Submitted to Bernoulli

Supplementary material to "Boundary adaptive local polynomial conditional density estimators"

MATIAS D. CATTANEO[1,a], RAJITA CHANDAK[1,b], MICHAEL JANSSON[2,c] and XINWEI MA[3,d]

[1]*Department of Operations Research and Financial Engineering, Princeton University, Princeton NJ, United States,* [a]*cattaneo@princeton.edu,* [b]*rchandak@princeton.edu*
[2]*Department of Economics, UC Berkeley, Berkeley CA, United States; Center for Research in Econometric Analysis of Time Series, Denmark,* [c]*mjansson@econ.berkeley.edu*
[3]*Department of Economics, UC San Diego, La Jolla CA, United States,* [d]*x1ma@ucsd.edu*

This Supplementary Material contains general theoretical results encompassing those discussed in the main paper, includes proofs of those general results, and discusses additional methodological and technical results. A companion `R` package is available at https://nppackages.github.io/lpcde/.

Keywords: Conditional density estimation; confidence bands; local polynomial methods; specification testing; strong approximation; uniform inference

Contents

1. Setup

Let $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$ be continuously distributed random variables supported on $\mathcal{X} = [0,1]^d$ and $\mathcal{Y} = [0,1]$, respectively. We are interested in estimating the conditional distribution function and its derivatives:

$$\theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) = \frac{\partial^\mu}{\partial y^\mu} \frac{\partial^{\boldsymbol{\nu}}}{\partial \mathbf{x}^{\boldsymbol{\nu}}} F(y|\mathbf{x}),$$

where $\mu \in \mathbb{N}$, and $\boldsymbol{\nu} \in \mathbb{N}^d$ representing multi-indices. (In the main paper we only consider the estimation of conditional density and derivatives thereof, that is, we set $\boldsymbol{\nu} = 0$ and $\mu = \vartheta + 1 \geq 1$.)

To present our estimation strategy, we start from $\theta_{0,\boldsymbol{\nu}}$, the conditional distribution function and its derivatives with respect to the conditioning variable, and apply the local polynomial method:

$$\widehat{\frac{\partial^{\boldsymbol{\nu}}}{\partial \mathbf{x}^{\boldsymbol{\nu}}}} F(y|\mathbf{x}) = \mathbf{e}_{\boldsymbol{\nu}}^\mathsf{T} \hat{\boldsymbol{\gamma}}(y|\mathbf{x}), \qquad \hat{\boldsymbol{\gamma}}(y|\mathbf{x}) = \operatorname*{argmin}_{\mathbf{v} \in \mathbb{R}^{\mathfrak{q}_d+1}} \sum_{i=1}^n \Big(\mathbb{1}(y_i \leq y) - \mathbf{q}(\mathbf{x}_i - \mathbf{x})^\mathsf{T} \mathbf{v} \Big)^2 L_h(\mathbf{x}_i; \mathbf{x}),$$

where and $\mathbf{e}_{\boldsymbol{\nu}}^\mathsf{T}$ is a basis vector extracting the corresponding estimate. We can write the solution in closed form as

$$\widehat{\frac{\partial^{\boldsymbol{\nu}}}{\partial \mathbf{x}^{\boldsymbol{\nu}}}} F(y|\mathbf{x}) = \mathbf{e}_{\boldsymbol{\nu}}^\mathsf{T} \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \Big(\frac{1}{n} \sum_{i=1}^n \mathbb{1}(y_i \leq y) \mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big) L_h(\mathbf{x}_i; \mathbf{x}) \Big),$$

where

$$\hat{\mathbf{S}}_{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big) \mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^\mathsf{T} L_h(\mathbf{x}_i; \mathbf{x}).$$

To estimate $\theta_{\mu,\boldsymbol{\nu}}$, we further smooth via local polynomials along the y-direction:

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) = \mathbf{e}_\mu^\mathsf{T} \hat{\boldsymbol{\beta}}(y|\mathbf{x}), \qquad \hat{\boldsymbol{\beta}}(y|\mathbf{x}) = \operatorname*{argmin}_{\mathbf{u} \in \mathbb{R}^{\mathfrak{p}+1}} \sum_{i=1}^n \left(\hat{F}(y_i|\mathbf{x}) - \mathbf{p}(y_i - y)^\mathsf{T} \mathbf{u} \right)^2 K_h(y_i; y).$$

We can write the solution in closed-form as

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) = \mathbf{e}_\mu^\mathsf{T} \hat{\mathbf{S}}_y^{-1} \hat{\mathbf{R}}_{y,\mathbf{x}} \hat{\mathbf{S}}_{\mathbf{x}}^{-1} \mathbf{e}_{\boldsymbol{\nu}},$$

where

$$\hat{\mathbf{S}}_y = \frac{1}{n} \sum_{i=1}^n \mathbf{p}\Big(\frac{y_i - y}{h}\Big) \mathbf{p}\Big(\frac{y_i - y}{h}\Big)^\mathsf{T} K_h(y_i; y), \qquad \text{and}$$

$$\hat{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{n^2 h^{\mu+|\boldsymbol{\nu}|}} \sum_{j=1}^{n}\sum_{i=1}^{n} \mathbb{1}(y_i \le y_j)\mathbf{p}\Big(\frac{y_j - y}{h}\Big)K_h(y_j;y)\mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathsf{T}} L_h(\mathbf{x}_i;\mathbf{x}).$$

While in the above we considered local polynomial regressions along both the $\mathbf{x}$- and y-directions, it is also possible to employ a local smoothing technique. To be precise, let G be some function such that the following Lebesgue-Stieltjes integration is well-defined, then an alternative estimator can be constructed as

$$\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) = \mathbf{e}_\mu^{\mathsf{T}}\check{\boldsymbol{\beta}}(y|\mathbf{x}), \qquad \check{\boldsymbol{\beta}}(y|\mathbf{x}) = \operatorname*{argmin}_{\mathbf{v}\in\mathbb{R}^{\mathfrak{p}+1}} \int \left(\hat{F}(u|\mathbf{x}) - \mathbf{p}(u-y)^{\mathsf{T}}\mathbf{v}\right)^2 K_h(u;y)\mathrm{d}G(u),$$

which has the solution

$$\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) = \mathbf{e}_\mu^{\mathsf{T}}\mathbf{S}_y^{-1}\bar{\mathbf{R}}_{y,\mathbf{x}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}},$$

where

$$\mathbf{S}_y = \int_{\mathcal{Y}} \mathbf{p}\Big(\frac{u-y}{h}\Big)\mathbf{p}\Big(\frac{u-y}{h}\Big)^{\mathsf{T}} K_h(u;y)\mathrm{d}G(u), \quad \text{and}$$

$$\bar{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{nh^{\mu+|\boldsymbol{\nu}|}} \sum_{i=1}^{n} \left(\int_{\mathcal{Y}} \mathbb{1}(y_i \le u)\mathbf{p}\Big(\frac{u-y}{h}\Big)K_h(u;y)\mathrm{d}G(u)\right)\mathbf{q}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{h}\Big)^{\mathsf{T}} L_h(\mathbf{x}_i;\mathbf{x}).$$

1.1. Notation

Limits are taken with respect to the sample size tending to infinity and the bandwidth shrinking to zero (i.e., $n \to \infty$ and $h \to 0$). For two positive sequences, $a_n \precsim b_n$ implies that $\limsup_{n\to\infty} |a_n/b_n| < \infty$. Similarly, $a_n \precsim_{\mathbb{P}} b_n$ means $|a_n/b_n|$ is asymptotically bounded in probability. We also adopt the small-o and big-O notation: $a_n = O_{\mathbb{P}}(b_n)$ is just $a_n \precsim_{\mathbb{P}} b_n$, and $a_n = o_{\mathbb{P}}(b_n)$ means a_n/b_n converges to zero in probability. Constants that do not depend on the sample size or the bandwidth will be denoted by $\mathfrak{c}$, $\mathfrak{c}_1$, $\mathfrak{c}_2$, etc.

We introduce another notation, $O_{\mathtt{TC}}$, which not only provides an asymptotic order but also controls the tail probability. To be specific, $a_n = O_{\mathtt{TC}}(b_n)$ if for any $\mathfrak{c}_1 > 0$, there exists some $\mathfrak{c}_2$ such that

$$\limsup_{n\to\infty} n^{\mathfrak{c}_1}\mathbb{P}\left[a_n \ge \mathfrak{c}_2 b_n\right] < \infty.$$

Here the subscript, $\mathtt{TC}$, stands for "tail control." Finally, let $\mathbf{X} = (\mathbf{x}_1^{\mathsf{T}}, \cdots, \mathbf{x}_n^{\mathsf{T}})^{\mathsf{T}}$ and $\mathbf{Y} = (y_1, \cdots, y_n)^{\mathsf{T}}$ be the data matrices.

- $F(y|\mathbf{x})$ and $f(y|\mathbf{x})$: the conditional distribution and density functions of y_i (at y) given $\mathbf{x}_i = \mathbf{x}$. The marginal distributions and densities are denoted by F_y, $F_{\mathbf{x}}$, f_y, and $f_{\mathbf{x}}$, respectively.
- y and $\mathbf{x}$: the evaluation points.
- $\mathcal{X} = [0,1]^d$ and $\mathcal{Y} = [0,1]$, the support of $\mathbf{x}_i$ and y_i, respectively.
- h: the bandwidth sequence.
- K: the kernel function, and L is the product kernel: $L(\mathbf{x}) = K(x_1)K(x_2)\cdots K(x_d)$.
- $\mathbf{p}$, $\mathbf{q}$: polynomial expansions.
- $\mathbf{P}$ and $\mathbf{Q}$: defined as $\mathbf{p}(\cdot)K(\cdot)$ and $\mathbf{q}(\cdot)L(\cdot)$, respectively.
- $\mathbf{e}_\mu$ and $\mathbf{e}_{\boldsymbol{\nu}}$: standard basis vectors extracting the μ-th and $\boldsymbol{\nu}$-th element in the expansion of $\mathbf{p}$ and $\mathbf{q}$ for univariate and multivariate arguments, respectively.

- $G(\cdot)$ the weighting function used in $\check{\theta}_{\mu,\nu}$, with its Lebesgue density denoted by $g(\cdot)$.
- Some matrices

$$\mathbf{S}_y = \int_{\frac{\mathcal{Y}-y}{h}} \mathbf{p}(u)\,\mathbf{P}(u)^{\mathsf{T}}\, g(y+hu)\mathrm{d}u, \qquad \hat{\mathbf{S}}_y = \frac{1}{nh}\sum_{i=1}^{n} \mathbf{p}\Big(\frac{y_i-y}{h}\Big)\mathbf{P}\Big(\frac{y_i-y}{h}\Big)^{\mathsf{T}},$$

$$\mathbf{c}_{y,\ell} = \int_{\frac{\mathcal{Y}-y}{h}} \frac{u^{\ell}}{\ell!}\mathbf{P}(u)\, g(y+hu)\mathrm{d}u, \qquad \hat{\mathbf{c}}_{y,\ell} = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\ell!}\left(\frac{y_i-y}{h}\right)^{\ell} \mathbf{P}\left(\frac{y_i-y}{h}\right)^{\mathsf{T}},$$

$$\mathbf{S}_{\mathbf{x}} = \int_{\frac{\mathcal{X}-\mathbf{x}}{h}} \mathbf{q}(\mathbf{v})\,\mathbf{Q}(\mathbf{v})^{\mathsf{T}}\, f_{\mathbf{x}}(\mathbf{x}+h\mathbf{v})\mathrm{d}\mathbf{v}, \qquad \hat{\mathbf{S}}_{\mathbf{x}} = \frac{1}{nh^d}\sum_{i=1}^{n} \mathbf{q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}},$$

$$\mathbf{c}_{\mathbf{x},\mathbf{m}} = \int_{\frac{\mathcal{X}-\mathbf{x}}{h}} \frac{\mathbf{v}^{\mathbf{m}}}{\mathbf{m}!}\mathbf{Q}(\mathbf{v})\, f_{\mathbf{x}}(\mathbf{x}+h\mathbf{v})\mathrm{d}\mathbf{v}, \qquad \hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}} = \frac{1}{nh^d}\sum_{i=1}^{n} \frac{1}{\mathbf{m}!}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathbf{m}} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right),$$

$$\mathbf{T}_{\mathbf{x}} = \int_{\frac{\mathcal{X}-\mathbf{x}}{h}} \mathbf{Q}(\mathbf{v})\,\mathbf{Q}(\mathbf{v})^{\mathsf{T}}\, f_{\mathbf{x}}(\mathbf{x}+h\mathbf{v})\mathrm{d}\mathbf{v}, \qquad \hat{\mathbf{T}}_{\mathbf{x}} = \frac{1}{nh^d}\sum_{i=1}^{n} \mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathsf{T}},$$

$$\mathbf{T}_y = \iint_{\frac{\mathcal{Y}-y}{h}} \min(u_1,u_2)\mathbf{P}(u_1)\,\mathbf{P}(u_2)^{\mathsf{T}}\, g(y+hu_1)g(y+hu_2)\mathrm{d}u_1\mathrm{d}u_2,$$

$$\widehat{\mathbf{T}}_y = \frac{1}{n^2h^3}\sum_{i,j=1}^{n} \big(\min(y_i,y_j)-y\big)\mathbf{P}\Big(\frac{y_i-y}{h}\Big)\mathbf{P}\Big(\frac{y_j-y}{h}\Big)^{\mathsf{T}},$$

$$\hat{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{n^2h^{1+d+\mu+|\nu|}}\sum_{j=1}^{n}\sum_{i=1}^{n} \mathbb{1}(y_i\le y_j)\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}},$$

$$\bar{\mathbf{R}}_{y,\mathbf{x}} = \frac{1}{nh^{1+d+\mu+|\nu|}}\sum_{i=1}^{n} \left(\int_{\mathcal{Y}} \mathbb{1}(y_i\le u)\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\right)\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}}.$$

- Equivalent kernels:

$$\check{\mathscr{K}}^{\circ}_{\mu,\nu,h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\nu|}}\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_y^{-1}\left[\int_{\mathcal{Y}}\left(\mathbb{1}(a\le u)-\hat{F}(u|\mathbf{b})\right)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\nu},$$

$$\hat{\mathscr{K}}^{\circ}_{\mu,\nu,h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\nu|}}\mathbf{e}_{\mu}^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{1}(a\le y_j)-\hat{F}(y_j|\mathbf{b})\right)\frac{1}{h}\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\nu},$$

$$\mathscr{K}^{\circ}_{\mu,\nu,h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\nu|}}\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_y^{-1}\left[\int_{\mathcal{Y}}\left(\mathbb{1}(a\le u)-F(u|\mathbf{b})\right)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\nu},$$

$$\mathscr{K}_{\mu,\nu,h}(a,\mathbf{b};y,\mathbf{x}) = \frac{1}{h^{\mu+|\nu|}}\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_y^{-1}\left[\int_{\mathcal{Y}}\mathbb{1}(a\le u)\frac{1}{h}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{b}-\mathbf{x}}{h}\right)^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\nu}.$$

- Some rates:

$$\mathtt{r}_{\mathtt{B}} = h^{\mathfrak{q}+1-|\nu|} + h^{\mathfrak{p}+1-\mu}, \quad \mathtt{r}_{\mathtt{V}} = \sqrt{\frac{1}{nh^{d+2|\nu|+2\mu-1}}},$$

$$\mathtt{r}_{\mathtt{BE}} = \begin{cases} \dfrac{1}{\sqrt{nh^d}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \neq 0 \text{ or } 1 \\ \dfrac{1}{\sqrt{nh^{d+1}}} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases},$$

$$\mathtt{r}_{\mathtt{VE}} = h^{\mathfrak{q}+\frac{1}{2}} + \sqrt{\frac{\log(n)}{nh^{d+1}}}, \quad \mathtt{r}_{\mathtt{SE}} = \sqrt{\log(n)}\mathtt{r}_{\mathtt{VE}}, \quad \mathtt{r}_{\mathtt{SA}} = \left(\frac{\log^{d+1}(n)}{nh^{d+1}}\right)^{\frac{1}{2d+2}}.$$

1.2. Overview

In this subsection we provide an overview of the main results. Underlying assumptions and precise statements of the lemmas and theorems will be given in later sections. First consider $\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$, with a conditional expectation decomposition:

$$\begin{aligned}\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) &= h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\left[\frac{1}{n}\sum_{i=1}^n\left(\int_{\mathcal{Y}} F(u|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\right)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}} \\ &+ h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\left[\frac{1}{n}\sum_{i=1}^n\left(\int_{\mathcal{Y}}\Big(\mathbb{1}(y_i \leq u) - F(u|\mathbf{x}_i)\Big)\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\right)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}.\end{aligned}$$

As we will show in Section 2, the first term above consists of the centering of the estimator (i.e., the parameter of interest $\theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$) and the smoothing bias. The second term, on the other hand, gives the asymptotic representation of the estimator. To be precise, we have

$$\begin{aligned}\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \mathscr{K}^\circ_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y,\mathbf{x}\right) \\ &\quad + O_{\mathbb{P}}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu} + \sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\frac{\log(n)}{\sqrt{nh^d}}\right).\end{aligned}$$

As a result, we can focus on establishing properties of the the first term, which provides an equivalent kernel expression. Denote its variance by $\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$. Then we show that the standardized process,

$$\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{n\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}\sum_{i=1}^n \mathscr{K}^\circ_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y,\mathbf{x}\right),$$

is approximately normally distributed both pointwise and uniformly for $y \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$. To be even more precise, we establish a strong approximation result, meaning that there exists a copy $\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, and a Gaussian process $\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ with the same covariance structure, such that

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\right| = O_{\mathbb{P}}\left(\frac{\log^{d+1}(n)}{nh^{d+1}}\right)^{\frac{1}{2d+2}}.$$

Together with a feasible variance-covariance estimator, the strong approximation result not only allows us to construct confidence bands for the target parameter and test shape restrictions, but also provides an explicit characterization of the coverage error probability for those procedures.

Inside the remainder term, $h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu}$ is the order of the leading smoothing bias, and $\log(n)\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})/(nh^d)}$ arises from the linearization step which replaces the random matrix $\hat{\mathbf{S}}_{\mathbf{x}}$ by its large-sample analogue $\mathbf{S}_{\mathbf{x}}$. It is worth mentioning that the order of the remainder term is uniformly valid for $y \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$, which is why an extra logarithmic factor is present.

Now consider the other estimator, $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$. While it is not possible to take a conditional expectation, we can still "center" the estimator with the conditional distribution function. That is,

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) = h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_\mu^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\left[\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n F(y_j|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}}\right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$$

$$+ h^{-\mu-|\boldsymbol{\nu}|}\mathbf{e}_\mu^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\left[\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \Big(\mathbb{1}(y_i\le y_j) - F(y_j|\mathbf{x}_i)\Big)\frac{1}{h}\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}}\right]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}.$$

As before, the first term captures the target parameter and the smoothing bias. The analysis of the second term is more involved. Besides the asymptotic linear representation term, it also consists of a leave-in bias term (since the same observation is used twice) and a second order U-statistic. We show that the following expansion holds uniformly for $y \in \mathcal{Y}$ and $\mathbf{x} \in \mathcal{X}$:

$$\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \mathscr{K}^\circ_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})$$

$$+ O_{\mathbb{P}}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu} + \sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\frac{\log(n)}{\sqrt{nh^d}} + \frac{\log(n)}{\sqrt{n^2h^{d+2\mu+2|\boldsymbol{\nu}|+1}}}\right).$$

Here, the contribution of the U-statistic is represented by the order $\log(n)/\sqrt{n^2h^{d+2\mu+2|\boldsymbol{\nu}|+1}}$ in the remainder term. Interestingly, this term is negligible compared to the standard error, $\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}$, provided that $\log(n)/(nh^2) \to \infty$.

The above demonstrates that important large-sample properties of the local regression based estimator, $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$ — such as pointwise and uniform normal approximation — stem from the equivalent kernel representation. Here we note that the representation holds by setting $G = F_y$. In other words, $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$ is first-order asymptotically equivalent to $\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$ with the (infeasible) local smoothing using the marginal distribution F_y.

1.3. Assumptions

We make the following assumptions on the joint distribution, the kernel function, and the weighting G.

Assumption S-DGP (Data generating process). (i) $\{y_i,\mathbf{x}_i\}_{1\le i\le n}$ is a random sample from the absolutely continuous joint distribution F supported on $\mathcal{Y}\times\mathcal{X} = [0,1]^{1+d}$. (ii) The joint density, f, is continuous and is bounded away from zero. (iii) $\theta_{2,\mathbf{0}}$ exists and is continuous.

Assumption S-K (Kernel).
The kernel function K is nonnegative, symmetric, supported on $[-1,1]$, Lipschitz continuous, and integrates to one.

Assumption S-W (Weighting function).
The weighting function G is continuously differentiable with a Lebesgue density denoted by g.

2. Pointwise large-sample properties

We first present several uniform convergence results which will be used later to establish pointwise and uniform properties of our estimators.

Lemma 2.1 (Matrix convergence). *Let Assumptions S-DGP, S-K, and S-W hold with $h \to 0$, $nh^d/\log(n) \to \infty$, and $G = F_y$. Then*

$$\sup_{y\in\mathcal{Y}}\left|\hat{\mathbf{S}}_y - \mathbf{S}_y\right| = O_{\mathtt{TC}}\left(\sqrt{\frac{\log(n)}{nh}}\right), \qquad \sup_{y\in\mathcal{Y}}\left|\hat{\mathbf{c}}_{y,\ell} - \mathbf{c}_{y,\ell}\right| = O_{\mathtt{TC}}\left(\sqrt{\frac{\log(n)}{nh}}\right),$$

$$\sup_{\mathbf{x}\in\mathcal{X}}\left|\hat{\mathbf{S}}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}}\right| = O_{\mathtt{TC}}\left(\sqrt{\frac{\log(n)}{nh^d}}\right), \qquad \sup_{\mathbf{x}\in\mathcal{X}}\left|\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}} - \mathbf{c}_{\mathbf{x},\mathbf{m}}\right| = O_{\mathtt{TC}}\left(\sqrt{\frac{\log(n)}{nh^d}}\right),$$

$$\sup_{\mathbf{x}\in\mathcal{X}}\left|\hat{\mathbf{T}}_{\mathbf{x}} - \mathbf{T}_{\mathbf{x}}\right| = O_{\mathtt{TC}}\left(\sqrt{\frac{\log(n)}{nh^d}}\right).$$

If in addition that $nh^{d+1}/\log(n) \to \infty$, then

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\left(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}\right]\right)\right| = O_{\mathtt{TC}}\left(\mathtt{r}_1\right), \quad \text{where } \mathtt{r}_1 = \begin{cases}\sqrt{\dfrac{\log(n)}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}} & \text{if } \mu = 0 \\ \sqrt{\dfrac{\log(n)}{nh^{d+2\mu+2|\boldsymbol{\nu}|-1}}} & \text{if } \mu > 0\end{cases}.$$

We now follow the decomposition in Section 1.2 and study the leading bias of our estimators.

Lemma 2.2 (Bias). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$ and $nh^d/\log(n) \to \infty$. In addition, $\theta_{\mu',\boldsymbol{\nu}'}$ exists and is continuous for all $\mu' + |\boldsymbol{\nu}'| = \max\{\mathfrak{q}+1+\mu,\ \mathfrak{p}+1+|\boldsymbol{\nu}|\}$. Then*

$$\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\Big[\frac{1}{nh^{\mu+|\boldsymbol{\nu}|}}\sum_{i=1}^n\Big(\int_{\mathcal{Y}} F(u|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\Big]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$$
$$= \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) + \mathsf{B}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + o_\mathbb{P}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu}\right),$$

where

$$\mathsf{B}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \underbrace{h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y|\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^\mathsf{T}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}}_{B_{(i),\mathfrak{q}+1}(y,\mathbf{x})} + \underbrace{h^{\mathfrak{p}+1-\mu}\theta_{\mathfrak{p}+1,\boldsymbol{\nu}}(y|\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+1}^\mathsf{T}\mathbf{S}_y^{-1}\mathbf{e}_\mu}_{B_{(ii),\mathfrak{p}+1}(y,\mathbf{x})}.$$

Similarly,

$$\mathbf{e}_\mu^\mathsf{T}\hat{\mathbf{S}}_y^{-1}\Big[\frac{1}{n^2h^{\mu+|\boldsymbol{\nu}|}}\sum_{i=1}^n\sum_{j=1}^n F(y_j|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\Big]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$$
$$= \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) + \mathsf{B}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) + o_\mathbb{P}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu}\right).$$

For future reference, we define the order of the leading bias as

$$\mathtt{r}_{\mathtt{B}} = h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu}.$$

Remark 2.1 (Higher-order bias). Because the leading bias established in the lemma can be exactly zero, one may need to extract higher-order terms for bandwidth selection:

$$\begin{aligned}\mathsf{B}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) &= h^{\mathfrak{q}+1-|\boldsymbol{\nu}|}B_{(i),\mathfrak{q}+1}(y,\mathbf{x}) + h^{\mathfrak{p}+1-\mu}B_{(ii),\mathfrak{p}+1}(y,\mathbf{x}) \\ &\quad + h^{\mathfrak{q}+2-|\boldsymbol{\nu}|}B_{(i),\mathfrak{q}+2}(y,\mathbf{x}) + h^{\mathfrak{p}+2-\mu}B_{(ii),\mathfrak{p}+2}(y,\mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\boldsymbol{\nu}|}B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y,\mathbf{x}),\end{aligned}$$

where

$$B_{(i),\mathfrak{q}+2}(y,\mathbf{x}) = \sum_{|\mathbf{m}|=\mathfrak{q}+2}\theta_{\mu,\mathbf{m}}(y|\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}, \qquad B_{(ii),\mathfrak{p}+2}(y,\mathbf{x}) = \theta_{\mathfrak{p}+2,\boldsymbol{\nu}}(y|\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+2}^{\mathsf{T}}\mathbf{S}_{y}^{-1}\mathbf{e}_{\mu},$$

$$B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y,\mathbf{x}) = \mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_{y}^{-1}\mathbf{c}_{y,\mathfrak{p}+1}\left(\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mathfrak{p}+1,\mathbf{m}}(y|\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathsf{T}}\right)\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}.$$

Note that the last term, $h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\boldsymbol{\nu}|}B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y,\mathbf{x})$, is present only if $\mu=\mathfrak{p}$ and $|\boldsymbol{\nu}|=\mathfrak{q}$.

Next we study the leading variance of our estimator, defined as

$$\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})\right].$$

Lemma 2.3 (Variance). *Let Assumptions S-DGP, S-K and S-W hold with $h\to 0$ and $nh^d/\log(n)\to\infty$. Then*
(i) $\mu=0$ and $\theta_{0,\mathbf{0}}\neq 0$ or 1:

$$\mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}\theta_{0,\mathbf{0}}(y|\mathbf{x})(1-\theta_{0,\mathbf{0}}(y|\mathbf{x}))\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{T}_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\right) + O\left(\frac{1}{nh^{d+2|\boldsymbol{\nu}|-1}}\right).$$

(ii) $\mu=0$ and $\theta_{0,\mathbf{0}}=0$ or 1: $\mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})$ has the order $\frac{1}{nh^{d+2|\boldsymbol{\nu}|-1}}$.
(iii) $\mu>0$:

$$\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}}\theta_{1,0}(y|\mathbf{x})\left(\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_{y}^{-1}\mathbf{T}_{y}\mathbf{S}_{y}^{-1}\mathbf{e}_{\mu}\right)\left(\mathbf{e}_{\boldsymbol{\nu}}^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{T}_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\right) + O\left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|-2}}\right).$$

For future reference, we will define

$$\mathtt{r}_{\mathtt{V}} = \sqrt{\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}}}.$$

Remark 2.2 (Vanishing boundary variance when $\mu=0$). In case (ii), the true conditional distribution function is 0 or 1, which is why the leading variance shrinks faster. We do not provide a formula as the leading variance in this case takes a complicated form.

Now, we propose two estimators for the variance that are valid for all three cases of Lemma 2.3, and hence will be useful for establishing a self-normalized distributional approximation later. Define

$$\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x}) = \frac{1}{n^2}\sum_{i=1}^{n}\check{\mathscr{K}}^{\circ}_{\mu,\nu,h}(y_i,\mathbf{x}_i;y,\mathbf{x})^2, \quad \hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x}) = \frac{1}{n^2}\sum_{i=1}^{n}\hat{\mathscr{K}}^{\circ}_{\mu,\nu,h}(y_i,\mathbf{x}_i;y,\mathbf{x})^2.$$

Note that $\hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$ is simply the plug-in variance estimator for $\hat{\theta}_{\mu,\nu}(y|\mathbf{x})$ and $\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$ is the plug-in variance estimator for $\check{\theta}_{\mu,\nu}(y|\mathbf{x})$. The next lemma provides pointwise convergence results for the two variance estimators.

Lemma 2.4 (Variance estimation). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$ and $nh^{d+1}/\log(n) \to \infty$. In addition, $\theta_{0,\nu}$ exists and is continuous for all $|\nu| \le \mathfrak{q}+1$. Then*
(i) $\mu = 0$ and $\theta_{0,\mathbf{0}} \neq 0$ or 1:

$$\left|\frac{\check{\mathsf{V}}_{0,\nu}(y,\mathbf{x}) - \mathsf{V}_{0,\nu}(y,\mathbf{x})}{\mathsf{V}_{0,\nu}(y,\mathbf{x})}\right| = O_{\mathbb{P}}\Big(h^{\mathfrak{q}+1} + \sqrt{\frac{\log(n)}{nh^d}}\Big).$$

(ii) $\mu > 0$, or $\theta_{0,\mathbf{0}} = 0$ or 1:

$$\left|\frac{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x}) - \mathsf{V}_{\mu,\nu}(y,\mathbf{x})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\right| = O_{\mathbb{P}}\Big(h^{\mathfrak{q}+\frac{1}{2}} + \sqrt{\frac{\log(n)}{nh^{d+1}}}\Big).$$

Let $G = F_y$, then the same conclusions hold for $\hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$.

Next, we study the large-sample distributional properties of the infeasible, standardized statistic

$$\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x}) = \frac{1}{n\sqrt{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}}\sum_{i=1}^{n}\mathscr{K}^{\circ}_{\mu,\nu,h}(y_i,\mathbf{x}_i;y,\mathbf{x}).$$

Note that this is equivalent to the scaled asymptotic linear representation of the estimator.

Theorem 2.1 (Asymptotic normality). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$. Then*

$$\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\left[\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x}) \le u\right] - \Phi(u)\Big| = O\left(\mathtt{r}_{\mathtt{BE}}\right), \quad \textit{where } \mathtt{r}_{\mathtt{BE}} = \begin{cases} \dfrac{1}{\sqrt{nh^d}} & \textit{if } \mu = 0, \textit{ and } \theta_{0,\mathbf{0}} \neq 0 \textit{ or } 1 \\ \dfrac{1}{\sqrt{nh^{d+1}}} & \textit{if } \mu > 0, \textit{ or if } \theta_{0,\mathbf{0}} = 0 \textit{ or } 1 \end{cases}.$$

While the theorem focuses on asymptotic normality of the infeasible t-statistic, $\bar{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$, we show in the following remark that similar conclusions can be made for the t-statistics constructed with the estimators, $\hat{\theta}_{\mu,\nu}(y|\mathbf{x})$ and $\check{\theta}_{\mu,\nu}(y|\mathbf{x})$.

Remark 2.3 (Asymptotic normality of standardized statistics). We first introduce the statistic

$$\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) = \frac{\check{\theta}_{\mu,\nu}(y|\mathbf{x}) - \mathbb{E}\left[\check{\theta}_{\mu,\nu}(y|\mathbf{x})|\mathbf{X}\right]}{\sqrt{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}},$$

which is based on $\check{\theta}_{\mu,\nu}(y|\mathbf{x})$. (In the main paper we directly center all statistics at the target parameter $\theta_{\mu,\nu}$. For clarity, however, we will separate the discussion on distributional convergence from the smoothing bias in this supplementary material. This is reflected by the superscript "circle.") By combining the results of Lemmas 2.1 and 2.3, we have

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) - \bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})\right| = O_{\mathtt{TC}}\Big(\frac{\log(n)}{\sqrt{nh^d}}\Big).$$

As a result,

$$\sup_{u\in\mathbb{R}} \left|\mathbb{P}\left[\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) \le u\right] - \Phi(u)\right| = O\Big(\frac{\log(n)}{\sqrt{nh^d}} + \mathtt{r}_{\mathtt{BE}}\Big).$$

To present the pointwise distributional approximation result for the estimator $\hat{\theta}_{\mu,\nu}(y|\mathbf{x})$, we define the following statistic

$$\hat{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) = \frac{1}{nh^{d+\mu+|\nu|}\sqrt{\mathbb{V}_{\mu,\nu}(y,\mathbf{x})}} \sum_{i=1}^{n} \mathbf{e}_{\mu}^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\left[\frac{1}{n}\sum_{j=1}^{n}\left[\mathbb{1}(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right]\frac{1}{h}\mathbf{P}\Big(\frac{y_j - y}{h}\Big)\right] \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\nu}.$$

It is worth mentioning that $\hat{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$ is not exactly centered and therefore, it is not mean zero. Nevertheless, by the results of Lemmas 2.1 and 2.3, and the concentration inequality for second order U-statistics in Equation (3.5) of [4] (Lemmas 7 and 8 in the main paper), we have

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\hat{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) - \check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})\right| = O_{\mathtt{TC}}\Big(\frac{\log(n)}{\sqrt{nh^2}}\Big).$$

Then we can conclude that the coverage error satisfies

$$\sup_{u\in\mathbb{R}} \left|\mathbb{P}\left[\hat{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) \le u\right] - \Phi(u)\right| = O\Big(\frac{\log(n)}{\sqrt{nh^{d\vee 2}}} + \mathtt{r}_{\mathtt{BE}}\Big).$$

3. Uniform large-sample properties

To conduct statistical inference on the entire function $\theta_{\mu,\nu}$, such as constructing confidence bands or testing shape restrictions, we need uniform distributional approximations to our estimators. In this section, we will consider large-sample properties of our estimator which hold uniformly on $\mathcal{Y}\times\mathcal{X} = [0,1]^{d+1}$. In the following remark, we demonstrate that the local sample size is uniformly large on the support $\mathcal{Y}\times\mathcal{X}$.

Remark 3.1 (Local sample size). Consider an evaluation point $(y,\mathbf{x})$ in $\mathcal{Y}\times\mathcal{X}$. We can define the local sample size by

$$n_{y,\mathbf{x}} = \sum_{i=1}^{n} \mathbb{1}(|y_i - y| \le \mathfrak{c}_1 h)\mathbb{1}(|\mathbf{x}_i - \mathbf{x}| \le \mathfrak{c}_1 h).$$

We employed the Euclidean norm in the definition, which is innocuous for our purposes, as all norms are equivalent in finite dimensional spaces. For this reason, we also introduced the constant $\mathfrak{c}_1$. The purpose of this remark is to provide a uniform control on the local sample size. In particular, we have the following result: for some positive constant $\mathfrak{c}_2$ and any shrinking sequence $\mathfrak{r}$,

$$\mathbb{1}\left(\inf_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|n_{y,\mathbf{x}}\right| < \mathfrak{c}_2 \frac{\log(n)}{nh^{d+1}}\right) = O_{\mathtt{TC}}(\mathfrak{r}).$$

The result above builds on the lemma:

Lemma 3.1 (Probabilistic bound on the smallest multinomial cell). *Let* $\mathbf{z} = (z_1, z_2, \ldots, z_{J_n})^\top$ *follow a multinomial distribution with parameters n (number of trials),* J_n *(number of cells), and* $1/J_n$ *(probability for each cell),* $\delta_n \in (0,1)$*, and* $\pi_n = n/(J_n \log(n))$*. If* $\delta_n^2 \pi_n \to \infty$*, then for any* $\mathfrak{c}_1 > 0$*,*

$$\limsup_{n\to\infty} n^{\mathfrak{c}_1}\mathbb{P}\left[\min_{1\le j\le J_n} z_j < (1-\delta_n)\frac{n}{J_n}\right] < \infty.$$

We now establish the uniform convergence rate of our estimator.

Lemma 3.2 (Uniform rate of convergence). *Let Assumptions S-DGP, S-K and S-W hold with* $h \to 0$ *and* $nh^{d+1}/\log(n) \to \infty$*. In addition,* $\theta_{\mu',\boldsymbol{\nu}'}$ *exists and is continuous for all* $\mu' + |\boldsymbol{\nu}'| = \max\{\mathfrak{q}+1+\mu,\ \mathfrak{p}+1+|\boldsymbol{\nu}|\}$*. Then*
(i) $\mu = 0$*:*

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\check{\theta}_{0,\boldsymbol{\nu}}(y|\mathbf{x}) - \theta_{0,\boldsymbol{\nu}}(y|\mathbf{x})\right| = O_{\mathtt{TC}}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1} + \sqrt{\frac{\log(n)}{nh^{d+2|\boldsymbol{\nu}|}}}\right);$$

(ii) $\mu > 0$*:*

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})\right| = O_{\mathtt{TC}}\left(h^{\mathfrak{q}+1-|\boldsymbol{\nu}|} + h^{\mathfrak{p}+1-\mu} + \sqrt{\frac{\log(n)}{nh^{d+2\mu+2|\boldsymbol{\nu}|-1}}}\right).$$

The same conclusions hold for $\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$*.*

In the next lemma, we characterize the uniform convergence rate of the variance estimators introduced in the previous section.

Lemma 3.3 (Uniform variance estimation). *Let Assumptions S-DGP, S-K and S-W hold with* $h \to 0$ *and* $nh^{d+1}/\log(n) \to \infty$*. In addition,* $\theta_{0,\boldsymbol{\nu}}$ *exists and is continuous for all* $|\boldsymbol{\nu}| \le \mathfrak{q}+1$*. Then*

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\frac{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right| = O_{\mathtt{TC}}(\mathtt{r}_{\mathtt{VE}}), \quad \textit{where } \mathtt{r}_{\mathtt{VE}} = h^{\mathfrak{q}+\frac{1}{2}} + \sqrt{\frac{\log(n)}{nh^{d+1}}}.$$

Let $G = F_y$*, then the same conclusions hold for* $\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$*.*

Now, we introduce the Studentized processes for each of the estimators, $\hat{\theta}_{\mu,\boldsymbol{\nu}}$ and $\check{\theta}_{\mu,\boldsymbol{\nu}}$:

$$\check{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \sqrt{\frac{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}\check{\mathbb{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}), \qquad \hat{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \sqrt{\frac{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}\hat{\mathbb{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}).$$

In the following lemma we study the error that arises from the Studentization of our estimators.

Lemma 3.4 (Studentization error). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$ and $nh^{d+1}/\log(n) \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \le \mathfrak{q}+1$. Then*

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\check{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \check{\mathbb{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\right| = O_{\mathtt{TC}}\left(\mathtt{r}_{\mathtt{SE}}\right), \qquad \textit{where } \mathtt{r}_{\mathtt{SE}} = \sqrt{\log(n)}\,\mathtt{r}_{\mathtt{VE}}.$$

The same holds for $\hat{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \hat{\mathbb{S}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$.

Our next goal is to establish a uniform normal approximation to the process $\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$. See Appendix A.4 of the main paper for important properties of the equivalent kernel.

Theorem 3.1 (Strong approximation). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$ and $nh^{d+1}/\log(n) \to \infty$. Also assume $\mu \ge 1$. Define*

$$\mathtt{r}_{\mathtt{SA}} = \left(\frac{\log^{d+1} n}{nh^{d+1}}\right)^{\frac{1}{2d+2}}.$$

Then there exist two centered processes, $\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ and $\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, such that (i) $\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ and $\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ have the same distribution, (ii) $\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ is a Gaussian process and has the same covariance kernel as $\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, and (iii)

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\bar{\mathbb{S}}'_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) - \mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\right| = O_{\mathtt{TC}}\left(\mathtt{r}_{\mathtt{SA}}\right).$$

The Gaussian approximation in the above lemma is not feasible, as its covariance kernel depends on unknowns. To be more precise, the covariance kernel takes the form

$$\rho_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \mathbb{C}\mathrm{ov}\left[\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}), \bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')\right] = \frac{\mathsf{C}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\mathsf{V}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')}},$$

where

$$\mathsf{C}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{1}{n}\mathbb{C}\mathrm{ov}\Big[\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y,\mathbf{x}\right), \mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y',\mathbf{x}'\right)\Big].$$

We consider two estimators of the covariance kernel

$$\check{\rho}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{\check{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')}}, \quad \hat{\rho}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{\hat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y',\mathbf{x}')}},$$

and

$$\check{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{1}{n^2}\sum_{i=1}^{n} \check{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y,\mathbf{x}\right)\check{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y',\mathbf{x}'\right)$$

$$\hat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') = \frac{1}{n^2}\sum_{i=1}^{n} \hat{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y,\mathbf{x}\right)\hat{\mathscr{K}}^{\circ}_{\mu,\boldsymbol{\nu},h}\left(y_i,\mathbf{x}_i;y',\mathbf{x}'\right).$$

Lemma 3.5 (Uniform consistency of the correlation estimator). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$ and $nh^{d+1}/\log(n) \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \leq \mathfrak{q}+1$. Then*

$$\sup_{y,y'\in\mathcal{Y},\mathbf{x},\mathbf{x}'\in\mathcal{X}} \left|\check{\rho}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}') - \rho_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')\right| = O_{\mathtt{TC}}\left(\mathtt{r}_{\mathtt{VE}}\right),$$

where $\mathtt{r}_{\mathtt{VE}}$ is defined in Lemma 3.3. Let $G = F_y$, then the same conclusion holds for $\hat{\rho}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y',\mathbf{x}')$.

Lemma 3.6 (Gaussian comparison). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$, $nh^{d+1}/\log(n) \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \leq \mathfrak{q}+1$. Then conditional on the data there exists a centered Gaussian process, $\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ with unit variance and correlation function $\check{\rho}_{\mu,\boldsymbol{\nu}}$, and another centered Gaussian process, $\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ with unit variance and correlation kernel $\hat{\rho}_{\mu,\boldsymbol{\nu}}$, such that*

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big|\mathbf{Y},\mathbf{X}\Big] - \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big]\right| = O_{\mathbb{P}}\left(\log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right),$$

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big|\mathbf{Y},\mathbf{X}\Big] - \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\mathbb{G}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big]\right| = O_{\mathbb{P}}\left(\log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right).$$

Theorem 3.2 (Feasible normal approximation). *Let Assumptions S-DGP, S-K and S-W hold with $h \to 0$ and $nh^{d+1}/\log(n) \to \infty$. In addition, $\theta_{0,\boldsymbol{\nu}}$ exists and is continuous for all $|\boldsymbol{\nu}| \leq \mathfrak{q}+1$. Also assume $\mu \geq 1$. Then*

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big] - \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big|\mathbf{Y},\mathbf{X}\Big]\right| \\
&\qquad = O_{\mathbb{P}}\left(\sqrt{\log(n)}\,\mathtt{r}_{\mathtt{SA}} + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right), \\
&\sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\hat{\mathbb{T}}^{\circ}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big] - \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big|\mathbf{Y},\mathbf{X}\Big]\right| \\
&\qquad = O_{\mathbb{P}}\left(\sqrt{\log(n)}\,\mathtt{r}_{\mathtt{SA}} + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right).
\end{aligned}$$

4. Applications

4.1. Confidence bands

A natural corollary of Theorem 3.2 is that one can employ critical values computed from $\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ and $\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$ to construct confidence bands. To be very precise, define

$$\check{\mathtt{cv}}_{\mu,\boldsymbol{\nu}}(\alpha) = \inf\left\{u:\ \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big|\mathbf{Y},\mathbf{X}\Big] \geq 1-\alpha\right\},$$

$$\hat{\mathtt{cv}}_{\mu,\boldsymbol{\nu}}(\alpha) = \inf\left\{u:\ \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\hat{\mathbb{G}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})| \leq u\Big|\mathbf{Y},\mathbf{X}\Big] \geq 1-\alpha\right\}.$$

Then level $(1-\alpha)$ confidence bands can be constructed as

$$\check{C}_{\mu,\boldsymbol{\nu}}(1-\alpha) = \left\{\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) \pm \check{\mathtt{cv}}_{\mu,\boldsymbol{\nu}}(\alpha)\sqrt{\check{\mathbb{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}:\ (y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X}\right\},$$

$$\hat{C}_{\mu,\boldsymbol{\nu}}(1-\alpha) = \left\{\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) \pm \hat{\mathtt{cv}}_{\mu,\boldsymbol{\nu}}(\alpha)\sqrt{\hat{\mathbb{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}:\ (y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X}\right\},$$

whose coverage error is given in the following theorem.

Theorem 4.1 (Confidence band). *Consider the setting of Theorem 3.2. In addition, $\theta_{\mu',\boldsymbol{\nu}'}$ exists and is continuous for all $\mu' + |\boldsymbol{\nu}'| = \max\{\mathfrak{q}+1+\mu,\ \mathfrak{p}+1+|\boldsymbol{\nu}|\}$. Then*

$$\mathbb{P}\left[\theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) \in \check{C}_{\mu,\boldsymbol{\nu}}(1-\alpha),\ \forall (y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X}\right] \geq 1-\alpha - O\left(\sqrt{\log(n)}\left(\mathtt{r}_{\mathtt{SA}} + \frac{\mathtt{r}_{\mathtt{B}}}{\mathtt{r}_{\mathtt{V}}}\right) + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right),$$

$$\mathbb{P}\left[\theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) \in \hat{C}_{\mu,\boldsymbol{\nu}}(1-\alpha),\ \forall (y,\mathbf{x}) \in \mathcal{Y}\times\mathcal{X}\right] \geq 1-\alpha - O\left(\sqrt{\log(n)}\left(\mathtt{r}_{\mathtt{SA}} + \frac{\mathtt{r}_{\mathtt{B}}}{\mathtt{r}_{\mathtt{V}}}\right) + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right).$$

4.2. Parametric specification testing

In applications, it is not uncommon to estimate conditional densities or higher-order derivatives by specifying a parametric family of distributions. While such parametric restrictions may provide reasonable approximations, it is still worthwhile to conduct specification testing. To be specific, assume the researcher postulates the following class

$$\left\{\theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x};\boldsymbol{\gamma}):\ \boldsymbol{\gamma} \in \Gamma_{\mu,\boldsymbol{\nu}}\right\},$$

where $\Gamma_{\mu,\boldsymbol{\nu}}$ is some compact parameter space. We abstract away from the specifics of the estimation technique, and assume that the researcher also picks some estimator (maximum likelihood, minimum distance, etc.) $\hat{\boldsymbol{\gamma}}$. Under fairly mild conditions, the estimator will converge in probability to some (possibly pseudo-true) parameter $\bar{\boldsymbol{\gamma}}$ in the parameter space $\Gamma_{\mu,\boldsymbol{\nu}}$. As before, we will denote the true parameter as $\theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x})$, and consider the following competing hypotheses:

$$\mathsf{H}_0:\ \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x};\bar{\boldsymbol{\gamma}}) = \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) \quad \text{vs.} \quad \mathsf{H}_1:\ \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x};\bar{\boldsymbol{\gamma}}) \neq \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}).$$

The test statistics we employ takes the following form

$$\check{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x}) = \frac{\check{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x};\hat{\boldsymbol{\gamma}})}{\sqrt{\check{\mathbb{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}, \quad \hat{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x}) = \frac{\hat{\theta}_{\mu,\boldsymbol{\nu}}(y|\mathbf{x}) - \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x};\hat{\boldsymbol{\gamma}})}{\sqrt{\hat{\mathbb{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}}.$$

Theorem 4.2 (Parametric specification testing). *Consider the setting of Theorem 3.2. In addition, $\theta_{\mu',\boldsymbol{\nu}'}$ exists and is continuous for all $\mu' + |\boldsymbol{\nu}'| = \max\{\mathfrak{q}+1+\mu,\ \mathfrak{p}+1+|\boldsymbol{\nu}|\}$. Assume the parametric estimate satisfies*

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x};\hat{\boldsymbol{\gamma}}) - \theta_{\mu,\boldsymbol{\nu}}(y|\mathbf{x};\bar{\boldsymbol{\gamma}})\right| = O_{\mathtt{TC}}\left(\mathtt{r}_{\mathtt{PS}}\right),$$

for some $\mathtt{r}_{\mathtt{PS}}$. Then under the null hypothesis,

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\check{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x})| > \check{\mathtt{cv}}_{\mu,\boldsymbol{\nu}}(\alpha)\Big] \leq \alpha + O\left(\sqrt{\log(n)}\left(\mathtt{r}_{\mathtt{SA}} + \frac{\mathtt{r}_{\mathtt{B}}+\mathtt{r}_{\mathtt{PS}}}{\mathtt{r}_{\mathtt{V}}}\right) + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\right),$$

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\hat{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x})| > \hat{\mathtt{cv}}_{\mu,\nu}(\alpha)\Big] \le \alpha + O\Big(\sqrt{\log(n)}\Big(\mathtt{r}_{\mathtt{SA}} + \frac{\mathtt{r}_{\mathtt{B}} + \mathtt{r}_{\mathtt{PS}}}{\mathtt{r}_{\mathtt{V}}}\Big) + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\Big).$$

4.3. Testing shape restrictions

Now consider shape restrictions on the conditional density or its derivatives. Let $c(y,\mathbf{x})$ be a pre-specified function, and we study the following one-sided competing hypotheses.

$$\mathsf{H}_0 : \theta_{\mu,\nu}(y|\mathbf{x}) \le c(y,\mathbf{x}) \quad \text{vs.} \quad \mathsf{H}_1 : \theta_{\mu,\nu}(y|\mathbf{x}) > c(y,\mathbf{x}).$$

The statistic we employ takes the form

$$\check{\mathbb{T}}_{\mathtt{SR}}(y,\mathbf{x}) = \frac{\check{\theta}_{\mu,\nu}(y|\mathbf{x}) - c(y,\mathbf{x})}{\sqrt{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}, \quad \hat{\mathbb{T}}_{\mathtt{SR}}(y,\mathbf{x}) = \frac{\hat{\theta}_{\mu,\nu}(y|\mathbf{x}) - c(y,\mathbf{x})}{\sqrt{\hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}.$$

and we will reject the null hypothesis if the test statistic exceeds a critical value.

Theorem 4.3 (Shape restriction testing). *Consider the setting of Theorem 3.2. In addition, $\theta_{\mu',\nu'}$ exists and is continuous for all $\mu' + |\nu'| = \max\{\mathfrak{q}+1+\mu,\ \mathfrak{p}+1+|\nu|\}$. Then under the null hypothesis,*

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \check{\mathbb{T}}_{\mathtt{SR}}(y,\mathbf{x}) > \check{\mathtt{cv}}_{\mu,\nu}(\alpha)\Big] \le \alpha + O\Big(\sqrt{\log(n)}\Big(\mathtt{r}_{\mathtt{SA}} + \frac{\mathtt{r}_{\mathtt{B}}}{\mathtt{r}_{\mathtt{V}}}\Big) + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\Big),$$
$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \hat{\mathbb{T}}_{\mathtt{SR}}(y,\mathbf{x}) > \hat{\mathtt{cv}}_{\mu,\nu}(\alpha)\Big] \le \alpha + O\Big(\sqrt{\log(n)}\Big(\mathtt{r}_{\mathtt{SA}} + \frac{\mathtt{r}_{\mathtt{B}}}{\mathtt{r}_{\mathtt{V}}}\Big) + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}}\Big).$$

5. Bandwidth selection

We assume throughout this section that $\mu > 0$. Using the bias expression derived in Lemma 2.2, and the leading variance is as characterized in Lemma 2.3, we can derive precise expressions for bandwidth selection.

5.1. Pointwise asymptotic MSE minimization

Following from [3], the pointwise MSE-optimal bandwidth is defined as a minimizer of the following optimization problem

$$h^*_{\mathfrak{p},\mathfrak{q},\mu,\nu}(y,\mathbf{x}) = \operatorname*{argmin}_{h>0} \Big[\mathsf{V}_{\mu,\nu}(y,\mathbf{x}) + \mathsf{B}^2_{\mu,\nu}(y,\mathbf{x})\Big]$$

The solution to this equation gives an MSE-optimal bandwidth that depends on (i) the order of the polynomials, (ii) the order of the derivative to be estimated, and (iii) the position of the evaluation point.

Case 1: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu$*, odd*

In this case, both the leading bias constants, $B_{(i),\mathfrak{q}+1}(y, \mathbf{x})$ and $B_{(ii),\mathfrak{p}+1}(y, \mathbf{x})$, are nonzero. Therefore, the MSE-optimal bandwidth is

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \Big[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2 \Big] \\
&= \Big[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(\mathfrak{p} + \mathfrak{q} + 2 - \mu - |\boldsymbol{\nu}|) \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2} \frac{1}{n} \Big]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+|\boldsymbol{\nu}|+\mu+1}}.
\end{aligned}
$$

Case 2: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu$*, even; either* $\mathbf{x}$ *or* y *is at or near the boundary*

In this case, at least one of the leading bias constants, $B_{(i),\mathfrak{q}+1}(y, \mathbf{x})$ and $B_{(ii),\mathfrak{p}+1}(y, \mathbf{x})$, is nonzero. Therefore, the MSE-optimal bandwidth is the same as in Case 1:

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \Big[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+2-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2 \Big] \\
&= \Big[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(\mathfrak{p} + \mathfrak{q} + 2 - \mu - |\boldsymbol{\nu}|) \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2} \frac{1}{n} \Big]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+|\boldsymbol{\nu}|+\mu+1}}.
\end{aligned}
$$

Case 3: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu \neq 0$*, even; both* $\mathbf{x}$ *and* y *are interior*

In this case, both leading bias constants are zero. Therefore, the MSE-optimal bandwidth will depend on higher-order bias terms:

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \Big[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+4-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) \right)^2 \Big] \\
&= \Big[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(\mathfrak{p} + \mathfrak{q} + 4 - \mu - |\boldsymbol{\nu}|) \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) \right)^2} \frac{1}{n} \Big]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+|\boldsymbol{\nu}|+\mu+3}}.
\end{aligned}
$$

Case 4: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu = 0$*, even; both* $\mathbf{x}$ *and* y *are interior*

As in Case 3, both leading bias constants are zero. The difference, however, is that the leading bias will involve an extra term:

$$
\begin{aligned}
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) &= \operatorname*{argmin}_{h>0} \Big[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) \\
&\qquad + h^4 \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) + B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y, \mathbf{x}) \right)^2 \Big] \\
&= \Big[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{4 \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) + B_{(iii),\mathfrak{p}+1,\mathfrak{q}+1}(y, \mathbf{x}) \right)^2} \frac{1}{n} \Big]^{\frac{1}{d+2|\boldsymbol{\nu}|+2\mu+3}}.
\end{aligned}
$$

Case 5: $\mathfrak{q} - |\boldsymbol{\nu}| < \mathfrak{p} - \mu$*,* $\mathfrak{q} - |\boldsymbol{\nu}|$ *odd*

In this case, the leading bias will involve only one term:

$$
h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \operatorname*{argmin}_{h>0} \Big[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{2\mathfrak{q}+2-2|\boldsymbol{\nu}|} B_{(i),\mathfrak{q}+1}(y, \mathbf{x})^2 \Big]
$$

$$= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(2\mathfrak{q} + 2 - 2|\boldsymbol{\nu}|) B_{(i),\mathfrak{q}+1}(y, \mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{q}+2\mu+1}}.$$

Case 6: $\mathfrak{q} - |\boldsymbol{\nu}| = \mathfrak{p} - \mu - 1$, $\mathfrak{q} - |\boldsymbol{\nu}|$ *even;* $\mathbf{x}$ *is interior*

In this case, the leading bias will involve two terms:

$$h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+3-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2 \right]$$
$$= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(\mathfrak{p} + \mathfrak{q} + 3 - \mu - |\boldsymbol{\nu}|) \left(B_{(i),\mathfrak{q}+2}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+1}(y, \mathbf{x}) \right)^2} \frac{1}{n} \right]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+\mu+|\boldsymbol{\nu}|+2}}.$$

Case 7: $\mathfrak{q} - |\boldsymbol{\nu}| < \mathfrak{p} - \mu - 1$, $\mathfrak{q} - |\boldsymbol{\nu}|$ *even;* $\mathbf{x}$ *is interior*

In this case, the leading bias will involve only one term:

$$h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{2\mathfrak{q}+4-2|\boldsymbol{\nu}|} B_{(i),\mathfrak{q}+2}(y, \mathbf{x})^2 \right]$$
$$= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(2\mathfrak{q} + 4 - 2|\boldsymbol{\nu}|) B_{(i),\mathfrak{q}+2}(y, \mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{q}+2\mu+3}}.$$

Case 8: $\mathfrak{q} - |\boldsymbol{\nu}| > \mathfrak{p} - \mu$, $\mathfrak{p} - \mu$ *odd*

In this case, the leading bias will involve only one term:

$$h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{2\mathfrak{p}+2-2\mu} B_{(ii),\mathfrak{p}+1}(y, \mathbf{x})^2 \right]$$
$$= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(2\mathfrak{p} + 2 - 2\mu) B_{(ii),\mathfrak{p}+1}(y, \mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{p}+2|\boldsymbol{\nu}|+1}}.$$

Case 9: $\mathfrak{q} - |\boldsymbol{\nu}| - 1 = \mathfrak{p} - \mu$, $\mathfrak{q} - |\boldsymbol{\nu}|$ *even;* y *is interior*

In this case, the leading bias will involve two terms:

$$h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{\mathfrak{p}+\mathfrak{q}+3-\mu-|\boldsymbol{\nu}|} \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) \right)^2 \right]$$
$$= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(\mathfrak{p} + \mathfrak{q} + 3 - \mu - |\boldsymbol{\nu}|) \left(B_{(i),\mathfrak{q}+1}(y, \mathbf{x}) + B_{(ii),\mathfrak{p}+2}(y, \mathbf{x}) \right)^2} \frac{1}{n} \right]^{\frac{1}{d+\mathfrak{p}+\mathfrak{q}+\mu+|\boldsymbol{\nu}|+2}}.$$

Case 10: $\mathfrak{q} - |\boldsymbol{\nu}| - 1 > \mathfrak{p} - \mu$, $\mathfrak{p} - \mu$ *even;* y *is interior*

In this case, the leading bias will involve only one term:

$$h^*_{\mathfrak{p},\mathfrak{q},\mu,\boldsymbol{\nu}}(y, \mathbf{x}) = \underset{h>0}{\operatorname{argmin}} \left[\frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}} V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x}) + h^{2\mathfrak{p}+4-2\mu} B_{(ii),\mathfrak{p}+2}(y, \mathbf{x})^2 \right]$$
$$= \left[\frac{(d + 2|\boldsymbol{\nu}| + 2\mu - 1) V_{\mu,\boldsymbol{\nu}}(y, \mathbf{x})}{(2\mathfrak{p} + 4 - 2\mu) B_{(ii),\mathfrak{p}+2}(y, \mathbf{x})^2} \frac{1}{n} \right]^{\frac{1}{d+2\mathfrak{p}+2|\boldsymbol{\nu}|+3}}.$$

5.2. Rule-of-thumb bandwidth selection

This section outlines the methodology that the companion R package, `lpcde`, uses to construct the rule-of-thumb bandwidth selection.

The rule-of-thumb estimation uses the following assumptions in order to compute the optimal bandwidth:

- the data is jointly normal,
- **X** and **Y** are independent, and,
- $p - \mu = q - |\nu| = 1$.

Using these assumptions, each of the terms in the formula given in Case 1 of Section 5.1 are computed as follows:

1. The densities and relevant derivatives are evaluated based on the joint normal distribution assumption.
2. $\mathbf{S}_y$, $\mathbf{T}_y$, $\mathbf{T}_\mathbf{x}$ and $\mathbf{S}_\mathbf{x}$ matrices are computed by plugging in for the range of the data, the evaluation point, the respective marginal densities, and the kernel used.
3. Similarly, the $\mathbf{c}_y$ and $\mathbf{c}_\mathbf{x}$ vectors are computed by using the range of the data, the evaluation point, kernel function, and the respective marginal densities.
4. Bias and variance estimates are constructed using the relevant entries of the vectors and matrices.

6. Alternative variance estimators

6.1. V-statistic variance estimator

We propose here an alternative variance estimator that is quick to implement in practice. We start by first observing that the estimator $\hat{\theta}_{\mu,\nu}(y|\mathbf{x})$ is a V-statistic. That is,

$$\begin{aligned}\hat{\theta}_{\mu,\nu}(y|\mathbf{x}) &= \frac{1}{n^2 h^{1+\mu+d+|\nu|}} \sum_{i,j} \mathbb{1}(y_i \le y_j) \mathbf{e}_\mu^\top \hat{\mathbf{S}}_y^{-1} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{Q}^\top \left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right) \hat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{e}_\nu \\ &= \frac{1}{n^2}\sum_{i=1}^{n} a(y_i, y) b(\mathbf{x}_i, \mathbf{x}) + \frac{1}{n^2}\sum_{i\neq j} \mathbb{1}(y_i \le y_j) a(y_j, y) b(\mathbf{x}_i, \mathbf{x}),\end{aligned} \tag{6.1}$$

where,

$$a(y_i, y) = h^{-1-\mu} \mathbf{e}_\mu^\top \hat{\mathbf{S}}_y^{-1} \mathbf{P}\left(\frac{y_i - y}{h}\right), \qquad b(\mathbf{x}_i, \mathbf{x}) = h^{-d-|\nu|} \mathbf{e}_\nu^\top \hat{\mathbf{S}}_\mathbf{x}^{-1} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right).$$

Note that $a(\cdot)$ and $b(\cdot)$ are scalar functions that are non-zero only for data points that are within h distance of the evaluation point. The second term in (6.1) can now be symmetrized and treated as a U-statistic. Applying the Hoeffding decomposition to the symmetrized version of the second term and plugging it back into Equation 6.1, we get

$$\begin{aligned}\hat{\theta}_{\mu,\nu}(y|\mathbf{x}) &= \frac{1}{n}\mathbb{E}\left[a(y_i, y) b(\mathbf{x}_i, \mathbf{x})\right] + \frac{n-1}{n}\mathbb{E}[u_{i,j}] \\ &\quad + \frac{1}{n^2}\sum_{i=1}^{n}\left(a(y_i, y) b(\mathbf{x}_i, \mathbf{x}) - \mathbb{E}\left[a(y_i, y) b(\mathbf{x}_i, \mathbf{x})\right]\right) + \frac{n-1}{n} L_{\mu,\nu}(y, \mathbf{x}) + \frac{n-1}{n} W_{\mu,\nu}(y, \mathbf{x})\end{aligned} \tag{6.2}$$

where

$$u_{i,j} = \frac{1}{2}\left(\mathbb{1}(y_i \le y_j)a(y_j, y)b(\mathbf{x}_i, \mathbf{x}) + \mathbb{1}(y_j \le y_i)a(y_i, y)b(\mathbf{x}_i, \mathbf{x})\right),$$

and

$$L_{\mu,\nu}(y, \mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} 2\left(\mathbb{E}\left[u_{i,j}|y_i, \mathbf{x}_i\right] - \mathbb{E}\left[u_{i,j}\right]\right),$$

$$W_{\mu,\nu}(y, \mathbf{x}) = \binom{n}{2}^{-1} \sum_{i,j=1, i\neq j}^{n} \left(u_{i,j} - \mathbb{E}\left[u_{i,j}|y_i, \mathbf{x}_i\right] - \mathbb{E}\left[u_{i,j}|y_j, \mathbf{x}_j\right] + \mathbb{E}\left[u_{i,j}\right]\right).$$

Dependence on polynomial orders is suppressed for notational simplicity. Since each of the terms in (6.1) are orthogonal, the variance of the estimator can be expressed as the sum of the variance of each of the terms on the right hand side. Furthermore, we note that the first three terms and $W_{\mu,\nu}(y, \mathbf{x})$ have higher-order variance. Thus, we only need to look at the variance of $L_{\mu,\nu}(y, \mathbf{x})$.

$$\mathbb{V}[L_{\mu,\nu}(y, \mathbf{x})] = \mathbb{V}\Big[\frac{2}{n}\sum_{i=1}^{n}\left(\mathbb{E}\left[u_{i,j} \mid y_i, \mathbf{x}_i\right] - \mathbb{E}\left[u_{i,j}\right]\right)\Big] = \frac{1}{n}\mathbb{V}\Big[2\mathbb{E}\left[u_{i,j} \mid y_i, \mathbf{x}_i\right] - 2\mathbb{E}\left[u_{i,j}\right]\Big]$$

where we know

$$2\mathbb{E}\left[u_{i,j} \mid y_i, \mathbf{x}_i\right] = \int_{\mathcal{Y}} \mathbb{1}(y_i \le u)a(u, y)\mathrm{d}F(u|\mathbf{x}_i)b(\mathbf{x}_i, \mathbf{x}) + F(y_i|\mathbf{x}_i)a(y_i, y)b(\mathbf{x}_i, \mathbf{x}).$$

We can expand and simplify this to get

$$\begin{aligned}
\mathbb{V}\left[L_{\mu,\nu}(y, \mathbf{x})\right] &= \mathbb{E}\left[\left(\int_{\mathcal{Y}} \mathbb{1}(y_i \le u)a(u, y)\mathrm{d}F(u|\mathbf{x}_i)b(\mathbf{x}_i, \mathbf{x}) + F(y_i|\mathbf{x}_i)a(y_i, y)b(\mathbf{x}_i, \mathbf{x})\right)^2\right] \\
&= \mathbb{E}\Bigg[\iint_{\mathcal{Y}} \mathbb{1}(y_i \le \min\{u, v\})a(u, y)a(v, y)\mathrm{d}F(u|\mathbf{x}_i)\mathrm{d}F(v|\mathbf{x}_i)b^2(\mathbf{x}_i, \mathbf{x}) \\
&\quad + \int_{\mathcal{Y}} \mathbb{1}(y_i \le u)a(u, y)\mathrm{d}F(u|\mathbf{x}_i)F(y_i|\mathbf{x}_i)a(y_i, y)b^2(\mathbf{x}_i, \mathbf{x}) \\
&\quad + (F(y_i|\mathbf{x}_i)a(y_i, y)b(\mathbf{x}_i, \mathbf{x}))^2\Bigg].
\end{aligned}$$

Note that this expression is identical to the variance expression derived in the proof of Lemma 2.3. This leads to a natural alternative jackknife covariance estimator,

$$\hat{\mathrm{C}}_{\mu,\nu}(y, \mathbf{x}, y', \mathbf{x}') = \frac{1}{n}\sum_{i=1}^{n} \hat{L}_{(i),\mu,\nu}(y, \mathbf{x})\hat{L}_{(i),\mu,\nu}(y', \mathbf{x}').$$

where

$$\hat{L}_{(i),\mu,\nu}(y, \mathbf{x}) = \frac{2}{n-1}\sum_{j\neq i}\left(u_{i,j} - \hat{\theta}_{\mu,\nu}(y|\mathbf{x})\right).$$

In particular, note that if the two evaluation points are equivalent, we return the variance estimator,

$$\hat{\mathsf{C}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x},y,\mathbf{x}) \equiv \hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{n-1}\sum_{i=1}^{n}\hat{L}^2_{(i),\mu,\boldsymbol{\nu}}(y,\mathbf{x}).$$

6.2. Asymptotic variance estimator

Another alternative variance estimator is a sample version of the asymptotic variance derived in Lemma 2.3. That is, each of the matrices in the formula are replaced with sample analogs. That is,
(i) $\mu = 0$:

$$\hat{\mathsf{V}}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}\hat{\theta}_{0,\mathbf{0}}(y|\mathbf{x})(1-\hat{\theta}_{0,\mathbf{0}}(y|\mathbf{x}))\Big(\mathbf{e}_{\boldsymbol{\nu}}^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\Big)$$

(ii) $\mu > 0$:

$$\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{nh^{d+2|\boldsymbol{\nu}|+2\mu-1}}\hat{\theta}_{1,\mathbf{0}}(y|\mathbf{x})\Big(\mathbf{e}_{\mu}^{\mathsf{T}}\hat{\mathbf{S}}_{y}^{-1}\hat{\mathbf{T}}_{y}\hat{\mathbf{S}}_{y}^{-1}\mathbf{e}_{\mu}\Big)\Big(\mathbf{e}_{\boldsymbol{\nu}}^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\hat{\mathbf{T}}_{\mathbf{x}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}\Big).$$

The covariance can be estimated using similar idea.

7. Proofs

7.1. Proof of Lemma 2.1

Part (i). See the proof of Lemma 1 in the main paper (i.e., Appendix A.3).

Part (ii). Next consider $\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_{y}^{-1}\big(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}]\big)$, which takes the form

$$\frac{1}{nh^{\mu+|\boldsymbol{\nu}|}}\sum_{i=1}^{n}\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_{y}^{-1}\int_{\frac{\mathcal{Y}-y}{h}}\Big(\mathbb{1}(y_i\le y+hu) - F(y+hu|\mathbf{x}_i)\Big)\mathbf{P}(u)g(y+hu)\mathrm{d}u\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}}.$$

It is straightforward to see that

$$\left|\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_{y}^{-1}\int_{\frac{\mathcal{Y}-y}{h}}\Big(\mathbb{1}(y_i\le y+hu) - F(y+hu|\mathbf{x}_i)\Big)\mathbf{P}(u)g(y+hu)\mathrm{d}u\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}}\right| \le C'h^{-d}$$

for some C' that holds uniformly for $y\in\mathcal{Y}$ and $\mathbf{x}\in\mathcal{X}$. We also have the following bound on the variance

$$\mathbb{V}\left[\mathbf{e}_{\mu}^{\mathsf{T}}\mathbf{S}_{y}^{-1}\int_{\frac{\mathcal{Y}-y}{h}}\Big(\mathbb{1}(y_i\le y+hu) - F(y+hu|\mathbf{x}_i)\Big)\mathbf{P}(u)g(y+hu)\mathrm{d}u\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^{\mathsf{T}}\right]$$
$$\le C'\begin{cases} h^{-d} & \text{if } \mu = 0 \\ h^{-d+1} & \text{if } \mu > 0. \end{cases}.$$

Consider the first case above ($\mu = 0$). By a discretization $\{(y_\ell, \mathbf{x}_\ell) : 1 \le \ell \le M_n\}$ of $\mathcal{Y} \times \mathcal{X}$, we have the probabilistic bound due to Bernstein's inequality

$$\mathbb{P}\Big[h^{\mu+|\boldsymbol{\nu}|} \max_{1\le\ell\le M_n} \Big|\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}])\Big| > \mathfrak{c}_1\mathtt{r}\Big] \le 2\exp\Big\{-\frac{1}{2}\frac{\mathfrak{c}_1^2\log(n)}{C' + \frac{1}{3}\mathfrak{c}_1 C'\mathtt{r}} + \log(M_n)\Big\},$$

provided that we set $\mathtt{r} = \sqrt{\log(n)/(nh^d)}$.$M_n$ is at most polynomial in n, and the error from discretization can be ignored. This concludes the proof for the $\mu = 0$ case.

For $\mu > 0$, we set $\mathtt{r} = \sqrt{\log(n)/(nh^{d-1})}$, and the probabilistic bound takes the form

$$\begin{aligned}
&\mathbb{P}\Big[h^{\mu+|\boldsymbol{\nu}|} \max_{1\le\ell\le M_n} \Big|\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}])\Big| > \mathfrak{c}_1\mathtt{r}\Big] \\
&\le 2\exp\Big\{-\frac{1}{2}\frac{\mathfrak{c}_1^2 n^2\mathtt{r}^2}{nC'h^{-d+1} + \frac{1}{3}\mathfrak{c}_1 C'h^{-d}n\mathtt{r}} + \log(M_n)\Big\} = 2\exp\Big\{-\frac{1}{2}\frac{\mathfrak{c}_1^2\log(n)}{C' + \frac{\mathfrak{c}_1 C'}{3}\sqrt{\frac{\log(n)}{nh^{d+1}}}} + \log(M_n)\Big\}.
\end{aligned}$$

This concludes the proof for the second case, where $\mu > 0$.

7.2. Proof of Lemma 2.2

The conditional expectation of $\bar{\mathbf{R}}_{y,\mathbf{x}}$ in $\check{\theta}_{\mu,\boldsymbol{\nu}}$ is

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{nh^{\mu+|\boldsymbol{\nu}|}}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}\mathbb{1}(y_i\le u)\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big]\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\Big|\mathbf{X}\Big] \\
&= \frac{1}{nh^{\mu+|\boldsymbol{\nu}|}}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}F(u|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big]\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}.
\end{aligned}$$

To proceed, we employ a Taylor expansion of the conditional distribution function to order s:

$$F(u|\mathbf{x}_i) = \sum_{\ell+|\mathbf{m}|\le s}\theta_{\ell,\mathbf{m}}(y|\mathbf{x})\frac{1}{\ell!\mathbf{m}!}(u-y)^\ell(\mathbf{x}_i-\mathbf{x})^\mathbf{m} + o\Big(\sum_{\ell+|\mathbf{m}|=s}|u-y|^\ell|\mathbf{x}_i-\mathbf{x}|^\mathbf{m}\Big).$$

Then, the conditional expectation can be simplified as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}F(u|\mathbf{x}_i)\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big]\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T} \\
&= \sum_{\ell+|\mathbf{m}|\le s}\theta_{\ell,\mathbf{m}}(y|\mathbf{x})\Big[\int_{\mathcal{Y}}\frac{1}{\ell!}(u-y)^\ell\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big]\Big[\frac{1}{n}\sum_{i=1}^n\frac{1}{h^d}\frac{1}{\mathbf{m}!}(\mathbf{x}_i-\mathbf{x})^\mathbf{m}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\Big] \\
&\quad + o\Big(\sum_{\ell+|\mathbf{m}|=s}\Big[\int_{\mathcal{Y}}\frac{1}{\ell!}|u-y|^\ell\frac{1}{h}\Big|\mathbf{P}\Big(\frac{u-y}{h}\Big)\Big|\mathrm{d}G(u)\Big]\Big[\frac{1}{n}\sum_{i=1}^n\frac{1}{h^d}\frac{1}{\mathbf{m}!}|\mathbf{x}_i-\mathbf{x}|^\mathbf{m}\Big|\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big|\Big]\Big) \\
&= \sum_{\ell+|\mathbf{m}|\le s}h^{\ell+|\mathbf{m}|}\theta_{\ell,\mathbf{m}}(y|\mathbf{x})\mathbf{c}_{y,\ell}\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}}^\mathsf{T} + o_{\mathbb{P}}(h^s).
\end{aligned}$$

We note that

$$\mathbf{S}_y^{-1}\mathbf{c}_{y,\ell} = \mathbf{e}_\ell \text{ for all } 0 \le \ell \le \mathfrak{p},$$

and

$$\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}} = \mathbf{e}_{\mathbf{m}} \text{ for all } 0 \le |\mathbf{m}| \le \mathfrak{q}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}[\check{\theta}_{\mu,\nu}(y|\mathbf{x})|\mathbf{X}] \\
&= \theta_{\mu,\nu}(y|\mathbf{x}) + h^{\mathfrak{q}+1-|\nu|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y|\mathbf{x})\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}}^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu + h^{\mathfrak{p}+1-\mu}\theta_{\mathfrak{p}+1,\nu}(y|\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+1}^{\mathsf{T}}\mathbf{S}_y^{-1}\mathbf{e}_\mu \\
&\quad + o_{\mathbb{P}}\left(h^{\mathfrak{q}+1-|\nu|} + h^{\mathfrak{p}+1-\mu}\right).
\end{aligned}$$

By Lemma 2.1, the second term on the right-hand side satisfies

$$h^{\mathfrak{q}+1-|\nu|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y|\mathbf{x})\hat{\mathbf{c}}_{\mathbf{x},\mathbf{m}}^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu = h^{\mathfrak{q}+1-|\nu|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y|\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu + O_{\mathbb{P}}\Big(h^{\mathfrak{q}+1-|\nu|}\sqrt{\frac{\log(n)}{nh^d}}\Big),$$

which means we can denote the leading bias as

$$\mathsf{B}_{\mu,\nu}(y,\mathbf{x}) = h^{\mathfrak{q}+1-|\nu|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y|\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu + h^{\mathfrak{p}+1-\mu}\theta_{\mathfrak{p}+1,\nu}(y|\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+1}^{\mathsf{T}}\mathbf{S}_y^{-1}\mathbf{e}_\mu.$$

For the second claim of this lemma, we again consider a Taylor expansion

$$F(y_j|\mathbf{x}_i) = \sum_{\ell+|\mathbf{m}|\le s}\theta_{\ell,\mathbf{m}}(y|\mathbf{x})\frac{1}{\ell!\mathbf{m}!}(y_j-y)^\ell(\mathbf{x}_i-\mathbf{x})^{\mathbf{m}} + o\Big(\sum_{\ell+|\mathbf{m}|=s}|y_j-y|^\ell|\mathbf{x}_i-\mathbf{x}|^{\mathbf{m}}\Big).$$

Then

$$\begin{aligned}
&\frac{1}{n^2h^{d+1+\mu+|\nu|}}\sum_{i,j=1}^n\mathbf{e}_\mu^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\left[F(y_j|\mathbf{x}_i)\mathbf{P}\left(\frac{y_j-y}{h}\right)\right]\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu \\
=&\frac{1}{n^2h^{d+1+\mu+|\nu|}}\sum_{i,j=1}^n\mathbf{e}_\mu^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\Big[\sum_{\ell+|\mathbf{m}|\le s}\theta_{\ell,\mathbf{m}}(y|\mathbf{x})\frac{1}{\ell!\mathbf{m}!}(y_j-y)^\ell(\mathbf{x}_i-\mathbf{x})^{\mathbf{m}}\mathbf{P}\left(\frac{y_j-y}{h}\right)\Big]\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathsf{T}}\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu \\
&\quad + o\Big(\frac{1}{n^2h^{d+1+\mu+|\nu|}}\mathbf{e}_\mu^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\sum_{i,j=1}^n\Big[\sum_{\ell+|\mathbf{m}|=s}|y_j-y|^\ell|\mathbf{x}_i-\mathbf{x}|^{\mathbf{m}}\Big|\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\Big|\Big|\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big|\Big]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu\Big) \\
=&\theta_{\mu,\nu}(y|\mathbf{x}) + h^{\mathfrak{q}+1-|\nu|}\sum_{|\mathbf{m}|=\mathfrak{q}+1}\theta_{\mu,\mathbf{m}}(y|\mathbf{x})\mathbf{c}_{\mathbf{x},\mathbf{m}}^{\mathsf{T}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu + h^{\mathfrak{p}+1-\mu}\theta_{\mathfrak{p}+1,\nu}(y|\mathbf{x})\mathbf{c}_{y,\mathfrak{p}+1}^{\mathsf{T}}\mathbf{S}_y^{-1}\mathbf{e}_\mu \\
&\quad + o_{\mathbb{P}}(h^{\mathfrak{q}+1-|\nu|} + h^{\mathfrak{p}+1-\mu}).
\end{aligned}$$

7.3. Proof of Lemma 2.3

Let $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$ and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$.

$$
\begin{aligned}
&\mathbb{V}\Big[\int_{\mathcal{Y}}\Big[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i)\Big]\mathbf{c}_1^\mathsf{T}\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\mathbf{c}_2\Big]\\
&=\mathbb{E}\Big[\mathbb{V}\Big[\int_{\mathcal{Y}}\Big[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i)\Big]\mathbf{c}_1^\mathsf{T}\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)^\mathsf{T}\mathbf{c}_2\Big|\mathbf{X}\Big]\Big]\\
&=\mathbb{E}\Big[\iint_{\frac{\mathcal{Y}-y}{h}}\Big(F(y+h(u_1\wedge u_2)|\mathbf{x}_i) - F(y+hu_1|\mathbf{x}_i)F(y+hu_2|\mathbf{x}_i)\Big)\mathbf{c}_1^\mathsf{T}\mathbf{P}(u_1)\,\mathbf{c}_1^\mathsf{T}\mathbf{P}(u_2)\\
&\qquad\qquad g(y+hu_1)g(y+hu_2)\mathrm{d}u_1\mathrm{d}u_2\Big(\mathbf{c}_2^\mathsf{T}\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big)^2\Big] \qquad (*)
\end{aligned}
$$

We make a further expansion:

$$
\begin{aligned}
&F(y+h(u_1\wedge u_2)|\mathbf{x}_i) - F(y+hu_1|\mathbf{x}_i)F(y+hu_2|\mathbf{x}_i)\\
&= F(y|\mathbf{x}_i)(1-F(y|\mathbf{x}_i)) + h(u_1\wedge u_2)f(y|\mathbf{x}_i) - h(u_1+u_2)f(y|\mathbf{x}_i)F(y|\mathbf{x}_i) + O(h^2).
\end{aligned}
$$

Note that the remainder term, $O(h^2)$, holds uniformly for $y \in \mathcal{Y}$ and $\mathbf{x}_i \in \mathcal{X}$ since the conditional distribution function is assumed to have bounded second derivative. Therefore,

$$
\begin{aligned}
(*) &= \Big(\mathbf{c}_1^\mathsf{T}\mathbf{c}_{y,0}\mathbf{c}_{y,0}^\mathsf{T}\mathbf{c}_1\Big)\mathbb{E}\Big[F(y|\mathbf{x}_i)(1-F(y|\mathbf{x}_i))\Big(\mathbf{c}_2^\mathsf{T}\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big)^2\Big]\\
&\quad + h\Big(\mathbf{c}_1^\mathsf{T}\mathbf{T}_y\mathbf{c}_1\Big)\mathbb{E}\Big[f(y|\mathbf{x}_i)\Big(\mathbf{c}_2^\mathsf{T}\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big)^2\Big]\\
&\quad - h\Big[\mathbf{c}_1^\mathsf{T}\Big(\mathbf{c}_{y,1}\mathbf{c}_{y,0}^\mathsf{T} + \mathbf{c}_{y,0}\mathbf{c}_{y,1}^\mathsf{T}\Big)\mathbf{c}_1\Big]\mathbb{E}\Big[f(y|\mathbf{x}_i)F(y|\mathbf{x}_i)\Big[\mathbf{c}_2^\mathsf{T}\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big]^2\Big] + O\Big(\frac{1}{h^{d-2}}\Big)\\
&= \mathbf{e}_\mu^\mathsf{T}\mathbf{e}_0\mathbb{E}\Big[F(y|\mathbf{x}_i)(1-F(y|\mathbf{x}_i))\Big(\mathbf{c}_2^\mathsf{T}\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big)^2\Big]\\
&\quad + h\Big(\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\mathbf{T}_y\mathbf{S}_y^{-1}\mathbf{e}_\mu\Big)\mathbb{E}\Big[f(y|\mathbf{x}_i)\Big(\mathbf{c}_2^\mathsf{T}\frac{1}{h^d}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big)^2\Big] + O\Big(\frac{1}{h^{d-2}}\Big).
\end{aligned}
$$

To conclude the proof, we note that two scenarios can arise: $\mu = 0$ and $\mu > 0$. In the second case,

$$
(*) = \frac{1}{h^{d-1}}\theta_{1,\mathbf{0}}(y|\mathbf{x})\Big(\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\mathbf{T}_y\mathbf{S}_y^{-1}\mathbf{e}_\mu\Big)\Big(\mathbf{e}_\nu^\mathsf{T}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{T}_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu\Big) + O\Big(\frac{1}{h^{d-2}}\Big).
$$

The first case is more involved. If $\theta_{0,\mathbf{0}}(y|\mathbf{x}) \ne 0, 1$, then

$$
(*) = \frac{1}{h^d}\theta_{0,\mathbf{0}}(y|\mathbf{x})(1-\theta_{0,\mathbf{0}}(y|\mathbf{x}))\Big(\mathbf{e}_\nu^\mathsf{T}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{T}_{\mathbf{x}}\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu\Big) + O_{\mathbb{P}}\Big(\frac{1}{h^{d-1}}\Big).
$$

If $\theta_{0,\mathbf{0}}(y|\mathbf{x}) = 0$ or 1, then a further expansion is needed, which is why an extra h will be present in the leading variance.

7.4. Proof of Lemma 2.4

Consistency of $\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$. For the purposes of this proof, let $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$, and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$. To start, consider

$$
\begin{aligned}
&\frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}\Big(\mathbb{1}(y_i\le u)-\hat F(u|\mathbf{x}_i)\Big)\frac{1}{h}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big]^2\\
&=\frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}\Big(\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\Big)\frac{1}{h}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big]^2 &&\text{(I)}\\
&-\frac{2}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big]^2\iint_{\frac{\mathcal{Y}-y}{h}}\Big(\mathbb{1}(y_i\le u_1)-F(u_1|\mathbf{x}_i)\Big)\Big(\hat F(u_2|\mathbf{x}_i)-F(u_2|\mathbf{x}_i)\Big)\\
&\qquad\qquad\mathbf{c}_1^{\mathsf{T}}\mathbf{P}(u_1)\,\mathbf{c}_1^{\mathsf{T}}\mathbf{P}(u_2)\,g(y+hu_1)g(y+hu_2)\mathrm{d}u_1\mathrm{d}u_2 &&\text{(II)}\\
&+\frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}\Big(\hat F(u|\mathbf{x}_i)-F(u|\mathbf{x}_i)\Big)\frac{1}{h}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big]^2. &&\text{(III)}
\end{aligned}
$$

First consider term (III). With the uniform convergence result for the estimated conditional distribution function, it is clear that

$$
|\text{(III)}|\precsim_{\mathbb{P}}\begin{cases}\mathsf{V}_{0,\nu}(y,\mathbf{x})\left(h^{2\mathfrak{q}+2}+\dfrac{\log(n)}{nh^d}\right) & \text{if } \mu=0,\text{ and } \theta_{0,\mathbf{0}}\neq 0\text{ or }1\\ \mathsf{V}_{\mu,\nu}(y,\mathbf{x})\left(h^{2\mathfrak{q}+1}+\dfrac{\log(n)}{nh^{d+1}}\right) & \text{if } \mu>0,\quad\text{or } \theta_{0,\mathbf{0}}=0\text{ or }1\end{cases}.
$$

Now we study term (I), which is clearly unbiased for $\mathsf{V}_{\mu,\nu}(y,\mathbf{x})$. Therefore, we compute its variance.

$$
\begin{aligned}
\mathbb{V}[\text{(I)}]&=\frac{1}{n^3h^{4d+4\mu+4|\nu|}}\mathbb{V}\Big[\Big(\int_{\mathcal{Y}}\Big(\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\Big)\frac{1}{h}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big)^2\Big]\\
&\le\frac{1}{n^3h^{4d+4\mu+4|\nu|}}\mathbb{E}\Big[\prod_{j=1}^4\Big[\int_{\mathcal{Y}}\Big(\mathbb{1}(y_i\le u_j)-F(u_j|\mathbf{x}_i)\Big)\frac{1}{h}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\Big(\frac{u_j-y}{h}\Big)\mathrm{d}G(u_j)\Big]\Big[\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big]^4\Big].
\end{aligned}
$$

With iterative expectation (by conditioning on $\mathbf{x}_i$), the above further reduces to

$$
\begin{aligned}
\mathbb{V}[\text{(I)}]&=\frac{1}{n^3h^{3d+4\mu+4|\nu|}}\theta_{0,\mathbf{0}}(1-\theta_{0,\mathbf{0}})(1-3\theta_{0,\mathbf{0}}(1-\theta_{0,\mathbf{0}}))\left[\mathbf{c}_1^{\mathsf{T}}\mathbf{c}_{y,0}\right]^4\mathbb{E}\left[\frac{1}{h^d}\Big[\mathbf{c}_2^{\mathsf{T}}\mathbf{R}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big]^4\right]\\
&\quad+O\left(\frac{h}{n^3h^{4d+4\mu+4|\nu|}}\right).
\end{aligned}
$$

In other words,

$$
\Big|\text{(I)}-\mathsf{V}_{\mu,\nu}(y,\mathbf{x})\Big|\precsim_{\mathbb{P}}\begin{cases}\mathsf{V}_{0,\nu}(y,\mathbf{x})\sqrt{\dfrac{1}{nh^d}} & \text{if } \mu=0,\text{ and } \theta_{0,\mathbf{0}}\neq 0\text{ or }1\\ \mathsf{V}_{\mu,\nu}(y,\mathbf{x})\sqrt{\dfrac{1}{nh^{d+1}}} & \text{if } \mu>0,\quad\text{or } \theta_{0,\mathbf{0}}=0\text{ or }1\end{cases}.
$$

Finally, we consider (II). Using the Cauchy-Schwartz inequality, we have

$$|(\mathrm{II})|^2 \leq |(\mathrm{I})| \cdot |(\mathrm{III})| \,.$$

As a result,

$$|(\mathrm{II})| \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{h^{2\mathfrak{q}+2}+\dfrac{\log(n)}{nh^d}} & \text{if } \mu=0, \text{ and } \theta_{0,\mathbf{0}}\neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{h^{2\mathfrak{q}+1}+\dfrac{\log(n)}{nh^{d+1}}} & \text{if } \mu>0, \ \text{ or } \theta_{0,\mathbf{0}}=0 \text{ or } 1 \end{cases}.$$

To conclude the proof for $\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor $1+O_{\mathbb{P}}(1/\sqrt{nh^d})$. See Lemma 2.1.

Consistency of $\hat{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$. For the purposes of this proof, let $\hat{\mathbf{c}}_1=\hat{\mathbf{S}}_y^{-1}\mathbf{e}_\mu$, $\mathbf{c}_1=\mathbf{S}_y^{-1}\mathbf{e}_\mu$, $\hat{\mathbf{c}}_2=\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$, and $\mathbf{c}_2=\mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$. We first consider the following decomposition

$$\begin{aligned}
&\frac{1}{n^2h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}\left(\mathbb{1}\left(y_i\leq u\right)-\hat{F}(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}\hat{F}_y(u)\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2\\
=&\frac{1}{n^2h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{i=1}^n\Big[\frac{1}{n^2}\sum_{j,k=1}^n\left(\mathbb{1}\left(y_i\leq y_j\right)-F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\left(y_i\leq y_k\right)-F(y_k|\mathbf{x}_i)\right)\\
&\qquad\frac{1}{h^2}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_k-y}{h}\right)\Big]\Big[\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2 &&\text{(I)}\\
-&\frac{2}{n^2h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{i=1}^n\Big[\frac{1}{n^2}\sum_{j,k=1}^n\left(\hat{F}(y_j|\mathbf{x}_i)-F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\left(y_i\leq y_k\right)-F(y_k|\mathbf{x}_i)\right)\\
&\qquad\frac{1}{h^2}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_k-y}{h}\right)\Big]\Big[\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2 &&\text{(II)}\\
+&\frac{1}{n^2h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{i=1}^n\Big[\frac{1}{n^2}\sum_{j,k=1}^n\left(\hat{F}(y_j|\mathbf{x}_i)-F(y_j|\mathbf{x}_i)\right)\left(\hat{F}(y_k|\mathbf{x}_i)-F(y_k|\mathbf{x}_i)\right)\\
&\qquad\frac{1}{h^2}\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_k-y}{h}\right)\Big]\Big[\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2. &&\text{(III)}
\end{aligned}$$

By the uniform convergence rate of the estimated conditional distribution function, we have

$$|(\mathrm{III})| \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\left(h^{2\mathfrak{q}+2}+\dfrac{\log(n)}{nh^d}\right) & \text{if } \mu=0, \text{ and } \theta_{0,\mathbf{0}}\neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\left(h^{2\mathfrak{q}+1}+\dfrac{\log(n)}{nh^{d+1}}\right) & \text{if } \mu>0, \ \text{ or } \theta_{0,\mathbf{0}}=0 \text{ or } 1 \end{cases}.$$

Next we consider (II). Using the Cauchy-Schwartz inequality, we have

$$|(\mathrm{II})|^2 \leq |(\mathrm{I})| \cdot |(\mathrm{III})| \,.$$

Finally, consider term (I), which has the expansion

$$
\begin{aligned}
\text{(I)} = {} & \frac{1}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|+2}} \sum_{\substack{i,j,k=1\\ \text{distinct}}}^{n} \Big[\left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i)\right) \left(\mathbb{1}\left(y_i \le y_k\right) - F(y_k|\mathbf{x}_i)\right) \\
& \qquad \mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_k - y}{h}\right) \Big] \left[\mathbf{c}_2^{\mathsf{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2 && \text{(I.1)} \\
& + \frac{2}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|+2}} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} \Big[\left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i)\right) \left(1 - F(y_i|\mathbf{x}_i)\right) \mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right) \\
& \qquad \mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right) \Big] \left[\mathbf{c}_2^{\mathsf{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2 && \text{(I.2)} \\
& + \frac{1}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|+2}} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} \left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i)\right)^2 \left[\mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right)\right]^2 \left[\mathbf{c}_2^{\mathsf{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2 && \text{(I.3)} \\
& + \frac{1}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|+2}} \sum_{i=1}^{n} \left(1 - F(y_i|\mathbf{x}_i)\right)^2 \left[\mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right)\right]^2 \left[\mathbf{c}_2^{\mathsf{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2 . && \text{(I.4)}
\end{aligned}
$$

Then,

$$
\begin{aligned}
|\text{(I.2)}| &\le \frac{2}{n^4 h^{2d+2\mu+2|\boldsymbol{\nu}|+2}} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} \left|\mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_j - y}{h}\right)\right| \cdot \left|\mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right)\right| \left[\mathbf{c}_2^{\mathsf{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2 \\
&\le \frac{2}{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|}} \left[\frac{1}{nh} \sum_{i=1}^{n} \left|\mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right)\right|\right] \left[\frac{1}{nh^{d+1}} \sum_{i=1}^{n} \left|\mathbf{c}_1^{\mathsf{T}} \mathbf{P}\left(\frac{y_i - y}{h}\right)\right| \left[\mathbf{c}_2^{\mathsf{T}} \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2\right] \\
&\precsim_{\mathbb{P}} \frac{1}{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|}} \left(1 + \sqrt{\frac{1}{nh}}\right) \left(1 + \sqrt{\frac{1}{nh^{d+1}}}\right) \precsim \frac{1}{n^2 h^{d+2\mu+2|\boldsymbol{\nu}|}} \\
&\precsim \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) \dfrac{1}{n} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \ne 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \dfrac{1}{nh} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases} .
\end{aligned}
$$

Using similar techniques, one can show that

$$
\begin{aligned}
|\text{(I.3)}| &\precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) \dfrac{1}{nh} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \ne 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \dfrac{1}{nh^2} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases} \\
|\text{(I.4)}| &\precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x}) \dfrac{1}{n^2 h} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \ne 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) \dfrac{1}{n^2 h^2} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases} .
\end{aligned}
$$

To streamline the remaining derivation, define

$$\phi_{j,i} = \frac{1}{h}\left(\mathbb{1}\left(y_i \le y_j\right) - F(y_j|\mathbf{x}_i)\right)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_j - y}{h}\right), \quad \phi_i = \mathbb{E}[\phi_{j,i}|y_i, \mathbf{x}_i], \quad \psi_i = \left[\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right]^2.$$

Then

$$\begin{aligned}
\text{(I.1)} &= \frac{1}{n^4 h^{2d+2\mu+2|\nu|}} \sum_{\substack{i,j,k=1\\ \text{distinct}}}^{n} \phi_{j,i}\phi_{k,i}\psi_i \\
&= \underbrace{\frac{1}{n^4 h^{2d+2\mu+2|\nu|}} \sum_{\substack{i,j,k=1\\ \text{distinct}}}^{n} (\phi_{j,i}-\phi_i)(\phi_{k,i}-\phi_i)\psi_i}_{\text{(I.1.1)}} + \left(2 + O\left(\frac{1}{n}\right)\right) \underbrace{\frac{1}{n^3 h^{2d+2\mu+2|\nu|}} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} (\phi_{j,i}-\phi_i)\phi_i\psi_i}_{\text{(I.1.2)}} \\
&\quad + \left(1 + O\left(\frac{1}{n}\right)\right) \underbrace{\frac{1}{n^2 h^{2d+2\mu+2|\nu|}} \sum_{i=1}^{n} \phi_i^2\psi_i}_{\text{(I.1.3)}}.
\end{aligned}$$

We have studied the term (I.1.3) in the proof for $\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$. In particular,

$$\left|\text{(I.1.3)} - \mathsf{V}_{\mu,\nu}(y,\mathbf{x})\right| \precsim_{\mathbb{P}} \begin{cases} \mathsf{V}_{0,\nu}(y,\mathbf{x})\sqrt{\dfrac{1}{nh^d}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\nu}(y,\mathbf{x})\sqrt{\dfrac{1}{nh^{d+1}}} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases}.$$

Term (I.1.1) is a mean zero third order U-statistic. Consider its variance

$$\left(\frac{1}{nh^{d+2\mu+2|\nu|}}\right)^2 \mathbb{E}\Big[\frac{1}{n^6 h^{2d}} \sum_{\substack{i,j,k=1\\ \text{distinct}}}^{n} \sum_{\substack{i',j',k'=1\\ \text{distinct}}}^{n} (\phi_{j,i}-\phi_i)(\phi_{k,i}-\phi_i)(\phi_{j',i'}-\phi_{i'})(\phi_{k',i'}-\phi_{i'})\psi_i\psi_{i'}\Big].$$

The above expectation is non-zero only in three scenarios: $(j = j', k = k', i \neq i')$, $(j = j', k = k', i = i')$ or $(j = i', k = k', i = j')$. Therefore,

$$\begin{aligned}
|\text{(I.1.1)}| &\precsim_{\mathbb{P}} \left(\frac{1}{nh^{d+2\mu+2|\nu|}}\right)\left(\frac{1}{nh} + \sqrt{\frac{1}{n^3 h^{d+2}}}\right) \precsim \left(\frac{1}{nh^{d+2\mu+2|\nu|}}\right)\frac{1}{nh} \\
&\precsim \begin{cases} \mathsf{V}_{0,\nu}(y,\mathbf{x})\dfrac{1}{nh} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \neq 0 \text{ or } 1 \\ \mathsf{V}_{\mu,\nu}(y,\mathbf{x})\dfrac{1}{nh^2} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases}.
\end{aligned}$$

Finally consider (I.1.2), which has a mean of zero. Its variance is

$$\mathbb{V}\left[\text{(I.1.2)}\right] = \left(\frac{1}{nh^{d+2\mu+2|\nu|}}\right)^2 \mathbb{E}\Big[\frac{1}{n^4 h^{2d}} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} \sum_{\substack{i',j'=1\\ \text{distinct}}}^{n} (\phi_{j,i}-\phi_i)(\phi_{j',i'}-\phi_{i'})\phi_i\psi_i\phi_{i'}\psi_{i'}\Big]$$

$$= \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)^2 \mathbb{E}\Big[\frac{1}{n^4h^{2d}} \sum_{\substack{i,i',j=1\\ \text{distinct}}}^{n} (\phi_{j,i}-\phi_i)(\phi_{j,i'}-\phi_{i'})\phi_i\psi_i\phi_{i'}\psi_{i'}\Big]$$

$$+ \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)^2 \mathbb{E}\Big[\frac{1}{n^4h^{2d}} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} (\phi_{j,i}-\phi_i)^2\phi_i^2\psi_i^2\Big]$$

$$+ \mathbb{E}\Big[\frac{1}{n^4h^{2d}} \sum_{\substack{i,j=1\\ \text{distinct}}}^{n} (\phi_{j,i}-\phi_i)(\phi_{i,j}-\phi_j)\phi_i\psi_i\phi_j\psi_j\Big]$$

$$\precsim \left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)^2 \left(\frac{1}{nh} + \frac{1}{n^2h^{d+1}}\right).$$

In addition, an extra h factor emerges if $\mu > 0$, or if $\theta_{0,\mathbf{0}} = 0$ or 1. As a result,

$$|(\text{I}.1.2)| \precsim_{\mathbb{P}} \begin{cases} \mathrm{V}_{0,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\dfrac{1}{nh}} & \text{if } \mu = 0, \text{ and } \theta_{0,\mathbf{0}} \neq 0 \text{ or } 1 \\ \mathrm{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})\sqrt{\dfrac{1}{nh^2}} & \text{if } \mu > 0, \quad \text{or } \theta_{0,\mathbf{0}} = 0 \text{ or } 1 \end{cases}.$$

To conclude the proof for $\hat{\mathrm{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_1$ by $\mathbf{c}_1$ and $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor $1 + O_{\mathbb{P}}(1/\sqrt{nh^d})$. See Lemma 2.1.

7.5. Proof of Theorem 2.1

We will write

$$\bar{\mathbb{S}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})}{\sqrt{\mathbb{V}\left[h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})\right]}}.$$

Define $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$ and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_{\boldsymbol{\nu}}$.

To apply the Berry-Esseen theorem, we first compute the third moment

$$\mathbb{E}\left[\left|h^{d+\mu+|\boldsymbol{\nu}|}\mathscr{K}^{\circ}_{\mu,\boldsymbol{\nu},h}(y_i,\mathbf{x}_i;y,\mathbf{x})\right|^3\right]$$

$$= \mathbb{E}\Big[\Big(\prod_{j=1}^{3}\Big|\int_{\frac{\mathcal{Y}-y}{h}} \Big(\mathbb{1}(y_i \le y + hu_j) - F(y+hu_j|\mathbf{x}_i)\Big)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}(u_j)\,\mathrm{d}G(u_j)\Big|\Big)\Big|\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big|^3\Big].$$

The leading term in the above is simply

$$\mathbb{E}\Big[\Big(\prod_{j=1}^{3}\Big|\int_{\frac{\mathcal{Y}-y}{h}} \Big(\mathbb{1}(y_i \le y) - F(y|\mathbf{x}_i)\Big)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}(u_j)\,\mathrm{d}G(u_j)\Big|\Big)\Big|\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\Big(\frac{\mathbf{x}_i-\mathbf{x}}{h}\Big)\Big|^3\Big]$$

$$= \left|\mathbf{c}_1^{\mathsf{T}}\mathbf{c}_{y,0}\right|^3 \mathbb{E}\Big[\Big|\Big(\mathbb{1}(y_i \le y) - F(y|\mathbf{x}_i)\Big)\Big|^3\Big|\mathbf{c}_2^{\mathsf{T}}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big|^3\Big]$$

$$= \left|\mathbf{c}_1^\top \mathbf{c}_{y,0}\right|^3 \mathbb{E}\Big[\Big(\theta_{0,\mathbf{0}}(y|\mathbf{x})(1-\theta_{0,\mathbf{0}}(y,\mathbf{x}))(2\theta_{0,\mathbf{0}}(y|\mathbf{x})^2 - 2\theta_{0,\mathbf{0}}(y|\mathbf{x})+1)\Big)\left|\mathbf{c}_2^\top \mathbf{Q}\left(\frac{\mathbf{x}_i - \mathbf{x}}{h}\right)\right|^3\Big] = O(h^d).$$

Note that the above will be exactly zero in cases (ii) and (iii) of Lemma 2.3.

7.6. Omitted details of Remark 2.3

Approximation and coverage error of $\check{\mathbb{S}}^\circ_{\mu,\nu}(y,\mathbf{x})$. To start,

$$\check{\mathbb{S}}^\circ_{\mu,\nu}(y,\mathbf{x}) - \bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x}) = \frac{1}{nh^{\mu+|\nu|}\sqrt{\mathbb{V}_{\mu,\nu}(y,\mathbf{x})}} \sum_{i=1}^n \mathbf{e}_\mu^\top \mathbf{S}_y^{-1}\Big[\int_{\mathcal{Y}}\Big[\mathbb{1}(y_i \le u) - F(u|\mathbf{x}_i)\Big]\frac{1}{h}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big]$$
$$\frac{1}{h^d}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\top \left(\hat{\mathbf{S}}_{\mathbf{x}}^{-1} - \mathbf{S}_{\mathbf{x}}^{-1}\right)\mathbf{e}_\nu.$$

By allowing the constant $\mathfrak{c}_1$ to take possibly different values in each term, we have

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\check{\mathbb{S}}^\circ_{\mu,\nu}(y,\mathbf{x}) - \bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})\right| \ge \mathfrak{c}_1 \frac{\log(n)}{\sqrt{nh^d}}\Big]$$
$$\le \mathbb{P}\Big[\sup_{\mathbf{x}\in\mathcal{X}}\left|\hat{\mathbf{S}}_{\mathbf{x}} - \mathbf{S}_{\mathbf{x}}\right| > \mathfrak{c}_1\sqrt{\frac{\log(n)}{nh^d}}\Big] + \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{\mathbf{e}_\mu^\top \mathbf{S}_y^{-1}\left(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}\right]\right)}{\sqrt{\mathbb{V}_{\mu,\nu}(y,\mathbf{x})}}\right| > \mathfrak{c}_1\sqrt{\log(n)}\Big] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3},$$

where the conclusions follow from the uniform rates established in Lemma 2.1 and the variance calculations in Lemma 2.3. Next, we consider the normal approximation error. Note that

$$\mathbb{P}\left[\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x}) \le u - \mathfrak{c}_1\frac{\log(n)}{\sqrt{nh^d}}\right] - \mathfrak{c}_2 n^{-\mathfrak{c}_3} \le \mathbb{P}\left[\check{\mathbb{S}}^\circ_{\mu,\nu}(y,\mathbf{x}) \le u\right] \le \mathbb{P}\left[\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x}) \le u + \mathfrak{c}_1\frac{\log(n)}{\sqrt{nh^d}}\right] + \mathfrak{c}_2 n^{-\mathfrak{c}_3},$$

which means

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\check{\mathbb{S}}^\circ_{\mu,\nu}(y,\mathbf{x}) \le u\right] - \Phi(u)\right| \precsim \frac{\log(n)}{\sqrt{nh^d}} + \mathtt{r}_{\mathtt{BE}},$$

where $\mathtt{r}_{\mathtt{BE}}$ is defined in Theorem 2.1.

Approximation and coverage error of $\hat{\mathbb{S}}^\circ_{\mu,\nu}(y,\mathbf{x})$. To begin with, we decompose the double sum into

$$\frac{1}{n^2h^{d+1}}\sum_{i,j=1}^n \left[\mathbb{1}(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right]\mathbf{P}\Big(\frac{y_j-y}{h}\Big)\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\top$$
$$= \frac{1}{n^2h^{d+1}}\sum_{i=1}^n \left(\left[1 - F(y_i|\mathbf{x}_i)\right]\mathbf{P}\Big(\frac{y_i-y}{h}\Big) - \int_{\mathcal{Y}}\left[\mathbb{1}(y_i\le u) - F(u|\mathbf{x}_i)\right]\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\right)\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\top \tag{I}$$
$$+ \frac{1}{n^2h^{d+1}}\sum_{\substack{i,j=1\\ \text{distinct}}}^n \left(\left[\mathbb{1}(y_i \le y_j) - F(y_j|\mathbf{x}_i)\right]\mathbf{P}\Big(\frac{y_j-y}{h}\Big) - \int_{\mathcal{Y}}\left[\mathbb{1}(y_i\le u) - F(u|\mathbf{x}_i)\right]\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\right)$$
$$\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\top \tag{II}$$

$$+\frac{1}{nh^{d+1}}\sum_{i=1}^{n}\int_{\mathcal{Y}}\Big[\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\Big]\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathsf{T}}, \tag{III}$$

where we set $G=F_y$. Term (I) represents the leave-in bias, and it is straightforward to show that

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|(\mathrm{I})|>\mathfrak{c}_1\frac{1}{n}\Big(1+\sqrt{\frac{\log(n)}{nh^{d+1}}}\Big)\Big]\le\mathfrak{c}_2 n^{-\mathfrak{c}_3},$$

for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$. See Lemma 2.1 for the proof strategy.

Term (II) is a degenerate U-statistic. Define

$$u_{i,j}=\mathbf{c}_1^{\mathsf{T}}\left(\Big[\mathbb{1}(y_i\le y_j)-F(y_j|\mathbf{x}_i)\Big]\mathbf{P}\Big(\frac{y_j-y}{h}\Big)-\int_{\mathcal{Y}}\Big[\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\Big]\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\right)$$
$$\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^{\mathsf{T}}\mathbf{c}_2,$$

where $\mathbf{c}_1$ and $\mathbf{c}_2$ are arbitrary (fixed) vectors of conformable dimensions. Then we apply Equation (3.5) of [4] (Lemmas 7 and 8 in the main paper), which gives (the value of C' may change for each line)

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\sum_{i,j=1,\ i\neq j}^{n}u_{i,j}\Big|>t\Big]\le C'\exp\left\{-\frac{1}{C'}\min\left[\frac{t}{\sqrt{n^2h^{d+1}}},\ \frac{t^{2/3}}{(nh)^{1/3}},\ t^{1/2}\right]+\log(n)\right\}$$
$$=C'\exp\left\{-\frac{\sqrt{C}}{C'}\min\left[\log(n),\ \Big((\log(n))^2nh^d\Big)^{\frac{1}{3}},\ \Big((\log(n))^2n^2h^{d+1}\Big)^{\frac{1}{4}}\right]+\log(n)\right\}.$$

As a result,

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|(\mathrm{II})|>\mathfrak{c}_1\frac{\log(n)}{\sqrt{n^2h^{d+1}}}\Big]\le\mathfrak{c}_2n^{-\mathfrak{c}_3},$$

for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$.

We now collect the pieces. The difference between $\hat{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$ and $\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$ is

$$\hat{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})-\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$$
$$=\frac{1}{h^{\mu+|\nu|}\sqrt{\mathbb{V}_{\mu,\nu}(y,\mathbf{x})}}\mathbf{e}_\mu^{\mathsf{T}}\hat{\mathbf{S}}_y^{-1}\Big[(\mathrm{I})+(\mathrm{II})\Big]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu+\frac{1}{\sqrt{\mathbb{V}_{\mu,\nu}(y,\mathbf{x})}}\mathbf{e}_\mu^{\mathsf{T}}\left(\hat{\mathbf{S}}_y^{-1}-\mathbf{S}_y^{-1}\right)\Big[\bar{\mathbf{R}}_{y,\mathbf{x}}-\mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}\right]\Big]\hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu,$$

and the conclusion follows from Lemmas 2.1 and 2.3.

7.7. Omitted details of Remark 3.1

To show this result, we first partition the support $\mathcal{Y}\times\mathcal{X}$ into cubes with edge length $\mathfrak{c}_3h$, where the constant $\mathfrak{c}_3$ is chosen so that, for any $(y,\mathbf{x})$ in $\mathcal{Y}\times\mathcal{X}$, at least one of the cubes will be contained in the ball $\{y':|y'-y|\le\mathfrak{c}_1h\}\times\{\mathbf{x}':|\mathbf{x}'-\mathbf{x}|\le\mathfrak{c}_1h\}$. The number of cubes in this partition is $\lceil 1/(\mathfrak{c}_3h)^{d+1}\rceil$. Then the conclusion follows from Lemma 3.1.

Proof of Lemma 3.1. For simplicity let $c_n = (1-\delta_n)\frac{n}{J_n}$. We first employ the union bound

$$\mathbb{P}\Big[\min_{1\le j\le J_n} z_j < c_n\Big] \le J_n \cdot \mathbb{P}\left[z_j < c_n\right].$$

Note that $z_j \sim \mathrm{Binomial}(n; \frac{1}{J_n})$, and therefore

$$\mathbb{P}\left[z_j < c_n\right] = \mathbb{P}\Big[z_j - \frac{n}{J_n} < c_n - \frac{n}{J_n}\Big] \le \exp\left(-\frac{1}{2}\frac{\left(\frac{n}{J_n}-c_n\right)^2}{\frac{n}{J_n}\left(1-\frac{1}{J_n}\right)+\frac{1}{3}\left(\frac{n}{J_n}-c_n\right)}\right)$$
$$\le \exp\left(-\frac{3}{8}\frac{J_n}{n}\left(\frac{n}{J_n}-c_n\right)^2\right).$$

Then we have

$$\mathbb{P}\Big[\min_{1\le j\le J_n} z_j < c_n\Big] \le J_n \exp\left(-\frac{3\delta_n^2}{8}\frac{n}{J_n}\right) = \frac{1}{\pi_n \log n}\exp\left(-\left(\frac{3}{8}\delta_n^2\pi_n - 1\right)\log n\right).$$

Therefore, the above will vanish faster than any polynomial of n provided that $\delta_n^2\pi_n \to \infty$.

7.8. Proof of Lemma 3.2

Part (i) Convergence of $\check{\theta}_{\mu,\nu} - \theta_{\mu,\nu}$. Recall that we have the following decomposition of our estimator

$$\check{\theta}_{\mu,\nu} - \theta_{\mu,\nu}$$
$$= \frac{1}{nh^{1+d+\mu+|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}} F(u|\mathbf{x}_i)\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\mathsf{T}\hat{\mathbf{S}}_\mathbf{x}^{-1}\mathbf{e}_\nu - \theta_{\mu,\nu} \tag{I}$$
$$+ \frac{1}{nh^{1+d+\mu+|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}} \Big(\mathbb{1}(y_i\le u) - F(u|\mathbf{x}_i)\Big)\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\mathsf{T}\mathbf{S}_\mathbf{x}^{-1}\mathbf{e}_\nu \tag{II}$$
$$+ \frac{1}{nh^{1+d+\mu+|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}} \Big(\mathbb{1}(y_i\le u) - F(u|\mathbf{x}_i)\Big)\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\right]\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)^\mathsf{T}\Big(\hat{\mathbf{S}}_\mathbf{x}^{-1} - \mathbf{S}_\mathbf{x}^{-1}\Big)\mathbf{e}_\nu. \tag{III}$$

(I) is simply the conditional bias, whose order is given in Lemma 2.2. The convergence rate of (II) can be easily deduced from that of $\mathbf{e}_\mu^\mathsf{T}\mathbf{S}_y^{-1}\left(\bar{\mathbf{R}}_{y,\mathbf{x}} - \mathbb{E}\left[\bar{\mathbf{R}}_{y,\mathbf{x}}|\mathbf{X}\right]\right)$ in Lemma 2.1. Finally, it should be clear that (III) is negligible relative to (II).

Part (ii) Convergence of $\hat{\theta}_{\mu,\nu} - \theta_{\mu,\nu}$. This part follows from Remark 2.3.

7.9. Proof of Lemma 3.3

Uniform consistency of $\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$. For the purposes of this proof, let $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$, and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$. To start, consider

$$\frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}}\left(\mathbb{1}(y_i\le u)-\hat{F}(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2$$

$$= \frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}}\left(\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2 \qquad \text{(I)}$$

$$- \frac{2}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^{n}\iint_{\frac{\mathcal{Y}-y}{h}}\left(\mathbb{1}(y_i\le u_1)-F(u_1|\mathbf{x}_i)\right)\left(\hat{F}(u_2|\mathbf{x}_i)-F(u_2|\mathbf{x}_i)\right)\mathbf{c}_1^\mathsf{T}\mathbf{P}(u_1)\,\mathbf{c}_1^\mathsf{T}\mathbf{P}(u_2)$$

$$g(y+hu_1)g(y+hu_2)\mathrm{d}u_1\mathrm{d}u_2\left[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2 \qquad \text{(II)}$$

$$+ \frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^{n}\left[\int_{\mathcal{Y}}\left(\hat{F}(u|\mathbf{x}_i)-F(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2. \qquad \text{(III)}$$

First consider term (I). Clearly this term is unbiased for $\mathsf{V}_{\mu,\nu}(y,\mathbf{x})$. In the proof of Lemma 2.4, we showed that

$$\mathbb{V}\left[\left(\int_{\mathcal{Y}}\left(\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right)^2\right]\le C_1\begin{cases}h^d & \text{if } \mu=0\\ h^{d+1} & \text{if } \mu>0\end{cases}.$$

Also note that

$$\left(\int_{\mathcal{Y}}\left(\mathbb{1}(y_i\le u)-F(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}G(u)\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right)^2\le C_2.$$

In the above, the constants, C_1 and C_2, can be chosen to be independent of the evaluation point, the sample size, and the bandwidth. Then by a proper discretization of $\mathcal{Y}\times\mathcal{X}$, and applying the union bound and Bernstein's inequality, one has

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{\text{(I)}-\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\right|>\mathfrak{c}_1\mathtt{r}_1\right]\le\mathfrak{c}_2n^{-\mathfrak{c}_3},\qquad \mathtt{r}_1=\begin{cases}\sqrt{\dfrac{\log(n)}{nh^d}} & \text{if } \mu=0\\ \sqrt{\dfrac{\log(n)}{nh^{d+1}}} & \text{if } \mu>0\end{cases},$$

for some constants $\mathfrak{c}_1$, $\mathfrak{c}_2$, and $\mathfrak{c}_3$. In addition, $\mathfrak{c}_3$ can be made arbitrarily large by appropriate choices of $\mathfrak{c}_1$. See the proof of Lemma 2.1 for an example of this proof strategy.

Next consider term (III). With the uniform convergence result for the estimated conditional distribution function, it is clear that

$$\mathbb{P}\left[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{\text{(III)}}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\right|>\mathfrak{c}_1\mathtt{r}_3\right]\le\mathfrak{c}_2n^{-\mathfrak{c}_3},\qquad \mathtt{r}_3=h^{2\mathfrak{q}+1}+\frac{\log(n)}{nh^{d+1}}.$$

Finally, we consider (II). Using the Cauchy-Schwartz inequality, we have

$$|\text{(II)}|^2\le|\text{(I)}|\cdot|\text{(III)}|.$$

As a result,

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\text{(II)}}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\Big| > \mathfrak{c}_1\mathtt{r}_2\Big] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3}, \qquad \mathtt{r}_2 = \sqrt{1+\mathtt{r}_1}\sqrt{\mathtt{r}_3}.$$

To conclude the proof for $\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor $1+O_{\mathbb{P}}(\sqrt{\log(n)/(nh^d)})$. See Lemma 2.1.

Uniform consistency of $\hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$. For the purposes of this proof, let $\hat{\mathbf{c}}_1 = \hat{\mathbf{S}}_y^{-1}\mathbf{e}_\mu$, $\mathbf{c}_1 = \mathbf{S}_y^{-1}\mathbf{e}_\mu$, $\hat{\mathbf{c}}_2 = \hat{\mathbf{S}}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$, and $\mathbf{c}_2 = \mathbf{S}_{\mathbf{x}}^{-1}\mathbf{e}_\nu$. We consider the same decomposition used in the proof of Lemma 2.4:

$$\begin{aligned}
&\frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\int_{\mathcal{Y}}\left(\mathbb{1}\left(y_i\le u\right)-\hat{F}(u|\mathbf{x}_i)\right)\frac{1}{h}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{u-y}{h}\right)\mathrm{d}\hat{F}_y(u)\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2\\
=&\frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\frac{1}{n^2}\sum_{j,k=1}^n\left(\mathbb{1}\left(y_i\le y_j\right)-F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\left(y_i\le y_k\right)-F(y_k|\mathbf{x}_i)\right)\\
&\qquad\frac{1}{h^2}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_k-y}{h}\right)\Big]\Big[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2 &\text{(I)}\\
-&\frac{2}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\frac{1}{n^2}\sum_{j,k=1}^n\left(\hat{F}(y_j|\mathbf{x}_i)-F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\left(y_i\le y_k\right)-F(y_k|\mathbf{x}_i)\right)\\
&\qquad\frac{1}{h^2}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_k-y}{h}\right)\Big]\Big[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2 &\text{(II)}\\
+&\frac{1}{n^2h^{2d+2\mu+2|\nu|}}\sum_{i=1}^n\Big[\frac{1}{n^2}\sum_{j,k=1}^n\left(\hat{F}(y_j|\mathbf{x}_i)-F(y_j|\mathbf{x}_i)\right)\left(\hat{F}(y_k|\mathbf{x}_i)-F(y_k|\mathbf{x}_i)\right)\\
&\qquad\frac{1}{h^2}\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_k-y}{h}\right)\Big]\Big[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2. &\text{(III)}
\end{aligned}$$

By the uniform convergence rate for the estimated conditional distribution function, we have

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\text{(III)}}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\Big| > \mathfrak{c}_1\mathtt{r}_3\Big] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3}, \qquad \mathtt{r}_3 = h^{2\mathfrak{q}+1}+\frac{\log(n)}{nh^{d+1}}.$$

Employing the Cauchy-Schwartz inequality gives

$$|\text{(II)}|^2 \le |\text{(I)}|\cdot|\text{(III)}|.$$

As a result, a probabilistic order for term (II) follows that of terms (I) and (III).

Finally, consider term (I), which has the expansion

$$\begin{aligned}
\text{(I)} = &\frac{1}{n^4h^{2d+2\mu+2|\nu|+2}}\sum_{\substack{i,j,k=1\\ \text{distinct}}}^n\Big[\left(\mathbb{1}\left(y_i\le y_j\right)-F(y_j|\mathbf{x}_i)\right)\left(\mathbb{1}\left(y_i\le y_k\right)-F(y_k|\mathbf{x}_i)\right)\\
&\qquad\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_k-y}{h}\right)\Big]\Big[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\Big]^2 &\text{(I.1)}
\end{aligned}$$

$$+\frac{2}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|+2}}\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}\Big[\left(\mathbb{1}\left(y_i\le y_j\right)-F(y_j|\mathbf{x}_i)\right)\left(1-F(y_i|\mathbf{x}_i)\right)$$

$$\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right)\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_i-y}{h}\right)\Big]\left[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2 \tag{I.2}$$

$$+\frac{1}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|+2}}\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}\left(\mathbb{1}\left(y_i\le y_j\right)-F(y_j|\mathbf{x}_i)\right)^2\left[\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right)\right]^2\left[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2 \tag{I.3}$$

$$+\frac{1}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|+2}}\sum_{i=1}^{n}\left(1-F(y_i|\mathbf{x}_i)\right)^2\left[\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_i-y}{h}\right)\right]^2\left[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2. \tag{I.4}$$

Then,

$$|(\text{I.2})|\le\frac{1}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|+2}}\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}\left|\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right)\right|\cdot\left|\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_i-y}{h}\right)\right|\left[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2$$
$$\le\left(\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\right)\frac{2}{n}\left[\frac{1}{nh}\sum_{i=1}^{n}\left|\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_i-y}{h}\right)\right|\right]\left[\frac{1}{nh^{d+1}}\sum_{i=1}^{n}\left|\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_i-y}{h}\right)\right|\left[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2\right],$$

which means

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.2})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|\ge\mathfrak{c}_1\frac{1}{nh}\Big]\le\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

Using similar techniques, one can show that

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.3})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|\ge\mathfrak{c}_1\frac{1}{nh^2}\Big]\le\mathfrak{c}_2n^{-\mathfrak{c}_3},\quad\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\left|\frac{(\text{I.4})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\right|\ge\mathfrak{c}_1\frac{1}{n^2h^2}\Big]\le\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

To streamline the remaining derivation, define

$$\phi_{j,i}=\frac{1}{h}\left(\mathbb{1}\left(y_i\le y_j\right)-F(y_j|\mathbf{x}_i)\right)\mathbf{c}_1^\mathsf{T}\mathbf{P}\left(\frac{y_j-y}{h}\right),\quad\phi_i=\mathbb{E}[\phi_{j,i}|y_i,\mathbf{x}_i],\quad\psi_i=\left[\mathbf{c}_2^\mathsf{T}\mathbf{Q}\left(\frac{\mathbf{x}_i-\mathbf{x}}{h}\right)\right]^2.$$

Then

$$(\text{I.1})=\frac{1}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{\substack{i,j,k=1\\ \text{distinct}}}^{n}\phi_{j,i}\phi_{k,i}\psi_i$$
$$=\underbrace{\frac{1}{n^4h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{\substack{i,j,k=1\\ \text{distinct}}}^{n}(\phi_{j,i}-\phi_i)(\phi_{k,i}-\phi_i)\psi_i}_{(\text{I.1.1})}+\left(2+O\left(\frac{1}{n}\right)\right)\underbrace{\frac{1}{n^3h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}(\phi_{j,i}-\phi_i)\phi_i\psi_i}_{(\text{I.1.2})}$$

$$+\left(1+O\left(\frac{1}{n}\right)\right)\underbrace{\frac{1}{n^2h^{2d+2\mu+2|\boldsymbol{\nu}|}}\sum_{i=1}^{n}\phi_i^2\psi_i}_{\text{(I.1.3)}}.$$

By employing the same techniques in the proof for $\check{\mathsf{V}}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})$, we have that

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\text{(I.1.3)}-\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\Big|>\mathfrak{c}_1\mathtt{r}_1\Big]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3},\qquad \mathtt{r}_1=\begin{cases}\sqrt{\dfrac{\log(n)}{nh^d}} & \text{if }\mu=0\\ \sqrt{\dfrac{\log(n)}{nh^{d+1}}} & \text{if }\mu>0\end{cases}.$$

Term (I.1.2) admits the following decomposition:

$$\text{(I.1.2)}=\underbrace{\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\frac{n-1}{n^2}\sum_{j=1}^{n}\mathbb{E}\Big[\frac{1}{h^d}(\phi_{j,i}-\phi_i)\phi_i\psi_i\Big|y_j,\mathbf{x}_j\Big]}_{\text{(I.1.2.1)}}$$

$$+\underbrace{\frac{1}{nh^{d+2\mu+2|\boldsymbol{\nu}|}}\frac{1}{n^2h^d}\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}\left((\phi_{j,i}-\phi_i)\phi_i\psi_i-\mathbb{E}\Big[\frac{1}{h^d}(\phi_{j,i}-\phi_i)\phi_i\psi_i\Big|y_j,\mathbf{x}_j\Big]\right)}_{\text{(I.1.2.2)}}.$$

Using the same techniques of Lemmas 2.1 and 2.4, we have

$$\mu=0\qquad \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\text{(I.1.2.1)}}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\Big|>\mathfrak{c}_1\sqrt{\frac{\log(n)}{nh}}\Big]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3},$$

$$\mu>0\qquad \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\text{(I.1.2.1)}}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\Big|>\mathfrak{c}_1\sqrt{\frac{\log(n)}{nh^2}}\Big]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

Term (I.1.2.2) is a degenerate second order U-statistic. We adopt Equation (3.5) of [4] (Lemmas 7 and 8 in the main paper), which implies (see Remark 2.3 and its proof for an example)

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\text{(I.1.2.2)}}{\mathsf{V}_{\mu,\boldsymbol{\nu}}(y,\mathbf{x})}\Big|>\mathfrak{c}_1\sqrt{\frac{\log(n)}{n^2h^{d+3}}}\Big]\leq\mathfrak{c}_2n^{-\mathfrak{c}_3}.$$

To handle term (I.1.1), first consider the quantity $\phi_{j,i}-\phi_i$, which takes the form

$$\max_i\Big|\frac{1}{n}\sum_{j=1}^{n}(\phi_{j,i}-\phi_i)\Big|\leq\sup_{y'\in\mathcal{Y},\mathbf{x}'\in\mathcal{X}}\Big|\frac{1}{nh}\sum_{j=1}^{n}\Big[\left(\mathbb{1}\left(y'\leq y_j\right)-F(y_j|\mathbf{x}')\right)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\left(\frac{y_j-y}{h}\right)$$

$$-\int\left(\mathbb{1}\left(y'\leq u\right)-F(u|\mathbf{x}')\right)\mathbf{c}_1^{\mathsf{T}}\mathbf{P}\Big(\frac{u-y}{h}\Big)\mathrm{d}G(u)\Big]\Big|.$$

Then it is straightforward to show that

$$\mathbb{P}\Big[\max_i \sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \Big|\frac{1}{n}\sum_{j=1}^{n}(\phi_{j,i}-\phi_i)\Big| \geq \mathfrak{c}_1\sqrt{\frac{\log(n)}{nh}}\Big] \leq \mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

As a result,

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \Big|\frac{(\text{I.1.1})}{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}\Big| > \mathfrak{c}_1\frac{\log(n)}{nh^2}\Big] \leq \mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

To conclude the proof for $\hat{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})$, we note that replacing $\hat{\mathbf{c}}_1$ by $\mathbf{c}_1$ and $\hat{\mathbf{c}}_2$ by $\mathbf{c}_2$ only leads to an additional multiplicative factor $1+O_{\mathbb{P}}(\sqrt{\log(n)/nh^d})$. See Lemma 2.1.

7.10. Proof of Lemma 3.4

First consider $\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$. The difference between $\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$ and $\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$ is

$$\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) - \check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) = \left(\sqrt{\frac{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}-1\right)\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}).$$

From Lemma 3.3, we have

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \Big|\sqrt{\frac{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})}{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})}}-1\Big| > \mathfrak{c}_1\mathtt{r}_{\mathtt{VE}}\Big] \leq \mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

To close the proof, it is straightforward to verify that

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \big|\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})\big| > \mathfrak{c}_1\sqrt{\log(n)}\Big] \leq \mathfrak{c}_2 n^{-\mathfrak{c}_3},$$

which follows from the uniform convergence rate in Lemma 3.2. The same technique applies to the analysis of $\hat{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) - \hat{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x})$.

7.11. Proof of Theorem 3.1

See the proof of Theorem 2 in the main paper (i.e., Appendix A.6).

7.12. Proof of Lemma 3.5

Consider $\check{\rho}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')$. Note that we can decompose the difference into

$$\check{\rho}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}') - \rho_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}') = \frac{\check{\mathsf{C}}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\nu}(y',\mathbf{x}')}} - \frac{\mathsf{C}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})\mathsf{V}_{\mu,\nu}(y',\mathbf{x}')}}$$

$$= \underbrace{\frac{\check{\mathsf{C}}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}') - \mathsf{C}_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')}{\sqrt{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\nu}(y',\mathbf{x}')}}}_{\text{(I)}} + \underbrace{\rho_{\mu,\nu}(y,\mathbf{x},y',\mathbf{x}')\left(\sqrt{\frac{\mathsf{V}_{\mu,\nu}(y,\mathbf{x})\mathsf{V}_{\mu,\nu}(y',\mathbf{x}')}{\check{\mathsf{V}}_{\mu,\nu}(y,\mathbf{x})\check{\mathsf{V}}_{\mu,\nu}(y',\mathbf{x}')}} - 1\right)}_{\text{(II)}}.$$

The probabilistic order of the second term is given in Lemma 3.3.

Using similar techniques as in the proof of Lemma 2.1 or 3.3, it is also straightforward to verify that term (I) has the same order. That is,

$$\mathbb{P}\Big[\sup_{y,y'\in\mathcal{Y},\mathbf{x},\mathbf{x}'\in\mathcal{X}} |(\mathrm{I})| > \mathfrak{c}_1 \mathtt{r}_{\mathtt{VE}}\Big] \le \mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

7.13. Proof of Lemma 3.6

Consider an ε discretization of $\mathcal{Y}\times\mathcal{X}$, which is denoted by $\mathcal{A}_\varepsilon = \{(y_\ell, \mathbf{x}_\ell^\mathsf{T})^\mathsf{T} :\ 1\le \ell\le L\}$. Then one can define two Gaussian vectors, $\mathbf{z}, \check{\mathbf{z}} \in \mathbb{R}^L$, such that

$$\mathbb{C}\mathrm{ov}[z_\ell, z_{\ell'}] = \rho(y_\ell, \mathbf{x}_\ell, y_{\ell'}, \mathbf{x}_{\ell'}), \quad \mathbb{C}\mathrm{ov}[\check{z}_\ell, \check{z}_{\ell'}|\mathrm{Data}] = \check{\rho}(y_\ell, \mathbf{x}_\ell, y_{\ell'}, \mathbf{x}_{\ell'}).$$

Then we apply the Gaussian comparison result in Corollary 5.1 of [2] (Lemma 11 in the main paper) and the error rate in Lemma 3.5, which lead to

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big[\sup_{1\le\ell\le L}|\check{z}_\ell|\le u\Big|\mathbf{Y},\mathbf{X}\Big] - \mathbb{P}\Big[\sup_{1\le\ell\le L}|z_\ell|\le u\Big]\right| \\
&= \sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big[\sup_{1\le\ell\le L}|\check{\mathbb{G}}_{\mu,\nu}(y_\ell,\mathbf{x}_\ell)|\le u\Big|\mathbf{Y},\mathbf{X}\Big] - \mathbb{P}\Big[\sup_{1\le\ell\le L}|\mathbb{G}_{\mu,\nu}(y_\ell,\mathbf{x}_\ell)|\le u\Big]\right| \\
&\precsim_{\mathbb{P}} \left[h^{\frac{\mathfrak{q}}{2}} + \Big(\frac{\log(n)}{nh^{d+2}}\Big)^{\frac14}\right]\log(\frac{1}{\varepsilon}).
\end{aligned}$$

Since ε only enters the above error bound logarithmically, one can choose $\varepsilon = n^{-c}$ for some c large enough, so that the error that arises from discretization becomes negligible. The same applies to $\hat{\mathbb{G}}_{\mu,\nu}(y_\ell,\mathbf{x}_\ell)$.

7.14. Proof of Theorem 3.2

First consider $\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x})$. Since

$$\begin{aligned}
&\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})| - \sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x}) - \bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})| \le \sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x})| \\
&\le \sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})| + \sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x}) - \bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|,
\end{aligned}$$

then with Lemma 3.4,

$$\begin{aligned}
\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|\le u - \mathfrak{c}_1\mathtt{r}_{\mathtt{SE}}\Big] - \mathfrak{c}_2 n^{-\mathfrak{c}_3} &\le \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}^\circ_{\mu,\nu}(y,\mathbf{x})|\le u\Big] \\
&\le \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})|\le u + \mathfrak{c}_1\mathtt{r}_{\mathtt{SE}}\Big] + \mathfrak{c}_2 n^{-\mathfrak{c}_3}.
\end{aligned}$$

In the above, we also used the fact that the difference $\check{\mathbb{S}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) - \bar{\mathbb{S}}_{\mu,\nu}(y,\mathbf{x})$ is negligible compared to $\mathtt{r}_{\mathtt{SE}}$ (see Remark 2.3).

By applying Lemma 3.1,

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\mathbb{G}_{\mu,\nu}(y,\mathbf{x})| \le u - \mathfrak{c}_1(\mathtt{r}_{\mathtt{SE}} + \mathtt{r}_{\mathtt{SA}})\Big] - \mathfrak{c}_2 n^{-\mathfrak{c}_3} \le \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})| \le u\Big]$$
$$\le \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\mathbb{G}_{\mu,\nu}(y,\mathbf{x})| \le u + \mathfrak{c}_1(\mathtt{r}_{\mathtt{SE}} + \mathtt{r}_{\mathtt{SA}})\Big] + \mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

Finally, we apply the Gaussian comparison result in Lemma 3.6, which implies that

$$\sup_{u\in\mathbb{R}}\bigg|\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})| \le u\Big] - \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\check{\mathbb{G}}_{\mu,\nu}y,\mathbf{x}| \le u\Big|\mathbf{Y},\mathbf{X}\Big]\bigg|$$
$$\lesssim_{\mathbb{P}} \mathfrak{c}_2 n^{-\mathfrak{c}_3} + \log(n)\sqrt{\mathtt{r}_{\mathtt{VE}}} + \sup_{u\in\mathbb{R}} \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\mathbb{G}_{\mu,\nu}(y,\mathbf{x})| \in [u, u + \mathfrak{c}_1(\mathtt{r}_{\mathtt{SE}} + \mathtt{r}_{\mathtt{SA}})]\Big].$$

Finally, due to Theorem 2.1 of [1] (Lemma 12 in the main paper), we have

$$\sup_{u\in\mathbb{R}} \mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} |\mathbb{G}_{\mu,\nu}(y,\mathbf{x})| \in [u, u + \mathfrak{c}_1(\mathtt{r}_{\mathtt{SE}} + \mathtt{r}_{\mathtt{SA}})]\Big] \lesssim \sqrt{\log(n)}(\mathtt{r}_{\mathtt{SE}} + \mathtt{r}_{\mathtt{SA}}).$$

7.15. Proof of Theorem 4.1

Note that $\theta_{\mu,\nu}(y|\mathbf{x})$ falls into the confidence band $\check{C}_{\mu,\nu}(1-\alpha)$ if and only if

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\frac{\check{\theta}_{\mu,\nu}(y|\mathbf{x}) - \theta_{\mu,\nu}(y|\mathbf{x})}{\sqrt{\check{\mathbb{V}}_{\mu,\nu}(y,\mathbf{x})}}\right| \le \check{\mathtt{cv}}_{\mu,\nu}(\alpha).$$

A sufficient condition would then be

$$\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x})\right| + \sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}} \left|\frac{\mathbb{E}[\check{\theta}_{\mu,\nu}(y|\mathbf{x})|\mathbf{X}] - \theta_{\mu,\nu}(y|\mathbf{x})}{\sqrt{\check{\mathbb{V}}_{\mu,\nu}(y,\mathbf{x})}}\right| \le \check{\mathtt{cv}}_{\mu,\nu}(\alpha).$$

The conclusion then follows from Theorem 3.2 and the bias calculation in Lemma 2.2. The same analysis applies to $\hat{C}_{\mu,\nu}(1-\alpha)$.

7.16. Proof of Theorem 4.2

To start, we decompose the test statistic into

$$\check{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x}) = \check{\mathbb{T}}^{\circ}_{\mu,\nu}(y,\mathbf{x}) + \frac{\mathbb{E}[\check{\theta}_{\mu,\nu}(y|\mathbf{x})] - \theta_{\mu,\nu}(y|\mathbf{x})}{\sqrt{\check{\mathbb{V}}_{\mu,\nu}(y,\mathbf{x})}} + \frac{\theta_{\mu,\nu}(y|\mathbf{x}) - \theta_{\mu,\nu}(y|\mathbf{x};\hat{\boldsymbol{\gamma}})}{\sqrt{\check{\mathbb{V}}_{\mu,\nu}(y,\mathbf{x})}}.$$

Then by the leading bias order in Lemma 2.2 and the leading variance order in Lemma 2.3, we have that

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\mathbb{E}[\check{\theta}_{\mu,\nu}(y|\mathbf{x})]-\theta_{\mu,\nu}(y|\mathbf{x})}{\sqrt{\check{\mathbb{V}}_{\mu,\nu}(y,\mathbf{x})}}\Big|>\mathfrak{c}_1\frac{\mathtt{r}_{\mathtt{B}}}{\mathtt{r}_{\mathtt{V}}}(1+\mathtt{r}_{\mathtt{VE}})\Big]\leq\mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

Similarly, under the null hypothesis,

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}\Big|\frac{\theta_{\mu,\nu}(y|\mathbf{x})-\theta_{\mu,\nu}(y|\mathbf{x};\hat{\boldsymbol{\gamma}})}{\sqrt{\check{\mathbb{V}}_{\mu,\nu}(y,\mathbf{x})}}\Big|>\mathfrak{c}_1\frac{\mathtt{r}_{\mathtt{PS}}}{\mathtt{r}_{\mathtt{V}}}(1+\mathtt{r}_{\mathtt{VE}})\Big]\leq\mathfrak{c}_2 n^{-\mathfrak{c}_3}.$$

Then we have the following error bound

$$\begin{aligned}&\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x})|\leq u\Big]-\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{G}}_{\mu,\nu}(y,\mathbf{x})|\leq u\Big|\mathbf{Y},\mathbf{X}\Big]\Big|\\&\lesssim_{\mathbb{P}}\sqrt{\log(n)}\left(\mathtt{r}_{\mathtt{SE}}+\mathtt{r}_{\mathtt{SA}}+\frac{\mathtt{r}_{\mathtt{B}}+\mathtt{r}_{\mathtt{PS}}}{\mathtt{r}_{\mathtt{V}}}\right)+(\log(n))\sqrt{\mathtt{r}_{\mathtt{VE}}}.\end{aligned}$$

As a result,

$$\mathbb{P}\Big[\sup_{y\in\mathcal{Y},\mathbf{x}\in\mathcal{X}}|\check{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x})|>\mathtt{cv}_{\mu,\nu}(\alpha)\Big]\leq\alpha+\mathfrak{c}\left(\sqrt{\log(n)}\left(\mathtt{r}_{\mathtt{SE}}+\mathtt{r}_{\mathtt{SA}}+\frac{\mathtt{r}_{\mathtt{B}}+\mathtt{r}_{\mathtt{PS}}}{\mathtt{r}_{\mathtt{V}}}\right)+(\log(n))\sqrt{\mathtt{r}_{\mathtt{VE}}}\right).$$

The same strategy can be employed to establish results for $\hat{\mathbb{T}}_{\mathtt{PS}}(y,\mathbf{x})$.

7.17. Proof of Theorem 4.3

The conclusion follows directly from Theorem 4.1.

Funding

Cattaneo gratefully acknowledges financial support from the National Science Foundation through grants SES-1947805 and DMS-2210561, and from the National Institute of Health (R01 GM072611-16).

Jansson gratefully acknowledges financial support from the National Science Foundation through grant SES-1947662 and the research support of CREATES.

References

[1] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Stat.* **42** 1564–1597.

[2] CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Stat.* **50** 2562–2586.

[3] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.

[4] GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability II* Springer.