Biometrika (2013), **??**, ??, *pp*. 1–9

Printed in Great Britain

Advance Access publication on

Supplementary Materials for Doubly Robust Learning for Estimating Individualized Treatment with Censored Data

BY Y. Q. ZHAO

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, 53792, U.S.A

yqzhao@biostat.wisc.edu

D. ZENG

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, U.S.A

dzeng@bios.unc.edu

E. B. LABER, R. SONG

Department of Statistics, North Carolina State University, Raleigh, North Carolina, 27695, U.S.A

eblaber@ncsu.edu rsong@ncsu.edu

M. YUAN

Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin, 53792, U.S.A

myuan@stat.wisc.edu

M. R. KOSOROK

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, U.S.A

kosorok@unc.edu

1. PROOF OF LEMMA 1

We denote the true survival function for $\widetilde{T}$ and the true survival function for C by $S^*_{\widetilde{T}}(t \mid A, X)$ and $S^*_C(t \mid A, X)$ respectively. If the model for the survival time is correct, then $E^m_{\widetilde{T}}[t \mid A, X]$ is the true mean $E^*_{\tilde{T}}[t \mid A, X]$. Therefore,

$$
\begin{aligned}
&E\left[\frac{\Delta Y}{S_C^m(Y \mid A, X)}\right.\\
&\quad\left.-\int_{0\le t<\tau} E^m_{\tilde{T}}(T \mid T>t, A, X)\left\{\frac{dN_C(t)}{S_C^m(t \mid A, X)}+I(Y\ge t)\frac{dS_C^m(t \mid A, X)}{S_C^m(t \mid A, X)^2}\right\} \mid A, X\right]\\
&=E\left[\frac{S_C^*(t \mid A, X)T}{S_C^m(t \mid A, X)}-\int_{0\le t<\tau} E^*_{\tilde{T}}(T \mid T>t, A, X)S^*_{\tilde{T}}(t \mid A, X)d\left\{\frac{S_C^*(t \mid A, X)}{S_C^m(t \mid A, X)}\right\} \mid A, X\right]\\
&=E\left[\frac{S_C^*(t \mid A, X)T}{S_C^m(t \mid A, X)}\right.\\
&\quad\left.-\int_{0<t<\tau}\left\{\int_{t<u<\tau} u f^*_{\tilde{T}}(u \mid A, X)du+\tau S^*_{\tilde{T}}(\tau \mid A, X)\right\}d\left\{\frac{S_C^*(t \mid A, X)}{S_C^m(t \mid A, X)}\right\} \mid A, X\right]\\
&=\int_{0<u<\tau} u f^*_{\tilde{T}}(u \mid A, X)du+\tau S^*_{\tilde{T}}(\tau \mid A, X)\\
&=E(T \mid A, X),
\end{aligned}
$$

and $V^m(\mathcal{D}) = E[E(T \mid A, X)I\{A = \mathcal{D}(X)\}/\pi(A; X)] = V(\mathcal{D})$. On the other hand, if the model for the censoring time is correct, then $S_C^m(t \mid A, X)$ is the true $S_C^*(t \mid A, X)$. We obtain

$$
\begin{aligned}
&E\left[\frac{\Delta Y}{S_C^m(Y|A, X)}\right.\\
&\quad\left.-\int E^m_{\tilde{T}}(T \mid T>t, A, X)\left\{\frac{dN_C(t)}{S_C^m(t \mid A, X)}+I(Y\ge t)\frac{dS_C^m(t \mid A, X)}{S_C^m(t \mid A, X)^2}\right\} \mid A, X\right]\\
&=E\left[\frac{\Delta Y}{S_C^*(Y|A, X)}\right.\\
&\quad\left.-\int E^m_{\tilde{T}}(T \mid T>t, A, X)\left\{\frac{dN_C(t)}{S_C^*(t \mid A, X)}+I(Y\ge t)\frac{dS_C^*(t \mid A, X)}{S_C^*(t \mid A, X)^2}\right\} \mid A, X\right]\\
&=E\left\{\frac{\Delta Y}{S_C^*(Y|A, X)} \mid A, X\right\}\\
&=E(T \mid A, X).
\end{aligned}
$$

Thus $V^m(\mathcal{D}) = V(\mathcal{D})$.

2. PROOF OF THEOREM 1

First, it can be established that

$$\begin{aligned}
V(f^*) - V(\widehat{f}) &= V(f^*) - \sup_{f\in\mathcal{F}} V_R(f, S_C^m, E_{\widetilde{T}}^m) + \sup_{f\in\mathcal{F}} V_R(f, S_C^m, E_{\widetilde{T}}^m) - V_R(\widehat{f}, S_C^m, E_{\widetilde{T}}^m)\\
&\quad + V_R(\widehat{f}, S_C^m, E_{\widetilde{T}}^m) - V(\widehat{f})\\
&\le V(f^*) - V_R(f^*, S_C^m, E_{\widetilde{T}}^m) + \{\sup_{f\in\mathcal{F}} V_R(f, S_C^m, E_{\widetilde{T}}^m) - V_R(\widehat{f}, S_C^m, E_{\widetilde{T}}^m)\}\\
&\quad + V_R(\widehat{f}, S_C^m, E_{\widetilde{T}}^m) - V(\widehat{f})\\
&\le \sup_{f\in\mathcal{F}} V_R(f, S_C^m, E_{\widetilde{T}}^m) - V_R(\widehat{f}, S_C^m, E_{\widetilde{T}}^m)\\
&\quad + 2\sup_{f\in\mathcal{F}} |V_R(f, S_C^*, E_{\widetilde{T}}^*) - V_R(f, S_C^m, E_{\widetilde{T}}^m)|.
\end{aligned}$$

The first inequality follows since $V_R(f^*, S_C^m, E_{\widetilde{T}}^m) \le \sup_{f\in\mathcal{F}} V_R(f, S_C^m, E_{\widetilde{T}}^m)$. According to Lemma 2(a), $V_R(f^m, S_C^m, E_{\widetilde{T}}^m) = \sup_{f\in\mathcal{F}} V_R(f, S_C^m, E_{\widetilde{T}}^m)$, where $f^m = \text{argmin}_{f\in\mathcal{F}} E\{L_\phi(f, S_C^m, E_T^m)\}$. Hence, it suffices to derive the convergence rate of $V_R(f^m, S_C^m, E_{\widetilde{T}}^m) - V_R(\widehat{f}, S_C^m, E_{\widetilde{T}}^m)$.

Let $f_{\lambda_n}^m = \text{argmin}_{f\in\mathcal{H}_k}[E\{R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m)\phi\{Af(X)\}/\pi(A; X)\} + \lambda_n\|f\|_k^2]$. Then,

$$\begin{aligned}
n^{-1}\sum_{i=1}^n \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A_i\widehat{f}(X_i)\}}{\pi(A_i; X_i)} + \lambda_n\|\widehat{f}\|_k^2 \le n^{-1}\sum_{i=1}^n &\frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A_i f_{\lambda_n}^m(X_i)\}}{\pi(A_i; X_i)}\\
&+ \lambda_n\|f_{\lambda_n}^m\|_k^2.
\end{aligned}$$

Equivalently,

$$\begin{aligned}
&V_R(f^m, S_C^m, E_{\widetilde{T}}^m) - V_R(\widehat{f}, S_C^m, E_{\widetilde{T}}^m)\\
&\le a(\lambda_n) + \Bigg(n^{-1}\sum_{i=1}^n \Big[\lambda_n\|\widehat{f}\|_k^2\\
&\quad + \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A_i\widehat{f}(X_i)\}}{\pi(A_i; X_i)} - \lambda_n\|f_{\lambda_n}^m\|_k^2 - \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A_i f_{\lambda_n}^m(X_i)\}}{\pi(A_i; X_i)}\Big]\\
&\quad - E\Big[\lambda_n\|\widehat{f}\|_k^2 + \frac{R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A\widehat{f}(X)\}}{\pi(A; X)} - \lambda_n\|f_{\lambda_n}^m\|_k^2 - \frac{R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{Af_{\lambda_n}^m(X)\}}{\pi(A; X)}\Big]\Bigg)\\
&\quad + E\left[\frac{R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m)\phi\{A\widehat{f}(X)\}}{\pi(A; X)} - \frac{R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A\widehat{f}(X)\}}{\pi(A; X)}\right]\\
&\quad + E\left[\frac{R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{Af_{\lambda_n}^m(X)\}}{\pi(A; X)} - \frac{R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m)\phi\{Af_{\lambda_n}^m(X)\}}{\pi(A; X)}\right]\\
&= a(\lambda_n) + (I) + (II) + (III).
\end{aligned}$$

To bound (II) and (III), we consider the class of functions

$$\mathcal{B} = \{R\{Y, \Delta, S_C(\beta_C, \Lambda_{C0}), E_{\widetilde{T}}(\beta_T, \Lambda_{T0})\} : \beta_C \in \mathbb{R}^d, \|\beta_C - \beta_C^m\| < \delta_0,\ \beta_T \in \mathbb{R}^d, \\ \|\beta_T - \beta_T^m\| < \delta_0, \Lambda_{C0}, \Lambda_{T0} \text{ are bounded monotone functions in } [0, \tau], \\ \sup_t |\Lambda_{C0}(t) - \Lambda_{C0}^m(t)| < \delta_0,\ \sup_t |\Lambda_{T0}(t) - \Lambda_{T0}^m(t)| < \delta_0\},$$

where δ_0 is a small constant, and $\beta_T^m, \beta_C^m, \Lambda_{C0}^m(t), \Lambda_{T0}^m(t)$ are the limits of $\widehat{\beta}_T, \widehat{\beta}_C, \widehat{\Lambda}_{C0}$ and $\widehat{\Lambda}_{T0}$ based on the Cox models. Then $|R\{Y, \Delta, S_C(\beta_C, \Lambda_{C0}), E_{\widetilde{T}}(\beta_C, \Lambda_{C0})\}|/\pi(A, X)$ can be bounded from above by a constant, say M.

Since

$$n^{-1} \sum_{i=1}^n \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A_i \widehat{f}(X_i)\}}{\pi(A_i; X_i)} + \lambda_n \|\widehat{f}\|_k^2 \le n^{-1} \sum_{i=1}^n \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi(0)}{\pi(A_i; X_i)},$$

it follows that

$$\|\widehat{f}\|_k \le \Big[\lambda_n^{-1} n^{-1} \sum_{i=1}^n \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})}{\pi(A_i; X_i)}\Big]^{1/2} \le M \lambda_n^{-1/2},$$

where M^2 is a constant bounding the empirical average given that the outcomes are bounded. Similarly, $\|f_{\lambda_n}^m\|_k \le M\lambda_n^{-1/2}$, given that

$$\lambda_n \|f_{\lambda_n}^m\|_k^2 \le \inf_{f \in \mathcal{H}_k} \lambda_n \|f\|_k^2 + E\left[\frac{R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m)\phi\{Af(X)\}}{\pi(A; X)}\right] \le E\left\{\frac{R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m)\phi(0)}{\pi(A; X)}\right\}.$$

For every $f \in M\lambda_n^{-1/2} B_{\mathcal{H}_k}$, $|(1 - Af)^+| \le 1 + M\lambda_n^{-1/2} = B$. Thus,

$$\begin{aligned} &\left|E\left[\frac{R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m)\phi\{Af(X)\}}{\pi(A; X)} - \frac{R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{Af(X)\}}{\pi(A; X)}\right]\right| \\ &\le E\left[\left|\frac{\{1 - Af(X)\}^+}{\pi(A; X)}\right| |R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m) - R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})|\right] \\ &\le BE\{|R(Y, \Delta, S_C^m, E_{\widetilde{T}}^m) - R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})|\} \\ &= O_p(n^{-\gamma} \lambda_n^{-1/2}). \end{aligned}$$

We use empirical process theory to bound (I). Define the functional class

$$\mathcal{L} = \Big\{\lambda_n \|f\|_k^2 + \frac{R(Y, \Delta, S_C, E_{\widetilde{T}})\phi\{Af(X)\}}{\pi(A; X)} - \frac{R(Y, \Delta, S_C, E_{\widetilde{T}})\phi\{Af_{\lambda_n}^m(X)\}}{\pi(A; X)} - \lambda_n \|f_{\lambda_n}^m\|_k^2, \\ f \in M\lambda_n^{-1/2} B_{\mathcal{H}_k}, R(Y, \Delta, S_C, E_{\widetilde{T}}) \in \mathcal{B}\Big\},$$

and

$$\mathcal{G} = \{E(l) - l : E(l) = \varepsilon, l \in \mathcal{L}\}.$$

Let $Z = \sup_{g \in \mathcal{G}} n^{-1} \sum_{i=1}^n g(X_i)$. Since $E(g) = 0, g \in \mathcal{G}$, it follows from Lemma S.1, by setting $\rho = 1$, that

$$\mathrm{pr}\{Z \ge 2E(Z) + \sigma(Kb)^{1/2} n^{-1/2} + 2KBbn^{-1}\} \le e^{-b},$$

where $B = O(\lambda_n^{-1/2})$. Furthermore, $\sigma^2 \le c_n' \varepsilon$ following the arguments for proving Theorem 3.4 in Zhao et al. (2012), given that $E(l^2) \le c_n' E(l)$, where $c_n' = O(\lambda_n^{-1})$. In addition, for $f \in M\lambda_n^{-1/2} B_{\mathcal{H}_k}$,

$$E(Z) = E\Big\{\sup_{g\in\mathcal{G}} n^{-1}\sum_{i=1}^n g(X_i)\Big\} = E\Big[\sup_{E(l^2)\le c_n'\varepsilon}\Big|E\{l(X)\} - n^{-1}\sum_{i=1}^n l(X_i)\Big|\Big].$$

Since $|\beta_C - \beta_C^m|$ and $|\beta_T - \beta_T^m|$ are bounded by δ_0, they lie in a hypercube of $\mathbb{R}^{2d}$. Moreover, $\{\Lambda_{C0} : \sup_t |\Lambda_{C0}(t) - \Lambda_{C0}^m(t)| < \delta_0\}$ is a class of monotone functions, so is $\{\Lambda_{T0} : \sup_t |\Lambda_{T0}(t) - \Lambda_{T0}^m(t)| < \delta_0\}$. The function in $\mathcal{B}$ is Lipschitz continuous with respect to all these parameters and the Lipschitz constant is less than a constant W. There exists a constant K, depending on d, such that the bracketing number for $\mathcal{B}$ satisfies $N_{[\,\cdot\,]}\{\mathcal{B}, \epsilon W, L_2(P)\} \le K(\delta_0/\epsilon)^{2d+2}$. According to (10) in the main text, $\sup_{P_n} \log N\{\mathcal{G}, \epsilon, L_2(P_n)\} \le c_n \epsilon^{-p}$, and threfore

$$E(Z) \le c_p M \lambda_n^{-\frac{1}{2}} \max\Big\{(M^2\lambda_n c_n'\varepsilon)^{(2-p)/4} {c_n}^{1/2} n^{-1/2}, {c_n}^{2/(2+p)} n^{-2/(2+p)}\Big\},$$

where c_p is a constant depending on p. See Proposition 5.5 in Steinwart & Scovel (2007) and references therein. Consequently,

$$\begin{aligned}\mathrm{pr}\Big(\Big|\; n^{1/2}\ \Big[n^{-1}\sum_{i=1}^n l(X_i) - E\{l(X)\}\Big]\Big| &> (c_n'\varepsilon K b)^{1/2} n^{-1/2} + 2KBbn^{-1} \\ &+\ \ 2c_p M\lambda_n^{-1/2}\max\Big\{(M^2\lambda_n c_n'\varepsilon)^{(2-p)/4} {c_n}^{1/2} n^{-1/2}, {c_n}^{2/(2+p)} n^{-2/(2+p)}\Big\}\Big) \le e^{-b}.\end{aligned}$$

Let $\varepsilon^* > 0$ be the largest number that satisfies

$$\varepsilon = 2c_p M\lambda_n^{-1/2}(M^2\lambda_n c_n'\varepsilon)^{(2-p)/4} {c_n}^{1/2} n^{-1/2} + (c_n'\varepsilon K b)^{1/2} n^{-1/2}.$$

If $2c_p M\lambda_n^{-1/2}(M^2\lambda_n c_n'\varepsilon^*)^{(2-p)/4} {c_n}^{1/2} n^{-1/2} \ge (c_n'\varepsilon^* K b)^{1/2} n^{-1/2}$, then $\varepsilon^* \le 4c_p M\lambda_n^{-1/2}(M^2\lambda_n c_n'\varepsilon^*)^{(2-p)/4} {c_n}^{1/2} n^{-1/2}$, and thus

$$\varepsilon^* \le \Big\{4c_p M\lambda_n^{-1/2}(M^2\lambda_n c_n')^{(2-p)/4} {c_n}^{1/2} n^{-1/2}\Big\}^{4/(p+2)}.$$

Conversely, if $2c_p M\lambda_n^{-1/2}(M^2\lambda_n c_n'\varepsilon^*)^{(2-p)/4} {c_n}^{1/2} n^{-1/2} \le (c_n'\varepsilon^* K b)^{1/2} n^{-1/2}$, then $\varepsilon^* \le c_n' K b n^{-1}$.

Given that $\mathcal{L}$ is convex, if $l \in \mathcal{L}$ satisfies $n^{-1}\sum_{i=1}^n l(X_i) \le \alpha\varepsilon$ and $E\{l(X)\} \ge \varepsilon$, there exists $l' \in \mathcal{L}$ such that $n^{-1}\sum_{i=1}^n l'(X_i) \le \alpha\varepsilon$ and $E\{l'(X)\} = \varepsilon$. Thus, with probability at least $1 - e^{-b}$, every $l \in \mathcal{L}$ with $n^{-1}\sum_{i=1}^n l(X_i) \le \alpha\varepsilon$ satisfies $El \le \varepsilon$ (Bartlett et al., 2006; Steinwart & Scovel, 2007). Since

$$\begin{aligned}n^{-1}\sum_{i=1}^n \Big[\lambda_n\|\widehat{f}\|_k^2 &+ \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A_i\widehat{f}(X_i)\}}{\pi(A_i; X_i)} \\ &- \lambda_n\|f_{\lambda_n}^m\|_k^2 - \frac{R(Y_i, \Delta_i, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A_i f_{\lambda_n}^m(X_i)\}}{\pi(A_i; X_i)}\Big] \le 0 < \alpha\varepsilon,\end{aligned}$$

with probability at least $1 - e^{-b}$,

$$E\Big[\lambda_n\|\widehat{f}\|_k^2 + \frac{R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A\widehat{f}(X)\}}{\pi(A; X)} - \lambda_n\|f_{\lambda_n}^m\|_k^2 - \frac{R(Y, \Delta, \widehat{S}_C, \widehat{E}_{\widetilde{T}})\phi\{A f_{\lambda_n}^m(X)\}}{\pi(A; X)}\Big] \le \varepsilon.$$

It follows that,

$$
\begin{aligned}
\mathrm{pr}\Big[|(I)| > \Big\{4c_p M\lambda_n^{-1/2}(M^2\lambda_n c_n')^{(2-p)/4} {c_n}^{1/2} n^{-1/2}\Big\}^{4/(p+2)} \\
+ c_p M\lambda_n^{-1/2} {c_n}^{2/(2+p)} n^{-2/2+p} + c_n' K b n^{-1} + 2KBbn^{-1}\Big] \le 2e^{-b},
\end{aligned}
$$

with $c_n' = O(\lambda_n^{-1})$ and $B = O(\lambda_n^{-1/2})$. Using M_p as a new constant depending on p, we subsequently obtain the desired results.

LEMMA S.1 $\{$LEMMA A.1 FROM BARTLETT ET AL. (2006)$\}$. *There is an absolute constant K for which the following holds. Let $\mathcal{G}$ be a class of functions defined on $\mathcal{X}$ with $\sup_{g\in\mathcal{G}} \|g\|_\infty \le b$. Suppose that P is a probability distribution such that for every $g \in \mathcal{G}, Eg = 0$. Let $X_1, ..., X_n$ be independent random variables distributed according to P and set $\sigma^2 = \sup_{g\in\mathcal{G}} var(g)$. Define*

$$
Z = \sup_{g\in\mathcal{G}} n^{-1} \sum_{i=1}^{n} g(X_i).
$$

Then, for every $x > 0$ and every $\rho > 0$,

$$
\mathrm{pr}\Big\{Z \ge (1+\rho)E(Z) + \sigma(Kx)^{1/2} n^{-1/2} + K(1+\rho^{-1})bxn^{-1}\Big\} \le e^{-x}.
$$

3. CALCULATION OF PSEUDO-OUTCOME USING COX PROPORTIONAL HAZARDS MODELS

Estimates $\widehat{\beta}_T$ and $\widehat{\beta}_C$ are obtained by fitting a Cox model using basis Z_T for survival time $\widetilde{T}$, and basis Z_C for censoring time. Subsequently, for $0 < t < \tau$,

$$
\begin{aligned}
&\widehat{E}_{\widetilde{T}}(T \mid T > t, A_i, X_i) \\
&= -\frac{\int_{t<u<\tau} u d\widehat{S}_{\widetilde{T}}(u \mid A_i, X_i)}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)} + \frac{\tau \widehat{S}_{\widetilde{T}}(\tau \mid A_i, X_i)}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)} \\
&= \frac{\int_{t<u<\tau} u \widehat{S}_{\widetilde{T}}(u \mid A_i, X_i) d\widehat{\Lambda}_{\widetilde{T}}(u \mid A_i, X_i)}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)} + \frac{\tau \widehat{S}_{\widetilde{T}}(\tau \mid A_i, X_i)}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)} \\
&= \frac{1}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)} \sum_{k=1}^{n} \int_{t<u<\tau} \frac{u \widehat{S}_{\widetilde{T}}(u \mid A_i, X_i) e^{\widehat{\beta}'_T Z_{Ti}} dN_{\widetilde{T}k}(u)}{\sum_{j=1}^{n} I(Y_j \ge u) e^{\widehat{\beta}'_T Z_{Tj}}} + \frac{\tau \widehat{S}_{\widetilde{T}}(\tau \mid A_i, X_i)}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)} \\
&= \frac{e^{\widehat{\beta}'_T Z_{Ti}}}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)} \sum_{k=1}^{n} \frac{Y_k \widehat{S}_{\widetilde{T}}(Y_k | A_i, X_i) I(t < Y_k < \tau)\Delta_k}{\sum_{j=1}^{n} I(Y_j \ge Y_k) e^{\widehat{\beta}'_T Z_{Tj}}} + \frac{\tau \widehat{S}_{\widetilde{T}}(\tau \mid A_i, X_i)}{\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i)}.
\end{aligned}
$$

The Breslow estimator for baseline hazard function $\Lambda_{\widetilde{T}0}(t)$ is

$$
\widehat{\Lambda}_{\widetilde{T}0}(t) = \int_0^t \frac{\sum_{k=1}^{n} dN_{\widetilde{T}k}(u)}{\sum_{j=1}^{n} I(Y_j \ge u) e^{\widehat{\beta}'_T Z_{Tj}}},
$$

and $\widehat{\Lambda}_{\widetilde{T}}(t \mid A_i, X_i) = \exp(\widehat{\beta}'_T Z_{Ti})\widehat{\Lambda}_{\widetilde{T}0}(t)$. Estimates for $S_{\widetilde{T}}(t \mid A_i, X_i)$ are obtained via

$$
\widehat{S}_{\widetilde{T}}(t \mid A_i, X_i) = \exp\{-\widehat{\Lambda}_{\widetilde{T}0}(t)\}^{\exp(\widehat{\beta}'_T Z_{Ti})}.
$$

The pseudo outcome $\widehat{R}_i$ using doubly robust methods is calculated as

$$
\begin{aligned}
\widehat{R}_i =& \frac{\Delta_i Y_i}{\widehat{S}_C(Y_i \mid A_i, X_i)} \\
& - \int \widehat{E}_{\widetilde{T}}(T \mid T > t, A_i, X_i) \left\{ \frac{dN_{Ci}(t)}{\widehat{S}_C(t \mid A_i, X_i)} - I(Y_i \geq t) \frac{d\widehat{\Lambda}_C(t \mid A_i, X_i)}{\widehat{S}_C(t \mid A_i, X_i)} \right\} \\
=& \frac{\Delta_i Y_i}{\widehat{S}_C(Y_i \mid A_i, X_i)} - \frac{\widehat{E}_{\widetilde{T}}(T \mid T > Y_i, A_i, X_i)(1 - \Delta_i) I(Y_i < \tau)}{\widehat{S}_C(Y_i \mid A_i, X_i)} \\
& + \sum_{k=1}^{n} \int \frac{\widehat{E}_{\widetilde{T}}(T \mid T > t, A_i, X_i) I(Y_i \geq t) e^{\widehat{\beta}'_C Z_{Ci}} dN_{Ck}(t)}{\widehat{S}_C(t \mid A_i, X_i) \sum_{j=1}^{n} I(Y_j \geq t) e^{\widehat{\beta}'_C Z_{Cj}}} \\
=& \frac{\Delta_i Y_i}{\widehat{S}_C(Y_i \mid A_i, X_i)} - \frac{\widehat{E}_{\widetilde{T}}(T \mid T > Y_i, Z_{Ti}) I(Y_i < \tau)(1 - \Delta_i)}{\widehat{S}_C(Y_i \mid A_i, X_i)} \\
& + \sum_{k=1}^{n} \frac{\widehat{E}_{\widetilde{T}}(T \mid T > Y_k, Z_{Ti}) I(Y_i \geq Y_k) I(Y_k < \tau) e^{\widehat{\beta}'_C Z_{Ci}} (1 - \Delta_k)}{\widehat{S}_C(Y_k \mid A_i, X_i) \sum_{j=1}^{n} I(Y_j \geq Y_k) e^{\widehat{\beta}'_C Z_{Cj}}}.
\end{aligned}
$$

4. ADDITIONAL SIMULATION RESULTS

Figures 1 and 2 are the boxplots of values of estimated rules for the simulation study in the main text, when $n = 200$ and $n = 400$. Similarly as the results under $n = 100$, they show favorable performances of inverse censoring weighted and doubly robust outcome weighted learning.

We present an additional simulation example. In this example, 30 independent covariates are generated from the uniform distribution between [0, 1]. We generate both survival time and censoring time from Cox models with nonlinear effects. The survival time T is the minimum of $\tau = 2$ and $\widetilde{T}$, where

$$
\begin{aligned}
\lambda_{\widetilde{T}}(t \mid A, X) = \lambda_{\widetilde{T}0}(t) \exp[&-0{\cdot}5 \sin(\pi X_2) + 0{\cdot}5 \sin(\pi X_3) \\
&+ \{0{\cdot}5 - 0{\cdot}5 \sin(\pi X_1) - 0{\cdot}5 \sin(\pi X_2)\} A],
\end{aligned}
$$

and $\lambda_{\widetilde{T}0}(t) = 2t$. The hazard rate function of C is

$$
\begin{aligned}
\lambda_C(t \mid A, X) = \lambda_{C0}(t) \exp[&0{\cdot}3 \cos(2\pi X_1) + 0{\cdot}5 \cos(2\pi X_2) + 0{\cdot}5 \cos(2\pi X_3) - \cos(2\pi X_4) \\
&+ \{1 - 1{\cdot}5 \cos(2\pi X_1) - 1{\cdot}5 \cos(2\pi X_2) + \cos(2\pi X_3)\} A],
\end{aligned}
$$

where $\lambda_{C0}(t) = 2t$. The censoring percentage is around 58%. The optimal decision boundary is nonlinear with $\mathcal{D}^*(X) = -\{0{\cdot}5 - 0{\cdot}5\sin(\pi X_1) - 0{\cdot}5\sin(\pi X_2)\}$. We implement the proposed methods using both linear and Gaussian kernels.

We use Cox regression to estimate survival and censoring probability, but with different working models. Specifically, a correctly specified model requires that we use the true sets of covariates in model fitting. If the model is incorrectly specified for survival time or censoring time, we use $(X_1, \ldots, X_{20}, A, X_1 A, \ldots, X_{20} A)$ as the basis. We plot the boxplots for the estimated values produced from different methods using 1000 replicates in Fig. 3. Since the underlying generative model is the same for different working models, we present all results in one figure for a better comparison in Fig. 3 instead of four subfigures as we did in the main text. As anticipated, the Cox regression model with correct basis leads to the highest value. However, the truth is usually unknown in reality. Indeed, there is severe bias in the values using the rules con-

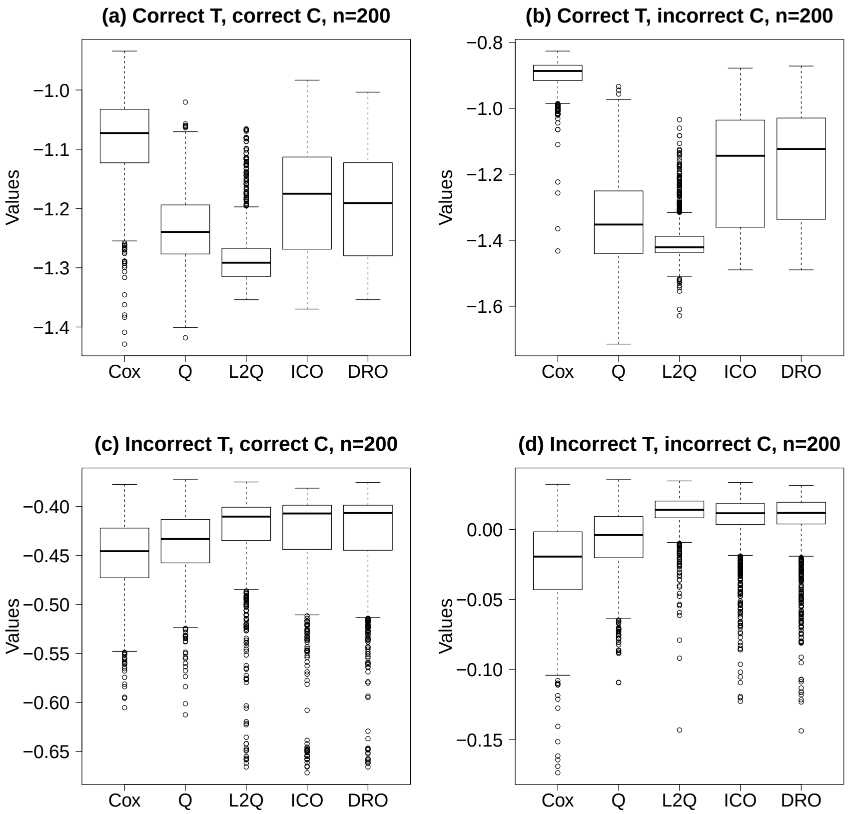

Fig. 1. Boxplots of values of estimated rules using different methods, representing the logarithm of the survival time with higher values being more preferable, $n = 200$. Cox: Cox model; Q: inverse censoring weighted Q-learning; L2Q: inverse censoring weighted L_2 Q-learning; ICO: inverse censoring weighted outcome weighted learning with linear kernel; DRO: doubly robust outcome weighted learning with linear kernel.

structed from the Cox model with incorrect basis. All methods with linear basis do not perform well, and the performances do not improve over increasing sample sizes. However, the gain from using the Gaussian kernel is pronounced, since the induced reproducing kernel Hilbert space is flexible enough to approximate the nonlinear treatment decision rule. When the censoring model is correctly specified, inverse censoring weighted outcome weighted learning with a Gaussian kernel yields a competitive performance. Doubly robust outcome weighted learning can further reduce variabilities of the decision rules with a correct survival model. Although variance tends to be larger in the values due to the estimated rules using the Gaussian kernel, the performances are satisfactory compared with the linear kernel.

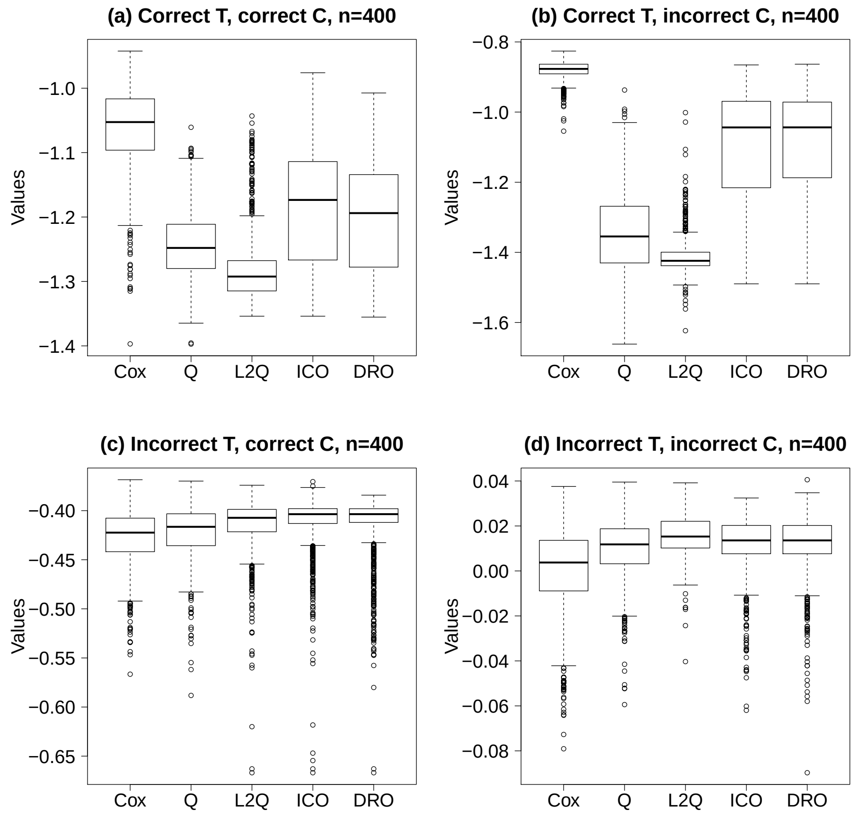

Fig. 2. Boxplots of values of estimated rules using different methods, representing the logarithm of the survival time with higher values being more preferable, $n = 400$. Cox: Cox model; Q: inverse censoring weighted Q-learning; L2Q: inverse censoring weighted L_2 Q-learning; ICO: inverse censoring weighted outcome weighted learning with linear kernel; DRO: doubly robust outcome weighted learning with linear kernel.

REFERENCES

BARTLETT, P. L., JORDAN, M. I. & MCAULIFFE, J. D. (2006). Convexity, classification, and risk bounds. *J. Am. Statist. Assoc.* **101**, 138–156.

STEINWART, I. & SCOVEL, C. (2007). Fast rates for support vector machines using gaussian kernels. *Ann. Statist.* **35**, 575–607.

ZHAO, Y. Q., ZENG, D., RUSH, A. J. & KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Am. Statist. Assoc.* **107**, 1106–1118.

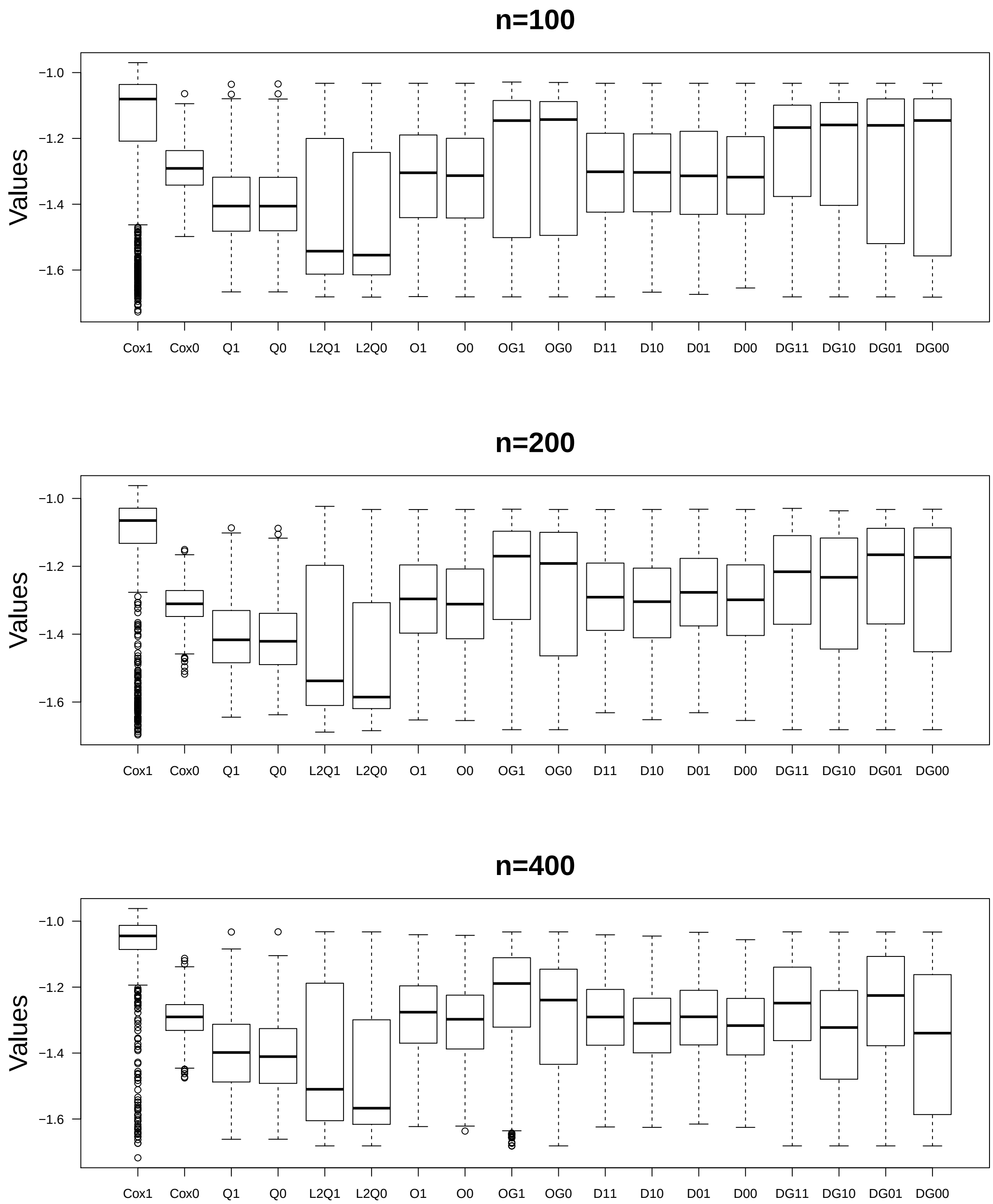

Fig. 3. Boxplots of values of estimated rules using different methods. Cox1(0): Cox model with correct (incorrect) basis; Q1(0), L2Q1(0): Q-learning or L_2 Q-learning with correct (incorrect) censoring weights; O1(0) $\{$OG1(0)$\}$: inverse censoring weighted outcome weighted learning with linear (Gaussian) kernel, correct (incorrect) censoring weights; D11(10) $\{$DG11(10)$\}$: doubly robust outcome weighted learning with linear (Gaussian) kernel, correct survival model and correct (incorrect) censoring model, and D01(00) $\{$DG01(00)$\}$ follows similarly.