Appendix B Proofs of Theorems 1 - 3. (PDF file)

Proof of Theorem 1.

Let us consider θ_0 implicitly defined by (2) and fix some other θ. Then, using Jensen's inequality, (5) and (10),

$$\begin{aligned} l(\theta) - l(\theta_0) &= E(\tilde{l}_n(\theta_0)) - E(\tilde{l}_n(\theta)) = E\left(\log\left(\frac{f_\theta(Y|\theta' X)}{f_{\theta_0}(Y|\theta_0' X)} \right) \right) \\ &\leq \log\left(E\left(\frac{f_\theta(Y|\theta' X)}{f_{\theta_0}(Y|\theta_0' X)} \right) \right) \\ &= \log\left(\int \int \frac{f_\theta(y|\theta' \mathbf{x})}{f_{\theta_0}(y|\theta_0' \mathbf{x})} f_{\theta_0}(y|\theta_0' \mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} dy \right) \\ &= \log\left(\int \left(\int f_\theta(y|\theta' \mathbf{x}) dy \right) f_X(\mathbf{x}) d\mathbf{x} \right) = \log 1 = 0. \end{aligned}$$

So $l(\theta) \leq l(\theta_0)$ $\forall \theta$, which completes the proof. □

Proof of Lemma 1. Using (11) and (9) we have

$$l^{[1]}(\theta_0) = 0 \text{ and } \hat{l}_n^{[1]}(\hat{\theta}_n) = 0.$$

Now a Taylor expansion gives

$$l^{[1]}(\theta_0) = 0 = \hat{l}_n^{[1]}(\hat{\theta}_n) = \hat{l}_n^{[1]}(\theta_0) + \hat{l}_n^{[2]}(\hat{\theta}_n^*)(\hat{\theta}_n - \theta_0),$$

where $\hat{\theta}_n^*$ is between $\hat{\theta}_n$ and θ_0. This completes the proof. □

The proof of Theorem 2 is postponed. Let us first state and prove an auxiliary lemma.

Lemma 2. *Under the conditions in Theorem 3 we have*

$$\tilde{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0) \xrightarrow{P} 0$$

and

$$\sqrt{n}(\tilde{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0)) \to \mathcal{N}(0, \Sigma_1)$$

Proof. According to (5) and since $l^{[1]}(\theta_0) = E(\tilde{l}_n^{[1]}(\theta_0)) = 0$, we have

$$\sqrt{n}(\tilde{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{f_\theta^{[1]}(Y_i|\theta' X_i)_{\theta=\theta_0}}{f_{\theta_0}(Y_i|\theta_0' X_i)} \right].$$

Then, $\tilde{l}_n^{[1]}(\theta) - l^{[1]}(\theta)$ is a sum of i.i.d. random vectors and the law of large numbers gives the convergence to zero in probability. Now the central limit theorem leads to

$$\sqrt{n}(\tilde{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0)) \to \mathcal{N}(0, \Sigma_1)$$

The matrix Σ_1 is easily computed:

$$\begin{aligned}\Sigma_1 &= E\left[\left(\frac{f_\theta^{[1]}(Y_1|\theta' X_1)_{\theta=\theta_0}}{f_{\theta_0}(Y_1|\theta_0' X_1)}\right)\left(\frac{f_\theta^{[1]}(Y_1|\theta' X_1)_{\theta=\theta_0}}{f_{\theta_0}(Y_1|\theta' X_1)}\right)'\right] \\ &= \int (\nabla_\theta \log(f_\theta(z|\theta'\mathbf{x}))_{\theta=\theta_0})(\nabla_\theta \log(f_\theta(z|\theta'\mathbf{x}))_{\theta=\theta_0})' f(\mathbf{x}, z) d\mathbf{x} dz.\end{aligned}$$

Observe that the last equation is a consequence of Lemma 4 below. □

Proof of Theorem 3. In view of Lemma 1, the term

$$\hat{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0) = \alpha_n(\theta_0) + \beta_n(\theta_0),$$

has to be studied, where

$$\alpha_n(\theta_0) = \hat{l}_n^{[1]}(\theta_0) - \tilde{l}_n^{[1]}(\theta_0) \tag{20}$$

$$\beta_n(\theta_0) = \tilde{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0).$$

To deal with $\beta_n(\theta_0)$, we use Lemma 2, to prove

$$\sqrt{n}\beta_n(\theta_0) \to \mathcal{N}(0, \Sigma_1)$$

and

$$\beta_n(\theta_0) \to 0 \text{ in probability.}$$

Concerning $\alpha_n(\theta)$, using (5), (8) and (6), we have

$$\begin{aligned}\hat{l}_n(\theta_0) - \tilde{l}_n(\theta_0) &= \\ &= \frac{1}{n}\sum_{i=1}^n \Big[(\log(\hat{r}^{-i}(\theta_0' X_i, Y_i)) - \log(f_{\theta_0}(\theta_0' X_i, Y_i))) + \\ &\qquad (\log(f_{\theta_0' X}(\theta_0' X_i)) - \log(\hat{s}^{-i}(\theta_0' X_i)))\Big],\end{aligned}$$

which, in view of (20), implies

$$\begin{aligned}\hat{l}_n^{[1]}(\theta_0) - \tilde{l}_n^{[1]}(\theta_0) &= \\ &= \frac{1}{n}\sum_{i=1}^{n}\Bigg[\left(\frac{\hat{r}^{-i,[1]}(\theta_0' X_i, Y_i)}{\hat{r}^{-i}(\theta_0' X_i, Y_i)} - \frac{f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}(\theta_0' X_i, Y_i)}\right) + \\ &\qquad \left(\frac{f_{\theta_0' X}^{[1]}(\theta_0' X_i)}{f_{\theta_0' X}(\theta_0' X_i)} - \frac{\hat{s}^{-i,[1]}(\theta_0' X_i)}{\hat{s}^{-i}(\theta_0' X_i)}\right)\Bigg].\end{aligned}$$

Now using

$$\frac{1}{\hat{r}^{-i}} = \frac{1}{f_{\theta_0}} + \frac{f_{\theta_0} - \hat{r}^{-i}}{f_{\theta_0}\hat{r}^{-i}}, \quad \frac{1}{\hat{s}^{-i}} = \frac{1}{f_{\theta_0' X}} + \frac{f_{\theta_0' X} - \hat{s}^{-i}}{f_{\theta_0' X}\hat{s}^{-i}}$$

we obtain

$$\hat{l}_n^{[1]}(\theta_0) - \tilde{l}_n^{[1]}(\theta_0) = \sum_{k=1}^{8} A_{kn} + o_P(n^{-1/2}), \tag{21}$$

where

$$\begin{aligned}
A_{1n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{r}^{-i,[1]}(\theta_0' X_i, Y_i) - f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}(\theta_0' X_i, Y_i)} \\
A_{2n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}^2(\theta_0' X_i, Y_i)}(f_{\theta_0}(\theta_0' X_i, Y_i) - \hat{r}^{-i}(\theta_0' X_i, Y_i)) \\
A_{3n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{f_{\theta_0' X}^{[1]}(\theta_0' X_i) - \hat{s}^{-i,[1]}(\theta_0' X_i)}{f_{\theta_0' X}(\theta_0' X_i)} \\
A_{4n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{f_{\theta_0' X}^{[1]}(\theta_0' X_i)}{f_{\theta_0' X}^2(\theta_0' X_i)}(\hat{s}^{-i}(\theta_0' X_i) - f_{\theta_0' X}(\theta_0' X_i)) \\
A_{5n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{r}^{-i,[1]}(\theta_0' X_i, Y_i) - f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i))(f_{\theta_0}(\theta_0' X_i, Y_i) - \hat{r}^{-i}(\theta_0' X_i, Y_i))}{f_{\theta_0}^2(\theta_0' X_i, Y_i)} \\
A_{6n} &= \frac{1}{n}\sum_{i=1}^{n}\nabla_\theta \log(\hat{r}^{-i}(\theta' X_i, Y_i))|_{\theta=\theta_0}\frac{(f_{\theta_0}(\theta_0' X_i, Y_i) - \hat{r}^{-i}(\theta_0' X_i, Y_i))^2}{f_{\theta_0}^2(\theta_0' X_i, Y_i)} \\
A_{7n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(f_{\theta_0' X}^{[1]}(\theta_0' X_i) - \hat{s}^{-i,[1]}(\theta_0' X_i))(\hat{s}^{-i}(\theta_0' X_i) - f_{\theta_0' X}(\theta_0' X_i))}{f_{\theta_0' X}^2(\theta_0' X_i)} \\
A_{8n} &= \frac{1}{n}\sum_{i=1}^{n}\nabla_\theta \log(\hat{s}^{-i}(\theta' X_i))|_{\theta=\theta_0}\frac{(f_{\theta_0' X}(\theta_0' X_i) - \hat{s}^{-i}(\theta_0' X_i))^2}{f_{\theta_0' X}^2(\theta_0' X_i)}
\end{aligned}$$

To deal with each of these terms separately, let us define the following functions:

$$\tilde{r}(\theta'\mathbf{x}, y) = \frac{1}{h_1 h_2}\int K\left(\frac{\theta'(\mathbf{x}-\mathbf{u})}{h_1}\right) K\left(\frac{y-v}{h_2}\right) f(\mathbf{u}, v) d\mathbf{u} dv, \tag{22}$$

$$\tilde{s}(\theta'\mathbf{x}) = \frac{1}{h_1}\int K\left(\frac{\theta'(\mathbf{x}-\mathbf{u})}{h_1}\right) f(\mathbf{u}, v) d\mathbf{u} dv$$

For instance, to deal with A_{1n} we telescope using $\tilde{r}^{[1]}(\theta_0' X_i, Y_i)$ to obtain $A_{1n} = B_{1n} + C_{1n}$, where

$$B_{1n} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{r}^{-i,[1]}(\theta_0' X_i, Y_i) - \tilde{r}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}(\theta_0' X_i, Y_i)} \tag{23}$$

$$C_{1n} = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{r}^{[1]}(\theta_0' X_i, Y_i) - f_{\theta}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}(\theta_0' X_i, Y_i)} \tag{24}$$

and $A_{2n} = B_{2n} + C_{2n}$, where

$$B_{2n} = \frac{1}{n}\sum_{i=1}^{n} \frac{f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}^2(\theta_0' X_i, Y_i)} (\tilde{r}^{-i}(\theta_0' X_i, Y_i) - \hat{r}(\theta_0' X_i, Y_i)) \tag{25}$$

$$C_{2n} = \frac{1}{n}\sum_{i=1}^{n} \frac{f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}^2(\theta_0' X_i, Y_i)} (f_{\theta_0}(\theta_0' X_i, Y_i) - \tilde{r}(\theta_0' X_i, Y_i)).$$

In a parallel way $A_{in} = B_{in} + C_{in}$ for $i = 3, 4$. Just defining $B_n = B_{1n} + B_{2n} + B_{3n} + B_{4n}$, $C_n = C_{1n} + C_{2n} + C_{3n} + C_{4n}$ and $D_n = A_{5n} + A_{6n} + A_{7n} + A_{8n}$ and using equation (21) we have

$$\alpha_n(\theta_0) = B_n + C_n + D_n + o_P(n^{-1/2}).$$

Now using Lemmas 5, 8 and 9 implies that

$$\sqrt{n} C_n = o_P(1), \quad \sqrt{n} B_n = o_P(1), \quad \sqrt{n} D_n = o_P(1).$$

Finally, Lemma 10 can be used to prove $\hat{l}_n^{[2]}(\theta) \to l^{[2]}(\theta)$. This completes the proof. □

Next we consider the gradients of f_θ. We have the following results:

Lemma 3. *Under A6, we have*

a)

$$\frac{\partial}{\partial\theta_k} f_\theta(\theta'\mathbf{x}, y)_{\theta=\theta_0} = \frac{\partial}{\partial t}\left(f_{\theta_0}(t, y)\left[\mathbf{x}_k - E(X_k|\theta_0' X = t, Y = y)\right]\right)_{t=\theta_0'\mathbf{x}}.$$

b)

$$\frac{\partial}{\partial\theta_k} f_{\theta' X}(\theta'\mathbf{x})_{\theta=\theta_0} = \frac{\partial}{\partial t}\left(f_{\theta_0' X}(t)\left[\mathbf{x}_k - E(X_k|\theta_0' X = t)\right]\right)_{t=\theta_0'\mathbf{x}}$$

Proof. We will prove only part a). The proof of b) is very similar. Let us consider $\theta_0 = (\theta_{01}, ..., \theta_{0d})'$ and $\theta_{0,\delta}^{(k)} = \theta_0 + (0, ..., \delta, ..., 0)'$, where the δ is in position k. Then, for every $\mathbf{x} = (x_1, ..., x_d)'$ and $y \in \mathbb{R}$, we have

$$\frac{\partial}{\partial\theta_k} f_\theta(\theta'\mathbf{x}, y)_{\theta=\theta_0} = \beta_1 + \beta_2 + \beta_3, \tag{26}$$

where

$$\beta_1 = \lim_{\delta\to 0} \frac{f_{\theta_{0,\delta}^{(k)}}(\theta_0'\mathbf{x}, y) - f_{\theta_0}(\theta_0'\mathbf{x}, y)}{\delta} \tag{27}$$

$$\beta_2 = \lim_{\delta\to 0} \frac{f_{\theta_0}(\theta_{0,\delta}^{(k)'}\mathbf{x}, y) - f_{\theta_0}(\theta_0'\mathbf{x}, y)}{\delta} \tag{28}$$

$$\beta_3 = \lim_{\delta\to 0}\left[\frac{f_{\theta_{0,\delta}^{(k)'}}(\theta_{0,\delta}^{(k)'}\mathbf{x}, y) - f_{\theta_0}(\theta_{0,\delta}^{(k)'}\mathbf{x}, y)}{\delta} - \frac{f_{\theta_{0,\delta}^{(k)}}(\theta_0'\mathbf{x}, y) - f_{\theta_0}(\theta_0'\mathbf{x}, y)}{\delta}\right]. \tag{29}$$

Using two Taylor expansions, the term in (29) becomes

$$\begin{aligned}\beta_3 &= \lim_{\delta\to 0}\left[\frac{\partial}{\partial\theta_k} f_\theta(\theta_{0,\delta}^{(k)'}\mathbf{x}, y) - \frac{\partial}{\partial\theta_k} f_\theta(\theta_0'\mathbf{x}, y)\right] \\ &+ \lim_{\delta\to 0}\frac{\delta}{2}\left[\frac{\partial^2}{\partial\theta_k^2} f_\theta(\theta_{0,\delta}^{(k)'}\mathbf{x}, y)_{\theta=\tilde\theta} - \frac{\partial^2}{\partial\theta_k^2} f_\theta(\theta_0'\mathbf{x}, y)_{\theta=\tilde{\tilde\theta}}\right] = 0\end{aligned}$$

where $\tilde\theta = \theta_0 + (0, ..., \tilde\delta, ..., 0)'$, $\tilde{\tilde\theta} = \theta_0 + (0, ..., \tilde{\tilde\delta}, ..., 0)'$ and $\tilde\delta$ and $\tilde{\tilde\delta}$ are intermediate points between 0 and δ and the last step comes from the continuity of the first and second partial derivatives (see Condition A6).

A direct inspection of (28) leads to

$$\beta_2 = \frac{\partial}{\partial\theta_k} f_{\theta_0}(\theta'\mathbf{x}, y)_{\theta=\theta_0} = x_k \frac{\partial}{\partial t} f_{\theta_0}(t, y)_{t=\theta_0'\mathbf{x}}.$$

On the other hand

$$\beta_1 = \frac{\partial}{\partial\theta_k} f_\theta(\theta_0'\mathbf{x}, y)_{\theta=\theta_0} = \frac{\partial}{\partial u}\left(\frac{\partial}{\partial v}\left(\frac{\partial}{\partial\theta_k} P(\theta' X \le u, Y \le v)_{\theta=\theta_0}\right)_{v=y}\right)_{u=\theta_0'\mathbf{x}}. \tag{30}$$

Let $f^{X_k,\theta_0'X,Y}(x_k, u, \tilde{y})$ denote the density of $(X_k, \theta_0'X, Y)$. Now using standard algebra for the inner partial derivative in (30), we obtain

$$\begin{aligned}
&\frac{\partial}{\partial\theta_k} P(\theta' X \le u, Y \le v)\big|_{\theta=\theta_0} = \\
&= \lim_{\delta\to 0} \frac{P(\theta_0'X + \delta X_k \le u, Y \le v) - P(\theta_0'X \le u, Y \le v)}{\delta} \\
&= \int_{-\infty}^{v}\int_{-\infty}^{\infty} \lim_{\delta\to 0}\frac{1}{\delta}\left[\int_{-\infty}^{u-\delta z_1} f^{X_k,\theta_0'X,Y}(z_1, z_2, z_3)dz_2 \right. \\
&\quad \left. - \int_{-\infty}^{u} f^{X_k,\theta_0'X,Y}(z_1, z_2, z_3)dz_2\right] dz_1 dz_3 \\
&= -\int_{-\infty}^{v}\int_{-\infty}^{\infty} z_1 f^{X_k,\theta_0'X,Y}(z_1, u, z_3)dz_1 dz_3.
\end{aligned} \tag{31}$$

Now using (31) in (30) gives

$$\beta_1 = -\frac{\partial}{\partial u}\left(f_{\theta_0}(u, y)E(X_k|\theta_0'X = u, Y = y)\right)_{t=\theta_0'\mathbf{x}} \tag{32}$$

Using (29), (28) and (32) in (26) the proof of a) is concluded. □

Lemma 4. *Under A2, we have*

a)

$$\nabla_\theta f_\theta(y|\theta'\mathbf{x})_{\theta=\theta_0} = \left[\mathbf{x} - E(X|\theta_0'X = \theta_0'\mathbf{x})\right]\frac{\partial}{\partial t} f_{\theta_0}(y|t)_{t=\theta_0'\mathbf{x}}.$$

b)

$$E(\nabla_\theta[\log f_\theta(Y_i|\theta' X_i)]_{\theta=\theta_0}|Y_i, \theta_0'X_i) = 0$$

Proof. Part a) is an immediate consequence of Lemma 3. For part b), A2 implies

$$\begin{aligned}
&E\left([X_i - E(X|\theta_0'X = \theta_0'X_i)]|Y_i, \theta_0'X_i\right) \\
&= E(X|\theta_0'X = \theta_0'X_i) - E(X|\theta_0'X = \theta_0'X_i) = 0.
\end{aligned}$$

This completes the proof. □

LEMMA 5. *Under A2, A4 and A5,* $\sqrt{n}C_n = O_P(\sqrt{n}h_1^4 + \sqrt{n}h_2^2) = o_P(1)$.

Proof. First, using several changes of variables, we have

$$\begin{aligned}\tilde{r}(\theta_0'X_i, Y_i)1_{\{Y_i \le a_n\}} &= \frac{1}{h_1 h_2}\int\int K\left(\frac{\theta_0'X_i - u}{h_1}\right)K\left(\frac{Y_i - v}{h_2}\right)f_{\theta_0}(u,v)dudv \\ &= \int\int K(z_1)\,K(z_2)\,f_{\theta_0}(\theta_0'X_i - h_1 z_1, Y_i - h_2 z_2)dz_1 dz_2\end{aligned}$$

and

$$\begin{aligned}&\tilde{r}^{[1]}(\theta_0'X_i, Y_i)1_{\{Y_i \le a_n\}} \\ =\ & \tfrac{1}{h_1^2 h_2}\int\int (X_i - E(X|\theta_0'X = u))K'\left(\tfrac{\theta_0'X_i - u}{h_1}\right)K\left(\tfrac{Y_i - v}{h_2}\right)f_{\theta_0}(u,v)dudv \\ =\ & \tfrac{1}{h_1}\int (X_i - E(X|\theta_0'X = \theta_0'X_i - h_1 z_1))K'(z_1)K(z_2)f_{\theta_0}(\theta_0'X_i - h_1 z_1, Y_i - h_2 z_2)dz_1 dz_2.\end{aligned}$$

Using A5, the function $f_\theta(u,v)$ is three times differentiable. Now applying a Taylor expansion and using $\int z^a K(z)dz = 0$ for a odd, $\int K(z)dz = 1$ and $d_K = \int z^2 K(z)dz < \infty$, together with $\int z^b K'(z)dz = 0$ for b even, $\int zK'(z)dz = -1$ and $\int z^3 K'(z)dz = -3\int z^2 K(z)dz$ we obtain

$$\begin{aligned}\tilde{r}(\theta_0'X_i, Y_i) &= f_{\theta_0}(\theta_0'X_i, Y_i) + \frac{d_K h_1^2}{2}\frac{\partial^2}{\partial^2 u}f_{\theta_0}(u, Y_i)_{u=\theta_0'X_i} \\ &\quad + \frac{d_K h_2^2}{2}\frac{\partial^2}{\partial^2 v}\left(f_{\theta_0}(\theta_0'X_i, v)1_{\{Y_i \le a_n\}}\right)_{v=Y_i} + O_P(h_1^4 + h_1^2 h_2^2 + h_2^4). \qquad (33)\end{aligned}$$

Similarly, Lemma 3 implies that

$$\begin{aligned}\tilde{r}^{[1]}(\theta_0'X_i, Y_i) &= f_{\theta_0}^{[1]}(\theta_0'X_i, Y_i) + \frac{d_K h_1^2}{2}\frac{\partial^3}{\partial^3 u}\left[(X_i - E(X|\theta_0'X = u))f_{\theta_0}(u, Y_i)\right]_{u=\theta_0'X_i} \\ &+ \frac{d_K h_2^2}{2}\frac{\partial}{\partial u}\frac{\partial^2}{\partial^2 v}\left[(X_i - E(X|\theta_0'X = u))f_{\theta_0}(u,v)\right]_{u=\theta_0'X_i, v=Y_i} \\ &+ O_P(h_1^4 + h_1^2 h_2^2 + h_2^4). \qquad (34)\end{aligned}$$

Now starting from (24), using (33) and (34) and repeating similar steps for $\tilde{s}(\theta_0'X_i)$, we obtain

$$C_n = \tilde{C}_{1n} + \tilde{C}_{2n} + O_P(h_1^4 + h_2^2), \qquad (35)$$

where

$$\tilde{C}_{1n} = \frac{1}{n}\sum_{i=1}^{n}\frac{d_K h_1^2}{2}\tilde{C}_{1i},$$

$$\tilde{C}_{2n} = \frac{1}{n}\sum_{i=1}^{n}\frac{d_K h_1^2}{2}\tilde{C}_{2i},$$

$$\begin{aligned}\tilde{C}_{1i} &= \frac{1}{f_{\theta_0}(\theta_0' X_i, Y_i)}\frac{d^3}{du^3}\Big([X_i - E(X|\theta_0' X = u)]f_{\theta_0}(u, Y_i)\Big)_{u=\theta_0' X_i} \\ &\quad -\frac{1}{f_{\theta_0' X}(\theta_0' X_i)}\frac{d^3}{du^3}\Big([X_i - E(X|\theta_0' X = u)]f_{\theta_0' X}(u)\Big)_{u=\theta_0' X_i}\end{aligned}$$

and

$$\tilde{C}_{2i} = -\frac{\nabla_\theta f_\theta(\theta' X_i, Y_i)_{\theta=\theta_0}}{f^2_{\theta_0}(Y_i, \theta_0' X_i)}\frac{d^2}{du^2}f_{\theta_0}(u, Y_i)|_{u=\theta_0' X_i} + \frac{\nabla_\theta f_{\theta' X}(\theta' X_i)_{\theta=\theta_0}}{f^2_{\theta_0' X}(\theta_0' X_i)}f''_{\theta_0' X}(\theta_0' X_i).$$

We show that $E(\tilde{C}_{kn}) = 0$ and $Var(\tilde{C}_{kn}) = o(n^{-1})$ for $k = 1, 2$. Since $f_{\theta_0}(z|\theta_0'\mathbf{x}) = f(z|\mathbf{x})$, we have that $f(\mathbf{x}, z) = f_{\theta_0}(\theta_0'\mathbf{x}, z)f_X(\mathbf{x})/f_{\theta_0' X}(\theta_0'\mathbf{x})$. This equation and condition A2 can be used to obtain:

$$\begin{aligned}E\left(\tilde{C}_{1i}\right) &= \int \frac{1}{f_{\theta_0}(\theta_0'\mathbf{x}, z)}\frac{d^3}{du^3}\Big([\mathbf{x} - E(X|\theta_0' X = u)]f_{\theta_0}(u, y)\Big)_{u=\theta_0'\mathbf{x}} f(y|\mathbf{x})f_X(\mathbf{x})d\mathbf{x}dy \\ &\quad -\int \frac{1}{f_{\theta_0' X}(\theta_0'\mathbf{x})}\frac{d^3}{du^3}\Big([\mathbf{x} - E(X|\theta_0' X = u)]f_{\theta_0' X}(u)\Big)_{u=\theta_0'\mathbf{x}} f_X(\mathbf{x})d\mathbf{x} = 0.\end{aligned}$$

Moreover, the $\tilde{C}_{1i}$ are i.i.d. and, since $h_1^2 \to 0$, $Var(\sqrt{n}\tilde{C}_{1n}) \to 0$. Hence $\sqrt{n}\tilde{C}_{1n} \to 0$ in probability. Similar arguments can be used to conclude $\sqrt{n}\tilde{C}_{2n} \to 0$. Now using (35), we have $\sqrt{n}C_n = O_P(h_1^2 + \sqrt{n}(h_1^4 + h_2^2)) = o_P(1)$,which completes the proof. □

LEMMA 6. *Under the conditions in Theorem 3 we have* $B_{1n} = o_P(n^{-1/2})$.

Proof. Starting from (23) and using (22) and (7) we obtain

$$\begin{aligned}B_{1n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{f_{\theta_0}(\theta_0' X_i, Y_i)}\Bigg[\frac{1}{h_1^2 h_2(n-1)}\sum_{j\neq i}(X_i - X_j)K'\left(\frac{\theta_0' X_i - \theta_0' X_j}{h_1}\right)K\left(\frac{Y_i - Y_j}{h_2}\right) \\ &\quad - \frac{1}{h_1^2 h_2}\int\int(X_i - \mathbf{u})K'\left(\frac{\theta_0' X_i - \theta_0'\mathbf{u}}{h_1}\right)K\left(\frac{Y_i - v}{h_2}\right)f(\mathbf{u}, v)\Bigg]d\mathbf{u}dv \\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}H_{ij},\end{aligned}$$

where $K'(t)$ denotes the derivative of K with respect to t and

$$\begin{aligned} H_{ij} &= \frac{1}{h_1^2 h_2}\frac{1}{f_{\theta_0}(\theta_0' X_i, Y_i)}\Bigg[(X_i - X_j)K'\left(\frac{\theta_0' X_i - \theta_0' X_j}{h_1}\right)K\left(\frac{Y_i - Y_j}{h_2}\right) \\ &\quad - \int\int (X_i - \mathbf{u})K'\left(\frac{\theta_0' X_i - \theta_0'\mathbf{u}}{h_1}\right)K\left(\frac{Y_i - v}{h_2}\right) f(\mathbf{u}, v)d\mathbf{u}dv\Bigg] \end{aligned}$$

Let us define

$$\bar{H}_j = E(H_{ij}|X_j, Y_j, \delta_j).$$

Then $B_{1n} = B_{1n1} + B_{1n2}$, where

$$B_{1n1} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}(H_{ij} - \bar{H}_j) \tag{36}$$

$$B_{1n2} = \frac{1}{n}\sum_{j=1}^{n}\bar{H}_j. \tag{37}$$

For the term in (36) it is evident that

$$E(H_{ij} - \bar{H}_j) = E(H_{ij} - \bar{H}_j|X_i, Y_i) = E(H_{ij} - \bar{H}_j|X_j, Y_j) = 0,$$

so $E(B_{1n1}) = 0$. On the other hand, long but straightforward calculations can be performed to compute the variance of $B_{1n1}^{(m)}$, the m-th component of the random vector B_{1n1}, which results:

$$\begin{aligned} Var(B_{1n1}^{(m)}) &= E({B_{1n1}^{(m)}}^2) \\ &= \frac{1}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\sum_{k=1}^{n}\sum_{l=1, l\neq k}^{n} E((H_{ij}^{(m)} - \bar{H}_j^{(m)})(H_{kl}^{(m)} - \bar{H}_l^{(m)})) \\ &= \frac{1}{n(n-1)}[E[(H_{12}^{(m)} - \bar{H}_2^{(m)})^2)] + E[(H_{12}^{(m)} - \bar{H}_2^{(m)})(H_{21}^{(m)} - \bar{H}_1^{(m)})]]. \end{aligned}$$

Using standard algebra, the last two expectations can be proved to be of order $O(h_1^{-3}h_2^{-1})$. As a consequence $Var(B_{1n1}^{(m)}) = O(n^{-2}h_1^{-3}h_2^{-1})$. Now Condition A4 implies $B_{1n1}^{(m)} = o_P(n^{-1/2})$. The term in (37) can be handled in a similar way to prove $E(B_{1n2}) = 0$ and $Var(B_{1n2}^{(m)}) = O(n^{-1}h_1)$. Thus, A4 implies $B_{1n2}^{(m)} = o_P(n^{-1/2})$. This concludes the proof. □

Now we state a similar lemma.

LEMMA 7. *Under the conditions in Theorem 3 we have* $B_{2n} = o_P(n^{-1/2})$, $B_{3n} = o_P(n^{-1/2})$ *and* $B_{4n} = o_P(n^{-1/2})$.

Proof. Given the similarities between the term B_{1n} in (23) and the term B_{2n} in (25), and also the terms B_{3n} and B_{4n}, the proof is parallel to that of Lemma 6, obtaining $B_{2n} = B_{2n1} + B_{2n2}$, $B_{3n} = B_{3n1} + B_{3n2}$, $B_{4n} = B_{4n1} + B_{4n2}$, where $E[B_{jnk}] = 0$, for $j = 2, 3, 4$ and $k = 1, 2$ and $Var(B_{2n1}) = O(n^{-2}h_1^{-1}h_2^{-1})$, $Var(B_{2n2}) = O(n^{-1}h_1^2)$, $Var(B_{3n1}) = O(n^{-2}h_1^{-3}), Var(B_{2n2}) = O(n^{-1}h_1), Var(B_{4n1}) = O(n^{-2}h_1^{-1}), Var(B_{4n2}) = O(n^{-1}h_1^2)$. Now assumption A4 implies that the orders in the six variance terms are all $o(n^{-1})$, which concludes the proof. □

LEMMA 8. *Under the conditions in Theorem 3,* $B_n = o_P(n^{-1/2})$.

Proof. The proof makes use of the definition of $B_n = B_{1n} + B_{2n} + B_{3n} + B_{4n}$ and Lemmas 6 and 7. □

LEMMA 9. *Under the conditions in Theorem 3,* $D_n = o_P(n^{-1/2})$.

Proof. Each of the terms A_{5n}, A_{6n}, A_{7n}, A_{8n} can be bounded using the Cauchy-Schwarz inequality. Let us consider A_{5n}:

$$|A_{5n}| \leq A_{5n1}^{1/2} A_{5n2}^{1/2}, \tag{38}$$

where

$$\begin{aligned} A_{5n1} &= \frac{1}{n}\sum_{i=1}^{n} \frac{(\hat{r}^{-i,[1]}(\theta_0' X_i, Y_i) - f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i))^2}{f_{\theta_0}^2(\theta_0' X_i, Y_i)} \\ A_{5n2} &= \frac{1}{n}\sum_{i=1}^{n} \frac{((f_{\theta_0}(\theta_0' X_i, Y_i) - \hat{r}^{-i}(\theta_0' X_i, Y_i))^2}{f_{\theta_0}^2(\theta_0' X_i, Y_i)} \end{aligned}$$

These two terms can be expanded as

$$A_{5n1} = A_{5n11} + A_{5n12} + A_{5n13}, \tag{39}$$

$$A_{5n2} = A_{5n21} + A_{5n22} + A_{5n23}, \tag{40}$$

where

$$
\begin{aligned}
A_{5n11} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{r}^{-i,[1]}(\theta_0'X_i,Y_i)-\tilde{r}^{[1]}(\theta_0'X_i,Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i,Y_i)}\\
A_{5n12} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(\tilde{r}^{[1]}(\theta_0'X_i,Y_i)-f_{\theta_0}^{[1]}(\theta_0'X_i,Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i,Y_i)}\\
A_{5n13} &= \frac{2}{n}\sum_{i=1}^{n}\frac{(\hat{r}^{-i,[1]}(\theta_0'X_i,Y_i)-\tilde{r}^{[1]}(\theta_0'X_i,Y_i))(\tilde{r}^{[1]}(\theta_0'X_i,Y_i)-f_{\theta_0}^{[1]}(\theta_0'X_i,Y_i))}{f_{\theta_0}^2(\theta_0'X_i,Y_i)}\\
A_{5n21} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(f_{\theta_0}(\theta_0'X_i,Y_i)-\tilde{r}(\theta_0'X_i,Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i,Y_i)}\\
A_{5n22} &= \frac{1}{n}\sum_{i=1}^{n}\frac{(\tilde{r}(\theta_0'X_i,Y_i)-\hat{r}^{-i}(\theta_0'X_i,Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i,Y_i)}\\
A_{5n23} &= \frac{2}{n}\sum_{i=1}^{n}\frac{(f_{\theta_0}(\theta_0'X_i,Y_i)-\tilde{r}(\theta_0'X_i,Y_i))(\tilde{r}(\theta_0'X_i,Y_i)-\hat{r}^{-i}(\theta_0'X_i,Y_i))}{f_{\theta_0}^2(\theta_0'X_i,Y_i)}.
\end{aligned}
$$

The terms A_{5n11}, A_{5n12}, A_{5n21} and A_{5n22} can be treated similarly to B_{1n}, C_{1n}, B_{2n}, C_{2n}, respectively. On the other hand, A_{5n13} (respectively A_{5n23}) can be bounded, using the Cauchy-Schwarz inequality, in terms of A_{5n11} and A_{5n12} (respectively A_{5n21} and A_{5n22}). All in all, the results in $A_{5njk} = o_P(n^{-1/2})$ for $j = 1, 2$ and $k = 1, 2$. In view of (39) (40) we have $A_{5n1} = o_P(n^{-1/2})$ and $A_{5n2} = o_P(n^{-1/2})$, which, using (38), gives $A_{5n} = o_P(n^{-1/2}$.

Following parallel arguments, straightforward but tedious algebra can be used to prove that $A_{6n} = o_P(n^{-1/2})$, $A_{7n} = o_P(n^{-1/2})$and $A_{8n} = o_P(n^{-1/2})$. Since $D_n = A_{5n} + A_{6n} + A_{7n} + A_{8n}$, this concludes the proof. □

LEMMA 10. *Under the conditions in Theorem 3,* $\hat{l}_n^{[2]}(\theta) \to l^{[2]}(\theta)$.

Proof. We have

$$
\begin{aligned}
\hat{l}_n^{[2]}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\Bigg(\frac{\hat{r}^{-i,[2]}(\theta'X_i,Y_i)}{\hat{r}^{-i}(\theta'X_i,Y_i)}-\frac{\hat{r}^{-i,[1]}(\theta'X_i,Y_i)(\hat{r}^{-i,[1]}(\theta'X_i,Y_i))'}{\hat{r}^{-i}(\theta'X_i,Y_i)^2}-\frac{\hat{s}^{-i,[2]}(\theta'X_i)}{\hat{s}^{-i}(\theta'X_i)}\\
&+ \frac{\hat{s}^{-i,[1]}(\theta'X_i)(\hat{s}^{-i,[1]}(\theta'X_i))'}{\hat{s}^{-i}(\theta'X_i)^2}\Bigg).
\end{aligned}
$$

Since

$$\hat{r}^{-i,[2]}(y,\theta' x) = \frac{1}{nh_1^3 h_2}\sum_{j=1}^{n}(x-X_j)(x-X_j)'K''\left(\frac{\theta' x-\theta' X_j}{h_1}\right)K\left(\frac{y-Y_j}{h_2}\right).$$

it is easy to show, that

$$\hat{r}^{-i,[2]}(\theta' x, y) \to \frac{\partial^2}{\partial u^2}\left\{f_\theta(u,y)E((\mathbf{x}-X)(\mathbf{x}-X)'|\theta' X=u)\right\}_{u=\theta'\mathbf{x}} \quad \text{in probability.}$$

Moreover, using Lemma 3, it can be proved that

$$\frac{\partial^2}{\partial u^2}\left\{f_\theta(u,u)E((\mathbf{x}-X)(\mathbf{x}-X)'|\theta' X=u)\right\}_{u=\theta' x} = f_\theta^{[2]}(\theta'\mathbf{x},y).$$

Similarly, it can be shown that $\hat{s}^{[2]}(\theta'\mathbf{x}) \to f_{\theta' X}^{[2]}(\theta'\mathbf{x})$ in probability. This completes the proof. □

Proof of Theorem 2. In view of Lemmas 1, 2 and 10, it remains to be shown that $\sum_{k=1}^{8} A_{kn} \xrightarrow{P} 0$. But this can be proved in line with Lemmas 5, 8 and 9 but even more simply, since only convergence in probability to zero (and no rate) is required. □