Suppose that Assumptions* A.1–A.4 *hold, with* $q \geq 2s-1$, $r \geq s-1+L$, $1/(2s) < \delta < 1/(2L)$, *and* $t \geq 2$. *Then*

$$\begin{aligned} V = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \bigg\{ & \frac{\partial n}{\partial h}\big(h(Z_i)\big)' \tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) \\ & - \mathbb{E}_Z\bigg[\frac{\partial n}{\partial h}\big(h(Z)\big)' \tilde{Y} f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i})\bigg]\bigg\} + o_p(1). \end{aligned} \tag{A.8}$$

(*To be clear here, we index the expectation by the random variable that the expectation is taken over, in this case* Z.)

Before stating some additional lemmas that will be used for proving Theorem 5.3, we need some additional definitions. Define

$$\overline{g}^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i, X_j)\frac{\phi_c(\Phi_c^{-1}(F_W(W_i)), \Phi_c^{-1}(F_X(X_j)); \rho)}{\phi_c(\Phi_c^{-1}(F_W(W_i)))\phi_c(\Phi_c^{-1}(F_X(X_j)))},$$

$$\hat{\beta}_g^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{g}(W_i, X_j)\frac{\phi_c(\Phi_c^{-1}(F_W(W_i)), \Phi_c^{-1}(F_X(X_j)); \rho)}{\phi_c(\Phi_c^{-1}(F_W(W_i)))\phi_c(\Phi_c^{-1}(F_X(X_j)))},$$

$$\hat{\beta}_W^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i, X_j)\frac{\phi_c(\Phi_c^{-1}(\hat{F}_W(W_i)), \Phi_c^{-1}(F_X(X_j)); \rho)}{\phi_c(\Phi_c^{-1}(\hat{F}_W(W_i)))\phi_c(\Phi_c^{-1}(F_X(X_j)))},$$

$$\hat{\beta}_X^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i, X_j)\frac{\phi_c(\Phi_c^{-1}(F_W(W_i)), \Phi_c^{-1}(\hat{F}_X(X_j)); \rho)}{\phi_c(\Phi_c^{-1}(F_W(W_i)))\phi_c(\Phi_c^{-1}(\hat{F}_X(X_j)))}.$$

LEMMA A.24. *Suppose Assumptions* 4.1–5.2 *hold with* $q \geq 2s+2$, $r \geq s+3$, *and* $1/(2s) < \delta < 1/4$. *Then*

$$\hat{\beta}^{\text{cm}}(\rho, 0) - \beta^{\text{cm}}(\rho, 0) = \left(\hat{\beta}_g^{\text{cm}} - \overline{g}^{\text{cm}}\right) + \left(\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}}\right) + \left(\hat{\beta}_X^{\text{cm}} - \overline{g}^{\text{cm}}\right) + \left(\overline{g}^{\text{cm}} - \beta^{\text{cm}}(\rho, 0)\right) + o_p\left(N^{-1/2}\right).$$

LEMMA A.25. *Suppose Assumptions* 4.1–5.2 *hold. Then*

$$\hat{\beta}_g^{\text{cm}} - \overline{g}^{\text{cm}} = \frac{1}{N}\sum_{i=1}^{N}\psi_g^{\text{cm}}(Y_i, W_i, X_i) + o_p\left(N^{-1/2}\right),$$

where

$$\psi_g^{\text{cm}}(y, w, x) = \frac{f_W(w)\cdot f_X(x)}{f_{WX}(w, x)}\left(y - g(w, x)\right)\omega(w, x).$$

LEMMA A.26. *Suppose Assumptions* 4.1–5.2 *hold. Then*

$$\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} = \frac{1}{N}\sum_{i=1}^{N}\psi_W^{\text{cm}}(Y_i, W_i, X_i) + o_p\left(N^{-1/2}\right),$$

where

$$\psi_W^{\text{cm}}(y, w, x) = \int\int g(s, t)e_W(s, t)\left(1(w \leq s) - F_W(s)\right)f_W(s)f_X(t)\,\mathrm{d}s\,\mathrm{d}t.$$

LEMMA A.27. *Suppose Assumptions* 4.1–5.2 *hold. Then*

$$\hat{\beta}_X^{\text{cm}} - \overline{g}^{\text{cm}} = \frac{1}{N}\sum_{i=1}^{N}\psi_X^{\text{cm}}(Y_i, W_i, X_i) + o_p\left(N^{-1/2}\right),$$

where

$$\psi_X^{\mathrm{cm}}(y,w,x)=\int\int g(s,t)e_X(s,t)\big(1(x\le t)-F_X(t)\big)f_W(s)f_X(t)\,\mathrm{d}s\,\mathrm{d}t.$$

Lemma A.28. *Suppose Assumptions* 4.1–5.2 *hold. Then*

$$\overline{g}^{\mathrm{cm}}-\beta^{\mathrm{cm}}(\rho,0)=\frac{1}{N}\sum_{i=1}^{N}\psi_0^{\mathrm{cm}}(Y_i,W_i,X_i)+o_p\big(N^{-1/2}\big),$$

where

$$\begin{aligned}\psi_0^{\mathrm{cm}}(w,x)=&\big(\mathbb{E}\big[g(W,x)\cdot\omega(W,x)\big]-\beta^{\mathrm{cm}}(\rho,0)\big)\\&+\big(\mathbb{E}\big[g(w,X)\cdot\omega(w,X)\big]-\beta^{\mathrm{cm}}(\rho,0)\big).\end{aligned}\tag{A.9}$$

The following theorem is a simplifed version of the V-statistics results in Lehmann (1999).

Theorem A.4 (V-Statistics). *Suppose $Z_1,\dots,Z_N$ are independent and identically distributed random vectors with dimension K, with support $\mathbb{Z}\subset\mathbb{R}^K$. Let $\psi:\mathbb{Z}^K\times\mathbb{Z}^K\mapsto\mathbb{R}$ be a real-valued function. Define*

$$\begin{aligned}&\theta=\mathbb{E}\big[\psi(Z_1,Z_2)\big],\qquad \psi_1(z)=\mathbb{E}\big[\psi(z,Z)\big],\qquad \psi_2(z)=\mathbb{E}\big[\psi(Z,z)\big],\\&\sigma^2=\mathrm{Cov}\big(\psi(Z_1,Z_2),\psi(Z_1,Z_3)\big)+\mathrm{Cov}\big(\psi(Z_2,Z_1),\psi(Z_1,Z_3)\big)\\&\qquad+\mathrm{Cov}\big(\psi(Z_1,Z_2),\psi(Z_3,Z_1)\big)+\mathrm{Cov}\big(\psi(Z_2,Z_1),\psi(Z_3,Z_1)\big),\end{aligned}$$

and

$$V=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\psi(Z_i,Z_j).$$

Then if $0<\sigma^2<\infty$,

$$V=\frac{1}{N}\sum_{i=1}^{N}\big\{\big(\psi_1(Z_i)-\theta\big)+\big(\psi_2(Z_i)-\theta\big)\big\}+o_p\big(N^{-1/2}\big)$$

and

$$\sqrt{N}\cdot(V-\theta)\xrightarrow{d}\mathcal{N}\big(0,\sigma^2\big).$$

Appendix B: Proofs of theorems stated in the main text

In this appendix, we provide proofs of the results in the main text. We use the lemmas and theorems stated in Appendix A throughout. Some details that are omitted here may be found in Appendix C.

PROOF OF THEOREM 4.1. Define

$$V_{\lambda,i} = \lambda \cdot X_i \cdot d(W_i) + W_i,$$
$$h(\lambda, a) = \mathrm{pr}(V_\lambda \le a) = F_{V_\lambda}(a), \quad \text{and} \quad k(w, x, \lambda) = h\big(\lambda, \lambda \cdot x \cdot d(w) + w\big).$$

First we focus on

$$\beta^{\mathrm{lr,v}}(\lambda) = \mathbb{E}\big[g\big(F_W^{-1}\big(F_{V_\lambda}(V_{\lambda,i})\big), X\big)\big] = \mathbb{E}\big[g\big(F_W^{-1}\big(k(W_i, X_i, \lambda)\big), X_i\big)\big].$$

We then prove four results. First, we show that for small λ, $\beta^{\mathrm{lr,v}}(\lambda)$ and $\beta^{\mathrm{lr}}(\lambda)$ are close or

$$\beta^{\mathrm{lr,v}}(\lambda) = \beta^{\mathrm{lr}}(\lambda) + o(\lambda). \tag{B.1}$$

Second, we show that

$$\begin{aligned}\beta^{\mathrm{lr,v}}(\lambda) &= \mathbb{E}\big[g(W, X)\big] \\ &\quad + \mathbb{E}\left[\frac{\partial g}{\partial w}(W_i, X_i)\frac{1}{f_W(W_i)}\big(k(W_i, X_i, \lambda) - k(W_i, X_i, 0)\big)\right] + o(\lambda).\end{aligned} \tag{B.2}$$

Next we show that β^{lc} has the two representations in Theorem 4.1. In particular, the third part of the proof shows that $\beta^{\mathrm{lc,v}} = \frac{\partial \beta^{\mathrm{lc,v}}}{\partial \lambda}(0)$ satisfies

$$\beta^{\mathrm{lc,v}} = \mathbb{E}\left[\frac{\partial g}{\partial w}(W_i, X_i) \cdot \big(X_i \cdot d(W_i) - \mathbb{E}\big[X_i \cdot d(W_i)|W_i\big]\big)\right]. \tag{B.3}$$

Fourth, we show that $\beta^{\mathrm{lc,v}}$ satisfies

$$\beta^{\mathrm{lc,v}} = \mathbb{E}\left[\delta(W_i, X_i) \cdot \frac{\partial^2 g}{\partial w\, \partial x}(W_i, X_i)\right]. \tag{B.4}$$

We start with the proof of (B.1). Define

$$u(w, x, \lambda) = \lambda \cdot x \cdot d(w)^{1-|\lambda|} + \sqrt{1-\lambda^2} \cdot w \quad \text{and} \quad v(w, x, \lambda) = \lambda \cdot x \cdot d(w) + w.$$

Then

$$\sup_{w \in \mathbb{W}, x \in \mathbb{X}} \big|u(w, x, \lambda) - v(w, x, \lambda)\big| = O\big(\lambda^2\big).$$

Define also

$$h_U(\lambda, a) = \mathrm{pr}(U_\lambda \le a) \quad \text{and} \quad k_U(w, x, \lambda) = h_U\big(\lambda, u(w, x, \lambda)\big).$$

Then

$$\sup_a \big|h_U(\lambda, a) - h(\lambda, a)\big| = O\big(\lambda^2\big)$$

and

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\big|k_U(w,x,\lambda)-k(w,x,\lambda)\big|=O\big(\lambda^2\big).$$

Combined with the smoothness assumptions, this implies that

$$\begin{aligned}\beta^{\text{lc,v}}(\lambda)-\beta^{\text{lc}}(\lambda)&=\mathbb{E}\big[g\big(F_W^{-1}\big(k(W_i,X_i,\lambda)\big),X_i\big)\big]-\mathbb{E}\big[g\big(F_W^{-1}\big(k_U(W_i,X_i,\lambda)\big),X_i\big)\big]\\&=O\big(\lambda^2\big).\end{aligned}$$

This finishes the proof of (B.1).

Next, we prove (B.2). Let c_1 and c_2 satisfy

$$\sup_{x,w,\gamma,\lambda}\big|k(w,x,\lambda+\gamma)-k(w,x,\lambda)\big|\le c_1\cdot\gamma$$

and

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2}{\partial w^2}g(w,x)\right|\le c_2,$$

respectively. Then applying Lemma A.1 with $f(a)=g(F_W^{-1}(a),x)$ and $h(\lambda)=k(w,x,\lambda)$, we obtain

$$\begin{aligned}&\left|g\big(F_W^{-1}\big(k(w,x,\lambda)\big),x\big)-\left(g\big(F_W^{-1}\big(k(w,x,0)\big),x\big)\right.\right.\\&\qquad\left.\left.+\frac{\frac{\partial}{\partial w}g(F_W^{-1}(k(w,x,0)),x)}{f_W(F_W^{-1}(k(w,x,0)))}\big(k(w,x,\lambda)-k(w,x,0)\big)\right)\right|\\&\quad\le c_2c_1^2\lambda^2=o(\lambda).\end{aligned}$$

Since the bound does not depend on x and w, we can average over W and X, and it follows that

$$\begin{aligned}&\left|\mathbb{E}\big[g\big(F_W^{-1}\big(k(W,X,\lambda)\big),X\big)\big]-\mathbb{E}\big[g(W,X)\big]\right.\\&\quad\left.-\mathbb{E}\left[\frac{\frac{\partial}{\partial w}g(W,X)}{f_W(W)}\big(k(W,X,\lambda)-W\big)\right]\right|=o(\lambda),\end{aligned}$$

where we also use the fact that $k(w,x,0)=F_W(w)$. This finishes the proof of (B.2).

Now we prove (B.3). By definition,

$$\begin{aligned}h(\lambda,a)=\Pr(V_{\lambda,i}<a)&=\Pr(V_{\lambda,i}<a,W_i<w_m)+\Pr(V_{\lambda,i}<a,W_i\ge w_m)\\&=\Pr\big(\lambda\cdot X_i\cdot d(W_i)+W_i\le a,W_i<w_m\big)\\&\quad+\Pr\big(\lambda\cdot X_i\cdot d(W_i)+W_i\le a,W_i\ge w_m\big)\end{aligned}$$

$$
\begin{aligned}
&= \Pr\big(\lambda \cdot X_i \cdot (W_i - w_l) + W_i \leq a, W_i < w_m\big) \\
&\quad + \Pr\big(\lambda \cdot X_i \cdot (w_u - W_i) + W_i \leq a, W_i \geq w_m\big) \\
&= \Pr\left(W_i \leq \min\left(w_m, \frac{a + \lambda \cdot X_i \cdot w_l}{1 + \lambda \cdot X_i}\right)\right) \\
&\quad + \Pr\left(w_m \leq W_i \leq \frac{a - \lambda \cdot X_i \cdot w_u}{1 - \lambda \cdot X_i}\right).
\end{aligned}
$$

For λ sufficiently close to zero, we can write this as

$$
\begin{aligned}
h(\lambda, a) &= 1_{a>w_m} \cdot \Pr(W_i \leq w_m) + 1_{a \leq w_m} \cdot \Pr\left(W_i \leq \frac{a + \lambda \cdot X_i \cdot w_l}{1 + \lambda \cdot X_i}\right) \\
&\quad + 1_{a>w_m} \cdot \Pr\left(w_m < W_i \leq \frac{a - \lambda \cdot X_i \cdot w_u}{1 - \lambda \cdot X_i}\right) \\
&= 1_{a \leq w_m} \cdot \Pr\left(W_i \leq \frac{a + \lambda \cdot X_i \cdot w_l}{1 + \lambda \cdot X_i}\right) \\
&\quad + 1_{a>w_m} \cdot \Pr\left(W_i \leq \frac{a - \lambda \cdot X_i \cdot w_u}{1 - \lambda \cdot X_i}\right) \\
&= 1_{a \leq w_m} \cdot \mathbb{E}\left[\Pr\left(W_i \leq \frac{a + \lambda \cdot X_i \cdot w_l}{1 + \lambda \cdot X_i} \middle| X_i\right)\right] \\
&\quad + 1_{a>w_m} \cdot \mathbb{E}\left[\Pr\left(W_i \leq \frac{a - \lambda \cdot X_i \cdot w_u}{1 - \lambda \cdot X_i} \middle| X_i\right)\right] \\
&= 1_{a \leq w_m} \cdot \int F_{W|X}\left(\frac{a + \lambda \cdot z \cdot w_l}{1 + \lambda \cdot z} \middle| z\right) f_X(z)\,\mathrm{d}z \\
&\quad + 1_{a>w_m} \cdot \int F_{W|X}\left(\frac{a - \lambda \cdot z \cdot w_u}{1 - \lambda \cdot z} \middle| z\right) f_X(z)\,\mathrm{d}z.
\end{aligned}
$$

Substituting $a = \lambda \cdot x \cdot d(w) + w$, we get

$$
\begin{aligned}
k(w, x, \lambda) &= 1_{\lambda \cdot x \cdot d(w) + w \leq w_m} \cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot d(w) + w + \lambda \cdot z \cdot w_l}{1 + \lambda \cdot z} \middle| z\right) f_X(z)\,\mathrm{d}z \\
&\quad + 1_{\lambda \cdot x \cdot d(w) + w > w_m} \cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot d(w) + w - \lambda \cdot z \cdot w_u}{1 - \lambda \cdot z} \middle| z\right) f_X(z)\,\mathrm{d}z \\
&= 1_{\lambda \cdot x \cdot (w - w_l) + w \leq w_m} 1_{w \leq w_m} \\
&\quad \cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w - w_m) + w + \lambda \cdot z \cdot w_l}{1 + \lambda \cdot z} \middle| z\right) f_X(z)\,\mathrm{d}z \\
&\quad + 1_{\lambda \cdot x \cdot (w_u - w) + w \leq w_m} 1_{w > w_m} \\
&\quad \cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w_u - w) + w + \lambda \cdot z \cdot w_l}{1 + \lambda \cdot z} \middle| z\right) f_X(z)\,\mathrm{d}z \\
&\quad + 1_{\lambda \cdot x \cdot (w - w_l) + w > w_m} \cdot 1_{w \leq w_m}
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w - w_m) + w - \lambda \cdot z \cdot w_u}{1 - \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z \\
&+ 1_{\lambda \cdot x \cdot (w_u - w) + w > w_m} \cdot 1_{w > w_m} \\
&\cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w_u - w) + w - \lambda \cdot z \cdot w_u}{1 - \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z \\
= {} & 1_{w \le w_m(1+\lambda x w_l/w_m)/(1+\lambda x)} \\
&\cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w - w_m) + w + \lambda \cdot z \cdot w_l}{1 + \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z \\
&+ 0 \cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w_u - w) + w + \lambda \cdot z \cdot w_l}{1 + \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z \\
&+ 1_{w_m(1+\lambda x w_l/w_m)/(1+\lambda x) \le w \le w_m} \\
&\cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w - w_m) + w - \lambda \cdot z \cdot w_u}{1 - \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z \\
&+ 1_{w \ge w_m} \cdot \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w_u - w) + w - \lambda \cdot z \cdot w_u}{1 - \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z.
\end{aligned}
$$

The last equality uses the following four facts: (i) $\lambda \cdot x \cdot (w - w_l) + w \le w_m$ implies $w \le w_m(1 + \lambda x w_l/w_m)/(1 + \lambda x) \le w_m$, (ii) $\lambda \cdot x \cdot (w_u - w) + w \le w_m$ implies $w \le w_m(1 - \lambda x w_u/w_m)/(1 - \lambda x) < w_m$, (iii) $\lambda \cdot x \cdot (w - w_l) + w > w_m$ implies $w \ge w_m(1 + \lambda x w_l/w_m)/(1 + \lambda x)$, and (iv) $\lambda \cdot x \cdot (w_u - w) + w > w_m$ implies $w \ge w_m(1 - \lambda x w_u/w_m)/(1 - \lambda x)$.

Now we will look at

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\partial g}{\partial w}(W_i, X_i)\frac{1}{f_W(W_i)}k(W_i, X_i, \lambda)\right] \\
&= \int_{x_l}^{x_u} \int_{w_l}^{w_u} \frac{\partial g}{\partial w}(w, x)\frac{1}{f_W(w)}k(w, x, \lambda) f_{W,X}(w, x)\,\mathrm{d}w\,\mathrm{d}x.
\end{aligned}
$$

Substituting the three terms of $k(w, x, \lambda)$ in here, we get

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\partial g}{\partial w}(W_i, X_i)\frac{1}{f_W(W_i)}k(W_i, X_i, \lambda)\right] \\
&= \int_{x_l}^{x_u} \int_{w_l}^{w_m(1+\lambda x w_l/w_m)/(1+\lambda x)} \frac{\frac{\partial g}{\partial w}(w, x)}{f_W(w)} \\
&\quad \times \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w - w_l) + w + \lambda \cdot z \cdot w_l}{1 + \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z f_{W,X}(w, x)\,\mathrm{d}w\,\mathrm{d}x
\end{aligned}
\tag{B.5}
$$

$$
\begin{aligned}
&\quad + \int_{x_l}^{x_u} \int_{w_m(1+\lambda x w_l/w_m)/(1+\lambda x)}^{w_m} \frac{\frac{\partial g}{\partial w}(w, x)}{f_W(w)} \\
&\quad \times \int F_{W|X}\left(\frac{\lambda \cdot x \cdot (w - w_l) + w - \lambda \cdot z \cdot w_u}{1 - \lambda \cdot z}\Big| z\right) f_X(z)\,\mathrm{d}z f_{W,X}(w, x)\,\mathrm{d}w\,\mathrm{d}x
\end{aligned}
\tag{B.6}
$$

$$
\begin{aligned}
&+\int_{x_l}^{x_u}\int_{w_m}^{w_u}\frac{\frac{\partial g}{\partial w}(w,x)}{f_W(w)} \\
&\times\int F_{W|X}\left(\frac{\lambda\cdot x\cdot(w_u-w)+w-\lambda\cdot z\cdot w_u}{1-\lambda\cdot z}\Big|z\right)f_X(z)\,\mathrm{d}zf_{W,X}(w,x)\,\mathrm{d}w\,\mathrm{d}x.
\end{aligned}
\tag{B.7}
$$

Next, we take the derivative with respect to λ for each of these three terms and evaluate that derivative at $\lambda=0$. For the first term, (B.5), this derivative consists of two terms: one that corresponds to the derivative with respect to the λ in the bounds of the integral, and one that corresponds to the derivative with respect to λ in the integrand. For the second term, we only have the term that corresponds to the derivative with respect to the λ in the bounds of the integral, since the other term vanishes when we evaluate it at $\lambda=0$. The third term, (B.7), only has λ in the integrand. So

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda}\mathbb{E}\left[\frac{\partial g}{\partial w}(W_i,X_i)\frac{1}{f_W(W_i)}k(W_i,X_i,\lambda)\right]\Bigg|_{\lambda=0} \\
&=(w_l-w_m)\cdot\mathbb{E}\left[\frac{\partial}{\partial w}g(w_m,X_i)\Big|W_i=w_m\right] \\
&\quad+\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x)\frac{1}{f_W(w)} \\
&\quad\times\int_{x_l}^{x_u}f_{W|X}(w|z)\big(x\cdot(w-w_l)+z\cdot w_l-z\cdot w\big)f_X(z)\,\mathrm{d}zf_{W,X}(w,x)\,\mathrm{d}w\,\mathrm{d}x \\
&\quad-(w_l-w_m)\cdot\mathbb{E}\left[\frac{\partial}{\partial w}g(w_m,X_i)\Big|W_i=w_m\right] \\
&\quad+\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x)\frac{1}{f_W(w)} \\
&\quad\times\int_{x_l}^{x_u}f_{W|X}(w|z)\big(x\cdot(w_u-w)+z\cdot w-z\cdot w_u\big)f_X(z)\,\mathrm{d}zf_{W,X}(w,x)\,\mathrm{d}w\,\mathrm{d}x \\
&=\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x) \\
&\quad\times\int_{x_l}^{x_u}f_{X|W}(z|w)\big(x\cdot d(w,x)-z\cdot d(w,z)\big)\,\mathrm{d}zf_{W,X}(w,x)\,\mathrm{d}w\,\mathrm{d}x \\
&\quad+\int_{x_l}^{x_u}\int_{w_l}^{w_m}\frac{\partial}{\partial w}g(w,x) \\
&\quad\times\int_{x_l}^{x_u}f_{X|W}(z|w)\big(x\cdot d(w,x)-z\cdot d(w,z)\big)\,\mathrm{d}zf_{W,X}(w,x)\,\mathrm{d}w\,\mathrm{d}x \\
&=\int_{x_l}^{x_u}\int_{w_l}^{w_u}\frac{\partial}{\partial w}g(w,x) \\
&\quad\times\int_{x_l}^{x_u}f_{X|W}(z|w)\big(x\cdot d(w,x)-z\cdot d(w,z)\big)\,\mathrm{d}zf_{W,X}(w,x)\,\mathrm{d}w\,\mathrm{d}x
\end{aligned}
$$

$$= \int_{x_l}^{x_u} \int_{w_l}^{w_u} \frac{\partial}{\partial w} g(w, x)$$
$$\times \big(X_i \cdot d(w, X_i) - \mathbb{E}\big[X_i \cdot d(w, X_i)|W_i = w\big]\big) f_{W,X}(w, x)\,\mathrm{d}w\,\mathrm{d}x$$
$$= \mathbb{E}\bigg[\frac{\partial g}{\partial w}(W_i, X_i) \cdot \big(X_i \cdot d(W_i, X_i) - \mathbb{E}\big[X_i \cdot d(W_i, X_i)|W_i\big]\big)\bigg] = \beta^{\mathrm{lc,v}}.$$

This finishes the proof of (B.3).

Finally, we show (B.4) by showing the equality of

$$\beta^{\mathrm{lc,v}} = \mathbb{E}\bigg[\frac{\partial g}{\partial w}(W_i, X_i) \cdot \big(X_i \cdot d(W_i) - \mathbb{E}\big[X_i \cdot d(W_i)|W_i\big]\big)\bigg] \tag{B.8}$$

and

$$\mathbb{E}\bigg[\delta(W_i, X_i) \cdot \frac{\partial^2 g}{\partial w\,\partial x}(W_i, X_i)\bigg]. \tag{B.9}$$

Define

$$b(w) = \mathbb{E}\bigg[\frac{\partial g}{\partial w}(w, X_i) \cdot \big(X_i \cdot d(w) - \mathbb{E}\big[X_i \cdot d(w)|W_i = w\big]\big)\bigg|W_i = w\bigg]$$
$$= \mathbb{E}\bigg[\frac{\partial g}{\partial w}(w, X_i) \cdot d(w) \cdot \big(X_i - \mathbb{E}[X_i|W_i = w]\big)\bigg|W_i = w\bigg],$$

so that $\beta^{\mathrm{lc,v}} = \mathbb{E}[b(W)]$. Apply Lemma A.2, with $h(x) = \frac{\partial g}{\partial w}(w, x) \cdot d(w)$, to get

$$b(w) = \mathbb{E}\bigg[\frac{\partial^2}{\partial w\,\partial x} g(w, X) \cdot \delta(w, X)\bigg],$$

with

$$\delta(w, x) = d(w) \cdot \frac{F_{X|W}(x|w) \cdot (1 - F_{X|W}(x|w))}{f_{X|W}(x|w)}$$
$$\cdot \big(\mathbb{E}[X|X > x, W = w] - \mathbb{E}[X|X \le x, W = w]\big).$$

Thus

$$\beta^{\mathrm{lc,v}} = \mathbb{E}\big[b(W)\big] = \mathbb{E}\bigg[\frac{\partial^2}{\partial w\,\partial x} g(W, X) \cdot \delta(W, X)\bigg].$$ □

Proof of Theorem 5.1. We apply Lemmas A.15–A.18. The assumptions in the theorem imply that the conditions for those lemmas are satisfied. □

The proof of Theorem 5.2 is essentially the same as that for Theorem 5.1 and is omitted.

Proof of Theorem 5.3. We apply Lemmas A.24–A.28 to get an asymptotic linear representation for $\hat{\beta}^{\mathrm{cm}}(\rho, \tau)$. The assumptions in the theorem imply that the conditions for the applications of these lemmas are satisfied. Therefore, by Lemma A.24, we have

$$\hat{\beta}^{\mathrm{cm}}(\rho,0)=\beta^{\mathrm{cm}}(\rho,0)+\big(\hat{\beta}^{\mathrm{cm}}_g-\overline{g}^{\mathrm{cm}}\big)+\big(\hat{\beta}^{\mathrm{cm}}_W-\overline{g}^{\mathrm{cm}}\big)+\big(\hat{\beta}^{\mathrm{cm}}_X-\overline{g}^{\mathrm{cm}}\big)$$
$$+\big(\overline{g}^{\mathrm{cm}}-\beta^{\mathrm{cm}}(\rho,0)\big)+o_p\big(N^{-1/2}\big).$$

By Lemmas A.25–A.28, this is equal to

$$\beta^{\mathrm{cm}}(\rho,0)+\frac{1}{N}\sum_{i=1}^{N}\big\{\psi^{\mathrm{cm}}_g(Y_i,W_i,X_i)+\psi^{\mathrm{cm}}_W(Y_i,W_i,X_i)$$
$$+\psi^{\mathrm{cm}}_X(Y_i,W_i,X_i)+\psi^{\mathrm{cm}}_0(Y_i,W_i,X_i)\big\}+o_p\big(N^{-1/2}\big)$$
$$=\beta^{\mathrm{cm}}(\rho,0)+\frac{1}{N}\sum_{i=1}^{N}\psi(Y_i,W_i,X_i)+o_p\big(N^{-1/2}\big),$$

with $\psi^{\mathrm{cm}}_g(y,w,x)$ given in (5.42), $\psi^{\mathrm{cm}}_W(y,w,x)$ given in (5.43), $\psi^{\mathrm{cm}}_X(y,w,x)$ given in (5.44), $\psi^{\mathrm{cm}}_0(y,w,x)$ given in (5.41), and $\psi(y,w,x)$ given in (5.45). Then we have an asymptotic linear representation for $\hat{\beta}^{\mathrm{cm}}(\rho,\tau)$:

$$\hat{\beta}^{\mathrm{cm}}(\rho,\tau)=\tau\cdot\overline{Y}+(1-\tau)\cdot\hat{\beta}^{\mathrm{cm}}(\rho,0)$$
$$=\beta^{\mathrm{cm}}(\rho,\tau)+\tau\cdot\big(\overline{Y}-\beta^{\mathrm{cm}}(\rho,1)\big)+(1-\tau)\cdot\big(\hat{\beta}^{\mathrm{cm}}(\rho,0)-\beta^{\mathrm{cm}}(\rho,0)\big)$$
$$=\beta^{\mathrm{cm}}(\rho,\tau)+\tau\cdot\big(\overline{Y}-\beta^{\mathrm{cm}}(\rho,1)\big)+(1-\tau)\cdot\frac{1}{N}\sum_{i=1}^{N}\psi(Y_i,W_i,X_i).$$

Since by the law of large numbers, $\overline{Y}\to\beta^{\mathrm{cm}}(\rho,1)$ and $\sum_i\psi(Y_i,W_i,X_i)/N\to\mathbb{E}[\psi(Y_i,W_i,X_i)]=0$, it follows that $\hat{\beta}^{\mathrm{cm}}(\rho,\tau)\to\beta^{\mathrm{cm}}(\rho,\tau)$. By the central limit theorem, the second part of the theorem follows. □

Proof of Theorem 5.4. The proof uses Lemmas A.13, A.14, A.19, A.20, and A.23.

By the conditions on q, r, s, and δ, Lemma A.13 implies that for some $\eta>1/4$,

$$\sup_{w\in\mathbb{W}}\big|\hat{m}(w)-m(w)\big|=o_p\big(N^{-\eta}\big).$$

Moreover, by the same conditions, Lemma A.14 implies that for some $\eta>1/4$,

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial\hat{g}}{\partial w}(w,x)-\frac{\partial g}{\partial w}(w,x)\right|=o_p\big(N^{-\eta}\big).$$

Then the conditions for Lemma A.19 are satisfied, so we can write

$$\sqrt{N}\big(\hat{\beta}^{\mathrm{lc}}-\beta^{\mathrm{lc}}\big)$$
$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot\big(X-m(W_i)\big)-\sqrt{N}\cdot\beta^{\mathrm{lc}}$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial\hat{g}}{\partial w}(W_i,X_i)\cdot d(W_i)\cdot\big(X-m(W_i)\big)$$

$$
\begin{aligned}
&- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X - m(W_i)\big) \\
&+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X - \hat{m}(W_i)\big) \\
&- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X - m(W_i)\big) + o_p(1).
\end{aligned}
$$

By Lemma A.20,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial w} \hat{g}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big) \\
&\qquad - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial w} g(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big) \\
&\quad = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi_g^{\mathrm{lc}}(Y_i, W_i, X_i) + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\psi_g^{\mathrm{lc}}(y, w, x) = &-\frac{1}{f_{W,X}(w, x)} \frac{\partial f_{W,X}(w, x)}{\partial W} \big(y - g(w, x)\big) d(w) \big(x - m(w)\big) \\
&- \frac{\partial m(w)}{\partial W} d(w) \big(y - g(w, x)\big) + \frac{\partial}{\partial w} d(w) \big(x - m(w)\big) \big(y - g(w, x)\big).
\end{aligned}
$$

By Lemma A.23,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial w} g(W_i, X_i) \cdot d(W_i) \cdot \hat{m}(W_i) \\
&\qquad - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial w} g(W_i, X_i) \cdot d(W_i) \cdot m(W_i) \\
&\quad = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi_m^{\mathrm{lc}}(Y_i, W_i, X_i) + o_p(1),
\end{aligned}
$$

where

$$
\psi_m^{\mathrm{lc}}(y, w, x) = \mathbb{E}\left[\frac{\partial g(w, X_i)}{\partial W} \middle| W_i = w \right] \cdot d(w) \cdot \big(x - m(w)\big).
$$

Combining these results implies that

$$
\sqrt{N}\big(\hat{\beta}^{\mathrm{lc}} - \beta^{\mathrm{lc}}\big) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi^{\mathrm{lc}}(Y_i, W_i, X_i) + o_p(1),
$$

with

$$\psi^{\text{lc}}(y,w,x) = \left(\frac{\partial g(w,x)}{\partial w} \cdot d(w) \cdot \big(x - m(w)\big) - \beta^{\text{lc}}\right) + \psi_g^{\text{lc}}(y,w,x) + \psi_m^{\text{lc}}(y,w,x).$$

Using the law of large numbers then implies the first result in the theorem, and using the central limit theorem implies the second result in the theorem. □

Appendix C: Proofs of results listed in Appendix A

In the following proofs, c is a generic constant.

Proof of Lemma A.1. Because $f(\cdot)$ is twice continuously differentiable on $\mathbb{X}$, a compact subset of $\mathbb{R}$, it follows that for all $a, b \in \mathbb{X}$, by a Taylor series expansion,

$$f(b) = f(a) + \frac{\partial f}{\partial x}(a) \cdot (b-a) + \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}(c) \cdot (b-a)^2$$

for some $c \in \mathbb{X}$. Hence

$$\left| f\big(g(\lambda)\big) - \left(f\big(g(0)\big) + \frac{\partial f}{\partial x}\big(g(0)\big) \cdot \big(g(\lambda) - g(0)\big)\right)\right| \le \frac{1}{2} \cdot \sup_{x\in\mathbb{X}} \left| \frac{\partial^2 f}{\partial x^2}(x)\right| \cdot \big(g(\lambda) - g(0)\big)^2.$$

By the Lipschitz condition on $g(\lambda)$, this is bounded by

$$\frac{1}{2} \cdot \sup_{x\in\mathbb{X}} \left| \frac{\partial^2 f}{\partial x^2}(x)\right| \cdot c^2 \cdot \lambda^2.$$

□

Proof of Lemma A.2. Let $\mu = \mathbb{E}[X]$ and write $h(x) = h(x_l) + \int_{x_l}^{x} \frac{\partial}{\partial x} h(z)\,\mathrm{d}z$. Then

$$\begin{aligned}
&\operatorname{Cov}\big(h(X), X\big) \\
&= \mathbb{E}\big[h(X) \cdot (X-\mu)\big] \\
&= \mathbb{E}\left[\left(h(x_l) + \int_{x_l}^{X} \frac{\partial}{\partial x} h(z)\,\mathrm{d}z\right) \cdot (X-\mu)\right] \\
&= \mathbb{E}\left[\int_{x_l}^{X} \frac{\partial}{\partial x} h(z)\,\mathrm{d}z \cdot (X-\mu)\right] \\
&= \mathbb{E}\left[\int_{x_l}^{x_u} \mathbf{1}_{X>z} \cdot \frac{\partial}{\partial x} h(z)\,\mathrm{d}z \cdot (X-\mu)\right] \\
&= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot \mathbb{E}\big[\mathbf{1}_{X>z} \cdot (X-\mu)\big]\,\mathrm{d}z
\end{aligned}$$

$$= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot \mathbb{E}[X - \mu | X > z] \cdot \Pr(X > z)\, \mathrm{d}z$$

$$= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot F_X(z) \cdot \big(1 - F_X(z)\big) \cdot \big(\mathbb{E}[X|X > z] - \mathbb{E}[X|X \le z]\big)\, \mathrm{d}z$$

$$= \int_{x_l}^{x_u} \frac{\partial}{\partial x} h(z) \cdot \frac{F_X(z) \cdot (1 - F_X(z))}{f_X(z)} \cdot \big(\mathbb{E}[X|X > z] - \mathbb{E}[X|X \le z]\big) f_X(z)\, \mathrm{d}z$$

$$= \mathbb{E}\left[\frac{\partial}{\partial x} h(X) \cdot \gamma(X)\right]. \qquad \square$$

Proof of Lemma A.7. By the triangle inequality,

$$\sup_{x \in \mathbb{X}} N^{\delta} \cdot \big|\hat{F}_Y^{-1}\big(\hat{F}_X(x)\big) - F_Y^{-1}\big(F_X(x)\big)\big|$$

$$\le \sup_{x \in \mathbb{X}} N^{\delta} \cdot \big|\hat{F}_Y^{-1}\big(\hat{F}_X(x)\big) - F_Y^{-1}\big(\hat{F}_X(x)\big)\big| + \sup_{x \in \mathbb{X}} N^{\delta} \cdot \big|F_Y^{-1}\big(\hat{F}_X(x)\big) - F_Y^{-1}\big(\hat{F}_X(x)\big)\big|$$

$$\le \sup_{q \in [0,1]} N^{\delta} \cdot \big|\hat{F}_Y^{-1}(q) - F_Y^{-1}(q)\big| + \sup_{x \in \mathbb{X}, y \in \mathbb{Y}} N^{\delta} \cdot \frac{1}{f_Y(y)} \big|\hat{F}_X(x) - F_X(x)\big|.$$

The first term is $o_p(1)$ by Lemma A.4, and the second by the fact that $f_Y(y)$ is bounded away from zero, in combination with Lemma A.3. $\square$

Proof of Lemma A.8. By the triangle inequality,

$$\sup_{y \in \mathbb{Y}, x \le N^{-\delta}, x+y \in \mathbb{Y}} N^{\eta} \cdot \big|\hat{F}_Y(y + x) - \hat{F}_Y(y) - f_Y(y) \cdot x\big|$$

$$\le \sup_{y \in \mathbb{Y}, x \le N^{-\delta}, x+y \in \mathbb{Y}} N^{\eta} \cdot \big|\hat{F}_Y(y + x) - \big(F_Y(y + x) - F_Y(y)\big)\big|$$

$$+ \sup_{y \in \mathbb{Y}, x \le N^{-\delta}, x+y \in \mathbb{Y}} N^{\eta} \cdot \big|F_Y(y + x) - F_Y(y) - f_Y(y) \cdot x\big|.$$

The first term on the right-hand side converges to zero in probability by Lemma A.5. To show that the second term converges to zero, note that

$$\sup_{y \in \mathbb{Y}, x \le N^{-\delta}, x+y \in \mathbb{Y}} N^{\eta} \cdot \big|F_Y(y + x) - F_Y(y) - f_Y(y) \cdot x\big|$$

$$\le \sup_{y \in \mathbb{Y}, x \le N^{-\delta}, x+y \in \mathbb{Y}, \lambda \in [0,1]} N^{\eta} \cdot \big|f_Y(y + \lambda x) \cdot x - f_Y(y) \cdot x\big|$$

$$\le \sup_{y \in \mathbb{Y}, z \in \mathbb{Y}, x \le N^{-\delta}, x+y \in \mathbb{Y}, \lambda \in [0,1]} N^{\eta} \cdot \left|\frac{\partial f_Y}{\partial y}(z)\right| \cdot \lambda x^2$$

$$\le \sup_{y \in \mathbb{Y}, x \le N^{-\delta}} N^{\eta} x^2 \frac{\partial f_Y}{\partial y}(y) \to 0,$$

because $\frac{\partial f_Y}{\partial y}(y)$ is bounded, $x < N^{-\delta}$, and $\delta > \eta/2$. $\square$

PROOF OF LEMMA A.12. By the inequality $|a| \geq |b| - |a-b|$,

$$\inf_{z\in\mathbb{Z}}\left|\hat{h}(z)\right| \geq \inf_{z\in\mathbb{Z}}\left|h(z)\right| - \sup_{z\in\mathbb{Z}}\left|\hat{h}(z)-h(z)\right|,$$

from which the result follows. □

The proofs of Lemmas A.13 and A.14 follow directly from Theorem 7.1 in IR.

PROOF OF LEMMA A.15. First note that by the assumptions in the lemma, the conditions for Lemma A.14 are satisfied. Moreover, by the assumption that $0<\delta<1/6$, it follows that $O_p(b^N) = o_p(N^{-\eta})$ for $\eta < \delta\cdot s$, and $O_p(\ln(N)N^{-1}b_N^{-2}) = O_p(\ln(N)N^{-1+2\delta}) = o_p(1)$, $O_p(\ln(N)N^{-1}b_N^{-4}) = O_p(\ln(N)N^{-1+4\delta}) = o_p(N^{-\eta})$ for $\eta < 1-4\delta$, and $O_p(\ln(N)\times N^{-1}b_N^{-6}) = O_p(\ln(N)N^{-1+6\delta}) = o_p(1)$. Hence the results from Lemma A.14 imply

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\hat{g}(w,x)-g(w,x)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^2}\right)^{1/2} + b_N^s\right) = o_p(1), \tag{C.1}$$

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial\hat{g}}{\partial w}(w,x)-\frac{\partial g}{\partial w}(w,x)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^4}\right)^{1/2} + b_N^s\right) = o_p\left(N^{-\eta}\right) \tag{C.2}$$

for $\eta < \min(1-4\delta, \delta\cdot s)$, and

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2\hat{g}}{\partial w^2}(w,x)-\frac{\partial^2 g}{\partial w^2}(w,x)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^6}\right)^{1/2} + b_N^s\right) = o_p(1). \tag{C.3}$$

Now

$$\begin{aligned}
&\hat{\beta}^{\text{pam}} - \beta^{\text{pam}} \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{g}\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) - \mathbb{E}\left[g\left(F_W^{-1}\left(F_X(X)\right), X\right)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{g}\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N}g\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) && \text{(C.4)} \\
&\quad - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{g}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N}g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\right) && \text{(C.5)} \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\hat{g}\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N}g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right) && \text{(C.6)} \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}g\left(\hat{F}_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N}g\left(F_W^{-1}\left(\hat{F}_X(X_i)\right), X_i\right) && \text{(C.7)} \\
&\quad - \left(\frac{1}{N}\sum_{i=1}^{N}g\left(\hat{F}_W^{-1}\left(F_X(X_i)\right), X_i\right) - \frac{1}{N}\sum_{i=1}^{N}g\left(F_W^{-1}\left(F_X(X_i)\right), X_i\right)\right) && \text{(C.8)}
\end{aligned}$$

$$+\frac{1}{N}\sum_{i=1}^{N} g\big(\hat{F}_W^{-1}\big(F_X(X_i)\big), X_i\big) - \frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big) \tag{C.9}$$

$$+\frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(\hat{F}_X(X_i)\big), X_i\big) - \frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big) \tag{C.10}$$

$$+\frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big) - \mathbb{E}\big[g\big(F_W^{-1}\big(F_X(X)\big), X\big)\big]. \tag{C.11}$$

Since (C.6) is equal to $\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}$, (C.9) equals $\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}$, (C.10) equals $\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}$, and (C.11) equals $\overline{g}^{\text{pam}} - \beta^{\text{pam}}$, we only need to show that the sum of (C.4) and (C.5), and that of (C.7) and (C.8) are $o_p(N^{-1/2})$.

First consider the sum of (C.4) and (C.5) that is equal to

$$\frac{1}{N}\sum_{i=1}^{N} \hat{g}\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big), X_i\big) - \frac{1}{N}\sum_{i=1}^{N} \hat{g}\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)$$
$$-\left(\frac{1}{N}\sum_{i=1}^{N} g\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big), X_i\big) - \frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\right).$$

By a second order Taylor series expansion of $\hat{g}$ and g in $F_W^{-1}(F_X(X_i))$ this is, for some $\tilde{W}_i$ and $\overline{W}_i$, equal to

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big)$$
$$+\frac{1}{2N}\sum_{i=1}^{N} \frac{\partial^2 \hat{g}}{\partial w^2}(\tilde{W}_i, X_i)\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big)^2 \tag{C.12}$$

$$-\frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big)$$
$$-\frac{1}{2N}\sum_{i=1}^{N} \frac{\partial^2 g}{\partial w^2}(\overline{W}_i, X_i)\cdot\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big)^2 \tag{C.13}$$

$$=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial \hat{g}}{\partial w}\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big) - \frac{\partial g}{\partial w}\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\right)$$
$$\times\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big) + o_p\big(N^{-1/2}\big)$$

$$\le \sup_{x\in\mathbb{X}}\left|\frac{\partial \hat{g}}{\partial w}\big(F_W^{-1}\big(F_X(x)\big), x\big) - \frac{\partial g}{\partial w}\big(F_W^{-1}\big(F_X(x)\big), x\big)\right| \tag{C.14}$$

$$\times \sup_{x\in\mathbb{X}}\big|\big(\hat{F}_W^{-1}\big(\hat{F}_X(x)\big) - F_W^{-1}\big(F_X(x)\big)\big)\big| + o_p\big(N^{-1/2}\big). \tag{C.15}$$

We used the fact that (C.13) is $o_p(N^{-1/2})$ because $\partial^2 g(w,x)/\partial w^2$ is bounded and because $\sup_{x\in\mathbb{X}}(\hat{F}_W^{-1}(\hat{F}_X(x)) - F_W^{-1}(F_X(x)))^2$ is $o_p(N^{-1/2})$ by Lemma A.7. Also, (C.12) is $o_p(N^{-1/2})$ by the same argument because the bandwidth choice implies $\sup_{w\in\mathbb{W},x\in\mathbb{X}} |\partial^2\hat{g}(w,x)/\partial w^2 - \partial^2 g(w,x)/\partial w^2| = o_p(1)$ by (C.3), so that

$$\begin{aligned}\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2\hat{g}(w,x)}{\partial w^2}\right| &\leq \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2 g(w,x)}{\partial w^2}\right| + \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2\hat{g}(w,x)}{\partial w^2} - \frac{\partial^2 g(w,x)}{\partial w^2}\right| \\ &= \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial^2 g(w,x)}{\partial w^2}\right| + o_p(1).\end{aligned}$$

Finally, by Lemma A.7,

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\big(\hat{F}_X(x)\big) - F_W^{-1}\big(F_X(x)\big)\right| = o_p\big(N^{-1/2+\eta}\big)$$

for all $\eta > 0$. By the assumption of the lemma,

$$\sup_{x\in\mathbb{X}}\left|\frac{\partial\hat{g}}{\partial w}\big(F_W^{-1}\big(F_X(x)\big),x\big) - \frac{\partial g}{\partial w}\big(F_W^{-1}\big(F_X(x)\big),x\big)\right| = o_p\big(N^{-\eta}\big)$$

for some $\eta > 0$. We conclude that the sum of (C.4) and (C.5) is $o_p(N^{-1/2})$.

Next, consider the sum of (C.7) and (C.8) that is bounded by

$$\begin{aligned}\sup_{x\in\mathbb{X}}\Big|&\big[g\big(\hat{F}_W^{-1}\big(\hat{F}_X(x)\big),x\big) - g\big(F_W^{-1}\big(\hat{F}_X(x)\big),x\big)\big] \\ &- \big[g\big(\hat{F}_W^{-1}\big(F_X(x)\big),x\big) - g\big(F_W^{-1}\big(F_X(x)\big),x\big)\big]\Big|.\end{aligned}$$

By a second order Taylor series expansion, with intermediate values $\tilde{W}(x)$ and $\overline{W}(x)$, and the triangle inequality, this is bounded by

$$\begin{aligned}\sup_{x\in\mathbb{X}}\Big|&\frac{\partial g}{\partial w}\big(F_W^{-1}\big(\hat{F}_X(x)\big),x\big)\big[\hat{F}_W^{-1}\big(\hat{F}_X(x)\big) - F_W^{-1}\big(\hat{F}_X(x)\big)\big] \\ &- \frac{\partial g}{\partial w}\big(F_W^{-1}\big(F_X(x)\big),x\big)\big[\hat{F}_W^{-1}\big(F_X(x)\big) - F_W^{-1}\big(F_X(x)\big)\big]\Big| \\ &+ \sup_{x\in\mathbb{X}}\frac{1}{2}\left|\frac{\partial^2 g}{\partial w^2}\big(\tilde{W}(x),x\big)\big[\hat{F}_W^{-1}\big(\hat{F}_X(x)\big) - F_W^{-1}\big(\hat{F}_X(x)\big)\big]^2\right| \\ &+ \frac{1}{2}\sup_{x\in\mathbb{X}}\left|\frac{\partial^2 g}{\partial w^2}\big(\overline{W}(x),x\big)\big[\hat{F}_W^{-1}\big(F_X(x)\big) - F_W^{-1}\big(F_X(x)\big)\big]^2\right|,\end{aligned}$$

where because the second derivative of $g(w,x)$ is bounded on $\mathbb{W}\times\mathbb{X}$, by Lemma A.4, the expression on the last line is $o_p(N^{-1/2})$. The first term is bounded by

$$\begin{aligned}\sup_{x\in\mathbb{X}}\Big|&\left[\frac{\partial g}{\partial w}\big(F_W^{-1}\big(\hat{F}_X(x)\big),x\big) - \frac{\partial g}{\partial w}\big(F_W^{-1}\big(F_X(x)\big),x\big)\right] \\ &\times\big[\hat{F}_W^{-1}\big(\hat{F}_X(x)\big) - F_W^{-1}\big(\hat{F}_X(x)\big)\big]\Big|\end{aligned}$$

$$+\sup_{x\in\mathbb{X}}\left|\frac{\partial g}{\partial w}\big(F_W^{-1}\big(F_X(x)\big),x\big)\right.$$
$$\left.\times\left[\hat{F}_W^{-1}\big(\hat{F}_X(x)\big)-F_W^{-1}\big(\hat{F}_X(x)\big)-\hat{F}_W^{-1}\big(F_X(x)\big)+F_W^{-1}\big(F_X(x)\big)\right]\right|.$$

By a first order Taylor series expansion of $\frac{\partial g}{\partial w}(F_W^{-1}(\hat{F}_X(x)),x)$ in $F_X(x)$, we have, because the second derivative of $g(w,x)$ is bounded and the density of W is bounded from 0 on its support, that by Lemmas A.4 and A.3, the expression on the first line is $o_p(N^{-1/2})$. The bound on the expression in the second line is proportional to

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\big(\hat{F}_X(x)\big)-F_W^{-1}\big(\hat{F}_X(x)\big)-\hat{F}_W^{-1}\big(F_X(x)\big)+F_W^{-1}\big(F_X(x)\big)\right|.$$

This expression is bounded by

$$\sup_{x\in\mathbb{X}}\left|\frac{1}{f_W(F_W^{-1}(\hat{F}_X(x)))}\left[\hat{F}_W\big(F_W^{-1}\big(\hat{F}_X(x)\big)\big)-\hat{F}_X(x)\right]\right.$$
$$\left.-\frac{1}{f_W(F_W^{-1}(F_X(x)))}\left[\hat{F}_W\big(F_W^{-1}\big(F_X(x)\big)\big)-F_X(x)\right]\right|$$
$$+\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\big(\hat{F}_X(x)\big)-F_W^{-1}\big(\hat{F}_X(x)\big)\right.$$
$$\left.-\frac{1}{f_W(F_W^{-1}(\hat{F}_X(x)))}\left[\hat{F}_W\big(F_W^{-1}\big(\hat{F}_X(x)\big)\big)-\hat{F}_X(x)\right]\right|$$
$$+\sup_{x\in\mathbb{X}}\left|\hat{F}_W^{-1}\big(F_X(x)\big)-F_W^{-1}\big(F_X(x)\big)\right.$$
$$\left.-\frac{1}{f_W(F_W^{-1}(F_X(x)))}\left[\hat{F}_W\big(F_W^{-1}\big(F_X(x)\big)\big)-F_X(x)\right]\right|.$$

By Lemma A.7, the expressions in the last two lines are $o_p(N^{-1/2})$. The expression in the first line is bounded by

$$\sup_{x\in\mathbb{X}}\left|\left[\frac{1}{f_W(F_W^{-1}(\hat{F}_X(x)))}-\frac{1}{f_W(F_W^{-1}(F_X(x)))}\right]\left[\hat{F}_W\big(F_W^{-1}\big(\hat{F}_X(x)\big)\big)-\hat{F}_X(x)\right]\right|$$
$$+\sup_{x\in\mathbb{X}}\left|\frac{1}{f_W(F_W^{-1}(F_X(x)))}\right.$$
$$\left.\times\left[\hat{F}_W\big(F_W^{-1}\big(\hat{F}_X(x)\big)\big)-\hat{F}_X(x)-\hat{F}_W\big(F_W^{-1}\big(F_X(x)\big)\big)+F_X(x)\right]\right|.$$

The expression in the first line is bounded by

$$\sup_{x\in\mathbb{X}}\left|\frac{1}{f_W(F_W^{-1}(\hat{F}_X(x)))}-\frac{1}{f_W(F_W^{-1}(F_X(x)))}\right|\times\sup_{x\in\mathbb{X}}\left|\hat{F}_W\big(F_W^{-1}\big(\hat{F}_X(x)\big)\big)-\hat{F}_X(x)\right|.$$

By a first order Taylor series expansion of $\frac{1}{f_W(F_W^{-1}(\hat{F}_X(x)))}$ in $F_X(x)$, the fact that $f_W(w)$ is bounded from 0 and its derivative is bounded on $\mathbb{W}$, and Lemma A.3, the first factor is $o_p(N^{-\delta})$ for all $\delta < 1/2$ and by Lemma A.3, the same is true for the second factor, so that the product is $o_p(N^{-1/2})$. Because $f_W(w)$ is bounded from 0 on $\mathbb{W}$, the expression on the second line has a bound that is proportional to

$$\sup_{x\in\mathbb{X}}\left|\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right) - \hat{F}_X(x) - \hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right) + F_X(x)\right|.$$

We rewrite this as

$$\begin{aligned}
&\sup_{x\in\mathbb{X}}\left|\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right) - \hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)\right. \\
&\qquad \left. - \left(F_W\left(F_W^{-1}\left(\hat{F}_X(x)\right) - F_W\left(F_W^{-1}\left(F_X(x)\right)\right)\right)\right)\right| \\
&\quad \le \sup_{x\in\mathbb{X}}\left|\hat{F}_W\left(F_W^{-1}\left(\hat{F}_X(x)\right)\right) - \hat{F}_W\left(F_W^{-1}\left(F_X(x)\right)\right)\right. \\
&\qquad \left. - \left(F_W\left(F_W^{-1}\left(\hat{F}_X(x)\right) - F_W\left(F_W^{-1}\left(F_X(x)\right)\right)\right)\right)\right| 1_{\sup_{x\in\mathbb{X}}|F_W^{-1}(\hat{F}_X(x)) - F_W^{-1}(F_X(x))|\le N^{-\delta}} \\
&\qquad + 4\cdot 1_{\sup_{x\in\mathbb{X}}|F_W^{-1}(\hat{F}_X(x)) - F_W^{-1}(F_X(x))| > N^{-\delta}}.
\end{aligned}$$

By Lemma A.7 and the mean value theorem, the final term is $o_p(1)$ if $1/3 < \delta < 1/2$. By

$$F_W^{-1}\left(\hat{F}_X(x)\right) = F_W^{-1}\left(F_X(x)\right) + \left[F_W^{-1}\left(\hat{F}_X(x)\right) - F_W^{-1}\left(F_X(x)\right)\right],$$

and defining $\overline{w} = F_W^{-1}(F_X(x))$ and $\tilde{w} = F_W^{-1}(\hat{F}_X(x)) - F_W^{-1}(F_X(x))$, we have that the first term on the right-hand side is bounded by

$$\sup_{\overline{w}\in\mathbb{W}, |\tilde{w}|\le N^{-\delta}, \tilde{w}+\overline{w}\in\mathbb{W}}\left|\hat{F}_W(\overline{w}+\tilde{w}) - \hat{F}_W(\overline{w}) - \left(F_W(\overline{w}+\tilde{w}) - F_W(\overline{w})\right)\right| = o_p\left(N^{-2/3}\right)$$

by Lemma A.5, with $1/3 < \delta < 1/2$, $\eta = 2/3$, so that we finally conclude that the sum of (C.7) and (C.8) is $o_p(N^{-1/2})$. □

Proof of Lemma A.16. The proof involves checking the conditions for Theorem A.2 from IR (given in Appendix A in this supplement) and simplifying the conclusions from that theorem to the case at hand.

Define

$$\begin{aligned}
&h_1(w,x) = f_{WX}(w,x) \quad \text{and} \quad h_2(w,x) = f_{WX}(w,x)\cdot g(w,x), \\
&n(h) = \frac{h_2}{h_1},
\end{aligned}$$

so that

$$\begin{aligned}
&\omega(x) = 1, \\
&\frac{\partial n}{\partial h_1}(h) = -\frac{h_2}{(h_1)^2} = -\frac{g(F_W^{-1}(F_X(x)), x)}{f_{WX}(F_W^{-1}(F_X(x)), x)},
\end{aligned}$$

$$\frac{\partial n}{\partial h_2}(h) = \frac{1}{h_1} = \frac{1}{f_{WX}(F_W^{-1}(F_X(x)), x)},$$

$$t(x) = F_W^{-1}\big(F_X(x)\big), \qquad \frac{\partial}{\partial x}t(x) = \frac{f_X(x)}{f_W(F_W^{-1}(F_X(x)))},$$

$$\alpha_1(x) = -g\big(F_W^{-1}\big(F_X(x)\big), x\big), \qquad \alpha_2(x) = 1.$$

With $\tilde{Y}_i = (\tilde{Y}_{i1}\ \tilde{Y}_{i2})' = (1\ Y_i)'$, we have

$$\alpha(x)'\tilde{y} = y - g\big(F_W^{-1}\big(F_X(x)\big), x\big).$$

Applying the results in Theorem A.2, we have

$$\int_{\mathbb{U}_2} K\left(\frac{W_i - t(X_i)}{b_N} + \frac{\partial t}{\partial x}(X_i)\cdot u_2, u_2\right) \mathrm{d}u_2$$
$$= \int_u K\left(u, \frac{W_i - F_W^{-1}(F_X(X_i))}{b_N} + \frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u\right) \mathrm{d}u.$$

Substituting this into the result from Theorem A.2, we get

$$\begin{aligned}
&\sqrt{N}\big(\hat{\theta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\big) \\
&= \frac{1}{\sqrt{N}b_N}\sum_{i=1}^N\Bigg(\big(Y_i - g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\big) \\
&\quad \cdot \int_u K\left(\frac{W_i - F_W^{-1}(F_X(X_i))}{b_N} + \frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u, u\right)\mathrm{d}u \\
&\quad - \mathbb{E}\Bigg[\big(Y - g\big(F_W^{-1}\big(F_X(X)\big), X\big)\big) \\
&\quad \cdot \int_u K\left(\frac{W - F_W^{-1}(F_X(X))}{b_N} + \frac{f_X(X)}{f_W(F_W^{-1}(F_X(X)))}\cdot u, u\right)\mathrm{d}u\Bigg]\Bigg) \\
&\quad + o_p(1).
\end{aligned}$$

Adding and subtracting $g(W_i, X_i)$ in both terms, this is equal to

$$\begin{aligned}
&\frac{1}{\sqrt{N}b_N}\sum_{i=1}^N\Bigg\{\big(Y_i - g(W_i, X_i)\big) \\
&\quad \cdot \int_u K\left(\frac{W_i - F_W^{-1}(F_X(X_i))}{b_N} + \frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u, u\right)\mathrm{d}u \\
&\quad - \mathbb{E}\Bigg[\big(Y - g(W, X)\big) \\
&\quad \cdot \int_u K\left(\frac{W - F_W^{-1}(F_X(X))}{b_N} + \frac{f_X(X)}{f_W(F_W^{-1}(F_X(X)))}\cdot u, u\right)\mathrm{d}u\Bigg]\Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{N}b_N}\sum_{i=1}^{N}\Bigg\{\big(g(W_i,X_i)-g\big(F_W^{-1}\big(F_X(X_i)\big),X_i\big)\big)\\
&\cdot\int_u K\bigg(\frac{W_i-F_W^{-1}(F_X(X_i))}{b_N}+\frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u,u\bigg)\,\mathrm{d}u\\
&-\mathbb{E}\bigg[\big(g(W,X)-g\big(F_W^{-1}\big(F_X(X)\big),X\big)\big)\\
&\cdot\int_u K\bigg(\frac{W-F_W^{-1}(F_X(X))}{b_N}+\frac{f_X(X)}{f_W(F_W^{-1}(F_X(X)))}\cdot u,u\bigg)\,\mathrm{d}u\bigg]\Bigg\}\\
&+o_p(1)\\
=&\frac{1}{\sqrt{N}b_N}\sum_{i=1}^{N}\big(Y_i-g(W_i,X_i)\big)\\
&\cdot\int_{u_2} K\bigg(\frac{W_i-F_W^{-1}(F_X(X_i))}{b_N}+\frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u_2,u_2\bigg)\,\mathrm{d}u_2\\
&+\frac{1}{\sqrt{N}b_N}\sum_{i=1}^{N}\Bigg\{\big(g(W_i,X_i)-g\big(F_W^{-1}\big(F_X(X_i)\big),X_i\big)\big)\\
&\cdot\int_{u_2} K\bigg(\frac{W_i-F_W^{-1}(F_X(X_i))}{b_N}+\frac{f_X(X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot u_2,u_2\bigg)\,\mathrm{d}u_2\\
&-\mathbb{E}\bigg[\big(g(W,X)-g\big(F_W^{-1}\big(F_X(X)\big),X\big)\big)\\
&\cdot\int_{u_2} K\bigg(\frac{W_i-F_W^{-1}(F_X(X))}{b_N}+\frac{f_X(X)}{f_W(F_W^{-1}(F_X(X)))}\cdot u_2,u_2\bigg)\,\mathrm{d}u_2\bigg]\Bigg\}\\
&+o_p(1).
\end{aligned}
$$

Having checked the conditions for Theorem A.2, the second part of the result in the lemma follows directly from the second part of the theorem. □

Proof of Lemma A.17. We prove the result in three parts. First, we show

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}g\big(\hat{F}_W^{-1}\big(F_X(X_i)\big),X_i\big)-\frac{1}{N}\sum_{i=1}^{N}g\big(F_W^{-1}\big(F_X(X_i)\big),X_i\big)\\
&=\frac{1}{N}\sum_{i=1}^{N}g_W\big(F_W^{-1}\big(F_X(X_i)\big),X_i\big)\cdot\big(\hat{F}_W^{-1}\big(F_X(X_i)\big)-F_W^{-1}\big(F_X(X_i)\big)\big)\\
&\quad+o_p\big(N^{-1/2}\big).
\end{aligned}
\tag{C.16}
$$

Second, we prove that

$$\frac{1}{N}\sum_{i=1}^{N} g_W\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\cdot\big(\hat{F}_W^{-1}\big(F_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)), X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_W\big(F_W^{-1}\big(F_X(X_i)\big)\big) - F_X(X_i)\big) + o_p\big(N^{-1/2}\big). \tag{C.17}$$

Third, we show that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)), X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_W\big(F_W^{-1}\big(F_X(X_i)\big)\big) - F_X(X_i)\big)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\psi_W^{\text{pam}}(W_i) + o_p\big(N^{-1/2}\big). \tag{C.18}$$

Together these three claims, (C.16)–(C.18), imply the result in the lemma.

First, we prove (C.16):

$$\bigg|\frac{1}{N}\sum_{i=1}^{N} g\big(\hat{F}_W^{-1}\big(F_X(X_i)\big), X_i\big) - \frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)$$
$$- \frac{1}{N}\sum_{i=1}^{N} g_W\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\cdot\big(\hat{F}_W^{-1}\big(F_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big)\bigg|$$
$$\le \sup_{x\in\mathbb{X}}\big|g\big(\hat{F}_W^{-1}\big(F_X(x)\big), x\big) - g\big(F_W^{-1}\big(F_X(x)\big), x\big)$$
$$- g_W\big(F_W^{-1}\big(F_X(x)\big), x\big)\cdot\big(\hat{F}_W^{-1}\big(F_X(x)\big) - F_W^{-1}\big(F_X(x)\big)\big)\big|$$
$$\le \frac{1}{2}\cdot\sup_{w\in\mathbb{W}, x\in\mathbb{X}}\left|\frac{\partial^2}{\partial w^2}g(w,x)\right|\cdot\sup_{q\in[0,1]}\big|\hat{F}_W^{-1}(q) - F_W^{-1}(q)\big|^2.$$

By Lemma A.3, it follows that for all $\delta < 1/2$, $\sup_{q\in[0,1]} N^{\delta}\cdot|\hat{F}_W^{-1}(q) - F_W^{-1}(q)| = o_p(1)$. In combination with the fact that $\frac{\partial^2 g}{\partial w^2}(w,x)$ is bounded, this implies that

$$\sup_{w\in\mathbb{W}, x\in\mathbb{X}}\left|\frac{\partial^2 g}{\partial w^2}(w,x)\right|\cdot\sup_{q\in[0,1]}\big|\hat{F}_W^{-1}(q) - F_W^{-1}(q)\big|^2 = o_p\big(N^{-1/2}\big).$$

This finishes the proof of (C.16).

Next, we prove (C.17),

$$\bigg|\frac{1}{N}\sum_{i=1}^{N} g_W\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big)\cdot\big(\hat{F}_W^{-1}\big(F_X(X_i)\big) - F_W^{-1}\big(F_X(X_i)\big)\big)$$
$$+ \frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)), X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_W\big(F_W^{-1}\big(F_X(X_i)\big)\big) - F_X(X_i)\big)\bigg|$$

$$
\begin{aligned}
&\leq \sup_{w\in\mathbb{W},x\in\mathbb{X},q\in[0,1]}\Big|g_W(w,x)\cdot\big(\hat{F}_W^{-1}(q)-F_W^{-1}(q)\big)\\
&\quad+\frac{g_W(w,x)}{f_W(F_W^{-1}(q))}\cdot\big(\hat{F}_W\big(F_W^{-1}(q)\big)-q\big)\Big|\\
&\leq \sup_{w\in\mathbb{W},x\in\mathbb{X}}\big|g_W(w,x)\big|\\
&\quad\cdot\sup_{q\in[0,1]}\Big|\big(\hat{F}_W^{-1}(q)-F_W^{-1}(q)\big)+\frac{1}{f_W(F_W^{-1}(q))}\cdot\big(\hat{F}_W\big(F_W^{-1}(q)\big)-q\big)\Big|,
\end{aligned}
$$

so that Lemma A.6 implies that (C.17) holds.

Finally, let us prove (C.18):

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_W\big(F_W^{-1}\big(F_X(X_i)\big)\big)-F_X(X_i)\big)\\
&=\frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\left(\frac{1}{N}\sum_{j=1}^{N}1_{W_j\leq F_W^{-1}(F_X(X_i))}-F_X(X_i)\right)\\
&=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(1_{F_W(W_j)\leq F_X(X_i)}-F_X(X_i)\big).
\end{aligned}
$$

This is a two-sample V-statistic. The projection is the sample average of the sum of the expectation over W_j if we fix $X_i=x$ (this expectation is zero) and the expectation over X_i if we fix $W_j=w$, which gives $\psi_W^{\text{pam}}(w)$. Thus,

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_W\big(F_W^{-1}\big(F_X(X_i)\big)\big)-F_X(X_i)\big)\\
&=\frac{1}{N}\sum_{i=1}^{N}\psi_W^{\text{pam}}(W_i)+o_p\big(N^{-1/2}\big),
\end{aligned}
$$

which is the claim in (C.18). □

Proof of Lemma A.18. We prove this result in two steps. First we prove

$$
\begin{aligned}
&\Big|\frac{1}{N}\sum_{i=1}^{N}g\big(F_w^{-1}\big(\hat{F}_X(X_i)\big),X_i\big)-\frac{1}{N}\sum_{i=1}^{N}g\big(F_w^{-1}\big(F_X(X_i)\big),X_i\big)\\
&\quad-\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_w(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_X(X_i)-F_X(X_i)\big)\Big| \qquad \text{(C.19)}\\
&=o_p\big(N^{-1/2}\big).
\end{aligned}
$$

Second, we prove

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_X(X_i)-F_X(X_i)\big)\\&\quad=\frac{1}{N}\sum_{i=1}^{N}\psi_X^{\text{pam}}(X_i)+o_p\big(N^{-1/2}\big).\end{aligned}\tag{C.20}$$

Together these two results imply the claim in Lemma A.18.

First, we prove (C.19). By a second order Taylor series expansion, using the fact that $g(w,x)$ is at least twice continuously differentiable,

$$\begin{aligned}&\Bigg|\frac{1}{N}\sum_{i=1}^{N}g\big(F_w^{-1}\big(\hat{F}_X(X_i)\big),X_i\big)-\frac{1}{N}\sum_{i=1}^{N}g\big(F_w^{-1}\big(F_X(X_i)\big),X_i\big)\\&\qquad-\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_X(X_i)-F_X(X_i)\big)\Bigg|\\&\quad\le\sup_{x\in\mathbb{X}}\Bigg|g\big(F_W^{-1}\big(\hat{F}_X(x)\big),x\big)-g\big(F_W^{-1}\big(F_X(x)\big),x\big)\\&\qquad-\frac{g_W(F_W^{-1}(F_X(x)),x)}{f_W(F_W^{-1}(F_X(x)))}\cdot\big(\hat{F}_X(x)-F_X(x)\big)\Bigg|\\&\quad\le\frac{1}{2}\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\frac{\partial^2 g}{\partial w^2}(w,x)}{f_W(w)}-\frac{g_W(w,x)\cdot\frac{\partial f}{\partial w}(w)}{(f_W(w))^2}\right|\sup_{x\in\mathbb{X}}\big|\hat{F}_X(x)-F_X(x)\big|^2\\&\quad=o_p\big(N^{-1/2}\big)\end{aligned}$$

by Lemma A.3. This finishes the proof of (C.19).

Second, we prove (C.20):

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\hat{F}_X(X_i)-F_X(X_i)\big)\\&\quad=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{g_W(F_W^{-1}(F_X(X_i)),X_i)}{f_W(F_W^{-1}(F_X(X_i)))}\cdot\big(\mathbf{1}_{X_j\le X_i}-F_X(X_i)\big).\end{aligned}$$

This is a one-sample V-statistic. To obtain the projection, we first fix $X_i=x$ and take the expectation over X_j. This gives 0 for all x. Second, we fix $X_j=x$ and take the expectation over X_i. This gives $\psi_X^{\text{pam}}(x)$ defined above. This finishes the proof of (C.20), and thus completes the proof of Lemma A.18. □

PROOF OF LEMMA A.19. Adding and subtracting terms, we have

$$
\begin{aligned}
&\hat{\beta}^{\mathrm{lc}} - \beta^{\mathrm{lc}} \\
&\quad = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - \hat{m}(W_i)\big) \\
&\qquad - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - \hat{m}(W_i)\big) \qquad \text{(C.21)} \\
&\qquad - \left(\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big)\right. \\
&\qquad \left. - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big)\right) \qquad \text{(C.22)} \\
&\qquad + \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big) \\
&\qquad - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big) \qquad \text{(C.23)} \\
&\qquad + \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - \hat{m}(W_i)\big) \\
&\qquad - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big) \qquad \text{(C.24)} \\
&\qquad + \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big) - \beta^{\mathrm{lc}}. \qquad \text{(C.25)}
\end{aligned}
$$

Because (C.23) is equal to $\beta_g^{\mathrm{lc}} - \overline{g}^{\mathrm{lc}}$, (C.24) is equal to $\beta_m^{\mathrm{lc}} - \overline{g}^{\mathrm{lc}}$, and (C.25) is equal to $\overline{g}^{\mathrm{lc}} - \beta^{\mathrm{lc}}$, it follows that it is sufficient for the proof of Lemma A.19 to show that the sum of (C.21) and (C.22) is $o_p(N^{-1/2})$. We can write the sum of (C.21) and (C.22) as

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - \hat{m}(W_i)\big) \\
&\qquad - \frac{1}{N}\sum_{i=1}^{N} \frac{\partial g}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - \hat{m}(W_i)\big) \\
&\qquad - \left(\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \hat{g}}{\partial w}(W_i, X_i) \cdot d(W_i) \cdot \big(X_i - m(W_i)\big)\right.
\end{aligned}
$$

$$- \frac{1}{N}\sum_{i=1}^{N}\frac{\partial g}{\partial w}(W_i, X_i)\cdot d(W_i)\cdot\big(X_i - m(W_i)\big)\Bigg)$$

$$= \frac{1}{N}\sum_{i=1}^{N} d(W_i)\cdot\left(\frac{\partial \hat{g}}{\partial w}(W_i, X_i) - \frac{\partial g}{\partial w}(W_i, X_i)\right)\cdot\big(m(W_i) - \hat{m}(W_i)\big)$$

$$\le \sup_{w\in\mathbb{W}}\big|d(w)\big|\cdot\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{\partial \hat{g}}{\partial w}(w,x) - \frac{\partial g}{\partial w}(x,x)\right|\cdot\sup_{w\in\mathbb{W}}\big|\hat{m}(w) - m(w)\big|$$

$$= C\cdot o_p\big(N^{-\eta}\big)\cdot o_p\big(N^{-\eta}\big)$$

for some $\eta > 1/4$, and so this expression is $o_p(N^{-1/2})$. □

Proof of Lemma A.20. The proof consists of checking the conditions for Theorem A.1 and specializing the result in Theorem A.1 to the case in the lemma.

We apply Theorem A.1 with $z = (z_1\ z_2)' = (w\ x)'$, $Z_i = (W_i\ X_i)'$, $\omega(z) = d(z_1)\cdot(z_2 - m(z_1)) = d(w)\cdot(x - m(w))$ (so that $\omega(z)$ goes smoothly to zero on the boundary of $\mathbb{Z}$), $L = 2$, and $\lambda = \binom{1}{0}$. Then $\{\kappa : \kappa \le \lambda\} = \{\kappa_0, \kappa_1\} = \left\{\binom{0}{0}, \binom{1}{0}\right\}$ and

$$h^{[\lambda]}(w,x) = \begin{pmatrix} h_1^{(\kappa_0)}(w,x) \\ h_2^{(\kappa_0)}(w,x) \\ h_1^{(\kappa_1)}(w,x) \\ h_2^{(\kappa_1)}(w,x) \end{pmatrix},$$

with

$$h_1^{(\kappa_0)}(w,x) = f_{WX}(w,x),$$
$$h_2^{(\kappa_0)}(w,x) = f_{WX}(w,x)\cdot g(w,x),$$
$$h_1^{(\kappa_1)}(w,x) = \frac{\partial}{\partial w} f_{WX}(w,x),$$
$$h_2^{(\kappa_1)}(w,x) = g(w,x)\cdot\frac{\partial}{\partial w} f_{WX}(w,x) + f_{WX}(w,,x)\cdot\frac{\partial}{\partial w} g(w,x).$$

The functional of interest is

$$n\big(h^{[\lambda]}\big) = \frac{\partial}{\partial w} g(\cdot) = \frac{h_2^{(\kappa_1)}}{h_1^{(\kappa_0)}} - \frac{h_2^{(\kappa_0)}\cdot h_1^{(\kappa_1)}}{(h_1^{(\kappa_0)})^2}.$$

The derivatives of this functional are

$$\frac{\partial}{\partial h_1^{(\kappa_0)}} n\big(h^{[\lambda]}\big) = -\frac{h_2^{(\kappa_1)}}{(h_1^{(\kappa_0)})^2} + 2\frac{h_2^{(\kappa_0)}\cdot h_1^{(\kappa_1)}}{(h_1^{(\kappa_0)})^3}$$

$$= -\frac{f_{WX}(w,x)\cdot\frac{\partial}{\partial w} g(w,x) + g(w,x)\cdot\frac{\partial}{\partial w} f_{WX}(w,x)}{(f_{WX}(w,x))^2}$$

$$+2\frac{g(w,x)\cdot f_{WX}(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{(f_{WX}(w,x))^3}$$

$$=-\frac{\frac{\partial}{\partial w}g(w,x)}{f_{WX}(w,x)}+\frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{(f_{WX}(w,x))^2},$$

$$\frac{\partial}{\partial h_2^{(\kappa_0)}}n\big(h^{[\lambda]}\big)=-\frac{h_1^{(\kappa_1)}}{(h_1^{(\kappa_0)})^2}=-\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{(f_{WX}(w,x))^2},$$

$$\frac{\partial}{\partial h_1^{(\kappa_1)}}n\big(h^{[\lambda]}\big)=-\frac{h_2^{(\kappa_0)}}{(h_1^{(\kappa_0)})^2}=-\frac{g(w,x)\cdot f_{WX}(w,x)}{(f_{WX}(w,x))^2}=-\frac{g(w,x)}{f_{WX}(w,x)},$$

$$\frac{\partial}{\partial h_2^{(\kappa_1)}}n\big(h^{[\lambda]}\big)=\frac{1}{h_1^{(\kappa_0)}}=\frac{1}{f_{WX}(w,x)},$$

$$\alpha_{\kappa_0,1}(w,x)=d(w)\cdot\big(x-m(w)\big)\cdot f_W(w,x)\cdot\left(-\frac{\frac{\partial}{\partial w}g(w,x)}{f_{WX}(w,x)}+\frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{(f_{WX}(w,x))^2}\right)$$

$$=d(w)\cdot\big(x-m(w)\big)\cdot\left(-\frac{\partial}{\partial w}g(w,x)+\frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\right),$$

$$\alpha_{\kappa_0,2}(w,x)=d(w)\cdot\big(x-m(w)\big)\cdot f_{WX}(w,x)\cdot\left(-\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{(f_{WX}(w,x))^2}\right)$$

$$=-d(w)\cdot\big(x-m(w)\big)\cdot\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)},$$

$$\alpha_{\kappa_1,1}(w,x)=d(w)\cdot\big(x-m(w)\big)\cdot f_{WX}(w,x)\cdot\left(-\frac{g(w,x)}{f_{WX}(w,x)}\right)$$

$$=-d(w)\cdot\big(x-m(w)\big)\cdot g(w,x),$$

$$\alpha_{\kappa_1,2}(w,x)=d(w)\cdot\big(x-m(w)\big)\cdot f_{WX}(w,x)\cdot\frac{1}{f_{WX}(w,x)}=d(w)\cdot\big(x-m(w)\big),$$

$$(-1)^{|\kappa_0|}\alpha_{\kappa_0,1}^{(\kappa_0)}(w,x)=\alpha_{\kappa_0,1}(w,x)$$

$$=d(w)\cdot\big(x-m(w)\big)\cdot\left(-\frac{\partial}{\partial w}g(w,x)+\frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\right),$$

$$(-1)^{|\kappa_0|}\alpha^{(\kappa_0)}_{\kappa_0,2}(w,x) = \alpha_{\kappa_0,2}(w,x) = -d(w)\cdot\big(x-m(w)\big)\cdot\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)},$$

$$\begin{aligned}(-1)^{|\kappa_1|}\alpha^{(\kappa_1)}_{\kappa_1,1}(w,x) &= \frac{\partial}{\partial w}\big(d(w)\cdot\big(x-m(w)\big)\cdot g(w,x)\big)\\ &= d(w)\cdot\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}g(w,x)\\ &\quad + g(w,x)\cdot\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}d(w)\\ &\quad - g(w,x)\cdot d(w)\cdot\frac{\partial}{\partial w}m(w),\\ (-1)^{|\kappa_1|}\alpha^{(\kappa_1)}_{\kappa_1,2}(w,x) &= -\frac{\partial}{\partial w}\big(d(w)\cdot\big(x-m(w)\big)\big)\\ &= -\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}d(w) + d(w)\cdot\frac{\partial}{\partial w}m(w).\end{aligned}$$

Then

$$\begin{aligned}&\sum_{\kappa\le\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\alpha^{(\kappa)}_{\kappa m}(w,x)\tilde{y}_{im}\\ &\quad = (-1)^{|\kappa_0|}\alpha^{(\kappa_0)}_{\kappa_0,1}(w,x) + Y_i\cdot(-1)^{|\kappa_0|}\alpha^{(\kappa_0)}_{\kappa_0,2}(w,x)\\ &\qquad + (-1)^{|\kappa_1|}\alpha^{(\kappa_1)}_{\kappa_1,1}(w,x) + Y_i\cdot(-1)^{|\kappa_1|}\alpha^{(\kappa_1)}_{\kappa_1,2}(w,x)\\ &\quad = d(w)\cdot\big(x-m(w)\big)\cdot\left(-\frac{\partial}{\partial w}g(w,x) + \frac{g(w,x)\cdot\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\right)\\ &\qquad - Y_i\cdot d(w)\cdot\big(x-m(w)\big)\cdot\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\\ &\qquad + d(w)\cdot\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}g(w,x)\\ &\qquad + g(w,x)\cdot\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}d(w) - g(w,x)\cdot d(w)\cdot\frac{\partial}{\partial w}m(w)\\ &\qquad + Y_i\cdot\left(-\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}d(w) + d(w)\cdot\frac{\partial}{\partial w}m(w)\right)\\ &\quad = -\big(Y-g(W,X)\big)\cdot\left(\frac{\frac{\partial}{\partial w}f_{WX}(W,X)}{f_{WX}(W,X)}\cdot d(w)\cdot\big(X-m(W)\big)\right.\\ &\qquad \left. + \big(X-m(W)\big)\cdot\frac{\partial}{\partial w}d(W) - d(W)\cdot\frac{\partial}{\partial w}m(W)\right).\end{aligned}$$

Since

$$\mathbb{E}\Bigg[-\big(y-g(w,x)\big)\cdot\left(\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\cdot d(w)\cdot\big(x-m(w)\big)\right.$$
$$\left.\left.+\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}d(w)-d(w)\cdot\frac{\partial}{\partial w}m(w)\right)\right]=0,$$

it follows that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{\kappa\leq\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\mathbb{E}\big[\alpha_{\kappa m}^{(\kappa)}(W_i,X_i)\tilde{Y}_{im}\big]\right)=0$$

and, therefore,

$$\begin{aligned}\sqrt{N}\big(\hat{\beta}_g^{\mathrm{lc}}-\overline{g}^{\mathrm{lc}}\big)&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{\kappa\leq\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\alpha_{\kappa m}^{(\kappa)}(W_i,X_i)\tilde{Y}_{im}\right)\\&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_g^{\mathrm{lc}}(Y_i,W_i,X_i),\end{aligned}$$

where

$$\psi_g^{\mathrm{lc}}(y,w,x)=-\big(y-g(w,x)\big)\cdot\left(\frac{\frac{\partial}{\partial w}f_{WX}(w,x)}{f_{WX}(w,x)}\cdot d(w)\cdot\big(x-m(w)\big)\right.$$
$$\left.+\big(x-m(w)\big)\cdot\frac{\partial}{\partial w}d(w)-d(w)\cdot\frac{\partial}{\partial w}m(w)\right).$$

□

Proof of Lemma A.21. We start with the inequality

$$\sup_{w\in\mathbb{W}}\left|\frac{1}{\hat{f}_W(w)}\big(\hat{f}_W(w)-f_W(w)\big)^2\right|\leq\frac{\Big(\sup_{w\in\mathbb{W}}|\hat{f}_W(w)-f_W(w)|\Big)^2}{\inf_{w\in W}|\hat{f}_W(w)|}.$$

Under the stated restriction on δ, the bandwidth sequence satisfies

$$\frac{N^{1/4}}{\sqrt{\ln(N)}}b_N^{1/2}\to\infty,\qquad N^{1/4}b_N^{s}\to 0,$$

which, by Lemma A.11, implies

$$\Big(\sup_{w\in\mathbb{W}}\big|\hat{f}_W(w)-f_W(w)\big|\Big)^2=o_p\big(N^{-1/2}\big).$$

Now observe that the denominator is bounded away from zero since, by the Triangle Inequality, we have $|\hat{f}_W(w)|+|f_W(w)|\geq|\hat{f}_W(w)-f_W(w)|$ and, therefore, $\inf_{w\in W}|\hat{f}_W(w)|\geq\sup_{w\in\mathbb{W}}|\hat{f}_W(w)-f_W(w)|-\inf_{w\in W}|f_W(w)|\geq\inf_{w\in W}|f_W(w)|-\sup_{w\in\mathbb{W}}|\hat{f}_W(w)-f_W(w)|$.

By Assumption 4.1, $\inf_{w\in W}|f_W(w)|$ is bounded away from zero, with the result then following. □

Proof of Lemma A.22. We start with the inequality

$$\sup_{w\in\mathbb{W}}\left|\frac{1}{\hat{h}_2(w)}\big(\hat{h}_1(w)-h_1(w)\big)\big(\hat{h}_2(w)-h_2(w)\big)\right| \leq \frac{\sup_{w\in\mathbb{W}}|\hat{h}_1(w)-h_1(w)|\times\sup_{w\in\mathbb{W}}|\hat{h}_2(w)-h_2(w)|}{\inf_{w\in W}|\hat{h}_2(W_i)|}.$$

The remainder of the proof is along the lines of that to Lemma A.21. □

Proof of Lemma A.23. Let $h(w)=(h_1(w),h_2(w))'=(m(w)\cdot f_W(w),f_W(w))'$. Then

$$\hat{\beta}_m^{\mathrm{lc}}=\frac{1}{N}\sum_{i=1}^N g_W(W_i,X_i)\cdot d(W_i)\cdot\left(X_i-\frac{\hat{h}_{1,\mathrm{nip}}(W_i)}{\hat{h}_{2,\mathrm{nip}}(W_i)}\right)$$

$$=\frac{1}{N}\sum_{i=1}^N g_W(W_i,X_i)\cdot d(W_i)\cdot\big(X_i-m(W_i)\big) \tag{C.26}$$

$$-\frac{1}{N}\sum_{i=1}^N g_W(W_i,X_i)\cdot d(W_i)\cdot\left(\frac{\hat{h}_{1,\mathrm{nip}}(W_i)}{\hat{h}_{2,\mathrm{nip}}(W_i)}-\frac{h_1(W_i)}{h_2(W_i)}\right). \tag{C.27}$$

Expanding the ratio[1] in (C.27) yields

$$\frac{1}{N}\sum_{i=1}^N g_W(W_i,X_i)\cdot d(W_i)\cdot\left(\frac{h_1(W_i)}{h_2(W_i)}-\frac{\hat{h}_{1,\mathrm{nip}}(W_i)}{\hat{h}_{2,\mathrm{nip}}(W_i)}\right)$$

$$=-\frac{1}{N}\sum_{i=1}^N g_W(W_i,X_i)\cdot d(W_i)\cdot\frac{1}{f_W(W_i)}\big(\hat{h}_{1,\mathrm{nip}}(W_i)-m(W_i)\hat{h}_{2,\mathrm{nip}}(W_i)\big) \tag{C.28}$$

$$-\frac{1}{N}\sum_{i=1}^N g_W(W_i,X_i)\cdot d(W_i)\cdot\frac{h_1(W_i)}{h_2(W_i)^2\hat{h}_{2,\mathrm{nip}}(W_i)}\big(\hat{h}_{2,\mathrm{nip}}(W_i)-h_2(W_i)\big)^2 \tag{C.29}$$

$$+\frac{1}{N}\sum_{i=1}^N g_W(W_i,X_i)\cdot d(W_i)\cdot\frac{h_1(W_i)}{h_2(W_i)\hat{h}_{2,\mathrm{nip}}(W_i)}\big(\hat{h}_{1,\mathrm{nip}}(W_i)-h_1(W_i)\big)\big(\hat{h}_{2,\mathrm{nip}}(W_i)-h_2(W_i)\big). \tag{C.30}$$

[1] The ratio expansion is of the form

$$\frac{\hat{a}}{\hat{b}}-\frac{a}{b}=\frac{1}{b}\left(\hat{a}-\frac{a}{b}\hat{b}\right)+\frac{a}{b^2\hat{b}}(\hat{b}-b)^2-\frac{a}{b\hat{b}}(\hat{a}-a)(\hat{b}-b).$$

First consider (C.29). By Lemma A.21,

$$\left|\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i)\cdot d(W_i)\cdot \frac{h_1(W_i)}{h_2(W_i)^2\hat{h}_{2,\text{nip}}(W_i)}\big(\hat{h}_{2,\text{nip}}(W_i)-h_2(W_i)\big)^2\right|$$
$$\leq \sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\frac{1}{f_W(w)}g_W(w,x)\cdot d(w)\cdot m(w)\right| \sup_{w\in\mathbb{W}}\left|\frac{1}{\hat{f}_W(w)}\big(\hat{f}_W(w)-f_W(w)\big)^2\right|$$
$$= o_p\big(N^{-1/2}\big)$$

if the NIP estimator is uniformly $o_p(N^{-1/4})$, which holds if $\frac{1}{4s} < \delta < \frac{1}{8}$. An analogous application of Lemma A.12 can be used to show that (C.30) is $o_p(N^{-1/2})$ under the same condition.

Now consider (C.28). We express it as the sum of a variance and a bias term:

$$-\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i)\cdot d(W_i)$$
$$\cdot \frac{1}{f_W(W_i)}\big(\hat{h}_{1,\text{nip}}(W_i)-\mathbb{E}\big[\hat{h}_{1,\text{nip}}(W_i)\big]-m(W_i)\big(\hat{h}_{2,\text{nip}}(W_i)-\mathbb{E}\big[\hat{h}_{2,\text{nip}}(W_i)\big]\big)\big)$$
$$+\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i)\cdot d(W_i)$$
$$\cdot \frac{1}{f_W(W_i)}\big(h_1(W_i)-\mathbb{E}\big[\hat{h}_{1,\text{nip}}(W_i)\big]-m(W_i)\big(h_2(W_i)-\mathbb{E}\big[\hat{h}_{2,\text{nip}}(W_i)\big]\big)\big).$$

The bias term is $O_p(N^{-1/2})$ if $\delta > \frac{1}{2s}$. After substitution of the NIP estimator, the variance term is

$$-\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i)\cdot d(W_i)$$
$$\cdot \frac{1}{f_W(W_i)}\sum_{j=0}^{s-1}\sum_{|\mu|=j}\frac{1}{\mu!}\Big(\hat{h}_{1,NW}^{(\mu)}-\mathbb{E}\big[\hat{h}_{1,NW}^{(\mu)}\big(r_b(W_i)\big)\big]$$
$$-m(W_i)\big(\hat{h}_{2,NW}^{(\mu)}\big(r_b(W_i)\big)-\mathbb{E}\big[\hat{h}_{2,NW}^{(\mu)}\big(r_b(W_i)\big)\big]\big)\big(W_i-r_b(W_i)\big)^{\mu}\Big).$$

We consider separately

$$-\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i)\cdot d(W_i)$$
$$\cdot \frac{1}{f_W(W_i)}\sum_{j=0}^{s-1}\sum_{|\mu|=j}\frac{1}{\mu!}\big(\hat{h}_{1,NW}^{(\mu)}\big(r_b(W_i)\big)-\mathbb{E}\big[\hat{h}_{1,NW}^{(\mu)}\big(r_b(W_i)\big)\big]\big)\big(W_i-r_b(W_i)\big)^{\mu} \tag{C.31}$$

and

$$\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i)\cdot d(W_i) \cdot \frac{m(W_i)}{f_W(W_i)}\sum_{j=0}^{s-1}\sum_{|\mu|=j}\frac{1}{\mu!}\big(\hat{h}_{2,NW}^{(\mu)}\big(r_b(W_i)\big) - \mathbb{E}\big[\hat{h}_{2,NW}^{(\mu)}\big(r_b(W_i)\big)\big]\big)\big(W_i - r_b(W_i)\big)^{\mu}. \tag{C.32}$$

We show that (C.31) is asymptotically equivalent to an average. The same method shows that (C.32) is also asymptotically equivalent to an average, but we omit the details. The expression (C.31) is a linear combination of terms,

$$\begin{aligned} D_\mu &= -\frac{1}{N}\sum_{i=1}^{N} g_W(W_i, X_i)\cdot d(W_i) \\ &\quad \cdot \frac{1}{f_W(W_i)}\big(\hat{h}_{1,NW}^{(\mu)}\big(r_b(W_i)\big) - \mathbb{E}\big[\hat{h}_{1,NW}^{(\mu)}\big(r_b(W_i)\big)\big]\big)\big(W_i - r_b(W_i)\big)^{\mu} \\ &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{N,\mu}(W_i, X_i, X_j, W_j), \end{aligned}$$

with

$$\begin{aligned} a_{N,\mu}(W_i, X_i, X_j, W_j) = \frac{g_W(W_i, X_i)d(W_i)}{f_W(W_i)}\bigg(&\frac{1}{b_N^{1+|\mu|}}X_j K^{(\mu)}\bigg(\frac{W_j - r_{b_N}(W_i)}{b_N}\bigg) \\ &- \mathbb{E}\bigg[\frac{1}{b_N^{1+|\mu|}}X_j K^{(\mu)}\bigg(\frac{W_j - r_{b_N}(W_i)}{b_N}\bigg)\bigg]\bigg)\big(W_i - r_{b_N}(W_i)\big)^{\mu}. \end{aligned}$$

Therefore, D_μ is a V-statistic with a kernel that depends on N so that the usual projection theorem does not apply directly. Instead, we derive the projection directly. First, we bound the second moments of $a_{N,\mu}(W_i, X_i, X_j, W_j)$. For $j \neq i$, we have

$$\begin{aligned} &\mathbb{E}\big[a_{N,\mu}(W_i, X_i, X_j, W_j)^2\big] \\ &\quad\le C\frac{\sup_{w\in\mathbb{W}}|w - r_b(w)|^{2|\mu|}}{b_N^{2|\mu|+2}}\mathbb{E}\bigg[X_j^2 K^{(\mu)}\bigg(\frac{W_j - r_{b_N}(W_i)}{b_N}\bigg)^2\bigg] \\ &\quad\le \frac{C}{b_N^2}\mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{W_j - r_{b_N}(W_i)}{b_N}\bigg)^2\bigg] \end{aligned}$$

because the conditional variance of X given W is bounded. Because given $W_i = \tilde{w}$,

$$\mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{W_j - r_{b_N}(W_i)}{b_N}\bigg)^2\bigg|W_i = \tilde{w}\bigg] = \int_{\mathbb{W}} K^{(\mu)}\bigg(\frac{w - r_{b_N}(\tilde{w})}{b_N}\bigg)^2 f_W(w)\,\mathrm{d}w,$$

we have, by a change of variables to $t=(w-r_{b_N}(\tilde{w}))/b_N$ with Jacobian b_N and the boundedness of $K^{(\mu)}(t)$ and $f_W(w)$, that this integral is bounded by Cb_N and we conclude that

$$\mathbb{E}\big[a_{N,\mu}(W_i,X_i,X_j,W_j)^2\big]=O\big(b_N^{-1}\big).$$

For $j=i$, we have

$$\begin{aligned}
&\mathbb{E}\big[a_{N,\mu}(W_i,X_i,X_i,W_i)^2\big]\\
&\quad=\frac{1}{b_N^{2+2|\mu|}}\mathbb{E}\bigg[\frac{g_W(W_i,X_i)^2d(W_i)^2}{f_W(W_i)^2}X_i^2K^{(\mu)}\bigg(\frac{W_i-r_{b_N}(W_i)}{b_N}\bigg)^2\big(W_i-r_{b_N}(W_i)\big)^{2\mu}\bigg]\\
&\qquad+\frac{1}{b_N^{2+2|\mu|}}\mathbb{E}\bigg[\frac{g_W(W_i,X_i)^2d(W_i)^2}{f_W(W_i)^2}\\
&\qquad\times\mathbb{E}\bigg[XK^{(\mu)}\bigg(\frac{W-r_{b_N}(W_i)}{b_N}\bigg)\bigg]^2\big(W_i-r_{b_N}(W_i)\big)^{2\mu}\bigg]\\
&\qquad-\frac{2}{b_N^{2+2|\mu|}}\mathbb{E}\bigg[\frac{g_W(W_i,X_i)^2d(W_i)^2}{f_W(W_i)^2}X_iK^{(\mu)}\bigg(\frac{W_i-r_{b_N}(W_i)}{b_N}\bigg)\\
&\qquad\times\mathbb{E}\bigg[XK^{(\mu)}\bigg(\frac{W-r_{b_N}(W_i)}{b_N}\bigg)\bigg]\big(W_i-r_{b_N}(W_i)\big)^{2\mu}\bigg].
\end{aligned}$$

The first term on the right-hand side is bounded by

$$\begin{aligned}
\frac{C}{b_N^2}\mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{W_i-r_{b_N}(W_i)}{b_N}\bigg)^2\bigg]&=\frac{C}{b_N^2}\int_{\mathbb{W}_{b_N}^I}K^{(\mu)}\bigg(\frac{w-r_{b_N}(w)}{b_N}\bigg)^2f_W(w)\,\mathrm{d}w\\
&\quad+\frac{C}{b_N^2}\int_{\mathbb{W}\setminus\mathbb{W}_{b_N}^I}K^{(\mu)}\bigg(\frac{w-r_{b_N}(w)}{b_N}\bigg)^2f_W(w)\,\mathrm{d}w,
\end{aligned}$$

where $\mathbb{W}_{b_N}^I$ is the internal set of the support. Because the argument of $K^{(\mu)}$ is 0 on the interior set, the first integral is obviously $O(b_N^{-2})$. The second integral is

$$\begin{aligned}
&\frac{C}{b_N^2}\int_{w_l}^{w_l+b_N}K^{(\mu)}\bigg(\frac{w-w_l}{b_N}-1\bigg)^2f_W(w)\,\mathrm{d}w\\
&\quad+\frac{C}{b_N^2}\int_{w_u-b_N}^{w_u}K^{(\mu)}\bigg(\frac{w-w_u}{b_N}+1\bigg)^2f_W(w)\,\mathrm{d}w.
\end{aligned}$$

Because the kernel has support $[-1,1]$ and its derivatives up to order μ are bounded so that

$$\begin{aligned}
&K^{(\mu)}\bigg(\frac{w-w_l}{b_N}-1\bigg)^2\le C1_{w_l\le w\le w_l+2b_N},\\
&K^{(\mu)}\bigg(\frac{w-w_u}{b_N}+1\bigg)^2\le C1_{w_u-2b_N\le w\le w_u},
\end{aligned}$$

the second integral by the boundedness of f_W is $O(b_N^{-1})$.

The second term on the right-hand side is bounded by

$$\frac{C}{b_N^2}\int_{\mathbb{W}}\left(\int_{\mathbb{W}} K^{(\mu)}\left(\frac{w-r_{b_N}(\tilde{w})}{b_N}\right)f_W(w)\,\mathrm{d}w\right)^2 f_W(\tilde{w})\,\mathrm{d}\tilde{w}$$
$$\leq \frac{C}{b_N^2}\int_{\mathbb{W}}\int_{\mathbb{W}} K^{(\mu)}\left(\frac{w-r_{b_N}(\tilde{w})}{b_N}\right)^2 f_W(w)f_W(\tilde{w})\,\mathrm{d}w\,\mathrm{d}\tilde{w}.$$

This integral is $O(b_N^{-1})$ by a change of variables with Jacobian b_N in the inner integral. The third term on the right-hand side is bounded by

$$\frac{C}{b_N^2}\left|\int_{\mathbb{W}} K^{(\mu)}\left(\frac{\tilde{w}-r_{b_N}(\tilde{w})}{b_N}\right)\int_{\mathbb{W}} K^{(\mu)}\left(\frac{w-r_{b_N}(\tilde{w})}{b_N}\right)f_W(w)\,\mathrm{d}w f_W(\tilde{w})\,\mathrm{d}\tilde{w}\right| = O\big(b_N^{-1}\big)$$

by a change of variables in the inner integral. We conclude that

$$\mathbb{E}\big[a_{N,\mu}(W_i,X_i,X_i,W_i)^2\big] = O\big(b_N^{-2}\big).$$

The next step is to express D_μ as an average. Define

$$c_{N,\mu}(X_j,W_j) = \frac{1}{b_N^{1+|\mu|}}\cdot\int_{\mathbb{X}}\int_{\mathbb{W}}\frac{g_W(w,x)d(x)}{f_W(w)}\left(X_j K^{(\mu)}\left(\frac{W_j-r_{b_N}(w)}{b_N}\right)\right.$$
$$\left.-\,\mathbb{E}\left[XK^{(\mu)}\left(\frac{W-r_{b_N}(w)}{b_N}\right)\right]\right)(w-r_{b_N}w)^\mu f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x$$

and

$$E_\mu = -\frac{1}{N}\sum_{j=1}^N c_{N,\mu}(X_j,W_j).$$

Then

$$D_\mu - E_\mu = \frac{N(N-1)}{N^2}(D_{\mu,1}-E_\mu) + \left(\frac{N(N-1)}{N^2}-1\right)E_\mu + D_{\mu,2},$$

with

$$D_{\mu,1} = -\frac{1}{N(N-1)}\sum_{i\neq j=1}^N a_{N,\mu}(W_i,X_i,X_j,W_j),$$

$$D_{\mu,2} = -\frac{1}{N^2}\sum_{i=1}^N a_{N,\mu}(W_i,X_i,X_i,W_i).$$

Now

$$D_{\mu,1} - E_\mu = -\frac{1}{N(N-1)}\sum_{i\neq j=1}^N\big(a_{N,\mu}(W_i,X_i,X_j,W_j) - c_{N,\mu}(X_j,W_j)\big),$$

with

$$\mathbb{E}\big[\big(a_{N,\mu}(W_i, X_i, X_j, W_j) - c_{N,\mu}(X_j, W_j)\big) \\ \times \big(a_{N,\mu}(W_{i'}, X_{i'}, X_{j'}, W_{j'}) - c_{N,\mu}(X_{j'}, W_{j'})\big)\big] = 0$$

if (i) $i \neq i'$, $j \neq j'$, (ii) $i = i'$, $j \neq j'$, and (iii) $i \neq i'$, $j = j'$, because

$$\mathbb{E}\big[a_{N,\mu}(W_i, X_i, X_j, W_j)|W_i, X_i\big] = 0,$$
$$\mathbb{E}\big[a_{N,\mu}(W_i, X_i, X_j, W_j)\big] = 0,$$
$$\mathbb{E}\big[a_{N,\mu}(W_i, X_i, X_j, W_j)|X_j, W_j\big] = c_{N,\mu}(W_j, X_j),$$
$$\mathbb{E}\big[c_{N,\mu}(W_j, X_j)\big] = 0.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\big[(D_{\mu,1} - E_\mu)^2\big] \\
&\quad = \frac{1}{N^2(N-1)^2} \sum_{i \neq j} \sum_{i' \neq j'} \mathbb{E}\big[\big(a_{N,\mu}(W_i, X_i, X_j, W_j) - c_{N,\mu}(X_j, W_j)\big) \\
&\qquad \times \big(a_{N,\mu}(W_{i'}, X_{i'}, X_{j'}, W_{j'}) - c_{N,\mu}(X_{j'}, W_{j'})\big)\big] \\
&\quad = \frac{1}{N^2(N-1)^2} \sum_{i \neq j} \mathbb{E}\big[\big(a_{N,\mu}(W_i, X_i, X_j, W_j) - c_{N,\mu}(X_j, W_j)\big)^2\big].
\end{aligned}$$

Because

$$\begin{aligned}
&\mathbb{E}\big[\big(a_{N,\mu}(W_i, X_i, X_j, W_j) - c_{N,\mu}(X_j, W_j)\big)^2\big] \\
&\quad = \mathbb{E}\big[\big(a_{N,\mu}(W_i, X_i, X_j, W_j)\big)^2\big] - \mathbb{E}\big[\big(c_{N,\mu}(X_j, W_j)\big)^2\big] \\
&\quad \leq \mathbb{E}\big[\big(a_{N,\mu}(W_i, X_i, X_j, W_j)\big)^2\big] = O\big(b_N^{-1}\big),
\end{aligned}$$

we have

$$\mathbb{E}\big[(D_{\mu,1} - E_\mu)^2\big] = O\big(N^{-2} b_N^{-1}\big)$$

so that

$$\frac{N(N-1)}{N^2}(D_{\mu,1} - E_\mu) = O_p\big(N^{-1} b_N^{-1/2}\big).$$

Also

$$\begin{aligned}
&\mathbb{E}\big[c_{N,\mu}(X_j, W_j)^2\big] \\
&\quad \leq \frac{1}{b_N^{2+2|\mu|}} \mathbb{E}\bigg[\bigg(\int_{\mathbb{X}} \int_{\mathbb{W}} \frac{g_W(w, x) d(x)}{f_W(w)} \\
&\qquad \times X_j K^{(\mu)}\bigg(\frac{W_j - r_{b_N}(w)}{b_N}\bigg)(w - r_{b_N} w)^\mu f_{WX}(w, x)\, \mathrm{d}w\, \mathrm{d}x\bigg)^2\bigg]
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{b_N^{2+2|\mu|}} \mathbb{E}\biggl[\int_{\mathbb{X}} \int_{\mathbb{W}} \frac{g_W(w,x)^2 d(x)^2}{f_W(w)^2} \\
&\quad \times X_j^2 K^{(\mu)}\biggl(\frac{W_j - r_{b_N}(w)}{b_N}\biggr)^2 (w - r_{b_N} w)^{2\mu} f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x\biggr] \\
&\leq \frac{C}{b_N^2} \int_{\mathbb{W}} \int_{\mathbb{X}} \int_{\mathbb{W}} K^{(\mu)}\biggl(\frac{\tilde{w} - r_{b_N}(w)}{b_N}\biggr)^2 f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x f_W(\tilde{w})\,\mathrm{d}\tilde{w} \\
&= O\bigl(b_N^{-1}\bigr)
\end{aligned}
$$

by a change of variables in the outer integral, so that

$$
\biggl(\frac{N(N-1)}{N^2} - 1\biggr) E_\mu = O_p\bigl(N^{-1} b_N^{-1/2}\bigr).
$$

Finally,

$$
\begin{aligned}
\mathbb{E}\bigl[|D_{\mu,2}|\bigr] &\leq \frac{1}{N}\mathbb{E}\bigl[\bigl|a_{N,\mu}(W_i,X_i,X_i,W_i)\bigr|\bigr] \leq \frac{1}{N}\sqrt{\mathbb{E}\bigl[a_{N,\mu}(W_i,X_i,X_i,W_i)^2\bigr]} \\
&= O\bigl(N^{-1} b_N^{-1}\bigr)
\end{aligned}
$$

so that

$$
D_{\mu,2} = O_p\bigl(N^{-1} b_N^{-1}\bigr).
$$

Therefore, if $\delta < 1/2$, then

$$
D_\mu = E_\mu + o_p\bigl(N^{-1/2}\bigr).
$$

Under the same condition, (C.32) is a linear combination of terms

$$
\begin{aligned}
F_\mu &= \frac{1}{N}\sum_{i=1}^{N} \frac{g_W(W_i,X_i) d(W_i) m(W_i)}{f_W(W_i)} \bigl(\hat{h}_{2,NW}^{(\mu)}\bigl(r_b(W_i)\bigr) - \mathbb{E}\bigl[\hat{h}_{2,NW}^{(\mu)}\bigl(r_b(W_i)\bigr)\bigr]\bigr) \\
&\quad \times \bigl(W_i - r_b(W_i)\bigr)^\mu \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} e_{N,\mu}(W_i,X_i,W_j),
\end{aligned}
$$

with

$$
\begin{aligned}
e_{N,\mu}(W_i,X_i,W_j) = \frac{g_W(W_i,X_i) d(W_i) m(W_i)}{f_W(W_i)} &\biggl(\frac{1}{b_N^{1+|\mu|}} K^{(\mu)}\biggl(\frac{W_j - r_{b_N}(W_i)}{b_N}\biggr) \\
&- \mathbb{E}\biggl[\frac{1}{b_N^{1+|\mu|}} K^{(\mu)}\biggl(\frac{W - r_{b_N}(W_i)}{b_N}\biggr)\biggr]\biggr) \bigl(W_i - r_{b_N}(W_i)\bigr)^\mu
\end{aligned}
$$

such that

$$
F_\mu = G_\mu + o_p\bigl(N^{-1/2}\bigr)
$$

with

$$G_\mu = \frac{1}{N}\sum_{i=1}^{N} f_{N,\mu}(W_j)$$

and

$$f_{N,\mu}(W_j) = \frac{1}{b_N^{1+|\mu|}}\cdot\int_{\mathbb{X}}\int_{\mathbb{W}} \frac{g_W(w,x)d(x)m(w)}{f_W(w)}\biggl(K^{(\mu)}\biggl(\frac{W_j - r_{b_N}(w)}{b_N}\biggr) \\ - \mathbb{E}\biggl[K^{(\mu)}\biggl(\frac{W - r_{b_N}(w)}{b_N}\biggr)\biggr]\biggr)(w - r_{b_N}w)^\mu f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x.$$

The final step is to show that

$$E_0 = -\frac{1}{N}\sum_{j=1}^{N} c_{N,\mu}(X_j, W_j) = -\frac{1}{N}\sum_{j=1}^{N}\bigl(\zeta_j - \mathbb{E}[\zeta_j]\bigr) + o_p\bigl(N^{-1/2}\bigr)$$

with

$$\zeta_j = X_j\mathbb{E}\bigl[g_W(W_j, X)d(X)|W_j\bigr],$$

where the expectation is over the conditional distribution of X given W, and

$$G_0 = \frac{1}{N}\sum_{j=1}^{N} c_{N,\mu}(X_j, W_j) = \frac{1}{N}\sum_{j=1}^{N}\bigl(\xi_j - \mathbb{E}[\xi_j]\bigr) + o_p\bigl(N^{-1/2}\bigr)$$

with

$$\xi_j = m(W_j)\mathbb{E}\bigl[g_W(W_j, X)d(X)|W_j\bigr]$$

and

$$E_\mu = o_p\bigl(N^{-1/2}\bigr), \qquad G_\mu = o_p\bigl(N^{-1/2}\bigr)$$

for $|\mu| \geq 1$. We only consider E_0 and E_μ. The proof for G_0 and G_μ is analogous. Define

$$\psi_{N,\mu,j} = \frac{1}{b_N^{1+|\mu|}}\int_{\mathbb{X}}\int_{\mathbb{W}} \frac{g_W(w,x)d(x)}{f_W(w)} \\ \times X_j K^{(\mu)}\biggl(\frac{W_j - r_{b_N}(w)}{b_N}\biggr)(w - r_{b_N}w)^\mu f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x$$

so that $c_{N,\mu}(X_j, W_j) = \psi_{N,\mu,j} - \mathbb{E}[\psi_{N,\mu,j}]$. Now

$$\psi_{N,0,j} = \psi_{N,0,j,0} + \psi_{N,0,j,1}$$

with

$$\psi_{N,0,i,0} = \frac{1}{b_N}\int_{\mathbb{X}}\int_{w_l+b_N}^{w_u-b_N} \frac{g_W(w,x)d(x)}{f_W(w)} X_j K\biggl(\frac{W_j - w}{b_N}\biggr) f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x$$

and

$$\psi_{N,0,j,1} = \frac{1}{b_N}\int_{\mathbb{X}}\int_{w_l}^{w_l+b_N} \frac{g_W(w,x)d(x)}{f_W(w)} X_j K\left(\frac{W_j - w_l}{b_N}\right) f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x + \frac{1}{b_N}\int_{\mathbb{X}}\int_{w_u-b_n}^{w_u} \frac{g_W(w,x)d(x)}{f_W(w)} X_j K^{(\mu)}\left(\frac{W_j - w_u}{b_N}\right) f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x$$

so that

$$E_0 = -\frac{1}{N}\sum_{j=1}^N\bigl(\psi_{N,0,j,0} - \mathbb{E}[\psi_{N,0,j,0}]\bigr) - \frac{1}{N}\sum_{j=1}^N\bigl(\psi_{N,0,j,1} - \mathbb{E}[\psi_{N,0,j,1}]\bigr).$$

Obviously

$$\mathbb{E}\left[\left(-\frac{1}{N}\sum_{i=1}^N\bigl(\psi_{N,0,j,0} - \mathbb{E}[\psi_{N,0,j,0}]\bigr) + \frac{1}{N}\sum_{j=1}^N\bigl(\zeta_j - \mathbb{E}[\zeta_j]\bigr)\right)^2\right] \leq \frac{1}{N}\mathbb{E}\bigl[(\psi_{N,0,j,0} - \zeta_j)^2\bigr].$$

By a change of variables to $t = (W_j - w)/b_N$ with Jacobian b_N,

$$\psi_{N,0,j,0} = \int_{\mathbb{X}}\int_{-1}^1 1_{1+(W_j-w_u)/b_N \leq t \leq -1+(W_j-w_l)/b_N} \frac{g_W(W_j - b_N t, x)d(x)}{f_W(W_j - b_N t)} \times X_j K(t) f_{WX}(W_j - b_N t, x)\,\mathrm{d}t\,\mathrm{d}x$$

so that

$$\begin{aligned}
&|\psi_{N,0,j,0} - \zeta_i| \\
&\leq \int_{\mathbb{X}}\int_{-1}^1 1_{1+(W_j-w_u)/b_N \leq t \leq -1+(W_j-w_l)/b_N} \\
&\quad\times \left|\frac{g_W(W_j - b_N t, x)d(x)}{f_W(W_j - b_N t)} f_{WX}(W_j - b_N t, x) - \frac{g_W(W_j, x)d(x)}{f_W(W_j)} f_{WX}(W_j, x)\right| \\
&\quad\times |X_j|\bigl|K(t)\bigr|\,\mathrm{d}t\,\mathrm{d}x \\
&\quad+ \int_{\mathbb{X}}\left|\frac{g_W(W_j, x)d(x)}{f_W(W_j)} f_{WX}(W_j, x)\right|\mathrm{d}x \\
&\quad\times |X_j|\int_{-1}^1 |1_{1+(W_j-w_u)/b_N \leq t \leq -1+(W_j-w_l)/b_N} - 1|\bigl|K(t)\bigr|\,\mathrm{d}t.
\end{aligned}$$

By the mean value theorem, the first term on the right-hand side is $b_N|X_j|p(W_j)$ with $p(W_j)$ a (generic) bounded function of W_j. The second term on the right-hand side is $|X_j|p(W_j)(1 - \Pr(w_l + 2b_N \leq W_j \leq w_u - 2b_N))$. Therefore,

$$|\psi_{N,0,j,0} - \zeta_i| \leq |X_j|p(W_j)\bigl(b_N + \bigl(1 - \Pr(w_l + 2b_N \leq W_j \leq w_u - 2b_N)\bigr)\bigr)$$

so that

$$\mathbb{E}\bigl[(\psi_{N,0,j,0} - \zeta_j)^2\bigr] = O(b_N)$$

and

$$\left|-\frac{1}{N}\sum_{i=1}^{N}\left(\psi_{N,0,j,0}-\mathbb{E}[\psi_{N,0,j,0}]\right)+\frac{1}{N}\sum_{j=1}^{N}\left(\zeta_j-\mathbb{E}[\zeta_j]\right)\right|=o_p\left(N^{-1/2}\right)$$

if $\delta<\frac{1}{2}$. For $\psi_{N,0,j,1}$, we consider the first term on the right-hand side:

$$\begin{aligned}&\left|K\left(\frac{W_j-w_l}{b_N}\right)X_j\frac{1}{b_N}\int_{w_l}^{w_l+b_N}\int_{\mathbb{X}}\frac{g_W(w,x)d(x)}{f_W(w)}f_{WX}(w,x)\,\mathrm{d}x\,\mathrm{d}w\right|\\&\quad\le C|X_j|1_{w_l\le W_j\le w_l+b_N}.\end{aligned}$$

For the other term on the right-hand side, we get a similar bound and we conclude that

$$\mathbb{E}\left[\psi_{N,0,j,1}^2\right]=O(b_N)$$

so that if $\delta<\frac{1}{2}$, then

$$\left|\frac{1}{N}\sum_{j=1}^{N}\left(\psi_{N,0,j,1}-\mathbb{E}[\psi_{N,0,j,1}]\right)\right|=o_p\left(N^{-1/2}\right).$$

Finally, if $\mu\ge 1$, then

$$\begin{aligned}\psi_{N,\mu,j}=&\frac{1}{b_N^{1+|\mu|}}\int_{\mathbb{X}}\int_{w_l}^{w_l+b_N}\frac{g_W(w,x)d(x)}{f_W(w)}\\&\times X_jK^{(\mu)}\left(\frac{W_j-w_l}{b_N}\right)(w-w_l)^{\mu}f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x\\&+\frac{1}{b_N^{1+|\mu|}}\int_{\mathbb{X}}\int_{w_u-b_N}^{w_u}\frac{g_W(w,x)d(x)}{f_W(w)}\\&\times X_jK^{(\mu)}\left(\frac{W_j-w_u}{b_N}\right)(w-w_u)^{\mu}f_{WX}(w,x)\,\mathrm{d}w\,\mathrm{d}x.\end{aligned}$$

The first term on the right-hand side is bounded by

$$\begin{aligned}&\left|K^{(\mu)}\left(\frac{W_j-w_l}{b_N}\right)\right||X_j|\frac{1}{b_N}\int_{\mathbb{X}}\int_{w_l}^{w_l+b_N}\left|\frac{g_W(w,x)d(x)}{f_W(w)}f_{WX}(w,x)\right|\mathrm{d}w\,\mathrm{d}x\\&\quad\le C|X_j|1_{w_l\le W_j\le w_l+b_N}\end{aligned}$$

so that

$$\mathbb{E}\left[\psi_{N,\mu,j}^2\right]=O(b_N)$$

and, therefore,

$$E_\mu=-\sum_{j=1}^{N}\left(\psi_{N,\mu,j}-\mathbb{E}[\psi_{N,\mu,j}]\right)=o_p\left(N^{-1/2}\right)$$

if $\delta < \frac{1}{2}$. □

Proof of Theorem A.3. Because the class of doubly averaged estimators has not been considered previously, we provide a somewhat detailed proof. The proof consists of four steps. In the first, we approximate the estimator by a linear function of the kernel estimator $\hat{h}_{\text{nip},s}$ (linearization). Formally, we show that

$$\begin{aligned} V = {} & \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'} \big(h(Z_{1j}, Z_{2k})\big)\big(\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k}) - h(Z_{1j}, Z_{2k})\big) \\ & + O_p\big(\sqrt{N}|\hat{h}_{\text{nip},s} - h|^2\big). \end{aligned} \tag{C.33}$$

By the assumptions and Lemma A.11, the remainder term is $o_p(1)$.

In the second step, we express the difference between the linearized estimator and the estimand as the sum of a bias term (that is asymptotically negligible) and a variance term (bias–variance decomposition). The bias term will be shown to satisfy

$$\begin{aligned} & \left| \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'} \big(h(Z_{1j}, Z_{2k})\big)\big(\mathbb{E}\big[\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k})\big] - h(Z_{1j}, Z_{2k})\big) \right| \\ & = O\big(\sqrt{N} b_N^p\big). \end{aligned} \tag{C.34}$$

By the assumption on the bandwidth rate, the remainder term is $o(1)$. Note that by $\mathbb{E}[\hat{h}(Z_{i1}, Z_{i2})]$, we mean the expectation of $\hat{h}(z_1, z_2)$, evaluated at $z_1 = Z_{1i}$ and $z_2 = Z_{2j}$: the expectation is taken over the estimator of the function $h(\cdot)$.

The second step leaves us with

$$V = W + O_p\big(\sqrt{N}|\hat{h}_{\text{nip},s} - h|^2\big) + O\big(\sqrt{N} b_N^p\big),$$

where

$$W = \frac{1}{N\sqrt{N}} \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{\partial n}{\partial h'} \big(h(Z_{1j}, Z_{2k})\big)\big(\hat{h}(Z_{1j}, Z_{2k}) - \mathbb{E}\big[\hat{h}_{\text{nip},s}(Z_{1j}, Z_{2k})\big]\big). \tag{C.35}$$

Define

$$\nu(z_1, z_2) = \frac{\partial n}{\partial h'} \big(h(z_1, z_2)\big)$$

and

$$\begin{aligned} & a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) \\ & = \nu(Z_{1j}, Z_{2k})' \cdot \left(\frac{1}{b_N^{L+|\mu|}} \tilde{Y}_i K^{(\mu)} \left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N} \right) \right. \\ & \quad \left. - \mathbb{E}_{\tilde{Y}Z} \left[\frac{1}{b_N^{L+|\mu|}} \tilde{Y} K^{(\mu)} \left(\frac{Z - r_{b_N}(Z_{1j}, Z_{2k})}{b_N} \right) \right] \right) \cdot \left(\begin{pmatrix} Z_{1j} \\ Z_{2k} \end{pmatrix} - r_{b_N}(Z_{1j}, Z_{2k}) \right)^{\mu} \end{aligned}$$

so that

$$W = \sum_{\mu:|\mu|\leq s-1} \frac{1}{\mu!} W_\mu, \tag{C.36}$$

where

$$W_\mu = \frac{1}{N^2\sqrt{N}} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}).$$

Define

$$\begin{aligned} c_{N,\mu}(\tilde{Y}_i, Z_i) = \frac{1}{b_N^{L+|\mu|}} \int_{\mathcal{Z}_2}\int_{\mathcal{Z}_1} \nu(z_1, z_2)' \biggl(\tilde{Y}_i K^{(\mu)}\biggl(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\biggr) \\ - \mathbb{E}_{\tilde{Y}Z}\biggl[\tilde{Y} K^{(\mu)}\biggl(\frac{Z - r_{b_N}(z_1, z_2)}{b_N}\biggr)\biggr]\biggr) \\ \times \bigl((z_1' \; z_2')' - r_{b_N}(z_1, z_2)\bigr)^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2)\, \mathrm{d}z_1\, \mathrm{d}z_2 \end{aligned}$$

and

$$U_\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_{N,\mu}(\tilde{Y}_i, Z_i),$$

or, equivalently,

$$\begin{aligned} U_\mu = \sqrt{N} \int_{\mathcal{Z}_2}\int_{\mathcal{Z}_1} \nu(z_1, z_2)' \bigl(\hat{h}_{NW}^{(\mu)}\bigl(r_{b_N}(z_1, z_2)\bigr) - \mathbb{E}\bigl[\hat{h}_{NW}^{(\mu)}\bigl(r_{b_N}(z_1, z_2)\bigr)\bigr]\bigr) \\ \times \bigl((z_1' \; z_2')' - r_{b_N}(z_1, z_2)\bigr)^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2)\, \mathrm{d}z_1\, \mathrm{d}z_2. \end{aligned}$$

In the third step, we show that

$$W_\mu = U_\mu + O_p\bigl(N^{-1/2} b_N^{-L}\bigr). \tag{C.37}$$

In the fourth step, we show that

$$\begin{aligned} U_0 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \biggl\{ \frac{\partial n}{\partial h}\bigl(h(Z_i)\bigr)' \tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) \\ - \mathbb{E}_Z\biggl[\frac{\partial n}{\partial h}\bigl(h(Z)\bigr)' \tilde{Y} f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i})\biggr]\biggr\} + o_p(1), \end{aligned} \tag{C.38}$$

which gives us the representation in the theorem.

In the fifth and final step, we show that we can ignore U_μ for μ such that $|\mu| \geq 1$, because for such μ,

$$U_\mu = O_p(b_N). \tag{C.39}$$

Proving these statements implies the result in the theorem.

Now we turn to proving each of the statements (C.33), (C.34), (C.37), (C.38), and (C.39).

STEP 1 (Linearization).

In the first step of the proof, we prove equality (C.33). First define

$$d(z_1, z_2) \equiv n\big(\hat{h}_{\mathrm{nip},s}(z_1, z_2)\big) - n\big(h(z_1, z_2)\big) - \frac{\partial n}{\partial h'}\big(h(z_1, z_2)\big)\big(\hat{h}_{\mathrm{nip},s}(z_1, z_2) - h(z_1, z_2)\big).$$

By a second order Taylor series expansion of $n(\hat{h}_{\mathrm{nip},s}(Z_{1i}, Z_{2j}))$ around $h(Z_{1i}, Z_{2j})$, we have

$$\begin{aligned}
\big|d(z_1, z_2)\big| &= \frac{1}{2}\bigg|\big(\hat{h}_{\mathrm{nip},s}(z_1, z_2) - h(z_1, z_2)\big)' \\
&\quad \times \frac{\partial^2 n}{\partial h\, \partial h'}\big(\overline{h}(z_1, z_2)\big)\big(\hat{h}_{\mathrm{nip},s}(z_1, z_2) - h(z_1, z_2)\big)\bigg| \\
&\leq \sup_z \left|\frac{\partial^2 n}{\partial h\, \partial h'}\big(h(z)\big)\right| \cdot \big|\hat{h}_{\mathrm{nip},s}(z_1, z_2) - h(z_1, z_2)\big|^2 \\
&\leq C \cdot |\hat{h}_{\mathrm{nip},s} - h|^2,
\end{aligned}$$

with $\overline{h}(z_1, z_2)$ intermediate between $\hat{h}_{\mathrm{nip},s}(z_1, z_2)$ and $h(z_1, z_2)$ so that

$$\left|\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N} d(Z_{1j}, Z_{2k})\right| \leq C|\hat{h}_{\mathrm{nip},s} - h|^2 = O_p\big(|\hat{h}_{\mathrm{nip},s} - h|^2\big).$$

Hence,

$$\begin{aligned}
&\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\big[n\big(\hat{h}_{\mathrm{nip},s}(Z_{1j}, Z_{2k})\big) - n\big(h(Z_{1j}, Z_{2k})\big)\big] \\
&\quad = \frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}\big(h(Z_{1j}, Z_{2k})\big)\big(\hat{h}_{\mathrm{nip},s}(Z_{1j}, Z_{2k}) - h(Z_{1j}, Z_{2k})\big) \\
&\qquad + \frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N} d(Z_{1j}, Z_{2k}) \\
&\quad = \frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}\big(h(Z_{1j}, Z_{2k})\big)\big(\hat{h}_{\mathrm{nip},s}(Z_{1j}, Z_{2k}) - h(Z_{1j}, Z_{2k})\big) \\
&\qquad + O_p\big(\sqrt{N}|\hat{h}_{\mathrm{nip},s} - h|^2\big)
\end{aligned}$$

so that the linearization remainder has the same stochastic order as $\sqrt{N}|\hat{h}_{\mathrm{nip},s} - h_0|^2$.

Step 2 (Bias–variance decomposition).

In the second step of the proof, we verify equation (C.34). Define

$$E=\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}\big(h(Z_{1j},Z_{2k})\big)\big(\hat{h}_{\mathrm{nip},s}(Z_{1j},Z_{2k})-h(Z_{1j},Z_{2k})\big)$$

so that

$$V=E+O_p\big(\sqrt{N}|\hat{h}_{\mathrm{nip},s}-h|^2\big).$$

We decompose E into a bias and variance part, $E=E_{\mathrm{bias}}+W$, where

$$E_{\mathrm{bias}}=\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}\big(h(Z_{1j},Z_{2k})\big)\big(\mathbb{E}\big[\hat{h}_{\mathrm{nip},s}(Z_{1j},Z_{2k})\big]-h(Z_{1j},Z_{2k})\big)$$

and W is defined in (C.35). The bias part is bounded by

$$\begin{aligned}&\left|\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial n}{\partial h'}\big(h(Z_{1j},Z_{2k})\big)\big(\mathbb{E}\big[\hat{h}_{\mathrm{nip},s}(Z_{1j},Z_{2k})\big]-h(Z_{1j},Z_{2k})\big)\right|\\&\quad\le\sup_{z\in\mathbb{Z}}\left|\frac{\partial n}{\partial h}\big(h(z)\big)\right|\sqrt{N}\big|\mathbb{E}[\hat{h}_{\mathrm{nip},s}]-h\big|=O\big(\sqrt{N}b_N^p\big)\end{aligned}$$

due to smoothness of the function and Lemma A.9.

Step 3 (Projection).

In the third step of the proof, we prove equation (C.37), $W_\mu=U_\mu+O_p(N^{-1/2}b_N^{-L})$. This is the most complicated step. First, note that

$$W_\mu=\frac{1}{N^2\sqrt{N}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})$$

is a third order V-statistic with kernel (that depends on N) $a_{N,\mu}$. We show that this V-statistic is asymptotically equivalent to a projection that is a single sum. Because the kernel depends on N, we cannot use a standard result.

The projection of W_μ is

$$U_\mu=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}c_{N,\mu}(\tilde{Y}_i,Z_i),$$

with

$$\begin{aligned}&c_{N,\mu}(\tilde{Y}_i,Z_i)\\&\quad=\frac{1}{b_N^{L+|\mu|}}\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1}\nu(z_1,z_2)'\end{aligned}$$

$$\times \left(\tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N} \right) - \mathbb{E}_{\tilde{Y}Z}\left[\tilde{Y} K^{(\mu)}\left(\frac{Z - r_{b_N}(z_1, z_2)}{b_N} \right) \right] \right)$$
$$\times \left((z_1' \; z_2')' - r_{b_N}(z_1, z_2) \right)^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2) \, \mathrm{d}z_1 \, \mathrm{d}z_2.$$

The projection remainder is

$$W_\mu - U_\mu = \frac{N(N-1)(N-2)}{N^3}(W_{\mu,1} - U_\mu) + \left(\frac{N(N-1)(N-2)}{N^3} - 1 \right) U_\mu + W_{\mu,2} + W_{\mu,3} + W_{\mu,4} + W_{\mu,5} \tag{C.40}$$

with

$$W_{\mu,1} \equiv \frac{\sqrt{N}}{N(N-1)(N-2)} \sum_{i \neq j \neq k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}),$$

$$W_{\mu,2} \equiv \frac{\sqrt{N}}{N^3} \sum_{i = j \neq k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k}),$$

$$W_{\mu,3} \equiv \frac{\sqrt{N}}{N^3} \sum_{i = k \neq j} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2i}),$$

$$W_{\mu,4} \equiv \frac{\sqrt{N}}{N^3} \sum_{i \neq j = k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2j}),$$

$$W_{\mu,5} \equiv \frac{\sqrt{N}}{N^3} \sum_{i = j = k} a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k}).$$

We prove that the projection remainder $W_\mu - U_\mu = O_p(N^{-1/2} b_N^{-L})$ by proving the following six equalities:

$$W_{\mu,1} - U_\mu = O_p\left(N^{-1} b_N^{-L/2}\right), \tag{C.41}$$

$$\left(\frac{N(N-1)(N-2)}{N^3} - 1 \right) U_\mu = O_p\left(N^{-1} b_N^{-L/2}\right), \tag{C.42}$$

$$W_{\mu,2} = O_p\left(N^{-1/2} b_N^{-L+L_2/2}\right), \tag{C.43}$$

$$W_{\mu,3} = O_p\left(N^{-1/2} b_N^{-L+L_1/2}\right), \tag{C.44}$$

$$W_{\mu,4} = O_p\left(N^{-3} b_N^{-L/2}\right), \tag{C.45}$$

$$W_{\mu,5} = O_p\left(N^{-1/2} b_N^{-L}\right). \tag{C.46}$$

To prove these results, we establish bounds on the second moment of $a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})$. This will be relatively straightforward if $i \neq j$ and $i \neq k$. The derivation of the bound is more involved if $i = j$ and/or $i = k$. We could simplify the proof by omitting these observations and redefining the estimator by restricting the averaging to observations with

$i \neq j$ and $i \neq k$. This would amount to redefining the kernel estimator in (A.7) by omitting observations $i = j$ and $i = k$ in $\hat{h}_{\text{nip},s}$. We will keep these observations and derive bounds on all second moments. We derive the following bounds, considering four separate cases (note that the bounds do not depend on μ):

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2\big] = O\big(b_N^{-L}\big), \quad j \neq i, \text{ and } k \neq i, \tag{C.47}$$

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})^2\big] = O\big(b_N^{-2L}\big), \quad i = j = k, \tag{C.48}$$

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})^2\big] = O\big(b_N^{-2L+L_2}\big), \quad k \neq i = j, \tag{C.49}$$

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2\big] = O\big(b_N^{-2L+L_1}\big), \quad j \neq i = k. \tag{C.50}$$

Step 3A (Equation (C.47)).

First, if $j \neq i$ and $k \neq i$, then

$$\begin{aligned}
&\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2\big] \\
&\quad \leq \frac{1}{b_N^{2L+2|\mu|}} \mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\bigg)^2 \\
&\qquad \times \nu(Z_{1j}, Z_{2k})' \tilde{Y}_i \tilde{Y}_i' \nu(Z_{1j}, Z_{2k}) \big((Z_{1j}' \; Z_{2k}')' - r_{b_N}(Z_{1j}, Z_{2k})\big)^{2\mu}\bigg] \\
&\quad \leq \frac{\sup_{z \in \mathbb{Z}} |z - r_{b_N}(z)|^{2\mu}}{b_N^{2L+2|\mu|}} \mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\bigg)^2 \\
&\qquad \times \nu(Z_{1j}, Z_{2k})' \tilde{Y}_i \tilde{Y}_i' \nu(Z_{1j}, Z_{2k})\bigg] \\
&\quad \leq \frac{1}{b_N^{2L}} \mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\bigg)^2 \\
&\qquad \times \nu(Z_{1j}, Z_{2k})' \tilde{Y}_i \tilde{Y}_i' \nu(Z_{1j}, Z_{2k})\bigg].
\end{aligned}$$

Now by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\bigg)^2 \nu(Z_{1j}, Z_{2k})' \tilde{Y}_i \tilde{Y}_i' \nu(Z_{1j}, Z_{2k})\bigg] \\
&\quad = \mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\bigg)^2 \big(\nu(Z_{1j}, Z_{2k})' \tilde{Y}_i\big)^2\bigg] \\
&\quad \leq \mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\bigg)^2 \big|\nu(Z_{1j}, Z_{2k})\big|^2 |\tilde{Y}_i|^2\bigg] \\
&\quad = \mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\bigg)^2 \big|\nu(Z_{1j}, Z_{2k})\big|^2 \mathbb{E}\big[|\tilde{Y}_i|^2 \big| Z_i\big]\bigg].
\end{aligned}$$

By Assumption 4.1, $\mathbb{E}[|\tilde{Y}|^2|Z=z]$ and ν are bounded on $\mathbb{Z}$ so that it is bounded by (condition on Z_{1j} and Z_{2k})

$$C\mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_1, Z_2)}{b_N}\right)^2\right] = C\int_{\mathbb{Z}} K^{(\mu)}\left(\frac{z - r_{b_N}(Z_1, Z_2)}{b_N}\right)^2 f_Z(z)\,\mathrm{d}z$$

and by a change of variables to $t = (z - r_{b_N}(Z_1, Z_2))/b_N$ with Jacobian b_N^L; thus, we obtain

$$\begin{aligned} &Cb_N^L \int_{\{t|t=(z-r_{b_N}(Z_1,Z_2))/b_N, z\in\mathbb{Z}\}} K^{(\mu)}(t)^2 f_Z\big(b_N t + r_{b_N}(Z_1, Z_2)\big)\,\mathrm{d}t \\ &\quad \le C_1 b_N^L \int_{\mathcal{U}} K^{(\mu)}(t)^2\,\mathrm{d}t \le C_2 b_N^L \end{aligned}$$

by Assumptions 4.1 and 5.1. We conclude that

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})^2\big] = O\big(b_N^{-L}\big). \tag{C.51}$$

The same proof and the same bound holds if $j \ne k \ne i$ or $j = k \ne i$.

Step 3B (Equation (C.48)).

Next, we consider $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})^2]$, where we note that $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})] \ne 0$. Because

$$\begin{aligned} &a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i}) \\ &\quad = \frac{1}{b_N^{L+|\mu|}}\Bigg[\nu(Z_i)'\tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_i)}{b_N}\right)\big(Z_i - r_{b_N}(Z_i)\big)^\mu \\ &\qquad - \mathbb{E}_Z\left(\nu(Z_i)'^{(\mu)}\left(\frac{Z - r_{b_N}(Z_i)}{b_N}\right)\big(Z_i - r_{b_N}(Z_i)\big)^\mu\right)\Bigg], \end{aligned}$$

we have

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})^2\big] \tag{C.52}$$

$$= \frac{1}{b_N^{2L+2|\mu|}}\mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_i)}{b_N}\right)^2 \big(\nu(Z_i)'\tilde{Y}_i\big)^2\big(Z_i - r_{b_N}(Z_i)\big)^{2\mu}\right] \tag{C.53}$$

$$\begin{aligned} &- \frac{2}{b_N^{2L+2|\mu|}}\mathbb{E}_{Z_i}\Bigg[\Bigg\{\nu(Z_i)'g(Z_i)K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_i)}{b_N}\right) \\ &- \mathbb{E}_Z\left[\nu(Z)g(Z)K^{(\mu)}\left(\frac{Z - r_{b_N}(Z_i)}{b_N}\right)\right]\big(Z_i - r_{b_N}(Z_i)\big)^{2\mu}\Bigg\}\Bigg] \end{aligned} \tag{C.54}$$

$$\begin{aligned} &+ \frac{1}{b_N^{2L+2|\mu|}}\mathbb{E}_{Z_i}\Bigg[\left(\nu(Z_i)'\mathbb{E}_Z\left(g(Z)K^{(\mu)}\left(\frac{Z - r_{b_N}(Z_i)}{b_N}\right)\right)\right)^2 \\ &\times \big(Z_i - r_{b_N}(Z_i)\big)^{2\mu}\Bigg]. \end{aligned} \tag{C.55}$$

By Assumption 4.1 and smoothness, (C.53) is bounded by

$$\frac{C}{b_N^{2L}}\int_{\mathbb{Z}} K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\,\mathrm{d}z = \frac{C}{b_N^{2L}}\int_{\mathbb{Z}^I_{b_N}} K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\,\mathrm{d}z + \frac{C}{b_N^{2L}}\int_{\mathbb{Z}\setminus\mathbb{Z}^I_{b_N}} K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\,\mathrm{d}z.$$

If $r_b(z)$ is the projection on the internal set, then $z-r_b(z)=0$ if z is in the internal set. Therefore,

$$\frac{C}{b_N^{2L}}\int_{\mathbb{Z}^I_{b_N}} K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\,\mathrm{d}z \leq \frac{CK^{(\mu)}(0)^2}{b_N^{2L}}.$$

Next we consider the second integral. If $s\in\mathbb{Z}^B_{b_N}\equiv\mathbb{Z}\setminus\mathbb{Z}^I_{b_N}$, then at least one component of z is in the boundary region. We can subdivide $\mathbb{Z}^B_{b_N}$ into disjoint subsets $\mathbb{Z}^B_{b_N,p}$, $p=1,\dots,2^L-1$, and in each such subset, $L_p\geq 1$ components of z are within b_N from the boundary. We further partition $\mathbb{Z}^B_{b_N,p}$ into disjoint sets $\mathbb{Z}^B_{b_N,p,r}$, $r=1,\dots,2^{L_p}$, with $0\leq K_r\leq L_p$ components with $z_{ll}\leq Z_l\leq z_{ll}+b_N$ and the remaining L_p-K_r components with $z_{ul}-b_N\leq Z_l\leq z_{ul}$. Without loss of generality, we assume that the first K_r components of z are near the lower bound, the next L_p-K_r are near the upper bound, and the rest are in the internal region, so that

$$\begin{aligned}
&\frac{C}{b_N^{2L}}\int_{\mathbb{Z}^B_{b_N,p,r}} K^{(\mu)}\left(\frac{z-r_{b_N}(z)}{b_N}\right)^2 f_Z(z)\,\mathrm{d}z \\
&= \int_{z_{l1}}^{z_{l1}+b_N}\cdots\int_{z_{l,K_r}}^{z_{l,K_r}+b_N}\int_{z_{u,K_r+1}-b_N}^{z_{u,K_r+1}}\cdots \\
&\quad \int_{z_{u,L_p}-b_N}^{z_{u,L_p}}\int_{z_{l,L_p+1}+b_N}^{z_{u,L_p+1}-b_N}\cdots\int_{z_{lL}+b_N}^{z_{uL}-b_N}\prod_{l=1}^{K_r}\mathcal{K}_l^{(\mu_l)}\left(\frac{Z_l-z_{ll}}{b_N}-1\right)^2 \\
&\quad\times\prod_{l=K_r+1}^{L_p}\mathcal{K}_l^{(\mu_l)}\left(\frac{Z_l-z_{ul}}{b_N}+1\right)^2\prod_{l=L_p+1}^{L}\mathcal{K}_l^{(\mu_l)}(0)^2 f_Z(z)\,\mathrm{d}z.
\end{aligned}$$

Because the support of the kernel is $[-1,1]$ and by Assumption 5.1, $\mathcal{K}_l^{(\mu_l)}$ is bounded on this support, we have

$$\begin{aligned}
&\mathcal{K}_l^{(\mu_l)}\left(\frac{Z_l-z_{ll}}{b_N}-1\right)\leq C\cdot 1(z_{ll}\leq z_l\leq z_{ll}+2b_N),\\
&\mathcal{K}_l^{(\mu_l)}\left(\frac{z_l-z_{ul}}{b_N}+1\right)\leq C\cdot 1(z_{ul}-2b_N\leq z_l\leq z_{ul})
\end{aligned}$$

and substitution gives the upper bound

$$\frac{C_1}{b_N^{2L}} \int_{z_{l1}}^{z_{l1}+b_N} \cdots \int_{z_{lK_r}}^{z_{lK_r}+b_N} \int_{z_{u,K_r+1}-b_N}^{z_{u,K_r+1}} \cdots \int_{z_{u,L_p}-b_N}^{z_{u,L_p}} \prod_{l=1}^{K_r} 1(z_{ll} \le z_l \le z_{ll} + 2b_N)$$
$$\times \prod_{l=K_r+1}^{L_p} 1(z_{ul} - 2b_N \le z_l \le z_{ul}) f_Z(z_1, \ldots z_{L_p})\, \mathrm{d}z_1 \cdots \mathrm{d}z_{L_p} \le \frac{C_2}{b_N^{2L-L_p}}.$$

Because $L_p \ge 1$, the integral over the boundary region is $O(b_N^{-2L+1})$. Combining the results, we have that

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})^2\big] = O\big(b_N^{-2L}\big), \tag{C.56}$$

which is larger than the bound in (C.51) and could be a reason to omit the terms $i = j = k$ (and redefine the kernel estimator).

STEP 3C (Equation (C.49)).

Third, we consider $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})^2]$. Again we have $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})] \ne 0$. We have

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})^2\big]$$
$$= \frac{1}{b_N^{2L+2|\mu|}} \mathbb{E}\bigg[K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1i}, Z_{2k})}{b_N}\bigg)^2$$
$$\times \nu(Z_{1i}, Z_{2k})' \tilde{Y}_i \tilde{Y}_i' \nu(Z_{1i}, Z_{2k}) \big(\big(Z_{1i}'\ Z_{2k}'\big)' - r_{b_N}(Z_{1i}, Z_{2k})\big)^{2\mu}\bigg] \tag{C.57}$$
$$- \frac{2}{b_N^{2L+2|\mu|}} \mathbb{E}_{Z_i Z_{2k}}\bigg[\nu(Z_{1i}, Z_{2k})' g(Z_{1i}, Z_{2k}) K^{(\mu)}\bigg(\frac{Z_i - r_{b_N}(Z_{1i}, Z_{2k})}{b_N}\bigg)$$
$$\times \mathbb{E}_Z\bigg(g(Z)'^{(\mu)}\bigg(\frac{Z - r_{b_N}(Z_{1i}, Z_{2k})}{b_N}\bigg)\bigg)\big(\big(Z_{1i}'\ Z_{2k}'\big)' - r_{b_N}(Z_{1i}, Z_{2k})\big)^{2\mu}\bigg] \tag{C.58}$$
$$+ \frac{1}{b_N^{2L+2|\mu|}} \mathbb{E}_{Z_i Z_{2k}}\bigg[\bigg(\nu(Z_{1i}, Z_{2k})' \mathbb{E}_Z\bigg(g(Z) K^{(\mu)}\bigg(\frac{Z - r_{b_N}(Z_{1i}, Z_{2k})}{b_N}\bigg)\bigg)\bigg)^2$$
$$\times \big(\big(Z_{1i}'\ Z_{2k}'\big)' - r_{b_N}(Z_{1i}, Z_{2k})\big)^{2\mu}\bigg]. \tag{C.59}$$

By Assumptions 4.1 and smoothness, (C.58) is bounded by

$$\frac{C}{b_N^{2L}} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}} K^{(\mu)}\bigg(\frac{z - r_{b_N}(z_1, \tilde{z}_2)}{b_N}\bigg)^2 f_Z(z) f_{Z_2}(\tilde{z}_2)\, \mathrm{d}z\, \mathrm{d}\tilde{z}_2$$
$$= \frac{C}{b_N^{2L}} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}_{b_N,1}^I} \int_{\mathbb{Z}_2} K^{(\mu)}\bigg(\frac{z - r_{b_N}(z_1, \tilde{z}_2)}{b_N}\bigg)^2 f_Z(z_1, z_2) f_{Z_2}(\tilde{z}_2)\, \mathrm{d}z_2\, \mathrm{d}z_1\, \mathrm{d}\tilde{z}_2$$
$$+ \frac{C}{b_N^{2L}} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1 \setminus \mathbb{Z}_{b_N,1}^I} \int_{\mathbb{Z}_2} K^{(\mu)}\bigg(\frac{z - r_{b_N}(z_1, \tilde{z}_2)}{b_N}\bigg)^2 f_Z(z_1, z_2) f_{Z_2}(\tilde{z}_2)\, \mathrm{d}z_2\, \mathrm{d}z_1\, \mathrm{d}\tilde{z}_2.$$

Because $z_1 - r_b(z_1, \tilde{z}_2) = 0$ if $z_1 \in \mathbb{Z}^I_{b_N,1}$, the first term on the right-hand side is equal to

$$\frac{CK_1^{(\mu_1)}(0)}{b_N^{2L}} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}^I_{b_N,1}} \int_{\mathbb{Z}_2} \mathcal{K}^{(\mu_2)}\left(\frac{z_2 - r_{b_N}(z_1, \tilde{z}_2)}{b_N}\right)^2 f_Z(z_1, z_2) f_{Z_2}(\tilde{z}_2)\,\mathrm{d}z_2\,\mathrm{d}z_1\,\mathrm{d}\tilde{z}_2$$
$$= O\big(b_N^{-2L+L_2}\big)$$

(where $\mathcal{K}(u)$ is the univariate kernel) by a change of variables to $t_2 = (z_2 - r_{b_N}(z_1, \tilde{z}_2))/b_N$ with Jacobian $b_N^{L_2}$. For the second integral, we partition $\mathbb{Z}^B_{1,b_N} \equiv \mathbb{Z}_1 \setminus \mathbb{Z}^I_{b_N,1}$ into sets $\mathbb{Z}^B_{1,b_N,p}$, $p = 1, \dots, 2^{L_1} - 1$, in which $1 \le L_{1p} \le L_1$ components of z_1 are in the boundary region. Each $\mathbb{Z}^B_{1,b_N,p}$ is partitioned further into sets $\mathbb{Z}^B_{1,b_N,p,r}$, $r = 1, \dots, 2^{L_{1p}}$, in which $0 \le K_{1r} \le L_{1r}$ components of z_1 are near the lower boundary, $L_{1r} - K_{1r}$ are near the upper boundary, and the remaining $L_1 - L_{1p}$ components are in the internal set. Hence, if we assume, without loss of generality, that the first K_r components of z_1 are near the lower boundary, the next $L_{1p} - K_{1r}$ are near the upper boundary, and the remaining components are in the internal set, then

$$\frac{C}{b_N^{2L}} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1 \setminus \mathbb{Z}^I_{b_N,1}} \int_{\mathbb{Z}_2} K^{(\mu)}\left(\frac{z - r_{b_N}(z_1, \tilde{z}_2)}{b_N}\right)^2 f_Z(z_1, z_2) f_{Z_2}(\tilde{z}_2)\,\mathrm{d}z_2\,\mathrm{d}z_1\,\mathrm{d}\tilde{z}_2$$
$$= \frac{C}{b_N^{2L}} \int_{\mathbb{Z}_2} \int_{z_{l,11}}^{z_{l,l1}+b_N} \cdots \int_{z_{l1,K_{1r}}}^{z_{l1,K_{1r}}+b_N} \int_{z_{u1,K_{1r}+1}-b_N}^{z_{u1,K_{1r}+1}} \cdots$$
$$\int_{z_{u1,L_{1p}}-b_N}^{z_{u1,L_{1p}}} \int_{z_{l1,L_{1p}+1}+b_N}^{z_{u1,L_{1p}+1}-b_N} \cdots \int_{z_{l1,L_1}+b_N}^{z_{u1,L_1}-b_N} \int_{\mathbb{Z}_2} \prod_{l=1}^{K_{1r}} \mathcal{K}_l^{(\mu_l)}\left(\frac{z_{1l} - z_{l1l}}{b_N} - 1\right)^2$$
$$\times \prod_{l=K_{1r}+1}^{L_{1p}} \mathcal{K}_l^{(\mu_l)}\left(\frac{z_{1l} - z_{u1l}}{b_N} + 1\right)^2 \prod_{l=L_{1p}+1}^{L_1} \mathcal{K}_l^{(\mu_l)}(0)^2$$
$$\times K_2^{(\mu_2)}\left(\frac{z_2 - r_{b_N}(z_1, \tilde{z}_2)}{b_N}\right)^2 f_Z(z_1, z_2) f_{Z_2}(\tilde{z}_2)\,\mathrm{d}z_2\,\mathrm{d}z_1\,\mathrm{d}\tilde{z}_2.$$

After a change of variables to $t_2 = (z_2 - r_{b_N}(z_1, \tilde{z}_2))/b_N$, with Jacobian $b_N^{L_2}$ we have by analogous argument as above that this term is $O(b_N^{-2L+L_2+L_{1p}})$. Because $L_{1p} \ge 1$, we have, by combining the results,

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})^2\big] = O\big(b_N^{-2L+L_2}\big). \tag{C.60}$$

Step 3D (Equation (C.50)).

An analogous argument gives

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2i})^2\big] = O\big(b_N^{-2L+L_1}\big). \tag{C.61}$$

This finishes the derivation of the bounds on the second moments of the kernel of the V-statistic.

Now we turn to the proofs of equalities (C.41)–(C.46).

STEP 3E (Equation (C.41)).

For the first term,

$$W_{\mu,1} - U_{\mu} = \frac{\sqrt{N}}{N(N-1)(N-2)} \times \sum_{i\neq j\neq k} \big(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i)\big) \tag{C.62}$$

so that

$$\mathbb{E}\big[(W_{\mu,1} - U_{\mu})^2\big] = \frac{N}{N^2(N-1)^2(N-2)^2} \sum_{i\neq j\neq k} \sum_{i'\neq j'\neq k'} \mathbb{E}\big[\big(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i)\big) \times \big(a_{N,\mu}(\tilde{Y}_{i'}, Z_{i'}, Z_{1j'}, Z_{2k'}) - c_{N,\mu}(\tilde{Y}_{i'}, Z_{i'})\big)\big].$$

This expression can be simplified using

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})\big] = 0, \tag{C.63}$$

$$\mathbb{E}\big[c_{N,\mu}(\tilde{Y}_i, Z_i)\big] = 0, \tag{C.64}$$

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})|\tilde{Y}_i, Z_i\big] = c_{N,\mu}(\tilde{Y}_i, Z_i), \tag{C.65}$$

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j'}, Z_{2k})|Z_{2k}\big] = 0, \tag{C.66}$$

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j}, Z_{2k'})|Z_{1j}\big] = 0, \tag{C.67}$$

$$\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k})a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j}, Z_{2k'})|Z_{1j}, Z_{2k}\big] = 0. \tag{C.68}$$

Therefore,

$$\mathbb{E}\big[\big(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i)\big) \times \big(a_{N,\mu}(V_{i'}, Z_{i'}, Z_{1j'}, Z_{2k'}) - c_{N,\mu}(V_{i'}, Z_{i'})\big)\big] = 0$$

if $i \neq i'$, $j \neq j'$, $k \neq k'$ by (C.63) and (C.64), if $i = i'$, $j \neq j'$, $k \neq k'$ by (C.65), if $i \neq i'$, $j \neq j'$, $k = k'$ by (C.66), and if $i \neq i'$, $j = j'$, $k \neq k'$ by (C.67), and if $i \neq i'$, $j = j'$, $k = k'$ by (C.68). Using this, we obtain

$$\begin{aligned} \mathbb{E}\big[(W_{\mu,1} - U_{\mu})^2\big] &= \frac{N}{N^2(N-1)^2(N-2)^2} \sum_{i\neq j\neq k} \mathbb{E}\big[\big(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i)\big)^2\big] \\ &\quad + \frac{N}{N^2(N-1)^2(N-2)^2} \sum_{i\neq j\neq k\neq k'} \mathbb{E}\big[\big(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k}) - c_{N,\mu}(\tilde{Y}_i, Z_i)\big) \\ &\quad \times \big(a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2k'}) - c_{N,\mu}(\tilde{Y}_i, Z_i)\big)\big] \end{aligned} \tag{C.69}$$

$$+\frac{N}{N^2(N-1)^2(N-2)^2}\sum_{i\neq k\neq j\neq j'}\mathbb{E}\big[\big(a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})-c_{N,\mu}(\tilde{Y}_i,Z_i)\big)$$
$$\times\big(a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j'},Z_{2k})-c_{N,\mu}(\tilde{Y}_i,Z_i)\big)\big].$$

Because $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})|\tilde{Y}_i,Z_i]=c_{N,\mu}(\tilde{Y}_i,Z_i)$, we have $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})c_{N,\mu}(\tilde{Y}_i,Z_i)]=\mathbb{E}[c_{N,\mu}(\tilde{Y}_i,Z_i)^2]$ so that by the bounds on the second moment of $a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})$, given in (C.47)–(C.50),

$$\begin{aligned}
&\mathbb{E}\big[\big(a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})-c_{N,\mu}(\tilde{Y}_i,Z_i)\big)^2\big]\\
&=\mathbb{E}\big[(a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})^2\big]-\mathbb{E}\big[c_{N,\mu}(\tilde{Y}_i,Z_i)^2\big]\\
&\leq\mathbb{E}\big[(a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})^2\big]\\
&=O\big(b_N^{-L}\big).
\end{aligned}$$

Furthermore (note that $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k'})]=\mathbb{E}[(\mathbb{E}_{Z_2}[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_2)])^2]\geq 0$),

$$\begin{aligned}
&\mathbb{E}\big[\big(a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})-c_{N,\mu}(\tilde{Y}_i,Z_i)\big)\\
&\quad\times\big(a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k'})-c_{N,\mu}(\tilde{Y}_i,Z_i)\big)\big]\\
&=\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k'})\big]-\mathbb{E}\big[c_{N,\mu}(\tilde{Y}_i,Z_i)^2\big]\\
&\leq\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k'})\big]
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k})a_{N,\mu}(\tilde{Y}_i,Z_i,Z_{1j},Z_{2k'})\big]\\
&=\frac{1}{b_N^{2|\mu|+2L}}\mathbb{E}\bigg[\nu(Z_{1j},Z_{2k})'\tilde{Y}_i\tilde{Y}_i'\nu(Z_{1j},Z_{2k'})\\
&\quad\times K^{(\mu)}\bigg(\frac{Z_i-r_{b_N}(Z_{1j},Z_{2k})}{b_N}\bigg)K^{(\mu)}\bigg(\frac{Z_i-r_{b_N}(Z_{1j},Z_{2k'})}{b_N}\bigg)\\
&\quad\times\big(\big(Z_{1j}'\ Z_{2k}'\big)'-r_{b_N}(Z_{1j},Z_{2k})\big)^{\mu}\big(\big(Z_{1j}'\ Z_{2k'}'\big)'-r_{b_N}(Z_{1j},Z_{2k'})\big)^{\mu}\bigg]\\
&\quad-\frac{1}{b_N^{2|\mu|+2L}}\mathbb{E}\bigg[\nu(Z_{1j},Z_{2k})'\mathbb{E}_{\tilde{Y}Z}\bigg[\tilde{Y}K^{(\mu)}\bigg(\frac{S-r_{b_N}(Z_{1j},Z_{2k})}{b_N}\bigg)\bigg]\\
&\quad\times\mathbb{E}_{\tilde{Y}Z}\bigg[\tilde{Y}V'^{(\mu)}\bigg(\frac{Z-r_{b_N}(Z_{1j},Z_{2k'})}{b_N}\bigg)\bigg]\nu(Z_{1j},Z_{2k'})\\
&\quad\times\big(\big(Z_{1j}'\ Z_{2k}'\big)'-r_{b_N}(Z_{1j},Z_{2k})\big)^{\mu}\big(\big(Z_{1j}'\ Z_{2k'}'\big)'-r_{b_N}(Z_{1j},Z_{2k'})\big)^{\mu}\bigg]\\
&\leq\frac{1}{b_N^{2|\mu|+2L}}\mathbb{E}\bigg[\nu(Z_{1j},Z_{2k})'\tilde{Y}_i\tilde{Y}_i'\nu(Z_{1j},Z_{2k'})
\end{aligned}$$

$$\times K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k'})}{b_N}\right)$$
$$\times \left((Z'_{1j}\ Z'_{2k})' - r_{b_N}(Z_{1j}, Z_{2k})\right)^{\mu}\left((Z'_{1j}\ Z'_{2k'})' - r_{b_N}(Z_{1j}, Z_{2k'})\right)^{\mu}\Bigg],$$

because both expectations are nonnegative. By Assumptions 4.1 and smoothness, this is bounded by

$$\frac{C_1}{b_N^{2L}}\mathbb{E}\left[K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k})}{b_N}\right)K^{(\mu)}\left(\frac{Z_i - r_{b_N}(Z_{1j}, Z_{2k'})}{b_N}\right)\right] \le C_2 b_N^{-L}$$

by a change of variables to $t = (Z_i - r_{b_N}(Z_{1j}, Z_{2k}))/b_N$ with Jacobian b_N^L and Assumption 5.1. By interchanging the roles of j and k, we obtain a bound of the same order for the third term on the right-hand side of (C.69).

Combining these results, we find

$$\mathbb{E}\left[(W_{\mu,1} - U_\mu)^2\right] = O\left(N^{-2}b_N^{-L}\right) + O\left(N^{-1}b_N^{-L}\right) = O\left(N^{-1}b_N^{-L}\right) \tag{C.70}$$

so that by the Markov inequality, the first term in the projection remainder (C.40) is $O_p(N^{-1/2}b_N^{-L/2})$.

STEP 3F (Equation (C.42)).

For the second term of the projection remainder (C.40), we have by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\left[c_{N,\mu}(\tilde{Y}_i, Z_i)^2\right] \\
&\quad\le \frac{1}{b_N^{2L+2|\mu|}}\mathbb{E}\Bigg[\Bigg(\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1} \nu(z_1, z_2)'\tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right) \\
&\qquad \times \left((z'_1\ z'_2)' - r_{b_N}(z_1, z_2)\right)^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\Bigg)^2\Bigg] \\
&\quad\le \frac{1}{b_N^{2L+2|\mu|}}\mathbb{E}\Bigg[|\tilde{Y}_i|^2\Bigg(\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1} \left|\nu(z_1, z_2)\right|\left|K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right)\right| \\
&\qquad \times \left|\left((z'_1\ z'_2)' - r_{b_N}(z_1, z_2)\right)^{\mu}\right| f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\Bigg)^2\Bigg] \\
&\quad\le \frac{C_1}{b_N^{2L}}\mathbb{E}_{Z_i}\Bigg[\Bigg(\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1} \left|K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right)\right| f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\Bigg)^2 \\
&\qquad \times \mathbb{E}\left(|\tilde{Y}_i|^2 \middle| Z_i\right)\Bigg] \\
&\quad\le \frac{C_2}{b_N^{2L}}\int_{\mathbb{Z}}\Bigg(\int_{\mathbb{Z}_2}\int_{\mathbb{Z}_1} \left|K^{(\mu)}\left(\frac{\tilde{z} - r_{b_N}(z_1, z_2)}{b_N}\right)\right| f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\Bigg)^2 f_Z(\tilde{z})\,\mathrm{d}\tilde{z}
\end{aligned}$$

by Assumptions 4.1 and smoothness. By a change of variables $t = (\tilde{z} - r_{b_N}(z_1, z_2))/b_N$ with Jacobian b_N^L, we conclude that

$$\mathbb{E}\big[c_{N,\mu}(\tilde{Y}_i, Z_i)^2\big] = O\big(b_N^{-L}\big). \tag{C.71}$$

Therefore,

$$\mathbb{E}\big[U_\mu^2\big] = \mathbb{E}\big[c_{N,\mu}(\tilde{Y}_i, Z_i)^2\big] = O\big(b_N^{-L}\big)$$

so that the second term of the projection remainder is $O_p(N^{-1}b_N^{-L/2})$.

Step 3G (Equations (C.43)–(C.46)).

The other terms of the projection remainder can be bounded using (C.47)–(C.50). For the third term (note $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})] \neq 0$), by (C.47)–(C.50),

$$\begin{aligned}
\mathbb{E}\big[|W_{\mu,2}|\big] &\le \frac{\sqrt{N}(N-1)}{N^2}\mathbb{E}\big[\big|a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})\big|\big] \\
&\le \frac{\sqrt{N}(N-1)}{N^2}\sqrt{\mathbb{E}\big[\big|a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2k})\big|^2\big]} \\
&= O\big(N^{-1/2}b_N^{-L+L_2/2}\big)
\end{aligned}$$

so that term is $O_p(N^{-1/2}b_N^{-L+L_2/2})$. In the same way, by (C.47)–(C.50), the fourth term of the remainder is $O_p(N^{-1/2}b_N^{-L+L_1/2})$. For the fifth term (note $a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2j}) = 0$),

$$\mathbb{E}\big[W_{\mu,4}^2\big] = \frac{N^2(N-1)}{N^6}\mathbb{E}\big[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1j}, Z_{2j})^2\big] = O\big(N^{-3}b_N^{-L}\big)$$

so that term is $O_p(N^{-3/2}b_N^{-L/2})$. Finally, the sixth term (note $\mathbb{E}[a_{N,\mu}(\tilde{Y}_i, Z_i, Z_{1i}, Z_{2i})] \neq 0$) is by a similar argument as for the third term and by (C.47)–(C.50), $O_p(N^{-1/2} \times b_N^{-L})$. This is the largest term in the projection remainder.

This finishes the proof of

$$W_\mu = U_\mu + O_p\big(N^{-1/2}b_N^{-L}\big). \tag{C.72}$$

Note again that the remainder is smaller if we redefine the kernel estimators. In that case, the sixth term of the projection remainder is 0.

Step 4 (Asymptotic Distribution).

The fourth step in the proof is the derivation of the asymptotically normal distribution of the projection U_μ. In particular, we show that U_0 is asymptotically normal and we obtain the variance of that distribution. We show that U_μ/b_N also converges to a normal distribution for $|\mu| \ge 1$ so that $U_\mu = O_p(b_N)$ if $|\mu| \ge 1$. Because W in (C.36) is a linear combination of the W_μ that are asymptotically equivalent to the U_μ if a rate condition is met, W is asymptotically equivalent to U_0 under that rate condition. Define

$$\psi_{N,\mu,i} \equiv \frac{1}{b_N^{L+|\mu|}} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1} \nu(z_1, z_2)' \tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right) \times \left(\left(z_1' \; z_2'\right)' - r_{b_N}(z_1, z_2)\right)^{\mu} f_{Z_1}(z_1) f_{Z_2}(z_2) \, \mathrm{d}z_1 \, \mathrm{d}z_2$$

so that

$$c_{N,\mu}(\tilde{Y}_i, Z_i) = \psi_{N,\mu,i} - \mathbb{E}[\psi_{N,\mu,i}].$$

We have

$$\psi_{N,0,i} \equiv \frac{1}{b_N^L} \int_{\mathbb{Z}_2} \int_{\mathbb{Z}_1} \nu(z_1, z_2)' \tilde{Y}_i K\left(\frac{Z_i - r_{b_N}(z_1, z_2)}{b_N}\right) f_{Z_1}(z_1) f_{Z_2}(z_2) \, \mathrm{d}z_1 \, \mathrm{d}z_2.$$

The integration region $\mathbb{Z}_1 \times \mathbb{Z}_2$ can be partitioned into a set where all components of z_1 and z_2 are in the internal region, $\mathbb{Z}^I_{1,b_N} \times \mathbb{Z}^I_{2,b_N}$, and its complement, $\mathbb{Z}_1 \times \mathbb{Z}_2 \setminus \mathbb{Z}^I_{1,b_N} \times \mathbb{Z}^I_{2,b_N}$. We define

$$\psi_{N,0,i,0} \equiv \frac{1}{b_N^L} \int_{\mathbb{Z}^I_{2,b_N}} \int_{\mathbb{Z}^I_{1,b_N}} \nu(z_1, z_2)' \tilde{Y}_i K_1\left(\frac{Z_{1i} - z_1}{b_N}\right) K_2\left(\frac{Z_{2i} - z_2}{b_N}\right) \times f_{Z_1}(z_1) f_{Z_2}(z_2) \, \mathrm{d}z_1 \, \mathrm{d}z_2$$

and

$$U_{0,0} \equiv \frac{1}{\sqrt{N}} \sum_{I=1}^{N} \left(\psi_{N,0,i,0} - \mathbb{E}[\psi_{N,0,i,0}]\right).$$

We apply the Liapounov central limit theorem for triangular arrays that requires

$$\frac{N^2 (\mathbb{E}[|\psi_{N,0,i,0} - \mathbb{E}[\psi_{N,0,i,0}]|^3])^2}{N^3 \operatorname{Var}(\psi_{N,0,i,0})} \to 0$$

and a sufficient condition is that $\mathbb{E}[|\psi_{N,0,i,0}|^m] < \infty$ for $m = 1, 2, 3$. By a change of variables to $t_1 = (Z_{1i} - z_1)/b_N$ and $t_2 = (Z_{2i} - z_2)/b_N$ with Jacobians $b_N^{L_1}$ and $b_N^{L_2}$, respectively,

$$\begin{aligned} &|\psi_{N,0,i,0}|^m \\ &= \left| \int_{\mathcal{U}_1} \int_{\mathcal{U}_2} \prod_{l=1}^{L_1} 1\left(1 + \frac{Z_{1li} - z_{u1l}}{b_N} \le t_{1l} \le -1 + \frac{Z_{1li} - z_{l1l}}{b_N}\right) \right. \\ &\quad \times \prod_{l=1}^{L_2} 1\left(1 + \frac{Z_{2li} - z_{u2l}}{b_N} \le t_{2l} \le -1 + \frac{Z_{2li} - z_{l2l}}{b_N}\right) \\ &\quad \times \nu(Z_{1i} - b_N t_1, Z_{2i} - b_N t_2)' \tilde{Y}_i K_1(t_1) K_2(t_2) \\ &\quad \left. \times f_{Z_1}(Z_{1i} - b_N t_1) f_{Z_2}(Z_{2i} - b_N t_2) \, \mathrm{d}t_1 \, \mathrm{d}t_2 \right|^m \end{aligned}$$

$$\le \Bigg(\int_{\mathcal{U}_1}\int_{\mathcal{U}_2}\prod_{l=1}^{L_1} 1\bigg(1+\frac{Z_{1li}-z_{u1l}}{b_N}\le t_{1l}\le -1+\frac{Z_{1li}-z_{l1l}}{b_N}\bigg)$$
$$\times \prod_{l=1}^{L_2} 1\bigg(1+\frac{Z_{2li}-z_{u2l}}{b_N}\le t_{2l}\le -1+\frac{Z_{2li}-z_{l2l}}{b_N}\bigg)$$
$$\times \big|\nu(Z_{1i}-b_N t_1, Z_{2i}-b_N t_2)\big|\cdot|\tilde{Y}_i|\cdot\big|K_1(t_1)\big|\cdot\big|K_2(t_2)\big|$$
$$\times f_{Z_1}(Z_{1i}-b_N t_1) f_{Z_2}(Z_{2i}-b_N t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\Bigg)^m$$

by the Cauchy–Schwarz inequality. Because $\max\{-1, 1+\frac{Z_{jli}-z_{ujl}}{b_N}\}\le t_{jl}\le \min\{1, -1+\frac{Z_{jli}-z_{ljl}}{b_N}\}$, $j=1,2$, if and only if $z_{ljl}+b_N\le Z_{jli}-b_N t_{jl}\le z_{ujl}-b_N$, by Assumptions 4.1 and 5.1, and smoothness, we obtain

$$|\psi_{N,0,i,0}|^m \le C|\tilde{Y}_i|^m \tag{C.73}$$

and $\mathbb{E}[|\tilde{Y}|^3]$ is finite by Assumption 4.1. Therefore, the condition of the Liapounov theorem holds.

The above expressions also show that for almost all Z_{1i}, Z_{2i},

$$\psi_{N,0,i,0}\to \nu(Z_{1i}, Z_{2i})'\tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}),$$

and by (C.73), $\mathbb{E}[\psi^m_{N,0,i,0}]$ converges to the corresponding expectation by dominated convergence. The conclusion is that $U_{0,0}$ has the same asymptotic distribution as

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\big\{\nu(Z_{1i}, Z_{2i})'\tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i})\\ &\quad-\mathbb{E}\big[\nu(Z_1, Z_2)'\tilde{Y} f_{Z_1}(Z_1) f_{Z_2}(Z_2)\big]\big\}.\end{aligned}\tag{C.74}$$

We still have to derive the stochastic order of

$$U_{0,1}\equiv \frac{1}{\sqrt{N}}\sum_{I=1}^{N}\big(\psi_{N,0,i,1}-\mathbb{E}[\psi_{N,0,i,1}]\big)$$

with the integration region in $\psi_{N,0,i,1}$, that is, $\mathbb{Z}_1\times\mathbb{Z}_2\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}$, such that at least one component of z_1 or z_2 is in the boundary region. We partition $\mathbb{Z}_1\times\mathbb{Z}_2\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}$ into subsets $\mathbb{Z}^B_{1,b_N,p_1}\times\mathbb{Z}^B_{1,b_N,p_2}$, $p_1=1,\dots,2^{L_1}$, $p_2=1,\dots,2^{L_2}$, $\min\{p_1,p_2\}\ge 1$, and in each such set, $0\le L_{1p_1}\le L_1$, $0\le L_{2p_2}\le L_2$, $\min\{L_{1p_1}L_{1p_1}\}\ge 1$ components of z_1 and z_2 are near the boundary. We take, without loss of generality, $\mathbb{Z}^B_{1,b_N,1}=\mathbb{Z}^I_{1,b_N}$ and $\mathbb{Z}^B_{2,b_N,1}=\mathbb{Z}^I_{2,b_N}$ so that we exclude the set with $p_1=p_2=1$, because in that set, all components are in the internal region. For $j=1,2$, each $\mathbb{Z}^B_{j,b_N,p_j}$ is partitioned further into sets $\mathbb{Z}^B_{j,b_N,p_j,r_j}$, $r_j=1,\dots,2^{L_{1p_j}}$, in which $0\le K_{jr_j}\le L_{jr_j}$ components of z_j are near the lower boundary, $L_{jr_j}-K_{1r_j}$ are near the upper boundary, and the remaining $L_j-L_{jp_j}$ components are in the internal set. Without loss of generality, we assume that the first K_{jr_j} components of z_j are near the lower boundary, the next $L_{jp_j} - K_{jr_j}$ are near the upper boundary, and the remaining components are in the internal set, $j = 1, 2$. Therefore,

$$
\begin{aligned}
|\psi_{N,0,i,1}|^m &= \Bigg|\frac{1}{b_N^L}\int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}} \nu(z_1,z_2)'\tilde{Y}_i K\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right) \\
&\quad \times f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\Bigg|^m \\
&\le \left(\frac{1}{b_N^L}\int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}^I_{1,b_N}\times\mathbb{Z}^I_{2,b_N}} \big|\nu(z_1,z_2)\big||\tilde{Y}_i|\left|K\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)\right|\right. \\
&\quad \left.\times f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\right)^m \\
&\le \sum_{p_1}\sum_{p_2}\sum_{r_1}\sum_{r_2}\left(\frac{1}{b_N^L}\int_{z_{l21}}^{z_{l21}+b_N}\cdots\int_{z_{l2K_{2r_2}}}^{z_{l2K_{2r_2}}+b_N}\int_{z_{u2,K_{2r_2}+1}-b_N}^{z_{u2,K_{2r_2}+1}}\cdots\right. \\
&\quad \int_{z_{u2,L_{2p_2}}-b_N}^{z_{u2,L_{2p_2}}}\int_{z_{l2,L_{2p_2}+1}+b_N}^{z_{u2,L_{2p_2}+1}-b_N}\cdots\int_{z_{l2,L_2}+b_N}^{z_{u2,L_2}-b_N}\int_{z_{l,l1}}^{z_{l,l1}+b_N}\cdots \\
&\quad \int_{z_{l,1K_{1r_1}}}^{z_{l,1K_{1r_1}}+b_N}\int_{z_{u1,K_{1r_1}+1}-b_N}^{z_{u1,K_{1r_1}+1}}\cdots\int_{z_{u1,L_{1p_1}}-b_N}^{z_{u1,L_{1p_1}}}\int_{z_{u1,L_{1p_1}+1}+b_N}^{z_{u1,L_{1p_1}+1}-b_N}\cdots \\
&\quad \int_{z_{l1,L_1}+b_N}^{z_{u1,L_1}-b_N}\big|\nu(z_1,z_2)\big||\tilde{Y}_i|\prod_{l=1}^{K_{1r_1}}\left|\mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l1l}}{b_N}-1\right)\right| \\
&\quad \times \prod_{l=K_{1r_1}+1}^{L_{1p_1}}\left|\mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l1l}}{b_N}+1\right)\right|\prod_{l=L_{1p_1}+1}^{L_1}\left|\mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{1l}}{b_N}\right)\right| \\
&\quad \times \prod_{l=1}^{K_{2r_1}}\left|\mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{l2l}}{b_N}-1\right)\right|\prod_{l=K_{2r_2}+1}^{L_{2p_2}}\left|\mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{u2l}}{b_N}+1\right)\right| \\
&\quad \left.\times \prod_{l=L_{2p_2}+1}^{L_2}\left|\mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{2l}}{b_N}\right)\right| f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2\right)^m.
\end{aligned}
\tag{C.75}
$$

By a change of variables to $t_{1l} = (Z_{1li} - z_{1l})/b_N$, $l = L_{1p_1} + 1, L_1$ and $t_{2l} = (Z_{2li} - z_{2l})/b_N$, $l = L_{2p_2} + 1, L_2$ with Jacobian $b_N^{L-L_{1p_1}-L_{2p_2}}$, we have

$$
\begin{aligned}
|\psi_{N,0,i,1}|^m &\le \sum_{p_1}\sum_{p_2}\sum_{r_1}\sum_{r_2}\left(\frac{1}{b_N^{m(L_{1p_1}+L_{2p_2})}}\int_{z_{l,21}}^{z_{l,21}+b_N}\cdots\right. \\
&\quad \int_{z_{l,2K_{2r_2}}}^{z_{l,2K_{2r_2}}+b_N}\int_{z_{u2,K_{2r_2}+1}-b_N}^{z_{u2,K_{2r_2}+1}}\cdots\int_{z_{u2,L_{2p_2}}-b_N}^{z_{u2,L_{2p_2}}}\int_{-1}^{1}\cdots\int_{-1}^{1}\int_{z_{l,l1}}^{z_{l,l1}+b_N}\cdots
\end{aligned}
$$

$$
\int_{z_{l,1K_{1r_1}}}^{z_{l,1K_{1r_1}}+b_N} \int_{z_{u1,K_{1r_1}+1}-b_N}^{z_{u1,K_{1r_1}+1}} \cdots \int_{z_{u1,L_{1p_1}}-b_N}^{z_{u1,L_{1p_1}}} \int_{-1}^{1} \cdots
$$

$$
\int_{-1}^{1} \prod_{j=1}^{2} \prod_{l=L_{jp_j}+1}^{L_j} 1\Big(\frac{Z_{jli}-z_{ujl}}{b_N} + 1 \le t_{jl} \le \frac{Z_{jli}-z_{ljl}}{b_N} - 1 \Big)
$$

$$
\times \big| \nu(z_{11}, \dots, z_{1L_{1p_1}}, Z_{1,L_{1p_1}+1,i} - b_N t_{1L_{1p_1}+1}, \dots, Z_{1,L_1,i} - b_N t_{1L_1},
$$

$$
z_{21}, \dots, z_{2L_{2p_2}}, Z_{2,L_{2p_2}+1,i} - b_N t_{2L_{2p_2}+1}, \dots, Z_{2,L_2,i} - b_N t_{2L_2}) \big| |\tilde{Y}_i|
$$

$$
\times \prod_{l=1}^{K_{1r_1}} \Big| \mathcal{K}_{1l}\Big(\frac{Z_{1li}-z_{l1l}}{b_N} - 1 \Big) \Big| \prod_{l=K_{1r_1}+1}^{L_{1p_1}} \Big| \mathcal{K}_{1l}\Big(\frac{Z_{1li}-z_{l1l}}{b_N} + 1 \Big) \Big|
$$

$$
\times \prod_{l=1}^{K_{2r_1}} \Big| \mathcal{K}_{2l}\Big(\frac{Z_{2li}-z_{l2l}}{b_N} - 1 \Big) \Big| \prod_{l=K_{2r_2}+1}^{L_{2p_2}} \Big| \mathcal{K}_{2l}\Big(\frac{Z_{2li}-z_{u2l}}{b_N} + 1 \Big) \Big|
$$

$$
\times \prod_{l=L_{1p_1}+1}^{L_1} \big| \mathcal{K}_{1l}(t_{1l}) \big| \prod_{l=L_{2p_2}+1}^{L_2} \big| \mathcal{K}_{2l}(t_{2l}) \big|
$$

$$
\times f_{Z_1}(z_{11}, \dots, z_{1L_{1p_1}}, Z_{1,L_{1p_1}+1,i} - b_N t_{1L_{1p_1}+1}, \dots, Z_{1,L_1,i} - b_N t_{1L_1})
$$

$$
\times f_{Z_2}(z_{21}, \dots, z_{2L_{2p_2}}, Z_{2,L_{2p_2}+1,i} - b_N t_{2L_{2p_2}+1}, \dots, Z_{2,L_2,i} - b_N t_{2L_2})
$$

$$
\mathrm{d}z_{11} \cdots \mathrm{d}z_{1L_{1p_1}} \, \mathrm{d}t_{1,L_{1p_1}+1} \cdots \mathrm{d}t_{1L_1}
$$

$$
\mathrm{d}z_{21} \cdots \mathrm{d}z_{2L_{2p_2}} \, \mathrm{d}t_{2,L_{2p_2}+1} \cdots \mathrm{d}t_{2L_2} \Big)^m .
$$

In this integral, the function ν takes only values in the support $\mathbb{Z}$, and this function and the kernel functions are bounded by smoothness and Assumption 5.1 so that

$$
\begin{aligned}
&|\psi_{N,0,i,1}|^m \\
&\le C |\tilde{Y}_i|^m \sum_{p_1} \sum_{p_2} \sum_{r_1} \sum_{r_2} \frac{1}{b_N^{m(L_{1p_1}+L_{2p_2})}} \\
&\quad \times \Bigg(\prod_{l=1}^{K_{1r_1}} \Big| \mathcal{K}_{1l}\Big(\frac{Z_{1li}-z_{l,11}}{b_N} - 1 \Big) \Big| \prod_{l=K_{1r_1}+1}^{L_{1p_1}} \Big| \mathcal{K}_{1l}\Big(\frac{Z_{1li}-z_{l1l}}{b_N} + 1 \Big) \Big| \\
&\quad \times \prod_{l=1}^{K_{2r_1}} \Big| \mathcal{K}_{2l}\Big(\frac{Z_{2li}-z_{l2l}}{b_N} - 1 \Big) \Big| \prod_{l=K_{2r_2}+1}^{L_{2p_2}} \Big| \mathcal{K}_{2l}\Big(\frac{Z_{2li}-z_{u2l}}{b_N} + 1 \Big) \Big| \\
&\quad \times \int_{z_{l,21}}^{z_{l,21}+b_N} \cdots \int_{z_{l,2K_{2r_2}}}^{z_{l,2K_{2r_2}}+b_N} \int_{z_{u2,K_{2r_2}+1}-b_N}^{z_{u2,K_{2r_2}+1}} \cdots \int_{z_{u2,L_{2p_2}}-b_N}^{z_{u2,L_{2p_2}}} \int_{-1}^{1} \cdots
\end{aligned}
$$

$$\int_{-1}^{1}\int_{z_{l,l1}}^{z_{l,l1}+b_N}\cdots\int_{z_{l,1K_{1r_1}}}^{z_{l,1K_{1r_1}}+b_N}\int_{z_{u1,K_{1r_1}+1}-b_N}^{z_{u1,K_{1r_1}+1}}\cdots\int_{z_{u1,L_{1p_1}}-b_N}^{z_{u1,L_{1p_1}}}\int_{-1}^{1}\cdots$$
$$\int_{-1}^{1} f_{Z_1}(z_{11},\dots,z_{1L_{1p_1}},Z_{1,L_{1p_1}+1,i}-b_N t_{1L_{1p_1}+1},\dots,Z_{1,L_1,i}-b_N t_{1L_1})$$
$$\times f_{Z_2}(z_{21},\dots,z_{2L_{2p_2}},Z_{2,L_{2p_2}+1,i}-b_N t_{2L_{2p_2}+1},\dots,Z_{2,L_2,i}-b_N t_{2L_2})$$
$$\left.\mathrm{d}z_{11}\cdots\mathrm{d}z_{1L_{1p_1}}\,\mathrm{d}t_{1,L_{1p_1}+1}\cdots\mathrm{d}t_{1L_1}\,\mathrm{d}z_{21}\cdots\mathrm{d}z_{2L_{2p_2}}\,\mathrm{d}t_{2,L_{2p_2}+1}\cdots\mathrm{d}t_{2L_2}\right)^m.$$

Because the density is bounded, the integral is bounded by $Cb_N^{L_{1p_1}+L_{2p_2}}$. Moreover, because the kernel has support $[-1,1]^L$ and is bounded on that support, we have that

$$\prod_{l=1}^{K_{1r_1}}\left|\mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l1l}}{b_N}-1\right)\right|\prod_{l=K_{1r_1}+1}^{L_{1p_1}}\left|\mathcal{K}_{1l}\left(\frac{Z_{1li}-z_{l1l}}{b_N}+1\right)\right|$$
$$\times\prod_{l=1}^{K_{2r_1}}\left|\mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{l2l}}{b_N}-1\right)\right|\prod_{l=K_{2r_2}+1}^{L_{2p_2}}\left|\mathcal{K}_{2l}\left(\frac{Z_{2li}-z_{u2l}}{b_N}+1\right)\right|$$
$$\le C\prod_{l=1}^{K_{1r_1}}1(z_{l1l}\le Z_{1li}\le z_{l1l}+2b_N)\prod_{l=K_{1r_1}+1}^{L_{1p_1}}1(z_{l1l}-2b_N\le Z_{1li}\le z_{l1l})$$
$$\times\prod_{l=1}^{K_{2r_1}}1(z_{l2l}\le Z_{2li}\le z_{l2l}+2b_N)\prod_{l=K_{2r_2}+1}^{L_{2p_2}}1(z_{u2l}-2b_N\le Z_{2li}\le z_{u2l}).$$

Therefore,

$$|\psi_{N,0,i,1}|^m\le C|\tilde{Y}_i|^m\sum_{p_1}\sum_{p_2}\sum_{r_1}\sum_{r_2}\prod_{l=1}^{K_{1r_1}}1(z_{l1l}\le Z_{1li}\le z_{l1l}+2b_N)$$
$$\times\prod_{l=K_{1r_1}+1}^{L_{1p_1}}1(z_{l1l}-2b_N\le Z_{1li}\le z_{u1l})$$
$$\times\prod_{l=1}^{K_{2r_1}}1(z_{l2l}\le Z_{2li}\le z_{l2l}+2b_N)\prod_{l=K_{2r_2}+1}^{L_{2p_2}}1(z_{u2l}-2b_N\le Z_{2li}\le z_{u2l}),$$

and because $\mathbb{E}[|\tilde{Y}|^3|Z=z]$ is bounded on $\mathbb{Z}$ and the density of Z is bounded, we have, because $L_{1p_1}+L_{2p_2}\ge 1$ for $m=1,2,3$,

$$\mathbb{E}\big[|\psi_{N,0,i,1}|^m\big]=O(b_N).$$

By the Liapounov central limit theorem, U_{01}/b_N converges in distribution and hence

$$U_{01} = O_p(b_N). \tag{C.76}$$

STEP 5 (Ignoring Higher Order Terms).

The final step is to show that U_μ is asymptotically negligible if $|\mu| \geq 1$. Note that if $|\mu| \geq 1$, then the integrand in $\psi_{N,\mu,i}$ is 0 if z_1 and z_2 are both in the internal region. Hence, we can take the integration region such that at least one component of either z_1 or z_2 is in the boundary region:

$$\begin{aligned}
|\psi_{N,\mu,i}|^m &= \left| \frac{1}{b_N^{L+|\mu|}} \int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}_{1,b_N}^I\times\mathbb{Z}_{2,b_N}^I} \nu(z_1,z_2)'\tilde{Y}_i K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)\right. \\
&\qquad \left. \times \left((z_1'\ z_2')' - r_{b_N}(z_1,z_2)\right)^\mu f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2 \right|^m \\
&\leq \left(\frac{1}{b_N^{L+|\mu|}} \int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}_{1,b_N}^I\times\mathbb{Z}_{2,b_N}^I} \left|\nu(z_1,z_2)\right|\left|\tilde{Y}_i\right| \left|K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)\right| \right. \\
&\qquad \left. \times \left|(z_1'\ z_2')' - r_{b_N}(z_1,z_2)\right|^{|\mu|} f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2 \right)^m \\
&\leq \left(\frac{1}{b_N^{L}} \int_{\mathbb{Z}_2\times\mathbb{Z}_1\setminus\mathbb{Z}_{1,b_N}^I\times\mathbb{Z}_{2,b_N}^I} \left|\nu(z_1,z_2)\right|\left|\tilde{Y}_i\right| \left|K^{(\mu)}\left(\frac{Z_i - r_{b_N}(z_1,z_2)}{b_N}\right)\right| \right. \\
&\qquad \left. \times f_{Z_1}(z_1) f_{Z_2}(z_2)\,\mathrm{d}z_1\,\mathrm{d}z_2 \right)^m .
\end{aligned}$$

We obtained a bound on the right-hand side in (C.75). Therefore, by the Liapounov central limit theorem, $\frac{U_\mu}{b_N}$ converges in distribution so that if $|\mu| \geq 1$, then

$$U_\mu = O_p(b_N). \tag{C.77}$$

By (C.33) (linearization), (C.34) (bias), (C.72) (projection), (C.76) (boundary remainder), and (C.77) (NIP remainder), we have that

$$\begin{aligned}
\sqrt{N}(\hat{\theta} - \theta) &= \frac{1}{N\sqrt{N}} \sum_{j=1}^N \sum_{k=1}^N \left(n\left(h_0(Z_{1j}, Z_{2k})\right) - \theta\right) \\
&\quad + \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{\partial n}{\partial h}\left(h_0(Z_i)\right)' \tilde{Y}_i f_{Z_1}(Z_{1i}) f_{Z_2}(Z_{2i}) \right. \\
&\quad \left. - \mathbb{E}_{\tilde{Y}Z}\left[\frac{\partial n}{\partial h}\left(h_0(S)\right)' \tilde{Y} f_{Z_1}(Z_1) f_{Z_2}(Z_2) \right] \right\} \\
&\quad + O_p\left(\sqrt{N}|\hat{h}_{\mathrm{nip},s} - h_0|^2\right) + O\left(\sqrt{N} b_N^p\right) + O_p\left(N^{-1} b_N^{-L}\right) + O_p(b_N).
\end{aligned} \tag{C.78}$$

The first term on the right-hand side is a V-statistic that is asymptotically equivalent to

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(\mathbb{E}[n\left(h_0(Z_{1i}, Z_2)\right) - \theta\right) + \mathbb{E}[n\left(h_0(Z_1, Z_{2i})\right) - \theta)\right\}.$$

□

Proof of Lemma A.24. Using Lemma A.14, the assumptions imply that

$$\sup_{w\in\mathbb{W},x\in\mathbb{X}}\left|\hat{g}(w,x) - g(w,x)\right| = O_p\left(\left(\frac{\ln(N)}{N\cdot b_N^2}\right)^{1/2} + b_N^s\right) = o_p\left(N^{-\eta}\right). \tag{C.79}$$

For $1/4 < \delta < 1/4s$, we can find an $\eta > 1/4$ such that this holds. Using the definitions preceding the statement of the lemma, we have, by adding and subtracting terms,

$$\hat{\beta}^{\text{cm}}(\rho,0) - \beta^{\text{cm}}(\rho,0) = \left(\hat{\beta}^{\text{cm}}(\rho,0) - \hat{\beta}_g^{\text{cm}}\right) \tag{C.80}$$

$$- \left(\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}}\right) \tag{C.81}$$

$$- \left(\hat{\beta}_X^{\text{cm}} - \overline{g}^{\text{cm}}\right) \tag{C.82}$$

$$+ \left(\hat{\beta}_g^{\text{cm}} - \overline{g}^{\text{cm}}\right) + \left(\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}}\right) + \left(\hat{\beta}_X^{\text{cm}} - \overline{g}^{\text{cm}}\right)$$

$$+ \left(\overline{g}^{\text{cm}} - \beta^{\text{cm}}(\rho,0)\right).$$

The result then follows if we can show that the sum of (C.80), (C.81), and (C.82) is $o_p(N^{-1/2})$. Define

$$\hat{\omega}^{\text{cm}}(w,x) = \frac{\phi_c(\Phi_c^{-1}(\hat{F}_W(w)), \Phi_c^{-1}(\hat{F}_X(x)); \rho)}{\phi_c(\Phi_c^{-1}(\hat{F}_W(w)))\phi_c(\Phi_c^{-1}(\hat{F}_X(x)))},$$

$$\hat{\omega}_W^{\text{cm}}(w,x) = \frac{\phi_c(\Phi_c^{-1}(\hat{F}_W(w)), \Phi_c^{-1}(F_X(x)); \rho)}{\phi_c(\Phi_c^{-1}(\hat{F}_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))},$$

and

$$\hat{\omega}_X^{\text{cm}}(w,x) = \frac{\phi_c(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(\hat{F}_X(x)); \rho)}{\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(\hat{F}_X(x)))}.$$

Then, using these definitions, we can write the sum of these three components as

$$\left(\hat{\beta}^{\text{cm}}(\rho,0) - \hat{\beta}_g^{\text{cm}}\right) - \left(\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}}\right) - \left(\hat{\beta}_X^{\text{cm}} - \overline{g}^{\text{cm}}\right)$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{g}(W_i, X_j)\left[\hat{\omega}^{\text{cm}}(W_i, X_j) - \omega^{\text{cm}}(W_i, X_j)\right]$$

$$- \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}g(W_i, X_j)\left[\hat{\omega}_W(W_i, X_j) - \omega^{\text{cm}}(W_i, X_j)\right]$$

$$- \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}g(W_i, X_j)\left[\hat{\omega}_X(W_i, X_j) - \omega^{\text{cm}}(W_i, X_j)\right]$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \hat{g}(W_i,X_j)\big[\hat{\omega}^{\mathrm{cm}}(W_i,X_j)-\omega^{\mathrm{cm}}(W_i,X_j)\big] \\
&\quad - \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\big[\hat{\omega}^{\mathrm{cm}}(W_i,X_j)-\omega^{\mathrm{cm}}(W_i,X_j)\big] \\
&\quad + \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\big[\hat{\omega}^{\mathrm{cm}}(W_i,X_j)-\hat{\omega}_W(W_i,X_j)\big] \\
&\quad - \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j)\big[\hat{\omega}_X(W_i,X_j)-\omega^{\mathrm{cm}}(W_i,X_j)\big] \\
&= \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \big[\hat{g}(W_i,X_j)-g(W_i,X_j)\big]\big[\hat{\omega}^{\mathrm{cm}}(W_i,X_j)-\omega^{\mathrm{cm}}(W_i,X_j)\big] \qquad \text{(C.83)} \\
&\quad + \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N g(W_i,X_j) \\
&\quad \times \big[\hat{\omega}^{\mathrm{cm}}(W_i,X_j)-\hat{\omega}_W(W_i,X_j)-\hat{\omega}_X(W_i,X_j)+\omega^{\mathrm{cm}}(W_i,X_j)\big]. \qquad \text{(C.84)}
\end{aligned}
$$

It remains to be shown that both (C.83) and (C.85) are $o_p(N^{-1/2})$.

Now define

$$
\begin{aligned}
&k(z_1,z_2) = \frac{\phi_c(\Phi_c^{-1}(z_1),\Phi_c^{-1}(z_2);\rho)}{\phi_c(\Phi_c^{-1}(z_1))\cdot\phi_c(\Phi_c^{-1}(z_2))} \quad \text{so that} \\
&\hat{\omega}^{\mathrm{cm}}(w,x) = k\big(\hat{F}_W(w),\hat{F}_X(x)\big). \qquad \text{(C.85)}
\end{aligned}
$$

By a second order Taylor expansion, we have

$$
\begin{aligned}
&\hat{\omega}^{\mathrm{cm}}(w,x)-\omega^{\mathrm{cm}}(w,x) \\
&= \frac{\partial k}{\partial z_1}\big(F_W(w),F_X(x)\big)\big(\hat{F}_W(w)-F_W(w)\big) \\
&\quad + \frac{\partial k}{\partial z_2}\big(F_W(w),F_X(x)\big)\big(\hat{F}_X(x)-F_X(x)\big) \\
&\quad + \frac{1}{2}\frac{\partial^2 k}{\partial z_1^2}\big(\overline{F}_W(w),\overline{F}_X(x)\big)\big(\hat{F}_W(w)-F_W(w)\big)^2 \\
&\quad + \frac{1}{2}\frac{\partial^2 k}{\partial z_2^2}\big(\overline{F}_W(w),\overline{F}_X(x)\big)\big(\hat{F}_X(x)-F_X(x)\big)^2 \\
&\quad + \frac{1}{2}\frac{\partial^2 k}{\partial z_1\,\partial z_2}\big(\overline{F}_W(w),\overline{F}_X(x)\big)\big(\hat{F}_W(w)-F_W(w)\big)\big(\hat{F}_X(x)-F_X(x)\big),
\end{aligned}
$$

with $\overline{F}_W(w)$ and $\overline{F}_X(x)$ intermediate values. By Lemma A.3, it follows that for any $0<\delta<1/2$, $\sup_x|\hat{F}_X(x)-F_X(x)| = o_p(N^{-\delta})$ and $\sup_w|\hat{F}_W(w)-F_W(w)| = o_p(N^{-\delta})$. In combination with the fact that $|\partial^2 k/\partial z_1^2|$, $|\partial^2 k/\partial z_2^2|$, and $|\partial^2 k/\partial z_1\,\partial z_2|$ are bounded, this implies that

$$\begin{aligned}
&\hat{\omega}^{\mathrm{cm}}(w,x)-\omega^{\mathrm{cm}}(w,x)\\
&\quad=\frac{\partial k}{\partial z_1}\big(F_W(w),F_X(x)\big)\big(\hat{F}_W(w)-F_W(w)\big)\\
&\qquad+\frac{\partial k}{\partial z_2}\big(F_W(w),F_X(x)\big)\big(\hat{F}_X(x)-F_X(x)\big)+o_p\big(N^{-1/2}\big).
\end{aligned}\tag{C.86}$$

The same argument implies that

$$\hat{\omega}_W(w,x)-\omega^{\mathrm{cm}}(w,x)=\frac{\partial k}{\partial z_1}\big(F_W(w),F_X(x)\big)\big(\hat{F}_W(w)-F_W(w)\big)+o_p\big(N^{-1/2}\big)$$

and

$$\hat{\omega}_X(w,x)-\omega^{\mathrm{cm}}(w,x)=\frac{\partial k}{\partial z_2}\big(F_W(w),F_X(x)\big)\big(\hat{F}_X(x)-F_X(x)\big)+o_p\big(N^{-1/2}\big).$$

Substituting in these results, it follows that (C.85) is $o_p(N^{-1/2})$.

Equation (C.86) also implies, by Lemma A.3, that

$$\hat{\omega}^{\mathrm{cm}}(w,x)-\omega^{\mathrm{cm}}(w,x)=o_p\big(N^{-1/4}\big).$$

In combination with (C.79), this implies that (C.83) is also $o_p(N^{-1/2})$. □

Proof of Lemma A.25. The proof of this lemma makes use of an application of Theorem A.3. Using the notation of that theorem, we have $Z_1=W$, $Z_2=X$, $\tilde{Y}=(Y,1)'$,

$$h(w,x)=\begin{pmatrix} g(w,x)\cdot f_{WX}(w,x)\\ f_{WX}(w,x)\end{pmatrix},$$

and

$$n\big(h(w,x)\big)=\frac{h_1(w,x)}{h_2(w,x)}\omega^{\mathrm{cm}}(w,x)=g(w,x)\cdot\omega^{\mathrm{cm}}(w,x).$$

In terms of this notation, we can write this in the form of Theorem A.3:

$$\hat{\beta}_g^{\mathrm{cm}}-\overline{g}^{\mathrm{cm}}=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}n\big(\hat{h}(W_i,X_j)\big)-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}n\big(h(W_i,X_j)\big).$$

We also have

$$\frac{\partial n}{\partial h}\big(h(w,x)\big)=\begin{pmatrix}\dfrac{1}{h_2(w,x)}\\ -\dfrac{h_1(w,x)}{h_2(w,x)^2}\end{pmatrix}\omega^{\mathrm{cm}}(w,x)=\begin{pmatrix}\dfrac{1}{f_{WX}(w,x)}\\ -\dfrac{g(w,x)}{f_{WX}(w,x)}\end{pmatrix}\omega^{\mathrm{cm}}(w,x)$$

and, hence,

$$\frac{\partial n}{\partial h}\big(h(w,x)\big)'\tilde{y}f_W(w)f_X(x)=\frac{f_W(w)f_X(x)}{f_{WX}(w,x)}\cdot\big(y-g(w,x)\big)\cdot\omega^{\mathrm{cm}}(w,x),$$

which is mean zero. Therefore, by the result of Theorem A.3, we have

$$\hat{\beta}_g^{\text{cm}} - \beta^{\text{cm}}(\rho, 0) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial n}{\partial h}\big(h(W_i, X_i)\big)' \tilde{Y}_i f_W(W_i) f_X(X_i)$$

$$- \mathbb{E}\left[\frac{\partial n}{\partial h}\big(h(W, X)\big)' \tilde{y} Y f_W(W) f_X(X)\right]$$

$$+ o_p\big(N^{-1/2}\big)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{f_W(W_i) f_X(X_i)}{f_{WX}(W_i, X_i)} \cdot \big(Y_i - g(W_i, X_i)\big) \cdot \omega^{\text{cm}}(W_i, X_i)$$

$$- \mathbb{E}\left[\frac{f_W(W) f_X(X)}{f_{WX}(W, X)} \cdot \big(Y - g(W, X)\big) \cdot \omega^{\text{cm}}(W, X)\right]$$

$$+ o_p\big(N^{-1/2}\big)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{f_W(W_i) f_X(X_i)}{f_{WX}(W_i, X_i)} \cdot \big(Y_i - g(W_i, X_i)\big) \cdot \omega^{\text{cm}}(W_i, X_i)$$

$$+ o_p\big(N^{-1/2}\big)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \psi_g^{\text{cm}}(Y_i, W_i, X_i) + o_p\big(N^{-1/2}\big) + o_p\big(N^{-1/2}\big).$$

□

Proof of Lemma A.26. Using the definition of $k(z_1, z_2)$ in (C.85) and the Taylor expansion in the proof of Lemma A.24, we have

$$\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i, X_j)$$

$$\times \frac{\partial k}{\partial z_1}\big(F_W(W_i), F_X(X_j)\big)\big(\hat{F}_W(W_i) - F_W(W_i)\big)$$

$$+ \frac{1}{2}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i, X_j)$$

$$\times \frac{\partial^2 k}{\partial z_1^2}\big(\overline{F}_W(W_i), \overline{F}_X(X_j)\big)\big(\hat{F}_W(W_i) - F_W(W_i)\big)^2.$$

By Lemma A.3, $\sup_w |\hat{F}_W(w) - F_W(w)| = o_p(N^{-\delta})$ for all $\delta < 1/2$, and using the fact that the second derivatives of $k(z_1, z_2)$ are bounded, this implies

$$\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i, X_j)\frac{\partial k}{\partial z_1}\big(F_W(W_i), F_X(X_i)\big)\big(\hat{F}_W(W_i) - F_W(W_i)\big)$$

$$+ o_p\big(N^{-1/2}\big).$$

Inspection of the definition of $e_W(w, x)$ shows that $e_W(w, x) = \frac{\partial k}{\partial s_1}(F_W(w), F_X(x))$ and, therefore,

$$\begin{aligned}\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} g(W_i, X_j)e_W(W_i, X_j)\big(\hat{F}_W(W_i) - F_W(W_i)\big) \\ &\quad + o_p\big(N^{-1/2}\big) \\ &= \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} g(W_i, X_j)e_W(W_i, X_j)\big(1(W_k \leq W_i) - F_W(W_i)\big) \\ &\quad + o_p\big(N^{-1/2}\big).\end{aligned}$$

This is, up to the $o_p(N^{-1/2})$ term, a third order V-statistic,

$$\hat{\beta}_W^{\text{cm}} - \overline{g}^{\text{cm}} = V + o_p\big(N^{-1/2}\big),$$

where

$$V = \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} \psi(W_i, X_i, W_j, X_j, W_k, X_k),$$

with

$$\psi(w_1, x_1, w_2, x_2, w_3, x_3) = g(w_1, x_2)e_W(w_1, x_2)\big(1(w_3 \leq w_1) - F_W(w_1)\big).$$

Define

$$\begin{aligned}\psi_1(w, x) &= \mathbb{E}\big[\psi(w, x, W_2, X_2, W_3, X_3)\big], \\ \psi_2(w, x) &= \mathbb{E}\big[\psi(W_1, X_1, w, x, W_3, X_3)\big], \\ \psi_3(w, x) &= \mathbb{E}\big[\psi(W_1, X_1, W_2, X_2, w, x)\big],\end{aligned}$$

and

$$\theta = \mathbb{E}\big[\psi(W_1, X_1, W_2, X_2, W_3, X_3)\big].$$

Using V-statistic theory, this V-statistic can be approximated as

$$\begin{aligned}V &= \frac{1}{N}\sum_{i=1}^{N}\big\{\big(\psi_1(W_i, X_i) - \theta\big) + \big(\psi_2(W_i, X_i) - \theta\big) + \big(\psi_3(W_i, X_i) - \theta\big)\big\} \\ &\quad + o_p\big(N^{-1/2}\big).\end{aligned}$$

Note that $\mathbb{E}[\psi(w_1, x_1, w_2, x_2, W, X)] = 0$. Hence, $\theta = 0$, $\psi_1(w, x) = 0$, and $\psi_2(w, x) = 0$. Thus,

$$\begin{aligned} V &= \frac{1}{N}\sum_{i=1}^{N}\psi_3(W_i, X_i) + o_p\big(N^{-1/2}\big) \\ &= \frac{1}{N}\sum_{k=1}^{N}\int\int g(s,t)e_W(s,t)\big(1(W_i \le s) - F_W(s)\big)f_W(s)f_X(t)\,\mathrm{d}s\,\mathrm{d}t + o_p\big(N^{-1/2}\big) \\ &= \frac{1}{N}\sum_{i=1}^{N}\psi_W^{\mathrm{cm}}(Y_i, W_i, X_i) + o_p\big(N^{-1/2}\big), \end{aligned}$$

as required. □

The proof of Lemma A.27 is entirely analogous to that of Lemma A.26 and, therefore, is omitted.

Proof of Lemma A.28. Define

$$\begin{aligned} &\psi(w, x) = g(w, x)\cdot\omega^{\mathrm{cm}}(w, x), \\ &\psi_1(w) = \mathbb{E}\big[\psi(w, X)\big] = \mathbb{E}\big[g(w, X)\cdot\omega^{\mathrm{cm}}(w, X)\big], \end{aligned}$$

and

$$\psi_2(x) = \mathbb{E}\big[\psi(W, x)\big] = \mathbb{E}\big[g(W, x)\cdot\omega^{\mathrm{cm}}(W, x)\big].$$

Then, by the V-statistic projection theorem, given as Theorem A.4 in Appendix A, it follows that

$$\begin{aligned} \overline{g}^{\mathrm{cm}} - \beta^{\mathrm{cm}}(\rho, 0) &= \frac{1}{N}\sum_{i=1}^{N}\big\{\big(\psi_1(W_i) - \beta^{\mathrm{cm}}(\rho, 0)\big) + \big(\psi_2(X_i) - \beta^{\mathrm{cm}}(\rho, 0)\big)\big\} \\ &\quad + o_p\big(N^{-1/2}\big) \\ &= \frac{1}{N}\sum_{i=1}^{N}\psi_0^{\mathrm{cm}}(Y_i, W_i, X_i) + o_p\big(N^{-1/2}\big). \end{aligned}$$

□

Proof of Theorem A.4. Define

$$\phi(z_1, z_2) = \big(\psi(z_1, z_2) + \psi(z_2, z_1)\big)/2.$$

Then

$$V = \sum_{i=1}^{N}\sum_{j=1}^{N}\phi(Z_i, Z_j)/N^2$$

is a V-statistic with a symmetric kernel. In the notation of Lehmann (1999),

$$\sigma_1^2 = \text{Cov}\big(\phi(Z_i, Z_j), \phi(Z_i, Z_k)\big)$$

for i, j, k distinct, which simplifies to $\sigma_1^2 = \sigma^2/4$. Therefore, by Theorems 6.1.2 (with $a = 2$) and 6.2.1 in Lehmann (1999), the result follows. □

Reference

Lehmann, E. (1999), *Elements of Large-Sample Theory*. Springer, New York. [12, 74]

Submitted December, 2009. Final version accepted May, 2013.