Maximum likelihood estimation for conditional distribution single-index models under censoring

Ewa Strzalkowska-Kominiak
Department of Mathematics, Facultad de Informática, Universidade da Coruña, A Coruña, Spain

Ricardo Cao
Department of Mathematics, Facultad de Informática, Universidade da Coruña, A Coruña, Spain

Abstract

A new likelihood approach is proposed for the problem of semiparametric estimation of a conditional distribution or density under censoring. Consistency and asymptotic normality for two versions of the maximum likelihood estimator of the parameter vector in the single index model are proved. The single-index model considered can be seen as a useful tool for credit scoring and estimation of the default probability in credit risk. A data-driven bandwidth selection procedure is proposed. It allows to choose the smoothing parameter involved in our approach. The finite sample performance of the estimators has been studied by simulations, where the new method has been compared with the method by Bouaziz and Lopez (2010) [1]. To the best of our knowledge this is the only existing competitor in this context. The simulation study shows the good behaviour of the proposed method.

Keywords: conditional density function, credit risk, kernel estimation, survival analysis

1. Motivation and background

The single-index model (SIM) is a flexible tool to incorporate the effect of a vector of covariates in a regression problem. By focusing on an index, the so-called “curse of dimensionality” is no longer a problem in the SIM, since multivariate nonparametric regression estimation is avoided. In other words, the SIM is as a reasonable compromise between a fully parametric and a fully nonparametric model.

Plenty of papers have dealt with the issue of estimating the parameter vector and the link function in SIM (see [6], [9], [12], [14], [15] and [23], among many other). Different methods are available for fitting the SIM: for instance, kernel smoothing, least-squares, average derivative estimation and sliced inverse regression. Parametric tests (see, for instance, [30]) and goodness-of-fit tests (see [8], [13], [27] and [32], among other) have been proposed for SIM. Generalizations of SIM has been also considered over the past few years. For instance [3], [28] and [29] studied estimation methods in partial linear single-index models. Empirical likelihood methods have been also applied for SIM (see, e.g., [33]). Finally SIM have been extended to survival analysis (as in [18]), including censored data (see [17], [19] and [20]).

In the contexts of conditional distribution function and conditional density the use of SIM has been much more limited. However, this is an important setting in applications to credit risk, where SIM’s can be used to jointly estimate the scoring and the probability of default of credits, since this probability can be expressed in terms of a conditional distribution function (see [2] for details). SIM’s have been also used for other purposes in credit risk (see, for instance, [21]).

To fix our notation, let the lifetime, $Z$, be a nonnegative random variable dependent on a vector of covariates $X = (X_1, ..., X_d)'$ and $f(z|x)$ the density function of $Z$ given $X = x$. Moreover, let

$$\theta_0 = (\theta_1, ..., \theta_d)'$$

where $d \geq 2$.
be a parameter vector with the property:
\[ f_{\theta_0}(z|x) = f(z|x), \]
where \( f_{\theta_0}(z|x) \) is the conditional density of \( Z \) given \( \theta_0X = \theta_0x \). Furthermore, let \( F(z|X = x) \) and \( F_{\theta_0}(z|\theta_0X = \theta_0x) \) be the conditional distribution functions, given \( X \) and \( \theta_0X \), respectively. As a consequence
\[ F_{\theta_0}(z|\theta_0X = \theta_0x) = F(z|X = x). \] (1)
Moreover, the random variable \( Z \) may be censored from the right by \( C \sim G \). Hence, we observe
\[ Y = \min(Z, C) \sim H \]
together with
\[ \delta = 1_{\{Z \leq C\}}. \]
For example, in credit risk applications (as in Cao, Vilar-Fernández and Devía [2]) the lifetime, \( Z \), could be the time to default of a credit and the vector \( X = (X_1, ..., X_d)' \) would include the client covariates that are relevant for estimating the probability of default. The linear combination \( \theta_0X \) is an index of the propensity to default of a credit with covariate vector \( x \). So, the interest is to estimate the parameter \( \theta_0 \), the conditional density, \( f_{\theta_0} \), and the distribution function, \( F_{\theta_0} \). In this context, Bouaziz and Lopez [1] have recently proposed a semiparametric procedure to estimate the conditional density, deriving consistency and asymptotic normality of some estimator of the index parameter in (1). This estimator is based on maximizing a regression-like analogue of the (uncomputable) theoretical likelihood.
In this paper a new maximum likelihood estimator is proposed. The idea is to compute some alternative observable version of the likelihood that takes into account explicitly the censoring mechanism. The rest of the paper is organized as follows. Section 2 presents the theoretical likelihood in this conditional context under censoring and some result that characterizes the true index vector as the maximizer of the expected value of this theoretical likelihood. The new likelihood version to be used in practice is presented in Section 3, where the maximum likelihood estimator is defined. Section 4 includes the main results for the maximum likelihood estimator for the index vector, namely its consistency and asymptotic normality. Section 5 is devoted to a simulation study that exhibits the good behaviour of the proposed approach, while Section 6 shows a real data example. Finally, Section 7 presents some conclusions and the proofs of the main results are collected in Appendix A.

### 2. Theoretical likelihood function

Suppose for a while that \( F_\theta \) is known except for the value of the index vector \( \theta \). We define the theoretical likelihood function as follows:
\[ \tilde{L}_n(\theta) = \prod_{i=1}^{n} \left( f_\theta(Y_i|\theta'X_i) \right)^{\delta_i} \left( 1 - F_\theta(Y_i|\theta'X_i) \right)^{1-\delta_i}, \]
and
\[ \tilde{l}_n(\theta) = \frac{1}{n} \log \left( \tilde{L}_n(\theta) \right) = \frac{1}{n} \sum_{i=1}^{n} \left( \delta_i \log f_\theta(Y_i|\theta'X_i) + (1 - \delta_i) \log(1 - F_\theta(Y_i|\theta'X_i)) \right). \] (2)
Set
\[ \tilde{\theta}_n = \arg \max_{\theta} \tilde{l}_n(\theta) \] (3)
and define the score function as the expected likelihood:
\[ l(\theta) = E(\tilde{l}_n(\theta)). \] (4)
Some assumptions are needed for the following result:
A1: \( Z \) is independent of \( C \)
A2: \( Z \) is conditionally independent of \( C \) given \( X \)
A3: \( \mathbb{P}(Z \leq C|X, Z) = \mathbb{P}(Z \leq C|Z) \)

Observe that assumption A3 allows some dependence between \( C \) and \( X \). Moreover, under (1) and A2, we have that \( Z \) is conditionally independent of \( C \) given \( \theta_0 X \).

The following result characterizes \( \theta_0 \) as the maximizer of the score function.

**Theorem 1.** Under A1-A3 we have

\[
\theta_0 = \arg \max_{\theta} l(\theta)
\]

**(Proof):**

We will show that for every \( \theta \) we have

\[
E(\hat{l}_n(\theta_0)) \geq E(\hat{l}_n(\theta)).
\]

According to (2) and since the data are i.i.d., we have

\[
E(\hat{l}_n(\theta_0)) - E(\hat{l}_n(\theta)) = E\left( \log \left( \frac{f_0(Y_1|\theta_0 X_1)}{f_0(Y_1|\theta' X_1)} \right)^{\delta_1} \right) + E\left( \log \left( \frac{1 - F_0(Y_1|\theta_0 X_1)}{1 - F_0(Y_1|\theta' X_1)} \right)^{1-\delta_1} \right)
\]

Moreover, using Jensen inequality

\[
E(\hat{l}_n(\theta_0)) - E(\hat{l}_n(\theta)) \geq - \log(A) - \log(B),
\]

where

\[
A = E\left( \left( \frac{f_0(Y_1|\theta_0 X_1)}{f_0(Y_1|\theta' X_1)} \right)^{\delta_1} \right)
\]

and

\[
B = E\left( \left( \frac{1 - F_0(Y_1|\theta_0 X_1)}{1 - F_0(Y_1|\theta' X_1)} \right)^{1-\delta_1} \right).
\]

Let \( f_1(x, c, z), f_2(c, z|x) \) and \( g(c|x) \) be the densities of \( (X, C, Z), (C, Z) \) given \( X \) and of \( C \) given \( X \), respectively.

According to A2, we have that \( f_1(x, c, z) = f_2(c, z|x)f_X(x) = g(c|x)f(z|x)f_X(x) \). Hence, by definition of \( \theta_0 \), we obtain

\[
A = \int f_0(\min(z,c)|x)g(c|x)f_X(x)dx dz dc
\]

\[
= \int f_0(\min(z,c)|\theta' x)g(c|x)f_X(x)dx dz dc
\]

\[
= \int f_0(\min(z,c)|x)g(c|x)f_X(x)dx dz dc + \int f_0(\min(z,c)|\theta' x)g(c|x)f_X(x)dx dz dc
\]

\[
= a + b,
\]

where \( a = \int g(c|x)f_X(x)F_0(c|\theta' x)dx dc \) and \( b = \int g(c|x)f_X(x)(1 - F(c|x))dx dc \)
Similarly,
\[
B = \int (1 - F_0(\min(z, c)|\theta' x))^{1_{(z \geq c)}} g(c|x) f(z|x) f_X(x) dx dz dc
\]
\[
= \int (1 - F_0(c|\theta' x)) g(c|x) f_X(x) dx dc + \int F(c|x) g(c|x) f_X(x) dx dc = 2 - a - b.
\]
Hence
\[
E(\hat{I}_n(\theta_0)) - E(\hat{I}_n(\theta)) \geq - \log((a + b)(2 - a - b)),
\]
where, \(a, b \in [0, 1].\) Finally, since the function \(x(2 - x)\) has a global maximum at \(x = 1\), we have that
\[
E(\hat{I}_n(\theta_0)) - E(\hat{I}_n(\theta)) \geq - \log((a + b)(2 - a - b)) \geq - \log(1) = 0.
\]
This completes the proof. 

3. Maximum likelihood estimation

Since, in practice, neither \(f_\theta\) nor \(F_0\) are known, they need to be estimated. Let \(K\) be a nonnegative kernel and \(h_1, h_2\) two positive bandwidths. Set
\[
F(t) = \mathbb{P}(Z \leq t), G(t) = \mathbb{P}(C \leq t) \quad \text{and} \quad H(t) = \mathbb{P}(Y \leq t)
\]

Together with
\[
\tau_H = \inf\{y : H(y) = 1\}.
\]
Furthermore, set
\[
\hat{f}_0(y|\theta' x) = \frac{1}{n_1 h_2} \int K \left( \frac{\theta' x - u}{h_2} \right) K \left( \frac{u - y}{h_2} \right) 1_{\{y \leq a_n\}} F_0^{\theta} (du, dv)
\]
and
\[
1 - \hat{F}_0(y|\theta' x) = \frac{1}{n_1} \int K \left( \frac{\theta' x - u}{h_1} \right) K \left( \frac{u - v}{h_2} \right) 1_{\{v \leq a_n\}} F_0^{\theta} (dv, du),
\]
where \(K(x) = \int_{-\infty}^{\infty} K(z) dz\) and \(a_n\) is a positive sequence, used to avoid problems with the right tail of the lifetime distribution, such that \(a_n \to \tau_H\) when \(n \to \infty.\) Moreover, \(F_0^{\theta}(u, v)\) is the Kaplan-Meier estimator of \((\theta' X, Z)\) defined as
\[
\int \varphi(u, v) F_0^{\theta}(du, dv) = \sum_{j=1}^{n} W_{jn} \varphi(\theta' X_j, Y_j),
\]
where
\[
W_{jn} = F_n(Y_j) - F_n(Y_j -)
\]
and \(F_n(t)\) is Kaplan-Meier estimator of \(F(t) = \mathbb{P}(Z \leq t)\). See, [16], [25] and [26] for details.

Moreover, it can be shown that
\[
W_{jn} = \frac{\delta_j}{n(1 - G_n(Y_j -))},
\]
where \(G_n(x)\) is the Kaplan-Meier estimator of \(G(t) = \mathbb{P}(C \leq t)\). See, e.g., [1] for details.
Finally, let $\hat{f}_{\theta}^{-i}(Y_i|\theta'X_i)$ and $1 - \hat{F}_{\theta}^{-i}(Y_i|\theta'X_i)$ be the estimators defined in (6) and (7), where the sums in definition of $\hat{F}_n^{\theta}(u,v)$ runs over $j \neq i$. Set

$$\hat{l}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left( \delta_i \log \hat{f}_{\theta}^{-i}(Y_i|\theta'X_i) + (1 - \delta_i) \log(1 - \hat{F}_{\theta}^{-i}(Y_i|\theta'X_i)) \right) 1_{\{Y_i \leq a_n, X_i \in A^c\}},$$

(8)

where $A^c$ is a set with the property $P(X_i \in A^c) \to 1$ for every $i = 1, ..., n$ when $n \to \infty$ and $c_n \to 0$, and

$$\hat{\theta}_n = \arg \max_{\theta} \hat{l}_n(\theta).$$

(9)

Finally, set

$$\hat{f}_{\theta}^{-i}(Y_i|\theta'X_i) = \frac{\hat{r}(\theta'X_i, Y_i)}{\hat{s}(\theta'X_i)}$$

(10)

and

$$1 - \hat{F}_{\theta}^{-i}(Y_i|\theta'X_i) = \frac{\hat{d}(\theta'X_i, Y_i)}{\hat{s}(\theta'X_i)}.$$  

(11)

where

$$\hat{r}(\theta'X_i, Y_i) = \frac{1}{h_1h_2} \sum_{j \neq i} \left( \frac{\delta_j}{n(1 - G_n(Y_j^-))} \right) K \left( \frac{\theta'X_i - \theta'X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right)$$

(12)

$$\hat{s}(\theta'X_i) = \frac{1}{h_1} \sum_{j \neq i} 1_{\{Y_j \leq a_n\}} \left( \frac{\delta_j}{n(1 - G_n(Y_j^-))} \right) K \left( \frac{\theta'X_i - \theta'X_j}{h_1} \right)$$

(13)

and

$$\hat{d}(\theta'X_i, Y_i) = \frac{1}{h_1} \sum_{j \neq i} 1_{\{Y_j \leq a_n\}} \left( \frac{\delta_j}{n(1 - G_n(Y_j^-))} \right) K \left( \frac{\theta'X_i - \theta'X_j}{h_1} \right) |Y_i - Y_j|.$$  

(14)

4. Main results

In this section we will study the properties of $\hat{\theta}_n$. To establish the main results we need to assume some further conditions:

**A4:** $E(X|\theta'_0X, C, Z) = E(X|\theta'_0)$

**A5:** $E(XX^T) < \infty$ componentwise.

The two bandwidths $h_1$, $h_2$ and the sequence $c_n$ should fulfill the following conditions

**A6:** $\sqrt{n}h_1^4 \to 0$, $\sqrt{n}h_2^4 \to 0$, $nh_1^6 \to \infty$ and $c_n n^{5/6}h_1^4h_2 \to \infty$ when $h_1, h_2 \to 0$ and $n \to \infty$.

**Remark 1.** Using $h_1 = c_1n^{-1/7}$, $h_2 = c_2n^{-1/3}$ and $c_n = 1/log(n)$, assumption A6 is fulfilled. These are the asymptotically optimal rates for the estimation of the first derivative of the conditional density and for the distribution function estimation, respectively.
Moreover, assume
A7: The derivatives \( \frac{\partial^k}{\partial u^k} f_\theta(u, v) \) and \( \frac{\partial^k}{\partial u^k} E(X|\theta) \) exists for \( k = 1, 2, 3 \) and \( l = 1, 2 \).

A8: The function \( h(\theta_1) = \frac{\partial}{\partial \theta} f_\theta(\theta'_1, x, y)_{\theta=\theta_0} \) is continuous and \( \frac{\partial^2}{\partial \theta^2} f_\theta(\theta'_1, x, y)_{\theta=\theta_0} \) exists.

A9: \( \int \frac{F(\varnothing)}{1 - C(t)} < \infty \) and \( \sup_x \int \frac{f_\theta(x, z)_{\theta=\theta_0}}{1 - C(t)} dt < \infty \).

Finally, let \( l(\theta) = \nabla \varnothing l(\theta)_{\theta=\theta_0} \) denote the gradient of \( l(\theta) \) over \( \theta \) evaluated in \( \theta_0 \). Further, let \( l^2(\theta) \) denote the Hessian matrix of \( l(\theta) \). Now we can state our first result.

**Lemma 1.** Under A1-A3, A7 and A8 we have
\[
\hat{\theta}_n - \theta_0 = - \left[ \tilde{l}_n^2(\hat{\theta}_n) \right]^{-1} (\tilde{l}_n^1(\hat{\theta}_n) - \tilde{l}_n^1(\theta_0)),
\]
where \( \hat{\theta}_n \) is between \( \hat{\theta}_n \) and \( \theta_0 \).

**Proof.** Using (5) and (9) we have
\[
l(\theta_0) = 0 \text{ and } \tilde{l}_n^1(\hat{\theta}_n) = 0.
\]
Now a Taylor expansion gives
\[
l(\theta_0) = 0 = \tilde{l}_n^1(\hat{\theta}_n) = \tilde{l}_n^1(\theta_0) + \tilde{l}_n^2(\hat{\theta}_n)(\hat{\theta}_n - \theta_0),
\]
where \( \hat{\theta}_n \) is between \( \hat{\theta}_n \) and \( \theta_0 \). This completes the proof.

**Theorem 2.** Let \( a_n = H^{-1}(1 - \frac{\Delta_n}{n}) \) with \( s_n \approx n^{2/3}(\log(n))^{2+\epsilon_1} \) for some \( \epsilon_1 > 0 \). Set \( A_{cn} = A_{cn}^r = \{ x : f_{\theta_n}(\theta'_n, x) > c_n \} \). Then, under A1-A9 and if \( l^2(\theta^*) \) is positive definite for \( \theta^* \) belonging to a neighborhood of \( \theta_0 \), we have
\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, \Sigma),
\]
where
\[
\Sigma = \Sigma_1 \Sigma_2, \quad \Sigma_1 = \left[ l^2(\theta_0) \right]^{-1}
\]
and
\[
\Sigma_2 = \int (1 - G(z - |x|))(\nabla \varnothing \log(f_\theta(z|\theta')_{\theta=\theta_0}))(\nabla \varnothing \log(f_\theta(z|\theta')_{\theta=\theta_0})) f(x, z) dx dz
\]
\[
+ \int (1 - F(c|x))(\nabla \varnothing \log(1 - F_\theta(c|\theta')_{\theta=\theta_0}))(\nabla \varnothing \log(1 - F_\theta(c|\theta')_{\theta=\theta_0})) g(x, c) dx dc.
\]

**Theorem 3.** Let \( a_n = Y_{r_n,n} \) with \( r_n = n - s_n \) and \( s_n \approx n^{2/3}(\log(n))^{2+\epsilon_1} \). Set \( A_{cn} = A_{cn}^r = \{ x : f_{\theta_n}(\theta'_n, x) > c_n \} \),
where \( \theta_n \) denotes a consistent estimator of \( \theta_0 \). Then, under A1-A9, if \( \text{supp}(K) \in [-a, a] \) and \( l^2(\theta^*) \) is positive definite for \( \theta^* \) belonging to a neighborhood of \( \theta_0 \), we have
\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, \Sigma),
\]
where \( \Sigma \) is defined as in Theorem 2.

**Theorem 4.** Under A1-A9, with \( a_n = r_H \), we have
\[
\hat{\theta}_n \to \theta_0 \text{ in probability}
\]
Remark 2. In Theorem 2 we assume that the quantile function $H^{-1}(u)$ and the set $A_0^n$ are known. This makes the trimming function $J(x, y) = 1_{\{y \leq a_n, x \in A_0^n\}}$ deterministic and the proofs more transparent. Nevertheless, Theorem 3 shows that we can replace the quantiles $H^{-1}(u)$ by order statistics and the set $A_0^n$ with an estimable $A_0^n$ based on preliminary estimator $\theta_n$. The $\theta_n$ can be found by setting $1_{\{X \in A_0^n\}} \equiv 1$ and maximizing the likelihood (8). See [6] and [17], for details. In praxis, since $J(X_i, Y_i)$ goes to one in probability for every $i = 1, \ldots, n$, we may take $J(X_i, Y_i) \equiv 1$. Such a choice of trimming function gives good results in the simulation study.

5. Simulations

Let us consider the model used by Bouaziz and Lopez [1]:

$$Z = \theta_0'X + \varepsilon,$$

where $\theta_0 = (1, 0.5, 1.4, 0.2)$, $X = (X_1, X_2, X_3, X_4)$,

$$X_i \sim \begin{cases} N(0, 1), & \text{with probability 0.2} \\ N(0.25, 2), & \text{with probability 0.8} \end{cases}$$

is a normal mixture for $i = 1, 2, 3, 4$ and $\varepsilon \sim N(0, |\theta_0'X|)$. Set $C \sim \exp(\lambda)$, so that we only observe $Y = \min(Z, C)$ together with $\delta = 1_{\{Z \leq C\}}$.

The goal is to estimate $\theta_0$ with $\hat{\theta}_n = (\hat{\theta}_n1, \hat{\theta}_n2, \hat{\theta}_n3) \equiv (1, \theta_n)$ and select the bandwidths $h_1$ and $h_2$. For the latter we consider two possible strategies:

1. Optimizing the likelihood function over two different bandwidths $h_1$ and $h_2$.

2. Optimizing the likelihood function over $h$ by setting $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$, where $\hat{\sigma}$ is the estimated standard deviation.

In the following sections we set $1_{\{Y_i \leq a_n, X_i \in A_0^n\}} \equiv 1$.

5.1. Model 1

We consider now the model from Bouaziz and Lopez [1], where $\lambda$ is constant (it does not depend on $X$). Hence $C$ is independent of $Z$ and the assumption A1 is fulfilled.

Tables 1-4 show Monte Carlo approximations for the bias, the variance and the mean squared error of the Bouaziz and Lopez (BL) estimator and the new estimator with the two bandwidth choices presented above. These results are for $n = 100$ and $n = 200$ together with $\lambda = 0.3$ (25% censoring) and $\lambda = 0.85$ (40% censoring). For the new estimator, the results are based on 500 trials.
Table 1: Estimated bias, variance and MSE for $\lambda = 0.3$, $n = 100$ and 500 trials. Different bandwidths $h_1$ and $h_2$

<table>
<thead>
<tr>
<th>$h_{n1}$</th>
<th>$h_{n2}$</th>
<th>$h_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.003</td>
<td>0.026</td>
</tr>
<tr>
<td>Variance</td>
<td>0.031</td>
<td>0.080</td>
</tr>
<tr>
<td>MSE</td>
<td>0.1350874</td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_{1} = h\sigma(\theta'X)$ and $h_{2} = h\sigma(Y)$

<table>
<thead>
<tr>
<th>$h_{n1}$</th>
<th>$h_{n2}$</th>
<th>$h_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.009</td>
<td>0.041</td>
</tr>
<tr>
<td>Variance</td>
<td>0.023</td>
<td>0.058</td>
</tr>
<tr>
<td>MSE</td>
<td>0.1014918</td>
<td></td>
</tr>
</tbody>
</table>

Bouaziz and Lopez [1]

<table>
<thead>
<tr>
<th>$h_{n1}$</th>
<th>$h_{n2}$</th>
<th>$h_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.07</td>
<td>0.221</td>
</tr>
<tr>
<td>Variance</td>
<td>0.034</td>
<td>0.074</td>
</tr>
<tr>
<td>MSE</td>
<td>0.182598</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Estimated bias, variance and MSE for $\lambda = 0.3$, $n = 200$ and 500 trials. Different bandwidths $h_1$ and $h_2$

<table>
<thead>
<tr>
<th>$h_{n1}$</th>
<th>$h_{n2}$</th>
<th>$h_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.007</td>
<td>-0.015</td>
</tr>
<tr>
<td>Variance</td>
<td>0.010</td>
<td>0.026</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0456373</td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_{1} = h\sigma(\theta'X)$ and $h_{2} = h\sigma(Y)$

<table>
<thead>
<tr>
<th>$h_{n1}$</th>
<th>$h_{n2}$</th>
<th>$h_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.003</td>
<td>-0.009</td>
</tr>
<tr>
<td>Variance</td>
<td>0.008</td>
<td>0.019</td>
</tr>
<tr>
<td>MSE</td>
<td>0.03511254</td>
<td></td>
</tr>
</tbody>
</table>

Bouaziz and Lopez [1]

<table>
<thead>
<tr>
<th>$h_{n1}$</th>
<th>$h_{n2}$</th>
<th>$h_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.034</td>
<td>0.107</td>
</tr>
<tr>
<td>Variance</td>
<td>0.007</td>
<td>0.011</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0364064</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Estimated bias, variance and MSE for $\lambda = 0.85$ $n = 100$ and 500 trials.

<table>
<thead>
<tr>
<th>Different bandwidths $h_1$ and $h_2$</th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.008</td>
<td>0.008</td>
<td>-0.008</td>
</tr>
<tr>
<td>Variance</td>
<td>0.054</td>
<td>0.136</td>
<td>0.039</td>
</tr>
<tr>
<td>MSE</td>
<td>0.2298826</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

| Bias                                | -0.003        | 0.029         | 0.0005        |
| Variance                            | 0.037         | 0.091         | 0.025         |
| MSE                                 | 0.1542782     |               |               |

Bouaziz and Lopez [1]

| Bias                                | 0.074         | 0.176         | 0.061         |
| Variance                            | 0.064         | 0.051         | 0.069         |
| MSE                                 | 0.2239023     |               |               |

Table 4: Estimated bias, variance and MSE for $\lambda = 0.85$ $n = 200$ and 500 trials.

<table>
<thead>
<tr>
<th>Different bandwidths $h_1$ and $h_2$</th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.005</td>
<td>-0.0046</td>
<td>-0.0047</td>
</tr>
<tr>
<td>Variance</td>
<td>0.021</td>
<td>0.043</td>
<td>0.0176</td>
</tr>
<tr>
<td>MSE</td>
<td>0.08184733</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

| Bias                                | 0.012         | 0.013         | -0.0009       |
| Variance                            | 0.016         | 0.033         | 0.015         |
| MSE                                 | 0.06476457    |               |               |

Bouaziz and Lopez [1]

| Bias                                | 0.043         | 0.14          | 0.021         |
| Variance                            | 0.018         | 0.022         | 0.014         |
| MSE                                 | 0.07533921    |               |               |

It is important to mention that the proposed estimator is mostly better than the one by Bouaziz and Lopez [1]. Furthermore, the new estimator with bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$ gives always the smallest MSE. The different performance between the BL estimator and the one proposed in here comes from the bias. It is worth mentioning that this difference increases when the sample size decreases.

5.2. Model 2

In this section we consider the same model as before, but with $\lambda = \lambda(X)$. Hence $C$ depends on $Z$ and assumption A1 is violated. More precisely, we take

$$\lambda(X) = \lambda_1|\theta_0'X|,$$

with $\lambda_1 = 0.15$ and $\lambda_1 = 0.65$, which gives, as above, 25% and 40% of censoring.

Tables 5-8 collect the results for the Monte Carlo approximations for the bias, the variance and the mean squared error of the new estimator. The sample size was set to $n = 100$ and $n = 200$, with 500 trials.
Table 5: Estimated bias, variance and MSE for $\lambda_1 = 0.15$ $n = 100$ and 500 trials.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.0088</td>
<td>0.0083</td>
<td>0.00002</td>
</tr>
<tr>
<td>Variance</td>
<td>0.041</td>
<td>0.087</td>
<td>0.030</td>
</tr>
<tr>
<td>MSE</td>
<td>0.1586619</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.016</td>
<td>0.029</td>
<td>0.00008</td>
</tr>
<tr>
<td>Variance</td>
<td>0.029</td>
<td>0.071</td>
<td>0.022</td>
</tr>
<tr>
<td>MSE</td>
<td>0.1247307</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Estimated bias, variance and MSE for $\lambda_1 = 0.15$ $n = 200$ and 500 trials.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.0048</td>
<td>-0.0044</td>
<td>-0.0018</td>
</tr>
<tr>
<td>Variance</td>
<td>0.012</td>
<td>0.034</td>
<td>0.012</td>
</tr>
<tr>
<td>MSE</td>
<td>0.05776507</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.0005</td>
<td>0.0081</td>
<td>0.0045</td>
</tr>
<tr>
<td>Variance</td>
<td>0.010</td>
<td>0.028</td>
<td>0.009</td>
</tr>
<tr>
<td>MSE</td>
<td>0.04728104</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Estimated bias, variance and MSE for $\lambda_1 = 0.65$ $n = 100$ and 500 trials.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.012</td>
<td>-0.031</td>
<td>-0.0048</td>
</tr>
<tr>
<td>Variance</td>
<td>0.168</td>
<td>0.693</td>
<td>0.081</td>
</tr>
<tr>
<td>MSE</td>
<td>0.942775</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.026</td>
<td>0.073</td>
<td>0.0008</td>
</tr>
<tr>
<td>Variance</td>
<td>0.079</td>
<td>0.219</td>
<td>0.051</td>
</tr>
<tr>
<td>MSE</td>
<td>0.3566779</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8: Estimated bias, variance and MSE for $\lambda = 0.65$ $n = 200$ and 500 trials.

<table>
<thead>
<tr>
<th>Different bandwidths $h_1$ and $h_2$</th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.004</td>
<td>-0.019</td>
<td>-0.0046</td>
</tr>
<tr>
<td>Variance</td>
<td>0.088</td>
<td>0.201</td>
<td>0.038</td>
</tr>
<tr>
<td>MSE</td>
<td>0.3281369</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

| Bias                                | 0.034       | 0.0436      | 0.007       |
| Variance                            | 0.049       | 0.111       | 0.036       |
| MSE                                 | 0.1997059   |             |             |

With 25% of censoring (Tables 5-6) the estimator is still behaving well, even if the assumption A1 is not fulfilled. A censoring of 40% causes a significant increase in MSE. Nevertheless, the estimator with $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$ still presents a reasonable behavior. It is worth mentioning that, although the estimator with a single smoothing parameter has a smaller variance, its bias tends to be larger.

5.3. Model 3

In this section we consider again model (15) with $\lambda = \lambda(X)$. As in Model 2, $Z \sim \mathcal{N}(\theta'_0X, |\theta'_0X|)$ and $C \sim \exp(\lambda_1|\theta'_0X|)$. This time, given $X$, $C$ and $Z$ are generated from the Gaussian copula with parameter 0.5. Hence $C$ depends on $Z$ and neither assumption A1 nor A2 is fulfilled. We take $\lambda_1 = 0.15$ and $\lambda_1 = 0.65$, which gives, as before, 25% and 40% of censoring.

Tables 9-12 collect the results for the Monte Carlo approximations for the bias, the variance and the mean squared error of the new estimator. The sample size was set to $n = 100$ and $n = 200$, with 500 trials.

Table 9: Estimated bias, variance and MSE for $\lambda = 0.15$ $n = 100$ and 500 trials.

<table>
<thead>
<tr>
<th>Different bandwidths $h_1$ and $h_2$</th>
<th>$\theta_{n1}$</th>
<th>$\theta_{n2}$</th>
<th>$\theta_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.005</td>
<td>0.022</td>
<td>0.003</td>
</tr>
<tr>
<td>Variance</td>
<td>0.036</td>
<td>0.088</td>
<td>0.029</td>
</tr>
<tr>
<td>MSE</td>
<td>0.1535400</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

| Bias                                | 0.014       | 0.036       | 0.004       |
| Variance                            | 0.027       | 0.064       | 0.021       |
| MSE                                 | 0.1139752   |             |             |
Table 10: Estimated bias, variance and MSE for $\lambda_1 = 0.15$ $n = 200$ and 500 trials.

<table>
<thead>
<tr>
<th>Different bandwidths $h_1$ and $h_2$</th>
<th>$\hat{\theta}_{n1}$</th>
<th>$\hat{\theta}_{n2}$</th>
<th>$\hat{\theta}_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.004</td>
<td>0.011</td>
<td>-0.002</td>
</tr>
<tr>
<td>Variance</td>
<td>0.016</td>
<td>0.036</td>
<td>0.012</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0647620</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

| Bias                                | 0.007                | 0.018                | 0.002                |
| Variance                            | 0.012                | 0.026                | 0.011                |
| MSE                                 | 0.0492625            |                      |                      |

Table 11: Estimated bias, variance and MSE for $\lambda_1 = 0.65$ $n = 100$ and 500 trials.

<table>
<thead>
<tr>
<th>Different bandwidths $h_1$ and $h_2$</th>
<th>$\hat{\theta}_{n1}$</th>
<th>$\hat{\theta}_{n2}$</th>
<th>$\hat{\theta}_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>0.009</td>
<td>-0.041</td>
<td>-0.021</td>
</tr>
<tr>
<td>Variance</td>
<td>0.086</td>
<td>0.350</td>
<td>0.074</td>
</tr>
<tr>
<td>MSE</td>
<td>0.5124920</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

| Bias                                | 0.049                | 0.069                | 0.0001               |
| Variance                            | 0.068                | 0.196                | 0.052                |
| MSE                                 | 0.3238710            |                      |                      |

Table 12: Estimated bias, variance and MSE for $\lambda_1 = 0.65$ $n = 200$ and 500 trials.

<table>
<thead>
<tr>
<th>Different bandwidths $h_1$ and $h_2$</th>
<th>$\hat{\theta}_{n1}$</th>
<th>$\hat{\theta}_{n2}$</th>
<th>$\hat{\theta}_{n3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-0.013</td>
<td>-0.012</td>
<td>0.004</td>
</tr>
<tr>
<td>Variance</td>
<td>0.071</td>
<td>0.160</td>
<td>0.054</td>
</tr>
<tr>
<td>MSE</td>
<td>0.2856553</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidths $h_1 = h\hat{\sigma}(\theta'X)$ and $h_2 = h\hat{\sigma}(Y)$

| Bias                                | 0.019                | 0.059                | 0.022                |
| Variance                            | 0.060                | 0.165                | 0.053                |
| MSE                                 | 0.2825447            |                      |                      |

The results in Tables 9-12 shows that if, additionally to A1, the assumption A2 is not fulfilled the MSE stays similar to MSE in Model 2. Similarly as in Model 2, the results are worse than in Model 1 especially when censoring is heavy. Consequently, the assumption A1 seems to be crucial for our model.

5.4. Asymptotic variance

In this section we compare the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ for the estimator presented in this paper with that given in [1]. Using Theorem 2, it can be shown that the asymptotic covariance matrix, $\Sigma$, of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can be written as
\[
\Sigma = \left[ \int (1 - G(z - |x|)) (\nabla_{\theta} \log(f_{\theta}(z|\theta' x)))_{\theta=\theta_0} (\nabla_{\theta} \log(f_{\theta}(z|\theta' x)))_{\theta=\theta_0}) f(x, z) dx dz + \int (1 - F(c|x)) (\nabla_{\theta} \log(1 - F_{\theta}(c|\theta' x)))_{\theta=\theta_0} (\nabla_{\theta} \log(1 - F_{\theta}(c|\theta' x)))_{\theta=\theta_0}) f(x, c) dc \right]^{-1} \tag{16}
\]

The asymptotic covariance matrix of the estimator proposed by Bouaziz and Lopez \[1\] is given by
\[
\Sigma_{\tau} = V_{\tau}^{-1} \Delta_{\tau}(f_1) V_{\tau}^{-1},
\]
where, using the notation of the present paper, \(f_1(x, y) = (f_{\theta_0}(y|\theta' x))^{-1}1_{\{f_{\theta_0}(x, y) > c\}} \nabla_{\theta} f_{\theta}(y|\theta' x)_{\theta=\theta_0}, \) \(f_{\theta}^{\tau}\) denotes the density of \(\theta' X\) and \(Z \in A^r\) with \(A^r \in (-\infty, \tau], \)
\[
V_{\tau} = E\left( \frac{\nabla_{\theta} f_{\theta}^{\tau}(Z|\theta' X)_{\theta=\theta_0} \nabla_{\theta} f_{\theta}^{\tau}(Z|\theta' X)_{\theta=\theta_0}}{(f_{\theta_0}^{\tau}(Z|\theta' X))^2} 1_{\{f_{\theta_0}^{\tau}(x, y) > c\}} 1_{\{Z \in A^r\}} \right)
\]
and
\[
\Delta_{\tau}(f_1) = Var(\psi(Y, \delta, X, f_11_{\{Y \in A^r\}}))
\]
Moreover, using Lemma 4, we have
\[
\int_y^\tau \int f_1(x, t) 1_{\{t \in A^r\}} F(dx, dt) = E(f_1(X, Z) 1_{\{Z \in [y, \tau]\}}) = 0.
\]
Hence
\[
\Delta_{\tau}(f_1) = Var\left( \frac{\delta f_1(X, Y) 1_{\{Y \in A^r\}}}{1 - G(Y -)} \right) = E\left( \frac{\nabla_{\theta} f_{\theta}^{\tau}(Z|\theta' X)_{\theta=\theta_0} \nabla_{\theta} f_{\theta}^{\tau}(Z|\theta' X)_{\theta=\theta_0}}{(1 - G(Z -))(f_{\theta_0}^{\tau}(Z|\theta_0' X))^2} 1_{\{f_{\theta_0}^{\tau}(x, y) > c\}} 1_{\{Z \in A^r\}} \right),
\]
where the last equation is a consequence of Lemma 4 and assumption A3 (Assumption 5 in \[1\]).

Note that if there is no censoring, \(C = \infty\), both asymptotic covariance matrices, \(\Sigma\) in (16) and \(\Sigma_{\tau}\) from \[1\] with \(\tau = \infty\) \(\text{and}\) \(c = 0\), reduce to the asymptotic covariance matrix in the complete data case
\[
\Sigma_0 = \left[ E( (\nabla_{\theta} \log(f_{\theta}(Z|\theta' X))_{\theta=\theta_0}) (\nabla_{\theta} \log(f_{\theta}(Z|\theta' X))_{\theta=\theta_0})' ) \right]^{-1}.
\]
See Theorem 2 in \[5\], for details.

Because of the complicated structure of the asymptotic covariance matrices, they cannot be compared in general. Nevertheless, we present a comparison based on Model (15). Tables 13-14 show the asymptotic variances \(diag(\Sigma)\) and \(diag(\Sigma_{\tau})\) for Model 1 for \(1_{\{f_{\theta_0}^{\tau}(x, y) > c\}} \equiv 1\) and different values of \(\tau\). More precisely, we choose the \(\tau = \tau^*\) (which differs depending on censoring) that gives the best results, as well as \(\tau = F^{-1}(0.9)\) (which coincides with Table 5 in \[1\]) and \(\tau = \infty\) (no trimming). Since there are no explicit expressions for the expectations, we approximate them via Monte Carlo based on 30000 trials.
Table 13: Asymptotic variances using $\lambda_1 = 0.3$ for Model 1.

<table>
<thead>
<tr>
<th>Present estimator: $\text{diag}(\Sigma)$ given in (16)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.148</td>
<td>1.154</td>
</tr>
<tr>
<td>Bouaziz and Lopez [1]: $\text{diag}(\Sigma_{\tau^<em>})$ with $\tau = \tau^</em> = F^{-1}(0.8)$</td>
<td>0.208</td>
</tr>
<tr>
<td></td>
<td>1.620</td>
</tr>
<tr>
<td>Bouaziz and Lopez [1]: $\text{diag}(\Sigma_{\tau})$ with $\tau = F^{-1}(0.9)$</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>1.701</td>
</tr>
<tr>
<td>Bouaziz and Lopez [1]: $\text{diag}(\Sigma_{\tau})$ with $\tau = \infty$</td>
<td>0.581</td>
</tr>
<tr>
<td></td>
<td>4.553</td>
</tr>
</tbody>
</table>

Consequently, the asymptotic variance for the present estimator is always better than that of [1] for Model 1. Moreover, the variance in [1] is strongly dependent on the choice of $\tau$, specially when censoring is heavy. Hence, in this particular example, the estimator presented in this paper is more efficient.

6. German Credit Data

In this section we apply our model to the German Credit data set which is publicly available on the internet page http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data). This data set includes information about 1000 credits from which 300 were classified as bad credits and 700 as good credits. In our model we use four covariates such as credit amount, checking account, time of employment and savings account. Let us denote with $X = (X_1, X_2, X_3, X_4)'$ the vector with the above mentioned covariates and with $\theta_0 = (1, \theta_2, \theta_3, \theta_4)'$ the index to be estimated.

Since some of the $X_i$’s are ordinal (interval) variables, in order to use our approach, we change them into numerical variables. To be more specific, $X_1$ is already a continuous variable denoting amount of credit in DM, $X_2 \in \{-0.05, 0.01, 0.25, 0\}$ denotes the checking account in thousands of DM, $X_3 \in \{0, 0.5, 2.5, 5.5, 8.5\}$ denotes the years of employment and $X_4 \in \{0, 0.05, 0.25, 0.75, 1.25\}$ denotes the savings account in thousands of DM.

Additionally, to the explanatory variable $X$, we set $Z$, as the time to default and denote with $\delta = 1$ the bad/defaulted credits and with $\delta = 0$ the good credits. The results are presented in Table 15

<table>
<thead>
<tr>
<th>$\hat{\theta}_1$</th>
<th>$\hat{\theta}_2$</th>
<th>$\hat{\theta}_3$</th>
<th>$\hat{\theta}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2091</td>
<td>0.2312</td>
<td>2.1891</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Figure 1 presents the estimated conditional survival function for
\[ x_1 = (\text{credit}, 0.1, 4, 0.25)' \] where \( \text{credit} \in \{1, 10\} \)

\[ x_2 = (2, \text{account}, 4, 0.25)' \] where \( \text{account} \in \{-0.05, 0.25\} \)

7. Conclusions

In this paper we have introduced an estimator of the parameter vector in a single-index model under censoring and proved its asymptotic properties. We compared our method with that presented by Bouaziz and Lopez [1]. The simulation study has shown that our estimator gives mostly better results and is more efficient than that presented in [1]. The main difference between the two approaches comes from the definition of the likelihood function. The theoretical likelihood in our approach, defined in (2), is commonly applied in the censored data setup. Bouaziz and Lopez [1], on the other hand, estimated the quantity \( E(\log f_Z(Z|\theta'X)J(X)1_{Z \in A}) \), where \( f_Z \) is the density of \( Z \) given \( \theta'X \) and \( Z \in A^\tau \) while \( J(X)1_{Z \in A} \) is a trimming function. This kind of likelihood, with \( A^\tau = \mathbb{R} \), is generally used in the complete data setup (see, e.g., [5]). The trimming function in the estimated likelihood is also present in our model. Nevertheless, it goes to 1 and hence vanishes when the sample size increases. The next crucial difference between our method and that presented in [1] has to do with bandwidth selection. In order to estimate the conditional density, Bouaziz and Lopez [1] used one smoothing parameter \( h \) for both \( Z \) and \( \theta'X \). Consequently they needed fourth order kernels in order to prove the asymptotic normality. On the contrary, our approach uses two different bandwidths, \( h_1 \) and \( h_2 \), which allows us to use positive second order kernels. The only drawback of our method is the need of Assumption A2, which requires the independence between \( Z \) and \( C \) given the explanatory variable \( X \). Nevertheless, both methods require Assumption A1, the unconditional independence between \( Z \) and \( C \). Note that in many practical examples the censoring variable \( C \) is independent of \( X \). In such a case Assumption A2 is a consequence of A1.

Acknowledgements

The authors acknowledge financial support from Ministerio de Economía y Competitividad Grants MTM2008-00166 and MTM2011-22392 (EU ERDF support included). Additionally, Ewa Strzalkowska-Kominiak acknowledges financial support from a Juan de la Cierva scholarship, from the Spanish Ministerio de Economía y Competitividad.
References


Appendix A. Proofs of Theorems 2 - 4.

Proof of Theorem 2.

According to Lemma 1, we need to study the quantity:

\[ \hat{\theta}_n \left[ \theta_0 \right] - \hat{\theta}_n \left[ \theta_0 \right] = \alpha_n \left[ \theta_0 \right] + \beta_n \left[ \theta_0 \right], \]

where

\[ \alpha_n \left[ \theta_0 \right] = \hat{\theta}_n \left[ \theta_0 \right] - \hat{\theta}_n \left[ \theta_0 \right], \]

\[ \beta_n \left[ \theta_0 \right] = \hat{\theta}_n \left[ \theta_0 \right] - \hat{\theta}_n \left[ \theta_0 \right]. \]

For \( \beta_n \left[ \theta_0 \right] \), according to Lemma 2, we have

\[ \sqrt{n} \beta_n \left[ \theta_0 \right] \rightarrow N \left( 0, \Sigma_2 \right) \]

and

\[ 16 \]
\[ \beta_n(\theta_0) \to 0 \] in probability.

Concerning \( \alpha_n(\theta) \), set
\[
d(\theta' X_i, Y_i) = \int_{Y_i}^{\infty} f_{\theta} (\theta' X_i, y) \, dy
\]
Hence, by (2), (8), (10) and (11), we have
\[
\hat{l}_n(\theta_0) - \hat{I}_n(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_i \left( \log(\hat{r}(\theta_0' X_i, Y_i)) - \log(f_{\theta_0}(\theta_0' X_i, Y_i)) + \log(f_{\theta}(\theta_0' X_i)) - \log(\hat{s}(\theta_0' X_i)) \right) \right.
\]
\[ + (1 - \delta_i) \left( \log(\hat{d}(\theta_0' X_i, Y_i)) - \log(d(\theta_0' X_i, Y_i)) \right) \right] 1_{\{X_i \in A^{\alpha_n}, Y_i \leq a_n\}}
\]
\[
+ \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_i \log f_{\theta_0}(Y_i|\theta_0' X_i) + (1 - \delta_i) \log(1 - F_{\theta_0}(Y_i|\theta_0' X_i)) \right] \left( 1_{\{Y_i > a_n, X_i \in A^{\alpha_n}\}} + 1_{\{X_i \in A^{\alpha_n}\}} \right)
\]
First note, that \( a_n \) and \( c_n \) are deterministic sequences such that \( a_n \to \tau_H \) and \( c_n \to 0 \). Then, using Lemma 4, the derivative with respect to \( \theta \) of the last term is of order \( o_p(n^{-1/2}) \). Hence
\[
\hat{l}_{n}^{1}(\theta_0) - \hat{l}_{n}^{1}(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_i \left( \frac{f_{\theta_0}(\theta_0' X_i, Y_i)}{\hat{r}(\theta_0' X_i, Y_i)} - \frac{f_{\theta_0}(\theta_0' X_i, Y_i)}{f_{\theta}(\theta_0' X_i, Y_i)} \right) + \left( \frac{f_{\theta}(\theta_0' X_i)}{f_{\theta_0}(\theta_0' X_i)} - \frac{\hat{s}(\theta_0' X_i)}{\hat{s}(\theta_0' X_i)} \right) \right.
\]
\[ + (1 - \delta_i) \left( \frac{d(\theta_0' X_i, Y_i)}{\hat{d}(\theta_0' X_i, Y_i)} - \frac{d(\theta_0' X_i, Y_i)}{d(\theta_0' X_i, Y_i)} \right) \right] 1_{\{X_i \in A^{\alpha_n}, Y_i \leq a_n\}} + o_p(n^{-1/2}).
\]
Furthermore, by repeated use of
\[
\frac{1}{\hat{r}} = \frac{1}{f_{\theta_0}} + \frac{f_{\theta_0} - \hat{r}}{f_{\theta_0} \hat{r}}
\]
and similar expansions for \( \hat{s} \) and \( \hat{d} \), we obtain
\[
\hat{l}_{n}^{1}(\theta_0) - \hat{l}_{n}^{1}(\theta_0) = \sum_{k=1}^{12} A_{kn} + o_p(n^{-1/2}),
\]
where
Now we need to consider each of these terms separately.

To deal with \( A_{in} \) for \( i = 1, \ldots, 6 \), let us define the following functions:

\[
\tilde{r}(\theta' x, y) = \frac{1}{h_1 h_2} \int K \left( \frac{\theta' x - u}{h_1} \right) K \left( \frac{y - v}{h_2} \right) f_\theta(u, v) 1_{\{v \leq a_n\}} du dv
\]

\[
\tilde{s}(\theta' x) = \frac{1}{h_1} \int K \left( \frac{\theta' x - u}{h_1} \right) f_\theta(u, v) 1_{\{v \leq a_n\}} du dv
\]

and

\[
\tilde{d}(\theta' x, y) = \frac{1}{h_1} \int K \left( \frac{\theta' x - u}{h_1} \right) \Xi \left( \frac{y - v}{h_2} \right) f_\theta(u, v) 1_{\{v \leq a_n\}} du dv.
\]
Moreover let

\[ A_{1n} = B_{1n} + C_{1n}, \]

where

\[ B_{1n} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{r^{[1]}(\theta_0', X_i, Y_i) - \hat{r}^{[1]}(\theta_0', X_i, Y_i)}{f_{\theta_0}(\theta_0', X_i, Y_i)} 1_{\{X_i \in A^{a_n}, Y_i \leq a_n\}} \]

\[ C_{1n} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{r^{[1]}(\theta_0', X_i, Y_i) - f^{[1]}_0(\theta_0', X_i, Y_i)}{f_{\theta_0}(\theta_0', X_i, Y_i)} 1_{\{X_i \in A^{a_n}, Y_i \leq a_n\}} \]

and

\[ A_{2n} = B_{2n} + C_{2n}, \]

where

\[ B_{2n} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{f^{[1]}_0(\theta_0', X_i, Y_i)}{f_{\theta_0}(\theta_0', X_i, Y_i)} \left( \hat{r}(\theta_0', X_i, Y_i) - \hat{r}(\theta_0', X_i, Y_i) \right) 1_{\{X_i \in A^{a_n}, Y_i \leq a_n\}} \]

\[ C_{2n} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{f^{[1]}_0(\theta_0', X_i, Y_i)}{f_{\theta_0}(\theta_0', X_i, Y_i)} \left( f_{\theta_0}(\theta_0', X_i, Y_i) - \hat{r}(\theta_0', X_i, Y_i) \right) 1_{\{X_i \in A^{a_n}, Y_i \leq a_n\}} \]

Similarly \( A_{in} = B_{in} + C_{in} \) for \( i = 3, 4, 5, 6 \). Set

\[ C_n = C_{1n} + C_{2n} + C_{3n} + C_{4n} + C_{5n} + C_{6n} \]

and

\[ B_n = B_{1n} + B_{2n} + B_{3n} + B_{4n} + B_{5n} + B_{6n}. \]

Now using Lemma 5,

\[ \sqrt{n} C_n = o_P(1). \]

Moreover, Lemma 13 leads to

\[ \sqrt{n} B_n = o_P(1). \]

Finally, Lemma 14 implies that \( \sqrt{n} \sum_{k=7}^{12} A_{kn} = o_P(1) \), while Lemma 15 shows that \( \hat{l}_n^{[2]}(\theta) \rightarrow l^{[2]}(\theta) \). This completes the proof.

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

\[ \frac{1}{n} l_n^{[1]}(\theta_0) - l^{[1]}(\theta_0) \overset{p}{\rightarrow} 0 \]

and

\[ \sqrt{n}(\frac{1}{n} l_n^{[1]}(\theta_0) - l^{[1]}(\theta_0)) \rightarrow N(0, \Sigma_1) \]
Proof.
According to (2) and since \( l^{(1)}(\theta_0) = E(l^{(1)}_n(\theta_0)) = 0 \), we have

\[
\sqrt{n}(l^{(1)}_n(\theta_0) - l^{(1)}(\theta_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \delta_i \frac{f^{(1)}_i(Y_i|\theta'X_i)}{\tilde{f}_0(Y_i|\theta'_0X_i)} - (1 - \delta_i) \frac{F^{(1)}_i(Y_i|\theta'X_i)_{\theta=\theta_0}}{1 - \tilde{f}_0(Y_i|\theta'_0X_i)} \right].
\]

Then, \( \sqrt{n}(l^{(1)}_n(\theta_0) - l^{(1)}(\theta_0)) \) is a sum of i.i.d. random vectors and the Law of Large Numbers gives the convergence to zero in probability. Moreover, by the Central Limit Theorem we have

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

To compute \( \Sigma_1 \), let \( f_\delta(x, z, c) \) be the density of \((X, Z, C)\), while \( f(x, z) \) and \( g(x, z) \) are the densities of \((X, Z)\) and \((X, C)\). Recall that \( Y = \min(Z, C)\) while \( \delta = 1_{(z \leq c)} \). Hence, using A1, A2 and A3, since \( \delta(1 - \delta) = 0 \), we obtain

\[
\Sigma_1 = \lim_{n \to \infty} E \left[ \left( \frac{f^{(1)}_i(Y_i|\theta'X_i)_{\theta=\theta_0}}{\tilde{f}_0(Y_i|\theta'_0X_i)} - (1 - \delta_i) \frac{F^{(1)}_i(Y_i|\theta'X_i)_{\theta=\theta_0}}{1 - \tilde{f}_0(Y_i|\theta'_0X_i)} \right)^2 \right]
\]

\[
= \int (1 - G(z - |x|))(\nabla_\theta \log(f_\theta(z|x))_{\theta=\theta_0})(\nabla_\theta \log(f_\theta(z|x))_{\theta=\theta_0})^t f(x, z) dx dz \\
+ \int (1 - F(c|x))(\nabla_\theta \log(1 - F_\theta(c|x))_{\theta=\theta_0})(\nabla_\theta \log(1 - F_\theta(c|x))_{\theta=\theta_0})^t g(x, c) dx dc.
\]

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

Next we consider the gradients of \( f_\theta \). We have following results:

**Lemma 3.** Under A8, we have

a) \[
\frac{\partial}{\partial \theta_k} f_\theta(\theta'x, y)_{\theta=\theta_0} = \frac{\partial}{\partial t} \left( f_{\theta_0}(t, y) \left[ x_k - E(X_k|\theta'_0X = t, Y = y) \right] \right)_{t=\theta_0}'x
\]

b) \[
\frac{\partial}{\partial \theta_k} f_{\theta'X}(\theta'x)_{\theta=\theta_0} = \frac{\partial}{\partial t} \left( f_{\theta'_0X}(t) \left[ x_k - E(X_k|\theta'_0X = t) \right] \right)_{t=\theta_0}'x
\]

**Proof.** Let \( \theta_0 = (\theta_{01}, ..., \theta_{0d})' \) and

\( \theta^* = \theta_0 + (0, ..., h, ..., 0)' \), where \( h \) is in position \( k \)

Then, for every \( x = (x_1, ..., x_d)' \) and \( y \in \mathbb{R} \), we have

\[
\frac{\partial}{\partial \theta_k} f_\theta(\theta'x, y)_{\theta=\theta_0} = A + B + C,
\]
Finally, we have
\[ f_{\theta}(\theta' x, y) - f_{\theta}(\theta' x, y) \]
\[ \frac{\partial}{\partial \theta} P(\theta' X \leq t, Y \leq \tilde{y}) = \frac{\partial}{\partial \theta} P(\theta' X + \theta_k X_k - \theta_{0k} X_k \leq t, Y \leq \tilde{y}) \]
\[ = \frac{\partial}{\partial \theta} \int_{-\infty}^{\tilde{y}} \int_{-\infty}^{t} f_{X_k, \theta'_0, Y}(x_k, u, z) du dz \]
\[ = \int_{-\infty}^{\tilde{y}} \left( \frac{\partial}{\partial \theta} \int_{-\infty}^{t} f_{X_k, \theta'_0, Y}(x_k, u, z) du \right) dx_k dz \]
\[ = \int_{-\infty}^{\tilde{y}} f_{X_k, \theta'_0, Y}(x_k, t - \theta_k x_k + \theta_{0k} x_k, z)(-x_k) dx_k dz \]
Thus,\[ \frac{\partial}{\partial \theta} P(\theta' X \leq t, Y \leq \tilde{y}) \theta = \theta_0 = \int_{-\infty}^{\tilde{y}} f_{X_k, \theta'_0, Y}(x_k, t, z)(-x_k) dx_k dz. \]
Finally,
Proof.

Lemma 5. This completes the proof.

Proof.

Lemma 4. Under A4, we have

\[ a) \quad \nabla_\theta \Phi_{\theta_0}(y|\theta_0, x)_{x=\theta_0} = \left[ x - E(X|\theta_0, X = \theta_0, x) \right] \frac{\partial}{\partial t} \Phi_{\theta_0}(y|t, \theta_0, x)_{x=\theta_0}. \]

\[ b) \quad E(\nabla_\theta \log \Phi_{\theta_0}(Z|\theta_0, X))_{x=\theta_0} = 0 \]

and

\[ E(\nabla_\theta \log(1 - F_{\theta_0}(C|\theta_0, X))_{x=\theta_0} = 0. \]

Proof.

Using A4, part a) is an immediate consequence of Lemma 3. As to b), by A4, we have

\[ E\left[ X_1 - E(X|\theta_0, X = \theta_0, x) \right]_{x=\theta_0} = E(X|\theta_0, X = \theta_0, x) - E(X|\theta_0, X = \theta_0, x) = 0. \]

This completes the proof.

Lemma 5. Under A2, A4, A6 and A7, we have following results:

\[ \sqrt{n}C_n = O_p(\sqrt{n}h_1^4 + \sqrt{n}h_2^2) = o_p(1). \]

Proof.

First, using a change of variable, we have

\[ \tilde{r}(\theta_0, X_i, Y_i)_{1|\{Y_i \leq a_n\}} = \frac{1}{h_1h_2} \int K\left( \frac{\theta_0, X_i - u}{h_1} \right) K\left( \frac{Y_i - u}{h_2} \right) f_{\theta_0}(u, v)_{1|\{v \leq a_n\}} du dv 1_{\{Y_i \leq a_n\}} \]

and

\[ \tilde{r}^{[1]}(\theta_0, X_i, Y_i)_{1|\{Y_i \leq a_n\}} = \frac{1}{h_1^2h_2} \int (X_i - E(X|\theta_0, X = u)) K\left( \frac{\theta_0, X_i - u}{h_1} \right) K\left( \frac{Y_i - v}{h_2} \right) f_{\theta_0}(u, v)_{1|\{v \leq a_n\}} du dv 1_{\{Y_i \leq a_n\}} \]

and

\[ = \int (X_i - E(X|\theta_0, X = u)) K\left( \frac{\theta_0, X_i - u}{h_1} \right) K\left( \frac{Y_i - v}{h_2} \right) f_{\theta_0}(u, v)_{1|\{v \leq a_n\}} du dv 1_{\{Y_i \leq a_n\}} \]

and

\[ = \int (X_i - E(X|\theta_0, X = \theta_0, X_i - h_1z_1, Y_i - h_2z_2)) K\left( \frac{Y_i - v}{h_2} \right) f_{\theta_0}(\theta_0, X_i - h_1z_1, Y_i - h_2z_2)_{1|\{Y_i \leq a_n\}} dz_1 dz_2 1_{\{Y_i \leq a_n\}} \]
Then, observe that \( \int z^a K(z) \, dz = 0 \) for \( a \) odd, \( \int K(z) \, dz = 1 \) and \( b_K = \int z^2 K(z) \, dz < \infty \), together with \( \int z^b K'(z) \, dz = 0 \) for \( b \) even, \( \int z K'(z) \, dz = -1 \) and \( \int z^2 K'(z) \, dz = -3 \int z^2 K(z) \, dz \).

Moreover, using A7, the function \( f_\theta(u, v)1_{\{v \leq \alpha_n\}} \) is three times differentiable on \([v, Y_i]\) if \( v \leq Y_i\) and on \([Y_i, v]\) if \( v > Y_i\). Hence, using a Taylor expansion, we have

\[
\tilde{r}(\theta_0 X_i, Y_i)1_{\{Y_i \leq \alpha_n\}} = f_{\theta_0}(\theta_0' X_i, Y_i)1_{\{Y_i \leq \alpha_n\}} + \frac{b_K h_1^2}{2} \frac{\partial^2}{\partial u^2} f_{\theta_0}(u, Y_i)_{u=\theta_0' X_i} 1_{\{Y_i \leq \alpha_n\}} \\
+ \frac{b_K h_2^2}{2} \frac{\partial^2}{\partial^2 v} (f_{\theta_0}(\theta_0' X_i, v)1_{\{Y_i \leq \alpha_n\}})_{v=Y_i} 1_{\{Y_i \leq \alpha_n\}} + O_P(h_1^4 + h_2^4 + h_2^4).
\]

Similarly, Lemma 3 additionally implies that

\[
\tilde{r}^{[1]}(\theta_0' X_i, Y_i)1_{\{Y_i \leq \alpha_n\}} = f_{\theta_0}^{[1]}(\theta_0' X_i, Y_i)1_{\{Y_i \leq \alpha_n\}} + \frac{b_K h_1^2}{2} \frac{\partial^2}{\partial u^2} ([X_i - E(X|\theta_0' X = u)] f_{\theta_0}(u, Y_i)_{u=\theta_0' X_i} 1_{\{Y_i \leq \alpha_n\}}) \\
+ \frac{b_K h_2^2}{2} \frac{\partial^2}{\partial u^2} ([X_i - E(X|\theta_0' X = u)] f_{\theta_0}(u, v)1_{\{v \leq \alpha_n\}})_{u=\theta_0' X_i, v=Y_i} 1_{\{Y_i \leq \alpha_n\}} + O_P(h_1^4 + h_2^4 + h_2^4).
\]

Finally, repeating the same steps for \( \tilde{s}(\theta_0' X_i) \) and \( \tilde{d}(\theta_0' X_i, Y_i) \) we obtain

\[
C_n = c_{1n} + c_{2n} + \tilde{c}_{1n} + O_P(h_1^4 + h_2^4)
\]

where

\[
c_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} h_1^2 b_K c_{1i} 1_{\{X_i \in A^{<n}\}},
\]

\[
c_{2n} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} h_1^2 b_K c_{2i} 1_{\{X_i \in A^{<n}\}},
\]

\[
\tilde{c}_{1n} = -\frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} h_1^2 b_K (c_{1i} + c_{2i}) 1_{\{X_i \in A^{<n}, Y_i > \alpha_n\}},
\]

\[
c_{1i} = \frac{\delta_i}{f_{\theta_0}(\theta_0' X_i, Y_i)} \frac{d^3}{du^3} \left([X_i - E(X|\theta_0' X = u)] f_{\theta_0}(u, Y_i)\right)_{u=\theta_0' X_i} \\
- \frac{1}{f_{\theta_0'}(X_i)} \frac{d^3}{du^3} \left([X_i - E(X|\theta_0' X = u)] f_{\theta_0'}(u)\right)_{u=\theta_0' X_i} \\
+ \frac{1 - \delta_i}{d(Y_i, \theta_0' X)} \int_{(Y_i, \infty)} \frac{d^3}{du^3} \left([X_i - E(X|\theta_0' X = u)] f_{\theta_0}(u, v)\right)_{u=\theta_0' X_i} \, dv
\]

and
\[ c_{2i} = -\delta \frac{\nabla \theta f_0(\theta' X_i, \theta' Y_i)}{f_{\theta_0}(\theta_0^i X_i)} \frac{d^2}{du^2} f_{\theta_0}(u, \theta_0^i Y_i) \bigg|_{u=\theta_0^i X_i} - \frac{\nabla \theta f_{\theta_0^i X}(\theta' X_i, \theta' Y_i)}{f_{\theta_0^i X}(\theta' X_i)} \frac{d^2}{du^2} f_{\theta_0^i X}(\theta_0^i X_i) \bigg|_{u=\theta_0^i X_i} \]

\[-(1 - \delta_i) \frac{\nabla \theta d(\theta' X_i, \theta' Y_i)}{d^2(\theta_0^i X_i, \theta_0^i Y_i)} \int_{(Y_i, \infty)} \frac{d^2}{du^2} f_{\theta_0}(u, \theta_0^i X_i) \nu(u, \theta_0^i X_i) \, du. \]

First we deal with \( \bar{e}_{1n} \). Lemma 2.3 in [24] leads to

\[ a_n = H^{-1}(1 - \frac{s_n}{n}) \geq Y_{n-\delta_n n}, \]

where \( \delta_n = [s_n/2] \). Hence

\[ \sqrt{n}|c_{1n}| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2} h_1^2 b_K|c_{1i} + c_{2i} 1_{(X_i \in A^\infty)} = O_P(n^{1/\theta} h_1^2 (\log(n))^{2+\varepsilon_1}) = o_P(1). \]

As to \( \bar{e}_{1n} \) and \( \bar{e}_{2n} \), we will show that \( E(\bar{e}_{1n}) = 0 \), \( Var(\bar{e}_{1n}) = o(n^{-1}) \) for \( k = 1, 2 \). For this set \( f_3(x, z, c) \) the density of \((X, Z, C)\) and \( g(c|x) \) the conditional density of \( C \) given \( X \). By A2 we have

\[ f_3(x, z, c) = g(c|x)f(z|x)f_X(x). \]

Moreover, since \( f_{\theta_0}(z|\theta_0^i x) = f(z|x) \), we have that \( f(x, z) = f_{\theta_0}(\theta_0^i x, z)f_X(x)/f_{\theta_0^i X}(\theta_0^i x) \) and \( \int f(z|x)dz = d(\theta_0^i x, c)/f_{\theta_0^i X}(\theta_0^i x) \). Hence

\[ E(\bar{c}_{1i} 1_{X_i \in A^\infty}) = \int \frac{1_{(z \leq c, x \in A^\infty)}}{f_{\theta_0^i X}(\theta_0^i x, c)} \frac{d^3}{du^3} \left( |z - E(X|\theta_0^i X = u)|f_{\theta_0}(u, z) \bigg|_{u=\theta_0^i x} g(c|x)f(z|x)f_X(x) \right) \, dx \, dz \, dc \]

\[-\int \frac{1_{(x \in A^\infty)}}{f_{\theta_0^i X}(\theta_0^i x)} \frac{d^3}{du^3} \left( |z - E(X|\theta_0^i X = u)|f_{\theta_0}(u, z) \bigg|_{u=\theta_0^i x} f_X(x) \right) \, dx \]

\[+ \int \frac{1_{(z \geq c, x \in A^\infty)}}{d(\theta_0^i x, c)} \int_{(c, \infty)} \frac{d^3}{du^3} \left( |z - E(X|\theta_0^i X = u)|f_{\theta_0}(u, v) \bigg|_{u=\theta_0^i x} g(c|x)f(z|x)f_X(x) \right) \, dv \]

where the last equation is consequence of change \( v \) to \( z \) in last coefficient and the fact that \( \int g(c|x)dx = 1 \). Moreover \( \bar{c}_{1i} \) are i.i.d. and, since \( h_1^2 \to 0 \), \( Var(\sqrt{n}\bar{c}_{1n}) \to 0 \). Hence \( \sqrt{n}\bar{c}_{1n} \to 0 \) in probability.

For \( \bar{c}_{2n}, \)

\[ E(\bar{c}_{2i} 1_{X_i \in A^\infty}) = -\int \frac{1_{(\leq c)}}{f_{\theta_0^i X}(\theta_0^i x, z)} \frac{d^2}{du^2} f_{\theta_0}(u, z) \bigg|_{u=\theta_0^i x} g(c|x)f(z|x)f_X(x) \right) 1_{(e \in A^\infty)} \, dx \, dz \, dc \]

\[+ \int \frac{\nabla \theta f_{\theta_0^i X}(\theta^i x)}{f_{\theta_0^i X}(\theta_0^i x)} \frac{d^2}{du^2} f_{\theta_0}(u, z) \bigg|_{u=\theta_0^i x} g(c|x)f(z|x)f_X(x) \right) 1_{(e \in A^\infty)} \, dx \]

\[-\int \frac{1_{(z \geq c)}}{d(\theta_0^i x, c)} \left( \int_{(c, \infty)} \frac{d^2}{du^2} f_{\theta_0}(u, v) \bigg|_{u=\theta_0^i x} g(c|x)f(z|x)f_X(x) \right) \right) 1_{(e \in A^\infty)} \, dx \]

\[= -\int \frac{1_{(\leq c)}}{f_{\theta_0^i X}(\theta_0^i x, z)} \frac{d^2}{du^2} f_{\theta_0}(u, z) \bigg|_{u=\theta_0^i x} g(c|x)f(z|x)f_X(x) \right) \frac{f_X(x)}{f_{\theta_0^i X}(\theta_0^i x)} \right) 1_{(e \in A^\infty)} \, dx \, dz \, dc \]

\[+ \int \frac{\nabla \theta f_{\theta_0^i X}(\theta^i x)}{f_{\theta_0^i X}(\theta_0^i x)} \frac{d^2}{du^2} f_{\theta_0}(u, z) \bigg|_{u=\theta_0^i x} g(c|x)f(z|x)f_X(x) \right) \frac{f_X(x)}{f_{\theta_0^i X}(\theta_0^i x)} \right) 1_{(e \in A^\infty)} \, dx \, dz \, dc \]
where the last equation is a consequence of Lemma 4.

Moreover, set $H_0(s) = \mathbb{P}(Y \leq s, \delta = 0)$. Then we have

$$G_n(t) - G(t) = (1 - G(t)) \frac{1}{n} \sum_{k=1}^{n} \psi(Y_k, \delta_k, t) + R_n(t), \tag{A.2}$$

where

$$\psi(Y_k, \delta_k, t) = \frac{1 - \delta_k}{1 - H(Y_k - t)} 1_{\{Y_k \leq t\}} - \int_{(0,t]} \frac{H_0(ds)}{1 - H(s -)} + \int_{(-\infty,t]} \frac{1(\{Y_k \leq s\} - H(s))}{(1 - H(s))^2} H_0(ds)$$

and for $r_n/n \uparrow, (n - r_n)/\log(n) \to \infty$ and each $\varepsilon > 0$

$$\sup_{t \leq Y_{r_n}} |R_n| = O \left( \frac{(\log(n))^{2+\varepsilon}}{2n - r_{2n}} \right) \text{ with probability 1.}$$
we obtain the following decomposition

\[
\sup_{t \leq Y_{n,n}} \left| \frac{G_n(t) - G(t)}{1 - G(t)} \right| = \begin{cases} 
O \left( \frac{\log(n)}{2n^{r/2}} \right) \quad & \text{with probability } 1 \\
O_P \left( \frac{1}{\sqrt{n}} \right) \quad & \text{.} 
\end{cases}
\tag{A.3}
\]

Proof.
The proof of (A.1) can be found, e.g., in Zhou [31]. Equation (A.2) is a consequence of Theorems 1.1 and 1.4 from Stute [24], while (A.3) follows from Theorem 2 from Csörgő [4]. Both rewritten for case where \( C \sim G \) is a variable of interest observed if \( 1 - \delta = 1 \).

Lemma 7. Under the conditions in Theorem 2 we have

\[
B_{1n} = \frac{1}{n} \sum_{j=1}^{n} H_{1j} + \frac{1}{n} \sum_{k=1}^{n} \bar{N}_{1k} + o_P(n^{-1/2}),
\]

where

\[
H_{1j} = \frac{-\delta j 1_{(Y_j \leq a_n)}}{1 - G(Y_j)} \int \left( (E(X_i|\theta_0 X_i = t) - X_j) (1 - G_{\theta_0}(Y_j - |t|)) 1_{ \{ f_{\theta_0 X}(t) > a_n \} } \right) dt = \theta_0 X_j
\]

\[
+ E \left( \frac{-\delta j 1_{(Y_j \leq a_n)}}{1 - G(Y_j)} \int \left( (E(X_i|\theta_0 X_i = t) - X_j) (1 - G_{\theta_0}(Y_j - |t|)) 1_{ \{ f_{\theta_0 X}(t) > a_n \} } \right) dt \right).
\]

and

\[
\bar{N}_{1k} = - \int \frac{d}{dt} \left( (E(X|\theta_0 X = t) - u)(1 - G_{\theta_0}(v - |t|)) 1_{ \{ f_{\theta_0 X}(t) > a_n \} } \right) \psi(Y_k, \delta_k, v) \frac{f(u, v) 1_{ \{ u \leq a_n \} } dv}{1 - G(v)}.
\]

Proof.

Let \( H_1(x, y) = P(X \leq x, Y \leq y, \delta = 1) \) and note that, under A1 and A3, we have

\[
H_1(dx, dy) = (1 - G(y-))F(dx, dy).
\]

Hence

\[
B_{1n} = \frac{1}{n} \sum_{i=1}^{n} \delta i 1_{(X_i, Y_i) \in A_{\theta_0}, Y_i \leq a_n} f_{\theta_0 X_i}, Y_i
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \frac{\delta X_i}{h^2 h^2} \int \frac{1}{1 - G_{\theta_0}(Y_j - |t|)} (X_i - X_j) K' \left( \frac{\theta_0 X_i - \theta_0 X_j}{1 - G_{\theta_0}(Y_j - |t|)} \right) \frac{Y_i - Y_j}{h^2} \int \frac{1_{\{v \leq a_n\}}}{1 - G(v)} H_1(du, dv)
\]

where \( K'(t) \) denotes the derivatives of \( K \) with respect to \( t \).

Finally, using

\[
\frac{1}{1 - G_{\theta_0}(v-)} - \frac{1}{1 - G(v-)} = \frac{G_{\theta_0}(v-)}{1 - G(v-)} + \frac{G(v-)}{(1 - G(v-))(1 - G_{\theta_0}(v-))}
\]

we obtain the following decomposition

26
Using (A.1) and (A.3) it is easy to prove that, choosing \( r_n = n - s_n \) and \( s_n = cn^{2/3}(\log(n))^{2+\varepsilon_1} \), we obtain \( \sqrt{n}b_{3n} = o_P(1) \). Moreover, we have

\[ b_{2n} = b_{21n} + b_{22n}, \]

where

\[ b_{21n} = \frac{1}{h_2^n b_{2n}^2} \sum_{i=1}^{n} \sum_{j \neq i} \int \left( X_i - u \right) K' \left( \frac{\theta_0 X_i - \theta_0 u}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \frac{1}{1 - G(Y_j -)} \sup_{0 \leq u \leq 1} H_1(du, dv) \text{1}_{\{f_{\theta_0 X_i} > u, Y_j \leq a_n\}} \]

As to \( b_{22n} \), according to Cs"{o}rg"{o} [4], for \( n \) large enough and with high probability

\[ a_n = H^{-1} \left( 1 - \frac{s_n}{n} \right) \leq Y_{r_n, n}, \]

where \( r_n = n - \lfloor s_n/5 \rfloor + 1 \) and \( s_n \approx n^{2/3}(\log(n))^{2+\varepsilon_1} \). Hence, using Lemma 6 with \( \varepsilon < \varepsilon_1 \), we obtain

\[ \sup_{t \leq a_n} |R_n(t)| \leq \sup_{t \leq Y_{r_n, n}} |R_n(t)| = o(n^{-2/3}) \]

and

\[ \sqrt{n}b_{22n} = o_P(n^{-1/6}h_1^{-1}) = o_P(1). \]

Concerning \( b_{1n} \), set

\[ B_{1n} = b_{1n} + b_{2n} + b_{3n}, \]
\[
H_{ij} = H(\theta_0 X_i, Y_i, \delta_i, \theta_0 X_j, Y_j, \delta_j)
\]
\[
= \frac{1}{h_1^2 h_2} \int_{\theta_0} \left[ \frac{1}{1 - G(y_{ij})} (X_i - X_j) K' \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \right] \text{1}_{\{Y_i \leq a_n\}}
\]
\[
- \int (X_i - u) K' \left( \frac{\theta_0 X_i - \theta_0 u}{h_1} \right) K \left( \frac{Y_i - v}{h_2} \right) \text{1}_{\{v \leq a_n\}} \text{1}_{\{f_{\theta_0}(\theta_0 X_i) > c_n, Y_i \leq a_n\}}
\]
and
\[
\bar{H}_j = E(H_{ij}|X_j, Y_j, \delta_j).
\]

Then
\[
\sqrt{n} b_{1n} = \sqrt{n} \sum_{i=1}^{n} \sum_{j \neq i} (H_{ij} - \bar{H}_j) + \sqrt{n} \sum_{i=1}^{n} \sum_{j \neq i} \bar{H}_j = \sqrt{n} b_{11n} + \sqrt{n} b_{12n}.
\]  
(A.4)

For the first term, observe that
\[
E(H_{ij} - \bar{H}_j) = E(H_{ij} - \bar{H}_j|X_i, Y_i, \delta_i) = E(H_{ij} - \bar{H}_j|X_j, Y_j, \delta_j) = 0.
\]

Since the goal is to show that the vector \(\sqrt{n} b_{12n} \in \mathbb{R}^d\) goes to zero in probability, it is enough to prove that componentwise. Let denote by \((v)_m\), for \(m = 1, ..., d\), the \(m\)th component of vector \(v \in \mathbb{R}^d\). Hence
\[
E\left( \frac{\delta_2}{(1 - G(y_{ij}))^2} \right) = \frac{1}{1 - G(y_{ij})}
\]
and, since \(Z_i \geq 0\), we have componentwise
\[
\frac{1}{n} E(H_{ij} - \bar{H}_j)^2_m \leq \frac{1}{n} E((H_{ij})^2_m) \leq \frac{1}{h_1^2 h_2} \int_{\theta_0} \left( \frac{\delta_1 (f_{\theta_0}(\theta_0 X_i) > c_n, 0 \leq Y_i \leq a_n, 0 \leq Y_2 \leq a_n)}{f_{\theta_0}(\theta_0 X_i, Y_i)} \right) \left( \frac{\delta_2}{(1 - G(y_{ij}))^2} \right)
\]
\[
(X_i - X_j)^2_m (K')^2 \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right)
\]
\[
= \frac{1}{h_1^2 h_2} \int f(u,v)(1 - G(v + h_2 z_2 -)) \text{1}_{\{f_{\theta_0}(\theta_0 u + h_1 z_1) > c_n\}} \text{1}_{\{f_{\theta_0}(\theta_0 u + h_1 z_1) > c_n\}} (K')^2 (z_1) K^2 (z_2)
\]
\[
E X_i - u)_{m} \theta_0 X_1 = \theta_0 u + h_1 z_1 \text{1}_{\{0 \leq v \leq a_n, 0 \leq v + h_2 z_2 \leq a_n\}} d z_1 d z_2 d u d v.
\]

Moreover, since \(f_{u,v} = f_{x,u}(\theta_0 u)\), the first term of a Taylor expansion is just
\[
\frac{a_n}{h_1 h_2} \int E[(X_1 - u)^2_m \theta_0 X_1 = \theta_0 u] \text{1}_{\{f_{\theta_0}(\theta_0 u) > c_n\}} \int f_x(u) du \int (K')^2 (z_1) K^2 (z_2) d z_1 d z_2.
\]
Hence, using A5, we obtain
\[
\frac{1}{n} E((H_{12})^2_m) \leq \frac{a_n}{c_n \sqrt{n} h_2} C,
\]
where $C$ is a constant.
Moreover, according to inequality (1) in [22], we have for $p = j/n$
\[
H^{-1}(p) \leq E(Y_{j:n}) + \sigma \frac{H_{j:n}(p)}{\sqrt{p(1-p)}},
\]
where $\sigma = Var(Y) < \infty$.
Furthermore, according to the inequality in Proposition 1 in [10],
\[
E(Y_{j:n}) = O(\sqrt{(j-1)/(n-j-1)}) = O((p-1/n)/(1-p-1/n)).
\]
Hence
\[
a_n \leq H^{-1}\left(\frac{n - [n^{2/3}]}{n}\right) = O(n^{1/6})
\]
and, using A6, we obtain $\frac{1}{n} E((H_{12})^2_m) = o(1)$ for $m = 1, \ldots, d$. A similar property holds for $E(H_{21} - \bar{H}_1)_m(H_{12} - \bar{H}_2)_m$. So that
\[
\sqrt{n} b_{12n} \overset{p}{\to} 0.
\]
For $\sqrt{n} b_{12n}$, using A2, we have
\[
\tilde{H}_j = E\left(\frac{1}{h_1^2 h_2} \int \frac{1_{[z \leq c]}}{f_{\theta_0}(x, z)} \frac{\delta_j}{1 - G(Y_j)} (E(X_i|\theta_0 X_i = x) - X_j) K' \left(\frac{x - \theta'_0 X_j}{h_1}\right) K \left(\frac{z - Y_j}{h_2}\right) f_{\theta_0}(x, z) g_{\theta_0}(c|x) 1_{\{f_{\theta_0}(x, z) > c, Y_j \leq a_n \}} dx dz d\theta_0 \right),
\]
\[
- \frac{1}{h_1^2 h_2} \int \int \frac{1_{[z \leq c]}}{f_{\theta_0}(x, z)} (E(X_i|\theta_0 X_i = x) - u) K' \left(\frac{x - \theta'_0 u}{h_1}\right) K \left(\frac{z - v}{h_2}\right) f_{\theta_0}(x, z) g_{\theta_0}(c|x) 1_{\{f_{\theta_0}(x, z) > c, x \leq a_n, z \leq a_n \}} dx dz d\theta_0 \left[X, Y_j, \delta_j \right],
\]
Finally, using a change of variable and a Taylor expansion, we obtain:
\[
\tilde{H}_j = \tilde{H}_{1j} + h_{1j},
\]
where
\[
\tilde{H}_{1j} = -\frac{\delta_j}{1 - G(Y_j)} \frac{d}{dt} [E(X_i|\theta_0 X_i = t) - X_j] (1 - G_{\theta_0}(Y_j - |t|)) 1_{\{f_{\theta_0, X}(t) > c_n\}} |t = \theta_0 X_j 1_{\{Y_j \leq a_n\}} + E\left(\frac{\delta_j 1_{[Y_j \leq a_n]}}{1 - G(Y_j)} \frac{d}{dt} [E(X_i|\theta_0 X_i = t) - X_j] (1 - G_{\theta_0}(Y_j - |t|)) 1_{\{f_{\theta_0, X}(t) > c_n\}} |t = \theta_0 X_j \right)
\]
and
\[
h_{1j} = o_p(1).
\]
Moreover, $h_{1j} = \tilde{H}_j - \tilde{H}_{1j}$. Hence $h_{1j}$ are i.i.d. with expectation 0. Hence $\frac{1}{\sqrt{n}} \sum_{j=1}^n h_{1j} \overset{p}{\to} 0$ in probability.
For $b_{21n}$, equation (A.1) leads to

\[
\sqrt{n} b_{21n} = \frac{\sqrt{n}}{h_1^2 h_2 n^3} \sum_{i=1}^{n} \sum_{j \neq i, k \neq i, j} \frac{\delta_i \delta_j (X_i - X_j)}{f_0 (\theta_0 X_i, Y_i)} K' \left( \frac{\theta'_0 X_i - \theta'_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \psi(Y_k, \delta_k, Y_j -) \frac{1}{1 - G(Y_j -)} 1\{f_0 (\theta'_0 X_i) > c_n, Y_i \leq a_n, Y_j \leq a_n\} + o_p(1)
\]

Set

\[
N_{ijk} = \frac{1}{h_1^2 h_2} \frac{\delta_i \delta_j (X_i - X_j)}{f_0 (\theta_0 X_i, Y_i)} K' \left( \frac{\theta'_0 X_i - \theta'_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \psi(Y_k, \delta_k, Y_j -) \frac{1}{1 - G(Y_j -)} 1\{f_0 (\theta'_0 X_i) > c_n, Y_i \leq a_n, Y_j \leq a_n\}
\]

Since $E(\psi(Y_k, \delta_k, t)) = 0$, we have

\[
E(N_{ijk}) = E(N_{ijk} | Y_i, \delta_i, X_i) = 0.
\]

Set

\[
\bar{N}_k = E(N_{ijk} | Y_k, \delta_k).
\]

Hence

\[
\sqrt{n} b_{21n} = \frac{\sqrt{n}}{n^3} \sum_{i=1}^{n} \sum_{j \neq i, k \neq i, j} (N_{ijk} - \bar{N}_k) + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \bar{N}_k + o_p(1).
\]

Moreover, since $E(\psi(Y_k, \delta_k, t -)) = 0$ we have, for every component $m = 1, ..., d$,

\[
E \left( \frac{\sqrt{n}}{n^3} \sum_{i=1}^{n} \sum_{j \neq i, k \neq i, j} (N_{ijk} - \bar{N}_k) \right)^2 = n^{-2} E(N_{123} - \bar{N}_3)_m^2 + n^{-1} E(N_{123} - \bar{N}_3)_m (N_{423} - \bar{N}_3)_m + n^{-1} E(N_{123} - \bar{N}_3)_m (N_{143} - \bar{N}_3)_m + n^{-1} E(N_{123} - \bar{N}_3)_m (N_{124} - \bar{N}_4)_m.
\]

Furthermore, we can show that

\[
E(\psi^2(Y_k, \delta_k, t -)) \leq 4 \frac{1 - F(t -)}{(1 - H(t -))^2}.
\]

Moreover, since $\int \frac{F(dt)}{1 - G(t -)} < \infty$, there exists some constant $K_1$, so that

\[
1 - F(t -) \leq K_1 (1 - G(t -))
\]

and

\[
1 - F(t -) \leq \sqrt{K_1} \sqrt{1 - H(t -)}.
\]

Hence

\[
E(\psi^2(Y_k, \delta_k, t -)) \leq \frac{4 \sqrt{K_1}}{(1 - H(t -))^{3/2}}
\]

Additionally,

\[
(1 - H(a_n -))^{-3/2} = \left( \frac{n}{s_n} \right)^{3/2} = o(n^{1/2}).
\]
Finally, for every \( m = 1, \ldots, d \), we obtain
\[
n^{-2}E((N_{123} - \bar{N}_3)_{m}^2) = o_p(n^{-2}h_1^{-3}h_2^{-1}n^{1/2}) = o_p(1)
\]
\[
n^{-1}E((N_{123} - \bar{N}_3)_{m}(N_{423} - \bar{N}_3)_{m}) = o_p(n^{-1}h_1^{-2}n^{1/2}) = o_p(1)
\]
\[
n^{-1}E((N_{123} - \bar{N}_3)_{m}(N_{143} - \bar{N}_3)_{m}) = o_p(n^{-1}h_1^{-2}n^{1/2}) = o_p(1)
\]
and, since \( E(\psi(Y_k, \delta_k, t-)) = 0 \),
\[
n^{-1}E((N_{123} - \bar{N}_3)_{m}(N_{124} - \bar{N}_4)_{m}) = 0.
\]
Hence, as before, the first term goes to zero in probability. As to the second, set
\[
\bar{N}_k = \frac{1}{h_1^2h_2}\int_{\mathbb{R}^2} \frac{1}{f_0(x, z)}(E(X|\theta_0^r X = x) - u)K\left(\frac{x - \theta_0 u}{h_1}\right)K\left(\frac{z - v}{h_2}\right)\psi(Y_k, \delta_k, v-) \, dudvdzdv.
\]
Using Taylor expansions we can show that \( \bar{N}_k = \bar{N}_{1k} + n_{1k} \), where
\[
\bar{N}_{1k} = - \int \frac{d}{dt} \left[ (E(X|\theta_0^r X = t) - u)(1 - G_0(v - t))1_{\{f_{0|X}(t) > a_n\}} \right]_{t = \theta_0} \psi(Y_k, \delta_k, v-) \, dudvdv
\]
and \( n_{1k} = o_p(1) \). Finally, \( n_{1k} \) are i.i.d. random variables with expectation 0. Hence \( \frac{1}{\sqrt{n}} \sum_{k=1}^n n_{1k} \rightarrow 0 \) in probability. This completes the proof.

Now we state some similar lemmas.

**Lemma 8.** Under the conditions in Theorem 2 we have
\[
B_{3n} = \frac{1}{n} \sum_{j=1}^n \bar{H}_{3j} + \frac{1}{n} \sum_{k=1}^n \bar{N}_{3k} + o_p(n^{-1/2}),
\]
where
\[
\bar{H}_{3j} = \frac{\delta_j1(Y_j \leq a_n)}{1 - G(Y_j-)} \frac{d}{dt} \left[ (E(X|\theta_0^r X = t) - X_j)1_{\{f_{0|X}(t) > a_n\}} \right]_{t = \theta_0} X_j
\]
and
\[
\bar{N}_{3k} = \int \frac{d}{dt} \left[ (E(X|\theta_0^r X = t) - u)1_{\{f_{0|X}(t) > a_n\}} \right]_{t = \theta_0} \psi(Y_k, \delta_k, v-) \, dudvdv.
\]

**Lemma 9.** Under the conditions in Theorem 2 we have
\[
B_{5n} = \frac{1}{n} \sum_{j=1}^n \bar{H}_{5j} + \frac{1}{n} \sum_{k=1}^n \bar{N}_{5k} + o_p(n^{-1/2}).
\]
where

\[
\bar{H}_{5j} = -\frac{\delta_j(1_{\{Y_i \leq a_n\}})}{1 - G(Y_j - t)} \frac{d}{dt} \left( (E(X_i|\theta_0 X_i = t) - X_j)G_{\theta_0}(Y_j - |t|1_{f^\theta_{0,X}(t) > a_n})|_{t=\theta_0 X_j} \right) + E \left( \frac{\delta_j(1_{\{Y_i \leq a_n\}})}{1 - G(Y_j - t)} \frac{d}{dt} \left( (E(X_i|\theta_0 X_i = t) - X_j)G_{\theta_0}(Y_j - |t|1_{f^\theta_{0,X}(t) > a_n})|_{t=\theta_0 X_j} \right) \right),
\]

and

\[
\bar{N}_{5k} = - \int \frac{d}{dt} (E(X|\theta_0^r X = t) - u)G_{\theta_0}(v - |t|1_{f^\theta_{0,X}(t) > a_n})|_{t=\theta_0^r u} \psi(Y_k, \delta, v-)1_{\{v \leq a_n\}} f(u, v) du dv.
\]

**Lemma 10.** Under the conditions in Theorem 2 we have

\[
B_{2n} = \frac{1}{n} \sum_{j=1}^{n} \bar{H}_{2j} + \frac{1}{n} \sum_{k=1}^{n} \bar{N}_{2k} + o_P(n^{-1/2}).
\]

where

\[
\bar{H}_{2j} = E \left( \frac{\delta_j(1_{\{Y_i \leq a_n\}})}{1 - G(Y_j - t)} \frac{d}{dt} \left( (E(\nabla \log f_\theta X(\theta' X_i)_{\theta=\theta_0}|\theta_0^r X_i = \theta_0^r X_j) - 1 - G_{\theta_0}(Y_j - |t|1_{f^\theta_{0,X}(t) > a_n})1_{f^\theta_{0,X}(\theta_0 X_j) > a_n}) \right) \right)
\]

and

\[
\bar{N}_{2k} = - \int E(\nabla \log f_\theta(\theta' X_i)_{\theta=\theta_0}|\theta_0^r X_i = u)1_{f^\theta_{0,X}(u) > a_n} f_{\theta_0}(u, v) \psi(Y_k, \delta, v-)1_{\{v \leq a_n\}} f(u, v) du dv.
\]

**Proof.**
The proof is similar to the proof of Lemma 7. Observe only, that using Lemma 4, we have

\[
E(\nabla \log f_\theta(\theta' X_i, Y_i)_{\theta=\theta_0}|\theta_0^r X_i, Y_i) = E(\nabla \log f_\theta(\theta' X_i|Y_i)_{\theta=\theta_0}|\theta_0^r X_i, Y_i)
+ E(\nabla \theta \log f_\theta X(\theta' X_i)_{\theta=\theta_0}|\theta_0^r X_i) = E(\nabla \theta \log f_\theta X(\theta' X_i)_{\theta=\theta_0}|\theta_0^r X_i)
\]

**Lemma 11.** Under the conditions in Theorem 2 we have

\[
B_{4n} = \frac{1}{n} \sum_{j=1}^{n} \bar{H}_{4j} + \frac{1}{n} \sum_{k=1}^{n} \bar{N}_{4k} + o_P(n^{-1/2}).
\]

where

\[
\bar{H}_{4j} = \frac{\delta_j(1_{\{Y_i \leq a_n\}})}{1 - G(Y_j - t)} \frac{d}{dt} \left( (E(\nabla \log f_\theta X(\theta' X_i)_{\theta=\theta_0}|\theta_0^r X_i = \theta_0^r X_j)1_{f^\theta_{0,X}(\theta_0 X_j) > a_n} \right)
\]

and

\[
\bar{N}_{4k} = \int E(\nabla \log f_\theta X(\theta' X_i)_{\theta=\theta_0}|\theta_0^r X_i = u)\psi(Y_k, \delta, v-)1_{\{v \leq a_n\}} f_{\theta_0}(u, v)1_{f^\theta_{0,X}(u) > a_n} du dv.
\]
Lemma 12. Under the conditions in Theorem 2 we have

\[ B_{n} = \frac{1}{n} \sum_{j=1}^{n} H_{0j} + \frac{1}{n} \sum_{k=1}^{n} \bar{N}_{0k} + o_p(n^{-1/2}). \]

where

\[ H_{0j} = -G_{0\theta_{0}}(Y_j - |\theta_{0} X_j|) \frac{\delta_{j} 1(Y_j \leq a_n)}{1 - G(Y_j -)} E(\nabla_{\theta} \log f_{\theta^j X_i}(\theta^j X_i)_{\theta = \theta_0} |\theta_{0} X_i = \theta_{0} X_j) 1 \{ f_{\hat{a}_{0} X_i}(\theta_{0} X_j) > c_n \} + E \left( G_{0\theta_{0}}(Y_j - |\theta_{0} X_j|) \frac{\delta_{j} 1(Y_j \leq a_n)}{1 - G(Y_j -)} E(\nabla_{\theta} \log f_{\theta^j X_i}(\theta^j X_i)_{\theta = \theta_0} |\theta_{0} X_i = \theta_{0} X_j) 1 \{ f_{\hat{a}_{0} X_i}(\theta_{0} X_j) > c_n \} \right). \]

and

\[ \bar{N}_{0k} = - \int E(\nabla_{\theta} \log f_{\theta^j X_i}(\theta^j X_i)_{\theta = \theta_0} |\theta_{0} X_i = u) G_{0\theta_{0}}(v - |u|) \psi(Y_k, \hat{\delta}_k, v -) G_{0\theta_{0}}(v - |u|) f_{\theta_{0}}(u, v) 1 \{ f_{\hat{a}_{0} X_i}(u) > c_n, v \leq a_n \} dxdudv. \]

Lemma 13. Under the conditions in Theorem 2 we have

\[ B_{1n} + B_{3n} + B_{5n} + B_{2n} + B_{4n} + B_{0n} = o_p(n^{-1/2}). \]

Proof. Using Lemmas 7-12, we can show that

\[ H_{1j} + H_{3j} + H_{5j} = H_{2j} + H_{4j} + H_{0j} = 0 \]

and

\[ \bar{N}_{1k} + \bar{N}_{3k} + \bar{N}_{5k} = \bar{N}_{2k} + \bar{N}_{4k} + \bar{N}_{6k} = 0. \]

Lemma 14. Under the conditions in Theorem 2 we have

\[ A_{7n} + A_{8n} + A_{9n} + A_{10n} + A_{11n} + A_{12n} = o_p(n^{-1/2}). \]

Proof. See Appendix Appendix B.

Lemma 15. Under the conditions in Theorem 2 we have

\[ \hat{l}^{[2]}(\theta) \rightarrow l^{[2]}(\theta) \]

Proof. We have

\[ \hat{l}^{[2]}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_{i} \left( \frac{\hat{r}^{[2]}(\theta^j X_i, Y_i)}{\hat{r}(\theta^j X_i, Y_i)} \frac{\hat{v}^{[1]}(\theta^j X_i, Y_i)(\hat{v}^{[1]}(\theta^j X_i, Y_i))^{t}}{\hat{v}^{2}(\theta^j X_i, Y_i)} - \frac{\hat{s}^{[2]}(\theta^j X_i)}{\hat{s}(\theta^j X_i)} + \frac{\hat{s}^{[1]}(\theta^j X_i)(\hat{s}^{[1]}(\theta^j X_i))^{t}}{\hat{s}^{2}(\theta^j X_i)} \right) \right] (Y_i \leq a_n, X_i \in A^{*n}) \]

Recall,
\[ \hat{r}^{[2]}(y, \theta' x) = \frac{1}{nh_1^2h_2} \sum_{j=1}^{n} 1_{(Y_j \leq a_n)} \frac{\delta_j}{1 - G(Y_j -)} (x - X_j)(x - X_j)^t K'' \left( \frac{\theta' x - \theta' X_j}{h_1} \right) K \left( \frac{y - Y_j}{h_2} \right). \]

Since \(a_n \to \tau_H\) and \(c_n \to 0\), it is easy to show, that

\[ \hat{r}^{[2]}(\theta' x, y) \to \frac{d^2}{dt^2} \left\{ f_{\theta}(t, y) E((x - X)(x - X)^t|\theta' X = t) \right\}_{t=\theta' x} \text{ in probability.} \]

Moreover, using Lemma 3, it can be proved that

\[ \frac{d^2}{dt^2} \left\{ f_{\theta}(y, t) E((x - X)(x - X)^t|\theta' X = t) \right\}_{t=\theta' x} = f_{\theta}^{[2]}(\theta' x, y). \]

Similarly, it can be shown that \(\hat{s}^{[2]}(\theta' x) \to f_{\theta}^{[2]}(\theta' x)\) and \(\hat{d}^2(\theta' x, y) \to d_{\theta}^{[2]}(\theta' x, y)\). This completes the proof.

**Proof of Theorem 4.**

In view of Lemma 2, It remains to show that

\[ \sum_{k=1}^{12} A_{k\theta} \to 0. \]

For this, note that for every \(\varepsilon > 0\) there exist a small \(c\) and a sequence \(b_n \to \tau_H\) so that

\[ 1 - H(b_n) = \frac{c}{n} \text{ and } \mathbb{P}(Y_i \leq b_n \text{ for } i = 1, ..., n) \geq 1 - \varepsilon. \]

Additionally, observe that the remainder \(R_n(Y_j)\) in Lemma 6 can be written as

\[ R_n(Y_j) = R_n(Y_j)1_{(Y_j \leq a_n - \varepsilon)} + R_n(Y_j)1_{(Y_j > a_n - \varepsilon)} \]

with \(\approx n^{1/6} (\log(n))^{2+\varepsilon/2}\). Hence

\[ R_n(Y_j) = o_p(n^{-1/6}) + O_p(1)1_{(Y_j > a_n - \varepsilon)}. \]  

(A.5)

The proof goes similarly to proofs of Lemmas 7-13 with \(a_n\) replaced by \(b_n\) and using (A.5) when necessary.

**Proof of Theorem 3.**

According to Proposition A.9 from [17] we may replace \(A^{\mu} = A^{\mu}_{a_n} = \{x : f_{\theta}(x, \theta') > c_n\}\) by \(A^{\mu} = A^{\mu}_{a_n} = \{x : f_{\theta}(x, \theta') > c_n\}\). The rest of the proof is similar to the proof of Theorem 2 but it is much longer. Since the proof is similar to the proof of Theorem 2 but it is much longer, we will present only parts of it in Appendix C.

**Appendix B. Proof of Lemma 14.**

Let first consider the terms in \(A_{7\theta}\) and \(A_{8n}\). Recall

\[ A_{7\theta} = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\hat{r}^{[1]}(\theta_0' X_i, Y_i) - \hat{r}(\theta_0' X_i, Y_i)}{f_{\theta_0}(\theta_0' X_i, Y_i)} \right] 1_{(X_i \in A^{\mu}, Y_i \leq a_n)} \]
and

\[ A_{kn} = \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_i \nabla_\theta \log(\tilde{r}(\theta' X_i, Y_i))|_{\theta = \theta_0} \frac{(f_{\theta_0}(\theta_0' X_i, Y_i) - \tilde{r}(\theta_0' X_i, Y_i))^2}{f_{\theta_0}'(\theta_0' X_i, Y_i)} \right] 1_{\{X_i \in A^{kn}, Y_i \leq a_n\}}. \]

Set

\[ M_n(X_i, Y_i) = \frac{\tilde{r}^{[1]}(\theta_0' X_i, Y_i) - \tilde{r}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}'(\theta_0' X_i, Y_i)} 1_{\{X_i \in A^{kn}\}} \]

\[ N_n(X_i, Y_i) = \frac{f_{\theta_0}(\theta_0' X_i, Y_i) - \tilde{r}(\theta_0' X_i, Y_i)}{f_{\theta_0}'(\theta_0' X_i, Y_i)} 1_{\{X_i \in A^{kn}, Y_i \leq a_n\}}. \]

Hence

\[ A_{kn} = \frac{1}{n} \sum_{i=1}^{n} \delta_i M_n(X_i, Y_i) N_n(X_i, Y_i) \]

\[ A_{kn} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \nabla_\theta \log(\tilde{r}(\theta' X_i, Y_i))|_{\theta = \theta_0} N_n^2(X_i, Y_i) \]

Moreover,

\[ M_n(X_i, Y_i) = M_{1n}(X_i, Y_i) + M_{2n}(X_i, Y_i), \]

where

\[ M_{1n}(X_i, Y_i) = \frac{\tilde{r}^{[1]}(\theta_0' X_i, Y_i) - \tilde{r}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}'(\theta_0' X_i, Y_i)} 1_{\{X_i \in A^{kn}, Y_i \leq a_n\}} \]

\[ M_{2n}(X_i, Y_i) = \frac{\tilde{r}^{[1]}(\theta_0' X_i, Y_i) - \tilde{r}^{[1]}(\theta_0' X_i, Y_i)}{f_{\theta_0}'(\theta_0' X_i, Y_i)} 1_{\{X_i \in A^{kn}, Y_i \leq a_n\}}. \]

Further \( \sup_{x,y} |M_{2n}(x,y)| = O_p(h_1^2 + h_2^2) \). As to \( M_{1n} \), recall

\[ M_{1n}(X_i, Y_i) = \frac{1}{f_{\theta_0}'(\theta_0' X_i, Y_i)} \left[ \frac{1}{h_1^2 h_2 n} \sum_{j \neq i} \delta_j \frac{1(Y_i \leq a_n)}{1 - G_n(Y_j - h_1)} (X_i - X_j) K' \left( \frac{\theta_0' X_i - \theta_0' X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \right. \]

\[ \left. - \frac{1}{h_1^2 h_2} (X_i - u) K' \left( \frac{\theta_0' X_i - \theta_0' u}{h_1} \right) K \left( \frac{Y_i - v}{h_2} \right) f(u, v) 1\{v \leq a_n\} dudv \right] \]

Hence

\[ M_{1n}(X_i, Y_i) = M_{11n}(X_i, Y_i) + M_{12n}(X_i, Y_i), \]

where

\[ 35 \]
\[ M_{11n}(X_i, Y_i) = \frac{1 \{ f_{\theta_0 X_i}(\theta_0 X_i) > c_n \}}{f_{\theta_0 X_i}(\theta_0 X_i, Y_i)} \int \frac{1}{h_1 h_2 n} \sum_{i \neq j} \delta_j 1_{\{ Y_j \leq a_n \}} (X_i - X_j) K' \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \]

\[ M_{12n}(X_i, Y_i) = \frac{1}{h_1 h_2 n} \sum_{i \neq j} \delta_j (X_i - X_j) K' \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \]

\[ G_n(Y_j -) - G(Y_j -) \]

\[ \sup_{t \leq Y_{r,n}} \left| \frac{G_n(t) - G(t)}{1 - G(t)} \right| = O(n^{-1/3}). \]

Hence, by choosing \( r_n = n - [s_n/5] + 1 \) and since \( a_n \leq Y_{r,n} \) with high probability, we have

\[ \sup_{x,y} |M_{11n}(x, y)| = O_P \left( h_1^{-1} n^{-1/3} \right). \]

and using Proposition 6 in [1],

\[ \sup_{x,y} |M_{11n}(x, y)| = O_P \left( n^{-1/2} h_1^{-3/2} h_2^{-1/2} (\log(n))^{1/2} \right). \]

For \( N_n \), we have

\[ N_n(X_i, Y_i) = N_{1n}(X_i, Y_i) + N_{2n}(X_i, Y_i), \]

where

\[ N_{1n}(X_i, Y_i) = \frac{\bar{r}(\theta_0 X_i, Y_i) - \bar{r}(\theta_0 X_i, Y_i)}{f_{\theta_0 X_i}(\theta_0 X_i, Y_i)} 1_{\{ X_i \in A^{c_n} \}} \]

\[ N_{2n}(X_i, Y_i) = \frac{f_{\theta_0 X_i}(\theta_0 X_i, Y_i) - \bar{r}(\theta_0 X_i, Y_i)}{f_{\theta_0 X_i}(\theta_0 X_i, Y_i)} 1_{\{ X_i \in A^{c_n} \}}. \]

Furthermore, \( \sup_{x,y} |N_{2n}(x, y)| = O_P \left( h_1^2 + h_2^2 \right) \). As to \( N_{1n} \), recall

\[ N_{1n}(X_i, Y_i) = \frac{\delta_i}{f_{\theta_0 X_i}(\theta_0 X_i, Y_i)} \int \frac{1}{h_1 h_2} K' \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) f(u, v) 1_{\{ v \leq a_n \}} du dv \]

\[ - \frac{1}{h_1 h_2 n} \sum_{i \neq j} \delta_j 1_{\{ Y_j \leq a_n \}} K' \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) 1_{\{ f_{\theta_0 X_i}(\theta_0 X_i) > c_n, Y_i \leq a_n \}} \]

Hence

\[ N_{1n}(X_i, Y_i) = N_{11n}(X_i, Y_i) + N_{12n}(X_i, Y_i), \]

\[ 36 \]
where

$$N_{11n}(X_i, Y_i) = \frac{1}{f_{0n}} \left[ \frac{1}{h_1 h_2} \int K \left( \frac{\theta_0 X_i - \theta_0 u}{h_1} \right) K \left( \frac{Y_i - u}{h_2} \right) f(u, v) 1_{\{v \leq a_n\}} \, du \, dv \right. \right.$$

$$\left. - \frac{1}{h_1 h_2 n} \sum_{j \neq i} \delta_j 1_{\{Y_j \leq a_n\}} \frac{\theta_0 X_i - \theta_0 X_j}{h_1} K \left( \frac{Y_i - Y_j}{h_2} \right) \right] 1_{\{f_{0n} > a_n, Y_i \leq a_n\}}$$

$$N_{12n}(X_i, Y_i) = \frac{1}{f_{0n}} \left[ \frac{1}{h_1 h_2 n} \sum_{j \neq i} \delta_j 1_{\{f_{0n} > a_n, Y_i \leq a_n\}} \frac{\theta_0 X_i - \theta_0 X_j}{h_1} K \left( \frac{Y_i - Y_j}{h_2} \right) \right. \right.$$

$$\left. \frac{G_n(Y_j) - G(Y_j)}{(1 - G(Y_j))(1 - G_n(Y_j))} 1_{\{Y_j \leq a_n\}} \right]$$

Moreover, using Proposition 6 in [1],

$$\sup_{x,y} |N_{11n}(x, y)| = O_p(n^{-1/2} h_1^{-1/2} h_2^{-1/2} (\log(n))^{1/2}).$$

and, by (A.3),

$$\sup_{x,y} |N_{12n}(x, y)| = O_p(n^{-1/3}).$$

Finally, it can be shown that

$$\sqrt{n} A_{8n} \overset{p}{\rightarrow} 0.$$

For $\sqrt{n} A_{17n}$ we need to show that

$$\sqrt{n}^{-1} \sum_{i=1}^{n} \delta_i M_{11n}(X_i, Y_i) N_{12n}(X_i, Y_i) \overset{p}{\rightarrow} 0 \quad \text{(B.1)}$$

$$\sqrt{n}^{-1} \sum_{i=1}^{n} \delta_i M_{12n}(X_i, Y_i) N_{11n}(X_i, Y_i) \overset{p}{\rightarrow} 0 \quad \text{(B.2)}$$

$$\sqrt{n}^{-1} \sum_{i=1}^{n} \delta_i M_{11n}(X_i, Y_i) N_{11n}(X_i, Y_i) \overset{p}{\rightarrow} 0 \quad \text{(B.3)}$$

$$\sqrt{n}^{-1} \sum_{i=1}^{n} \delta_i M_{11n}(X_i, Y_i) N_{2n}(X_i, Y_i) \overset{p}{\rightarrow} 0 \quad \text{(B.4)}$$

and

$$\sqrt{n}^{-1} \sum_{i=1}^{n} \delta_i M_{12n}(X_i, Y_i) N_{2n}(X_i, Y_i) \overset{p}{\rightarrow} 0 \quad \text{(B.5)}$$

As to (B.1) and (B.2), we choose $b_n = H^{-1}(1 - \frac{\varepsilon_1}{2})$ with $\varepsilon_1 \approx n^{5/6} (\log(n))^{2+\varepsilon_1}$ and write $1_{\{Y_j \leq a_n\}} = 1_{\{Y_j \leq b_n\}} + 1_{\{b_n < Y_j \leq a_n\}}$. The proof follows from (A.3) and uniform bounds for $M_{11n}$ and $N_{11n}$.  

37
For (B.3), set
\[
\psi_1(X_i, Y_i, X_j, Y_j, \delta_j) = \int K\left(\frac{\theta'_0 X_i - \theta'_0 u}{h_1}\right) K\left(\frac{Y_i - u}{h_2}\right) f(u, v) 1_{\{v \leq a_n\}} du dv
- \frac{\delta_j 1\{Y_j \leq a_n\}}{1 - G(Y_j)} K\left(\frac{\theta'_0 X_j - \theta'_0 X_j}{h_1}\right) K\left(\frac{Y_j - Y_j}{h_2}\right)
\]
and
\[
\psi_2(X_i, Y_i, X_j, Y_j, \delta_j) = \int (X_i - u) K'\left(\frac{\theta'_0 X_i - \theta'_0 u}{h_1}\right) K\left(\frac{Y_i - u}{h_2}\right) f(u, v) 1_{\{v \leq a_n\}} du dv
- \frac{\delta_j 1\{Y_j \leq a_n\}}{1 - G(Y_j)} (X_i - X_j) K'\left(\frac{\theta'_0 X_j - \theta'_0 X_j}{h_1}\right) K\left(\frac{Y_j - Y_j}{h_2}\right)
\]
Then the term (B.3) is equal to
\[
\frac{\sqrt{n}}{n^3 h_1^3 h_2^3} \sum_{i = 1}^{n} \sum_{j \neq i} \sum_{k \neq i} \frac{\delta_i}{f^{\delta}_{0}(\theta'_0 X_i, Y_i)} \psi_1(X_i, Y_i, X_j, Y_j, \delta_j) \psi_2(X_i, Y_i, X_k, Y_k, \delta_k).
\]
Moreover, for \(m = 1, 2\)
\[
E(\psi_m(X_i, Y_i, X_j, Y_j, \delta_j) | X_i, Y_i) = 0
\]
and
\[
E[\psi_m(X_i, Y_i, X_j, Y_j, \delta_j)] = O(h_1 h_2).
\]
Hence (B.3) is fulfilled. For (B.4) and (B.5), observe that
\[
N_{2n}(X_i, Y_i) = \left[ - h_i^2 \frac{1}{2} \frac{d^2}{dy^2} f_0(\theta' X_i, y)|_{y = Y_i} - h_i \frac{1}{2} \frac{d^2}{dx^2} f_0(x, Y_i)|_{x = \theta' X_i} \right] \int z^2 K(z) dz + O(h_i^4 + h_i^4).
\]
Hence, it remains to prove that
\[
\frac{\sqrt{n}}{n^3 h_1^3 h_2} \sum_{i = 1}^{n} \sum_{j \neq i} \sum_{k \neq i} \frac{\delta_i}{f^{\delta}_{0}(\theta'_0 X_i, Y_i)} \left[ - h_i^2 \frac{1}{2} \frac{d^2}{dy^2} f_0(\theta' X_i, y)|_{y = Y_i} - h_i \frac{1}{2} \frac{d^2}{dx^2} f_0(x, Y_i)|_{x = \theta' X_i} \right] \xrightarrow{p} 0
\]
and, using Lemma 6, that
\[
\frac{\sqrt{n}}{n^3 h_1^3 h_2} \sum_{i = 1}^{n} \sum_{j \neq i} \sum_{k \neq i} \sum_{k \neq i} \frac{\delta_i}{f^{\delta}_{0}(\theta'_0 X_i, Y_i)} K'\left(\frac{\theta'_0 X_i - \theta'_0 X_j}{h_1}\right) K\left(\frac{Y_i - Y_j}{h_2}\right) \psi(Y_k, \delta_k, Y_j)
\sum_{k \neq i} 1\{f^{\delta}_{0}(\theta'_0 X_k, Y_k) > c_n\} \left[ - h_i^2 \frac{1}{2} \frac{d^2}{dy^2} f_0(\theta' X_i, y)|_{y = Y_i} - h_i \frac{1}{2} \frac{d^2}{dx^2} f_0(x, Y_i)|_{x = \theta' X_i} \right] \xrightarrow{p} 0.
\]
The proof is similar to that of Lemma 10, with the difference that we have and additional term which goes to zero when \(n \rightarrow \infty\).

The remaining terms \(A_{kn}\), for \(k = 9, \ldots, 12\), can be handled similarly.
Appendix C. Proof of Theorem 3.

The proof follows the lines of that of Theorem 2. However the deterministic sequence \( a_n \) has to replaced by \( a_n = Y_{r_n:n} \). For this reason only three chosen auxiliary results are shown here.

**Lemma 16.** Under the conditions in Theorem 3 we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \delta_i \nabla_{\theta} \log f_\theta(Y_i|\theta'X_i)_{\theta=\theta_0} + (1 - \delta_i) \nabla_{\theta} \log (1 - F_\theta(Y_i|\theta'X_i))_{\theta=\theta_0} \right] 1\{Y_i > Y_{r_n:n}, X_i \in A' \} = o_p(1)
\]

**Proof.**

Set

\[
A_i = \delta_i \nabla_{\theta} \log f_\theta(Y_i|\theta'X_i)_{\theta=\theta_0} 1\{Y_i > Y_{r_n:n}, X_i \in A' \}.
\]

We will show that, for every component \( m = 1, \ldots, d \) of the vector \( A_i \in \mathbb{R}^d \), we have

\[
E\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \right)_m^2 \to 0.
\]

For this, let us consider

\[
E\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \right)_m^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E((A_i, A_j)_m) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left( \delta_i \delta_j \frac{\partial}{\partial \theta_m} \log f_\theta(Y_i|\theta'X_i)_{\theta=\theta_0} \right)
\]

\[
\frac{\partial}{\partial \theta_m} \log f_\theta(Y_j|\theta'X_j)_{\theta=\theta_0} 1\{H_n(Y_i) > \frac{r_n}{n}, H_n(Y_j) > \frac{r_n}{n} \} \{X_i, X_j \in A' \}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} E\left( \delta_i \delta_j \frac{\partial}{\partial \theta_m} \log f_\theta(Y_i|\theta'X_i)_{\theta=\theta_0} \right)
\]

\[
\frac{\partial}{\partial \theta_m} \log f_\theta(Y_j|\theta'X_j)_{\theta=\theta_0} 1\{H_n(Y_i) > \frac{r_n}{n}, H_n(Y_j) > \frac{r_n}{n} \} \{X_i, X_j \in A' \} + O(n^{-1/3}(\log(n))^{2+\varepsilon})
\]

Further, setting \( H_n^{-1}(v) = \frac{1}{n} \sum_{b \neq i} 1\{Y_i \leq v\} \), we have

\[
1\{H_n(Y_i) > \frac{r_n}{n}, H_n(Y_j) > \frac{r_n}{n}\} = 1\{H_n^{-1}(Y_i) > \frac{r_n}{n}, H_n^{-1}(Y_j) > \frac{r_n}{n}\} = 1\{H_n^{-1}(Y_i) > \frac{r_n}{n}, H_n^{-1}(Y_j) > \frac{r_n}{n}\}
\]

\[
+ 1\{H_n^{-1}(Y_i) > \frac{r_n}{n}, H_n^{-1}(Y_j) = \frac{r_n}{n}, Y_i < Y_j\}
\]

\[
+ 1\{H_n^{-1}(Y_i) = \frac{r_n}{n}, H_n^{-1}(Y_j) > \frac{r_n}{n}, Y_i \leq Y_j\}
\]

\[
= 1\{H_n^{-1}(Y_i) > \frac{r_n}{n}, H_n^{-1}(Y_j) > \frac{r_n}{n}\}
\]

Additionally, using Lemma 3 and since \( 1\{X_i \in A' \} = 1\{f_{0,X}(0_0, X_i) < c_n\} \) is known given \( \theta'_0, X_i \), we have

\[
E\left( \delta_i \nabla_{\theta} \log f_\theta(Y_i|\theta'X_i)_{\theta=\theta_0} 1\{H_n^{-1}(Y_i) > \frac{r_n}{n}\} C_i, Z_i, \theta'_0, X_i, Y_s \text{ for } s \neq i \right) = 0.
\]

This completes the proof.

Next we will deal with a version of Lemma 7, but for the choice \( a_n = Y_{r_n:n} \).
Lemma 17. Under the conditions in Theorem 3 we have

\[ B_{1n} = \frac{1}{n} \sum_{j=2}^{n} H_{1j} + \frac{1}{n} \sum_{k=3}^{n} N_{1k} + o_{\P}(n^{-1/2}), \]

where

\[ H_{1j} = -\frac{\delta_j 1\{H^{-1}(y_j) \leq \frac{c_n-2}{n}\}}{1 - G(Y_j)} \frac{d}{dt}(E(X_i | \theta_0 X_i = t) - X_j)(1 - \varphi_0(Y_j - |t|)1_{(f_{\theta_0,x}(t) > c_n)})|_{t=\theta_0 X_j} \]

\[ + E\left( \frac{\delta_j 1\{H^{-1}(y_j) \leq \frac{c_n-2}{n}\}}{1 - G(Y_j)} \frac{d}{dt}(E(X_i | \theta_0 X_i = t) - X_j)(1 - \varphi_0(Y_j - |t|)1_{(f_{\theta_0,x}(t) > c_n)})|_{t=\theta_0 X_j}| s \neq j \right), \]

and

\[ N_{1k} = -\int \frac{d}{dt}[(E(X|\theta_0 X = t) - u)(1 - \varphi_0(v - |t|)1_{(f_{\theta_0,x}(t) > c_n)})|_{t=\theta_0 u}] \frac{\psi(Y_k, \delta_k, v-)}{1 - G(v-)} \]

\[ 1\{H^{-1}(2,y) \leq \frac{c_n-1}{n}\}f(u,v)du dv. \]

Proof. As before

\[ B_{1n} = b_{1n} + b_{2n} + b_{3n}, \]

where

\[ b_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{f_{\theta_0}(\theta_0 X_i, Y_i)} \left[ \frac{1}{\sigma_i h_2 n} \sum_{j \neq i} \delta_j 1\{Y_j \leq Y_{\text{r.m.}}\} (X_j - X_i) K^\prime \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_j - Y_i}{h_2} \right) \right] \]

\[ - \frac{1}{\sigma_i h_2 n^2} \int (X_i - u) K^\prime \left( \frac{\theta_0 X_i - \theta_0 u}{h_1} \right) K \left( \frac{Y_i - u}{h_2} \right) \frac{1_{1|c_n, Y_i \leq Y_{\text{r.m.}}}|}{1 - G(v-)} H_1(du, dv) 1_{(f_{\theta_0,x}(\theta_0 X_i) > c_n, Y_i \leq Y_{\text{r.m.}})} \]

\[ b_{2n} = \frac{1}{\sigma_i h_2 n^2} \sum_{i=1}^{n} \delta_i \delta_j \frac{1}{f_{\theta_0}(\theta_0 X_i, Y_i)} \left[ K^\prime \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \frac{G_n(Y_j - ) - G(Y_j - )}{(1 - G(Y_j - ))^2} \right] \]

\[ 1_{(f_{\theta_0,x}(\theta_0 X_i) > c_n, Y_i \leq Y_{\text{r.m.}})} \]

\[ b_{3n} = \frac{1}{\sigma_i h_2 n^2} \sum_{i=1}^{n} \delta_i \delta_j \frac{1}{f_{\theta_0}(\theta_0 X_i, Y_i)} \left[ K^\prime \left( \frac{\theta_0 X_i - \theta_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \frac{(G_n(Y_j - ) - G(Y_j - ))^2}{(1 - G(Y_j - ))^2(1 - G_n(Y_j - ))} \right] \]

\[ 1_{(f_{\theta_0,x}(\theta_0 X_i) > c_n, Y_i \leq Y_{\text{r.m.}})} \]

We have

\[ \sqrt{n} h_{1n} := \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \sum_{j \neq i} H_{1ij}, \]

where
Moreover, and additionally, we define
\[ H_{ij} = H(\theta_0', X_i, Y_i, \delta_i, \theta_0', X_j, Y_j, \delta_j) \]
\[ = \frac{1}{h_1^2 h_2} \frac{\delta_i}{f_{X}(\theta_0^0, X_i)} \left[ \frac{\delta_j}{1 - G(Y_j - v)} (X_i - X_j) K' \left( \frac{\theta_0^0 X_i - \theta_0^0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \right] \left\{ Y_i \leq Y_{r \alpha, n} \right\} \]
\[ - \int (X_i - u) K' \left( \frac{\theta_0^0 X_i - \theta_0^0 u}{h_1} \right) K \left( \frac{Y_i - v}{h_2} \right) \left\{ v \leq Y_{r \alpha, n} \right\} \left[ H_1(du) dv \right] \]
\[ \left\{ f_{X}(\theta_0^0, X_i) > 0, Y_i \leq Y_{r \alpha, n} \right\}. \]

Additionally, we define
\[ H_{ij}^* = \frac{1}{h_1^2 h_2} \frac{\delta_i}{f_{X}(\theta_0^0, X_i)} \left[ \frac{\delta_j}{1 - G(Y_j - v)} (X_i - X_j) K' \left( \frac{\theta_0^0 X_i - \theta_0^0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \right] \left\{ H_n^{(i,j)}(Y_j) \leq \frac{\epsilon_{m - 2}}{n} \right\} \]
\[ - \int (X_i - u) K' \left( \frac{\theta_0^0 X_i - \theta_0^0 u}{h_1} \right) K \left( \frac{Y_i - v}{h_2} \right) \left\{ H_n^{(i,j)}(v) \leq \frac{\epsilon_{m - 2}}{n} \right\} \left[ H_1(du) dv \right] \left\{ H_n^{(i,j)}(Y_j) \leq \frac{\epsilon_{m - 2}}{n} \right\} . \]

As first we show that
\[ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} (H_{ij} - H_{ij}^*) = o_P(1). \] (C.1)

For this we consider following indicators
\[ 1\{Y_i \leq Y_{r \alpha, n}\} 1\{Y_j \leq Y_{r \alpha, n}\} = 1\{Y_i \leq Y_j\} 1\{H_n^{(i,j)}(Y_j) \leq \frac{\epsilon_{m - 2}}{n} \} \]
\[ + 1\{Y_i > Y_j\} 1\{H_n^{(i,j)}(Y_j) \leq \frac{\epsilon_{m - 2}}{n} \} \]
\[ = 1\{H_n^{(i,j)}(Y_j) \leq \frac{\epsilon_{m - 2}}{n} \} \]
\[ 1\{H_n^{(i,j)}(Y_j) \leq \frac{\epsilon_{m - 2}}{n} \} . \]

Moreover,
\[ 1\{H_n^{(i,j)}(v) \leq \frac{\epsilon_{m - 2}}{n} \} = 1\{Y_i \leq v\} 1\{H_n^{(i,j)}(Y_i) \leq \frac{\epsilon_{m - 2}}{n} \} + 1\{Y_i > v\} 1\{H_n^{(i,j)}(Y_i) \leq \frac{\epsilon_{m - 2}}{n} \}
\[ = 1\{Y_i \leq v\} 1\{H_n^{(i,j)}(Y_i) \leq \frac{\epsilon_{m - 2}}{n} \} + 1\{Y_i > v\} 1\{H_n^{(i,j)}(Y_i) \leq \frac{\epsilon_{m - 2}}{n} \}
\[ + 1\{Y_i \leq v\} 1\{H_n^{(i,j)}(v) \leq \frac{\epsilon_{m - 2}}{n} \} + 1\{Y_i > v\} 1\{H_n^{(i,j)}(v) \leq \frac{\epsilon_{m - 2}}{n} \}
\[ = 1\{H_n^{(i,j)}(v) \leq \frac{\epsilon_{m - 2}}{n} \} \]
Observe that
\[ 1_{\{Y_i < Y_j \leq v\}} 1_{\{H_n^{-1}(\cdot) = \frac{2n-1}{n}, H_n^{-1}(Y_i) = \frac{2n-1}{n}\}} = 0 \]
\[ 1_{\{Y_i \leq v \leq Y_j\}} 1_{\{H_n^{-1}(\cdot) = \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} = 0 \]
\[ 1_{\{Y_j > Y_i \}} 1_{\{H_n^{-1}(\cdot) \leq \frac{2n-1}{n}, H_n^{-1}(Y_i) = \frac{2n-1}{n}\}} = 0 \]
\[ 1_{\{Y_j > Y_i \}} 1_{\{H_n^{-1}(\cdot) \leq \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} = 0 \]
and
\[ 1_{\{Y_i > v \leq Y_j\}} 1_{\{H_n^{-1}(\cdot) = \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} = 0. \]

Hence
\[ 1_{\{H_n^{-1}(\cdot) \leq \frac{2n-1}{n}, H_n^{-1}(Y_i) = \frac{2n-1}{n}\}} = 1_{\{H_n^{-1}(\cdot) \leq \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} + 1_{\{Y_i \leq v \leq Y_j\}} 1_{\{H_n^{-1}(\cdot) = \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} + 1_{\{Y_j > Y_i \}} 1_{\{H_n^{-1}(\cdot) \leq \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} \]

Finally, we obtain
\[ H_{ij} = H_{ij}^* + T_{ij}, \]

where
\[ T_{ij} = -\frac{1}{h_1^2 h_2} \frac{1}{f_{\theta_0}(\theta_0' x_i, Y_i)} \int (X_i - u) K' \left( \frac{\theta_0' X_i - \theta_0' u}{h_1} \right) K \left( \frac{Y_i - v}{h_2} \right) \]
\[ \left( 1_{\{Y_i \leq v \leq Y_j\}} 1_{\{H_n^{-1}(\cdot) = \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} \right) \frac{1}{1 - G(v-)} H_1(du, dv). \]

Hence, the quantity in (C.1) equals
\[ \frac{\sqrt{n}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} T_{ij} = T_{1n} + O_P \left( \frac{n^{1/2}}{h_1^2 h_2} \frac{n - r_n}{n^2} \right), \]

where
\[ T_{1n} = \frac{1}{h_1^2 h_2} \frac{\sqrt{n}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{f_{\theta_0}(\theta_0' x_i, Y_i)} \int (X_i - u) K' \left( \frac{\theta_0' X_i - \theta_0' u}{h_1} \right) K \left( \frac{Y_i - v}{h_2} \right) \]
\[ 1_{\{Y_i \leq v \leq Y_j\}} 1_{\{H_n^{-1}(\cdot) = \frac{2n-2}{n}, H_n^{-1}(Y_i) = \frac{2n-2}{n}\}} \frac{1}{1 - G(v-)} H_1(du, dv) \]

and, since \( n - r_n \approx n^{2/3}(\log(n))^{2+\epsilon} \).
Further, we can rewrite the $T_{1n}$ as follow

$$T_{1n} = \frac{1}{h^2_1 h^2_2} \sqrt{n} \sum_{i=1}^{n} \sum_{j \neq i} \frac{\delta_i f_{\theta_0} (\theta_0, X_i) > c_n}{f_{\theta_0} (\theta_0, X_i, Y_i)} \int \frac{d}{dt} \left[ (X_i - E(X|\theta_0, X = t)) f_{\theta_0}(t, Y_i) \right]_{t=\theta_0} K(z_2)$$

Moreover, by change of variables and Taylor expansion, we obtain

$$T_{1n} = \frac{1}{h^2_1 h^2_2} \sqrt{n} \sum_{i=1}^{n} \sum_{j \neq i} \frac{\delta_i f_{\theta_0} (\theta_0, X_i) > c_n}{f_{\theta_0} (\theta_0, X_i, Y_i)} \int \frac{d}{dt} \left[ (X_i - E(X|\theta_0, X = t)) f_{\theta_0}(t, Y_i) \right]_{t=\theta_0} K(z_2)$$

Further, if $supp(K) \in [-a, a]$ and since $Y_j \geq Y_{r_{1,n}}$, we can show that

$$\frac{\sqrt{n}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} \delta_i f_{\theta_0} (\theta_0, X_i) > c_n \int \frac{d}{dt} \left[ (X_i - E(X|\theta_0, X = t)) f_{\theta_0}(t, Y_i) \right]_{t=\theta_0} K(z_2)$$

Next, set

$$\hat{H}^*_j = E(\hat{H}^*_{ij} | Y_j, \delta_j, X_j, Y_s \text{ for } s \neq i).$$

Hence

$$\sqrt{n} \hat{b}_{1n} := \frac{\sqrt{n}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} (H^*_{ij} - \hat{H}^*_j) + \frac{1}{\sqrt{n}} \sum_{j=2}^{n} \hat{H}^*_j + o_p(1).$$

As for the second part, we have
\[
\frac{1}{\sqrt{n}} \sum_{j=2}^{n} H^*_j = \frac{1}{h_1^2 \sqrt{n}} \sum_{j=2}^{n} \int \left(1 - G(y - \{x\}) f_{\theta_0}(x, y) \right) \frac{\delta_j}{1 - G(Y_j -)} (E(X|\theta'_0, X = x) - X_j) \\
K'(\frac{x - \theta_0}{h_1}) K\left(\frac{y - Y_j}{h_2}\right) 1_{\{H_n^{(1,1)}(Y_j) \leq \frac{c_n}{n}\}} \\
- \int (E(X|\theta'_0, X = x) - \mathbf{u}) K'(\frac{x - \theta_0}{h_1}) K\left(\frac{y - v}{h_2}\right) 1_{\{H_n^{(1,1)}(v) \leq \frac{c_n}{n}\}} f_\theta(x, y) dx dy \\
1_{\{H_n^{(1,1)}(y) \leq \frac{c_n}{n}\}} f_\theta(x, y) dx dy \\
= \frac{1}{\sqrt{n}} \sum_{j=2}^{n} \int \left(1 - G(Y_j + h_2 z_2 - \{\theta'_0 X_j + h_1 z_1\}) f_{\theta_0}(x, y) \right) \frac{\delta_j}{1 - G(Y_j -)} (E(X|\theta'_0, X = \theta'_0 X_j + h_1 z_1) - \mathbf{u}) \\
K'(z_1) K(z_2) 1_{\{H_n^{(1,1)}(Y_j + h_2 z_2) \leq \frac{c_n}{n}\}} \int (E(X|\theta'_0, X = \theta'_0 u + h_1 z_1) - \mathbf{u}) \\
K'(z_1) K(z_2) 1_{\{H_n^{(1,1)}(v) \leq \frac{c_n}{n}\}} f(u, v) dz_1 dz_2 \\
1_{\{H_n^{(1,1)}(v) \leq \frac{c_n}{n}\}} f(u, v) dz_1 dz_2 \\
= \frac{1}{\sqrt{n}} \sum_{j=2}^{n} \int \left(1 - G(|Y_j| - |E(X|\theta'_0, X = t) - X_j)| \right) \frac{\delta_j}{1 - G(Y_j -)} 1_{\{f_{\theta_0}(x, y) > c_n\}} \\
- \int \left(1 - G(|v - |t|)|E(X|\theta'_0, X = t) - \mathbf{u})| \right) \frac{\delta_j}{1 - G(Y_j -)} 1_{\{f_{\theta_0}(x, y) > c_n\}} \\
+ \frac{1}{\sqrt{n}} \sum_{j=2}^{n} h^*_j
\]

Set
\[
\frac{1}{\sqrt{n}} \sum_{j=2}^{n} \tilde{H}_{ij} = \frac{1}{\sqrt{n}} \sum_{j=2}^{n} \int \left(1 - G(|Y_j| - |E(X|\theta'_0, X = t) - X_j)| \right) \frac{\delta_j}{1 - G(Y_j -)} 1_{\{f_{\theta_0}(x, y) > c_n\}} \\
- \int \left(1 - G(|v - |t|)|E(X|\theta'_0, X = t) - \mathbf{u})| \right) \frac{\delta_j}{1 - G(Y_j -)} 1_{\{f_{\theta_0}(x, y) > c_n\}} \\
+ \alpha(1)
\]

Moreover, by Lemma 18
\[
\frac{1}{\sqrt{n}} \sum_{j=2}^{n} h_{ij}^* = \frac{1}{\sqrt{n}} \sum_{j=2}^{n} h_{ij}^2 \left[ \frac{d^2}{dt^2} G(Y_j - t)(E(X|\theta_0 X = t) - X_j) \right]_{t=\theta_0 X_j} 1_{(f_0, X|\theta_0 > c_n)} \frac{\delta 1_{[Y_j \leq a_n]}}{1 - G(Y_j - a_n)} \\
- \int \frac{d^3}{dt^3} G(v - t)(E(X|\theta_0 X = t) - u) \left[ \int f(u,v) dudv \right] \right]_{t=\theta_0 u} 1_{(f_0, X|\theta_0 > c_n)} \frac{\delta 1_{[v \leq a_n]}}{1 - G(Y_j - a_n)} \\
+ O_p(n^{1/2-\beta/2} h_1^2) + o_p(1),
\]

where \( a_n = H^{-1}(\frac{c_n}{n}) \). Finally, since the first part is a sum of i.i.d. random variables with variance going to zero as \( n \to \infty \), we have

\[
\frac{1}{\sqrt{n}} \sum_{j=2}^{n} h_{ij}^* = o_p(1).
\]

Next we show that

\[
\frac{\sqrt{n}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} (H^*_i - \bar{H}^*_i) = o_p(1).
\]

For this let us define

\[
H_{ij(kl)}^* = \frac{1}{h_1^2 h_2} \delta 1_{(f_0, X|\theta_0 > c_n)} \left[ \frac{\delta 1_{(H_n^{-1}(j,k,l)(Y_j) \leq \frac{c_n}{n})}}{1 - G(Y_j - a_n)} (X_i - X_j) \frac{\theta_0^t X_i - \theta_0^t X_j}{h_1} \right] K \left( \frac{Y_i - Y_j}{h_2} \right) - \int (X_i - u) K' \left( \frac{\theta_0^t X_i - \theta_0^t u}{h_1} \right) \frac{1}{1 - G(v - \frac{c_n}{n})} (H_i(du, dv)) \right] \frac{\delta 1_{(H_n^{-1}(j,k,l)(Y_j) \leq \frac{c_n}{n})}}{1 - G(Y_j - a_n)} \\
\]

and

\[
\bar{H}_{ij(kl)}^* = E(H_{ij(kl)}^*| Y_j, \delta_j, X_j, Y_s \text{ for } s \neq i).
\]

We have

\[
E([H_{ij(kl)}^* - \bar{H}_{ij(kl)}^*]| Y_j, \delta_j, X_j, Y_s \text{ for } s \neq i) = E([H_{ij(kl)}^* - \bar{H}_{ij(kl)}^*]| Y_j, \delta_i, X_i, Y_s \text{ for } s \neq j] = 0
\]

and given \( Y_s \text{ for } s \neq i, j, k, l \), we have that \( (H_{ij(kl)}^* - \bar{H}_{ij(kl)}^*) \) is independent of \( (H_{kl(ij)}^* - \bar{H}_{kl(ij)}^*) \). Hence, similarly as in the proof of Lemma 7, we can show that for every component \( m = 1, ..., d \)

\[
\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k=1}^{n} \sum_{l \neq k} E([H_{ij(kl)}^* - \bar{H}_{ij(kl)}^*]|m(H_{kl(ij)}^* - \bar{H}_{kl(ij)}^*)|m] = o(1). \quad (C.2)
\]

Therefore, it remains to prove that

\[
D = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k=1}^{n} \sum_{l \neq k} \left[ E([H_{ij(kl)}^* - \bar{H}_{ij(kl)}^*]|m(H_{kl(ij)}^* - \bar{H}_{kl(ij)}^*)|m - (H_{ij}^* - \bar{H}_{ij}^*)|m(H_{kl}^* - \bar{H}_{kl}^*)|m) \right] = o(1) \quad (C.3)
\]
The term in (C.3) can be handled in the following way

\[ D = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i}^{n} \sum_{l \neq k} E[(H^*_i (kl) - H^*_j (kl)) m(H^*_k (ij) - H^*_l (ij)) m - (H^*_i j - H^*_j) m(H^*_k l - H^*_l) m] \]

\[ = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i}^{n} \sum_{l \neq k} E\left\{ \left( (H^*_i (kl) - H^*_j (kl)) m - (H^*_i j - H^*_j) m \right) \left( H^*_k (ij) - H^*_l (ij) \right) m \right\} \]

\[ + (H^*_i (kl) - H^*_j (kl)) m \left( (H^*_k (ij) - H^*_l (ij)) m - (H^*_k) m - (H^*_l) m \right) \]

\[ + \left\{ (H^*_i j - H^*_j) m - (H^*_i (kl) - H^*_j (kl)) m \right\} \left\{ (H^*_k (ij) - H^*_l (ij)) m - (H^*_k l - H^*_l) m \right\} \]

Let us consider the first part

\[ \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i}^{n} \sum_{l \neq k} E\left\{ \left( (H^*_i (kl) - H^*_j (kl)) m - (H^*_i j - H^*_j) m \right) (H^*_k (ij) - H^*_l (ij)) m \right\} \]

First observe that, componentwise, \( E[(H^*_i j - H^*_j) | Y_j, \delta_i, X_j, Y_s \text{ for } s \neq i] = E[(H^*_i j - H^*_j) | Y_i, \delta_i, X_i, Y_s \text{ for } s \neq j] = 0 \)

- If \( i = k \neq j = l \), then

\[ \left\{ (H^*_i (kl) - H^*_j (kl)) m - (H^*_i j - H^*_j) m \right\} = 0 \]

- If \( i \neq k \neq j \neq l \), then

\[ E\left\{ \left( (H^*_i (kl) - H^*_j (kl)) m - (H^*_i j - H^*_j) m \right) (H^*_k (ij) - H^*_l (ij)) m \right\} = E[E(\ldots | X_k, \delta_k, X_l, \delta_l, Y_s \text{ for } s \neq i, j)] \]

\[ = E \left[ (H^*_k (ij) - H^*_l (ij)) m E\left\{ \left( (H^*_i (kl) - H^*_j (kl)) m - (H^*_i j - H^*_j) m \right) (X_k, \delta_k, X_l, \delta_l, Y_s \text{ for } s \neq i, j) \right\} \right] \]

Moreover,

\[ E\left\{ \left( (H^*_i (kl) - H^*_j (kl)) m - (H^*_i j - H^*_j) m \right) \right\} |X_k, \delta_k, X_l, \delta_l, Y_s \text{ for } s \neq i, j) \]

\[ = E(E\left\{ \left( (H^*_i (kl) - H^*_j (kl)) m - (H^*_i j - H^*_j) m \right) \right\} |X_s, \delta_s, Y_s \text{ for } s \neq i) |X_k, \delta_k, X_l, \delta_l, Y_s \text{ for } s \neq i, j) = 0 \]

- If \( i = k \neq j \neq l \), we have similarly

\[ E\left\{ \left( (H^*_i j (ld) - H^*_j (ld)) m - (H^*_i j - H^*_j) m \right) \right\} |X_k, \delta_k, X_l, \delta_l, Y_s \text{ for } s \neq i, j) \]

\[ = E(E\left\{ \left( (H^*_i j (ld) - H^*_j (ld)) m - (H^*_i j - H^*_j) m \right) \right\} |X_s, \delta_s, Y_s \text{ for } s \neq i) |X_k, \delta_k, X_l, \delta_l, Y_s \text{ for } s \neq i, j) \]

and

\[ E\left\{ \left( (H^*_i j (ld) - H^*_j (ld)) m - (H^*_i j - H^*_j) m \right) \right\} |X_s, \delta_s, Y_s \text{ for } s \neq j \}

\[ = E(E\left\{ \left( (H^*_i j (ld) - H^*_j (ld)) m - (H^*_i j - H^*_j) m \right) \right\} |X_s, \delta_s, Y_s \text{ for } s \neq j) |X_l, \delta_l, Y_s \text{ for } s \neq j) = 0 \]
Lemma 18. We have

\[ |1 - 1_n(Y) - 1_n(Y) - 1_n(Y) + 1_n(Y) + 1_n(\infty)Z | \leq A^{\delta_n}(Y) + 1_n(\infty)Z, \]  

where

\[ A^{\delta_n}(y) = \{ y : |H(y) - \frac{r_n}{n}| \leq \delta_n \} \]
\[
Z_n = \max_{i=1,\ldots,n} |H_n(Y_i) - H(Y_i)|.
\]

Set, \(\delta_n = n^{-\beta/2}\) and \(\beta < 1\). Hence

\[
1_{A_n}(Y_i) = O_P(n^{-\beta/2})
\]
and

\[
1_{(\delta_n,\infty)}(Z_n) = o_P(n^{-\alpha}). \quad (C.5)
\]

**Proof.**
The proof is based on Lemma B.2 in [6]. Observe that, \(Z_n \leq \sup_{t \in \mathbb{R}} |H_n(t) - H(t)|\) and, using the Dvoretzky-Kiefer-Wolfowitz inequality (see [7]) we have

\[
n^\alpha E \left( 1_{(\delta_n,\infty)}(Z_n) \right) \leq n^\alpha \mathbb{P}(D_n \geq \delta_n) \leq cn^\alpha \exp\left( -2n\delta_n^2 \right) \to 0.
\]