A 2-step estimation procedure for semiparametric mixture cure models

Supporting information

Eni Musta*, Valentin Patilea† and Ingrid Van Keilegom‡

*KdVI, University of Amsterdam, Netherlands † CREST, Ensai, France ‡ORSTAT, KU Leuven, Belgium

This supplement is organized as follows. Appendix S1 contains proofs of the asymptotic results in Section 4. Appendix S2 collects the simulation results for Section 5, which were omitted from the main paper to improve its readability.

S1 Proofs

Proof of Theorem 1. We start by showing that, when $\tilde{\gamma}_n - \gamma_0 = o_P(1)$, we have

$$\sup_{x\in\tilde{\mathcal{X}}} |\hat{\pi}(x) - \pi_0(x)| = o_P(1). \tag{S1}$$

Note that, independently of the choice of the trimming function,

$$\mathbb{E}\left[\left\{(1-\pi_0(X))\log\phi(\gamma^T X) + \pi_0(X)\log\left(1-\phi(\gamma^T X)\right)\right\}\tau(X)\right],$$

is maximized at $\gamma = \gamma_0$ because of condition (I1) and the fact that, for any $x \in \mathcal{X}$, the function

$$g_x(z) = \left\{\phi(\gamma_0^T x)\log\frac{z}{\phi(\gamma_0^T x)} + \left\{1-\phi(\gamma_0^T x)\right\}\log\frac{1-z}{1-\phi(\gamma_0^T x)}\right\}\tau(x), \quad z\in(0,1),$$

is strictly negative for $z \neq \phi(\gamma_0^T x)$ and $g_x(\phi(\gamma_0^T x)) = 0$.

Next we show uniform consistency of $\hat{\pi}_n$ using the decomposition (11) and restricting to $\tilde{\mathcal{X}}$ as in (12). We have

$$\begin{aligned}
\|\hat{\pi}_n - \pi_0\|_\infty &= \sup_{x\in\tilde{\mathcal{X}}} \left|\hat{g}_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x\right) - g_{\gamma_0}\left(\gamma_0^T x\right)\right| \\
&= \sup_{x\in\tilde{\mathcal{X}}} \left|\hat{g}_{\gamma_0}\left(\gamma_0^T x\right) - g_{\gamma_0}\left(\gamma_0^T x\right)\right| + \sup_{x\in\tilde{\mathcal{X}}} \left|\hat{g}_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x\right) - \hat{g}_{\gamma_0}\left(\gamma_0^T x\right)\right| \\
&\leq \sup_{x\in\tilde{\mathcal{X}}}\sup_{t\leq\tau_0} \left|\hat{F}_n\left(t \mid \gamma_0^T X = \gamma_0^T x\right) - F_T\left(t \mid \gamma_0^T X = \gamma_0^T x\right)\right| \\
&\quad + \|\tilde{\gamma}_n^T - \gamma_0\| \sup_{x\in\tilde{\mathcal{X}}}\sup_{\gamma\in G} \left|\nabla_\gamma \hat{g}_\gamma\left(\gamma^T x\right)\right|.
\end{aligned} \tag{S2}$$

The first term on the right hand side of the equation converges to zero by Theorem 4.1 in Van Keilegom and Akritas (1999). The second term converges to zero because of assumption (C1) and the fact that

$$\sup_{x\in\tilde{\mathcal{X}}}\sup_{\gamma\in G} \left|\nabla_\gamma \hat{g}_\gamma\left(\gamma^T x\right)\right| = O_P(1). \tag{S3}$$

Indeed, (S3) follows from assumption (C9) and

$$\sup_{x\in\tilde{\mathcal{X}}}\sup_{\gamma\in G} \left|\nabla_\gamma \hat{g}_\gamma\left(\gamma^T x\right) - \nabla_\gamma g_\gamma\left(\gamma^T x\right)\right| = o_P(1), \tag{S4}$$

which can be proved as in Lemma A.2 in Lopez et al. (2013). Their estimator $\hat{G}_\theta(t|\lambda(\theta,x))$ is the same as our $\hat{F}_n(t|\gamma^T X = \gamma^T x)$ if we replace θ by γ, consider $\lambda(\theta,x) = \theta^T x$ and exchange T with C (they are interested in the conditional distribution of C given $\lambda(\theta,X)$). This concludes the proof of (S1).

The consistency of $\hat{\gamma}_n$ then follows similarly to Theorem 1 in Musta et al. (2022). There, in order to obtain strong consistency of $\hat{\gamma}_n$, it is required that $\sup_{x\in\tilde{\mathcal{X}}} |\hat{\pi}(x) - \pi_0(x)| \to 0$ almost surely. Here we restrict to convergence in probability and conclude via Theorem 1 in Chen et al. (2003) formulated for M-estimators. □

Proof of Theorem 2. The result follows from Theorem 2 in Musta et al. (2022) once we show that the assumptions (AN1)-(AN4) of that paper hold. Note that the introduction of the trimming function $\tau(\cdot)$ would not change anything in the proof. It just allows us to restrict to the set $\tilde{\mathcal{X}}$ where the density of the index is bounded from below, in order to apply standard results from the literature. The assumptions (AN1) and (AN3) of Musta et al. (2022) are the same as assumptions (C2), (N1) and (N2) here. It remains to verify assumptions (AN2) and (AN4) which for completeness we state below:

(AN2) $\pi_0(\cdot)$ belongs to a class of functions Π such that

$$\int_0^\infty \sqrt{\log N(\epsilon, \Pi, \|\cdot\|_\infty)}\, \mathrm{d}\epsilon < \infty,$$

where $N(\epsilon, \Pi, \|\cdot\|_\infty)$ denotes the ϵ-covering number of the space Π with respect to $\|\pi\|_\infty = \sup_{x\in\tilde{\mathcal{X}}} |\pi(x)|$.

(AN4) The estimator $\hat{\pi}(\cdot)$ satisfies the following properties:

(i) $\mathbb{P}\left(\hat{\pi}(\cdot) \in \Pi\right) \to 1$.

(ii) $\|\hat{\pi}(x) - \pi_0(x)\|_\infty = o_P(n^{-1/4})$.

(iii) There exists a function Ψ such that

$$\mathbb{E}^*\left[(\hat{\pi}(X) - \pi_0(X))\left(\frac{1}{\phi(\gamma_0^T X)} + \frac{1}{1-\phi(\gamma_0^T X)}\right)\phi'(\gamma_0^T X) X \tau(X)\right] = \frac{1}{n}\sum_{i=1}^n \Psi(Y_i, \Delta_i, X_i, Z_i) + R_n,$$

where $\mathbb{E}^*$ denotes the conditional expectation given the sample, taken with respect to the generic variable X, $\|R_n\| = o_P(n^{-1/2})$ and $\mathbb{E}[\Psi(Y, \Delta, X, Z)] = 0$.

Step 1. For (AN2) we use the class of functions

$$\Pi = \left\{ f : \mathcal{X} \to [0,1],\ f(x) = g(\gamma^T x) \text{ for some } \gamma \in G \text{ and } g \in \tilde{\Pi}_\gamma \right\},$$

where $\tilde{\Pi}_\gamma$ is the space of continuously differentiable functions g from $\mathcal{X}_\gamma$ (defined in assumption C5) to $[0,1]$ such that $\sup_{u\in\mathcal{X}_\gamma} |g'(u)| \le M$ and

$$\sup_{\substack{u_1 \neq u_2 \\ u_1, u_2 \in \mathcal{X}_\gamma}} \frac{|g'(u_1) - g'(u_2)|}{|u_1 - u_2|^\xi} \le M,$$

for some $M > 0$ and $\xi \in (0,1]$ independent of γ. The norm that we consider on Π is the sup norm. By assumption (A2) we have $\pi_0 \in \Pi$. Next we compute the ϵ-covering number of Π. Let $\gamma_1, \ldots, \gamma_{N_1}$ be a δ-covering of the compact $G \in \mathbb{R}^p$ with respect to the l_1 norm. We have $N_1 \le \frac{K_1}{\delta^p}$ for some constant $K_1 > 0$. Consider the class $\tilde{\Pi}$ of continuously differentiable functions g from $\mathcal{U} = \{\gamma^T x \mid \gamma \in G, x \in \mathcal{X}\}$ to $[0,1]$ such that $\sup_{u\in\mathcal{U}} |g'(u)| \le M$ and

$$\sup_{\substack{u_1 \neq u_2 \\ u_1, u_2 \in \mathcal{U}}} \frac{|g'(u_1) - g'(u_2)|}{|u_1 - u_2|^\xi} \le M.$$

We have in particular that if $g \in \tilde{\Pi}$ then the restriction of g to $\mathcal{X}_\gamma$ belongs to the class $\tilde{\Pi}_\gamma$ for any $\gamma \in \mathcal{G}$. Let $g_1, \ldots, g_{N_2}$ be a ρ-covering of $\tilde{\Pi}$ with respect to the sup norm. Since $\mathcal{U}$ is bounded and convex (G can be chosen to be a small neighborhood of γ_0), from Theorem 2.7.1 in van der Vaart and Wellner (1996), we have

$$\log N_2 \le K_2 \left(\frac{1}{\rho}\right)^{1/(1+\xi)},$$

for some constant $K_2 > 0$. For $1 \le i \le N_1$ and $1 \le j \le N_2$, define $f_{i,j}(x) = g_j(\gamma_i^T x)$. We show that $\{f_{i,j}\}$, $1 \le i \le N_1$ and $1 \le j \le N_2$ is an ϵ-covering of Π. Let $f \in \Pi$, $f(x) = g(\gamma^T x)$. From Whitney's theorem it

follows that $g \in \tilde{\Pi}_\gamma$ can be extended to a function $\bar{g} \in \tilde{\Pi}$, i.e. $\bar{g}(u) = g(u)$ for $u \in \mathcal{X}_\gamma$. Let i and j be such that $\sup_{u\in\mathcal{U}} |\bar{g}(u) - g_j(u)| \leq \rho$ and $\|\gamma - \gamma_i\| \leq \delta$. Then we have

$$\begin{aligned}\|f - f_{i,j}\|_\infty &\leq \sup_{x\in\mathcal{X}} \left|g\left(\gamma^T x\right) - g_j\left(\gamma^T x\right)\right| + \sup_{x\in\mathcal{X}} \left|g_j\left(\gamma^T x\right) - g_j\left(\gamma_i^T x\right)\right| \\ &\leq \sup_{u\in\mathcal{U}} |\bar{g}(u) - g_j(u)| + K_3\|\gamma - \gamma_i\|_1 \\ &\leq \rho + MK_3\delta.\end{aligned}$$

Hence, if we take $\rho = \epsilon/2$ and $\delta = \epsilon/MK_3$ then we get an ϵ-covering of Π. It follows that

$$\log N(\epsilon, \Pi, \|\cdot\|_\infty) \leq c_1 - p\log\epsilon + c_2\left(\frac{1}{\epsilon}\right)^{1/(1+\xi)} \qquad c_1, c_2 > 0,$$

and as a result (AN2) is satisfied.

Step 2. For assumption (AN4))(i), note that $\hat{\pi}_n(x) = \hat{g}_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T x)$ with $\hat{g}_\gamma$ as in (9). From consistency of $\tilde{\gamma}_n$ it follows that $\tilde{\gamma}_n \in G$. Let $\hat{g}'_{\tilde{\gamma}_n}$ denote the derivative of the function $u \mapsto \hat{g}_{\tilde{\gamma}_n}(u)$. Let g_γ be as in (10). Next we show that

$$\sup_{x\in\tilde{\mathcal{X}}} |\hat{g}'_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T x) - g'_{\gamma_0}(\gamma_0^T x)| = o_P(1), \tag{S5}$$

and

$$\sup_{\substack{x_1,x_2\in\tilde{\mathcal{X}} \\ \tilde{\gamma}_n^T x_1 \neq \tilde{\gamma}_n^T x_2}} \frac{\left|\hat{g}'_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x_1\right) - g'_{\gamma_0}\left(\gamma_0^T x_1\right) - \hat{g}'_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x_2\right) + g'_{\gamma_0}\left(\gamma_0^T x_2\right)\right|}{|\tilde{\gamma}_n^T x_1 - \tilde{\gamma}_n^T x_2|^\xi} = o_P(1), \tag{S6}$$

from which we can derive that (AN4)(i) is satisfied since $\sup_{x\in\tilde{\mathcal{X}}} |g'_{\gamma_0}(\gamma_0^T x)| < \infty$ and

$$\sup_{\substack{x_1,x_2\in\tilde{\mathcal{X}} \\ \gamma_0^T x_1 \neq \gamma_0^T x_2}} \frac{\left|g'_{\gamma_0}\left(\gamma_0^T x_2\right) - g'_{\gamma_0}\left(\gamma_0^T x_1\right)\right|}{|\gamma_0^T x_1 - \gamma_0^T x_2|^\xi} < \infty.$$

From Theorems 4.1 and 4.2 in Van Keilegom and Akritas (1999) we have

$$\sup_{x\in\tilde{\mathcal{X}}} |\hat{g}'_{\gamma_0}(\gamma_0^T x) - g'_{\gamma_0}(\gamma_0^T x)| = o_P(1), \tag{S7}$$

and

$$\sup_{\substack{x_1,x_2\in\tilde{\mathcal{X}} \\ \gamma_0^T x_1 \neq \gamma_0^T x_2}} \frac{\left|\hat{g}'_{\gamma_0}\left(\gamma_0^T x_1\right) - g'_{\gamma_0}\left(\gamma_0^T x_1\right) - \hat{g}'_{\gamma_0}\left(\gamma_0^T x_2\right) + g'_{\gamma_0}\left(\gamma_0^T x_2\right)\right|}{|\gamma_0^T x_1 - \gamma_0^T x_2|^\xi} = o_P(1), \tag{S8}$$

by conditioning on the variable $\gamma_0^T X$. Next we need to deal with the fact that we are using $\tilde{\gamma}_n^T X$ instead of $\gamma_0^T X$.

We can write

$$\log \hat{g}_{\tilde{\gamma}_n}(u) = \sum_{i=1}^n \Delta_i \log\left(1 - W_i(u; \tilde{\gamma}_n)\right),$$

where

$$W_i(u; \tilde{\gamma}_n) = k\left(\frac{\tilde{\gamma}_n^T X_i - u}{b}\right) \Bigg/ \left(\sum_{j=1}^n \mathbb{1}_{\{T_j \geq T_i\}} k\left(\frac{\tilde{\gamma}_n^T X_j - u}{b}\right)\right).$$

Hence

$$\hat{g}'_{\tilde{\gamma}_n}(u) = \hat{g}_{\tilde{\gamma}_n}(u) \sum_{i=1}^n \Delta_i \frac{W'_i(u; \tilde{\gamma}_n)}{1 - W_i(u; \tilde{\gamma}_n)}. \tag{S9}$$

Using the triangular inequality we obtain

$$\begin{aligned}
&\sup_{x\in\tilde{\mathcal{X}}} \left|\hat{g}'_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T x) - \hat{g}'_{\gamma_0}(\gamma_0^T x)\right| \\
&\leq \sup_{x\in\tilde{\mathcal{X}}} \left|\hat{g}_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T x) - \hat{g}_{\gamma_0}(\gamma_0^T x)\right| \sup_{x\in\tilde{\mathcal{X}}} \left|\sum_{i=1}^n \Delta_i \frac{W_i'(\tilde{\gamma}_n^T x;\tilde{\gamma}_n)}{1-W_i(\tilde{\gamma}_n^T x;\tilde{\gamma}_n)}\right| \\
&+ \sup_{x\in\tilde{\mathcal{X}}} \left|\hat{g}'_{\gamma_0}(\gamma_0^T x)\right| \sup_{x\in\mathcal{X}} \left|\sum_{i=1}^n \Delta_i \left\{\frac{W_i'(\tilde{\gamma}_n^T x;\tilde{\gamma}_n)}{1-W_i(\tilde{\gamma}_n^T x;\tilde{\gamma}_n)} - \frac{W_i'(\gamma_0^T x;\gamma_0)}{1-W_i(\gamma_0^T x;\gamma_0)}\right\}\right|.
\end{aligned} \tag{S10}$$

As is (S2), we have

$$\sup_{x\in\tilde{\mathcal{X}}} \left|\hat{g}_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T x) - \hat{g}_{\gamma_0}(\gamma_0^T x)\right| = O_P(n^{-1/2}).$$

Moreover, $\sup_{x\in\tilde{\mathcal{X}}} |\hat{g}'_{\gamma_0}(\gamma_0^T x)| = O_P(1)$ and, as in the proof of Proposition 4.1 in Van Keilegom and Akritas (1999) it can be seen that

$$\sup_{x\in\tilde{\mathcal{X}}} \left|\sum_{i=1}^n \Delta_i \frac{W_i'(\gamma_0^T x;\gamma_0)}{1-W_i(\gamma_0^T x;\gamma_0)}\right| = O_P(1).$$

and

$$\sup_{x\in\tilde{\mathcal{X}}} \left|\sum_{i=1}^n \Delta_i \nabla_\gamma \left(\frac{W_i'(\gamma^T x;\gamma)}{1-W_i(\gamma^T x;\gamma)}\right)\Big|_{\gamma_0}\right| = O_P(1/b).$$

From a Taylor expansion and $\tilde{\gamma}_n - \gamma_0 = O_P(n^{-1/2})$ we conclude that the left hand side of (S10) converges to zero. Together with (S7) we obtain (S5).

Next we show that

$$\sup_{\substack{x_1,x_2\in\tilde{\mathcal{X}}\\ \tilde{\gamma}_n^T x_1 \neq \tilde{\gamma}_n^T x_2}} \frac{\left|\hat{g}'_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x_1\right) - \hat{g}'_{\gamma_0}\left(\gamma_0^T x_1\right) - \hat{g}'_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x_2\right) + \hat{g}'_{\gamma_0}\left(\gamma_0^T x_2\right)\right|}{|\tilde{\gamma}_n^T x_1 - \tilde{\gamma}_n^T x_2|^\xi} = o_P(1).$$

Note that by (S9) and the fact that $\sup_{x\in\mathcal{X}} |\hat{g}'_\gamma(\gamma^T x)| = O_P(1)$ for both $\gamma = \gamma_0$ and $\gamma = \tilde{\gamma}_n$, it is sufficient to consider

$$\sup_{\substack{x_1,x_2\in\tilde{\mathcal{X}}\\ \tilde{\gamma}_n^T x_1 \neq \tilde{\gamma}_n^T x_2}} \frac{\left|V_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x_1\right) - V_{\gamma_0}\left(\gamma_0^T x_1\right) - V_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x_2\right) + V_{\gamma_0}\left(\gamma_0^T x_2\right)\right|}{|\tilde{\gamma}_n^T x_1 - \tilde{\gamma}_n^T x_2|^\xi}, \tag{S11}$$

where

$$V_\gamma(u) = \sum_{i=1}^n \Delta_i \frac{W_i'(u;\gamma)}{1-W_i(u;\gamma)}.$$

In addition we can also restrict the supremum over x_1, x_2 such that $\tilde{\gamma}_n^T x_1 \neq \tilde{\gamma}_n^T x_2$ and $\gamma_0^T x_1 \neq \gamma_0^T x_2$ because otherwise we would have $|\tilde{\gamma}_n^T x_1 - \tilde{\gamma}_n^T x_2| \leq c n^{-1/2}$ and (S11) would obviously hold for ξ sufficiently small by using a Taylor expansion and $\sup_u V'_\gamma(u) = O_P(1/b)$. Note also that

$$\sup_{\substack{x_1,x_2\in\tilde{\mathcal{X}}\\ \tilde{\gamma}_n^T x_1 \neq \tilde{\gamma}_n^T x_2,\, \gamma_0^T x_1 \neq \gamma_0^T x_2}} \frac{|\gamma_0^T x_1 - \gamma_0^T x_2|}{|\tilde{\gamma}_n^T x_1 - \tilde{\gamma}_n^T x_2|} = O_P(1),$$

hence the denominator can be replaced by $|\gamma_0^T x_1 - \gamma_0^T x_2|^\xi$. By Taylor expansion we obtain that the largest order term in (S11) is

$$(\tilde{\gamma}_n - \gamma_0) \sup_{\substack{x_1,x_2\in\tilde{\mathcal{X}}\\ \gamma_0^T x_1 \neq \gamma_0^T x_2}} \sum_{i=1}^n \Delta_i \frac{\nabla_\gamma \left(\frac{W_i'(\gamma^T x_1;\gamma)}{1-W_i(\gamma^T x_1;\gamma)}\right)\Big|_{\gamma_0} - \nabla_\gamma \left(\frac{W_i'(\gamma^T x_2;\gamma)}{1-W_i(\gamma^T x_2;\gamma)}\right)\Big|_{\gamma_0}}{|\gamma_0^T x_1 - \gamma_0^T x_2|^\xi}.$$

Using Lemma A.1. in Van Keilegom and Akritas (1999) and that for a smooth function l (in our case l is the kernel function or its derivatives) we have

$$\frac{\left|l(\frac{x-z}{b}) - l(\frac{x-z}{b})\right|}{|x-y|^{\xi}} = O(b^{-\xi}),$$

uniformly over x, y, z, we obtain that the expression in the previous equation is of the order $O_P(n^{-1/2}b^{-1-\xi}) = o_P(1)$. Together with (S8) this yields (S6).

Step 3. Assumption (AN4)(ii) can be checked using (S2). The first term on the right hand side of that equation is of order $o_P(n^{-1/4})$ because of Theorem 4.1 in Van Keilegom and Akritas (1999) and the assumption (C6). The second term on the right hand side of (S2) also is of order $o_P(n^{-1/4})$ because of assumptions (N3), (C9) and (S4).

Step 4. For (AN4)(iii), using the decomposition (11) we have

$$\begin{aligned}
&\mathbb{E}^*\left[(\hat{\pi}(X) - \pi_0(X))\left(\frac{1}{\phi(\gamma_0^T X)} + \frac{1}{1-\phi(\gamma_0^T X)}\right)\phi'(\gamma_0^T X)X\tau(X)\right] \\
&= \mathbb{E}^*\left[\frac{\hat{F}_n(\tau_0 \mid \gamma_0^T X) - F_T(\tau_0 \mid \gamma_0^T X)}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right] \\
&\quad + \mathbb{E}^*\left[\frac{\hat{g}_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T X) - \hat{g}_{\gamma_0}(\gamma_0^T X)}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right].
\end{aligned} \tag{S12}$$

Since $\sup_{x\in\mathcal{X}} H([t,\infty)|\gamma_0^T x) < 1$ by condition (8), from Theorem 3.2 in Du and Akritas (2002) it follows as in Musta et al. (2022) (see proof of Theorem 5) that the first term on the right hand side of (S12) is equal to $n^{-1}\sum_{i=1}^n \psi(Y_i, \Delta_i, X_i) + R_n$ with $\|R_n\| = o_P(n^{-1/2})$ and ψ as in (13). For the second term in (S12), by mean value theorem we write

$$\begin{aligned}
\hat{g}_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x\right) - \hat{g}_{\gamma_0}\left(\gamma_0^T x\right) &= (\tilde{\gamma}_n - \gamma_0)^T \nabla_\gamma \hat{g}_\gamma\left(\gamma^T x\right)\big|_{\gamma^*} \\
&= (\tilde{\gamma}_n - \gamma_0)^T \nabla_\gamma \hat{g}_\gamma\left(\gamma^T x\right)\big|_{\gamma^*},
\end{aligned}$$

for some γ^* such that $\|\gamma^* - \gamma_0\| \le \|\tilde{\gamma}_n - \gamma_0\|$. Using (S4) and assumption (C9), we obtain

$$\hat{g}_{\tilde{\gamma}_n}\left(\tilde{\gamma}_n^T x\right) - \hat{g}_{\gamma_0}\left(\gamma_0^T x\right) = (\tilde{\gamma}_n - \gamma_0)^T \nabla_\gamma g_\gamma\left(\gamma^T x\right)\big|_{\gamma_0} + o_P(n^{-1/2}),$$

where the $o_P(n^{-1/2})$ term is uniform with respect to $x \in \tilde{\mathcal{X}}$. It follows that

$$\begin{aligned}
&\mathbb{E}^*\left[\frac{\hat{g}_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T X) - \hat{g}_{\gamma_0}(\gamma_0^T X)}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right] \\
&= \mathbb{E}\left[\frac{\nabla_\gamma g_\gamma(\gamma^T X)^T\big|_{\gamma_0}}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right](\tilde{\gamma}_n - \gamma_0) + o_P(n^{-1/2}),
\end{aligned}$$

because E^* is just the expectation with respect to the variable X. By definition of $g_\gamma(\gamma^T x)$ and a Taylor expansion we have

$$\begin{aligned}
g_\gamma\left(\gamma^T x\right) &= \mathbb{E}\left[1 - \phi\left(\gamma_0^T X\right)|\gamma^T X = \gamma^T x\right] \\
&= 1 - \phi\left(\gamma^T x\right) + (\gamma - \gamma_0)^T \mathbb{E}\left[X|\gamma^T X = \gamma^T x\right]\phi'\left(\gamma^T x\right) \\
&\quad + \frac{1}{2}(\gamma - \gamma_0)^T \mathbb{E}\left[\phi''\left(\gamma_*^T x\right) XX^T|\gamma^T X = \gamma^T x\right](\gamma - \gamma_0),
\end{aligned}$$

for some $\|\gamma_* - \gamma_0\| \le \|\gamma - \gamma_0\|$. Hence

$$\nabla_\gamma g_\gamma\left(\gamma^T x\right)\big|_{\gamma_0} = -\phi'\left(\gamma_0^T x\right)x + \mathbb{E}\left[X|\gamma_0^T X = \gamma_0^T x\right]\phi'\left(\gamma_0^T x\right).$$

and consequently

$$\begin{aligned}
&\mathbb{E}\left[\frac{\nabla_\gamma g_\gamma(\gamma^T X)^T\big|_{\gamma_0}}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right]\\
&= -\mathbb{E}\left[\frac{\phi'(\gamma_0^T X)X^T}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right]\\
&\quad + \mathbb{E}\left[\frac{\mathbb{E}[X^T|\gamma_0^T X]\phi'(\gamma_0^T X)}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right]\\
&= \mathbb{E}\left[\frac{\phi'(\gamma_0^T X)^2\left\{\mathbb{E}[X^T|\gamma_0^T X]\mathbb{E}[X\tau(X)|\gamma_0^T X]-\mathbb{E}[X^T X\tau(X)|\gamma_0^T X]\right\}}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\right],
\end{aligned}$$

We denote this expression by Q. This yields

$$\mathbb{E}^*\left[\frac{\hat{g}_{\tilde{\gamma}_n}(\tilde{\gamma}_n^T X)-\hat{g}_{\gamma_0}(\gamma_0^T X)}{\phi(\gamma_0^T X)\left\{1-\phi(\gamma_0^T X)\right\}}\phi'(\gamma_0^T X)X\tau(X)\right] = Q(\tilde{\gamma}_n-\gamma_0)+o_P(n^{-1/2}),$$

and by assumption (N3)

$$\begin{aligned}
&\mathbb{E}^*\left[(\hat{\pi}(X)-\pi_0(X))\left(\frac{1}{\phi(\gamma_0^T X)}+\frac{1}{1-\phi(\gamma_0^T X)}\right)\phi'(\gamma_0^T X)X\tau(X)\right]\\
&= \frac{1}{n}\sum_{i=1}^n \Psi(Y_i,\Delta_i,X_i)+\frac{1}{n}\sum_{i=1}^n Q\zeta(Y_i,\Delta_i,X_i,Z_i)+R_n,
\end{aligned}$$

with $\|R_n\| = o_P(n^{-1/2})$. This means that assumption (AN4)(iii) of Musta et al. (2022) is satisfied with $\Psi(Y,\Delta,X,Z) = \psi(Y,\Delta,X,Z)+Q\zeta(Y,\Delta,X,Z)$.

This concludes the verification of the assumptions of Theorem 3 in Musta et al. (2022). It also follows that the covariance matrix is given by

$$\Sigma_\gamma = \left(\Gamma_1^T\Gamma_1\right)^{-1}\Gamma_1^T V\Gamma_1\left(\Gamma_1^T\Gamma_1\right)^{-1} = \Gamma_1^{-1}V\Gamma_1^{-1},$$

where $V = \mathrm{Var}(\Psi(Y,\Delta,X,Z))$ and Γ_1 is defined as in (14). □

S2 Additional simulation results

In this section we provide additional simulation results which were not included in Section 5 in order to simplify exposition and improve readability of the paper. The simulation settings have been described in Section 5 and a summary of these results can also be found there.

		Scenario 1			Scenario 2			Scenario 3		
n	Par.	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
200	γ_1	−0.007	0.246	0.246	−0.013	0.062	0.062	0.088	0.049	0.057
		0.243	0.534	0.593	0.058	0.092	0.095	0.096	0.053	0.062
		-1.233	0.034	1.553	-0.627	0.029	0.421	-0.228	0.037	0.089
	γ_2	−0.009	0.217	0.217	−0.035	0.116	0.117	−0.043	0.089	0.091
		0.201	0.302	0.343	0.096	0.146	0.155	0.049	0.099	0.102
		-0.380	0.044	0.188	-0.15	0.053	0.076	-0.017	0.065	0.065
	γ_3	−0.050	0.406	0.408	−0.028	0.205	0.206	−0.031	0.152	0.153
		0.189	0.717	0.753	0.108	0.242	0.253	0.063	0.174	0.178
		-0.552	0.094	0.399	-0.259	0.092	0.159	-0.077	0.118	0.124
	β_1	−0.006	0.015	0.015	0.005	0.024	0.024	0.015	0.034	0.034
		−0.009	0.016	0.016	−0.002	0.025	0.025	0.009	0.034	0.034
		-0.059	0.012	0.016	-0.042	0.019	0.02	-0.014	0.029	0.03
	β_2	0.008	0.035	0.035	0.008	0.050	0.050	0.012	0.071	0.072
		0.003	0.036	0.036	−0.002	0.052	0.052	0.002	0.073	0.073
		-0.025	0.029	0.029	-0.019	0.04	0.041	-0.006	0.062	0.062
400	γ_1	−0.004	0.109	0.109	−0.002	0.036	0.036	0.000	0.026	0.026
		0.129	0.161	0.178	0.040	0.043	0.045	0.004	0.027	0.027
		-1.236	0.015	1.543	-0.623	0.014	0.403	-0.317	0.018	0.119
	γ_2	−0.015	0.094	0.095	−0.023	0.059	0.059	−0.012	0.047	0.048
		0.101	0.108	0.118	0.059	0.059	0.063	0.057	0.052	0.055
		-0.383	0.022	0.169	-0.154	0.025	0.048	-0.001	0.035	0.035
	γ_3	−0.008	0.201	0.201	−0.019	0.110	0.110	−0.010	0.084	0.084
		0.116	0.247	0.261	0.065	0.115	0.119	0.060	0.088	0.091
		-0.550	0.048	0.350	-0.264	0.051	0.121	-0.065	0.062	0.066
	β_1	0.005	0.008	0.008	0.013	0.012	0.012	0.011	0.017	0.017
		0.003	0.008	0.008	0.009	0.012	0.012	0.006	0.017	0.017
		-0.052	0.006	0.009	-0.034	0.009	0.011	-0.018	0.014	0.015
	β_2	0.007	0.019	0.019	0.012	0.026	0.027	0.015	0.034	0.034
		0.004	0.019	0.019	0.005	0.027	0.027	0.008	0.035	0.035
		-0.024	0.015	0.016	-0.014	0.021	0.021	-0.002	0.029	0.029

Table S1: Bias, variance and MSE of $\hat{\gamma}$ and $\hat{\beta}$ for the proposed 2-step approach (first rows), `smcure` (second rows) and multivarite presmoothing of Musta et al. (2022) (third rows) for Model 1 and $n = 200, 400$.

		Scenario 1			Scenario 2			Scenario 3		
n	Par.	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
200	γ_1	0.049	0.248	0.250	0.011	0.114	0.114	0.000	0.244	0.244
		0.103	0.268	0.278	0.021	0.121	0.122	0.097	0.389	0.398
	γ_2	0.017	0.073	0.074	0.010	0.047	0.047	0.090	0.091	0.099
		−0.081	0.088	0.094	−0.040	0.051	0.052	−0.154	0.169	0.193
	γ_3	0.025	0.947	0.948	−0.009	0.171	0.171	−0.102	0.299	0.310
		0.167	1.895	1.923	0.035	0.177	0.178	0.084	0.444	0.451
	γ_4	0.007	0.273	0.273	0.005	0.150	0.150	0.031	0.270	0.271
		−0.021	0.303	0.303	−0.011	0.162	0.162	−0.040	0.459	0.461
	β_1	−0.003	0.014	0.014	−0.010	0.021	0.021	−0.074	0.069	0.075
		0.002	0.014	0.014	−0.007	0.021	0.021	−0.016	0.085	0.086
	β_2	0.029	0.050	0.051	0.043	0.069	0.071	0.099	0.236	0.245
		0.027	0.051	0.051	0.042	0.069	0.071	0.074	0.269	0.274
	β_3	−0.009	0.044	0.044	−0.014	0.052	0.052	−0.034	0.220	0.221
		−0.008	0.045	0.045	−0.013	0.052	0.052	−0.028	0.252	0.253
400	γ_1	0.029	0.111	0.112	0.010	0.055	0.055	0.017	0.107	0.107
		0.055	0.111	0.114	0.013	0.056	0.056	0.058	0.129	0.133
	γ_2	0.016	0.040	0.040	0.006	0.022	0.022	0.061	0.046	0.049
		−0.043	0.039	0.041	−0.025	0.022	0.023	−0.082	0.064	0.071
	γ_3	−0.030	0.155	0.156	−0.017	0.079	0.080	−0.111	0.142	0.154
		0.030	0.160	0.161	0.012	0.084	0.084	0.006	0.163	0.163
	γ_4	0.010	0.114	0.114	0.015	0.066	0.067	0.030	0.128	0.129
		−0.006	0.119	0.119	0.007	0.068	0.068	−0.003	0.167	0.167
	β_1	0.004	0.007	0.007	−0.001	0.008	0.008	−0.037	0.030	0.032
		0.007	0.007	0.007	0.001	0.008	0.008	0.004	0.036	0.036
	β_2	0.010	0.024	0.024	0.018	0.029	0.029	0.063	0.104	0.108
		0.009	0.024	0.024	0.017	0.029	0.029	0.043	0.112	0.114
	β_3	0.001	0.021	0.021	−0.005	0.026	0.026	−0.015	0.100	0.100
		0.001	0.022	0.022	−0.005	0.026	0.026	−0.012	0.112	0.112

Table S2: Bias, variance and MSE of $\hat{\gamma}$ and $\hat{\beta}$ for the proposed 2-step approach (first rows) and `smcure` (second rows) in Model 2 and for $n = 200, 400$.

		Scenario 1			Scenario 2			Scenario 3		
n	Par.	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
200	γ_1	−0.056	0.287	0.290	−0.016	0.123	0.123	−0.010	0.097	0.097
		0.431	4.843	5.028	0.068	0.144	0.149	0.011	0.098	0.099
	γ_2	0.083	0.041	0.048	0.043	0.033	0.035	0.022	0.030	0.031
		−0.029	0.071	0.072	−0.014	0.040	0.040	−0.012	0.032	0.032
	γ_3	−0.049	0.177	0.179	−0.021	0.105	0.106	0.004	0.093	0.093
		0.105	0.268	0.279	0.049	0.120	0.123	0.043	0.097	0.099
	γ_4	−0.080	0.216	0.222	−0.041	0.131	0.133	−0.047	0.122	0.124
		0.159	1.233	1.259	0.046	0.151	0.154	0.010	0.131	0.131
	γ_5	0.123	0.325	0.340	0.051	0.158	0.161	0.021	0.136	0.137
		−0.359	4.817	4.946	−0.062	0.183	0.187	−0.036	0.140	0.141
	β_1	0.003	0.011	0.011	0.003	0.016	0.016	0.016	0.025	0.025
		0.003	0.012	0.012	0.003	0.017	0.017	0.016	0.025	0.025
	β_2	0.022	0.034	0.034	0.017	0.045	0.046	0.028	0.075	0.076
		0.010	0.036	0.036	0.008	0.048	0.048	0.023	0.077	0.077
	β_3	−0.048	0.044	0.046	−0.042	0.060	0.061	−0.030	0.093	0.094
		−0.032	0.048	0.049	−0.027	0.063	0.064	−0.023	0.095	0.095
400	γ_1	−0.045	0.149	0.151	−0.009	0.069	0.070	−0.020	0.046	0.047
		0.124	0.213	0.228	0.043	0.071	0.073	−0.003	0.045	0.045
	γ_2	0.063	0.021	0.025	0.034	0.016	0.017	0.023	0.015	0.015
		−0.018	0.030	0.030	−0.008	0.018	0.018	−0.003	0.016	0.016
	γ_3	−0.041	0.098	0.099	−0.025	0.058	0.059	0.001	0.050	0.050
		0.052	0.116	0.119	0.019	0.061	0.062	0.026	0.052	0.052
	γ_4	−0.088	0.111	0.119	−0.052	0.069	0.071	−0.020	0.059	0.059
		0.029	0.149	0.149	0.007	0.076	0.076	0.019	0.060	0.060
	γ_5	0.103	0.165	0.175	0.047	0.083	0.085	0.033	0.068	0.069
		−0.079	0.226	0.232	−0.030	0.085	0.086	−0.013	0.066	0.067
	β_1	−0.001	0.005	0.005	0.001	0.007	0.007	0.003	0.010	0.010
		−0.001	0.005	0.005	0.001	0.007	0.007	0.003	0.010	0.010
	β_2	0.007	0.017	0.017	0.006	0.023	0.023	0.002	0.034	0.034
		0.000	0.018	0.018	0.001	0.024	0.024	−0.001	0.035	0.035
	β_3	−0.017	0.021	0.021	−0.012	0.028	0.028	−0.007	0.042	0.042
		−0.002	0.022	0.022	−0.003	0.029	0.029	−0.002	0.043	0.043

Table S3: Bias, variance and MSE of $\hat{\gamma}$ and $\hat{\beta}$ for the proposed 2-step approach (first rows) and `smcure` (second rows) in Model 3 and for $n = 200, 400$.

		Scenario 1			Scenario 2			Scenario 3		
n	Par.	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
200	γ_1	0.086	0.237	0.245	0.037	0.238	0.239	0.010	0.229	0.229
		0.110	0.273	0.285	0.077	0.311	0.317	0.013	0.310	0.310
		-0.013	0.224	0.224	-0.105	0.196	0.207	-0.098	0.223	0.233
	γ_2	0.010	0.058	0.058	0.042	0.065	0.067	0.006	0.065	0.069
		−0.069	0.064	0.069	−0.091	0.077	0.085	−0.073	0.068	0.073
		0.112	0.056	0.068	0.193	0.040	0.077	0.168	0.042	0.071
	γ_3	−0.025	0.131	0.132	−0.018	0.122	0.122	−0.021	0.122	0.122
		0.000	0.153	0.153	0.033	0.162	0.164	0.027	0.155	0.156
		-0.084	0.095	0.102	-0.082	0.082	0.089	-0.079	0.092	0.099
	γ_4	0.008	0.095	0.095	0.007	0.108	0.109	−0.007	0.110	0.110
		−0.025	0.110	0.111	−0.049	0.138	0.140	−0.075	0.156	0.161
		-0.041	0.090	0.092	-0.116	0.086	0.099	-0.125	0.09	0.106
	γ_5	−0.026	0.212	0.212	−0.055	0.199	0.202	−0.038	0.193	0.194
		0.019	0.245	0.245	0.020	0.287	0.287	0.045	0.268	0.270
		-0.051	0.201	0.204	-0.060	0.158	0.162	-0.015	0.179	0.179
	γ_6	−0.024	0.176	0.177	−0.026	0.172	0.172	−0.038	0.163	0.165
		0.031	0.203	0.204	0.068	0.213	0.217	0.063	0.209	0.213
		-0.009	0.184	0.184	-0.008	0.158	0.158	0.017	0.187	0.187
	β_1	0.004	0.008	0.008	0.000	0.019	0.019	−0.001	0.035	0.035
		0.004	0.008	0.008	0.003	0.019	0.019	0.001	0.037	0.037
		0.003	0.008	0.008	-0.003	0.018	0.018	-0.006	0.034	0.034
	β_2	−0.011	0.019	0.019	−0.001	0.049	0.049	0.010	0.095	0.096
		−0.011	0.019	0.019	0.004	0.053	0.053	0.021	0.103	0.104
		-0.007	0.019	0.019	0.042	0.043	0.044	0.060	0.079	0.083
	β_3	0.000	0.033	0.033	0.014	0.083	0.083	0.021	0.144	0.144
		−0.001	0.033	0.033	0.006	0.088	0.088	0.007	0.155	0.155
		0.000	0.032	0.032	-0.005	0.074	0.074	-0.002	0.128	0.128

Table S4: Bias, variance and MSE of $\hat{\gamma}$ and $\hat{\beta}$ for the proposed 2-step approach (first rows), `smcure` (second rows) and the multivariate presmoothing of Musta et al. (2022) (third rows) in Model 4 for $n = 200$.

		Scenario 1			Scenario 2			Scenario 3		
n	Par.	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
400	γ_1	0.042	0.115	0.116	0.016	0.105	0.105	0.003	0.118	0.118
		0.056	0.124	0.127	0.036	0.119	0.121	0.007	0.125	0.125
		-0.015	0.118	0.118	-0.058	0.111	0.114	-0.042	0.12	0.122
	γ_2	0.025	0.027	0.027	0.045	0.036	0.038	0.047	0.036	0.039
		−0.029	0.028	0.029	−0.040	0.032	0.034	−0.038	0.032	0.034
		0.092	0.028	0.036	0.172	0.022	0.052	0.146	0.026	0.047
	γ_3	0.002	0.064	0.064	−0.022	0.062	0.062	−0.018	0.063	0.063
		0.024	0.072	0.072	0.009	0.073	0.073	0.012	0.074	0.074
		-0.062	0.050	0.054	-0.088	0.049	0.056	-0.080	0.056	0.062
	γ_4	0.010	0.045	0.045	0.016	0.054	0.054	0.018	0.061	0.061
		−0.014	0.049	0.049	−0.019	0.058	0.059	−0.018	0.069	0.070
		-0.021	0.046	0.047	-0.085	0.046	0.053	-0.090	0.050	0.058
	γ_5	0.000	0.101	0.101	−0.019	0.097	0.097	−0.012	0.107	0.108
		0.034	0.111	0.112	0.033	0.111	0.112	0.038	0.119	0.120
		-0.029	0.099	0.100	-0.039	0.093	0.094	-0.002	0.095	0.095
	γ_6	−0.039	0.084	0.086	−0.045	0.081	0.084	−0.041	0.090	0.092
		−0.002	0.094	0.094	0.015	0.091	0.092	0.018	0.098	0.098
		-0.026	0.089	0.090	-0.029	0.092	0.093	-0.017	0.106	0.106
	β_1	0.000	0.004	0.004	0.003	0.009	0.009	−0.002	0.015	0.015
		0.000	0.004	0.004	0.005	0.009	0.009	0.001	0.015	0.015
		-0.001	0.004	0.004	-0.001	0.009	0.009	-0.004	0.014	0.014
	β_2	−0.003	0.009	0.009	−0.007	0.023	0.023	−0.011	0.045	0.045
		−0.002	0.009	0.009	−0.004	0.024	0.024	−0.006	0.047	0.047
		0.000	0.009	0.009	0.031	0.02	0.021	0.043	0.037	0.039
	β_3	0.005	0.016	0.017	0.012	0.039	0.039	0.015	0.072	0.072
		0.004	0.017	0.017	0.006	0.040	0.040	0.006	0.074	0.074
		0.005	0.016	0.016	0.002	0.036	0.036	-0.004	0.063	0.063

Table S5: Bias, variance and MSE of $\hat{\gamma}$ and $\hat{\beta}$ for the proposed 2-step approach (first rows), `smcure` (second rows) and the multivariate presmoothing of Musta et al. (2022) (third rows) in Model 4 for $n = 400$.

	`smcure`			Step 2			Step 3			Step 4		
Par.	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
γ_1	0.241	0.531	0.589	−0.007	0.246	0.246	−0.007	0.247	0.247	−0.010	0.247	0.247
γ_2	0.201	0.302	0.342	−0.006	0.219	0.219	−0.005	0.217	0.217	−0.006	0.219	0.219
γ_3	0.189	0.713	0.748	−0.046	0.405	0.407	−0.047	0.417	0.419	−0.052	0.425	0.427
β_1	−0.009	0.016	0.016	−0.006	0.015	0.015	−0.006	0.015	0.015	−0.006	0.015	0.015
β_2	0.003	0.036	0.036	0.008	0.035	0.035	0.009	0.035	0.035	0.009	0.035	0.035

Table S6: Bias, variance and MSE of $\hat{\gamma}$ and $\hat{\beta}$ for `smcure`, the proposed 2-step approach and its iterations (3-step and 4-step) for Model 1, scenario 1 and $n = 200$.

		$n = 200$			$n = 400$		
Mod.	Par.	Bias	Var.	MSE	Bias	Var.	MSE
5	γ_1	−0.154	0.187	0.211	−0.087	0.100	0.107
		0.514	0.393	0.657	0.437	0.152	0.343
	γ_2	−0.304	0.277	0.370	0.213	0.144	0.189
		0.623	0.464	0.852	0.507	0.184	0.441
	γ_3	−0.232	0.203	0.257	−0.166	0.108	0.136
		0.297	0.655	0.744	0.235	0.176	0.231
	γ_4	0.207	0.188	0.231	0.132	0.112	0.130
		−0.326	0.377	0.483	−0.281	0.166	0.245
	β_1	−0.281	0.051	0.130	−0.296	0.025	0.112
		−0.354	0.056	0.181	−0.353	0.027	0.151
	β_2	−0.137	0.043	0.062	−0.144	0.022	0.043
		−0.174	0.046	0.076	−0.169	0.023	0.052
	β_3	0.221	0.051	0.099	0.210	0.023	0.067
		0.257	0.054	0.120	0.235	0.024	0.079
4	γ_1	0.036	0.250	0.251	0.016	0.116	0.116
		0.097	0.360	0.370	0.041	0.133	0.135
	γ_2	0.059	0.066	0.070	0.050	0.036	0.039
		−0.089	0.092	0.100	−0.038	0.038	0.040
	γ_3	−0.029	0.150	0.151	−0.023	0.075	0.075
		0.025	0.222	0.223	0.008	0.091	0.091
	γ_4	0.005	0.109	0.109	0.018	0.055	0.055
		−0.059	0.149	0.152	−0.018	0.061	0.062
	γ_5	−0.057	0.204	0.207	−0.022	0.101	0.102
		0.022	0.307	0.308	0.032	0.120	0.121
	γ_6	−0.029	0.204	0.205	−0.047	0.103	0.105
		0.075	0.287	0.293	0.014	0.122	0.122
	β_1	−0.025	0.028	0.029	−0.017	0.011	0.012
		−0.004	0.030	0.030	−0.002	0.012	0.012
	β_2	0.017	0.068	0.069	0.007	0.032	0.032
		0.010	0.074	0.074	0.002	0.033	0.033
	β_3	0.004	0.054	0.054	−0.001	0.024	0.024
		0.011	0.058	0.059	0.002	0.025	0.025
	β_4	0.013	0.094	0.095	0.012	0.043	0.043
		0.004	0.103	0.103	0.005	0.044	0.044
	β_5	−0.001	0.112	0.112	0.011	0.047	0.047
		−0.019	0.119	0.119	0.000	0.049	0.049

Table S7: Bias, variance and MSE of $\hat{\gamma}$ and $\hat{\beta}$ for the proposed 2-step approach (first rows) and `smcure` (second rows) in Model 5 and Model 4 with misspecification.

References

Xiaohong Chen, Oliver Linton, and Ingrid Van Keilegom (2003) Estimation of semiparametric models when the criterion function is not426 smooth, Econometrica, 71(5):1591–1608.

Yunling Du and MG Akritas. Uniform strong representation of the conditional kaplan-meier process. Mathematical Methods of434 Statistics, 11(2):152–182, 2002

Olivier Lopez, Valentin Patilea, and Ingrid Van Keilegom (2013) Single index regression models in the presence of censoring depending on the covariates, Bernoulli, 19(3):721–747.

Eni Musta, Valentin Patilea, and Ingrid Van Keilegom (2022) A presmoothing approach for estimation in the semiparametric Cox mixture cure model. Bernoulli, 28(4):2689–2715.

Aad W van der Vaart and John A Wellner. Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag,486 New York, 1996

Ingrid Van Keilegom and Michael G. Akritas (1999) Transfer of tail information in censored regression models, Annals of Statistics, 27: 1745–1784.