Supplement to "Kernel-Based Semiparametric Estimators: Small Bandwidth Asymptotics and Bootstrap Consistency"*

Matias D. Cattaneo† Michael Jansson‡

February 6, 2018

Abstract

This supplement includes additional results not reported in the main paper to conserve space. First, it discusses in detail the examples of semiparametric estimators analyzed in the main paper, and also introduces and discusses a new example of interest: 'Hit Rate', which involves a non-differentiable functional of the nonparametric component and is briefly mentioned in the simulation section of the main paper. Second, it reports a technical lemma useful to handle kernel-based nonparametric estimators, which may be of independent interest.

*The first author gratefully acknowledges financial support from the National Science Foundation (SES 1122994 and SES 1459931). The second author gratefully acknowledges financial support from the National Science Foundation (SES 1124174 and SES 1459967) and the research support of CREATES (funded by the Danish National Research Foundation under grant no. DNRF78).

†Department of Economics and Department of Statistics, University of Michigan.

‡Department of Economics, UC Berkeley and *CREATES*.

Contents

SA.1 Example 1: Average Density 2

SA.1.1 Average Density Estimator . . . 3

SA.1.2 Integrated Square Density Estimator . . . 6

SA.1.3 Locally Robust Estimator . . . 11

SA.2 Example 2: Inverse Probability Weighting 14

SA.2.1 Condition AL . . . 18

SA.2.2 Condition AS . . . 21

SA.2.3 Condition AN . . . 22

SA.2.4 Bandwidth Selection . . . 27

SA.2.5 Condition AL* . . . 27

SA.2.6 Condition AS* . . . 29

SA.2.7 Condition AN* . . . 30

SA.3 Example 3: Hit Rate 31

SA.3.1 Condition AL . . . 32

SA.3.2 Condition AS . . . 33

SA.3.3 Condition AN . . . 34

SA.3.4 Bandwidth Selection . . . 35

SA.3.5 Condition AL* . . . 35

SA.3.6 Condition AS* . . . 35

SA.3.7 Condition AN* . . . 36

SA.4 Uniform Convergence Rates for Kernel-Based Estimators 37

SA.4.1 Proof of Lemma SA-1 . . . 38

SA.1 Example 1: Average Density

This section considers the estimand

$$\theta_0 = \mathbb{E}[\gamma_0(z)] = \int_{\mathbb{R}^d} \gamma_0(u)^2 du,$$

where γ_0 denotes the Lebesgue density of a random vector $z \in \mathbb{R}^d$. In many respects this can be seen as the simplest possible semiparametric problem; that is, it can be viewed as a semiparametric analogue of the problem of estimating the mean of a distribution in parametric mathematical statistics. For our purposes, the example is attractive because it provides a straightforward illustration of several interesting features of semiparametric estimators, as already mentioned in the main text.

Suppose $z_1, \ldots, z_n$ are *i.i.d.* copies of z. We consider three distinct estimators, each of which employs the kernel-based density estimator

$$\hat{\gamma}_n(z) = \frac{1}{n}\sum_{j=1}^{n} K_n(z - z_j), \qquad K_n(u) = \frac{1}{h_n^d} K\left(\frac{u}{h_n}\right),$$

where K is a kernel and h_n is a bandwidth. The estimators considered are: (i) the average density estimator $\hat{\theta}_n^{\texttt{AD}} = n^{-1}\sum_{i=1}^{n} \hat{\gamma}_n(z_i)$; (ii) the integrated square density estimator $\hat{\theta}_n^{\texttt{ISD}} = \int_{\mathbb{R}^d} \hat{\gamma}_n(u)^2 du$; and (iii) the "locally robust" estimator $\hat{\theta}_n^{\texttt{LR}} = 2\hat{\theta}_n^{\texttt{AD}} - \hat{\theta}_n^{\texttt{ISD}}$.

To obtain primitive bandwidth conditions for the high-level conditions of Theorems 1 and 2, suppose that for some $P > d/2$, the following regularity conditions are satisfied:

- γ_0 is $P+1$ times differentiable, and γ_0 and its first $P+1$ derivatives are bounded, continuous and square integrable.
- K is even and bounded with $\int_{\mathbb{R}^d} |K(u)| (1 + \|u\|^{P+1}) du < \infty$ and

$$\int_{\mathbb{R}^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du = \begin{cases} 1, & \text{if } l_1 = \cdots = l_d = 0, \\ 0, & \text{if } (l_1, \ldots, l_d)' \in \mathbb{Z}_+^d \text{ and } l_1 + \cdots + l_d < P \end{cases}.$$

The smoothness assumption on γ_0 can be relaxed substantially (e.g., Giné and Nickl (2008) and the references therein), but the stated assumption is sufficient for our purposes.

As discussed in the paper, this example is used to illustrate three main findings. First, $\hat{\theta}_n^{\texttt{AD}}$ sheds light on the Stochastic Equicontinuity condition featuring prominently in existing "master theorems". To be specific, $\hat{\theta}_n^{\texttt{AD}}$ illustrates the consequences of relaxing the Stochastic Equicontinuity condition and shows how the weaker Asymptotic Separability condition is useful to that end; in the case of $\hat{\theta}_n^{\texttt{AD}}$, $\mathcal{B}_n^{\texttt{LI}} \neq 0$ and $\mathcal{B}_n^{\texttt{NL}} = 0$. Second, $\hat{\theta}_n^{\texttt{ISD}}$ shows that changing the form of the estimating equation can have important implications for small bandwidth biases; in the case of $\hat{\theta}_n^{\texttt{ISD}}$, $\mathcal{B}_n^{\texttt{LI}} = 0$ but $\mathcal{B}_n^{\texttt{NL}} \neq 0$. Finally, $\hat{\theta}_n^{\texttt{LR}}$ shows that "locally robust" estimators are not necessarily robust to small bandwidths; in the case of $\hat{\theta}_n^{\texttt{LR}}$, $\mathcal{B}_n^{\texttt{LI}} \neq 0$ and $\mathcal{B}_n^{\texttt{NL}} \neq 0$.

SA.1.1 Average Density Estimator

When verifying the conditions of Theorems 1 and 2 for $\hat{\theta}_n^{\texttt{AD}}$, we set $d_\gamma = 1$, $x(z,\theta) = z$, $w(z,\theta) = 1$, and let $\hat{\theta}_n^{\texttt{AD}}$ be defined by $\hat{G}_n(\hat{\theta}_n^{\texttt{AD}}, \hat{\gamma}_n) = 0$, where $g(z,\theta,\gamma) = g^{\texttt{AD}}(z,\theta,\gamma) = \gamma(z) - \theta$ is a linear functional of γ.

Because

$$\hat{\theta}_n^{\texttt{AD}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}_n(z_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_n(z_i - z_j),$$

the estimator can be represented as a second-order V-statistic and is therefore very tractable. For this reason (and others), the estimator has been widely studied. We include it here in part because it provides a dramatic demonstration of the fragility of Stochastic Equicontinuity with respect to bandwidth choice. It also illustrates how to verify sufficient conditions, and their relationship to necessary conditions, in a very simple and transparent case.

If the bandwidth satisfies $nh_n^{2P} \to 0$ and $nh_n^d \to \infty$, we show here that the assumptions of Theorems 1 and 2 are satisfied and that

$$\sqrt{n}(\hat{\theta}_n^{\texttt{AD}} - \theta_0 - \mathfrak{B}_n^{\texttt{AD}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0), \qquad \Sigma_0 = 4\mathbb{V}[\gamma_0(z)],$$

with

$$\mathfrak{B}_n^{\texttt{AD}} = \frac{1}{nh_n^d} K(0).$$

Because $\sqrt{n}\mathfrak{B}_n^{\texttt{AD}} = K(0)/\sqrt{nh_n^{2d}}$, the condition $nh_n^d \to \infty$ is weak enough to permit $\mathfrak{B}_n^{\texttt{AD}} \neq o(n^{-1/2})$. On the other hand, $\sqrt{n}(\hat{\theta}_n^{\texttt{AD}} - \theta_0 - \mathfrak{B}_n^{\texttt{AD}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ reduces to $\sqrt{n}(\hat{\theta}_n^{\texttt{AD}} - \theta_0) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ when imposing conditions requiring $nh_n^{2d} \to \infty$, so it is necessary to guard against this when the goal is to obtain the more refined results given by Theorems 1 and 2.

SA.1.1.1 Condition AL

Condition AL holds with $\mathcal{J}_n = \mathcal{J}_0 = 1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.1.1.2 Condition AS

Because

$$g_n(z,\gamma) = g_n(z,\gamma_n) + g_{n,\gamma}(z)[\gamma - \gamma_n], \qquad g_{n,\gamma}(z)[\eta] = (1 - n^{-1})\eta(z),$$

Condition AS holds with $\bar{g}_n = g_n$ if $\mathbb{V}(g_{n,\gamma}(z_i)[\hat{\gamma}_n^j]) = o(n)$ whenever i and j are distinct. More precisely, the first part of Condition AS is automatically satisfied, and the second part becomes

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} [\hat{\gamma}_n^{(i)}(z_i) - 2\gamma_n(z_i) + \theta_n] = o_{\mathbb{P}}(1),$$

where

$$\hat{\gamma}_n^{(i)}(z) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} K_n(z-z_j) \qquad \text{and} \qquad \theta_n = \int_{\mathbb{R}^d} \gamma_n(u)\gamma_0(u)du.$$

A simple variance calculation now shows that Condition AS is satisfied if $nh_n^d \to \infty$, because then

$$\mathbb{V}(g_{n,\gamma}(z_i)[\hat{\gamma}_n^j]) = (1-n^{-1})^2\mathbb{V}[K_n(z_i - z_j) - \gamma_n(z_j)] = O(h_n^{-d}) = o(n)$$

whenever i and j are distinct.

SA.1.1.3 Condition SE

If $h_n \to 0$ and if $nh_n^d \to \infty$, then, using Condition AS,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\hat{\gamma}_n(z_i) - \gamma_n(z_i) - \gamma_0(z_i) + \theta_0] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[n^{-1}K_n(0) + (1-n^{-1})\hat{\gamma}_n^{(i)}(z_i) - \gamma_n(z_i) - \gamma_0(z_i) + \theta_0] \\
&= \frac{K(0)}{\sqrt{nh_n^{2d}}} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[(1-n^{-1})(2\gamma_n(z_i) - \theta_n) - \gamma_n(z_i) - \gamma_0(z_i) + \theta_0] + o_{\mathbb{P}}(1) \\
&= \frac{K(0)}{\sqrt{nh_n^{2d}}} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\gamma_n(z_i) - \gamma_0(z_i) - (\theta_n - \theta_0)] + o_{\mathbb{P}}(1) = \frac{K(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1),
\end{aligned}$$

where the last equality uses $\mathbb{E}(|\gamma_n(z) - \gamma_0(z) - (\theta_n - \theta_0)|^2) = o(1)$.

As a consequence, Stochastic Equicontinuity requires $nh_n^{2d} \to \infty$ in this example; that is, because the calculations are based on an exact decomposition we are able to give necessary conditions for Condition SE.

SA.1.1.4 Condition AN

As in the paper, we have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\gamma_n) + \bar{G}_n(\hat{\gamma}_n^{(i)}) - \bar{G}_n(\gamma_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n(z_i) + \sqrt{n}\hat{\mathcal{B}}_n,$$

where

$$\psi_n(z) = 2[\gamma_n^+(z) - \theta_n^+], \qquad \hat{\mathcal{B}}_n = \theta_n^+ - \theta_0,$$

$$\gamma_n^+(z) = n^{-1}K_n(0) + (1-n^{-1})\gamma_n(z), \qquad \theta_n^+ = \int_{\mathbb{R}^d}\gamma_n^+(u)\gamma_0(u)du.$$

Suppose $h_n \to 0$. Then $\psi_n(z) \to \psi(z)$ for every z, and, by the dominated convergence theorem,

$$\mathbb{E}(|\psi_n(z) - \psi_0(z)|^2) \to 0, \qquad \psi_0(z) = 2[\gamma_0(z) - \theta_0].$$

Therefore, it follows from Lemma 3 that Condition AN is satisfied with $\Omega_0 = 4\mathbb{V}[\gamma_0(z)]$.

Furthermore, $\mathbb{E}\hat{\mathcal{B}}_n = \mathcal{B}_n^{\mathtt{S}} + \mathcal{B}_n^{\mathtt{LI}} + \mathcal{B}_n^{\mathtt{NL}}$, where

$$\mathcal{B}_n^{\mathtt{NL}} = 0, \qquad \mathcal{B}_n^{\mathtt{LI}} = G_n(\gamma_n) - G_0(\gamma_n) = \frac{1}{nh_n^d}K(0) + O(n^{-1}),$$

and

$$\mathcal{B}_n^{\mathtt{S}} = G_0(\gamma_n) = h_n^P \mathcal{B}_0^{\mathtt{S}} + o(h_n^P),$$

$$\mathcal{B}_0^{\mathtt{S}} = (-1)^P \sum_{|p|=P} \frac{1}{p!}\left(\int_{\mathbb{R}^d} u^p K(u)du\right)\left(\int_{\mathbb{R}^d} \gamma_0(u)\left(\partial^p \gamma_0(u)\right)du\right),$$

where, for $p = (p_1, \ldots p_d)' \in \mathbb{Z}_+^d$, the definition of $\mathcal{B}_0^{\mathtt{S}}$ uses the multi-index notation

$$|p| = p_1 + \cdots + p_d, \qquad p! = p_1! \cdots p_d!, \qquad u^p = u_1^{p_1} \cdots u_d^{p_d}, \qquad \partial^p = \frac{\partial^{|p|}}{\partial^{p_1} \cdots \partial^{p_d}}.$$

As a consequence, we can set $\mathcal{B}_n = K(0)/(nh_n^d)$ provided that $nh_n^{2P} \to 0$.

In summary, if $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then the conditions of Theorem 1 are satisfied and $\sqrt{n}(\hat{\theta}_n^{\mathtt{AD}} - \theta_0 - \mathfrak{B}_n^{\mathtt{AD}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ holds with $\Sigma_0 = 4\mathbb{V}[\gamma_0(z)]$ and $\mathfrak{B}_n^{\mathtt{AD}} = O(n^{-1}h_n^{-d})$.

SA.1.1.5 Bandwidth Selection

Assuming $\mathcal{B}_0^{\mathtt{SB}} \neq 0$ and $\mathcal{B}_0^{\mathtt{S}} \neq 0$, we can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

$$h_{\mathtt{opt}} = \begin{cases} \left(\frac{|\mathcal{B}_0^{\mathtt{SB}}|}{|\mathcal{B}_0^{\mathtt{S}}|}\frac{1}{n}\right)^{\frac{1}{P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}}) \neq \operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \\ \left(\frac{d}{P}\frac{|\mathcal{B}_0^{\mathtt{SB}}|}{|\mathcal{B}_0^{\mathtt{S}}|}\frac{1}{n}\right)^{\frac{1}{P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}}) = \operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \end{cases}, \qquad \mathcal{B}_0^{\mathtt{SB}} = K(0).$$

SA.1.1.6 Condition AL*

Condition AL* holds with $\mathcal{J}_n^* = \mathcal{J}_0^* = 1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.1.1.7 Condition AS*

Because

$$g_n^*(z,\gamma) = g_n^*(z,\hat{\gamma}_n) + g_{n,\gamma}^*(z)[\gamma - \hat{\gamma}_n], \qquad g_{n,\gamma}^*(z)[\eta] = (1-n^{-1})\eta(z),$$

Condition AS* holds with $\bar{g}_n^* = g_n^*$ if $\mathbb{V}^*(g_{n,\gamma}^*(z_{i,n}^*)[\hat{\gamma}_n^{*,j}]) = o_{\mathbb{P}}(n)$ whenever i and j are distinct. A sufficient condition for this to occur is that $nh_n^d \to \infty$, because then

$$\mathbb{E}\mathbb{V}^*(g_{n,\gamma}^*(z_{i,n}^*)[\hat{\gamma}_n^{*,j}]) = (1-n^{-1})^2 \mathbb{E}\mathbb{V}^*[\mathcal{K}_n(z_{i,n}^* - z_{j,n}^*) - \hat{\gamma}_n(z_{j,n}^*)] = O(h_n^{-d}) = o(n)$$

whenever i and j are distinct.

SA.1.1.8 Condition AN*

We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_{i,n}^*,\hat{\gamma}_n)+\bar{G}_n^*(\hat{\gamma}_n^{*,(i)})-\bar{G}_n^*(\hat{\gamma}_n)]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n^*(z_{i,n}^*)+\sqrt{n}\hat{\mathcal{B}}_n^*,$$

where

$$\psi_n^*(z_{i,n}^*)=2[\hat{\gamma}_n^+(z_{i,n}^*)-\hat{\theta}_n^+],\qquad \hat{\mathcal{B}}_n^*=n^{-1}K_n(0)-n^{-1}\hat{\theta}_n,$$

$$\hat{\gamma}_n^+(z_{i,n}^*)=n^{-1}K_n(0)+(1-n^{-1})\hat{\gamma}_n(z_{i,n}^*),\qquad \hat{\theta}_n^+=\mathbb{E}^*\hat{\gamma}_n^+(z_{i,n}^*).$$

Suppose $h_n\to 0$ and $nh_n^d\to\infty$. Then $\hat{\mathcal{B}}_n^*=K(0)/(nh_n^d)+o_{\mathbb{P}}(n^{-1/2})$ because $\hat{\theta}_n=O_{\mathbb{P}}(1)$. Moreover, because $\hat{\theta}_n-\theta_n\to_{\mathbb{P}}0$, $\mathbb{E}^*[|\psi_n^*(z_{i,n}^*)-\psi_n(z_{i,n}^*)|^2]=o_{\mathbb{P}}(1)$ also holds provided

$$\frac{1}{n}\sum_{i=1}^{n}|\hat{\gamma}_n(z_i)-\gamma_n(z_i)|^2=o_{\mathbb{P}}(1).$$

A sufficient condition for this to occur is that $\max_{1\le i\le n}|\hat{\gamma}_n(z_i)-\gamma_n(z_i)|=o_{\mathbb{P}}(1)$, which in turn will hold if $nh_n^d/\log n\to\infty$, as can be established using Lemma SA-1 below. Sufficiency of the slightly weaker condition $nh_n^d\to\infty$ can be demonstrated by using a direct calculation to show that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}|\hat{\gamma}_n(z_i)-\gamma_n(z_i)|^2\right]=O(n^{-1}h_n^{-d}).$$

In other words, if $h_n\to 0$ and if $nh_n^d\to\infty$, then Condition AN* holds with $\Omega_0^*=\Omega_0$ and $\mathcal{B}_n^*=K(0)/(nh_n^d)$.

In summary, if $nh_n^{2P}\to 0$ and if $nh_n^d\to\infty$, then the conditions of Theorem 2 are satisfied.

SA.1.2 Integrated Square Density Estimator

Like $\hat{\theta}_n^{\mathtt{AD}}$, the estimator $\hat{\theta}_n^{\mathtt{ISD}}$ can be represented as a second-order V-statistic:

$$\begin{aligned}\hat{\theta}_n^{\mathtt{ISD}} &= \int_{\mathbb{R}^d}\hat{\gamma}_n(u)^2du=\int_{\mathbb{R}^d}\left(\frac{1}{n}\sum_{j=1}^{n}K_n(u-x_j)\right)^2du=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{\mathbb{R}^d}K_n(u-x_i)K_n(u-x_j)du\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{K}_n(x_i-x_j),\end{aligned}$$

where

$$\bar{K}_n(v)=\frac{1}{h_n^d}\int_{\mathbb{R}^d}K(u)K\left(\frac{v}{h_n}-u\right)du.$$

For our purposes, however, it is more attractive to analyze $\hat{\theta}_n^{\mathtt{ISD}}$ with the help of Theorems 1 and

2. To do so, we set $d_\gamma = 1$, $x(z,\theta) = z$, $w(z,\theta) = 1$, and let $\hat{\theta}_n^{\texttt{ISD}}$ be defined by $\hat{G}_n(\hat{\theta}_n^{\texttt{ISD}}, \hat{\gamma}_n) = 0$, where $g(z,\theta,\gamma) = g^{\texttt{ISD}}(z,\theta,\gamma) = \int_{\mathbb{R}^d} \gamma(u)^2 du - \theta$ is a non-linear functional of γ.

If the bandwidth satisfies $nh_n^{2P} \to 0$ and $nh_n^d \to \infty$, we show here that the assumptions of Theorems 1 and 2 are satisfied and that

$$\sqrt{n}(\hat{\theta}_n^{\texttt{ISD}} - \theta_0 - \mathfrak{B}_n^{\texttt{ISD}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0), \qquad \Sigma_0 = 4\mathbb{V}[\gamma_0(z)],$$

with

$$\mathfrak{B}_n^{\texttt{ISD}} = \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n) du dv = O(n^{-1}h_n^{-d}).$$

As in the case of $\hat{\theta}_n^{\texttt{AD}}$, the condition $nh_n^d \to \infty$ is weak enough to permit $\mathfrak{B}_n^{\texttt{ISD}} \neq o(n^{-1/2})$, while $\sqrt{n}(\hat{\theta}_n^{\texttt{ISD}} - \theta_0 - \mathfrak{B}_n^{\texttt{ISD}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ reduces to $\sqrt{n}(\hat{\theta}_n^{\texttt{ISD}} - \theta_0) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ when imposing conditions requiring $nh_n^{2d} \to \infty$, so once again it is necessary to guard against this when the goal is to obtain the more refined results given by Theorem 1 and 2.

SA.1.2.1 Condition AL

Condition AL holds with $\mathcal{J}_n = \mathcal{J}_0 = 1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.1.2.2 Condition AS

Let a quadratic approximation to g_n be given by

$$\bar{g}_n(z,\gamma) = g_n(z,\gamma_n) + g_{n,\gamma}(z)[\gamma - \gamma_n] + \frac{1}{2} g_{n,\gamma\gamma}[\gamma - \gamma_n, \gamma - \gamma_n],$$

where

$$g_n(z,\gamma_n) = \int_{\mathbb{R}^d} [n^{-1}K_n(u-z) + (1-n^{-1})\gamma_n(u)]^2 du - \theta_0,$$

$$g_{n,\gamma}(z)[\eta] = 2(1-n^{-1}) \int_{\mathbb{R}^d} [n^{-1}K_n(u-z) + (1-n^{-1})\gamma_n(u)]\eta(u) du,$$

$$g_{n,\gamma\gamma}[\eta, \varphi] = 2(1-n^{-1})^2 \int_{\mathbb{R}^d} \eta(u)\varphi(u) du.$$

The first part of Condition AS holds directly, without any remainder term, because the quadratic approximation is exact. The second part of Condition AS follows from Lemma 2 if $nh_n^d \to \infty$ because simple variance calculations show that if i, j, and k are distinct, then

$$\mathbb{V}[g_{n,\gamma}(z_i)[\hat{\gamma}_n^j - \gamma_n]] = O(h_n^{-d}) = o(n), \qquad \mathbb{V}(g_{n,\gamma\gamma}[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^k - \gamma_n]) = O(h_n^{-2d}) = o(n^2),$$

$$\mathbb{V}(\mathbb{E}(g_{n,\gamma\gamma}[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n] | z_i)) = O(h_n^{-2d}) = o(n^2), \qquad \mathbb{V}(g_{n,\gamma\gamma}[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n]) = O(h_n^{-3d}) = o(n^3).$$

SA.1.2.3 Condition AN

We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\gamma_n)+\bar{G}_n(\hat{\gamma}_n^{(i)})-\bar{G}_n(\gamma_n)]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n(z_i)+\sqrt{n}\hat{\mathcal{B}}_n,$$

where

$$\bar{G}_n(\gamma)=G_n(\gamma_n)+G_{n,\gamma}[\gamma-\gamma_n]+\frac{1}{2}G_{n,\gamma\gamma}[\gamma-\gamma_n,\gamma-\gamma_n],$$

$$G_n(\gamma_n)=\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}[n^{-1}K_n(u-v)+(1-n^{-1})\gamma_n(u)]^2du\gamma_0(v)dv-\theta_0,$$

$$\begin{aligned}G_{n,\gamma}[\eta] &= 2(1-n^{-1})\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}[n^{-1}K_n(u-z)+(1-n^{-1})\gamma_n(u)]\eta(u)du\gamma_0(v)dv\\ &= 2(1-n^{-1})\int_{\mathbb{R}^d}\gamma_n(u)\eta(u)du,\end{aligned}$$

$$G_{n,\gamma\gamma}[\eta,\varphi]=2(1-n^{-1})^2\int_{\mathbb{R}^d}\eta(u)\varphi(u)du,$$

and

$$\psi_n(z)=g_n(z,\gamma_n)-G_n(\gamma_n)+\delta_n(z),\qquad \delta_n(z)=2(1-n^{-1})\int_{\mathbb{R}^d}\gamma_n(u)[K_n(u-z)-\gamma_n(u)]du,$$

$$\hat{\mathcal{B}}_n=G_n(\gamma_n)+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}G_{n,\gamma\gamma}[\hat{\gamma}_n^{(i)}-\gamma_n,\hat{\gamma}_n^{(i)}-\gamma_n].$$

Suppose $h_n\to 0$ and $nh_n^d\to\infty$. Then, for every z,

$$\begin{aligned}g_n(z,\gamma_n) &= \int_{\mathbb{R}^d}[n^{-1}K_n(u-z)+(1-n^{-1})\gamma_n(u)]^2du-\theta_0\\ &= n^{-2}\int_{\mathbb{R}^d}K_n(u-z)^2du+2n^{-1}(1-n^{-1})\int_{\mathbb{R}^d}K_n(u-z)\gamma_n(u)du\\ &\quad +(1-n^{-1})^2\int_{\mathbb{R}^d}\gamma_n(u)^2du-\theta_0\\ &= O(n^{-2}h_n^{-d}+n^{-1}+h_n^P)\to 0\end{aligned}$$

and hence

$$\psi_n(z)\to\psi_0(z)=\delta_0(z),\qquad \delta_0(z)=2[\gamma_0(z)-\theta_0].$$

Therefore, $\mathbb{E}(|\psi_n(z)-\psi_0(z)|^2)\to 0$ by the dominated convergence theorem. Moreover, if i and j are distinct, then

$$\mathbb{V}(G_{n,\gamma\gamma}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n])=O(h_n^{-2d})=o(n^2),\qquad \mathbb{V}(G_{n,\gamma\gamma}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^j-\gamma_n])=O(h_n^{-d})=o(n),$$

so it follows from Lemma 3 that Condition AN is satisfied with $\Omega_0=4\mathbb{V}[\gamma_0(z)]$.

Finally, consider the biases $\mathcal{B}_n^{\mathtt{S}}, \mathcal{B}_n^{\mathtt{LI}}$, and $\mathcal{B}_n^{\mathtt{NL}}$. In this case,

$$\mathcal{B}_n^{\mathtt{S}} = h_n^P \mathcal{B}_0^{\mathtt{S}} + o(h_n^P) + O(n^{-1})$$

with

$$\mathcal{B}_0^{\mathtt{S}} = 2(-1)^P \sum_{|p|=P} \frac{1}{p!} \left(\int_{\mathbb{R}^d} u^p K(u) du \right) \left(\int_{\mathbb{R}^d} \gamma_0(u) \left(\partial^p \gamma_0(u)\right) du \right);$$

that is, to the order h_n^P the smoothing bias of $\hat{\theta}_n^{\mathtt{ISD}}$ is twice that of $\hat{\theta}_n^{\mathtt{AD}}$. In addition, $\mathcal{B}_n^{\mathtt{LI}} = 0$ and

$$\begin{aligned} \mathcal{B}_n^{\mathtt{NL}} &= \frac{1}{2n} \mathbb{E} G_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] = \frac{1}{2n} \mathbb{E} G_{n,\gamma\gamma}[\hat{\gamma}_n^i, \hat{\gamma}_n^i] + O(n^{-1}) \\ &= \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n) du dv + O(n^{-1} + n^{-2} h_n^{-d}), \end{aligned}$$

where the second equality uses $\mathbb{E} G_{n,\gamma\gamma}[\hat{\gamma}_n^i, \gamma_n] = O(1)$ and $G_{n,\gamma\gamma}[\gamma_n, \gamma_n] = O(1)$, and the last equality uses

$$\mathbb{E} G_{n,\gamma\gamma}[\hat{\gamma}_n^i, \hat{\gamma}_n^i] = 2(1 - n^{-1})^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{1}{h_n^{2d}} K \left(\frac{u - v}{h_n} \right)^2 du \gamma_0(v) dv.$$

As a consequence, we can set

$$\begin{aligned} \mathcal{B}_n &= \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n) du dv \\ &= \frac{1}{nh_n^d} \int_{\mathbb{R}^d} K(u)^2 du + o(n^{-1} h_n^{-d}), \end{aligned}$$

provided that $nh_n^{2P} \to 0$.

In summary, if $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then the conditions of Theorem 1 are satisfied and $\sqrt{n}(\hat{\theta}_n^{\mathtt{ISD}} - \theta_0 - \mathfrak{B}_n^{\mathtt{ISD}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ holds with $\Sigma_0 = 4\mathbb{V}[\gamma_0(z)]$ and $\mathfrak{B}_n^{\mathtt{ISD}} = O(n^{-1} h_n^{-d})$.

SA.1.2.4 Bandwidth Selection

Assuming $\mathcal{B}_0^{\mathtt{SB}} \neq 0$ and $\mathcal{B}_0^{\mathtt{S}} \neq 0$, we can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

$$h_{\mathtt{opt}} = \begin{cases} \left(\frac{|\mathcal{B}_0^{\mathtt{SB}}|}{|\mathcal{B}_0^{\mathtt{S}}|} \frac{1}{n} \right)^{\frac{1}{P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}}) \neq \operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \\ \left(\frac{d}{P} \frac{|\mathcal{B}_0^{\mathtt{SB}}|}{|\mathcal{B}_0^{\mathtt{S}}|} \frac{1}{n} \right)^{\frac{1}{P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}}) = \operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \end{cases}, \qquad \mathcal{B}_0^{\mathtt{SB}} = \int_{\mathbb{R}^d} K(u)^2 du.$$

SA.1.2.5 Condition AL*

Condition AL* holds with $\mathcal{J}_n^* = \mathcal{J}_0^* = 1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.1.2.6 Condition AS*

Define the (exact) quadratic approximation

$$\bar{g}_n^*(z,\gamma) = g_n^*(z,\hat{\gamma}_n) + g_{n,\gamma}^*(z)[\gamma - \hat{\gamma}_n] + \frac{1}{2}g_{n,\gamma\gamma}^*[\gamma - \hat{\gamma}_n, \gamma - \hat{\gamma}_n],$$

with

$$g_n^*(z,\hat{\gamma}_n) = \int_{\mathbb{R}^d} [n^{-1}K_n(u-z) + (1-n^{-1})\hat{\gamma}_n(u)]^2 du - \hat{\theta}_n,$$

$$g_{n,\gamma}^*(z)[\eta] = 2(1-n^{-1}) \int_{\mathbb{R}^d} [n^{-1}K_n(u-z) + (1-n^{-1})\hat{\gamma}_n(u)]\eta(u) du,$$

$$g_{n,\gamma\gamma}^*[\eta,\varphi] = 2(1-n^{-1})^2 \int_{\mathbb{R}^d} \eta(u)\varphi(u) du.$$

Condition AS* holds if $nh_n^d \to \infty$, because then the conditions of Lemma 5 hold: If i, j, and k are distinct, then

$$\mathbb{V}^*[g_{n,\gamma}^*(z_{i,n}^*)[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n]] = O_{\mathbb{P}}(h_n^{-d}) = o_{\mathbb{P}}(n),$$

$$\mathbb{V}^*(g_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,k} - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-2d}) = o_{\mathbb{P}}(n^2),$$

$$\mathbb{V}^*(\mathbb{E}^*(g_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,j} - \hat{\gamma}_n]|z_{i,n}^*)) = O_{\mathbb{P}}(h_n^{-2d}) = o_{\mathbb{P}}(n^2),$$

$$\mathbb{V}^*(g_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,j} - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-3d}) = o_{\mathbb{P}}(n^3).$$

SA.1.2.7 Condition AN*

We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n [g_n^*(z_{i,n}^*,\hat{\gamma}_n) + \bar{G}_n^*(\hat{\gamma}_n^{*,(i)}) - \bar{G}_n^*(\hat{\gamma}_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_n^*(z_{i,n}^*) + \sqrt{n}\hat{\mathcal{B}}_n^*,$$

where

$$\psi_n^*(z_{i,n}^*) = g_n^*(z_{i,n}^*,\hat{\gamma}_n) - G_n^*(\hat{\gamma}_n) + \delta_n^*(z_{i,n}^*) = \delta_n^*(z_{i,n}^*),$$

$$\delta_n^*(z_{i,n}^*) = 2\int_{\mathbb{R}^d} \hat{\gamma}_n(u)[K_n(u - z_{i,n}^*) - \hat{\gamma}_n(u)] du,$$

$$\hat{\mathcal{B}}_n^* = G_n^*(\hat{\gamma}_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^n G_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,(i)} - \hat{\gamma}_n, \hat{\gamma}_n^{*,(i)} - \hat{\gamma}_n].$$

Assuming $h_n \to 0$ and $nh_n^d \to \infty$, the assumptions of Lemma 6 are satisfied. In particular,

$$\mathbb{E}^*[|\psi_n^*(z_{i,n}^*) - \psi_n(z_{i,n}^*)|^2] = \mathbb{E}^*[|\delta_n^*(z_{i,n}^*) - \delta_n(z_{i,n}^*)|^2] + O_{\mathbb{P}}(n^{-1}) = o_{\mathbb{P}}(1)$$

and

$$\hat{\mathcal{B}}_n^* = \mathbb{E}^*\hat{\mathcal{B}}_n^* + o_{\mathbb{P}}(n^{-1/2}) = \frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n) du dv + o_{\mathbb{P}}(n^{-1/2}).$$

In other words, if $h_n \to 0$ and if $nh_n^d \to \infty$, then Condition AN* holds with $\Omega_0^* = \Omega_0$ and

$$\mathcal{B}_n^* = \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n) du dv.$$

In summary, if $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then the conditions of Theorem 2 are satisfied.

SA.1.3 Locally Robust Estimator

When verifying the conditions of Theorems 1 and 2 for $\hat{\theta}_n^{\texttt{LR}}$, we set $d_\gamma = 1$, $x(z, \theta) = z$, $w(z, \theta) = 1$, and let $\hat{\theta}_n^{\texttt{LR}}$ be defined by $\hat{G}_n(\hat{\theta}_n^{\texttt{LR}}, \hat{\gamma}_n) = 0$, where

$$g(z, \theta, \gamma) = g^{\texttt{LR}}(z, \theta, \gamma) = 2g^{\texttt{AD}}(z, \theta, \gamma) - g^{\texttt{ISD}}(z, \theta, \gamma),$$

which is a "locally robust" estimating equation because, with ∇_γ denoting the appropriate functional derivative, $\nabla_\gamma \mathbb{E}[g(z, \theta_0, \gamma)]|_{\gamma_0} = 0$.

If the bandwidth satisfies $nh_n^{4P} \to 0$ and $nh_n^d \to \infty$, we show here that the assumptions of Theorems 1 and 2 are satisfied and that

$$\sqrt{n}(\hat{\theta}_n^{\texttt{LR}} - \theta_0 - \mathfrak{B}_n^{\texttt{LR}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0), \qquad \Sigma_0 = 4\mathbb{V}[\gamma_0(z)],$$

with

$$\mathfrak{B}_n^{\texttt{LR}} = 2\mathfrak{B}_n^{\texttt{AD}} - \mathfrak{B}_n^{\texttt{ISD}} = \frac{1}{nh_n^d}\left(2K(0) - \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n) du dv\right) = O(n^{-1}h_n^{-d}).$$

Once again, $\sqrt{n}\mathfrak{B}_n^{\texttt{LR}} = O(1/\sqrt{nh_n^{2d}})$ and therefore the condition $nh_n^d \to \infty$ is weak enough to permit $\mathfrak{B}_n^{\texttt{LR}} \neq o(n^{-1/2})$. On the other hand, as before, $\sqrt{n}(\hat{\theta}_n^{\texttt{LR}} - \theta_0 - \mathfrak{B}_n^{\texttt{LR}}) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ reduces to $\sqrt{n}(\hat{\theta}_n^{\texttt{LR}} - \theta_0) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ when imposing conditions requiring $nh_n^{2d} \to \infty$. Importantly, this example shows that $\hat{\theta}_n^{\texttt{LR}}$ has both leave-in and non-linearity small bandwidth biases in general.

SA.1.3.1 Condition AL

Condition AL holds with $\mathcal{J}_n = \mathcal{J}_0 = 1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.1.3.2 Condition AS

In this case, an (exact) quadratic approximation to g_n is given by

$$\bar{g}_n(z, \gamma) = g_n(z, \gamma_n) + g_{n,\gamma}(z)[\gamma - \gamma_n] + \frac{1}{2} g_{n,\gamma\gamma}[\gamma - \gamma_n, \gamma - \gamma_n],$$

with

$$g_n(z, \gamma_n) = 2\gamma_n(z) - \int_{\mathbb{R}^d} [n^{-1}K_n(u - z) + (1 - n^{-1})\gamma_n(u)]^2 du - \theta_0,$$

$$g_{n,\gamma}(z)[\eta] = 2(1-n^{-1})\eta(z) - 2(1-n^{-1})\int_{\mathbb{R}^d}[n^{-1}K_n(u-z) + (1-n^{-1})\gamma_n(u)]\eta(u)du,$$

$$g_{n,\gamma\gamma}[\eta,\varphi] = -2(1-n^{-1})^2\int_{\mathbb{R}^d}\eta(u)\varphi(u)du.$$

The first part of Condition AS holds directly, without any remainder term because the quadratic approximation above is exact. Next, if $nh_n^d \to 0$, simple variance calculations show that if i, j, and k are distinct, then

$$\mathbb{V}[g_{n,\gamma}(z_i)[\hat{\gamma}_n^j - \gamma_n]] = O(h_n^{-d}) = o(n), \qquad \mathbb{V}(g_{n,\gamma\gamma}[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^k - \gamma_n]) = O(h_n^{-2d}) = o(n^2),$$

$$\mathbb{V}(\mathbb{E}(g_{n,\gamma\gamma}[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n]|z_i)) = O(h_n^{-2d}) = o(n^2), \qquad \mathbb{V}(g_{n,\gamma\gamma}[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n]) = O(h_n^{-3d}) = o(n^3),$$

and hence Condition AS holds via Lemma 2.

SA.1.3.3 Condition AN

We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n[g_n(z_i,\gamma_n) + \bar{G}_n(\hat{\gamma}_n^{(i)}) - \bar{G}_n(\gamma_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_n(z_i) + \sqrt{n}\hat{\mathcal{B}}_n,$$

where

$$\psi_n(z) = 4[\gamma_n^+(z) - \theta_n^+] - 2(1-n^{-1})\int_{\mathbb{R}^d}\gamma_n(u)[K_n(u-z) - \gamma_n(u)]du,$$

$$\begin{aligned}\hat{\mathcal{B}}_n &= 2(\theta_n^+ - \theta_0) - \left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}[n^{-1}K_n(u-z) + (1-n^{-1})\gamma_n(u)]^2 du\gamma_0(v)dv - \theta_0\right)\\ &\quad -(1-n^{-1})^2\frac{1}{n}\sum_{i=1}^n\int_{\mathbb{R}^d}[\hat{\gamma}_n^{(i)}(u) - \gamma_n(u)]^2du.\end{aligned}$$

Suppose $h_n \to 0$ and $nh_n^d \to \infty$. Proceeding as above, we have

$$\psi_n(z) \to \psi_0(z) = 2[\gamma_0(z) - \theta_0]$$

for each z, and therefore $\mathbb{E}[|\psi_n(z) - \psi_0(z)|^2] \to 0$ by the dominated convergence theorem. Moreover, if i and j are distinct, then

$$\mathbb{V}(G_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]) = O(h_n^{-2d}) = o(n^{-2}), \qquad \mathbb{V}(G_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^j - \gamma_n]) = O(h_n^{-d}) = o(n^{-1}),$$

so it follows from Lemma 3 that Condition AN is satisfied with $\Omega_0 = 4\mathbb{V}[\gamma_0(z)]$.

Finally, consider $\mathcal{B}_n^{\mathtt{S}}, \mathcal{B}_n^{\mathtt{LI}}$, and $\mathcal{B}_n^{\mathtt{NL}}$. It follows from the results for $\hat{\theta}_n^{\mathtt{AD}}$ and $\hat{\theta}_n^{\mathtt{ISD}}$ that

$$\mathcal{B}_n^{\mathtt{LI}} = \frac{1}{nh_n^d}2K(0) + O(n^{-1}),$$

$$\mathcal{B}_n^{\texttt{NL}} = -\frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K\left(v\right)^2\gamma_0(u-vh_n)dudv + O(n^{-1}+n^{-2}h_n^{-d}),$$

and $\mathcal{B}_n^{\texttt{S}} = o(h_n^P)$. The latter can be sharpened because it follows from a direct calculation that

$$\begin{aligned}\mathcal{B}_n^{\texttt{S}} &= 2\left[\theta_n^+ - \theta_0\right] - \left(\int_{\mathbb{R}^d}\gamma_n(u)^2du - \theta_0\right) = -\int_{\mathbb{R}^d}[\gamma_n(u)-\gamma_0(u)]^2du \\ &= h_n^{2P}\mathcal{B}_0^{\texttt{S}} + o(h_n^{2P}),\end{aligned}$$

where

$$\begin{aligned}\mathcal{B}_0^{\texttt{S}} &= -\int_{\mathbb{R}^d}\left[\sum_{|p|=P}\left(\int_{\mathbb{R}^d}u^pK(u)du\right)\left(\frac{1}{p!}\partial^p\gamma_0(v)\right)\right]^2 dv \\ &= -\sum_{|p|=P,|s|=P}\frac{1}{p!s!}\left(\int_{\mathbb{R}^d}u^pK(u)du\right)\left(\int_{\mathbb{R}^d}u^sK(u)du\right)\left(\int_{\mathbb{R}^d}\left(\partial^p\gamma_0(u)\right)\left(\partial^s\gamma_0(u)\right)du\right).\end{aligned}$$

As a consequence, we can set

$$\mathcal{B}_n = \frac{1}{nh_n^d}\left(2K(0) - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K\left(v\right)^2\gamma_0(u-vh_n)dudv\right)$$

provided that $nh_n^{4P}\to 0$.

In summary, if $nh_n^{4P}\to 0$ and if $nh_n^d\to\infty$, then the conditions of Theorem 1 are satisfied and $\sqrt{n}(\hat{\theta}_n^{\texttt{LR}} - \theta_0 - \mathfrak{B}_n^{\texttt{LR}}) \rightsquigarrow \mathcal{N}(0,\Sigma_0)$ holds with $\Sigma_0 = 4\mathbb{V}[\gamma_0(z)]$ and $\mathfrak{B}_n^{\texttt{LR}} = O(n^{-1}h_n^{-d})$.

SA.1.3.4 Bandwidth Selection

Assuming $\mathcal{B}_0^{\texttt{SB}}\neq 0$ and $\mathcal{B}_0^{\texttt{S}}\neq 0$, we can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

$$h_{\texttt{opt}} = \begin{cases}\left(\frac{|\mathcal{B}_0^{\texttt{SB}}|}{|\mathcal{B}_0^{\texttt{S}}|}\frac{1}{n}\right)^{\frac{1}{2P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\texttt{SB}})\neq\operatorname{sgn}(\mathcal{B}_0^{\texttt{S}}) \\ \left(\frac{d}{2P}\frac{|\mathcal{B}_0^{\texttt{SB}}|}{|\mathcal{B}_0^{\texttt{S}}|}\frac{1}{n}\right)^{\frac{1}{2P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\texttt{SB}})=\operatorname{sgn}(\mathcal{B}_0^{\texttt{S}})\end{cases}, \qquad \mathcal{B}_0^{\texttt{SB}} = 2K(0) - \int_{\mathbb{R}^d}K(u)^2du.$$

SA.1.3.5 Condition AL*

Condition AL* holds with $\mathcal{J}_n^* = \mathcal{J}_0^* = 1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.1.3.6 Condition AS*

Define the (exact) quadratic approximation

$$\bar{g}_n^*(z,\gamma) = g_n^*(z,\hat{\gamma}_n) + g_{n,\gamma}^*(z)[\gamma-\hat{\gamma}_n] + \frac{1}{2}g_{n,\gamma\gamma}^*[\gamma-\hat{\gamma}_n,\gamma-\hat{\gamma}_n],$$

with

$$g_n^*(z, \hat{\gamma}_n) = 2\hat{\gamma}_n(z) - \int_{\mathbb{R}^d} [n^{-1} K_n(u-z) + (1-n^{-1})\hat{\gamma}_n(u)]^2 du - \hat{\theta}_n,$$

$$g_{n,\gamma}^*(z)[\eta] = 2(1-n^{-1})\eta(z) - 2(1-n^{-1}) \int_{\mathbb{R}^d} [n^{-1} K_n(u-z) + (1-n^{-1})\hat{\gamma}_n(u)]\eta(u) du,$$

$$g_{n,\gamma\gamma}^*[\eta, \varphi] = -2(1-n^{-1})^2 \int_{\mathbb{R}^d} \eta(u)\varphi(u) du.$$

Condition AS* holds if $nh_n^d \to \infty$, because then the conditions of Lemma 5 hold: If i, j, and k are distinct, then

$$\mathbb{V}^*[g_{n,\gamma}(z_{i,n}^*)[\gamma_n^{*,j} - \hat{\gamma}_n]] = O_{\mathbb{P}}(h_n^{-d}) = o_{\mathbb{P}}(n),$$

$$\mathbb{V}^*(g_{n,\gamma\gamma}[\gamma_n^{*,j} - \hat{\gamma}_n, \gamma_n^{*,k} - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-2d}) = o_{\mathbb{P}}(n^2),$$

$$\mathbb{V}^*[\mathbb{E}^*(g_{n,\gamma\gamma})[\gamma_n^{*,j} - \hat{\gamma}_n, \gamma_n^{*,j} - \hat{\gamma}_n]|z_{i,n}^*)] = O_{\mathbb{P}}(h_n^{-2d}) = o_{\mathbb{P}}(n^2),$$

$$\mathbb{V}^*(g_{n,\gamma\gamma}^*[\gamma_n^{*,j} - \hat{\gamma}_n, \gamma_n^{*,j} - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-3d}) = o_{\mathbb{P}}(n^3).$$

SA.1.3.7 Condition AN*

It follows directly from the calculations above that if $h_n \to 0$ and if $nh_n^d \to \infty$, then Condition AN* holds with $\Omega_0^* = \Omega_0$ and

$$\mathcal{B}_n^* = \frac{1}{nh_n^d} \left(2K(0) - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n) du dv \right).$$

In summary, if $nh_n^{4P} \to 0$ and if $nh_n^d \to \infty$, then the conditions of Theorem 2 are satisfied.

SA.2 Example 2: Inverse Probability Weighting

This example is also discussed in the paper. It illustrates two important features that are absent in the average density example: (i) the parameter of interest is (implicitly) defined via a possibly non-differentiable moment condition (i.e., Condition AL does not hold automatically), and (ii) the unknown regression function is estimated using local polynomial estimators. Overidentification of the parameter of interest could also be handled in this example, but we abstract from this additional complication to save some space. Finally, see also the results in Cattaneo, Crump, and Jansson (2013) concerning large sample distribution theory robust to (possibly) small bandwidths in the context of weighted average derivatives for a simpler example of a non-linear (in the nonparametric component) semiparametric problem that also fits into our general framework

Suppose $z_1, \ldots, z_n$ are *i.i.d.* copies of $z = (y, t, x')'$, where $y \in \mathbb{R}$ is a scalar dependent variable, $t \in \{0, 1\}$ is a binary indicator, and $x \in \mathbb{X} \subseteq \mathbb{R}^d$ is a continuous covariate with density f_0. Assuming the estimand $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ is the unique solution to an equation of the form

$$\mathbb{E}\left[\frac{t}{q_0(x)}m(y;\theta)\right] = 0, \qquad q_0(x) = \mathbb{E}(t|x) = \mathbb{P}[t=1|x],$$

where m is a known $\mathbb{R}^{d_\theta}$-valued function, an IPW estimator $\hat{\theta}_n$ of θ_0 is one that satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\frac{t_i}{\hat{q}_n(x_i)}m(y_i;\hat{\theta}_n) = o_{\mathbb{P}}(n^{-1/2}),$$

where $\hat{q}_n$ is an estimator of (the propensity score) q_0.

In what follows, we assume that q_0 is estimated using a local polynomial estimator of order $P > 3d/4 - 1$. To describe this estimator, define $d_P = (P+d-1)!/[P!(d-1)!]$, and let $b_P(x) \in \mathbb{R}^{d_P}$ denote the P-th order polynomial basis expansion based on $x = (x_1, \ldots, x_d)' \in \mathbb{R}^d$; that is,

$$b_P(x) = \begin{pmatrix} 1 \\ [x]^1 \\ \vdots \\ [x]^P \end{pmatrix}, \qquad [x]^p = \begin{pmatrix} x_1^p \\ x_1^{p-1}x_2 \\ \vdots \\ x_d^p \end{pmatrix}.$$

In other words, the basis vector $b_P(x)$ is defined by $b_P(x) = (1, [x']^1, \ldots, [x']^P)'$ with

$$[x']^p = \left[x_1^{\ell_1}x_2^{\ell_2}\cdots x_d^{\ell_d} : |\ell| = \ell_1 + \ell_2 + \cdots + \ell_d = p, \quad \ell = (\ell_1, \ell_2, \cdots, \ell_d) \in \mathbb{Z}_+^d\right],$$

assumed to be ordered lexicographically.

The local polynomial estimator (of order P) of $q_0(x)$ is given by

$$\hat{q}_n(x) = e_P'\hat{\xi}_n(x), \qquad \hat{\xi}_n(x) = \arg\min_{\xi \in \mathbb{R}^{d_P}} \sum_{i=1}^{n}(t_i - b_P(x_i - x)'\xi)^2 K_n(x_i - x),$$

where e_P is the first unit vector in $\mathbb{R}^{d_P}$, $K_n(u) = K(u/h_n)/h_n^d$, $h_n = o(1)$ is a bandwidth, and K is a kernel. For our purposes, it is convenient to work with the representation $\hat{q}_n(x) = q(x;\hat{\gamma}_n)$, where

$$q(x;\gamma) = e_P'(\mathrm{vec}_P^{-1}[\gamma_x(x)])^{-1}\gamma_t(x), \qquad \gamma = (\gamma_x', \gamma_t')',$$

$$\hat{\gamma}_{x,n}(x) = \mathrm{vec}_P[\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{x,n}(x_i - x)], \qquad \mathcal{K}_{x,n}(u) = b_{P,n}(u)b_{P,n}(u)'K_n(u),$$

$$\hat{\gamma}_{t,n}(x) = \frac{1}{n}\sum_{i=1}^{n}t_i\mathcal{K}_{t,n}(x_i - x), \qquad \mathcal{K}_{t,n}(u) = b_{P,n}(u)K_n(u),$$

with $\mathrm{vec}_P : \mathbb{R}^{d_P \times d_P} \to \mathbb{R}^{d_P^2}$ denoting the vectorization operator, $\mathrm{vec}_P^{-1} : \mathbb{R}^{d_P^2} \to \mathbb{R}^{d_P \times d_P}$ denoting the inverse of vec_P, and defining $b_{P,n}(x) = b_P(x/h_n)$.

Defining

$$\gamma_{x,0}(x) = f_0(x)\,\mathrm{vec}_P[\int_{\mathbb{R}^d} \mathcal{K}_x(u)du], \qquad \mathcal{K}_x(u) = b_P(u)b_P(u)'K(u),$$

$$\gamma_{t,0}(x) = q_0(x)f_0(x)\int_{\mathbb{R}^d} \mathcal{K}_t(u)du, \qquad \mathcal{K}_t(u) = b_P(u)K(u),$$

we note in passing that $\gamma_0 = (\gamma_{x,0}', \gamma_{t,0}')'$ satisfies $q_0(x) = q(x;\gamma_0)$.

Because $\hat{\gamma}_n$ is kernel-based, the associated IPW estimator $\hat{\theta}_n$ is a kernel-based two-step semiparametric estimator, which can be analyzed using the results of the paper by representing the defining property of $\hat{\theta}_n$ as

$$\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_{\mathbb{P}}(n^{-1}), \qquad \hat{W}_n = I_{d_\theta},$$

where

$$g(z,\theta,\gamma) = \frac{t}{q(x;\gamma)}m(y;\theta)$$

is neither linear in γ nor (necessarily) differentiable in θ.

We impose the following primitive regularity conditions:

- $\theta_0 \in \mathrm{int}(\Theta)$.
- $\mathbb{E}[\|t\cdot m(y;\theta_0)\|^4] + \sup_{x\in\mathbb{X}}\mathbb{E}[\|t\cdot m(y;\theta_0)\|^4|x]f_0(x) < \infty$ and $\mathbb{E}\left[\sup_{\theta\in\Theta}\|t\cdot m(y;\theta)\|^2\right] < \infty$.
- $\mathcal{M} = \{t\cdot m(y;\theta) : \theta\in\Theta\}$ satisfies the bracketing integral entropy condition

  $$J_{[],2}(\delta,\mathcal{M}) = \int_0^\delta \sqrt{\log N_{[],2}(\varepsilon,\mathcal{M})}d\varepsilon < \infty,$$

  with $N_{[],2}(\cdot,\mathcal{M})$ denoting the L_2-bracketing number for the class $\mathcal{M}$; for more details and precise definitions see, e.g., van der Vaart and Wellner (1996). Furthermore, as $\theta\to\theta_0$, $\mathbb{E}[t\|m(y;\theta)-m(y;\theta_0)\|] = O(\|\theta-\theta_0\|)$ and $\mathbb{E}[t\|m(y;\theta)-m(y;\theta_0)\|^2] = O(\|\theta-\theta_0\|^p)$ for some $p\in[1,2]$.
- $r_0(x;\theta) = \mathbb{E}[m(y;\theta)|x,t=1]$ is twice continuously differentiable in θ, with first and second bounded derivatives denoted by $\dot{r}_0(x;\theta)$ and $\ddot{r}_0(x;\theta)$, and $\mathbb{E}[\sup_{\|\theta-\theta_0\|\le\delta}\|\ddot{r}_0(x;\theta)\|] < \infty$ for some $\delta > 0$.
- f_0 is bounded away from zero on $\mathbb{X}$.
- q_0 is bounded away from zero and $P+2$ times continuously differentiable on $\mathbb{X}$.
- $\mathbb{V}[\psi_0(z)]$ is positive definite, where

  $$\psi_0(z) = \frac{t}{q_0(x)}m(y;\theta_0) - \frac{r_0(x;\theta_0)}{q_0(x)}(t-q_0(x)).$$

- K is even, compactly supported, and continuously differentiable.

With the possible exception of the third assumption, these assumptions are standard. The third assumption controls the "smoothness" of $\theta \mapsto m(y;\theta)$ and holds, in particular, if $m(y;\theta)$ is Lipschitz continuous in θ (and the implied Lipschitz constant is integrable). More generally, certain types of discontinuous-in-θ moment functions are also allowed, such as $m(y;\theta) = \mathbb{1}(y \leq \theta) - \tau$ for the τ-th quantile of y, $\tau \in (0,1)$; this function satisfies

$$\mathbb{E}[\|m(y;\theta) - m(y;\theta_0)\|] = \mathbb{E}[\mathbb{1}(\min\{\theta,\theta_0\} < y \leq \max\{\theta,\theta_0\})] = O(\|\theta - \theta_0\|),$$

provided that y is continuously distributed with bounded density.

Defining $\gamma_n = (\gamma_{x,n}', \gamma_{t,n}')'$ with

$$\gamma_{x,n}(x) = \mathrm{vec}_P[\int_{\mathbb{R}^d} \mathcal{K}_x(u) f_0(x + uh_n)du], \qquad \gamma_{t,n}(x) = \int_{\mathbb{R}^d} \mathcal{K}_t(u) q_0(x + uh_n) f_0(x + uh_n)du,$$

we also impose the following assumptions on the kernel-based nonparametric estimators:

- Uniform consistency:

$$\sup_{x\in\mathbb{X}} \|\hat{\gamma}_n(x) - \gamma_n(x)\| = o_{\mathbb{P}}(1), \qquad \sup_{x\in\mathbb{X}} \|\hat{\gamma}_n^*(x) - \hat{\gamma}_n(x)\| = o_{\mathbb{P}}(1).$$

- Empirical uniform rate of convergence:

$$\max_{1\leq i\leq n} \|\hat{\gamma}_n(x_i) - \gamma_n(x_i)\| = o_{\mathbb{P}}(n^{-1/6}), \qquad \max_{1\leq i,j\leq n} \|\hat{\gamma}_n^{(i)}(x_j) - \gamma_n(x_j)\| = o_{\mathbb{P}}(n^{-1/6}),$$

$$\max_{1\leq i\leq n} \|\hat{\gamma}_n^*(x_i) - \hat{\gamma}_n(x_i)\| = o_{\mathbb{P}}(n^{-1/6}), \qquad \max_{1\leq i,j\leq n} \|\hat{\gamma}_n^{*,(i)}(x_j) - \hat{\gamma}_n(x_j)\| = o_{\mathbb{P}}(n^{-1/6}).$$

- $\underline{\lim}_{n\to\infty} \inf_{x\in\mathbb{X}} q_n(x) > 0$, where $q_n(x) = q(x;\gamma_n)$.

Primitive conditions for these assumptions can be given using standard methods in the literature and Lemma SA-1 below. For example, using Lemma SA-1 below, we have

$$\max_{1\leq i\leq n} \|\hat{\gamma}_n(x_i) - \gamma_n(x_i)\| = O_{\mathbb{P}}(\sqrt{\log n}/\sqrt{nh_n^d}) = o_{\mathbb{P}}(n^{-1/6}),$$

provided that $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$, and similarly for the bootstrap and leave-one-out versions. Furthermore, these assumptions imply

$$\sup_{x\in\mathbb{X}} |\hat{q}_n(x) - q_n(x)| = o_{\mathbb{P}}(1), \qquad \max_{1\leq i\leq n} |\hat{q}_n(x_i) - q_n(x_i)| = o_{\mathbb{P}}(n^{-1/6}),$$

and similarly for the bootstrap and leave-one-out versions. If also $\sup_{x\in\mathbb{X}} |q_n(x) - q_0(x)| = o(1)$, then the third assumption is satisfied.

Finally, we assume throughout that $\hat{\theta}_n \to_{\mathbb{P}} \theta_0$ and $\hat{\theta}_n^* \to_{\mathbb{P}} \theta_0$. These consistency results can be established using standard techniques already available in the literature.

If the bandwidth satisfies $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$ and $nh_n^{2P+2} \to 0$, we show here that the assumptions of Theorems 1 and 2 are satisfied and that, for some Σ_0,

$$\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0).$$

where $\mathfrak{B}_n = O(n^{-1}h_n^{-d})$. Once again, the stated bandwidth conditions are weak enough to permit $\mathfrak{B}_n \neq o(n^{-1/2})$, while $\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ reduces to $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ when imposing conditions requiring $nh_n^{2d} \to \infty$.

SA.2.1 Condition AL

We apply Lemma 1 with $\rho = 3$ to verify Condition AL. In this example, $W_n = W_0 = I_{d_\theta}$ and

$$G(\theta, \gamma) = \mathbb{E}\left[\frac{t}{q(x;\gamma)} m(y;\theta)\right] = \mathbb{E}\left[\frac{q_0(x)}{q(x;\gamma)} r_0(x;\theta)\right], \qquad \dot{G}(\gamma) = \mathbb{E}\left[\frac{q_0(x)}{q(x;\gamma)} \dot{r}_0(x;\theta_0)\right].$$

The smoothness assumptions imposed on $r_0(x;\theta)$ imply that as $\theta \to \theta_0$,

$$\|G(\theta,\gamma) - G(\theta_0,\gamma) - \dot{G}(\gamma)(\theta - \theta_0)\| \le O(\|\theta - \theta_0\|^2) \int_{\mathbb{R}^d} \frac{q_0(u)}{|q(u;\gamma)|} f_0(u) du.$$

Also, setting $\dot{G}_n = \dot{G}(\gamma_n)$, we obtain

$$\begin{aligned} \|\dot{G}_n - \dot{G}_0\| &\le \int_{\mathbb{R}^d} \left| \frac{q(x;\gamma_n) - q(x;\gamma_0)}{q(x;\gamma_n) q(x;\gamma_0)} \right| \|\dot{r}_0(x;\theta_0)\| \, f_0(x) dx \\ &= O(1) \int_{\mathbb{R}^d} \|\gamma_n(x) - \gamma_0(x)\| \, f_0(x) dx = o(1), \end{aligned}$$

under the assumptions imposed and provided that $h_n \to 0$.

Condition (i). Holds by definition of the estimator.

Condition (ii). Using the calculations above, as $\theta \to \theta_0$,

$$\|G(\theta,\hat{\gamma}_n) - G(\theta_0,\hat{\gamma}_n) - \dot{G}(\hat{\gamma}_n)(\theta - \theta_0)\| = O_{\mathbb{P}}(\|\theta - \theta_0\|^2)$$

because $\sup_{x \in \mathbb{X}} \|\hat{q}_n(x) - q_n(x)\| = o_{\mathbb{P}}(1)$ and q_n is bounded away from zero for all n large enough. This implies, for every $\delta_n = o(1)$,

$$\sup_{\|\theta - \theta_0\| \le \delta_n} \frac{\|G(\theta,\hat{\gamma}_n) - G(\theta_0,\hat{\gamma}_n) - \dot{G}(\hat{\gamma}_n)(\theta - \theta_0)\|}{\|\theta - \theta_0\|^{3/2}} = O_{\mathbb{P}}(\delta_n^{1/2}) = o_{\mathbb{P}}(1).$$

Condition (iii). We have

$$\|\hat{G}_n(\theta,\hat{\gamma}_n) - G(\theta,\hat{\gamma}_n) - \hat{G}_n(\theta_0,\hat{\gamma}_n) + G(\theta_0,\hat{\gamma}_n)\| \leq \Delta_{1,n}(\theta) + \rho_n\Delta_{2,n}(\theta) + \Delta_{3,n}(\theta) + \rho_n\Delta_{4,n}(\theta),$$

where

$$\rho_n = \max_{1\leq i\leq n} \frac{|\hat{q}_n(x_i) - q_n(x_i)|}{|\hat{q}_n(x_i)q_n(x_i)|} = o_{\mathbb{P}}(n^{-1/6})$$

and

$$\Delta_{1,n}(\theta) = \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)}(m(y_i;\theta) - m(y_i;\theta_0)) - \mathbb{E}\left[\frac{t}{q_n(x)}(m(y;\theta) - m(y;\theta_0))\right]\right\|,$$

$$\Delta_{2,n}(\theta) = \left| \frac{1}{n}\sum_{i=1}^{n} t_i \|m(y_i;\theta) - m(y_i;\theta_0)\| - \mathbb{E}[t\,\|m(y;\theta) - m(y;\theta_0)\|]\right|,$$

$$\Delta_{3,n}(\theta) = \int_{\mathbb{R}^d} \frac{|\hat{q}_n(u) - q_n(u)|}{|\hat{q}_n(u)q_n(u)|}\|r_0(u;\theta) - r_0(u;\theta_0)\| f_0(u)du,$$

$$\Delta_{4,n}(\theta) = \mathbb{E}\left[t\|m(y;\theta) - m(y;\theta_0)\|\right].$$

In what follows, suppose $\delta_n = o(1)$. First,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{1,n}(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

because q_n is non-random and bounded away from zero for all n large enough, and because the class of n-varying functions $\mathcal{M}_n = \{tm(y;\theta)/q_n(x) : \theta \in \Theta\}$ satisfies $J_{[],2}(\epsilon_n, \mathcal{M}_n) \to 0$ for all $\epsilon_n \downarrow 0$.

Similarly, using $\rho_n = o_{\mathbb{P}}(1)$,

$$\rho_n \sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{2,n}(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

because the class of functions $\mathcal{M}_{\|\cdot\|} = \{t\,\|m(y;\theta) - m(y;\theta_0)\| : \theta \in \Theta\}$ satisfies $J_{[],2}(1, \mathcal{M}_{\|\cdot\|}) < \infty$.

Also,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\Delta_{3,n}(\theta)}{1 + n^{1/3}\|\theta-\theta_0\|} = O_{\mathbb{P}}(n^{-1/3}) \int_{\mathbb{R}^d} |\hat{q}_n(u) - q_n(u)| f_0(u)du = o_{\mathbb{P}}(n^{-1/2})$$

because $\sup_{x\in\mathbb{X}} |\hat{q}_n(x) - q_n(x)| = o_{\mathbb{P}}(1)$, q_n is bounded away from zero for all n large enough, $\sup_{x\in\mathbb{X}} \|r_0(x;\theta) - r_0(x;\theta_0)\| = O(\|\theta-\theta_0\|)$, and, using standard results for local polynomial regression estimators and the assumed bandwidth rate restrictions,

$$\int_{\mathbb{R}^d} |\hat{q}_n(u) - q_n(u)|^2 f_0(u)du = O_{\mathbb{P}}(n^{-1}h_n^{-d}) = o_{\mathbb{P}}(n^{-1/3}).$$

Finally, by arguments similar to those given above,

$$\rho_n \sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\Delta_{4,n}(\theta)}{1 + n^{1/3}\|\theta-\theta_0\|} = o_{\mathbb{P}}(n^{-1/6})O_{\mathbb{P}}(n^{-1/3}) = o_{\mathbb{P}}(n^{-1/2}).$$

Condition (iv). Follows directly by the results established in the following subsections, because

$$\begin{aligned}\hat{G}_n(\theta_0, \hat{\gamma}_n) &= \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{\hat{q}_n(x_i)} m(y_i;\theta_0) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)} m(y_i;\theta_0) - \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^2} m(y_i;\theta_0)[\hat{q}_n(x_i) - q_n(x_i)] \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^3} m(y_i;\theta_0)[\hat{q}_n(x_i) - q_n(x_i)]^2 \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^3 \hat{q}_n(x_i)} m(y_i;\theta_0)[\hat{q}_n(x_i) - q_n(x_i)]^3\end{aligned}$$

where

$$\frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)} m(y_i;\theta_0) = O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(n^{-1/3}),$$

$$\frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^2} m(y_i;\theta_0)[\hat{q}_n(x_i) - q_n(x_i)] = O_{\mathbb{P}}(n^{-1}h_n^{-d} + n^{-1/2}) = o_{\mathbb{P}}(n^{-1/3}),$$

$$\frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^3} m(y_i;\theta_0)[\hat{q}_n(x_i) - q_n(x_i)]^2 = O_{\mathbb{P}}(n^{-1}h_n^{-d}) + o_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(n^{-1/3}),$$

$$\begin{aligned}&\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^3 \hat{q}_n(x_i)} m(y_i;\theta_0)[\hat{q}_n(x_i) - q_n(x_i)]^3 \right\| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{|q_n(x_i)^3 \hat{q}_n(x_i)|} \|m(y_i;\theta_0)\| \|\hat{q}_n(x_i) - q_n(x_i)\|^3 \\ &= O_{\mathbb{P}}(1) \max_{1\leq i\leq n} \|\hat{\gamma}_n(x_i) - \gamma_n(x_i)\|^3 = O_{\mathbb{P}}(\sqrt{(\log n)^3}/\sqrt{n^3 h_n^{3d}}) = o_{\mathbb{P}}(n^{-1/3}),\end{aligned}$$

provided that $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

Condition (v). Holds by assumption because θ_0 is an interior point.

Condition (vi). We have $W_n = I_{d_\theta} = W_0$ and, using $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$,

$$\begin{aligned}\left\| \dot{G}(\hat{\gamma}_n) - \dot{G}_n \right\| &\leq \int_{\mathbb{R}^d} \left| \frac{\hat{q}_n(u) - q_n(u)}{\hat{q}_n(u) q_n(u)} \right| \|\dot{r}_0(u;\theta_0)\| f_0(u) dx \\ &= O_{\mathbb{P}}(1) \int_{\mathbb{R}^d} \|\hat{\gamma}_n(u) - \gamma_n(u)\| f_0(u) dx = O_{\mathbb{P}}(1/\sqrt{nh_n^d}) = o_{\mathbb{P}}(n^{-1/6}).\end{aligned}$$

Condition (vii). Suppose $\delta_n = O(n^{-1/3})$. When verifying condition (iii), we already showed

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{1,n}(\theta) = o_{\mathbb{P}}(n^{-1/2}), \qquad \rho_n \sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{2,n}(\theta) = o_{\mathbb{P}}(n^{-1/2}).$$

Proceeding as in condition (iii), for every $\delta_n = O(n^{-1/3})$, we also have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{3,n}(\theta) = O_{\mathbb{P}}(\delta_n)\int_{\mathbb{R}^d} |q(u;\hat{\gamma}_n) - q(u;\gamma_n)|f_0(u)du = o_{\mathbb{P}}(n^{-1/2})$$

and

$$\rho_n \sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{22,n}(\theta) = o_{\mathbb{P}}(n^{-1/6})O_{\mathbb{P}}(\delta_n) = o_{\mathbb{P}}(n^{-1/2}).$$

SA.2.2 Condition AS

A quadratic approximation to $g_n(z,\gamma)$ is given by

$$\bar{g}_n(z,\gamma) = g_n(z,\gamma_n) + g_{n,\gamma}(z)[\gamma-\gamma_n] + \frac{1}{2}g_{n,\gamma\gamma}(z)[\gamma-\gamma_n,\gamma-\gamma_n],$$

where

$$g_n(z,\gamma_n) = \frac{t}{q_n^+(x)}m(y;\theta_0), \qquad q_n^+(x) = e_P'\Upsilon_n^+(x),$$

$$\Upsilon_n^+(x) = \Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x), \qquad \Gamma_{x,n}^+(x) = \operatorname{vec}_P^{-1}(\gamma_{x,n}^+(x)),$$

$$\gamma_{x,n}^+(x) = (n-1)^{-1}K_n(0)\operatorname{vec}_P(e_Pe_P') + \gamma_{x,n}(x), \qquad \gamma_{t,n}^+(x) = (n-1)^{-1}K_n(0)e_P + \gamma_{t,n}(x),$$

and where, for $\eta = (\eta_x',\eta_t')'$ and $\varphi = (\varphi_x',\varphi_t')'$, the linear and quadratic terms are of the form

$$g_{n,\gamma}(z)[\eta] = -\frac{tm(y;\theta_0)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\eta_t(x) + \frac{tm(y;\theta_0)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Upsilon_n^+(x)$$

and

$$g_{n,\gamma\gamma}(z)[\eta,\varphi] = \sum_{\ell=1}^{10} g_{n,\gamma\gamma,\ell}(z)[\eta,\varphi],$$

respectively, with

$$g_{n,\gamma\gamma,1}(z)[\eta,\varphi] = -\frac{tm(y;\theta_0)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(x))\Upsilon_n^+(x),$$

$$g_{n,\gamma\gamma,2}(z)[\eta,\varphi] = -\frac{tm(y;\theta_0)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(x))\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Upsilon_n^+(x),$$

$$g_{n,\gamma\gamma,3}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\varphi_t(x),$$

$$g_{n,\gamma\gamma,4}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(x))\Gamma_{x,n}^+(x)^{-1}\eta_t(x),$$

$$g_{n,\gamma\gamma,5}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{q_n^+(x)^3}e_P'\Gamma_{x,n}^+(x)^{-1}\eta_t(x)\varphi_t(x)'\Gamma_{x,n}^+(x)^{-1}e_P,$$

$$g_{n,\gamma\gamma,6}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{q_n^+(x)^3}e_P'\Gamma_{x,n}^+(x)^{-1}\varphi_t(x)\eta_t(x)'\Gamma_{x,n}^+(x)^{-1}e_P,$$

$$g_{n,\gamma\gamma,7}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Upsilon_n^+(x)\Upsilon_n^+(x)'\operatorname{vec}_P^{-1}(\varphi_x(x))\Gamma_{x,n}^+(x)^{-1}e_P,$$

$$g_{n,\gamma\gamma,8}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(x))\Upsilon_n^+(x)\Upsilon_n^+(x)'\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}e_P,$$

$$g_{n,\gamma\gamma,9}(z)[\eta,\varphi] = -\frac{2tm(y;\theta_0)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}\eta_t(x)\Upsilon_n^+(x)'\operatorname{vec}_P^{-1}(\varphi_x(x))\Gamma_{x,n}^+(x)^{-1}e_P,$$

$$g_{n,\gamma\gamma,10}(z)[\eta,\varphi] = -\frac{2tm(y;\theta_0)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}\varphi_t(x)\Upsilon_n^+(x)'\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}e_P.$$

Suppose $h_n \to 0$. The first part of Condition AS is satisfied if

$$\max_{1\le i\le n} \|\hat{\gamma}_n^{(i)}(x_i) - \gamma_n(x_i)\| = o_{\mathbb{P}}(n^{-1/6}),$$

a sufficient condition for which is that $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

Moreover, the second part of Condition AS can be verified using Lemma 2, because if i, j, and k are distinct, then

$$\mathbb{V}(g_{n,\gamma}(z_i)[\hat{\gamma}_n^j - \gamma_n]) = O(h_n^{-d}) = o(n),$$

$$\mathbb{V}(g_{n,\gamma\gamma}(z_i)[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^k - \gamma_n]) = O(h_n^{-2d}) = o(n^2),$$

$$\mathbb{V}(\mathbb{E}(g_{n,\gamma\gamma}(z_i)[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n]|z_i)) = O(h_n^{-2d}) = o(n^2),$$

$$\mathbb{V}(g_{n,\gamma\gamma}(z_i)[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n]) = O(h_n^{-3d}) = o(n^3),$$

provided $nh_n^d \to \infty$.

In other words, Condition AS holds if $h_n \to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

SA.2.3 Condition AN

Suppose $h_n \to 0$ and $nh_n^d \to \infty$. Then,

$$\mathbb{E}\left[\|g_n(z,\gamma_n) - g_0(z,\gamma_0)\|^2\right] = \mathbb{E}\left[[q_n^+(x) - q_0(x)]^2 \frac{t}{(q_n^+(x)q_0(x))^2}\|m(y;\theta_0)\|^2\right] \to 0.$$

Also, for the correction term δ_n, we have:

$$\delta_n(z) = \delta_{1,n}(z) + \delta_{2,n}(z),$$

where

$$\begin{aligned}\delta_{1,n}(z) \quad &= \quad -\int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0)e_P'\Gamma_{x,n}^+(u)^{-1}\mathcal{K}_{t,n}(x-u)tf_0(u)du \\ &\quad\quad + \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0)e_P'\Gamma_{x,n}^+(u)^{-1}\mathcal{K}_{x,n}(x-u)\Upsilon_n^+(u)f_0(u)du,\end{aligned}$$

$$\begin{aligned}\delta_{2,n}(z) &= \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \gamma_{t,n}(u) f_0(u) du \\ &\quad - \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \Gamma_{x,n}(u) \Upsilon_n^+(u) f_0(u) du.\end{aligned}$$

We analyze each term. First,

$$\begin{aligned}\delta_{1,n}(z) &= -t \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \mathcal{K}_{t,n}(x-u) f_0(u) du \\ &\quad + \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \mathcal{K}_{x,n}(x-u) \Upsilon_n^+(u) f_0(u) du \\ &= -t \int_{\mathbb{R}^d} \frac{q_0(x-vh_n)}{q_n(x-vh_n)^2} r_0(x-vh_n;\theta_0) e_P' \Gamma_{x,n}(x-vh_n)^{-1} b_P(v) K(v) f_0(x-vh_n) dv \\ &\quad + \int_{\mathbb{R}^d} r_0(x-vh_n;\theta_0) \frac{q_0(x-vh_n)^2}{q_n(x-vh_n)^2} e_P' \Gamma_{x,n}(x-vh_n)^{-1} b_P(v) b_P(v)' K(v) f_0(x-vh_n) dv + o(1) \\ &\to -\frac{t}{q_0(x)} r_0(x;\theta_0) + r_0(x;\theta_0).\end{aligned}$$

Next,

$$\begin{aligned}\delta_{2,n}(z) &= \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \gamma_{t,n}(u) f_0(u) du \\ &\quad - \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \Gamma_{x,n}(u) \Upsilon_n^+(u) f_0(u) du \\ &= \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}(u)^{-1} \gamma_{t,n}(u) f_0(u) du \\ &\quad - \int_{\mathbb{R}^d} \frac{q_0(u)}{q_n(u)^2} r_0(u;\theta_0) q_0(u) f_0(u) du + o(1) \\ &\to 0.\end{aligned}$$

Therefore, $\delta_n(z) \to \delta_0(z)$ and

$$\mathbb{E}\left[\|\delta_n(z) - \delta_0(z)\|^2\right] \to 0, \qquad \delta_0(z) = -\frac{r_0(x;\theta_0)}{q_0(x)}(t - q_0(x)).$$

Putting the above together, we have

$$\mathbb{E}\left[\|\psi_n(z) - \psi_0(z)\|^2\right] \to 0.$$

It now follows from Lemma 3 that Condition AN is satisfied with $\Omega_0 = \mathbb{V}[\psi_0(z)]$, because standard bounding arguments can be used to show that if i and j are distinct, then

$$\mathbb{V}(G_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]) = O(h_n^{-2d}) = o(n^2),$$

$$\mathbb{V}(G_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^j - \gamma_n]) = O(h_n^{-d}) = o(n),$$

where the variance calculations use the representation

$$G_{n,\gamma\gamma}[\eta,\varphi]=\sum_{\ell=1}^{10}G_{n,\gamma\gamma,\ell}[\eta,\varphi],$$

with, letting $w_0(x)=q_0(x)r_0(x;\theta_0)f_0(x)$ to save some notation,

$$G_{n,\gamma\gamma,1}[\eta,\varphi]=-\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^2}e_P'\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\eta_x(u))\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(u))\Upsilon_n^+(u)du,$$

$$G_{n,\gamma\gamma,2}[\eta,\varphi]=-\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^2}e_P'\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(u))\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\eta_x(u))\Upsilon_n^+(u)du,$$

$$G_{n,\gamma\gamma,3}[\eta,\varphi]=\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^2}e_P'\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\eta_x(u))\Gamma_{x,n}^+(u)^{-1}\varphi_t(u)du,$$

$$G_{n,\gamma\gamma,4}[\eta,\varphi]=\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^2}e_P'\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(u))\Gamma_{x,n}^+(u)^{-1}\eta_t(u)du,$$

$$G_{n,\gamma\gamma,5}[\eta,\varphi]=\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^3}e_P'\Gamma_{x,n}^+(u)^{-1}\eta_t(u)\varphi_t(u)'\Gamma_{x,n}^+(u)^{-1}e_Pdu,$$

$$G_{n,\gamma\gamma,6}[\eta,\varphi]=\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^3}e_P'\Gamma_{x,n}^+(u)^{-1}\varphi_t(u)\eta_t(u)'\Gamma_{x,n}^+(u)^{-1}e_Pdu,$$

$$G_{n,\gamma\gamma,7}[\eta,\varphi]=\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^3}e_P'\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\eta_x(u))\Upsilon_n^+(u)\Upsilon_n^+(u)'\operatorname{vec}_P^{-1}(\varphi_x(u))\Gamma_{x,n}^+(u)^{-1}e_Pdu,$$

$$G_{n,\gamma\gamma,8}[\eta,\varphi]=\int_{\mathbb{R}^d}\frac{w_0(u)}{q_n^+(u)^3}e_P'\Gamma_{x,n}^+(u)^{-1}\operatorname{vec}_P^{-1}(\varphi_x(u))\Upsilon_n^+(u)\Upsilon_n^+(u)'\operatorname{vec}_P^{-1}(\eta_x(u))\Gamma_{x,n}^+(u)^{-1}e_Pdu,$$

$$G_{n,\gamma\gamma,9}[\eta,\varphi]=-\int_{\mathbb{R}^d}\frac{2w_0(u)}{q_n^+(u)^3}e_P'\Gamma_{x,n}^+(u)^{-1}\eta_t(u)\Upsilon_n^+(u)'\operatorname{vec}_P^{-1}(\varphi_x(u))\Gamma_{x,n}^+(u)^{-1}e_Pdu,$$

$$G_{n,\gamma\gamma,10}[\eta,\varphi]=-\int_{\mathbb{R}^d}\frac{2w_0(u)}{q_n^+(u)^3}e_P'\Gamma_{x,n}^+(u)^{-1}\varphi_t(u)\Upsilon_n^+(u)'\operatorname{vec}_P^{-1}(\eta_x(u))\Gamma_{x,n}^+(u)^{-1}e_Pdu.$$

Finally, to characterize $\mathcal{B}_n$, suppose $nh_n^{2P+2}\to 0$ and $nh_n^{3d/2}/(\log n)^{3/2}\to\infty$ and let

$$\Gamma_{x,n}(x)=\operatorname{vec}_P^{-1}(\gamma_{x,n}(x)),\qquad \vartheta_{P+1,n}(x)=\int_{\mathbb{R}^d}\mathcal{K}_t(u)[u]_{P+1}'f_0(x+uh_n)du,$$

$$q_0^{(P+1)}(x)'=\left[\frac{1}{\ell!}\partial^\ell q_0(x):|\ell|=P+1,\quad \ell=(\ell_1,\ell_2,\cdots,\ell_d)'\in\mathbb{Z}_+^d\right],$$

where for $\ell=(\ell_1,\ell_2,\cdots,\ell_d)'\in\mathbb{Z}_+^d$, the definition of $q_0^{(P+1)}(x)$ uses the multi-index notation

$$|\ell|=\ell_1+\cdots+\ell_d,\qquad \ell!=\ell_1!\cdots\ell_d!,\qquad \partial^\ell=\frac{\partial^{|\ell|}}{\partial^{\ell_1}\cdots\partial^{\ell_d}}.$$

Using

$$\Gamma_{x,n}(x) \to f_0(x)\Gamma_x, \qquad \Gamma_x = \int_{\mathbb{R}^d} \mathcal{K}_x(u)du,$$

$$\vartheta_{P+1,n}(x) \to f_0(x)\vartheta_{P+1}, \qquad \vartheta_{P+1} = \int_{\mathbb{R}^d} \mathcal{K}_t(u)[u]'_{P+1}du,$$

and the representation

$$\begin{aligned} q_n^+(x) - q_0(x) &= q_n(x) - q_0(x) \\ &\quad + (n-1)^{-1}K_n(0)e_P'\Gamma_{x,n}(x)^{-1}e_P\left[1 - \frac{q_n(x)}{1+(n-1)^{-1}K_n(0)e_P'\Gamma_{x,n}(x)^{-1}e_P}\right] \\ &\quad - \frac{(n-1)^{-2}K_n(0)^2}{1+(n-1)^{-1}K_n(0)e_P'\Gamma_{x,n}(x)^{-1}e_P}\left(e_P'\Gamma_{x,n}(x)^{-1}e_P\right)^2, \end{aligned}$$

the leading term in the expansion

$$\mathcal{B}_n^{\mathtt{S}} + \mathcal{B}_n^{\mathtt{LI}} = G_n(\gamma_n) = \mathbb{E}\left[\frac{q_0(x)}{q_n^+(x)}r_0(x;\theta_0)\right] = -\mathbb{E}\left[\frac{q_n^+(x)-q_0(x)}{q_0(x)}r_0(x;\theta_0)\right] + o(n^{-1/2})$$

admits the decomposition

$$-\mathbb{E}\left[\frac{q_n^+(x)-q_0(x)}{q_0(x)}r_0(x;\theta_0)\right] = \bar{\mathcal{B}}_n^{\mathtt{S}} + \bar{\mathcal{B}}_n^{\mathtt{LI}},$$

with

$$\begin{aligned} \bar{\mathcal{B}}_n^{\mathtt{S}} &= -h_n^{P+1}\int_{\mathbb{R}^d} e_P'\Gamma_{x,n}(u)^{-1}\vartheta_{P+1,n}(u)q_0^{(P+1)}(u)\frac{r_0(u;\theta_0)}{q_0(u)}f_0(u)du + o(h_n^{P+1}) \\ &= -h_n^{P+1}(e_P'\Gamma_x^{-1}\vartheta_{P+1})\int_{\mathbb{R}^d} q_0^{(P+1)}(u)\frac{r_0(u;\theta_0)}{q_0(u)}f_0(u)du + o(h_n^{P+1}) \end{aligned}$$

and

$$\begin{aligned} \bar{\mathcal{B}}_n^{\mathtt{LI}} &= -\frac{K(0)}{(n-1)h_n^d}\left[\int_{\mathbb{R}^d} e_P'\Gamma_{x,n}(u)^{-1}e_P\frac{1-q_0(u)}{q_0(u)}r_0(u;\theta_0)f_0(u)du + O(n^{-1}h_n^{-d}) + O(h_n^{P+1})\right] \\ &= -\frac{K(0)}{nh_n^d}(e_P'\Gamma_x^{-1}e_P)\int_{\mathbb{X}}\frac{1-q_0(u)}{q_0(u)}r_0(u;\theta_0)du + o(n^{-1/2}). \end{aligned}$$

To also characterize the nonlinearity bias, we use the representation

$$\begin{aligned} \frac{1}{2}\mathbb{E}G_{n,\gamma\gamma}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n] &= \sum_{\ell\in\{1,3,5,7,9\}}\mathbb{E}G_{n,\gamma\gamma,\ell}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n] \\ &= \sum_{\ell\in\{1,3,5,7,9\}}\mathbb{E}G_{n,\gamma\gamma,\ell}[\hat{\gamma}_n^i,\hat{\gamma}_n^i] + O(1), \end{aligned}$$

where the first equality holds because $G_{n,\gamma\gamma,\ell+1}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n] = G_{n,\gamma\gamma,\ell}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n]$ for $\ell\in\{1,3,5,7,9\}$, and the second equality uses $\mathbb{E}G_{n,\gamma\gamma,\ell}[\hat{\gamma}_n^i,\gamma_n]=O(1)$ and $\mathbb{E}G_{n,\gamma\gamma,\ell}[\gamma_n,\gamma_n]=O(1)$.

The term $\mathbb{E}G_{n,\gamma\gamma,1}[\hat{\gamma}_n^i, \hat{\gamma}_n^i] + \mathbb{E}G_{n,\gamma\gamma,3}[\hat{\gamma}_n^i, \hat{\gamma}_n^i]$ is asymptotically negligible because

$$
\begin{aligned}
& h_n^d(\mathbb{E}G_{n,\gamma\gamma,1}[\hat{\gamma}_n^i, \hat{\gamma}_n^i] + \mathbb{E}G_{n,\gamma\gamma,3}[\hat{\gamma}_n^i, \hat{\gamma}_n^i]) \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{w_0(u)}{q_n^+(u)^2} e_P'\Gamma_{x,n}^+(u)^{-1}\mathcal{K}_x(v)\Gamma_{x,n}^+(u)^{-1}[\mathcal{K}_t(v)q_0(u+vh_n) - \mathcal{K}_x(v)\Upsilon_n^+(u)]f_0(u+vh_n)dudv \\
&= O(n^{-1}h_n^d + h_n^{P+1}),
\end{aligned}
$$

where the second equality uses $\mathcal{K}_t(v)q_0(u+vh_n) - \mathcal{K}_x(v)\Upsilon_n^+(u) = O(n^{-1}h_n^d + h_n^{P+1})$.

Next, defining $\sigma_t^2(x) = \mathbb{V}[t|x] = q_0(x)(1-q_0(x))$ and using simple algebra, we have

$$
\begin{aligned}
& h_n^d(\mathbb{E}G_{n,\gamma\gamma,5}[\hat{\gamma}_n^i, \hat{\gamma}_n^i] + \mathbb{E}G_{n,\gamma\gamma,7}[\hat{\gamma}_n^i, \hat{\gamma}_n^i] + \mathbb{E}G_{n,\gamma\gamma,9}[\hat{\gamma}_n^i, \hat{\gamma}_n^i]) \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{w_0(u)}{q_n^+(u)^3} e_P'\Gamma_{x,n}^+(u)^{-1}\mathcal{K}_t(v)\sigma_t^2(u+vh_n)\mathcal{K}_t(v)'\Gamma_{x,n}^+(u)^{-1}e_P f_0(u+vh_n)dudv \\
&+ \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{w_0(u)}{q_n^+(u)^3} e_P'\Gamma_{x,n}^+(u)^{-1}\mathcal{K}_t(v)\left[q_0(u+vh_n) - \Upsilon_n^+(u)'b_P(v)\right]^2\mathcal{K}_t(v)'\Gamma_{x,n}^+(u)^{-1}e_P f_0(u+vh_n)dudv \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{w_0(u)}{q_n^+(u)^3} e_P'\Gamma_{x,n}^+(u)^{-1}\mathcal{K}_t(v)\sigma_t^2(u+vh_n)\mathcal{K}_t(v)'\Gamma_{x,n}^+(u)^{-1}e_P f_0(u+vh_n)dudv + O(h_n^{P+1}) \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{r_0(u;\theta_0)f_0(u)}{q_0(u)^2} e_P'\Gamma_{x,n}(u)^{-1}\mathcal{K}_t(v)\mathcal{K}_t(v)'\Gamma_{x,n}(u)^{-1}e_P\sigma_t^2(u+vh_n)f_0(u+vh_n)dudv \\
&+ O(n^{-1} + h_n^{P+1}),
\end{aligned}
$$

where the second equality uses $q_0(u+vh_n) - \Upsilon_n^+(u)'b_P(v) = O(h_n^{P+1})$.

Putting the results together, we have

$$
\begin{aligned}
\mathcal{B}_n^{\mathtt{NL}} &= \frac{1}{2n}\mathbb{E}G_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] \\
&= \frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{r_0(u;\theta_0)f_0(u)}{q_0(u)^2} e_P'\Gamma_{x,n}(u)^{-1}\mathcal{K}_t(v)\mathcal{K}_t(v)'\Gamma_{x,n}(u)^{-1}e_P\sigma_t^2(u+vh_n)f_0(u+vh_n)dudv \\
&\quad + o(n^{-1/2}).
\end{aligned}
$$

In particular, we can set

$$
\begin{aligned}
\mathcal{B}_n &= -\frac{K(0)}{nh_n^d}(e_P'\Gamma_x^{-1}e_P)\int_{\mathbb{X}} \frac{1-q_0(u)}{q_0(u)} r_0(u;\theta_0)du \\
&\quad + \frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{r_0(u;\theta_0)f_0(u)}{q_0(u)^2} e_P'\Gamma_{x,n}(u)^{-1}\mathcal{K}_t(v)\mathcal{K}_t(v)'\Gamma_{x,n}(u)^{-1}e_P\sigma_t^2(u+vh_n)f_0(u+vh_n)dudv
\end{aligned}
$$

in Condition AN.

In summary, if $nh_n^{2P+2} \to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$, then the conditions of Theorem 1 are satisfied and $\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ holds with $\Sigma_0 = \mathbb{V}[\psi_0(z)]$ and $\mathfrak{B}_n = O(n^{-1}h_n^{-d})$.

SA.2.4 Bandwidth Selection

As in the previous example, and assuming $\mathcal{B}_0^{\mathtt{SB}} \neq 0$ and $\mathcal{B}_0^{\mathtt{S}} \neq 0$, we can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

$$h_{\mathtt{opt}} = \begin{cases} \left(\frac{|\mathcal{B}_0^{\mathtt{SB}}|}{|\mathcal{B}_0^{\mathtt{S}}|} \frac{1}{n} \right)^{\frac{1}{P+1+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}}) \neq \operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \\ \left(\frac{d}{P+1} \frac{|\mathcal{B}_0^{\mathtt{SB}}|}{|\mathcal{B}_0^{\mathtt{S}}|} \frac{1}{n} \right)^{\frac{1}{P+1+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}}) = \operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \end{cases},$$

where, defining $\Lambda_x = \int_{\mathbb{R}^d} \mathcal{K}_t(u)\mathcal{K}_t(u)'du$,

$$\mathcal{B}_0^{\mathtt{SB}} = -\left[K(0)\left(e_P'\Gamma_x^{-1}e_P\right) - (e_P'\Gamma_x^{-1}\Lambda_x\Gamma_x^{-1}e_P)\right] \int_{\mathbb{X}} \frac{1-q_0(x)}{q_0(x)} r_0(x;\theta_0)dx$$

and

$$\mathcal{B}_0^{\mathtt{S}} = -(e_P'\Gamma_x^{-1}\vartheta_{P+1}) \int_{\mathbb{R}^d} q_0^{(P+1)}(u) \frac{r_0(u;\theta_0)}{q_0(u)} f_0(u)du.$$

SA.2.5 Condition AL*

We apply Lemma 4 with $\rho = 3$ to verify Condition AL*, following as closely as possible our calculations above for Lemma 1.

Conditions (i*)-(ii*). Are verified exactly like their counterparts in Lemma 1 were verified above.

Condition (iii*). We have

$$\begin{aligned} &\|\hat{G}_n^*(\theta,\hat{\gamma}_n^*) - G(\theta,\hat{\gamma}_n^*) - \hat{G}_n^*(\theta_0,\hat{\gamma}_n^*) + G(\theta_0,\hat{\gamma}_n^*)\| \\ &\leq \Delta_{1,n}^*(\theta) + \Delta_{1,n}(\theta) + \rho_n^*\Delta_{2,n}^*(\theta) + \rho_n^*\Delta_{2,n}(\theta) + \Delta_{3,n}^*(\theta) + \rho_n^*\Delta_{4,n}(\theta), \end{aligned}$$

where

$$\rho_n^* = \max_{1\leq i\leq n} \frac{|\hat{q}_n^*(x_{i,n}^*) - q_n(x_{i,n}^*)|}{|\hat{q}_n^*(x_{i,n}^*)q_n(x_{i,n}^*)|} = o_{\mathbb{P}}(n^{-1/2})$$

and

$$\Delta_{1,n}^*(\theta) = \left\| \frac{1}{n}\sum_{i=1}^n \frac{t_{i,n}^*}{q_n(x_{i,n}^*)}(m(y_{i,n}^*;\theta) - m(y_{i,n}^*;\theta_0)) - \mathbb{E}^*\left[\frac{t_{i,n}^*}{q_n(x_{i,n}^*)}(m(y_{i,n}^*;\theta) - m(y_{i,n}^*;\theta_0))\right] \right\|,$$

$$\Delta_{2,n}^*(\theta) = \left| \frac{1}{n}\sum_{i=1}^n t_{i,n}^*\|m(y_{i,n}^*;\theta) - m(y_{i,n}^*;\theta_0)\| - \mathbb{E}^*[t_{i,n}^*\|m(y_{i,n}^*;\theta) - m(y_{i,n}^*;\theta_0)\|] \right|,$$

$$\Delta_{3,n}^*(\theta) = \int_{\mathbb{R}^d} \frac{|\hat{q}_n^*(x) - q_n(x)|}{|\hat{q}_n^*(x)q_n(x)|}\|r_0(x;\theta) - r_0(x;\theta_0)\|f_0(x)dx.$$

In what follows, suppose $\delta_n = o(1)$. First,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta^*_{1,n}(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

because q_n is non-random and bounded away from zero for all n large enough, and because the class of n-varying functions $\mathcal{M}_n = \{tm(y;\theta)/q_n(x) : \theta \in \Theta\}$ satisfies $J_{[],2}(\epsilon_n, \mathcal{M}_n) \to 0$ for all $\epsilon_n \downarrow 0$.

Next, using $\rho^*_n = o_{\mathbb{P}}(1)$,

$$\rho^*_n \sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta^*_{2,n}(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

because the class of functions $\mathcal{M}_{\|\cdot\|} = \{t\,\|m(y;\theta) - m(y;\theta_0)\| : \theta \in \Theta\}$ satisfies $J_{[],2}(1, \mathcal{M}_{\|\cdot\|}) < \infty$.

Finally,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\Delta^*_{3,n}(\theta)}{1 + n^{1/3}\|\theta-\theta_0\|} = O_{\mathbb{P}}(n^{-1/3}) \int_{\mathbb{R}^d} |\hat{q}^*_n(u) - q_n(u)| f_0(u)du = o_{\mathbb{P}}(n^{-1/2})$$

because $\sup_{x\in\mathbb{X}} |\hat{q}^*_n(x) - q_n(x)| = o_{\mathbb{P}}(1)$, q_n is bounded away from zero for all n large enough, $\sup_{x\in\mathbb{X}} \|r_0(x;\theta) - r_0(x;\theta_0)\| = O(\|\theta-\theta_0\|)$, and, using standard results for local polynomial regression estimators and the assumed bandwidth rate restrictions,

$$\int_{\mathbb{R}^d} |\hat{q}^*_n(u) - q_n(u)|^2 f_0(u)du = O_{\mathbb{P}}(n^{-1}h_n^{-d}) = o_{\mathbb{P}}(n^{-1/3}).$$

Condition (iv*). Follows directly by the results established in the following sections, because

$$\begin{aligned}
\hat{G}^*_n(\theta_0, \hat{\gamma}^*_n) &= \frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}^*_n(x^*_{i,n})} m(y^*_{i,n};\theta_0) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}_n(x^*_{i,n})} m(y^*_{i,n};\theta_0) - \frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}_n(x^*_{i,n})^2} m(y^*_{i,n};\theta_0)[\hat{q}^*_n(x^*_{i,n}) - \hat{q}_n(x^*_{i,n})] \\
&\quad + \frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}_n(x^*_{i,n})^3} m(y^*_{i,n};\theta_0)[\hat{q}^*_n(x^*_{i,n}) - \hat{q}_n(x^*_{i,n})]^2 \\
&\quad - \frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}_n(x^*_{i,n})^3 \hat{q}^*_n(x^*_{i,n})} m(y^*_{i,n};\theta_0)[\hat{q}^*_n(x^*_{i,n}) - \hat{q}_n(x^*_{i,n})]^3,
\end{aligned}$$

where

$$\frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}_n(x^*_{i,n})} m(y^*_{i,n};\theta_0) = O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(n^{-1/3}),$$

$$\frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}_n(x^*_{i,n})^2} m(y^*_{i,n};\theta_0)[\hat{q}^*_n(x^*_{i,n}) - \hat{q}_n(x^*_{i,n})] = O_{\mathbb{P}}(n^{-1}h_n^{-d} + n^{-1/2}) = o_{\mathbb{P}}(n^{-1/3}),$$

$$\frac{1}{n}\sum_{i=1}^n \frac{t^*_{i,n}}{\hat{q}_n(x^*_{i,n})^3} m(y^*_{i,n};\theta_0)[\hat{q}^*_n(x^*_{i,n}) - \hat{q}_n(x^*_{i,n})]^2 = O_{\mathbb{P}}(n^{-1}h_n^{-d}) + o_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(n^{-1/3}),$$

$$
\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{t_{i,n}^*}{\hat{q}_n(x_{i,n}^*)^3\hat{q}_n^*(x_{i,n}^*)}m(y_{i,n}^*;\theta_0)[\hat{q}_n^*(x_{i,n}^*)-\hat{q}_n(x_{i,n}^*)]^3\right\| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\frac{t_{i,n}^*}{|\hat{q}_n(x_{i,n}^*)^3\hat{q}_n^*(x_{i,n}^*)|}\|m(y_{i,n}^*;\theta_0)\|\|\hat{q}_n^*(x_{i,n}^*)-\hat{q}_n(x_{i,n}^*)\|^3 \\
&= O_{\mathbb{P}}(1)\max_{1\leq i\leq n}\|\hat{\gamma}_n^*(x_i)-\hat{\gamma}_n(x_i)\|^3 = O_{\mathbb{P}}(\sqrt{(\log n)^3}/\sqrt{n^3h_n^{3d}}) = o_{\mathbb{P}}(n^{-1/3}),
\end{aligned}
$$

provided that $nh_n^{3d/2}/(\log n)^{3/2}\to\infty$.

Conditions (v*)-(vii*). Are verified exactly like their counterparts in Lemma 1 were verified above.

SA.2.6 Condition AS*

A quadratic approximation to g_n^* is given by

$$
\bar{g}_n^*(z,\gamma) = g_n^*(z,\hat{\gamma}_n) + g_{n,\gamma}^*(z)[\gamma-\hat{\gamma}_n] + \frac{1}{2}g_{n,\gamma\gamma}^*(z)[\gamma-\hat{\gamma}_n,\gamma-\hat{\gamma}_n],
$$

where

$$
g_n^*(z,\hat{\gamma}_n) = \frac{t}{\hat{q}_n^+(x)}m(y;\theta_0), \qquad \hat{q}_n^+(x) = e_P'\hat{\Upsilon}_n^+(x),
$$

$$
\hat{\Upsilon}_n^+(x) = \hat{\Gamma}_{x,n}^+(x)^{-1}\hat{\gamma}_{t,n}^+(x), \qquad \hat{\Gamma}_{x,n}^+(x) = \mathrm{vec}_P^{-1}(\hat{\gamma}_{x,n}^+(x)),
$$

$$
\hat{\gamma}_{x,n}^+(x) = (n-1)^{-1}K_n(0)\,\mathrm{vec}_P(e_Pe_P') + \hat{\gamma}_{x,n}(x), \qquad \hat{\gamma}_{t,n}^+(x) = (n-1)^{-1}K_n(0)e_P + \hat{\gamma}_{t,n}(x),
$$

and where, for $\eta = (\eta_x',\eta_t')'$ and $\varphi = (\varphi_x',\varphi_t')'$, the linear and quadratic terms are of the form

$$
g_{n,\gamma}^*(z)[\eta] = -\frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2}e_P'\hat{\Gamma}_{x,n}^+(x)^{-1}\eta_t(x) + \frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2}e_P'\hat{\Gamma}_{x,n}^+(x)^{-1}\,\mathrm{vec}_P^{-1}(\eta_x(x))\hat{\Upsilon}_n^+(x)
$$

and

$$
g_{n,\gamma\gamma}^*(z)[\eta,\varphi] = \sum_{\ell=1}^{10} g_{n,\gamma\gamma,\ell}^*(z)[\eta,\varphi],
$$

respectively, with

$$
g_{n,\gamma\gamma,1}^*(z)[\eta,\varphi] = -\frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2}e_P'\hat{\Gamma}_{x,n}^+(x)^{-1}\,\mathrm{vec}_P^{-1}(\eta_x(x))\hat{\Gamma}_{x,n}^+(x)^{-1}\,\mathrm{vec}_P^{-1}(\varphi_x(x))\hat{\Upsilon}_n^+(x),
$$

$$
g_{n,\gamma\gamma,2}^*(z)[\eta,\varphi] = -\frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2}e_P'\hat{\Gamma}_{x,n}^+(x)^{-1}\,\mathrm{vec}_P^{-1}(\varphi_x(x))\hat{\Gamma}_{x,n}^+(x)^{-1}\,\mathrm{vec}_P^{-1}(\eta_x(x))\hat{\Upsilon}_n^+(x)),
$$

$$
g_{n,\gamma\gamma,3}^*(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2}e_P'\hat{\Gamma}_{x,n}^+(x)^{-1}\,\mathrm{vec}_P^{-1}(\eta_x(x))\hat{\Gamma}_{x,n}^+(x)^{-1}\varphi_t(x),
$$

$$
g_{n,\gamma\gamma,4}^*(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2}e_P'\hat{\Gamma}_{x,n}^+(x)^{-1}\,\mathrm{vec}_P^{-1}(\varphi_x(x))\hat{\Gamma}_{x,n}^+(x)^{-1}\eta_t(x),
$$

$$g^*_{n,\gamma\gamma,5}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e_P' \hat{\Gamma}^+_{x,n}(x)^{-1} \eta_t(x) \varphi_t(x)' \hat{\Gamma}^+_{x,n}(x)^{-1} e_P,$$

$$g^*_{n,\gamma\gamma,6}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e_P' \hat{\Gamma}^+_{x,n}(x)^{-1} \varphi_t(x) \eta_t(x)' \hat{\Gamma}^+_{x,n}(x)^{-1} e_P,$$

$$g^*_{n,\gamma\gamma,7}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e_P' \hat{\Gamma}^+_{x,n}(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \hat{\Upsilon}^+_n(x) \hat{\Upsilon}^+_n(x)' \operatorname{vec}_P^{-1}(\varphi_x(x)) \hat{\Gamma}^+_{x,n}(x)^{-1} e_P,$$

$$g^*_{n,\gamma\gamma,8}(z)[\eta,\varphi] = \frac{tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e_P' \hat{\Gamma}^+_{x,n}(x)^{-1} \operatorname{vec}_P^{-1}(\varphi_x(x)) \hat{\Upsilon}^+_n(x) \hat{\Upsilon}^+_n(x)' \operatorname{vec}_P^{-1}(\eta_x(x)) \hat{\Gamma}^+_{x,n}(x)^{-1} e_P,$$

$$g^*_{n,\gamma\gamma,9}(z)[\eta,\varphi] = -\frac{2tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e_P' \hat{\Gamma}^+_{x,n}(x)^{-1} \eta_t(x) \hat{\Upsilon}^+_n(x)' \varphi_x(x) \hat{\Gamma}^+_{x,n}(x)^{-1} e_P,$$

$$g^*_{n,\gamma\gamma,10}(z)[\eta,\varphi] = -\frac{2tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e_P' \hat{\Gamma}^+_{x,n}(x)^{-1} \varphi_t(x) \hat{\Upsilon}^+_n(x)' \eta_x(x) \hat{\Gamma}^+_{x,n}(x)^{-1} e_P,$$

Suppose $h_n \to 0$. The first part of Condition AS* is satisfied if

$$\max_{1\le i,j\le n} \|\hat{\gamma}^{*,(i)}_n(x_j) - \hat{\gamma}_n(x_j)\| = o_{\mathbb{P}}(n^{-1/6}),$$

a sufficient condition for which is that $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

Moreover, the second part of Condition AS* can be verified using Lemma 5, because if i, j, and k are distinct, then

$$\mathbb{V}^*(g^*_{n,\gamma}(z^*_{i,n})[\hat{\gamma}^{*,j}_n - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-d}(1 + n^{-1}h_n^{-d})) = o_{\mathbb{P}}(n),$$

$$\mathbb{V}^*(g^*_{n,\gamma\gamma}(z^*_{i,n})[\hat{\gamma}^{*,j}_n - \hat{\gamma}_n, \hat{\gamma}^{*,k}_n - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-2d}(1 + n^{-1}h_n^{-d})^2) = o_{\mathbb{P}}(n^2),$$

$$\mathbb{V}^*(\mathbb{E}^*(g^*_{n,\gamma\gamma}(z^*_{i,n})[\hat{\gamma}^{*,j}_n - \hat{\gamma}_n, \hat{\gamma}^{*,j}_n - \hat{\gamma}_n]|z^*_{i,n})) = O_{\mathbb{P}}(h_n^{-2d}(1 + n^{-1}h_n^{-d})) = o_{\mathbb{P}}(n^2),$$

$$\mathbb{V}^*(g^*_{n,\gamma\gamma}(z^*_{i,n})[\hat{\gamma}^{*,j}_n - \hat{\gamma}_n, \hat{\gamma}^{*,j}_n - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-3d}(1 + n^{-1}h_n^{-d})) = o_{\mathbb{P}}(n^3),$$

provided $nh_n^d \to \infty$.

In other words, Condition AS* holds if $h_n \to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

SA.2.7 Condition AN*

Suppose $h_n \to 0$ and $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$. Then, using the fact that $\hat{\theta}_n \to_{\mathbb{P}} \theta_0$ and $\max_{1\le i\le n} \|\hat{\gamma}_n(x_i) - \gamma_n(x_i)\| = o_{\mathbb{P}}(1)$, it can be shown that

$$\mathbb{E}^*[\|g^*_n(z^*_{i,n}, \hat{\gamma}_n) - g_n(z^*_{i,n}, \gamma_n)\|^2] = o_{\mathbb{P}}(1), \qquad G^*_n(\hat{\gamma}_n) - G_n(\gamma_n) = o_{\mathbb{P}}(1).$$

Also, the correction term

$$\begin{aligned}\delta_n^*(z_{i,n}^*) &= -\frac{1}{n}\sum_{j=1}^{n}\frac{t_j m(y_j;\theta_0)}{\hat{q}_n^+(x_j)^2}e_P'\hat{\Gamma}_{x,n}^+(x_j)^{-1}\left[\mathcal{K}_{t,n}(x_{i,n}^*-x_j)t-\hat{\gamma}_{t,n}(x_j)\right]\\ &\quad+\frac{1}{n}\sum_{j=1}^{n}\frac{t_j m(y_j;\theta_0)}{\hat{q}_n^+(x_j)^2}e_P'\hat{\Gamma}_{x,n}^+(x_j)^{-1}\left[\mathcal{K}_{x,n}(x_{i,n}^*-x_j)-\hat{\Gamma}_{x,n}(x_j)\right]\hat{\Upsilon}_n^+(x_j)\end{aligned}$$

can be shown to satisfy

$$\mathbb{E}^*[\left\|\delta_n^*(z_{i,n}^*)-\delta_n(z_{i,n}^*)\right\|^2]=o_{\mathbb{P}}(1).$$

As a consequence,

$$\mathbb{E}^*[\|\psi_n^*(z_{i,n}^*)-\psi_n(z_{i,n}^*)\|^2]=o_{\mathbb{P}}(1),\qquad \psi_n^*(z_{i,n}^*)=g_n^*(z_{i,n}^*,\hat{\gamma}_n)-G_n^*(\hat{\gamma}_n)+\delta_n^*(z_{i,n}^*).$$

Moreover, if i and j are distinct, then

$$\mathbb{V}^*(G_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,i}-\hat{\gamma}_n,\hat{\gamma}_n^{*,i}-\hat{\gamma}_n])=O_{\mathbb{P}}(h_n^{-2d})=o_{\mathbb{P}}(n^2),\qquad \mathbb{V}^*(G_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,i}-\hat{\gamma}_n,\hat{\gamma}_n^{*,j}-\hat{\gamma}_n])=O_{\mathbb{P}}(h_n^{-d})=o_{\mathbb{P}}(n).$$

Finally, it can be shown that

$$\hat{\mathcal{B}}_n^*=\mathbb{E}^*\hat{\mathcal{B}}_n^*+o_{\mathbb{P}}(n^{-1/2})=\mathcal{B}_n^*+o_{\mathbb{P}}(n^{-1/2}),$$

where

$$\begin{aligned}\mathcal{B}_n^* &= -\frac{K(0)}{nh_n^d}(e_P'\Gamma_x^{-1}e_P)\int_{\mathbb{X}}\frac{1-q_0(u)}{q_0(u)}r_0(u;\theta_0)du\\ &\quad+\frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{r_0(u;\theta_0)f_0(u)}{q_0(u)^2}e_P'\Gamma_{x,n}(u)^{-1}\mathcal{K}_t(v)\mathcal{K}_t(v)'\Gamma_{x,n}(u)^{-1}e_P\sigma_t^2(u+vh_n)f_0(u+vh_n)dudv.\end{aligned}$$

In other words, if $h_n\to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2}\to\infty$, then the assumptions of Lemma 6 are satisfied and Condition AN* holds with $\Omega_0^*=\Omega_0$.

In summary, if $nh_n^{2P+2}\to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2}\to\infty$, then the conditions of Theorem 2 are satisfied.

SA.3 Example 3: Hit Rate

This example is Example 1 in Chen, Linton, and van Keilegom (2003), which corresponds to a particular instance of a so-called 'Hit Rate'. While simple in many respects, this example is interesting because it allows us to compare our results with previous influential work in a tractable setting, where the semiparametric estimator $\hat{\theta}_n$ is given in closed form but it involves a discontinuous functional of a kernel density estimator $\hat{\gamma}_n$. Thus, we illustrate how Condition AS (and AS*) can be verified in a non-smooth example to construct valid, more robust inference procedures, where standard empirical process methods cannot be applied to obtain asymptotic normality for two-step kernel-based semiparametric estimators when $\mathfrak{B}_n \neq o(n^{-1/2})$.

Suppose $z_1, \ldots, z_n$ are $i.i.d.$ copies of $z = (y, x')'$, where $y \in \mathbb{R}$ is a scalar and the vector $x \in \mathbb{R}^d$ is a continuous explanatory variable with density γ_0. Letting $\mathbb{1}(\cdot)$ denote the indicator function, the estimand is

$$\theta_0 = \mathbb{P}[y \geq \gamma_0(x)] = \mathbb{E}(\mathbb{1}[y \geq \gamma_0(x)]),$$

and the corresponding estimator is

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[y_i \geq \hat{\gamma}_n(x_i)], \qquad \hat{\gamma}_n(x) = \frac{1}{n}\sum_{j=1}^{n} K_n(x - x_j), \qquad K_n(u) = \frac{1}{h_n^d} K\left(\frac{u}{h_n}\right).$$

To study this estimator using our main results, we set $d_\gamma = 1,\ x(z, \theta) = z,\ w(z, \theta) = 1$, and let $\hat{\theta}_n$ be defined by $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = 0$, where $g(z, \theta, \gamma) = \mathbb{1}[y \geq \gamma(z)] - \theta$.

In what follows, we assume that for some $P > 3d/4$, the following regularity conditions hold:

- γ_0 is $P + 1$ times differentiable, and γ_0 and its first $P + 1$ derivatives are bounded and continuous.
- $F_{y|x}(\cdot|x)$, the conditional cdf of y given x, has three bounded (uniformly in x) derivatives.
- K is even and bounded with $\int_{\mathbb{R}^d} |K(u)|(1 + \|u\|^{P+1})du < \infty$ and

$$\int_{\mathbb{R}^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du = \begin{cases} 1, & \text{if } l_1 = \cdots = l_d = 0, \\ 0, & \text{if } (l_1, \ldots, l_d)' \in \mathbb{Z}_+^d \text{ and } l_1 + \cdots + l_d < P \end{cases}.$$

As in the average density example, the smoothness assumptions can be relaxed, but once again the stated assumption is sufficient for our purposes.

If the bandwidth satisfies $nh_n^{2P} \to 0$ and $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$, we show here that the assumptions of Theorems 1 and 2 are satisfied and that, for some Σ_0,

$$\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0).$$

where $\mathfrak{B}_n = O(n^{-1}h_n^{-d})$. Once again, the stated bandwidth conditions are weak enough to permit $\mathfrak{B}_n \neq o(n^{-1/2})$, while $\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ reduces to $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ when imposing conditions requiring $nh_n^{2d} \to \infty$.

SA.3.1 Condition AL

Condition AL holds with $\mathcal{J}_n = \mathcal{J}_0 = 1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.3.2 Condition AS

Define

$$\breve{g}_n(x,\gamma)=\mathbb{E}[g_n(z,\gamma)|x]-(1-\theta_0)=-F_{y|x}[n^{-1}K_n(0)+(1-n^{-1})\gamma(x)|x].$$

Being a defined through a projection, $\breve{g}_n(x,\gamma)$ is likely to be close to $g_n(z,\gamma)$ in the the appropriate sense and, indeed,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\hat{\gamma}_n^{(i)})-\breve{g}_n(x_i,\hat{\gamma}_n^{(i)})-g_n(z_i,\gamma_n)+\breve{g}_n(x_i,\gamma_n)]=o_{\mathbb{P}}(1)$$

if $\Delta_n=\max_{1\leq i\leq n}|\hat{\gamma}_n^{(i)}(x_i)-\gamma_n(x_i)|=o_{\mathbb{P}}(1)$, because then

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\hat{\gamma}_n^{(i)})-\breve{g}_n(x_i,\hat{\gamma}_n^{(i)})-g_n(z_i,\gamma_n)+\breve{g}_n(x_i,\gamma_n)]\right)^2|\mathcal{X}_n\right]\\ =\;&\frac{1}{n}\mathbb{V}\left(\sum_{i=1}^{n}[g_n(z_i,\hat{\gamma}_n^{(i)})-g_n(z_i,\gamma_n)]|\mathcal{X}_n\right)\leq\sup_{r,s}f_{y|x}(r|s)\Delta_n=o_{\mathbb{P}}(1),\end{aligned}$$

where $\mathcal{X}_n=(x_1,\ldots,x_n)'$, $f_{y|x}(\cdot|x)$ denotes the derivative of $F_{y|x}(\cdot|x)$, and

$$\hat{\gamma}_n^{(i)}(x)=\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}K_n(x-x_j).$$

Next, being smooth $\breve{g}_n(x,\gamma)$ admits the quadratic approximation

$$\bar{g}_n(x,\gamma)=\breve{g}_n(x,\gamma_n)+\breve{g}_{n,\gamma}(x)[\gamma-\gamma_n]+\frac{1}{2}\breve{g}_{n,\gamma\gamma}(x)[\gamma-\gamma_n,\gamma-\gamma_n],$$

where, letting $\dot{f}_{y|x}(\cdot|x)$ denote the derivative of $f_{y|x}(\cdot|x)$,

$$\breve{g}_{n,\gamma}(x)[\gamma]=-(1-n^{-1})f_{y|x}[\gamma_n^{+}(x)|x]\gamma(x),\qquad\gamma_n^{+}(x)=n^{-1}K_n(0)+(1-n^{-1})\gamma_n(x),$$

$$\breve{g}_{n,\gamma\gamma}(x)[\gamma,\eta]=-(1-n^{-1})^2\dot{f}_{y|x}[\gamma_n^{+}(x)|x]\gamma(x)\eta(x).$$

It follows from standard bounding arguments that if $\Delta_n=o_{\mathbb{P}}(n^{-1/6})$, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\breve{g}_n(x_i,\hat{\gamma}_n^{(i)})-\bar{g}_n(x_i,\hat{\gamma}_n^{(i)})-\breve{g}_n(x_i,\gamma_n)+\bar{g}_n(x_i,\gamma_n)]=o_{\mathbb{P}}(1).$$

As a consequence, the first part of Condition AS is satisfied if $\Delta_n=o_{\mathbb{P}}(n^{-1/6})$. By Lemma SA-1, the latter holds provided that $nh_n^{3d/2}/(\log n)^{3/2}\to\infty$. Moreover, if i, j, and k are distinct, then

$$\mathbb{V}(\breve{g}_{n,\gamma}(x_i)[\hat{\gamma}_n^{j}-\gamma_n])=O(h_n^{-d}),\qquad\mathbb{V}(\breve{g}_{n,\gamma\gamma}(x_i)[\hat{\gamma}_n^{j}-\gamma_n,\hat{\gamma}_n^{k}-\gamma_n])=O(h_n^{-2d}),$$

$$\mathbb{V}(\mathbb{E}(\breve{g}_{n,\gamma\gamma}(x_i)[\hat{\gamma}_n^j-\gamma_n,\hat{\gamma}_n^j-\gamma_n]|z_i))=O(h_n^{-2d}),\qquad \mathbb{V}(\breve{g}_{n,\gamma\gamma}(x_i)[\hat{\gamma}_n^j-\gamma_n,\hat{\gamma}_n^j-\gamma_n])=O(h_n^{-3d}),$$

and hence it follows from Lemma 2 that the second part of Condition AS holds provided $nh_n^d\to\infty$.

SA.3.3 Condition AN

We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\gamma_n)+\bar{G}_n(\hat{\gamma}_n^{(i)})-\bar{G}_n(\gamma_n)]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n(z_i)+\sqrt{n}\hat{\mathcal{B}}_n,$$

where

$$\psi_n(z)=g_n(z,\gamma_n)-G_n(\gamma_n)+\delta_n(x),\qquad \delta_n(x)=-(1-n^{-1})\int_{\mathbb{R}^d}f_{y|x}[\gamma_n^+(u)|u][K_n(u-x)-\gamma_n(u)]\gamma_0(u)du,$$

$$\hat{\mathcal{B}}_n=G_n(\gamma_n)+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\breve{G}_{n,\gamma\gamma}[\hat{\gamma}_n^{(i)}-\gamma_n,\hat{\gamma}_n^{(i)}-\gamma_n].$$

Suppose $h_n\to 0$ and $nh_n^d\to\infty$. Then

$$\psi_n(z)\to\psi_0(z)=g_0(z,\gamma_0)+\delta_0(x),\qquad \delta_0(x)=-f_{y|x}[\gamma_0(x)|x]\gamma_0(x)+\int_{\mathbb{R}^d}f_{y|x}[\gamma_0(u)|u]\gamma_0(u)^2du,$$

for every z, and it follows from the dominated convergence theorem that $\mathbb{E}[|\psi_n(x)-\psi_0(x)|^2]\to 0$. Also, the representation

$$\breve{G}_{n,\gamma\gamma}[\eta,\varphi]=-(1-n^{-1})^2\int_{\mathbb{R}^d}\dot{f}_{y|x}[\gamma_n^+(u)|u]\eta(u)\varphi(u)\gamma_0(u)du,$$

can be used to show that if i and j are distinct, then

$$\mathbb{V}(\breve{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n])=O(h_n^{-2d})=o(n^{-2}),\qquad \mathbb{V}(\breve{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^j-\gamma_n])=O(h_n^{-d})=o(n^{-1}),$$

so it follows from Lemma 3 that Condition AN is satisfied with $\Omega_0=\mathbb{V}[\psi_0(z)]$.

Under the bandwidth conditions imposed it can also be shown that

$$\mathcal{B}_n^{\mathtt{LI}}=-\frac{1}{nh_n^d}K(0)\int_{\mathbb{R}^d}f_{y|x}[\gamma_0(u)|u]\gamma_0(u)du+o(n^{-1/2}),$$

$$\mathcal{B}_n^{\mathtt{NL}}=-\frac{1}{nh_n^d}\frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\dot{f}_{y|x}[\gamma_0(v)|v]K(u)^2\gamma_0(v)\gamma_0\left(v-uh_n\right)dvdu+o(n^{-1/2}),$$

and

$$\mathcal{B}_n^{\mathtt{S}}=h_n^P\mathcal{B}_0^{\mathtt{S}}+o(h_n^P),\qquad \mathcal{B}_0^{\mathtt{S}}=(-1)^{P+1}\sum_{|p|=P}\frac{1}{p!}\left(\int_{\mathbb{R}}u^pK(u)du\right)\left(\int_{\mathbb{R}^d}f_{y|x}(\gamma_0(u)|u)\gamma_0(u)\left(\partial^p\gamma_0(u)\right)du\right).$$

As a consequence, we can set

$$\mathcal{B}_n = -\frac{1}{nh_n^d}\left(K(0)\int_{\mathbb{R}^d} f_{y|x}[\gamma_0(u)|u]\gamma_0(u)du + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\dot{f}_{y|x}[\gamma_0(v)|v]K(u)^2\gamma_0(v)\gamma_0\left(v-uh_n\right)dvdu\right).$$

In summary, if $nh_n^{2P}\to 0$ and if $nh_n^{3d/2+1}/(\log n)^{3/2}\to\infty$, then the conditions of Theorem 1 are satisfied and $\sqrt{n}(\hat{\theta}_n-\theta_0-\mathfrak{B}_n)\rightsquigarrow\mathcal{N}(0,\Sigma_0)$ holds with $\Sigma_0=\mathbb{V}[\psi_0(z)]$ and $\mathfrak{B}_n=O(n^{-1}h_n^{-d})$.

SA.3.4 Bandwidth Selection

As in the previous example, and assuming $\mathcal{B}_0^{\mathtt{SB}}\neq 0$ and $\mathcal{B}_0^{\mathtt{S}}\neq 0$, we can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

$$h_{\mathtt{opt}} = \begin{cases} \left(\frac{\left|\mathcal{B}_0^{\mathtt{SB}}\right|}{\left|\mathcal{B}_0^{\mathtt{S}}\right|}\frac{1}{n}\right)^{\frac{1}{P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}})\neq\operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \\ \left(\frac{d}{P}\frac{\left|\mathcal{B}_0^{\mathtt{SB}}\right|}{\left|\mathcal{B}_0^{\mathtt{S}}\right|}\frac{1}{n}\right)^{\frac{1}{P+d}} & \text{if } \operatorname{sgn}(\mathcal{B}_0^{\mathtt{SB}})=\operatorname{sgn}(\mathcal{B}_0^{\mathtt{S}}) \end{cases},$$

where

$$\mathcal{B}_0^{\mathtt{SB}} = -K(0)\int_{\mathbb{R}^d} f_{y|x}[\gamma_0(u)|u]\gamma_0(u)du - \frac{1}{2}\left(\int_{\mathbb{R}^d}K(u)^2du\right)\left(\int_{\mathbb{R}^d}\dot{f}_{y|x}[\gamma_0(u)|u]\gamma_0(u)^2du\right).$$

SA.3.5 Condition AL*

Condition AL* holds with $\mathcal{J}_n^*=\mathcal{J}_0^*=1$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term.

SA.3.6 Condition AS*

Let $\breve{g}_n^*(x,\gamma)=\breve{g}_n(x,\gamma)$ and define

$$\bar{g}_n^*(x,\gamma)=\breve{g}_n^*(x,\hat{\gamma}_n)+\breve{g}_{n,\gamma}^*(x)[\gamma-\hat{\gamma}_n]+\frac{1}{2}\breve{g}_{n,\gamma\gamma}^*(x)[\gamma-\hat{\gamma}_n,\gamma-\hat{\gamma}_n],$$

where

$$\breve{g}_{n,\gamma}^*(x)[\eta]=-(1-n^{-1})f_{y|x}[\hat{\gamma}_n^+(x)|x]\eta(x),\qquad \hat{\gamma}_n^+(x)=n^{-1}K_n(0)+(1-n^{-1})\hat{\gamma}_n(x),$$

$$\breve{g}_{n,\gamma\gamma}^*(x)[\eta,\varphi]=-(1-n^{-1})^2\dot{f}_{y|x}[\hat{\gamma}_n^+(x)|x]\eta(x)\varphi(x).$$

Defining $N_i=\sum_{j=1}^n\mathbb{1}(x_j^*=x_i)$ and using the fact (about the multinomial distribution) that $n^{-1}\sum_{i=1}^n N_i^2=O_{\mathbb{P}}(1)$, it can be shown that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n[g_n^*(z_{i,n}^*,\hat{\gamma}_n^{*,(i)})-\breve{g}_n^*(x_{i,n}^*,\hat{\gamma}_n^{*,(i)})-g_n^*(z_{i,n}^*,\hat{\gamma}_n)+\breve{g}_n^*(x_{i,n}^*,\hat{\gamma}_n)]=o_{\mathbb{P}}(1)$$

if $\Delta_n^* = o_{\mathbb{P}}(1)$, because then

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_{i,n}^*,\hat{\gamma}_n^{*,(i)}) - \breve{g}_n^*(x_{i,n}^*,\hat{\gamma}_n^{*,(i)}) - g_n^*(z_{i,n}^*,\hat{\gamma}_n) + \breve{g}_n^*(x_{i,n}^*,\hat{\gamma}_n)]\right)^2 |\mathcal{X}_n, \mathcal{X}_n^*\right]$$

$$= \frac{1}{n}\mathbb{V}\left(\sum_{i=1}^{n}[g_n^*(z_{i,n}^*,\hat{\gamma}_n^{*,(i)}) - g_n^*(z_{i,n}^*,\hat{\gamma}_n)]|\mathcal{X}_n, \mathcal{X}_n^*\right) \leq \sup_{r,s} f_{y|x}(r|s)\left(\frac{1}{n}\sum_{i=1}^{n}N_i^2\right)\Delta_n^* = o_{\mathbb{P}}(1),$$

where $\mathcal{X}_n^* = (x_{1,n}^*, \ldots, x_{n,n}^*)'$. Also, it follows from standard bounding arguments that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\breve{g}_n^*(x_{i,n}^*,\hat{\gamma}_n^{*,(i)}) - \bar{g}_n^*(x_{i,n}^*,\hat{\gamma}_n^{*,(i)}) - \breve{g}_n^*(x_{i,n}^*,\hat{\gamma}_n) + \bar{g}_n^*(x_{i,n}^*,\hat{\gamma}_n)] = o_{\mathbb{P}}(1)$$

provided $\Delta_n^* = o_{\mathbb{P}}(n^{-1/6})$. The latter rate condition holds when $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$, as can be verified using Lemma SA-1. As a consequence, the first part of Condition AS* is satisfied when $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$ and $h_n \to 0$. Moreover, it can be shown that if i, j, and k are distinct, then

$$\mathbb{V}^*(\breve{g}_{n,\gamma}^*(x_{i,n}^*)[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-d}),$$

$$\mathbb{V}^*(\breve{g}_{n,\gamma\gamma}^*(x_{i,n}^*)[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,k} - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-2d}),$$

$$\mathbb{V}^*(\mathbb{E}^*(\breve{g}_{n,\gamma\gamma}^*(x_{i,n}^*)[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,j} - \hat{\gamma}_n]|x_{i,n}^*)) = O_{\mathbb{P}}(h_n^{-2d}),$$

$$\mathbb{V}^*(\breve{g}_{n,\gamma\gamma}^*(x_{i,n}^*)[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,j} - \hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-3d}),$$

so it follows from Lemma 5 that the second part of Condition AS* will be satisfied provided $nh_n^d \to \infty$.

SA.3.7 Condition AN*

We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{\gamma}_n) + \bar{G}_n^*(\hat{\gamma}_n^{*,(i)}) - \bar{G}_n^*(\hat{\gamma}_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n^*(z_i^*) + \sqrt{n}\hat{\mathcal{B}}_n^*,$$

where

$$\psi_n^*(z) = g_n^*(z,\hat{\gamma}_n) - G_n^*(\hat{\gamma}_n) + \delta_n^*(z),$$

$$\delta_n^*(z_{i,n}^*) = \breve{G}_{n,\gamma}^*[\hat{\gamma}_n^{*,i} - \hat{\gamma}_n] = -(1-n^{-1})\frac{1}{n}\sum_{i=1}^{n}f_{y|x}[\hat{\gamma}_n^+(x_i)|x_i][K_n(x_i - x_{i,n}^*) - \hat{f}_n(x_i)],$$

$$\hat{\mathcal{B}}_n^* = G_n^*(\hat{\gamma}_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\breve{G}_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,(i)} - \hat{\gamma}_n, \hat{\gamma}_n^{*,(i)} - \hat{\gamma}_n].$$

Suppose $h_n \to 0$ and $nh_n^{\frac{3}{2}d}/(\log n)^{3/2} \to \infty$. Using Lemma A-2 and the fact that $\hat{\theta}_n \to_{\mathbb{P}} \theta_0$, it can be shown that

$$\mathbb{E}^*[|\psi_n^*(z_{i,n}^*) - \psi_n(z_{i,n}^*)|^2] = o_{\mathbb{P}}(1).$$

Also, the representation

$$\breve{G}^*_{n,\gamma\gamma}[\eta,\varphi] = -(1-n^{-1})^2 \frac{1}{n}\sum_{i=1}^{n} \dot{f}_{y|x}[\hat{\gamma}_n^+(x_i)|x_i]\eta(x_i)\varphi(x_i)$$

can be used to show that if i and j are distinct, then

$$\mathbb{V}^*(\breve{G}^*_{n,ff}[\hat{\gamma}_n^{*,i}-\hat{\gamma}_n,\hat{\gamma}_n^{*,i}-\hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-2d}), \qquad \mathbb{V}^*(\breve{G}^*_{n,ff}[\hat{\gamma}_n^{*,i}-\hat{\gamma}_n,\hat{\gamma}_n^{*,j}-\hat{\gamma}_n]) = O_{\mathbb{P}}(h_n^{-d}).$$

Finally, it can be shown that

$$\hat{\mathcal{B}}^*_n = \mathbb{E}^*\hat{\mathcal{B}}^*_n + o_{\mathbb{P}}(n^{-1/2}) = \mathcal{B}^*_n + o_{\mathbb{P}}(n^{-1/2}),$$

where

$$\mathcal{B}^*_n = -\frac{1}{nh_n^d}\left(K(0)\int_{\mathbb{R}^d} f_{y|x}[\gamma_0(u)|u]\gamma_0(u)du + \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \dot{f}_{y|x}[\gamma_0(v)|v]K(u)^2\gamma_0(v)\gamma_0\left(v-uh_n\right)dvdu\right).$$

In other words, if $h_n \to 0$ and if $nh_n^{3d/2}/\left(\log n\right)^{3/2} \to \infty$, then the assumptions of Lemma 6 are satisfied and Condition AN* holds with $\Omega_0^* = \Omega_0$.

In summary, if $nh_n^{2P} \to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$, then the conditions of Theorem 2 are satisfied.

SA.4 Uniform Convergence Rates for Kernel-Based Estimators

Various results on uniform convergence rates for kernel-based estimators are used to verify the conditions of Theorems 1 and 2 in the examples. The results utilized are all special cases of Lemma SA-1 below.

Suppose that for every n, $Z_{i,n} = (W_{i,n}, X_{i,n}')'$ $(i = 1,\ldots,n)$ are *i.i.d.* copies of $Z_n = (W_n, X')'$, where W_n is scalar and $X \in \mathbb{R}^d$ is continuous with bounded density f_X. The estimators we consider are of the form

$$\hat{\Psi}_n(x) = \frac{1}{n}\sum_{j=1}^{n} W_{j,n}\mathcal{K}_n(x-X_{j,n}), \qquad \hat{\Psi}_n^{(i)}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} W_{j,n}\mathcal{K}_n(x-X_{j,n}),$$

where $\mathcal{K}_n\left(x\right) = \mathcal{K}(x/h_n)/h_n^d$, $h_n = o\left(1\right)$ is a bandwidth, and $\mathcal{K}$ is a bounded and integrable (kernel-like) function.

Bootstrap analogs of these estimators are also of interest. Letting $\{Z^*_{1,n},\ldots,Z^*_{n,n}\}$ be a random sample with replacement from $\{Z_{1,n},\ldots,Z_{n,n}\}$, define

$$\hat{\Psi}^*_n(x) = \frac{1}{n}\sum_{j=1}^{n} W^*_{j,n}\mathcal{K}_n(x-X^*_{j,n}), \qquad \hat{\Psi}_n^{*,(i)}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} W^*_{j,n}\mathcal{K}_n(x-X^*_{j,n}).$$

Defining $\Psi_n(x) = \mathbb{E}\hat{\Psi}_n(x)$, the objective is to give conditions (on h_n, ρ_n, and the distribution of Z_n) under which

$$\max_{1\leq i\leq n} |\hat{\Psi}_n(X_{i,n}) - \Psi_n(X_{i,n})| = O_p(\rho_n), \tag{SA-1}$$

$$\max_{1\leq i\leq n} |\hat{\Psi}_n^{(i)}(X_{i,n}) - \Psi_n(X_{i,n})| = O_p(\rho_n), \tag{SA-2}$$

$$\max_{1\leq j\leq n} |\hat{\Psi}_n^{*}(X_{j,n}) - \hat{\Psi}_n(X_{j,n})| = O_p(\rho_n), \tag{SA-3}$$

$$\max_{1\leq i,j\leq n} |\hat{\Psi}_n^{*,(i)}(X_{j,n}) - \hat{\Psi}_n(X_{j,n})| = O_p(\rho_n). \tag{SA-4}$$

To give a succinct statement, let $Gam(\cdot)$ be the Gamma function and for $s > 0$, let

$$\mathcal{C}(s) = \sup_{n\geq 1}[\mathbb{E}(|W_n|^s) + \sup_{x\in\mathbb{R}^d} \mathbb{E}(|W_n|^s|X = x)f_X(x)].$$

Lemma SA-1 *(a) If $\mathcal{C}(S) < \infty$ for some $S \geq 2$ and if $n^{1-1/S}h_n^d/\log n \to \infty$, then (SA-1)–(SA-4) hold with $\rho_n = \max(\sqrt{\log n}/\sqrt{nh_n^d}, \log n/(n^{1-1/S}h_n^d))$.*
(b) If $\mathcal{C}(s) \leq Gam(s)H^s$ for some $H < \infty$ and every s and if $\underline{\lim}_{n\to\infty} nh_n^d/(\log n)^3 > 0$, then (SA-1)–(SA-4) hold with $\rho_n = \sqrt{\log n}/\sqrt{nh_n^d}$.
(c) If $\mathcal{C}(s) \leq H^s$ for some $H < \infty$ and every s and if $\underline{\lim}_{n\to\infty} nh_n^d/\log n > 0$, then (SA-1)–(SA-4) hold with $\rho_n = \sqrt{\log n}/\sqrt{nh_n^d}$.

SA.4.1 Proof of Lemma SA-1

For $i = 1, \ldots, n$, we have

$$\hat{\Psi}_n(X_{i,n}) = (1 - n^{-1})\hat{\Psi}_n^{(i)}(X_{i,n}) + n^{-1}\mathcal{K}_n(0)W_{i,n}$$

and therefore

$$\max_{1\leq i\leq n} |\hat{\Psi}_n(X_{i,n}) - \Psi_n(X_{i,n})| \leq \max_{1\leq i\leq n} |\hat{\Psi}_n^{(i)}(X_{i,n}) - \Psi_n(X_{i,n})| + R_n,$$

where

$$R_n = n^{-1}\mathcal{K}_n(0) \max_{1\leq i\leq n} |W_{i,n}| + n^{-1} \sup_{x\in\mathbb{R}^d} |\Psi_n(x)| = O(n^{-1}h_n^{-d}) \max_{1\leq i\leq n} |W_{i,n}| + O(\rho_n)$$

because $n\rho_n \to \infty$ and $\sup_{x\in\mathbb{R}^d} |\Psi_n(x)| \leq \mathcal{C}(1) \int_{\mathbb{R}^d} |\mathcal{K}(u)|du$. By Chebychev's inequality,

$$\mathbb{P}[\max_{1\leq i\leq n} |W_{i,n}| > M\tau_n] \leq n\mathbb{P}[|W_n| > M\tau_n] \leq \frac{n\mathcal{C}(S_n)}{M^{S_n}\tau_n^{S_n}}$$

for every M and every (S_n, τ_n). Therefore, $\max_i |W_{i,n}| = O_p(\tau_n)$ if the $\overline{\lim}_{n\to\infty}$ of the majorant can be made arbitrarily small by choosing S_n appropriately and making M large.

In case (a), setting $(S_n, \tau_n) = (S, n^{1/S})$ we have $\tau_n = O(nh_n^d \rho_n)$ and

$$\frac{n\mathcal{C}(S_n)}{M^{S_n}\tau_n^{S_n}} = \frac{\mathcal{C}(S)}{M^S},$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily small by making M large.

In case (b), setting $(S_n, \tau_n) = (\log n, \log n)$ we have $\tau_n = O(nh_n^d \rho_n)$ and

$$\frac{n\mathcal{C}(S_n)}{M^{S_n}\tau_n^{S_n}} = \frac{n\mathcal{C}(\log n)}{M^{\log n}(\log n)^{\log n}} \leq \frac{nGam(\log n)H^{\log n}}{M^{\log n}(\log n)^{\log n}} = \left(\frac{H}{M}\right)^{\log n} O(1/\sqrt{\log n}),$$

where the second equality uses Stirling's formula and the $\overline{\lim}_{n\to\infty}$ of the majorant can be made arbitrarily small by making M large.

In case (c), setting $(S_n, \tau_n) = (\log n, 1)$ we have $\tau_n = O(nh_n^d \rho_n)$ and

$$\frac{n\mathcal{C}(S_n)}{M^{S_n}\tau_n^{S_n}} = \frac{n\mathcal{C}(\log n)}{M^{\log n}} \leq n\left(\frac{H}{M}\right)^{\log n},$$

where the $\overline{\lim}_{n\to\infty}$ of the majorant can be made arbitrarily small by making M large.

In all cases, $R_n = O_p(\rho_n)$ because $\tau_n/(nh_n^d) = O(\rho_n)$. The proof of (SA-1) can therefore be completed by showing that (SA-2) holds.

Proof of (SA-2). With (S_n, τ_n) as before, let

$$\hat{\Psi}_n^{\tau,(i)}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} W_{j,n}^{\tau}\mathcal{K}_n(x - X_{j,n}), \qquad W_{j,n}^{\tau} = W_{j,n}1[|W_{j,n}| \leq C_\tau \tau_n],$$

where C_τ is a constant to be chosen. We have

$$\mathbb{P}[\hat{\Psi}_n^{(i)}(\cdot) \neq \hat{\Psi}_n^{\tau,(i)}(\cdot) \text{ for some } i] \leq \mathbb{P}[\max_{1\leq i\leq n}|W_{i,n}| > C_\tau\tau_n],$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily small by making C_τ large. Also,

$$\max_{1\leq i\leq n}\sup_{x\in\mathbb{R}^d}|\mathbb{E}[\hat{\Psi}_n^{(i)}(x) - \hat{\Psi}_n^{\tau,(i)}(x)]| = O(n^{-1/2}) = O(\rho_n)$$

because

$$\begin{aligned}\frac{n}{\tau_n}|\mathbb{E}[\hat{\Psi}_n^{(i)}(x) - \hat{\Psi}_n^{\tau,(i)}(x)]| &= \frac{n}{\tau_n}|\mathbb{E}[W_n 1(|W_n| > C_\tau\tau_n)\mathcal{K}_n(x - X)]| \\ &\leq \frac{n\mathcal{C}(S_n)}{C_\tau^{S_n}\tau_n^{S_n}} C_\tau \int_{\mathbb{R}^d}|\mathcal{K}(u)|du,\end{aligned}$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily small by making C_τ large. To show the desired result it therefore suffices to show that for every C_τ,

$$\max_{1\leq i\leq n}|\hat{\Psi}_n^{\tau,(i)}(X_{i,n})-\Psi_n^\tau(X_{i,n})|=O_p(\rho_n),\qquad \Psi_n^\tau(x)=\mathbb{E}\hat{\Psi}_n^\tau(x)=\mathbb{E}\hat{\Psi}_n^{\tau,(i)}(x).$$

For any M,

$$\begin{aligned}\mathbb{P}\left[\max_{1\leq i\leq n}|\hat{\Psi}_n^{\tau,(i)}(X_{i,n})-\Psi_n^\tau(X_{i,n})|>M\rho_n\right] &\leq n\max_{1\leq i\leq n}\mathbb{P}[|\hat{\Psi}_n^{\tau,(i)}(X_{i,n})-\Psi_n^\tau(X_{i,n})|>M\rho_n]\\ &\leq n\max_{1\leq i\leq n}\sup_{x\in\mathbb{R}^d}\mathbb{P}[|\hat{\Psi}_n^{\tau,(i)}(x)-\Psi_n^\tau(x)|>M\rho_n],\end{aligned}$$

where the last inequality uses the fact that X_i is independent of $\hat{\Psi}_n^{\tau,(i)}$. Because

$$|W_{j,n}^\tau\mathcal{K}_n(x-X_{j,n})-\Psi_n^\tau(x)|=O(\tau_n/h_n^d),\qquad \mathbb{V}[W_{j,n}^\tau\mathcal{K}_n(x-X_{j,n})]=O(1/h_n^d),$$

it follows from Bernstein's inequality that

$$n\max_{1\leq i\leq n}\sup_{x\in\mathbb{R}^d}\mathbb{P}[|\hat{\Psi}_n^{\tau,(i)}(x)-\Psi_n^\tau(x)|>M\rho_n]\leq 2n\exp\left[-\frac{M^2n\rho_n^2h_n^d}{O(1+M\rho_n\tau_n)}\right].$$

To complete the proof of ($SA-2$) it therefore suffices to show that

$$\overline{\lim}_{n\to\infty}\frac{1}{\log n}\frac{M^2n\rho_n^2h_n^d}{1+M\rho_n\tau_n}$$

can be made arbitrarily large by making M large.

In case (a), the desired result follows from the proof of Cattaneo, Crump, and Jansson (2013, Lemma B-1).

In case (b),

$$\frac{1}{\log n}\frac{M^2n\rho_n^2h_n^d}{1+M\rho_n\tau_n}=\frac{M^2}{1+MC_\tau\rho_n\log n},$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily large (by making M large) if $\rho_n\log n=\sqrt{(\log n)^3/(nh_n^d)}$ is bounded.

In case (c),

$$\frac{1}{\log n}\frac{M^2n\rho_n^2h_n^d}{1+M\rho_n\tau_n}=\frac{M^2}{1+MC_\tau\rho_n},$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily large (by making M large) if ρ_n is bounded.

Proof of (SA-3). For any M,

$$\mathbb{P}[\max_{1\leq i\leq n}|\hat{\Psi}_n^*(X_{i,n})-\hat{\Psi}_n(X_{i,n})|>M\rho_n]=\mathbb{E}\mathbb{P}^*[\max_{1\leq i\leq n}|\hat{\Psi}_n^*(X_{i,n})-\hat{\Psi}_n(X_{i,n})|>M\rho_n]$$

and

$$\mathbb{P}^*[\max_{1\le i\le n} |\hat{\Psi}_n^*(X_{i,n}) - \hat{\Psi}_n(X_{i,n})| > M\rho_n] \le n \sup_{x\in\mathbb{R}^d} \mathbb{P}^*[|\hat{\Psi}_n^*(x) - \hat{\Psi}_n(x)| > M\rho_n].$$

Because

$$|W_{j,n}^* \mathcal{K}_n(x - X_{j,n}^*) - \hat{\Psi}_n(x)| = O_p(h_n^{-d}\tau_n), \qquad \mathbb{V}^*[W_{j,n}^* \mathcal{K}_n(x - X_{j,n}^*)] = O_p(h_n^{-d}),$$

it follows from Bernstein's inequality that

$$\mathbb{P}^*[|\hat{\Psi}_n^*(x) - \hat{\Psi}_n(x)| > M\rho_n] \le 2\exp\left[-\frac{M^2 n\rho_n^2 h_n^d}{O_p(1 + M\rho_n\tau_n)}\right].$$

Validity of (SA-3) follows from this bound and the fact that

$$\overline{\lim}_{n\to\infty} \frac{1}{\log n} \frac{M^2 n\rho_n^2 h_n^d}{1 + M\rho_n\tau_n}$$

can be made arbitrarily large by making M large.

Proof of (SA-4). Because

$$\hat{\Psi}_n^{*,(i)}(x) = (1 - n^{-1})^{-1}\hat{\Psi}_n^*(x) - (n-1)^{-1} W_{i,n}^* \mathcal{K}_n(x - X_{i,n}^*),$$

we have the bound

$$(1 - n^{-1}) \max_{1\le i,j\le n} |\hat{\Psi}_n^{*,(i)}(X_{j,n}) - \hat{\Psi}_n(X_{j,n})| \le \max_{1\le j\le n} |\hat{\Psi}_n^*(X_{j,n}) - \hat{\Psi}_n(X_{j,n})| + R_n^*,$$

where

$$\begin{aligned} R_n^* &= n^{-1} \max_{1\le i\le n} |\hat{\Psi}_n(X_{i,n})| + n^{-1}\mathcal{K}_n(0) \max_{1\le i\le n} |W_{i,n}| \\ &\le n^{-1} \max_{1\le i\le n} |\hat{\Psi}_n(X_{i,n}) - \Psi_n(X_{i,n})| + n^{-1} \sup_{x\in\mathbb{R}^d} |\Psi_n(x)| + O(n^{-1}h_n^{-d}) \max_{1\le i\le n} |W_{i,n}| = O_p(\rho_n). \end{aligned}$$

In particular, (SA-4) holds because (SA-3) holds.

References

Cattaneo, M. D., R. K. Crump, and M. Jansson (2013): "Generalized Jackknife Estimators of Weighted Average Derivatives (with Discussion and Rejoinder)," *Journal of the American Statistical Association*, 108, 1243–1268.

Chen, X., O. Linton, and I. van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function is Not Smooth," *Econometrica*, 71, 1591–1608.

Giné, E., and R. Nickl (2008): "A Simple Adaptive Estimator of the Integrated Square of a Density," *Bernoulli*, 14, 47–61.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer, New York.