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

Matias D. Cattaneo† Michael Jansson‡

November 22, 2017

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

We provide details on the examples given in the main paper, which illustrate different features of our main results. This section considers the average density estimand

$$\theta_0 = \mathbb{E}[\gamma_0(x)] = \int_{\mathbb{R}^d} \gamma_0(x)^2 dx,$$

where $\gamma_0(x)$ denotes the Lebesgue density of a random vector $x \in \mathbb{R}^d$. In many respects this can be seen as the simplest possible semiparametric problem, that is, we view it as the analogue of the (sample) mean of a distribution in parametric mathematical statistics. This simple example provides straightforward illustration of several interesting features of semiparametric estimators, as already discussed in the main text briefly, and in more detail herein.

We consider three distinct semiparametric kernel-based estimators: (i) the plug-in sample average estimator $\hat{\theta}_n^{\texttt{AD}} = n^{-1}\sum_{i=1}^n \hat{\gamma}_n(x_i)$, where $\hat{\gamma}_n(x)$ is a kernel-based density estimator; (ii) the integrated square density estimator $\hat{\theta}_n^{\texttt{ISD}} = \int_{\mathbb{R}^d} \hat{\gamma}_n(x)^2 dx$; and (iii) the "locally robust" estimator $\hat{\theta}_n^{\texttt{LR}} = 2\hat{\theta}_n^{\texttt{AD}} - \hat{\theta}_n^{\texttt{ISD}}$.

As discussed in the main paper, this example is used to illustrate three main findings. First, $\hat{\theta}_n^{\texttt{AD}}$ shows that Stochastic Equicontinuity is a necessary condition, when a "master theorem" is applied directly to this estimator, and hence such a condition must be replaced; in this case, $\mathfrak{B}_n^{\texttt{LI}} \neq 0$ but $\mathfrak{B}_n^{\texttt{NL}} = 0$. This fact leads to our proposed weaker condition: Asymptotic Separability. Second, $\hat{\theta}_n^{\texttt{ISD}}$ shows that changing the form of the estimating equation can have important implications for small bandwidth biases; in this case, $\mathfrak{B}_n^{\texttt{LI}} = 0$ but $\mathfrak{B}_n^{\texttt{NL}} \neq 0$, as argued in Cattaneo, Crump, and Jansson (2013, Rejoinder). Finally, $\hat{\theta}_n^{\texttt{LR}}$ shows that "locally robust" estimators are not robust to small bandwidths; in this case, $\mathfrak{B}_n^{\texttt{LI}} \neq 0$ and $\mathfrak{B}_n^{\texttt{NL}} \neq 0$.

Suppose $x_1, \ldots, x_n$ are *i.i.d.* copies of a continuously distributed random vector $x \in \mathbb{R}^d$ with Lebesgue density γ_0. To obtain primitive bandwidth conditions for the conditions of Theorems 1 and 2, suppose that for some $P > d/2$, the following regularity conditions hold:

- γ_0 is P times differentiable, and γ_0 and its first P derivatives are bounded and continuous.
- K is even and bounded with $\int_{\mathbb{R}^d} |K(u)| (1 + \|u\|^P) du < \infty$ 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, though this is not the main focus of our paper. See, e.g., Giné and Nickl (2008) and references therein.

SA.1.1 Average Density Estimator $\hat{\theta}_n^{\texttt{AD}}$

A sample analogue kernel-based estimator of $\theta_0 = \mathbb{E}[\gamma_0(x)]$ is given by

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

When verifying the conditions of Theorems 1 and 2 for this example, we set $z = x$, $x(z, \theta) = z$, $w(z, \theta) = 1$, $\gamma_0(\cdot, \theta) = \gamma_0(\cdot)$, and let $\hat{\theta}_n$ be defined by $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = 0$, where $g(x, \theta, \gamma) = \gamma(x) - \theta = g^{\texttt{AD}}(x, \theta, \gamma)$ is a linear functional of $\gamma = f$.

Being a second-order V-statistic, the estimator is very tractable. Partly due to this tractability, this estimator has been widely studied, and we include here in part because it provides a dramatic demonstration of the fragility of Stochastic Equicontinuity (SE) with respect to bandwidth choice. It also illustrates how to verify sufficient conditions, and their relationship to necessary conditions, in arguably 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 conclusion of Theorem 1 holds with $\mathfrak{B}_n = \mathfrak{B}_n^{\texttt{AD}} + o(n^{-1/2})$,

$$\mathfrak{B}_n^{\texttt{AD}} = \frac{1}{nh_n^d}K(0) \qquad \textit{and} \qquad \Sigma_0 = 4\mathbb{V}[\gamma_0(x)].$$

Because

$$\sqrt{n}\mathfrak{B}_n = \sqrt{n}\mathfrak{B}_n^{\texttt{AD}} + o(1) = \frac{1}{\sqrt{nh_n^{2d}}}K(0) + o(n^{-1/2}),$$

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

SA.1.1.1 Condition AL

This condition holds with $\mathcal{J}_0 = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms. Therefore, $\mathfrak{B}_n = \mathcal{B}_n$ and $\Sigma_0 = \Omega_0$.

SA.1.1.2 Condition AS

Because

$$\bar{g}_n(x, \gamma) = g_n(x, \gamma_n) + g_{n,\gamma}(x)[\gamma - \gamma_n], \qquad g_{n,\gamma}(x)[\gamma] = (1 - n^{-1})\gamma(x),$$

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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\hat{\gamma}_n^{(i)}(x_i) - 2\gamma_n(x_i) + \theta_n] = o_{\mathbb{P}}(1), \qquad \theta_n = \int_{\mathbb{R}^d}\gamma_n(x)\gamma_0(x)dx,$$

where $\hat{\gamma}_n^{(i)}(x) = (n-1)^{-1} \sum_{j=1, j\neq i}^n K_n(x - x_j)$. A simple variance calculation now shows that Condition AS is satisfied if $nh_n^d \to \infty$, because

$$\mathbb{V}(g_{n,\gamma}(z_i)[\hat{\gamma}_n^j]) = (1-n^{-1})^2 \mathbb{V}[K_n(x_1 - x_2) - \gamma_n(x_1)] = O(1/h_n^d) = o(n).$$

SA.1.1.3 Comparison to SE

If $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then

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

where the last equality uses $\mathbb{E}(|\gamma_n(x) - \gamma_0(x)|^2) = o(1)$ and $\theta_n - \theta_0 = O(h_n^P) = o(n^{-1/2})$.

As a consequence, Stochastic Equicontinuity requires $nh_n^{2d} \to \infty$ in this example, which can not be improved upon. In other words, in this example, the calculations are based on an exact decomposition and hence give necessary conditions.

SA.1.1.4 Condition AN

We have:

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

where

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

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

If $h_n \to 0$, then $\psi_n(x) \to \psi(x)$ for every x, and it follows from the dominated convergence theorem that $\mathbb{E}\|\psi_n(z) - \psi_0(z)\|^2 \to 0$ with $\psi_0(x) = 2[\gamma_0(x) - \theta_0]$. Furthermore,

$$\mathfrak{B}_n = \mathcal{B}_n = \mathcal{B}_n^{\mathtt{S}} + \mathcal{B}_n^{\mathtt{LI}} + \mathcal{B}_n^{\mathtt{NL}},$$

$$\mathcal{B}_n^{\mathtt{S}} = h_n^P \mathcal{B}_0^{\mathtt{S}} + o(h_n^P), \qquad \mathcal{B}_n^{\mathtt{LI}} = \frac{1}{nh_n^d} \mathcal{B}_0^{\mathtt{LI}} + o(n^{-1/2}), \qquad \mathcal{B}_n^{\mathtt{NL}} = 0,$$

where, using standard multi-index notation,

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

and $\mathcal{B}_0^{\mathtt{LI}} = K(0)$ and $\mathcal{B}_0^{\mathtt{NL}} = 0$. Therefore, Condition AN is satisfied with $\Omega_0 = 4\mathbb{V}[\gamma_0(x)]$ if $h_n \to 0$, and $\mathcal{B}_n = K(0)/(nh_n^d) + o(n^{-1/2})$, 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 $\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ holds with $\mathfrak{B}_n = \mathcal{B}_n = \mathcal{B}_n^{\mathtt{LI}} = O(1/(nh_n^d))$.

SA.1.1.5 Bandwidth Selection.

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 the small bandwidth bias constant is

$$\mathcal{B}_0^{\mathtt{SB}} = \mathcal{B}_0^{\mathtt{LI}} + \mathcal{B}_0^{\mathtt{NL}} = K(0),$$

and the smoothing bias constant is

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

SA.1.1.6 Condition AL*

This condition holds with $\mathcal{J}_0 = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms.

SA.1.1.7 Condition AS*

Because

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

Condition AS* holds with $\bar{g}_n^* = g_n^*$ if $\mathbb{V}^*(\bar{g}_{n,\gamma}^*(z_i^*)[\gamma_n^{*,j}]) = o_{\mathbb{P}}(n)$. A sufficient condition for this to occur is that $nh_n^d \to \infty$, because then

$$\mathbb{E}\mathbb{V}^*(g_{n,\gamma}^*(z_i^*)[\gamma_n^{*,j}]) = (1 - n^{-1})^2 \mathbb{E}\mathbb{V}^*[\mathcal{K}_n(x_1^* - x_2^*) - \hat{\gamma}_n(x_1^*)] = O(1/h_n^d) = o(n).$$

SA.1.1.8 Condition AN*

We have:

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

where, defining $\hat{\gamma}_n^+(x)=n^{-1}K_n(0)+(1-n^{-1})\hat{\gamma}_n(x)$,

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

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

$$\frac{1}{n}\sum_{i=1}^{n}|\hat{\gamma}_n(x_i)-\gamma_n(x_i)|^2\to_{\mathbb{P}}0.$$

A sufficient condition for this to occur is that $\max_{1\le i\le n}|\hat{\gamma}_n(x_i)-\gamma_n(x_i)|=o_{\mathbb{P}}(1)$, which in turn will hold if $nh_n^d/\log n\to\infty$. This could be established using the technical 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 if $nh_n^d\to\infty$, then

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}|\hat{\gamma}_n(x_i)-\gamma_n(x_i)|^2\right]=O(n^{-1}h_n^{-d})\to 0.$$

In other words, Condition AN* holds if $h_n\to 0$ and if $nh_n^d\to\infty$.

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 $\hat{\theta}_n^{\texttt{ISD}}$

We now consider the plug-in kernel-based estimator of $\theta_0=\int_{\mathbb{R}^d}\gamma_0(x)^2dx$, given by

$$\hat{\theta}_n^{\texttt{ISD}}=\int_{\mathbb{R}^d}\hat{\gamma}_n(x)^2dx,\qquad \hat{\gamma}_n(x)=\frac{1}{n}\sum_{j=1}^{n}K_n(x-x_j),\qquad K_n(u)=\frac{1}{h_n^d}K\left(\frac{u}{h_n}\right).$$

When verifying the conditions of Theorems 1 and 2 for this example, we set $z=x$, $x(z,\theta)=z$, $w(z,\theta)=1$, $\gamma_0(\cdot,\theta)=\gamma_0(\cdot)$, and let $\hat{\theta}_n^{\texttt{ISD}}$ be defined by $\hat{G}_n(\hat{\theta}_n^{\texttt{ISD}},\hat{\gamma}_n)=0$, where $g(x,\theta,\gamma)=g^{\texttt{ISD}}(x,\theta,\gamma)=\int_{\mathbb{R}^d}\gamma(x)^2dx-\theta$ is a non-linear functional of $\gamma=f$, which does not involve an evaluation point.

The estimator $\hat{\theta}_n^{\texttt{ISD}}$ is also a second-order V-statistic, but unlike $\hat{\theta}_n^{\texttt{AD}}$, it will not exhibit leave-in bias. On the other hand, this estimator has a non-linearity bias. If the bandwidth satisfies $nh_n^{2P}\to 0$

and $nh_n^d \to \infty$, we show here that the conclusion of Theorem 1 holds with $\mathfrak{B}_n = \mathfrak{B}_n^{\mathtt{ISD}} + o(n^{-1/2})$,

$$\mathfrak{B}_n^{\mathtt{ISD}} = \frac{1}{nh_n^d}\int_{\mathbb{R}^d} K(u)^2 du \qquad \textit{and} \qquad \Sigma_0 = 4\mathbb{V}[\gamma_0(x)].$$

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

SA.1.2.1 Condition AL

This condition holds with $\mathcal{J}_0 = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms. Therefore, $\mathfrak{B}_n = \mathcal{B}_n$ and $\Sigma_0 = \Omega_0$.

SA.1.2.2 Condition AS

The estimator $\hat{\theta}_n$ can be analyzed using direct calculations, after observing that

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

where

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

However, instead of using the above V-statistic representation, here we verify the high-level conditions to illustrate our generic results. Therefore, define the (exact) quadratic approximation

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

with

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

$$\bar{g}_{n,\gamma}(x)[\gamma] = 2(1-n^{-1})\int_{\mathbb{R}^d} [n^{-1}K_n(u-x) + (1-n^{-1})\gamma_n(u)]\gamma(u)du,$$

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

The first part of Condition AS holds directly, without any remainder (cubic) term because the quadratic approximation above is exact. The second part of Condition AS follows from Lemma 2 if $nh_n^d \to 0$ because, simple variance calculations, give

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

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

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

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

where $i \neq j \neq k$.

SA.1.2.3 Condition AN

Recall that we have

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

where

$$\bar{G}_n(\gamma) = \mathbb{E}[g_n(z_i,\gamma)] = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}[n^{-1}K_n(u-x) + (1-n^{-1})\gamma_n(u)]^2 du\gamma_0(x)dx - \theta_0,$$

$$\bar{G}_{n,\gamma}[\gamma] = \mathbb{E}[\bar{g}_{n,\gamma}(z_i)[\gamma]] = 2(1-n^{-1})\int_{\mathbb{R}^d}\gamma_n(u)\gamma(u)du,$$

$$\bar{G}_{n,\gamma\gamma}[\gamma,\eta] = \mathbb{E}[\bar{g}_{n,\gamma\gamma}(z_i)[\gamma,\eta]] = \bar{g}_{n,\gamma\gamma}(x)[\gamma,\eta],$$

and

$$\psi_n(z) = g_n(z,\gamma_n) - \mathbb{E}g_n(z,\gamma_n) + \delta_n(x),$$

$$\delta_n(x) = 2(1-n^{-1})\int_{\mathbb{R}^d}\gamma_n(u)[K_n(u-x) - \gamma_n(u)]du,$$

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

In this example,

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

and hence, if $h_n \to 0$ and if $nh_n^d \to \infty$, then

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

for each z. Therefore, using the dominated convergence theorem, we verifies that $\mathbb{E}\|\psi_n(z) - \psi_0(z)\|^2 \to 0$, and hence the asymptotic linear representation, and establishes that $\Sigma_0 = 4\mathbb{V}[\gamma_0(x)]$.

Next, we have the representation

$$\bar{G}_{n,\gamma\gamma}[\gamma, \eta] = 2 \int_{\mathbb{R}^d} \gamma(u)\eta(u)du$$

and therefore, if $h_n \to 0$ and if $nh_n^d \to \infty$, it is easy to verify the conditions of Lemma 3, namely

$$\mathbb{E}(\|\bar{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]\|^2) = O(1/h_n^{2d}) = o(n^2),$$

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

where $i \neq j$.

Therefore, it remains to study the bias $\mathfrak{B}_n = \mathcal{B}_n = \mathcal{B}_n^{\mathtt{S}} + \mathcal{B}_n^{\mathtt{LI}} + \mathcal{B}_n^{\mathtt{NL}}$. In this case,

$$\mathbb{E}g_n(x, \gamma_n) = \int_{\mathbb{R}^d} \gamma_n(u)^2 du - \theta_0 + O(n^{-1}) = \mathcal{B}_n^{\mathtt{S}}$$

with $\mathcal{B}_n^{\mathtt{S}} = h_n^P \mathcal{B}_0^{\mathtt{S}} + o(h_n^P)$ where, using standard multi-index notation,

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

That is, the smoothing bias of $\hat{\theta}_n^{\mathtt{ISD}}$ is twice as large the smoothing bias of $\hat{\theta}_n^{\mathtt{AD}}$. In addition, we have $\mathcal{B}_n^{\mathtt{LI}} = 0$ and

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

and hence

$$\mathcal{B}_n^{\mathtt{NL}} = \frac{1}{nh_n^d}\mathcal{B}_0^{\mathtt{NL}} + o(n^{-1/2}), \qquad \mathcal{B}_0^{\mathtt{NL}} = \int_{\mathbb{R}^d} K(u)^2 du.$$

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

SA.1.2.4 Bandwidth Selection

We can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

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

where the small bandwidth bias is

$$\mathcal{B}_0^{\texttt{SB}} = \mathcal{B}_0^{\texttt{LI}} + \mathcal{B}_0^{\texttt{NL}} = \int_{\mathbb{R}^d} K(u)^2 du$$

and the smoothing bias is

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

SA.1.2.5 Condition AL*

This condition holds with $\mathcal{J} = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms. Therefore, $\mathfrak{B}_n = \mathcal{B}_n$ and $\Sigma_0 = \Omega_0$.

SA.1.2.6 Condition AS*

The (exact) quadratic approximation

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

with

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

$$\bar{g}_{n,\gamma}^*(x)[\gamma] = 2(1 - n^{-1}) \int_{\mathbb{R}^d} [n^{-1} K_n(u - x) + (1 - n^{-1}) \hat{\gamma}_n(u)] \gamma(u) du,$$

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

Condition AS* holds if $nh_n^d \to \infty$, because the conditions of Lemma 5 hold, since simple calculations verify

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

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

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

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

where, as before, $i \neq j \neq k$ and the quadratic approximation $\bar{g}_n^*(x, \hat{\gamma}_n)$ is exact.

SA.1.2.7 Condition AN*

We have:

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

where

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

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

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

Therefore, assuming $h_n \to 0$ and $nh_n^d/\log n \to \infty$, we have

$$\frac{1}{n}\sum_{i=1}^{n}|\psi_n^*(x_i)-\psi_n(x_i)|^2=\frac{1}{n}\sum_{i=1}^{n}|\delta_n^*(x_i)-\delta_n(x_i)|^2+O_{\mathbb{P}}(n^{-1})\to_{\mathbb{P}}0,$$

because $\max_{1\leq i\leq n}|\hat{\gamma}_n^*(x_i)-\hat{\gamma}_n(x_i)|=o_{\mathbb{P}}(1)$, which could be established using Lemma SA-1. Sufficiency of the slightly weaker condition $nh_n^d \to \infty$ can be demonstrated by using a direct calculation. Finally, employing Lemma 6, it is not difficult to show that under the side rate restrictions above, we have

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

In other words, Condition AN* holds if $h_n \to 0$ and if $nh_n^d/\log n \to \infty$.

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

SA.1.3 Locally Robust Estimator $\hat{\theta}_n^{\mathtt{LR}}$

Finally, we consider the "locally robust" estimator

$$\hat{\theta}_n^{\mathtt{LR}}=2\hat{\theta}_n^{\mathtt{AD}}-\hat{\theta}_n^{\mathtt{ISD}}=2\sum_{i=1}^{n}\hat{\gamma}_n(x)-\int_{\mathbb{R}^d}\hat{\gamma}_n(x)^2dx.$$

When verifying the conditions of Theorems 1 and 2 for this example, we set $z = x$, $x(z, \theta) = z$, $w(z, \theta) = 1$, $\gamma_0(\cdot, \theta) = \gamma_0(\cdot)$, and let $\hat{\theta}_n$ be defined by $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = 0$, where

$$\begin{aligned} g(x, \theta, \gamma) &= g^{\mathtt{LR}}(x, \theta, \gamma) = 2g^{\mathtt{AD}}(x, \theta, \gamma) - g^{\mathtt{ISD}}(x, \theta, \gamma) \\ &= 2\gamma(x) - \int_{\mathbb{R}^d} \gamma(x)^2 dx - \theta \\ &= 2\left[\gamma(x) - \theta\right] - \left[\int_{\mathbb{R}^d} \gamma(x)^2 dx - \theta\right] \end{aligned}$$

is a "locally robust" estimating equation. Specifically,

$$\nabla_\gamma \mathbb{E}[g(x, \theta_0, \gamma)]\big|_{\gamma_0} = 0,$$

where ∇_γ denotes the appropriate functional derivative (i.e., in most cases, Gateaux derivative).

If the bandwidth satisfies $nh_n^{4P} \to 0$ and $nh_n^d \to \infty$, we show here that the conclusion of Theorem 1 holds with

$$\mathfrak{B}_n = \frac{1}{nh_n^d}\left(2K(0) - \int_{\mathbb{R}^d} K(u)^2 du\right) \qquad \textit{and} \qquad \Sigma_0 = 4\mathbb{V}[\gamma_0(x)],$$

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

SA.1.3.1 Condition AL

This condition holds with $\mathcal{J}_0 = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms. Therefore, $\mathfrak{B}_n = \mathcal{B}_n$ and $\Sigma_0 = \Omega_0$.

SA.1.3.2 Condition AS

In this case, the (exact) quadratic approximation is

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

with

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

$$\bar{g}_{n,\gamma}(x)[\gamma] = 2(1 - n^{-1})\gamma(x) - 2(1 - n^{-1})\int_{\mathbb{R}^d} [n^{-1}K_n(u - x) + (1 - n^{-1})\gamma_n(u)]\gamma(u)du,$$

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

The first part of Condition AS holds directly, without any remainder (cubic) term because the quadratic approximation above is exact. Next, if $nh_n^d \to 0$, simple variance calculations, give

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

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

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

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

where, as before, $i \neq j \neq k$ and hence Condition AS holds via Lemma 2.

SA.1.3.3 Condition AN

We have:

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

where

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

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

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

Proceeding as above, we have

$$\psi_n(x) \to \psi_0(x) = 2[f(x) - \theta_0]$$

for each x, and therefore $\mathbb{E}[\|\psi_n(x) - \psi_0(x)\|^2] \to 0$ using the dominated convergence theorem. This establishes that $\Sigma_0 = 4\mathbb{V}[\gamma_0(x)]$. Furthermore, as above, if $h_n \to 0$ and if $nh_n^d \to \infty$, it is easy to verify the conditions of Lemma 3:

$$\mathbb{E}(\|\bar{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]\|^2) = O(1/h_n^{2d}) = o(n^2),$$

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

where, again, $i \neq j$.

Finally, it remains to study the biases $\mathcal{B}_n = \mathcal{B}_n^{\mathtt{S}} + \mathcal{B}_n^{\mathtt{LI}} + \mathcal{B}_n^{\mathtt{NL}}$. First, observe that

$$\mathcal{B}_n^{\mathtt{LI}} = \mathbb{E}g_n(x, \gamma_n) - \mathbb{E}g_0(x, \gamma_n) = \frac{1}{nh_n^d} 2K(0) + o\left(n^{-1}h^{-d}\right).$$

Second, we have

$$\begin{aligned}
\mathcal{B}_n^{\mathtt{S}} &= \mathbb{E}g_0(x, \gamma_n) = 2\left[\theta_n^+ - \theta_0\right] - \left(\int_{\mathbb{R}^d} \gamma_n(u)^2 du - \theta_0\right) + o\left(n^{-2}h_n^d + n^{-1}\right) \\
&= 2\left[\theta_n^+ - \theta_0\right] - \left(\int_{\mathbb{R}^d} [\gamma_n(u) - \gamma_0(u)]^2 du + 2\left[\theta_n^+ - \theta_0\right]\right) + o\left(n^{-2}h_n^d + n^{-1}\right) \\
&= -\int_{\mathbb{R}^d} [\gamma_n(u) - \gamma_0(u)]^2 du = O(h_n^{2P}),
\end{aligned}$$

and therefore

$$\mathcal{B}_n^{\mathtt{S}} = h_n^{2P}\mathcal{B}_0^{\mathtt{S}} + o(h_n^{2P}) + o\left(\frac{1}{nh^d}\right)$$

where, using standard multi-index notation,

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

Third, we have

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

In summary, if $nh_n^{4P} \to 0$ and if $nh_n^d \to \infty$, then the conditions of Theorem 1 are satisfied and $\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ holds with $\mathfrak{B}_n = \mathcal{B}_n = \mathcal{B}_n^{\mathtt{LI}} + \mathcal{B}_n^{\mathtt{NL}} = O(1/(nh_n^d))$.

SA.1.3.4 Bandwidth Selection

We can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

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

where the small bandwidth bias is

$$\mathcal{B}_0^{\mathtt{SB}} = \mathcal{B}_0^{\mathtt{LI}} + \mathcal{B}_0^{\mathtt{NL}} = 2K(0) - \int_{\mathbb{R}^d} K(u)^2 du$$

and the smoothing bias is

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

SA.1.3.5 Condition AL*

This condition holds with $\mathcal{J}_0 = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms. Therefore, $\mathfrak{B}_n = \mathcal{B}_n$ and $\Sigma_0 = \Omega_0$.

SA.1.3.6 Condition AS*

The (exact) quadratic approximation

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

with

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

$$\bar{g}_{n,\gamma}^*(x)[\gamma] = 2(1 - n^{-1})\gamma(x) - 2(1 - n^{-1}) \int_{\mathbb{R}^d} [n^{-1} K_n(u - x) + (1 - n^{-1})\hat{\gamma}_n(u)]\gamma(u) du,$$

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

Condition AS* holds if $nh_n^d \to \infty$, because the same calculations used previously verify

$$\mathbb{V}^*[\bar{g}_{n,\gamma}(x_i^*)[\gamma_n^{*,j} - \hat{\gamma}_n]] = O_{\mathbb{P}}(h_n^{-d}) = o_{\mathbb{P}}(n),$$

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

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

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

where, as before, $i \neq j \neq k$ and the quadratic approximation $\bar{g}_n^*(x, \hat{\gamma}_n)$ is exact.

SA.1.3.7 Condition AN*

It follows directly from the calculations above that the conditions of Theorem 2 are satisfied, provided that $nh_n^{4P} \to 0$ and if $nh_n^d / \log n \to \infty$.

SA.2 Example 2: Inverse Probability Weighting

This example is also discussed in the main paper. It illustrates two important features that are absent in the previous 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 same 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 γ_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 function of the same dimension as θ, 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. Define $r_0(x; \theta) = \mathbb{E}[m(y; \theta)|x, t = 1]$.

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,

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

That is, for $u = (u_1, u_2, \cdots, u_d)' \in \mathbb{R}^d$, the basis vector $b_P(u)$ is defined by $b_P(u) = (1, [u']^1, \ldots, [u']^P)'$ with

$$[u']^p = \left[u_1^{\ell_1} u_2^{\ell_2} \cdots u_d^{\ell_d} : |\ell| = \ell_1 + \ell_2 + \cdots + \ell_d = p, \quad \ell = (\ell_1, \ell_2, \cdots, \ell_d) \in \mathbb{Z}_+^d\right],$$

assumed to be ordered lexicographically without loss of generality. Also, let

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

and

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

where $b_{P,n}(x) = b_P(x/h_n)$, $K_n(z) = K(x/h_n)/h_n^d$, h_n is a bandwidth, K is a kernel, and where $\text{vec}_P : \mathbb{R}^{d_P \times d_P} \to \mathbb{R}^{d_P^2}$ is the vectorization operator. The local polynomial estimator (of order P) of $q_0(x)$ is given by $q(x;\hat{\gamma}_n) = e_P'\hat{\xi}_n(x)$, where

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

e_P is the first unit vector in $\mathbb{R}^{d_P}$, and $\text{vec}_P^{-1} : \mathbb{R}^{d_P^2} \to \mathbb{R}^{d_P \times d_P}$ is the inverse of vec_P. That is, $q(x;\hat{\gamma}_n) = e_P'\hat{\xi}_n(x)$ with

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

Because $\hat{\gamma}_n$ is kernel-based, the associated IPW estimator $\hat{\theta}_n$ is a kernel-based two-step semiparametric, which can be analyzed using the results of the previous sections 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

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

We also define

$$\gamma_{x,n}(x) = \mathbb{E}\hat{\gamma}_{x,n}(x) = \text{vec}_P[\int_{\mathbb{R}^d} b_P(u)b_P(u)'K(u)f_0(x + uh_n)du],$$

$$\gamma_{t,n}(x) = \mathbb{E}\hat{\gamma}_{t,n}(x) = \int_{\mathbb{R}^d} b_P(u)K(u)q_0(x + uh_n)f_0(x + uh_n)du,$$

and

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

$$\gamma_{t,0}(x) = [\int_{\mathbb{R}^d} b_P(u)K(u)du]q_0(x)f_0(x).$$

We also set

$$\xi_{P,0}(x) = \left[q_0(x), q_0^{(1)}(x)', q_0^{(2)}(x)', \cdots, q_0^{(P)}(x)'\right]'$$

with

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

using the usual multi-index notation

$$\ell! = \ell_1!\ell_2!\cdots\ell_d!, \qquad \partial^{\ell} = \frac{\partial^{|\ell|}}{\partial^{\ell_1}\partial^{\ell_2}\cdots\partial^{\ell_d}}.$$

When verifying the conditions of Theorems 1 and 2 for this example, we also set $x(z,\theta) = x$, $w(z,\theta) = 1$, $\gamma_0(\cdot,\theta) = \gamma_0(\cdot) = (\gamma_{x,0}(\cdot)', \gamma_{t,0}(\cdot)')'$, $\gamma_n(\cdot,\theta) = \gamma_n(\cdot) = (\gamma_{x,n}(\cdot)', \gamma_{t,n}(\cdot)')'$, and

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

$$q_n(x) = q(x;\gamma_n) = e_P'(\mathrm{vec}_P^{-1}[\gamma_{x,n}(x)])^{-1}\gamma_{t,n}(x),$$
$$q_0(x) = q(x;\gamma_0) = e_0'\Xi(\gamma_{x,0}(x))^{-1}\gamma_{t,0}(x).$$

In this example, $g(z,\theta,\gamma) = \frac{t}{q(x;\gamma)}m(y;\theta)$ is neither linear in γ nor is assumed differentiable in θ. We can handle over-identification, via Lemma 1, but we assume just-identification for simplicity.

The leave-one-out estimator $\hat{\gamma}_n^{(i)}(x_i) = (\hat{\gamma}_{x,n}^{(i)}(x)', \hat{\gamma}_{t,n}^{(i)}(x)')'$ is defined as below. That is,

$$\mathrm{vec}_P^{-1}(\hat{\gamma}_{x,n}(x)) = (n-1)^{-1}K_n(0)e_Pe_P' + \mathrm{vec}_P^{-1}(\hat{\gamma}_{x,n}^{(i)}(x)),$$

$$\hat{\gamma}_{x,n}(x) = (n-1)^{-1}K_n(0)e_P + \hat{\gamma}_{x,n}^{(i)}(x),$$
$$\hat{\gamma}_{t,n}(x) = t_i(n-1)^{-1}K_n(0)e_P + \hat{\gamma}_{t,n}^{(i)}(x),$$

where e_P is the first unit vector of the required, conformable length in the above displays.

We impose the following primitive regularity conditions to verify the assumptions of Theorems 1 and 2:

- $\theta_0 \in \mathrm{int}(\Theta)$.

- $\mathbb{E}[\|t\cdot m(y;\theta_0)\|^4] < \infty$, $\sup_{x\in\mathbb{X}}\mathbb{E}[\|t\cdot m(y;\theta_0)\|^4|x]f_0(x) < \infty$ and $\Sigma = \mathbb{V}[\psi_0(z)]$ is positive definite, where
$$\psi_0(z) = \frac{t}{q_0(x)}m(y;\theta_0) - \frac{r_0(x;\theta_0)}{q_0(x)}(t - q_0(x)),$$
where $x = (x_1, x_2, \cdots, x_d)$. The Lebesgue density f_0 is bounded away from zero.

- $\mathcal{M} = \{t\cdot m(y;\theta) : \theta\in\Theta\}$ satisfies the bracketing integral entropy condition $J_{[]}(1,\mathcal{M},L_2(P)) < \infty$, where
$$J_{[]}(\delta,\mathcal{F},L_2(P)) = \int_0^{\delta}\sqrt{\log N_{[]}(\delta,\mathcal{F},L_2(P))}d\varepsilon < \infty$$
with $N_{[]}(\delta,\mathcal{F},L_r(P))$ denoting the bracketing number for the class $\mathcal{F}$ under the usual $L_r(P)$ metric; for more details and precise definitions see, e.g., van der Vaart and Wellner (1996). Furthermore, suppose that $\mathbb{E}\left[\sup_{\theta\in\Theta}\|tm(y;\theta)\|^2\right] < \infty$, $\mathbb{E}[t\|m(y;\theta) - m(y;\theta_0)\|] \preceq \|\theta-\theta_0\|$, and $\mathbb{E}[t\|m(y;\theta) - m(y;\theta_0)\|^2] \preceq \|\theta-\theta_0\|^p$ for some $p\in[1,2]$.

- $r_0(\cdot;\theta)$ is twice continuously differentiable in θ, with first and second bounded derivatives denoted by $\dot{r}_0(x;\theta)$ and $\ddot{r}_0(\cdot;\theta)$, and $\mathbb{E}[\sup_{\|\theta-\theta_0\|\leq\delta}\|\ddot{r}_0(x;\theta)\|] \leq \infty$ for some $\delta > 0$.

- q_0 is bounded away from zero, and $P+2$ times continuously differentiable.

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

The third assumption controls the "smoothness" of $\theta \mapsto m(y;\theta)$, allowing for discontinuous (in θ) moment functions. It holds, in particular, if $m(y;\theta)$ is Lipschitz continuous in θ (and the implied Lipschitz constant is integrable). More generally, for example, the nondifferentiable moment condition $m(y;\theta) = \mathbb{1}(y \leq \theta) - \tau$ for the τ-th quantile of y, $\tau \in (0,1)$, satisfies

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

provided that y is continuously distributed with bounded density. The rest of the assumptions are standard.

We also impose the following assumptions on the kernel-based nonparametric estimator:

- Uniform consistency:

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

- Empirical uniform rate of convergence:

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

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

- Bounded away from zero: for some $q_{\min} > 0$,

$$\underline{\lim}_{n\to\infty}\inf_{x\in\mathbb{X}} q(x;\gamma_n) \geq q_{\min}.$$

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

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

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 $\sup_{x\in\mathbb{X}}\|\hat{q}_n(x) - q_n(x)\| = o_{\mathbb{P}}(1)$ and $\max_{1\leq i\leq n}\|\hat{q}_n(x_i) - q_n(x_i)\| = o_{\mathbb{P}}(n^{-1/6})$, and similarly for the bootstrap and leave-one-out versions. If, in addition, $\sup_{x\in\mathbb{X}}\|q_n(x) - q_0(x)\| = o(1)$ holds, then the third assumption is satisfied.

Finally, we assume throughout that $\hat{\theta}_n \to_{\mathbb{P}} \theta_0$. Similarly, for the bootstrap results we also assume that $\hat{\theta}_n^* \to_{\mathbb{P}} \theta_0$. This 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 conclusions of Theorem 1 and 2 hold with $\mathfrak{B}_n = \mathfrak{B}_n^{\mathtt{IPW}} + o(n^{-1/2})$,

$$\mathfrak{B}_n^{\mathtt{IPW}} = O\left(\frac{1}{nh_n^d}\right) \qquad and \qquad \Sigma_0 = 4\mathbb{V}[\psi_0(x)].$$

Therefore, once again, because

$$\sqrt{n}\mathfrak{B}_n = O\left(1/\sqrt{nh_n^{2d}}\right),$$

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

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

$$\hat{G}_n(\theta, \gamma) = \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q(x_i; \gamma)} m(y_i; \theta),$$

and

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

with $q_0(x) = q(x; \gamma_0)$ and $r_0(x; \theta) = \mathbb{E}[m(y; \theta)|x, t = 1]$.

Also, $r_0(x; \theta) = r_0(x; \theta_0) + \dot{r}_0(x; \theta_0)(\theta - \theta_0) + (\theta - \theta_0)'\ddot{r}_0(x; \tilde{\theta})(\theta - \theta_0)$, for some $\tilde{\theta}$ in between θ and θ_0, and hence

$$\|G(\theta, \gamma) - G(\theta_0, \gamma) - \dot{G}(\gamma)(\theta - \theta_0)\| \preceq \|\theta - \theta_0\|^2$$

with

$$\dot{G}(\gamma) = \mathbb{E}\left[\frac{q_0(x)}{q(x)} \dot{r}_0(x; \theta_0)\right].$$

Therefore, we set $\dot{G}_n = \dot{G}(\gamma_n)$ in this case, and obtain

$$\begin{aligned}
\|\dot{G}(\gamma_n) - \dot{G}(\gamma_0)\| &= \int \left|\frac{q(x; \gamma_n) - q(x; \gamma_0)}{q(x; \gamma_n)q(x; \gamma_0)}\right| \|\dot{r}_0(x; \theta_0)\| f_0(x)dx \\
&\preceq \int \|\gamma_n(x) - \gamma_0(x)\| f_0(x)dx = o(1),
\end{aligned}$$

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,

$$\begin{aligned}
&\|G(\theta,\hat{\gamma}_n) - G(\theta_0,\hat{\gamma}_n) - \dot{G}(\hat{\gamma}_n)(\theta-\theta_0)\| \\
&\leq \|\theta-\theta_0\|^2 \int \frac{t}{|\hat{q}_n(x)|} \|\ddot{r}_0(x;\theta_0)\| f_0(x)dx \preceq O_{\mathbb{P}}(1)\|\theta-\theta_0\|^2
\end{aligned}$$

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

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

Condition (iii). We have

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

where

$$\Delta_{1,n}(\theta) = \Delta_{11,n}(\theta) + \Delta_{12,n}(\theta)$$

with

$$\Delta_{11,n}(\theta) = \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{t_i}{q_n(x_i)}(m(y_i;\theta) - m(y_i;\theta_0)) - \mathbb{E}\left[\frac{t_i}{q_n(x_i)}(m(y_i;\theta) - m(y_i;\theta_0))\right] \right\|,$$

$$\Delta_{12,n}(\theta) = \left(\max_{1\leq i\leq n} \frac{\|\hat{q}_n(x_i) - q_n(x_i)\|}{\|\hat{q}_n(x_i)q_n(x_i)\|}\right) \left(\frac{1}{n}\sum_{i=1}^{n} t_i \|m(y_i;\theta) - m(y_i;\theta_0)\| - \mathbb{E}[t_i \|m(y_i;\theta) - m(y_i;\theta_0)\|]\right),$$

and

$$\Delta_{2,n}(\theta) = \Delta_{21,n}(\theta) + \Delta_{22,n}(\theta)$$

with

$$\Delta_{21,n}(\theta) = \int \frac{\|q(x;\hat{\gamma}_n) - q(x;\gamma_n)\|}{\|q(x;\hat{\gamma}_n)q(x;\gamma_n)\|} \|r_0(x;\theta) - r_0(x;\theta_0)\| f_0(x)dx,$$

$$\Delta_{22,n}(\theta) = \left(\max_{1\leq i\leq n} \frac{\|q(x_i;\hat{\gamma}_n) - q(x_i;\gamma_n)\|}{\|q(x_i;\hat{\gamma}_n)q(x_i;\gamma_n)\|}\right) \mathbb{E}\left[t_i\|m(y_i;\theta) - m(y_i;\theta_0)\|\right]$$

For the first term, for every $\delta_n = o(1)$, we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{11,n}(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

because $q_n(x)$ is non-random, $q_n(x)$ is bounded away from zero for all n large enough, and the class of n-varying functions $\mathcal{M}_n = \{t_i m(y_i;\theta)/q_n(x_i) : \theta \in \Theta\}$ satisfies easily the integral entropy condition $J_{[]}(\epsilon_n, \mathcal{M}_n, L_2(P)) \to 0$ for all $\epsilon_n \downarrow 0$.

For the second term, for every $\delta_n = o(1)$, we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{12,n}(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

because $\max_{1\leq i\leq n} \|\hat{q}_n(x_i) - q_n(x_i)\| = o_{\mathbb{P}}(1)$, $q_n(x)$ is bounded away from zero for all n large enough, and the class of functions $\mathcal{M}_{\|\cdot\|} = \{t_i \|m(y_i;\theta) - m(y_i;\theta_0)\| : \theta \in \Theta\}$ satisfies the integral entropy condition $J_{[]}(1, \mathcal{M}_{\|\cdot\|}, L_2(P)) < \infty$.

For the third term, for every $\delta_n = o(1)$, we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\Delta_{21,n}(\theta)}{1 + n^{1/3}\|\theta-\theta_0\|} = O_{\mathbb{P}}(n^{-1/3}) \int \|q(x;\hat{\gamma}_n) - q(x;\gamma_n)\| \|\dot{r}_0(x;\theta_0)\| f_0(x)dx = o_{\mathbb{P}}(n^{-1/2})$$

because $\sup_{x\in\mathbb{X}} \|\hat{q}_n(x) - q_n(x)\| = o_{\mathbb{P}}(1)$, $q_n(x)$ is bounded away from zero for all n large enough, $\|r_0(x;\theta) - r_0(x;\theta_0)\| \preceq \|\theta - \theta_0\|$, and

$$\int \|q(x;\hat{\gamma}_n) - q(x;\gamma_n)\|^2 f_0(x)dx = O_{\mathbb{P}}\left(\frac{1}{nh_n^d}\right) = o_{\mathbb{P}}(n^{-1/3}),$$

using standard results for local polynomial regression estimators and are assumed bandwidth rate restrictions.

For the fourth term, for every $\delta_n = o(1)$, we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\Delta_{22,n}(\theta)}{1 + n^{1/3}\|\theta-\theta_0\|} = O_{\mathbb{P}}(n^{-1/3}) \max_{1\leq i\leq n} \|q(x_i;\hat{\gamma}_n) - q(x_i;\gamma_n)\| = o_{\mathbb{P}}(n^{-1/2}),$$

by the same arguments given above.

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

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

where

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

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

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

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^3 \hat{q}_n(x_i)} m(y_i;\theta_0)[\hat{q}_n(x_i) - q_n(x_i)]^3 \right\| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \frac{t_i}{q_n(x_i)^3 |\hat{q}_n(x_i)|} \|m(y_i;\theta_0)\| \|\hat{q}_n(x_i) - q_n(x_i)\|^3 \\
&\preceq \max_{1\leq i\leq n} \|\hat{\gamma}_n(x_i) - \gamma_n(x_i)\|^3 \frac{1}{n} \sum_{i=1}^{n} \|m(y_i;\theta_0)\| = O_{\mathbb{P}}\left(\frac{(\log n)^{3/2}}{n^{3/2} h_n^{3d/2}} \right) = o_{\mathbb{P}}(n^{-1/3}),
\end{aligned}$$

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

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

because $\sup_{x\in\mathbb{X}} |\hat{q}_n(x_i) - q_n(x)| = o(1)$, $q_n(x)$ is bounded away from zero for all n large enough, and $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

Condition (vii). In condition (iii) we already showed

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{11,n}(\theta) = o_{\mathbb{P}}(n^{-1/2}).$$

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

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{21,n}(\theta) = O_{\mathbb{P}}(\delta_n) \int \|q(x;\hat{\gamma}_n) - q(x;\gamma_n)\| \|\dot{r}_0(x;\theta_0)\| f_0(x) dx = o_{\mathbb{P}}(n^{-1/2})$$

and

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{22,n}(\theta) = O_{\mathbb{P}}(\delta_n) \max_{1\leq i\leq n} \|q(x_i;\hat{\gamma}_n) - q(x_i;\gamma_n)\| = o_{\mathbb{P}}(n^{-1/2}).$$

SA.2.2 Condition AS

We now show that Condition AS holds using direct calculations and Lemma 2. A quadratic approximation to $g_n(z,\gamma)$ is given by

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

where

$$g_n(z,\gamma_n) = \frac{t_i}{q_n^+(x)} m(y_i;\theta_0), \qquad q_n^+(x) = e_P' \operatorname{vec}_P^{-1}(\gamma_{x,n}^+(x))^{-1}\gamma_{t,n}^+(x),$$

$$\gamma_{x,n}^+(x) = (n-1)^{-1}K_n(0)e_P + \gamma_{x,n}(x),$$

$$\gamma_{t,n}^+(x) = (n-1)^{-1}K_n(0)e_P + \gamma_{t,n}(x),$$

and for appropriate choice of linear and quadratic terms. To be precise, letting

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

to save some notation, we have the linear term

$$\begin{aligned}\bar{g}_{n,\gamma}(z)[\gamma] &= -\frac{tm(y;\theta_0)}{q_n^+(x)^2} e_P'\Gamma_{x,n}^{-1}(x)\gamma_t(x) \\ &\quad + \frac{tm(y;\theta_0)}{q_n^+(x)^2} e_P'\Gamma_{x,n}^{-1}(x)\operatorname{vec}_P^{-1}(\gamma_x(x))\Gamma_{x,n}^{-1}(x)\gamma_{t,n}^+(x)\end{aligned}$$

and the quadratic term

$$\bar{g}_{n,\gamma\gamma}(z)[\gamma,\eta] = \sum_{\ell=1}^{10} \bar{g}_{n,\gamma\gamma,\ell}(z)[\gamma,\eta]$$

with, using $\gamma = (\gamma_x', \gamma_t')'$ and $\eta = (\eta_x', \eta_t')'$,

$$\bar{g}_{n,\gamma\gamma,1}(z)[\gamma,\eta] = -\frac{tm(y;\theta_0)}{q_n^+(x)^2} e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\gamma_x(x))\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x),$$

$$\bar{g}_{n,\gamma\gamma,2}(z)[\gamma,\eta] = -\frac{tm(y;\theta_0)}{q_n^+(x)^2} e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\gamma_x(x))\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x),$$

$$\bar{g}_{n,\gamma\gamma,3}(z)[\gamma,\eta] = \frac{tm(y;\theta_0)}{q_n^+(x)^2} e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\gamma_x(x))\Gamma_{x,n}^+(x)^{-1}\eta_t(x),$$

$$\bar{g}_{n,\gamma\gamma,4}(z)[\gamma,\eta] = \frac{tm(y;\theta_0)}{q_n^+(x)^2} e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\gamma_t(x),$$

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

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

$$\bar{g}_{n,\gamma\gamma,7}(z)[\gamma,\eta]$$
$$= \frac{tm(y;\theta_0)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \Gamma_{x,n}^+(x)^{-1} e_P,$$

$$\bar{g}_{n,\gamma\gamma,8}(z)[\gamma,\eta]$$
$$= \frac{tm(y;\theta_0)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \Gamma_{x,n}^+(x)^{-1} e_P,$$

$$\bar{g}_{n,\gamma\gamma,9}(z)[\gamma,\eta] = -\frac{2tm(y;\theta_0)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \gamma_t(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \eta_x(x) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x),$$

$$\bar{g}_{n,\gamma\gamma,10}(z)[\gamma,\eta] = -\frac{2tm(y;\theta_0)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \eta_t(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \gamma_x(x) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x),$$

where in the above $q_n^+(x)$ can be treated as non-random because for any function $a(\cdot)$ we have:

$$\begin{aligned}
t \cdot a(q_n^+(x)) &= t \cdot a(e_P' \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x)) \\
&= t \cdot a\left(e_P' \left[\frac{K_n(0)}{(n-1)} e_P e_P' + \Gamma_{x,n}(x) \right]^{-1} \left[t \frac{K_n(0)}{(n-1)} e_P + \gamma_{t,n}(x) \right] \right) \\
&= t \cdot a\left(e_P' \left[\frac{K_n(0)}{(n-1)} e_P e_P' + \Gamma_{x,n}(x) \right]^{-1} \left[\frac{K_n(0)}{(n-1)} e_P + \gamma_{t,n}(x) \right] \right),
\end{aligned}$$

that is, whenever $t = 1$, the leave-in part of $\gamma_{t,n}^+(x)$ is non-random anymore.

If $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g_n(z_i, \hat{\gamma}_n^{(i)}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{g}_n(z_i, \hat{\gamma}_n^{(i)}) + o_{\mathbb{P}}(n^{-1/2}),$$

which gives the first part of of Condition AS.

Moreover, the second part of Condition AS is verified by Lemma 2 as well, because

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

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

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

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

provided $nh_n^d \to \infty$, with $i \neq j \neq k$.

Putting the above together, Condition AS is verified if $h_n \to 0$ and $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

SA.2.3 Condition AN

First we show that $\mathbb{E}[\|\psi_n(z) - \psi_0(z)\|^2] \to 0$. Recall that

$$g_0(z, \gamma_0) = \frac{t}{q_0(x)} m(y; \theta_0),$$

and note that

$$\begin{aligned}
\mathbb{E}\left[\|g_n(z,\gamma_n) - g_0(z,\gamma_n)\|^2\right] &= -\mathbb{E}\left[[q_n^+(x) - q_0(x)]^2 \frac{t}{(q_n^+(x)q_0(x))^2}\|m(y;\theta_0)\|^2\right] \\
&\preceq \mathbb{E}\left[\left(e_P'\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}(x) - q_0(x)\right)^2 \frac{t}{(q_n^+(x)q_0(x))^2}\|m(y(t);\theta_0)\|^2\right] \\
&+ \frac{K_n(0)^2}{(n-1)^2}\mathbb{E}\left[\frac{t}{(q_n^+(x)q_0(x))^2}\|m(y(t);\theta_0)\|^2\right] \\
&\to 0
\end{aligned}$$

provided that $nh_n^d \to \infty$ and $h_n \to 0$. Therefore, $\mathbb{E}\left[\|g_n(z,\gamma_n) - g_0(z,\gamma_0)\|^2\right] \to 0$.

Next, for the correction term $\delta_n(z_1) = \bar{G}_{n,\gamma}[\kappa_{n,1}]$ with $\bar{G}_{n,\gamma}[\cdot] = \mathbb{E}[\bar{g}_{n,\gamma}(z,\gamma_n)[\cdot]]$, we have:

$$\begin{aligned}
\delta_n(z) \quad = \quad &-\int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P'\Gamma_{x,n}^+(u)^{-1}\left[\bar{K}_{t,n}(x-u)t - \gamma_{t,n}(u)\right] f_0(u)du \\
&+\int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P'\Gamma_{x,n}^+(u)^{-1}\left[\bar{K}_{x,n}(x-u) - \Gamma_{x,n}(u)\right]\Gamma_{x,n}^+(u)^{-1}\gamma_{t,n}^+(u) f_0(u)dx
\end{aligned}$$

where we set

$$\bar{K}_{t,n}(x-u) = b_P\left(\frac{x-u}{h_n}\right) K_n(x-u),$$

$$\bar{K}_{x,n}(x-u) = b_P\left(\frac{x-u}{h_n}\right) b_P\left(\frac{x-u}{h_n}\right)' K_n(x-u).$$

Therefore,

$$\delta_n(z) = \delta_{1,n}(z) + \delta_{2,n}(z) + \delta_{3,n}(z) + \delta_{4,n}(z),$$

$$\delta_{1,n}(z) = -\int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P'\Gamma_{x,n}^+(u)^{-1}\bar{K}_{t,n}(x-u) t f_0(u)du,$$

$$\delta_{2,n}(z) = \int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P'\Gamma_{x,n}^+(u)^{-1}\gamma_{t,n}(u) f_0(u)du,$$

$$\delta_{3,n}(z) = \int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P'\Gamma_{x,n}^+(u)^{-1}\bar{K}_{x,n}(x-u)\Gamma_{x,n}^+(u)^{-1}\gamma_{t,n}^+(u) f_0(u)du,$$

$$\delta_{4,n}(z) = -\int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P'\Gamma_{x,n}^+(u)^{-1}\Gamma_{x,n}(u)\Gamma_{x,n}^+(u)^{-1}\gamma_{t,n}^+(u) f_0(u)du.$$

Then, we have

$$\begin{aligned}
\delta_{1,n}(z) &= -\int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \bar{K}_{t,n}(x-u) t f_0(u) du \\
&= -t \int \frac{q_0(x-vh_n)}{q_n^+(x-vh_n)^2} r_0(x-vh_n;\theta_0) e_P' \Gamma_{x,n}^+(x-vh_n)^{-1} b_P(v) K(v) f_0(x-vh_n) dv \\
&= -\frac{t}{q_0(x)} r_0(x;\theta_0) + O(n^{-1}h_n^{-d}),
\end{aligned}$$

because

$$\begin{aligned}
&\int e_P' \Gamma_{x,n}(x-vh_n)^{-1} b_P(v) K(v) f_0(x-vh_n) dv \\
&= e_P' \Gamma_{x,n}(x) \int b_P(v) K(v) f_0(x-vh_n) dv + o(1) = e_P' e_P + o(1) = 1.
\end{aligned}$$

Next,

$$\begin{aligned}
\delta_{2,n}(z) &= \int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \gamma_{t,n}(u) f_0(u) du \\
&= \int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}(u)^{-1} \gamma_{t,n}(u) f_0(u) du + O(n^{-1}h_n^{-d}) \\
&= \int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) q_0(u) f_0(u) du + O\left(\frac{1}{nh_n^d} + h_n^{P+1}\right).
\end{aligned}$$

Next,

$$\begin{aligned}
\delta_{3,n}(z) &= \int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \bar{K}_{x,n}(x-u) \Gamma_{x,n}^+(u)^{-1} \gamma_{t,n}^+(u) f_0(u) du \\
&= \int r_0(x-vh_n;\theta_0) \frac{q_0(x-vh_n)^2}{q_n^+(x-vh_n)^2} e_P' \Gamma_{x,n}(x-vh_n)^{-1} b_P(v) b_P(v)' K(v) f_0(x-vh_n) dv + o(1) \\
&\to r_0(x;\theta_0).
\end{aligned}$$

Finally,

$$\begin{aligned}
\delta_{4,n}(z) &= -\int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) e_P' \Gamma_{x,n}^+(u)^{-1} \Gamma_{x,n}(u) \Gamma_{x,n}^+(u)^{-1} \gamma_{t,n}^+(u) f_0(u) du \\
&= -\int \frac{q_0(u)}{q_n^+(u)^2} r_0(u;\theta_0) q_0(u) f_0(u) du + O(n^{-1}h_n^{-d}).
\end{aligned}$$

Therefore, $\delta_{2,n}(z) + \delta_{4,n}(z) \to 0$, and

$$\delta_n(z) \to -\frac{t}{q_0(x)} r_0(x;\theta_0) + r_0(x;\theta_0) = -\frac{r_0(x;\theta_0)}{q_0(x)}(t - q_0(x)).$$

Putting the above together, we have

$$\mathbb{E}\left[(\psi_n(z)-\psi_0(z))^2\right]\to 0,\qquad \psi_0(z)=\frac{t}{q_0(x)}m(y;\theta_0)-\frac{r_0(x;\theta_0)}{q_0(x)}(t-q_0(x)).$$

Next, we study the biases. First, we have

$$\begin{aligned}
\mathbb{E}[g_n(z,\gamma_n)] &= \mathbb{E}\left[\frac{t_i}{q_n^+(x)}m(y_i;\theta_0)\right]=\mathbb{E}\left[\frac{t_i}{q_n^+(x)}r_0(x_i;\theta_0)\right]\\
&= \mathbb{E}\left[\frac{t_i}{q_0(x)}r_0(x_i;\theta_0)\right]-\mathbb{E}\left[\left[q_n^+(x)-q_0(x)\right]\frac{t_i}{q_0(x)^2}r_0(x_i;\theta_0)\right]\left\{1+o(n^{-1/2})\right\}\\
&= -\mathbb{E}\left[\left[q_n^+(x)-q_0(x)\right]\frac{1}{q_0(x)}r_0(x_i;\theta_0)\right]\left\{1+o(n^{-1/2})\right\}
\end{aligned}$$

where

$$\begin{aligned}
q_n^+(x)-q_0(x) &= e_P'[\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)-\Gamma_{x,n}(x)^{-1}\gamma_{t,n}(x)]+e_P'\Gamma_{x,n}(x)^{-1}\gamma_{t,n}(x)-q_0(x)\\
&= t(n-1)^{-1}K_n(0)e_P'\Gamma_{x,n}(x)^{-1}e_P\\
&\quad -\frac{t(n-1)^{-2}K_n(0)^2}{1+(n-1)^{-1}K_n(0)e_P'\Gamma_{x,n}(x)^{-1}e_P}\left(e_P'\Gamma_{x,n}(x)^{-1}e_P\right)^2\\
&\quad -\frac{(n-1)^{-1}K_n(0)}{1+(n-1)^{-1}K_n(0)e_P'\Gamma_{x,n}(x)^{-1}e_P}e_P'\Gamma_{x,n}(x)^{-1}e_Pe_P'\Gamma_{x,n}(x)^{-1}\gamma_{t,n}(x)\\
&\quad +e_P'\Gamma_{x,n}(x)^{-1}\gamma_{t,n}(x)-q_0(x).
\end{aligned}$$

Therefore, if $nh^{3d/2}/(\log n)^{3/2}\to\infty$ and $h_n\to 0$, we have

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

where

$$\begin{aligned}
\mathcal{B}_n^{\mathtt{LI}} &= -\mathbb{E}\left[(n-1)^{-1}K_n(0)e_P'\Gamma_{x,n}(x)^{-1}e_P\frac{1}{q_0(x)}r_0(x;\theta_0)\right]\\
&= -\frac{K(0)}{nh_n^d}\int e_P'\Gamma_{x,n}(x)^{-1}e_P\frac{1}{q_0(x)}r_0(x;\theta_0)f_0(x)dx\\
&= -\frac{K(0)}{nh_n^d}e_P'\Gamma_x^{-1}e_P\int\frac{1}{q_0(x)}r_0(x;\theta_0)dx+O(h_n),
\end{aligned}$$

and

$$
\begin{aligned}
\mathcal{B}_n^{\mathtt{S}} &= -\mathbb{E}\left[[e_P'\Gamma_{x,n}(x)^{-1}\gamma_{t,n}(x) - q_0(x)]\frac{1}{q_0(x)}r_0(x;\theta_0)\right] \\
&= -\int [e_P'\Gamma_{x,n}(x)^{-1}\gamma_{t,n}(x) - q_0(x)]\frac{1}{q_0(x)}r_0(x;\theta_0)f_0(x)dx \\
&= -h_n^{P+1}\int \left[e_P'\Gamma_{x,n}(x)^{-1}\vartheta_{P,n}(x)q_0^{(P+1)}(x)\right]\frac{1}{q_0(x)}r_0(x;\theta_0)f_0(x)dx + O(h_n^{P+2}) \\
&= -h_n^{P+1}\left[\left(\int \frac{1}{q_0(x)}r_0(x;\theta_0)q_0^{(P+1)}(x)'f_0(x)dx\right)\vartheta_P'\Gamma_x^{-1}e_P + O(h_n + n^{-1}h_n^d)\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\Gamma_{x,n}(x)^{-1}\gamma_{t,n}(x) &= \Gamma_{x,n}(x)^{-1}\Gamma_{x,n}(x)H_n\xi_{P,0}(x) \\
&\quad + h_n^{P+1}\Gamma_{x,n}(x)^{-1}\vartheta_{P,n}(x)q_0^{(P+1)}(x) + O(h_n^{P+2})
\end{aligned}
$$

with $H_n = \mathrm{diag}(b_P(h_n))$,

$$
\Gamma_{x,n}(x) = \int b_P(u)b_P(u)'K(u)f_0(x+uh_n)du,
$$

$$
\vartheta_{P,n}(x) = \int b_P(u)[u]_{P+1}'K(u)f_0(x+uh_n)du,
$$

$$
\xi_{P,0}(x) = \left[q_0(x), q_0^{(1)}(x)', q_0^{(2)}(x)', \cdots, q_0^{(P)}(x)'\right]',
$$

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

Finally, we study the NL bias. We compute the quadratic functional first:

$$
\bar{G}_{n,\gamma\gamma}[\kappa,\lambda] = \mathbb{E}\bar{g}_{n,\gamma\gamma}(z)[\kappa,\lambda] = \sum_{\ell=1}^{10}\bar{G}_{n,\gamma\gamma,\ell}[\kappa,\lambda]
$$

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

$$
\bar{G}_{n,\gamma\gamma,1}[\kappa,\lambda] = -\int \frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\gamma_x(x))\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)dx,
$$

$$
\bar{G}_{n,\gamma\gamma,2}[\kappa,\lambda] = -\int \frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\gamma_x(x))\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)dx,
$$

$$
\bar{G}_{n,\gamma\gamma,3}[\kappa,\lambda] = \int \frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\gamma_x(x))\Gamma_{x,n}^+(x)^{-1}\eta_t(x),
$$

$$
\bar{G}_{n,\gamma\gamma,4}[\kappa,\lambda] = \int \frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\operatorname{vec}_P^{-1}(\eta_x(x))\Gamma_{x,n}^+(x)^{-1}\gamma_t(x)dx,
$$

$$\bar{G}_{n,\gamma\gamma,5}[\kappa,\lambda] = \int \frac{w_0(x)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \gamma_t(x) \eta_t(x)' \Gamma_{x,n}^+(x)^{-1} e_P dx,$$

$$\bar{G}_{n,\gamma\gamma,6}[\kappa,\lambda] = \int \frac{w_0(x)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \eta_t(x) \gamma_t(x)' \Gamma_{x,n}^+(x)^{-1} e_P dx,$$

$$\begin{aligned}&\bar{G}_{n,\gamma\gamma,7}[\kappa,\lambda]\\ &= \int \frac{w_0(x)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \Gamma_{x,n}^+(x)^{-1} e_P,\end{aligned}$$

$$\begin{aligned}&\bar{G}_{n,\gamma\gamma,8}[\kappa,\lambda]\\ &= \int \frac{w_0(x)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \Gamma_{x,n}^+(x)^{-1} e_P dx,\end{aligned}$$

$$\bar{G}_{n,\gamma\gamma,9}[\kappa,\lambda] = -\int \frac{2w_0(x)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \gamma_t(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \eta_x(x) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x) dx,$$

$$\bar{G}_{n,\gamma\gamma,10}[\kappa,\lambda] = -\int \frac{2w_0(x)}{q_n^+(x)^3} e_P' \Gamma_{x,n}^+(x)^{-1} \eta_t(x) \gamma_{t,n}^+(x)' \Gamma_{x,n}^+(x)^{-1} \gamma_x(x) \Gamma_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x) dx.$$

Next, using standard bounding arguments, we have

$$\mathbb{E}[\|\bar{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]\|^2] = O\left(\frac{1}{nh_n^{2d}}\right),$$

and

$$\mathbb{E}[\|\bar{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^j - \gamma_n]\|^2] = O\left(\frac{1}{nh_n^{d}}\right),$$

with $i \neq j$. Therefore, we need to analyze

$$\begin{aligned}&\frac{1}{2}\bar{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]\\ &= \bar{G}_{n,\gamma\gamma,1}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] + \bar{G}_{n,\gamma\gamma,3}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] + \bar{G}_{n,\gamma\gamma,5}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]\\ &\quad + \bar{G}_{n,\gamma\gamma,7}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] + \bar{G}_{n,\gamma\gamma,9}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n]\end{aligned}$$

and hence we have only 5 terms to consider. To give some interpretation to the kernel constants, we define:

$$\Gamma_{x,r} = \int_{\mathbb{R}^d} b_P(u) b_P(u)' K(u)^r du.$$

For the first term,

$$
\begin{aligned}
&\mathbb{E}\bar{G}_{n,\gamma\gamma,1}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n]\\
&=-\mathbb{E}\int\frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i-x)\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i-x)\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)dx+O(1)\\
&=-\int\int\frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(u-x)\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(u-x)\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)dxf_0(u)du+O(1)\\
&=-\frac{1}{h_n^d}\left[\int\int\frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_x(v)\Gamma_{x,n}^+(x)^{-1}\bar{K}_x(v)H_n\xi_{P,0}(x)f_0(x+vh_n)dxdv+O(h_n^{P+1})\right]\\
&=-\frac{1}{h_n^d}\left[\int r_0(x;\theta_0)dx\int e_P'\Gamma_x^{-1}\bar{K}_x(v)\Gamma_x^{-1}\bar{K}_x(v)e_Pdv+O(n^{-1}h_n^d+h_n)\right]+O(1)\\
&=-\frac{1}{h_n^d}\left(\int r_0(x;\theta_0)dx\right)\left(e_P'\Gamma_x^{-1}\int b_P(v)b_P(v)'\Gamma_x^{-1}b_P(v)K(v)^2dv\right)+\frac{1}{h_n^d}O(n^{-1}h_n^d+h_n),
\end{aligned}
$$

where we use

$$
\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)=H_n\xi_{P,0}(x)+O(n^{-1}h_n^{-d}+h_n^{p+1}),
$$

and where the $O(1)$ terms capture the centering terms. That is, the three additional terms are:

$$
\begin{aligned}
&-\mathbb{E}\int\frac{q_0(x)}{q_n^+(x)^2}r_0(x;\theta_0)e_P'\Gamma_{x,n}^+(x)^{-1}\Gamma_{x,n}(x)\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i-x)\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)f_0(x)dx,\\
&-\mathbb{E}\int\frac{q_0(x)}{q_n^+(x)^2}r_0(x;\theta_0)e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i-x)\Gamma_{x,n}^+(x)^{-1}\Gamma_{x,n}(x)\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)f_0(x)dx,\\
&+\mathbb{E}\int\frac{q_0(x)}{q_n^+(x)^2}r_0(x;\theta_0)e_P'\Gamma_{x,n}^+(x)^{-1}\Gamma_{x,n}(x)\Gamma_{x,n}^+(x)^{-1}\Gamma_{x,n}(x)\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)f_0(x)dx.
\end{aligned}
$$

For the second term,

$$
\begin{aligned}
&\mathbb{E}\bar{G}_{n,\gamma\gamma,3}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n]\\
&=\mathbb{E}\int\frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i-x)\Gamma_{x,n}^+(x)^{-1}\bar{K}_{y,n}(x_i-x)t_idx+O(1)\\
&=\mathbb{E}\int\frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i-x)\Gamma_{x,n}^+(x)^{-1}\bar{K}_{t,n}(x_i-x)q_0(x_i)dx+O(1)\\
&=\int\int\frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(u-x)\Gamma_{x,n}^+(x)^{-1}\bar{K}_{t,n}(x_i-x)q_0(u)f_0(u)dudx+O(1)\\
&=\frac{1}{h_n^d}\int\int\frac{w_0(x)}{q_n^+(x)^2}e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_x(v)\Gamma_{x,n}^+(x)^{-1}b_P(v)K(v)q_0(x+vh_n)f_0(x+vh_n)dvdx+O(1)\\
&=\frac{1}{h_n^d}\left[\int r_0(x;\theta_0)dx\int e_P'\Gamma_x^{-1}\bar{K}_x(v)\Gamma_x^{-1}b_P(v)K(v)dv+O(n^{-1}h_n^d+h_n)\right]\\
&=\frac{1}{h_n^d}\left(\int r_0(x;\theta_0)dx\right)\left(e_P'\Gamma_x^{-1}\int b_P(v)b_P(v)'\Gamma_x^{-1}b_P(v)K(v)^2dv\right)+\frac{1}{h_n^d}O(n^{-1}h_n^d+h_n).
\end{aligned}
$$

Therefore, we find:

$$\mathbb{E}\bar{G}_{n,\gamma\gamma,1}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] + \mathbb{E}\bar{G}_{n,\gamma\gamma,3}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] = \frac{1}{h_n^d} O(n^{-1}h_n^d + h_n).$$

For the third term,

$$\begin{aligned}
&\mathbb{E}\bar{G}_{n,\gamma\gamma,5}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] \\
&= \mathbb{E}\int \frac{w_0(x)}{q_n^+(x)^3} e_0'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{t,n}(x_i - x)t_i^2\bar{K}_{t,n}(x_i - x)'\Gamma_{x,n}^+(x)^{-1}e_0 dx + O(1) \\
&= \frac{1}{h_n^d}\int \frac{w_0(x)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}b_P(v)K(v)^2 b_P(v)'\Gamma_{x,n}^+(x)^{-1}e_P q_0(x+vh_n)f_0(x+vh_n)dudx + O(1) \\
&= \frac{1}{h_n^d}\left[\int \frac{1}{q_0(x)} r_0(x;\theta_0)dx \int e_P'\Gamma_x^{-1}b_P(v)K(v)^2 b_P(v)'\Gamma_x^{-1}e_P dv + O(n^{-1}h_n^d + h_n)\right] \\
&= \frac{1}{h_n^d}\left(\int \frac{1}{q_0(x)} r_0(x;\theta_0)dx\right)\left(e_P'\Gamma_{x,1}^{-1}\Gamma_{x,2}\Gamma_{x,1}^{-1}e_P\right) + \frac{1}{h_n^d}O(n^{-1}h_n^d + h_n).
\end{aligned}$$

For the fourth term,

$$\begin{aligned}
&\mathbb{E}\bar{G}_{n,\gamma\gamma,7}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] \\
&= \mathbb{E}\int \frac{w_0(x)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i - x)\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)\gamma_{t,n}^+(x)'\Gamma_{x,n}^+(x)^{-1}\bar{K}_{x,n}(x_i - x)\Gamma_{x,n}^+(x)^{-1}e_P dx \\
&= \frac{1}{h_n^d}\left[\int\int \frac{w_0(x)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}\bar{K}_x(v)H_n\xi_{P,0}(x)\xi_{P,0}(x)'H_n\bar{K}_x(v)\Gamma_{x,n}^+(x)^{-1}e_P f_0(x+vh_n)dxdv + O(h_n^{p+1})\right] \\
&= \frac{1}{h_n^d}\left[\int r_0(x;\theta_0)dx \int e_P'\Gamma_x^{-1}\bar{K}_x(v)e_P e_P'\bar{K}_x(v)\Gamma_x^{-1}e_P dv + O(n^{-1}h_n^d + h_n)\right] \\
&= \frac{1}{h_n^d}\left(\int r_0(x;\theta_0)dx\right)\left(e_P'\Gamma_{x,1}^{-1}\Gamma_{x,2}\Gamma_{x,1}^{-1}e_P\right) + \frac{1}{h_n^d}O(n^{-1}h_n^d + h_n).
\end{aligned}$$

For the last term,

$$\begin{aligned}
&\mathbb{E}\bar{G}_{n,\gamma\gamma,9}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] \\
&= -\mathbb{E}\int \frac{2w_0(x)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}\bar{\mathcal{K}}_{t,n}(x_i - x)q_0(x_i)\gamma_{t,n}^+(x)'\Gamma_{x,n}^+(x)^{-1}\bar{\mathcal{K}}_{x,n}(x_i - x)\Gamma_{x,n}^+(x)^{-1}\gamma_{t,n}^+(x)dx \\
&= -\frac{2}{h_n^d}\int\int \frac{w_0(x)}{q_n^+(x)^3} e_P'\Gamma_{x,n}^+(x)^{-1}b_P(v)K(v)\xi_{P,0}(x)'H_n\bar{\mathcal{K}}_x(v)H_n\xi_{P,0}(x)q_0(x+vh_n)f_0(x+vh_n)dvdx + \frac{1}{h_n^d}O(h_n^{P+1}) \\
&= -\frac{2}{h_n^d}\left(\int r_0(x;\theta_0)f_0(x)dx\right)\left(\int e_P'\Gamma_x^{-1}b_P(v)K(v)^2 dv\right) + \frac{1}{h_n^d}O(n^{-1}h_n^d + h_n) \\
&= \frac{1}{h_n^d}O(n^{-1}h_n^d + h_n),
\end{aligned}$$

because

$$\int r_0(x;\theta_0)f_0(x)dx = \mathbb{E}[r_0(x;\theta_0)] = 0.$$

Therefore,

$$\begin{aligned}
\mathcal{B}_n^{\texttt{NL}} &= \frac{1}{2n}\mathbb{E}\bar{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n] \\
&= \frac{1}{nh_n^d}\left(\int \frac{1+q_0(x)}{q_0(x)} r_0(x;\theta_0)dx\right)\left(e_P'\Gamma_{x,1}^{-1}\Gamma_{x,2}\Gamma_{x,1}^{-1}e_P\right) + O(n^{-1} + h_n^{1-d})
\end{aligned}$$

The final rate restrictions are: $nh_n^{3d/2}/(\log n)^{1/2} \to \infty$ and $nh_n^{2P+2} \to 0$, which forces

$$p > \frac{3}{4}d - 1.$$

Under these conditions, we obtain

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

with

$$\mathcal{B}_n = \mathcal{B}_n^{\texttt{LI}} + \mathcal{B}_n^{\texttt{NL}} = O(1/(nh_n^d)).$$

SA.2.4 Bandwidth Selection

As in the previous example, We can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

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

where the small bandwidth bias is

$$\begin{aligned}
\mathcal{B}_0^{\texttt{SB}} &= \mathcal{B}_0^{\texttt{LI}} + \mathcal{B}_0^{\texttt{NL}} \\
&= \left(-\int \frac{1}{q_0(x)} r_0(x;\theta_0)dx\right) K(0)\left(e_P'\Gamma_{x,1}^{-1}e_P\right) \\
&\quad + \left(\int \frac{1+q_0(x)}{q_0(x)} r_0(x;\theta_0)dx\right)\left(e_P'\Gamma_{x,1}^{-1}\Gamma_{x,2}\Gamma_{x,1}^{-1}e_P\right)
\end{aligned}$$

and the smoothing bias is

$$\mathcal{B}_0^{\texttt{S}} = -\left(\int \frac{1}{q_0(x)} r_0(x;\theta_0) q_0^{(P+1)}(x)' f_0(x)dx\right)\left(\vartheta_P'\Gamma_{x,1}^{-1}e_P\right).$$

SA.2.5 Condition AL*

We apply Lemma 4 with $\rho = 3$ to verify Condition AL*, following as close as possible our calculations above for Lemma 1. We have $\hat{\theta}_n^* - \theta_0 = o_{\mathbb{P}}(1)$, which can be established using classical results in the literature. Using Lemma SA-1 we can (give primitive conditions and) verify

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

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

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

Condition (i*). Holds by definition of the bootstrap analogue estimator.

Condition (ii*). Is verified exactly like condition (ii) in Lemma 1 was verified above.

Condition (iii*). We have

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

where

$$\Delta_{1,n}^*(\theta) = \Delta_{11,n}^*(\theta) + \Delta_{11,n}(\theta) + \Delta_{12,n}^*(\theta) + \Delta_{13,n}^*(\theta)$$

with

$$\Delta_{11,n}^*(\theta) = \left\|\frac{1}{n}\sum_{i=1}^n \left\{\frac{t_i^*}{q_n(x_i^*)}(m(y_i^*;\theta) - m(y_i^*;\theta_0)) - \frac{t_i}{q_n(x_i)}(m(y_i;\theta) - m(y_i;\theta_0))\right\}\right\|$$

$$\Delta_{11,n}(\theta) = \left\|\frac{1}{n}\sum_{i=1}^n \frac{t_i}{q_n(x_i)}(m(y_i;\theta) - m(y_i;\theta_0)) - \mathbb{E}\left[\frac{t_i}{q_n(x_i)}(m(y_i;\theta) - m(y_i;\theta_0))\right]\right\|,$$

$$\Delta_{12,n}^*(\theta) = \left(\max_{1\leq i\leq n} \frac{\|\hat{q}_n^*(x_i^*) - q_n(x_i^*)\|}{\|\hat{q}_n^*(x_i^*)q_n(x_i^*)\|}\right)\left(\frac{1}{n}\sum_{i=1}^n t_i^*\,\|m(y_i^*;\theta) - m(y_i^*;\theta_0)\| - t_i\,\|m(y_i;\theta) - m(y_i;\theta_0)\|\right),$$

$$\Delta_{13,n}^*(\theta) = \left(\max_{1\leq i\leq n} \frac{\|\hat{q}_n^*(x_i^*) - q_n(x_i^*)\|}{\|\hat{q}_n^*(x_i^*)q_n(x_i^*)\|}\right)\left(\frac{1}{n}\sum_{i=1}^n t_i\,\|m(y_i;\theta) - m(y_i;\theta_0)\| - \mathbb{E}\left[t_i\|m(y_i;\theta) - m(y_i;\theta_0)\|\right]\right),$$

and

$$\Delta_{2,n}^*(\theta) = \Delta_{21,n}^*(\theta) + \Delta_{22,n}^*(\theta)$$

with

$$\Delta^*_{21,n}(\theta) = \int \frac{\|q(x;\hat{\gamma}^*_n) - q(x;\gamma_n)\|}{\|q(x;\hat{\gamma}^*_n)q(x;\gamma_n)\|} \|r_0(x;\theta) - r_0(x;\theta_0)\| f_0(x)dx,$$

$$\Delta^*_{22,n}(\theta) = \left(\max_{1\leq i\leq n} \frac{\|\hat{q}^*_n(x^*_i) - q_n(x^*_i)\|}{\|\hat{q}^*_n(x^*_i)q_n(x^*_i)\|}\right) \mathbb{E}\left[t_i\|m(y_i;\theta) - m(y_i;\theta_0)\|\right]$$

For the first new term, for every $\delta_n = o(1)$, we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta^*_{11,n}(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

because $q_n(x)$ is non-random, $q_n(x)$ is bounded away from zero for all n large enough, and the class of n-varying functions $\mathcal{M}_n = \{t_i m(y_i;\theta)/q_n(x_i) : \theta \in \Theta\}$ satisfies easily the integral entropy condition $J_{[]}(\epsilon_n, \mathcal{M}_n, L_2(P)) = 0$ for all $\epsilon_n \downarrow 0$, using standard results for bootstrap empirical processes.

For the second term, for every $\delta_n = o(1)$, we showed above that

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta_{11,n}(\theta) = o_{\mathbb{P}}(n^{-1/2}).$$

For the third term, for every $\delta_n = o(1)$,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta^*_{12,n}(\theta) = o_{\mathbb{P}}(n^{-1/2}),$$

because $\max_{1\leq i\leq n} \|\hat{q}_n(x_i) - q_n(x_i)\| = o_{\mathbb{P}}(1)$, $q_n(x)$ is bounded away from zero for all n large enough, and the class of functions $\mathcal{M}_{\|\cdot\|} = \{t_i \|m(y_i;\theta) - m(y_i;\theta_0)\| : \theta \in \Theta\}$ satisfies the integral entropy condition $J_{[]}(1, \mathcal{M}_{\|\cdot\|}, L_2(P)) < \infty$.

For the fourth term, for every $\delta_n = o(1)$, using the same arguments as above we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \Delta^*_{13,n}(\theta) = o_{\mathbb{P}}(n^{-1/2}).$$

For the fifth term, for every $\delta_n = o(1)$, we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\Delta^*_{21,n}(\theta)}{1 + n^{1/3}\|\theta - \theta_0\|} = O_{\mathbb{P}}(n^{-1/3}) \int \|q(x;\hat{\gamma}^*_n) - q(x;\gamma_n)\| \|\dot{r}_0(x;\theta_0)\| f_0(x)dx = o_{\mathbb{P}}(n^{-1/2})$$

because $\sup_{x\in\mathbb{X}} \|\hat{q}_n(x) - q_n(x)\| = o_{\mathbb{P}}(1)$, $q_n(x)$ is bounded away from zero for all n large enough, $\|r_0(x;\theta) - r_0(x;\theta_0)\| \preceq \|\theta - \theta_0\|$, and

$$\int \|q(x;\hat{\gamma}^*_n) - q(x;\gamma_n)\|^2 f_0(x)dx = O_{\mathbb{P}}\left(\frac{1}{nh_n^d}\right) = o_{\mathbb{P}}(n^{-1/3}),$$

as discussed previously.

For the sixth term, for every $\delta_n = o(1)$, we have

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\Delta^*_{22,n}(\theta)}{1+n^{1/3}\|\theta-\theta_0\|} = O_{\mathbb{P}}(n^{-1/3}) \max_{1\leq i\leq n} \|\hat{q}^*_n(x^*_i) - q_n(x^*_i)\| = o_{\mathbb{P}}(n^{-1/2}),$$

by the same arguments given above.

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

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

where

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

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

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

$$\begin{aligned}
&\left\| \frac{1}{n}\sum_{i=1}^n \frac{t^*_i}{\hat{q}_n(x^*_i)^3\hat{q}^*_n(x^*_i)} m(y^*_i;\theta_0)[\hat{q}^*_n(x^*_i) - \hat{q}_n(x^*_i)]^3 \right\| \\
&\leq \frac{1}{n}\sum_{i=1}^n \frac{t^*_i}{\hat{q}_n(x^*_i)^3|\hat{q}^*_n(x^*_i)|} \|m(y^*_i;\theta_0)\| \|\hat{q}^*_n(x^*_i) - \hat{q}_n(x^*_i)\|^3 \\
&\preceq \max_{1\leq i\leq n} \|\hat{\gamma}^*_n(x^*_i) - \hat{\gamma}_n(x^*_i)\|^3 \frac{1}{n}\sum_{i=1}^n t^*_i \|m(y^*_i;\theta_0)\| = O_{\mathbb{P}}\left(\frac{(\log n)^{3/2}}{n^{3/2}h_n^{3d/2}}\right) = o_{\mathbb{P}}(n^{-1/3}),
\end{aligned}$$

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

Condition (v*). Is verified exactly like condition (v) in Lemma 1 was verified above.

Condition (vi*). Is verified exactly like condition (vi) in Lemma 1 was verified above.

Condition (vii*). Follows from the results used to verify Condition (iii*) above, as discussed when verifying Condition (vii) previously.

SA.2.6 Condition AS*

A "quadratic" (in $\gamma - \hat{\gamma}_n$) approximation to $g_n^*(z, \gamma)$ is given by

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

where the linear term is

$$\begin{aligned} \bar{g}_{n,\gamma}^*(z)[\gamma] &= -\frac{t}{\hat{q}_n^+(x)^2} m(y;\theta_0) e_P' \hat{\Gamma}_{x,n}^+(x)^{-1} \gamma_t(x) \\ &\quad + \frac{t}{\hat{q}_n^+(x)^2} m(y;\theta_0) e_P' \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} \hat{\gamma}_{t,n}^+(x), \end{aligned}$$

with $\gamma = (\gamma_x', \gamma_t')'$, $\hat{\Gamma}_{x,n}^+(x) = (n-1)^{-1} K_n(0) e_P e_P' + \hat{\Gamma}_{x,n}(x)$, $\hat{\gamma}_{t,n}^+(x) = (n-1)^{-1} K_n(0) e_P + \hat{\gamma}_{t,n}(x)$ and $\hat{q}_n^+(x) = e_P' \hat{\Gamma}_{x,n}^+(x)^{-1} \hat{\gamma}_{t,n}^+(x)$, and the quadratic term is

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

with

$$\bar{g}_{n,\gamma\gamma,1}^*(z)[\gamma, \eta] = -\frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2} e_P' \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} \hat{\gamma}_{t,n}^+(x),$$

$$\bar{g}_{n,\gamma\gamma,2}^*(z)[\gamma, \eta] = -\frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2} e_P' \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} \hat{\gamma}_{t,n}^+(x),$$

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

$$\bar{g}_{n,\gamma\gamma,4}^*(z)[\gamma, \eta] = \frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^2} e_P' \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} \gamma_t(x),$$

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

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

$$\begin{aligned} &\bar{g}_{n,\gamma\gamma,7}^*(z)[\gamma, \eta] \\ &= \frac{tm(y;\theta_0)}{\hat{q}_n^+(x)^3} e_P' \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\gamma_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} \gamma_{t,n}^+(x) \gamma_{t,n}^+(x)' \hat{\Gamma}_{x,n}^+(x)^{-1} \operatorname{vec}_P^{-1}(\eta_x(x)) \hat{\Gamma}_{x,n}^+(x)^{-1} e_P, \end{aligned}$$

$$\bar{g}^*_{n,\gamma\gamma,8}(z)[\gamma,\eta]$$
$$= \frac{tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e'_P \hat{\Gamma}^+_{x,n}(x)^{-1} \operatorname{vec}^{-1}_P(\eta_x(x)) \hat{\Gamma}^+_{x,n}(x)^{-1} \gamma^+_{t,n}(x) \gamma^+_{t,n}(x)' \hat{\Gamma}^+_{x,n}(x)^{-1} \operatorname{vec}^{-1}_P(\gamma_x(x)) \hat{\Gamma}^+_{x,n}(x)^{-1} e_P,$$

$$\bar{g}^*_{n,\gamma\gamma,9}(z)[\gamma,\eta] = -\frac{2tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e'_P \hat{\Gamma}^+_{x,n}(x)^{-1} \gamma_t(x) \gamma^+_{t,n}(x)' \hat{\Gamma}^+_{x,n}(x)^{-1} \eta_x(x) \hat{\Gamma}^+_{x,n}(x)^{-1} \gamma^+_{t,n}(x),$$

$$\bar{g}^*_{n,\gamma\gamma,10}(z)[\gamma,\eta] = -\frac{2tm(y;\theta_0)}{\hat{q}^+_n(x)^3} e'_P \hat{\Gamma}^+_{x,n}(x)^{-1} \eta_t(x) \gamma^+_{t,n}(x)' \hat{\Gamma}^+_{x,n}(x)^{-1} \gamma_x(x) \hat{\Gamma}^+_{x,n}(x)^{-1} \gamma^+_{t,n}(x),$$

Recall that $h_n \to 0$ and

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

using, for example, Lemma SA-1 below. Then, the first part of Condition AS* is satisfied provided that

$$\max_{1\le i,j\le n} \|\hat{\gamma}^{(i),*}_n(x^*_j) - \hat{\gamma}_n(x^*_j)\|^3 = o_{\mathbb{P}}(n^{-1/2}),$$

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

The second part of Condition AS* is satisfied provided $nh_n^d \to \infty$, because it can be shown that

$$\mathbb{V}^*[\bar{g}^*_{n,\gamma}(z^*_i)[\hat{\gamma}^{j,*}_n - \hat{\gamma}_n]] = O_{\mathbb{P}}\left[\frac{1}{h_n^d}\left(1 + \frac{1}{nh_n^d}\right)\right],$$

$$\mathbb{V}^*[\bar{g}^*_{n,\gamma\gamma}(z^*_i)[\hat{\gamma}^{j,*}_n - \hat{\gamma}_n, \hat{\gamma}^{k,*}_n - \hat{\gamma}_n]] = O_{\mathbb{P}}\left[\frac{1}{h_n^{2d}}\left(1 + \frac{1}{nh_n^d}\right)^2\right].$$

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

$$\mathbb{V}^*[\bar{g}^*_{n,\gamma\gamma}(z^*_i)[\hat{\gamma}^{j,*}_n - \hat{\gamma}_n, \hat{\gamma}^{j,*}_n - \hat{\gamma}_n]] = O_{\mathbb{P}}\left[\frac{1}{h_n^{3d}}\left(1 + \frac{1}{nh_n^d}\right)\right],$$

where, as always, $i \neq j \neq k$.

Putting the above together, Lemma 5 implies Condition AS* if $h_n \to 0$ and $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

SA.2.7 Condition AN*

First we show that

$$\frac{1}{n}\sum_{i=1}^n \|\psi^*_n(z_i) - \psi_n(z_i)\|^2 \to_{\mathbb{P}} 0,$$

where

$$\psi^*_n(z) = g^*_n(z,\hat{\gamma}_n) - \mathbb{E}^*[g^*_n(z^*,\hat{\gamma}_n)] + \delta^*_n(z), \qquad \delta^*_n(z_i) = \bar{G}^*_{n,\gamma}[\hat{\gamma}^i_n - \gamma_n],$$

and hence

$$\begin{aligned}\delta_n^*(z) &= -\frac{1}{n}\sum_{i=1}^n \frac{t_i m(y_i;\theta_0)}{\hat{q}_n^+(x_i)^2} e_P'\hat{\Gamma}_{x,n}^+(x_i)^{-1}\left[\bar{K}_{t,n}(x-x_i)t - \hat{\gamma}_{t,n}(x_i)\right] \\ &\quad +\frac{1}{n}\sum_{i=1}^n \frac{t_i m(y_i;\theta_0)}{\hat{q}_n^+(x_i)^2} e_P'\hat{\Gamma}_{x,n}^+(x_i)^{-1}\left[\bar{K}_{x,n}(x-x_i) - \hat{\Gamma}_{x,n}(x_i)\right]\hat{\Gamma}_{x,n}^+(x_i)^{-1}\hat{\gamma}_{t,n}^+(x_i).\end{aligned}$$

Using the fact that $\hat{\theta}_n \to_{\mathbb{P}} \theta_0$ and $\max_{1\le j\le n}\|\hat{\gamma}_n(x_j) - \gamma_n(x_j)\| = o_{\mathbb{P}}(1)$, it can be shown that

$$\frac{1}{n}\sum_{i=1}^n \|g_n^*(z_i,\hat{\gamma}_n) - g_n(z_i,\gamma_n)\|^2 \to_{\mathbb{P}} 0,$$

and

$$\mathbb{E}^*[g_n^*(z^*,\hat{\gamma}_n)] - \mathbb{E}[g_n(z,\gamma_n)] \to_{\mathbb{P}} 0.$$

Also, it can be shown that

$$\max_{1\le i\le n}\|\delta_n^*(z_i) - \delta_n(z_i)\| \to_{\mathbb{P}} 0,$$

implying in particular that

$$\frac{1}{n}\sum_{i=1}^n \|\delta_n^*(z_i) - \delta_n(z_i)\|^2 \to_{\mathbb{P}} 0.$$

As a consequence, if $h_n \to 0$ and $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$, the first part of Condition AN holds.

Finally, employing direct calculations for kernel-based estimators we verify

$$\mathbb{V}^*(G_{n,\gamma\gamma}^*[\hat{\gamma}_n^{i,*} - \hat{\gamma}_n, \hat{\gamma}_n^{i,*} - \hat{\gamma}_n]) = O_{\mathbb{P}}(1/h_n^{2d}),$$

$$\mathbb{V}^*(G_{n,\gamma\gamma}^*[\hat{\gamma}_n^{i,*} - \hat{\gamma}_n, \hat{\gamma}_n^{j,*} - \hat{\gamma}_n]) = O_{\mathbb{P}}(1/h_n^{d}),$$

and

$$\begin{aligned}\mathbb{E}^*[B_n^*] &= \frac{1}{n}\sum_{i=1}^n g_n^*(z_i,\hat{\gamma}_n) + \frac{1}{2n}\sum_{i=1}^n \mathbb{E}^*\left[\bar{G}_{n,\gamma\gamma}^*[\hat{\gamma}_n^{*,(i)} - \hat{\gamma}_n, \hat{\gamma}_n^{*,(i)} - \hat{\gamma}_n]\right] \\ &= \mathcal{B}_n^{\mathtt{LI}} + \mathcal{B}_n^{\mathtt{NL}} + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

provided that $h_n \to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

Therefore, using Lemma 6, Condition AN* holds if $h_n \to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

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 to compare our results with previous influential work in literature 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 $z_1, \dots, z_n$ 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. The estimand is

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

and the corresponding estimator is

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

where $\mathbb{1}(\cdot)$ is the indicator function. In this example, $z = (y, x')'$, $x(z, \theta) = x$, $w(z, \theta) = 1$, $\gamma_0(\cdot, \theta) = \gamma_0(\cdot)$, and $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = 0$, where $g(z, \theta, \gamma) = \mathbb{1}[y \geq \gamma(z)] - \theta$, and $\gamma = f$.

To obtain primitive bandwidth conditions for the conditions of Theorems 1 and 2, suppose that for some $P > 3d/4$, the following regularity conditions hold:

- γ_0 is P times differentiable, and γ_0 and its first P 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 $\int_{\mathbb{R}^d} |K(u)|(1 + \|u\|^P)du < \infty$ and

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

Let $f_{y|x}(\cdot|x)$ and $\dot{f}_{y|x}(\cdot|x)$ denote its first and second derivatives of $F_{y|x}(\cdot|x)$. As in the average density example, the smoothness assumptions can be relaxed but this is not the main focus of our paper.

SA.3.1 Condition AL

This condition holds with $\mathcal{J}_0 = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms. Therefore, $\mathfrak{B}_n = \mathcal{B}_n$ and $\Sigma_0 = \Omega_0$.

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 $\mathcal{X}_n=(x_1,\ldots,x_n)'$ and $\hat{\gamma}_n^{(i)}(x)=(n-1)^{-1}\sum_{j=1,j\neq i}^{n}K_n(x-x_j)$. 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

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

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

where $\gamma_n^{+}(x)=n^{-1}K_n(0)+(1-n^{-1})\gamma_n(x),$. It follows from standard bounding arguments that

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

provided $\Delta_n=o_{\mathbb{P}}(n^{-1/6})$. These results, which employ Lemma SA-1, verify the first part of Condition AS, provided that $nh_n^{3d/2}/(\log n)^{3/2}\to\infty$. Moreover,

$$\mathbb{V}(\breve{g}_{n,\gamma}(x_i)[\hat{\gamma}_n^{j}-\gamma_n])=O(1/h_n^{d}),$$

$$\mathbb{V}(\breve{g}_{n,\gamma\gamma}(x_i)[\hat{\gamma}_n^{j}-\gamma_n,\hat{\gamma}_n^{k}-\gamma_n])=O(1/h_n^{2d}),$$

$$\mathbb{V}[\mathbb{E}(\breve{g}_{n,\gamma\gamma}(x_i)[\hat{\gamma}_n^{j}-\gamma_n,\hat{\gamma}_n^{j}-\gamma_n]|z_i)]=O(1/h_n^{2d}),$$

$$\mathbb{V}(\breve{g}_{n,\gamma\gamma}(x_i)[\hat{\gamma}_n^{j}-\gamma_n,\hat{\gamma}_n^{j}-\gamma_n])=O(1/h_n^{3d}),$$

where, as before, $i\neq j\neq k$ and hence Condition AS holds via Lemma 2, provided $nh_n^{d}\to\infty$.

SA.3.3 Condition AN

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

where
$$\psi_n(z)=g_n(z,\gamma_n)-\mathbb{E}g_n(z,\gamma_n)+\delta_n(x),$$
$$\delta_n(x)=-(1-n^{-1})\int_{\mathbb{R}^d}f_{y|x}[\gamma_n^+(r)|r][K_n(r-x)-\gamma_n(r)]\gamma_0(r)dr,$$

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

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

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

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 conditions of Lemma 3 are satisfied if $h_n\to 0$ and if $nh_n^d\to\infty$ because the representation

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

can be used to show that

$$\mathbb{E}(\|\breve{G}_{n,\gamma\gamma}[\hat{\gamma}_n^i-\gamma_n,\hat{\gamma}_n^i-\gamma_n]\|^2)=O(1/h_n^{2d}),$$

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

with $i\neq j$. Condition (AN) is therefore satisfied with $\Sigma_0=\mathbb{E}[\psi_0(z)^2]$, if $h_n\to 0$ and if $nh_n^d\to\infty$. It can be shown that, under the bandwidth conditions imposed, that $\hat{\mathcal{B}}_n=\mathcal{B}_n+o_{\mathbb{P}}(n^{-1/2})$, where $\mathcal{B}_n=\mathcal{B}_n^{\texttt{LI}}+\mathcal{B}_n^{\texttt{NL}}+\mathcal{B}_n^{\texttt{S}}$ with

$$\mathcal{B}_n^{\texttt{LI}}=-\frac{1}{nh_n^d}K(0)\left(\int_{\mathbb{R}^d}f_{y|x}[\gamma_0(x)|x]\gamma_0(x)dx\right),$$

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

and $\mathcal{B}_n^{\texttt{S}}=O(h_n^P)$. It follows that $\mathcal{B}_n^{\texttt{NL}}$ admits a polynomial-in-h_n expansion of the form

$$\mathcal{B}_n^{\texttt{NL}}=\frac{1}{nh_n^d}[\mathcal{B}_0^{\texttt{NL}}+\mathcal{B}_1^{\texttt{NL}}h_n^2+\mathcal{B}_2^{\texttt{NL}}h_n^4+\ldots],$$

the constants $\mathcal{B}_0^{\mathtt{NL}}, \mathcal{B}_1^{\mathtt{NL}}, \mathcal{B}_2^{\mathtt{NL}}, \ldots$ being functionals of K and the data generating process. Symmetry of the kernel implies that $\mathcal{B}_n^{\mathtt{NL}}$ is of order $1/(nh_n^d)$ and that the polynomial expansion of $nh_n^d\mathcal{B}_n^{\mathtt{NL}}$ involves only even powers of h_n.

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 $\mathfrak{B}_n = \mathcal{B}_n = \mathcal{B}_n^{\mathtt{LI}} + \mathcal{B}_n^{\mathtt{NL}} = O(1/(nh_n^d))$.

SA.3.4 Bandwidth Selection

As before, we have $\mathcal{B}_n^{\mathtt{S}} = h^P \mathcal{B}_0^{\mathtt{S}} + o(h^P)$ with

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

Therefore, in this example we can balance the leading bias terms to obtain a (second-order) optimal bandwidth selector:

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

where the small bandwidth bias is

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

and the smoothing bias is given above.

SA.3.5 Condition AL*

This condition holds with $\mathcal{J}_0 = I_{d_\theta}$, with $d_\theta = 1$ in this example, and without any $o_{\mathbb{P}}(1)$ terms. Therefore, $\mathfrak{B}_n = \mathcal{B}_n$ and $\Sigma_0 = \Omega_0$.

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

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

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

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

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

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

where $\mathcal{X}_n^* = (x_1^*, \ldots, x_n^*)'$. Also, it follows from standard bounding arguments that

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

provided $\Delta_n^* = o_{\mathbb{P}}(n^{-1/6})$. The latter rate result can be easily verified using Lemma SA-1. The above results verify 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

$$\mathbb{V}^*(\breve{g}_{n,\gamma}^*(x_i^*)[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n]) = O_{\mathbb{P}}(1/h_n^d),$$

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

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

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

so that the conditions of Lemma 5 also hold, and hence the second part of Condition AS* will be satisfied provided $nh_n^d \to \infty$.

SA.3.7 Condition AN*

We have:

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

where

$$\psi_n^*(z) = g_n^*(z, \hat{\gamma}_n) - \frac{1}{n}\sum_{i=1}^n g_n^*(z_i, \hat{\gamma}_n) + \delta_n^*(z),$$

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

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

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 $n^{-1}\sum_{i=1}^n \|\psi_n^*(z_i) - \psi_n(z_i)\|^2 \to_{\mathbb{P}} 0$. Also, the conditions of Lemma 6 are satisfied because the representation

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

can be used to show that

$$\mathbb{V}^*(\|\breve{G}_{n,ff}^*[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,j} - \hat{\gamma}_n]\|^2) = O_{\mathbb{P}}(1/h_n^{2d}),$$

$$\mathbb{V}^*(\|\breve{G}_{n,ff}^*[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,k} - \hat{\gamma}_n]\|^2) = O_{\mathbb{P}}(1/h_n^{d}),$$

with $j \neq k$. Finally, it can be shown that $\mathbb{E}^*(\hat{\mathcal{B}}_n^*) = \mathcal{B}_n + o_p\left(n^{-1/2}\right)$ if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$. In other words, Condition AN* holds if $h_n \to 0$ and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

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, usually via Lemmas 1–6. 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

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

and

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

where $h_n = o(1)$ 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

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

and

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

Defining $\Psi_n(\cdot) = \mathbb{E}\hat{\Psi}_n(\cdot)$ 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

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

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 $\rho_n = \sqrt{\log n}/\sqrt{nh_n^d}$.

<The condition $\mathcal{C}(s) \leq H^s$ (for some $H < \infty$ and every s) is satisfied when W_n is bounded (uniformly in n), so part (c) can be used to analyze $\hat{f}_n$ and its derivative and we use this part in all of the examples. Part (b) covers certain distributions with full support (e.g., sub-Gaussian distributions), but is not used in our examples. On the other hand, the $S = 4$ version of part (a) is used to verify Condition (AN*) in Example 2.>

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

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

because $n\rho_n \to \infty$ and $\sup_{x\in\mathbb{R}^d}|\Psi_n(x)| \le \mathcal{C}(1)\int_{\mathbb{R}^d}|\mathcal{K}(t)|dt$. By Chebychev's inequality,

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

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

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

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

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

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

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

where C_τ is a constant to be chosen. We have

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

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

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

because

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

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

$$\max_{1\le i\le n}|\hat{\Psi}_n^{\tau,(i)}(X_{i,n})-\Psi_n^\tau(X_{i,n})|=O_p(\rho_n)$$

for every C_τ, where $\Psi_n^\tau(x)=\mathbb{E}\hat{\Psi}_n^\tau(x)=\mathbb{E}\hat{\Psi}_n^{\tau,(i)}(x)$.

For any M,

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

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

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

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),

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

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,

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

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

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

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

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

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*, pp. 47–61.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer, New York.