Biometrics 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for "Estimation and Evaluation of Linear Individualized Treatment Rules to Guaranteed Performance"

by Xin Qiu, Donglin Zeng, and Yuanjia Wang

This paper has been submitted for consideration for publication in *Biometrics*

A1. Algorithms for Minimizing the Weighted Ramp Loss

[Figure 1 about here.]

To minimize the loss function of the weighted ramp loss,

$$L(f) = \sum_{i=1}^{n} C\frac{|W_i|h_s(Z_iA_if(\boldsymbol{X}_i))}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2, \tag{A.1}$$

express $h_s(u)$ as the difference of two convex functions. That is,

$$h_s(u) = h_{1,s}(u) - h_{2,s}(u) = (\frac{1}{2} - \frac{u}{s})_+ - (-\frac{1}{2} - \frac{u}{s})_+,$$

where function $(x)_+$ denotes the positive part of x. Let η_i denote $A_iZ_if(\boldsymbol{X}_i)$. Then the penalized weighted sum of ramp loss can be simplified as $L = \sum_{i=1}^{n} C\frac{|W_i|h_s(\eta_i)}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2$, and the minimization in (A.1) can be carried out in three steps:

- Step 1: Start with an initial value of $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta}_0$, which can be derived from the optimal rule estimated by the O-learning with hinge loss. Then, the initial value of $\boldsymbol{\eta}$ can be calculated and we denote it as $\boldsymbol{\eta}_0$.
- Step 2: Solve

$$\widehat{\boldsymbol{\beta}} = \arg\min \sum_{i=1}^{n} C\frac{|W_i|\{h_{1,s}(\eta_i) - \widehat{h}_{2,s}(\eta_i, \eta_i^0)\}}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2, \tag{A.2}$$

  where $\widehat{h}_{2,s}(\eta_i, \eta_i^0) = h_{2,s}(\eta_i^0) + h'_{2,s}(\eta_i^0)\eta_i$ and $h'_{2,s}(u) = \dfrac{-I(u/s < -1/2)}{s}$.
- Step 3: Compute $\boldsymbol{\eta}^0$ and update it in step 2 until the change in L is less than a pre-specified threshold.

In order to solve the optimization problem in Step 2, we introduce slack variables ξ_i to replace $h_{1,s}(\eta_i)$. Therefore, (A.2) is equivalent to minimize

$$\sum_{i=1}^{n} C\frac{|W_i|\{\xi_i - \widehat{h}'_{2,s}(\eta_i^0)\eta_i\}}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2, \ s.t. \ \ \xi_i \geqslant \frac{1}{2} - \frac{\eta_i}{s}, \text{ and } \ \xi_i \geqslant 0.$$

By adding two non-negative Lagrange multipliers $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$, we obtain

$$L = \sum_{i=1}^{n} C\frac{|W_i|\{\xi_i - \widehat{h}'_{2,s}(\eta_i^0)\eta_i\}}{\pi(A_i|\boldsymbol{X}_i)} + \frac{1}{2}||\boldsymbol{\beta}||^2 - \sum_{i=1}^{n}\alpha_i(\xi_i + \frac{\eta_i}{s} - \frac{1}{2}) - \sum_{i=1}^{n}\tau_i\xi_i.$$

Let $\boldsymbol{\gamma}$ be a vector with i-element $\gamma_i = \frac{|W_i|\widehat{h}'_{2,s}(\eta_i^0)}{\pi(A_i|\boldsymbol{X}_i)}$. Notice that $\eta_i = A_iZ_i(\beta_0 + \boldsymbol{X}_i^T\boldsymbol{\beta})$, and take

derivative with regard to β_0, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, we obtain the following equations

$$0 = \sum_{i=1}^{n} A_i Z_i \left(C\gamma_i + \frac{\alpha_i}{s} \right), \tag{A.3}$$

$$\boldsymbol{\beta} = \sum_{i=1}^{n} C \frac{|W_i| \widehat{h}'_{2,s}(\eta_i^0) A_i Z_i \boldsymbol{X}_i}{\pi(A_i|\boldsymbol{X}_i)} + \sum_{i=1}^{n} \alpha_i A_i Z_i \boldsymbol{X}_i / s = \sum_{i=1}^{n} A_i Z_i \left(C\gamma_i + \frac{\alpha_i}{s} \right) \boldsymbol{X}_i, \tag{A.4}$$

$$0 = \frac{|W_i|}{\pi(A_i|\boldsymbol{X}_i)} C - \alpha_i - \tau_i. \tag{A.5}$$

By (A.5), $\xi_i' s$ cancel out and the penalized weighted sum of ramp loss becomes

$$\begin{aligned}
L &= -\sum_{i=1}^{n} C\gamma_i \left\{ A_i Z_i (\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta}) \right\} + \frac{1}{2} ||\boldsymbol{\beta}||^2 - \sum_{i=1}^{n} \frac{\alpha_i}{s} \left\{ A_i Z_i (\beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta}) \right\} + \frac{1}{2} \sum_{i=1}^{n} \alpha_i \\
&= -\sum_{i=1}^{n} A_i Z_i \left(C\gamma_i + \frac{\alpha_i}{s} \right) \boldsymbol{X}_i^T \boldsymbol{\beta} + \frac{1}{2} \sum_{i=1}^{n} \alpha_i + \frac{1}{2} ||\boldsymbol{\beta}||^2 \qquad \text{by (A.3)} \\
&= -\frac{1}{2} ||\boldsymbol{\beta}||^2 + \frac{1}{2} \sum_{i=1}^{n} \alpha_i \qquad \text{by (A.4)} \\
&\propto \frac{1}{2} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \left(\sum A_i Z_i \frac{\alpha_i}{s} \boldsymbol{X}_i^T \sum A_i Z_i \frac{\alpha_i}{s} \boldsymbol{X}_i + 2 \sum A_i Z_i C\gamma_i \boldsymbol{X}_i^T \sum A_i Z_i \frac{\alpha_i}{s} \boldsymbol{X}_i \right) \\
&= -\frac{1}{2s^2} \boldsymbol{\alpha}^T \boldsymbol{Q} \boldsymbol{\alpha} + \frac{1}{2} \left(\mathbf{1} - 2C\boldsymbol{Q}\boldsymbol{\gamma}/s \right) \boldsymbol{\alpha},
\end{aligned}$$

where $\mathbf{Q}$ is a square matrix where $Q_{i,j} =< A_i Z_i \boldsymbol{X}_i, A_j Z_j \boldsymbol{X}_j >$.

Hence, the dual problem is

$$\min \quad \frac{1}{2s^2} \boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - \frac{1}{2} \left(\mathbf{1} - 2C\mathbf{Q}\boldsymbol{\gamma}/s \right)^T \boldsymbol{\alpha}, \tag{A.6}$$

subject to $0 \leqslant \alpha_i \leqslant C|W_i|/\pi(A_i|\boldsymbol{X}_i)$ and $\sum C A_i Z_i \gamma_i + \sum A_i Z_i \alpha_i / s = 0$. Thus, the optimization problem can be solved via quadratic programming. After obtaining α_i, the original coefficient can be derived by $\widehat{\boldsymbol{\beta}} = \sum A_i Z_i \left(C\boldsymbol{\gamma}_i + \frac{\alpha_i}{s} \right) \boldsymbol{X}_i$. Based on the KKT condition $\xi_i (C|W_i|/\pi(A_i|\boldsymbol{X}_i) - \alpha_i) = 0$, when $0 < \alpha_i < C|W_i|/\pi(A_i|\boldsymbol{X}_i)$, we have $\xi_i = 0$ and $A_i Z_i (\widehat{\beta}_0 + \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}) - \frac{1}{2} s = 0$. The intercept term $\widehat{\beta}_0$ can be calculated by taking the average of $\dfrac{s}{2A_i Z_i} - \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}$.

Therefore, we obtain the optimal linear ITR as

$$\widehat{f}_L^*(\boldsymbol{X}) = \widehat{\beta}_0 + \boldsymbol{X}^T \widehat{\boldsymbol{\beta}}.$$

A2. Asymptotic Properties

Let $\boldsymbol{X}$ denote a vector with one as the first component and the remaining components as feature variables. To emphasize that the tuning parameter s of ramp loss may depend on the sample size to establish asymptotic properties, we denote it by s_n in this section. We assume

(a) The true optimal linear function, $f_L^*(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}^*$, is the unique minimizer of $E\{RI(Af(\boldsymbol{X}) < 0)\}$ for $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}$ where $\|\boldsymbol{\beta}\| = 1$. Furthermore, there exists a positive constant δ such that $P(|\boldsymbol{X}^T\boldsymbol{\beta}^*| > \delta_0) = 1$.

(b) The joint densities of $(R, \boldsymbol{X})$ given $A = 1$ and -1 are twice-continuously differentiable.

(c) There exits a function $r(\boldsymbol{x})$ such that $\{\widehat{r}(\boldsymbol{x}) - r(\boldsymbol{x})\} = o((ns_n)^{-1/2})$ uniformly in $\boldsymbol{x}$.

(d) $(nC_n)^{-1} \to 0$, $ns_n \to \infty$, $ns_n^3 \to 0$, and $(ns_n)^{1/2}(nC_n)^{-1} \to 0$.

(e) There exits a unique minimizer, denoted by $\boldsymbol{\beta}_n$, that minimizes

$$E\left[|R - r(\boldsymbol{X})|h_s\left\{A\text{sign}(R - r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}\right\}/\pi(A|\boldsymbol{X})\right].$$

Assume that $\boldsymbol{\beta}_n$ belongs to a bounded set. Furthermore, let

$$IF_n(R, \boldsymbol{X}, A) = \left[\frac{\partial}{\partial\beta}E\left\{A(R - r(\boldsymbol{X}))\boldsymbol{X}/\pi(A|\boldsymbol{X})|Z(\boldsymbol{\beta}) = 0\right\}f_{Z(\boldsymbol{\beta})}(0)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_n}\right]^{-1}$$

$$\times\left[|R - r(\boldsymbol{X})|A\text{sign}(R - r(\boldsymbol{X}))\boldsymbol{X}(2s_n)^{-1}I(A\text{sign}(R - r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}_n \in [-s_n/2, s_n/2])/\pi(A|\boldsymbol{X})\right],$$

we assume that $s_n^{1/2}IF_n(R, \boldsymbol{X}, A)$ has a bounded third moment and converges to a random variable in $L_2(P)$ norm.

Condition (a) requires a separable boundary condition, but this condition can be further relaxed to allow $\boldsymbol{X}^T\boldsymbol{\beta}^*$ to have positive probability around the boundary and the density vanishes faster than a linear rate when close to the boundary. Condition (c) usually holds if we estimate $r(\boldsymbol{x})$ through some parametric models. In condition (d), s_n and C_n are the tuning parameters to be chosen depending on n, for example, $C_n = 1$ and $s_n = n^{-1/2}$. Condition (e) assumes the convergence of the minimizer associated with the ramp loss. Under these assumptions, we first show the consistency of ABLO, $\widehat{f}_L^*(\boldsymbol{x}) = \boldsymbol{x}^T\widehat{\boldsymbol{\beta}}$. The proof follows the standard M-estimation theory by Van der Vaart (2000). Let $\mathbf{P}_n$ denote the empirical measure,

then $\widehat{f}_L^*$ minimizes

$$\mathbf{P}_n \left[|R - \widehat{r}(\boldsymbol{X})| h_s \left\{ A\text{sign}(R - \widehat{r}(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} \right\} / \pi(A|\boldsymbol{X}) \right] + (2nC_n)^{-1} \|\boldsymbol{\beta}\|^2.$$

It is clear that from assumptions (a), (b) and (c),

$$\sup_{\boldsymbol{\beta}} \Big| \mathbf{P}_n \left[|R - \widehat{r}(\boldsymbol{X})| h_s \left\{ A\text{sign}(R - \widehat{r}(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} \right\} / \pi(A|\boldsymbol{X}) \right]$$

$$+ (2nC_n)^{-1} \|\beta\|^2 - E \left[|R - r(\boldsymbol{X})| h_s \left\{ A\text{sign}(R - r(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} \right\} / \pi(A|\boldsymbol{X}) \right] \Big| \to 0$$

almost surely. By condition (b) and (d), $E \left[|R - r(\boldsymbol{X})| h_s \left\{ A\text{sign}(R - r(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} \right\} / \pi(A|\boldsymbol{X}) \right]$ converges uniformly to $E \left[|R - r(\boldsymbol{X})| I \left\{ A\text{sign}(R - r(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} < 0 \right\} / \pi(A|\boldsymbol{X}) \right]$, which is equivalent to

$$E \left[RI(A\boldsymbol{X}^T \boldsymbol{\beta} < 0) / \pi(A|\boldsymbol{X}) \right] - E[(R - r(\boldsymbol{X}))^-] - r(\boldsymbol{X}).$$

This gives

$$\mathbf{P}_n \left[|R - \widehat{r}(\boldsymbol{X})| h_s \left\{ A\text{sign}(R - \widehat{r}(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} \right\} / \pi(A|\boldsymbol{X}) \right] + (2nC_n)^{-1} \|\boldsymbol{\beta}\|^2$$

$$\to E \left[RI(A\boldsymbol{X}^T \boldsymbol{\beta} < 0) / \pi(A|\boldsymbol{X}) \right] - E[(R - r(\boldsymbol{X}))^-] - r(\boldsymbol{X})$$

uniformly in $\boldsymbol{\beta}$. Since (a) implies f_L^* is also the unique minimizer of the latter limit for $\|\boldsymbol{\beta}\| = 1$, it yields that any convergent subsequence of $\widehat{\boldsymbol{\beta}}$ should converge to a limit proportional to $\boldsymbol{\beta}^*$. Therefore, we conclude that $\widehat{\boldsymbol{\beta}} / \|\widehat{\boldsymbol{\beta}}\|$ converges to $\boldsymbol{\beta}^*$ almost surely. Furthermore, by noting

$$\sup_{\boldsymbol{\beta}} \Big| \mathbf{P} \left[|R - \widehat{r}(\boldsymbol{X})| h_s \left\{ A\text{sign}(R - \widehat{r}(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} \right\} / \pi(A|\boldsymbol{X}) \right]$$

$$- E \left[|R - r(\boldsymbol{X})| h_s \left\{ A\text{sign}(R - r(\boldsymbol{X})) \boldsymbol{X}^T \boldsymbol{\beta} \right\} / \pi(A|\boldsymbol{X}) \right] \Big| \to 0,$$

we can easily show that $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n\|$ converges to zero almost surely.

To obtain the asymptotic normality for $\widehat{\boldsymbol{\beta}}$, we follow Koo et al. (2008) by noting $\widehat{\boldsymbol{\beta}}$ solves

$$\mathbf{P}_n \left[|R - \widehat{r}(\boldsymbol{X})| A\text{sign}(R - \widehat{r}(\boldsymbol{X})) \boldsymbol{X} h_s' \left\{ A\text{sign}(R - \widehat{r}(\boldsymbol{X}) \boldsymbol{X}^T \widehat{\boldsymbol{\beta}} \right\} / \pi(A|\boldsymbol{X}) \right] + (nC_n)^{-1} \widehat{\boldsymbol{\beta}} = 0.$$

This gives

$$\begin{aligned}
&\sqrt{ns_n}(\mathbf{P}_n - \mathbf{P})\left[|R-\widehat{r}(\boldsymbol{X})|A\text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}h_s'\left\{A\text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}/\pi(A|\boldsymbol{X})\right] \\
= \;& -(ns_n)^{1/2}(nC_n)^{-1}\widehat{\boldsymbol{\beta}} \\
& -(ns_n)^{1/2}\mathbf{P}\left[|R-\widehat{r}(\boldsymbol{X})|A\text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}h_s'\left\{A\text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}/\pi(A|\boldsymbol{X})\right] \\
= \;& o(1) - (ns_n)^{1/2}\frac{\partial}{\partial y}\mathbf{P}\left[|R-y|A\text{sign}(R-y)\boldsymbol{X}h_s'\left\{A\text{sign}(R-y))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}\frac{\widehat{r}(\boldsymbol{X})-r(\boldsymbol{X})}{\pi(A|\boldsymbol{X})}\right]\Big|_{y=r(\boldsymbol{X})} \\
& +(ns_n)^{1/2}s_n^{-1}\int_{-s_n/2}^{s_n/2}E\left[A(R-r(\boldsymbol{X}))\boldsymbol{X}/\pi(A|\boldsymbol{X})|Z(\widehat{\boldsymbol{\beta}})=z\right]dF_{Z(\widehat{\boldsymbol{\beta}})}(z),
\end{aligned}$$

where $Z(\boldsymbol{\beta})$ denotes the random variable $A\text{sign}(R-r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}$ and $F_{Z(\boldsymbol{\beta})}$ is its cumulative distribution function. From (b) and since $\boldsymbol{\beta}_n$ is the minimizer of the expected ramp loss, the last term is equal to

$$(ns_n)^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)\frac{\partial}{\partial\boldsymbol{\beta}}E\left[A(R-r(\boldsymbol{X}))\boldsymbol{X}/\pi(A|\boldsymbol{X})|Z(\boldsymbol{\beta})=0\right]f_{Z(\boldsymbol{\beta})}(0)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_n}+o(1).$$

Thus, the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)$ holds by noting that

$$\sqrt{ns_n}(\mathbf{P}_n-\mathbf{P})\left[|R-\widehat{r}(\boldsymbol{X})|A\text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}h_s'\left\{A\text{sign}(R-\widehat{r}(\boldsymbol{X}))\boldsymbol{X}^T\widehat{\boldsymbol{\beta}}\right\}/\pi(A|\boldsymbol{X})\right]$$

is equivalent to

$$\sqrt{ns_n}(\mathbf{P}_n-\mathbf{P})\left[|R-r(\boldsymbol{X})|A\text{sign}(R-r(\boldsymbol{X}))\boldsymbol{X}\frac{I(A\text{sign}(R-r(\boldsymbol{X}))\boldsymbol{X}^T\boldsymbol{\beta}_n\in[-s_n/2,s_n/2])}{2s_n\pi(A|\boldsymbol{X})}\right]$$

and therefore,

$$\sqrt{ns_n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)=\sqrt{ns_n}(\mathbf{P}_n-\mathbf{P})IF_n(R,\boldsymbol{X},A)+o_p(1).$$

The asymptotical normality of $\sqrt{ns_n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_n)$ follows from condition (e).

Lastly, we examine the diagnostic statistics for any estimated decision function, denoted as $\widehat{\delta}_{\mathcal{C}}(\widehat{f})$ in (7) of the main paper, where $\widehat{f}(\boldsymbol{x})=\boldsymbol{x}^T\widehat{\boldsymbol{\beta}}$ is an estimated rule converging to $f^*(\boldsymbol{x})$ uniformly in x. Note that we split the data into K folds, $\widehat{f}^{(-k)}$ is estimated without the kth part of data and $\widehat{\delta}_{\mathcal{C}}^{(k)}$ is computed using the kth part. Let n_k denote the sample size of the kth part of data and let $\mathbf{P}_{n_k}$ denote the empirical measure for the k part of data. Define by

$$\delta_{\mathcal{C}}^*=\frac{E\left[I(\boldsymbol{X}\in\mathcal{C},Af^*(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X})-I(\boldsymbol{X}\in\mathcal{C},Af^*(\boldsymbol{X})<0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X}\in\mathcal{C})]}$$

the subgroup benefit based on the optimal linear rule f^*. Since $\boldsymbol{\beta}_n/\|\boldsymbol{\beta}_n\| \to \boldsymbol{\beta}^*$, from condition (a), we have

$$\boldsymbol{\beta}^{*T}\boldsymbol{X}\boldsymbol{\beta}_n^T\boldsymbol{X}/\|\boldsymbol{\beta}_n\| > 0$$

with probability one. Therefore,

$$\delta_{\mathcal{C}}^* = \frac{E\left[I(\boldsymbol{X}\in\mathcal{C}, Af_n(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X}\in\mathcal{C}, Af_n(\boldsymbol{X})<0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X}\in\mathcal{C})]},$$

where $f_n(\boldsymbol{X}) = \boldsymbol{\beta}_n^T\boldsymbol{X}$.

Re-express $\widehat{\delta}_{\mathcal{C}}(\widehat{f}^{(-k)})$ as

$$\widehat{\delta}_{\mathcal{C}}^{(k)} = \frac{\mathbf{P}_{n_k} I(\boldsymbol{X}\in\mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X})}{\mathbf{P}_{n_k} I(\boldsymbol{X}\in\mathcal{C})} - \frac{\mathbf{P}_{n_k} I(\boldsymbol{X}\in\mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X})<0)R/\pi(A|\boldsymbol{X})}{\mathbf{P}_{n_k} I(\boldsymbol{X}\in\mathcal{C})}.$$

Since $\{I(X\in\mathcal{C}) : \mathcal{C}\in\{\mathcal{C}_1,...,\mathcal{C}_m\}\}$ and $\left\{Af(\boldsymbol{X})>0 : f=\boldsymbol{X}^T\boldsymbol{\beta}\right\}$ are VC-major classes,

$$\begin{aligned} & (\mathbf{P}_{n_k}-\mathbf{P})I(\boldsymbol{X}\in\mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X}) \\ = \;& (\mathbf{P}_{n_k}-\mathbf{P})I(\boldsymbol{X}\in\mathcal{C}, Af^*(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X}) + o_p(n_k^{-1/2}). \end{aligned}$$

We obtain

$$\begin{aligned} & \widehat{\delta}_{\mathcal{C}}(\widehat{f}^{(-k)}) - \delta_{\mathcal{C}}^* \\ = \;& \frac{(\mathbf{P}_{n_k}-\mathbf{P})I(\boldsymbol{X}\in\mathcal{C}, Af^*(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X})}{\mathbf{P}I(\boldsymbol{X}\in\mathcal{C})} - \frac{(\mathbf{P}_{n_k}-\mathbf{P})I(\boldsymbol{X}\in\mathcal{C}, Af^*(\boldsymbol{X})<0)R/\pi(A|X)}{\mathbf{P}I(\boldsymbol{X}\in\mathcal{C})} \\ - \;& \frac{E\left[I(\boldsymbol{X}\in\mathcal{C}, Af^*(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X}\in\mathcal{C}, Af^*(\boldsymbol{X})<0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X}\in\mathcal{C})]^2}(\mathbf{P}_{n_k}-\mathbf{P})I(\boldsymbol{X}\in\mathcal{C}) \\ + \;& \frac{E\left[I(\boldsymbol{X}\in\mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X}\in\mathcal{C}, A\widehat{f}^{(-k)}(\boldsymbol{X})<0)R/\pi(A|\boldsymbol{X})\right]}{E[I(X\in\mathcal{C})]^2} \\ & - \frac{E\left[I(\boldsymbol{X}\in\mathcal{C}, Af_n(\boldsymbol{X})>0)R/\pi(A|\boldsymbol{X}) - I(\boldsymbol{X}\in\mathcal{C}, Af_n(X)<0)R/\pi(A|\boldsymbol{X})\right]}{E[I(\boldsymbol{X}\in\mathcal{C})]^2} \\ + \;& o_p(n_k^{-1/2}). \end{aligned}$$

Using the smooth condition in (b) and the expansion for $\widehat{\boldsymbol{\beta}}^{(-k)}$ around $\widehat{\beta}_n$ from the previous asymptotic proof, we can show that the difference in the last two terms has a convergence rate faster than $n_k^{-1/2}$, given $n_k = o(ns_n)$, and furthermore, when $n_k\to\infty$,

$$\sqrt{n_k}\left(\widehat{\delta}_{\mathcal{C}}^{(k)} - \delta_{\mathcal{C}}^*\right) \to_d \mathcal{G}(\mathcal{C}),$$

where $\mathcal{G}(\mathcal{C})$ is a tight Gaussian process indexed by $\mathcal{C}$ with mean zero. After averaging over all folds and assuming K is fixed, similar argument shows that $\sqrt{n}\left(\widehat{\delta}_{\mathcal{C}} - \delta^*_{\mathcal{C}}\right) \to_d \widetilde{\mathcal{G}}(\mathcal{C})$ for some tight Gaussian process $\widetilde{\mathcal{G}}$, where $\widehat{\delta}_{\mathcal{C}} = \frac{1}{K}\sum_k \widehat{\delta}^{(k)}_{\mathcal{C}}$. Note that these results apply to ABLO $\widehat{f}_L$, or other $\widehat{f}$ estimated from minimizing a weighted hinge loss as in O-learning or predictive modeling.

If f^*_L is also the global optimal rule, that is $f^*_L = f^*$, then $\delta^*_{\mathcal{C}} > 0$ for any $\mathcal{C}$ and any $\boldsymbol{X}$. Therefore, the confidence interval for $\delta^*_{\mathcal{C}}$ will be expected to be within $(0, \infty)$ when n is sufficiently large. We can also construct a test for $H_0 : \delta^*_{\mathcal{C}} \geqslant 0$ vs $H_a : \delta^*_{\mathcal{C}} < 0$ using this asymptotic distribution.

A3. Computing the Theoretical Optimal Linear Rule

Here we derive the theoretical optimal linear rule f^*_L in the class of all linear rules $f \in \mathcal{L}$ under our simulation settings in Section 3. Let G be the latent class identifier in the simulations. Define $G|(\boldsymbol{X}, W, V, A, \boldsymbol{U}) = G|\boldsymbol{X}$ as the class number, which only depends on $\boldsymbol{X} = (X^1, X^2, \cdots, X^p)$, where $X^j|G = k \sim N(\mu_k, 1)$ for $j = 1, \cdots, p$, and $k = 1, 2, 3, 4$. For a given treatment decision rule f, the expected value function under the decision rule is

$$\begin{aligned} & E\left[\frac{R}{\pi(A|\boldsymbol{X})}\left\{I(Af(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\right\}\right] \\ = \;& E\left[I(f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\left\{E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = 1) - E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = -1)\right\}\right] \\ + \;& E\big\{E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = -1)\big\}. \end{aligned}$$

Because $E\left\{E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A = -1)\right\}$ is a constant which doesn't depend on f, maximizing the expected value function is equivalent to maximizing $E\left\{I(f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\Omega(\boldsymbol{X}, W)\right\}$, where under the simulation model for $E(R|\boldsymbol{X}, V, W, \boldsymbol{U}, A)$ we can obtain

$$\begin{aligned} \Omega(\boldsymbol{X}, W) \;=\; & P(G = 1|\boldsymbol{X})\left\{\delta_1 + (\alpha_{11} - \alpha_{21})W\right\} + P(G = 2|\boldsymbol{X})\left\{\delta_2 + (\alpha_{12} - \alpha_{22})W\right\} \\ + \;& P(G = 3|\boldsymbol{X})\left\{-\delta_1 + (\alpha_{13} - \alpha_{23})W\right\} + P(G = 4|\boldsymbol{X})\left\{-\delta_2 + (\alpha_{14} - \alpha_{24})W\right\}. \end{aligned}$$

Next, we show that V and $\boldsymbol{U}$ are independent of optimal linear decision rule f_L^*. Let $f_L^*(\boldsymbol{X}, W)$ maximizes the value function in class $\mathcal{L}$. For any fixed V and $\boldsymbol{U}$,

$$\begin{aligned} E\left\{I[f_L^*(\boldsymbol{X}, W) > 0]\Omega(\boldsymbol{X}, W)\right\} &\geqslant E\left[I\{f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0\}\Omega(\boldsymbol{X}, W)\right] \\ &= E\left[I\{f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0\}\Omega(\boldsymbol{X}, W)|V, \boldsymbol{U}\right]. \end{aligned}$$

Therefore, take expectation of the inequality to obtain

$$\begin{aligned} E\left[E\left\{I(f_L^*(\boldsymbol{X}, W) > 0)\Omega(\boldsymbol{X}, W)\right\}\right] &\geqslant E\left[E\left\{I(f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0)\Omega(\boldsymbol{X}, W)|V, \boldsymbol{U}\right\}\right] \\ &= E\left[I\{f(\boldsymbol{X}, V, W, \boldsymbol{U}) > 0\}\Omega(\boldsymbol{X}, W)\right]. \end{aligned}$$

Thus we can ignore the independent noise variables while maximizing the value function.

Under linear transformation,

$$\boldsymbol{X} \to (\frac{X_s}{\sqrt{p}}, \widetilde{x_2}, \cdots, \widetilde{x_p}),$$

where $X_s = X^1 + X^2 + \cdots + X^p$, and $\widetilde{x_2}, \cdots, \widetilde{x_p}$ are orthogonal to X_s, the objective function becomes

$$\int\int I\left\{f(X_s, \widetilde{x_2}, \cdots, \widetilde{x_p}, W) > 0\right\}\Omega(X_s, W)e^{-\frac{X_s^2}{2p} - \frac{\widetilde{x}_2^2}{2} \cdots - \frac{\widetilde{x}_s^2}{2}} dX_s f(W) dW d\widetilde{x}_2 \cdots d\widetilde{x}_s,$$

where

$$\begin{aligned} \Omega(X_s, W) = &e^{\mu_1 X_s - \frac{p\mu_1^2}{2}}\left\{\delta_1 + (\alpha_{11} - \alpha_{21})W\right\} + e^{\mu_2 X_s - \frac{p\mu_2^2}{2}}\left\{\delta_2 + (\alpha_{12} - \alpha_{22})W\right\} \\ &+ e^{\mu_3 X_s - \frac{p\mu_3^2}{2}}\left\{-\delta_1 + (\alpha_{13} - \alpha_{23})W\right\} + e^{\mu_4 X_s - \frac{p\mu_4^2}{2}}\left\{-\delta_2 + (\alpha_{14} - \alpha_{24})W\right\}. \end{aligned}$$

Because $(\widetilde{x_2}, \cdots, \widetilde{x_p})$ are independent noise variables, as shown before, the optimal linear rule only depends on X_s and W. The objective function is thus equivalent to

$$\int\int I\left\{f(X_s, W) > 0\right\}\Omega(X_s, W) dX_s f(W) dW.$$

As $X_s \sim \frac{1}{4}N(\mu_1, p) + \frac{1}{4}N(\mu_2, p) + \frac{1}{4}N(\mu_3, p) + \frac{1}{4}N(\mu_4, p)$ and $W \sim N(0, 1)$, where μ_k is the mean of X^p in the kth class. Monte Carlo method can be applied to find the optimal linear rule f_L^*.

A4. Additional Simulation Results

[Figure 2 about here.]

We performed additional simulations to vary the strength of the informative feature variable W, such that its effects in different settings are $\boldsymbol{\alpha} = \begin{bmatrix} 1 & 1 & 0.3 & 0.6 \\ 0.5 & 0.5 & 0.3 & 0.6 \end{bmatrix}$.

Results from 500 replicates are summarizied in Table A.1, Figure A.3, and A.4. ABLO with linear kernel has the highest optimal treatment classification accuracy regardless of the sample size for both settings, and also estimates the ITR benefit closest to the true global maximal value of 0.705 on the overall sample. PM, Q-learning, and O-learning underestimate the ITR benefit, especially when the sample size is smaller ($N = 400$ training, 400 testing). Thus they do not achieve the maximal value of the theoretical optimal linear rule. The performance of estimating subgroup ITR benefit is similar to the overall sample. ABLO outperforms other methods with subgroup ITR benefit closer to the true global maximal value (e.g., in groups $W \in [-0.5, 0.5]$ and $W > 0.5$).

[Table 1 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

A5. Additional Results for STAR*D

The final STAR*D ITR estimated by ABLO using full data can be expressed as

$$\begin{aligned}\widehat{f}(\boldsymbol{X}) = &- 12.97 + 0.30 * sex + 1.27 * white + 0.79 * black + 2.77 * depression + 0.05 * age \\ &+ 0.26 * qids.start - 3.40 * qids.slope + 2.39 * preference,\end{aligned}$$

and treat a patient with SSRI if $\widehat{f} > 0$; otherwise treat with a non-SSRI if $\widehat{f} \leqslant 0$. The variable "sex" was coded as one for female and "preference" was coded as one for switch and zero for no preference.

[Table 2 about here.]

A6. Sensitivity to the Starting Values of ABLO

To evaluate the sensitivity of the algorithm to starting values, we include the algorithm convergence path of two example datasets in terms of value function and weighted ramp loss function. In Figure A.5, lines indicate convergence paths given different initial values. In the first example dataset, the algorithm converges to the same value function and ramp loss. However, the algorithm converges fastest if starting with O-leaning estimates. In the second dataset, the algorithm is more sensitive to different starting values, but the one starting with O-learning estimates performs the best, which is also the proposed starting values for ABLO.

[Figure 5 about here.]

References

Koo, J.-Y., Lee, Y., Kim, Y., and Park, C. (2008). A bahadur representation of the linear support vector machine. *The Journal of Machine Learning Research* **9,** 1343–1368.

Liu, Y., Wang, Y., Kosorok, M., Zhao, Y., and Zeng, D. (2014). Robust hybrid learning for estimating personalized dynamic treatment regimens. *arXiv preprint arXiv:1611.02314* .

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge university press.

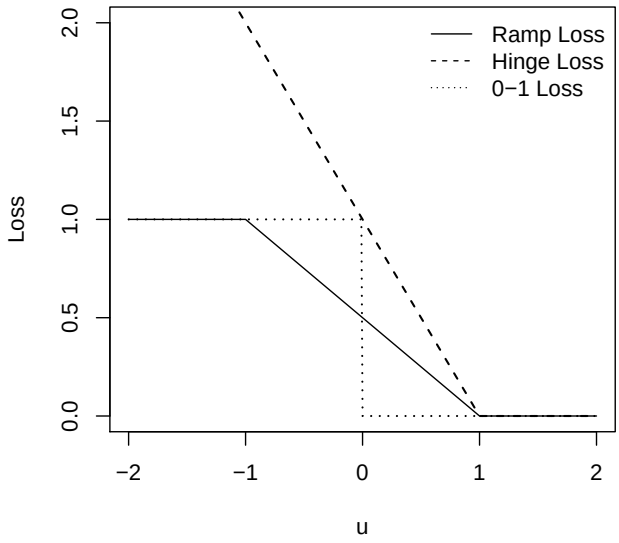

Figure A.1. Different approximation functions of the zero-one loss.

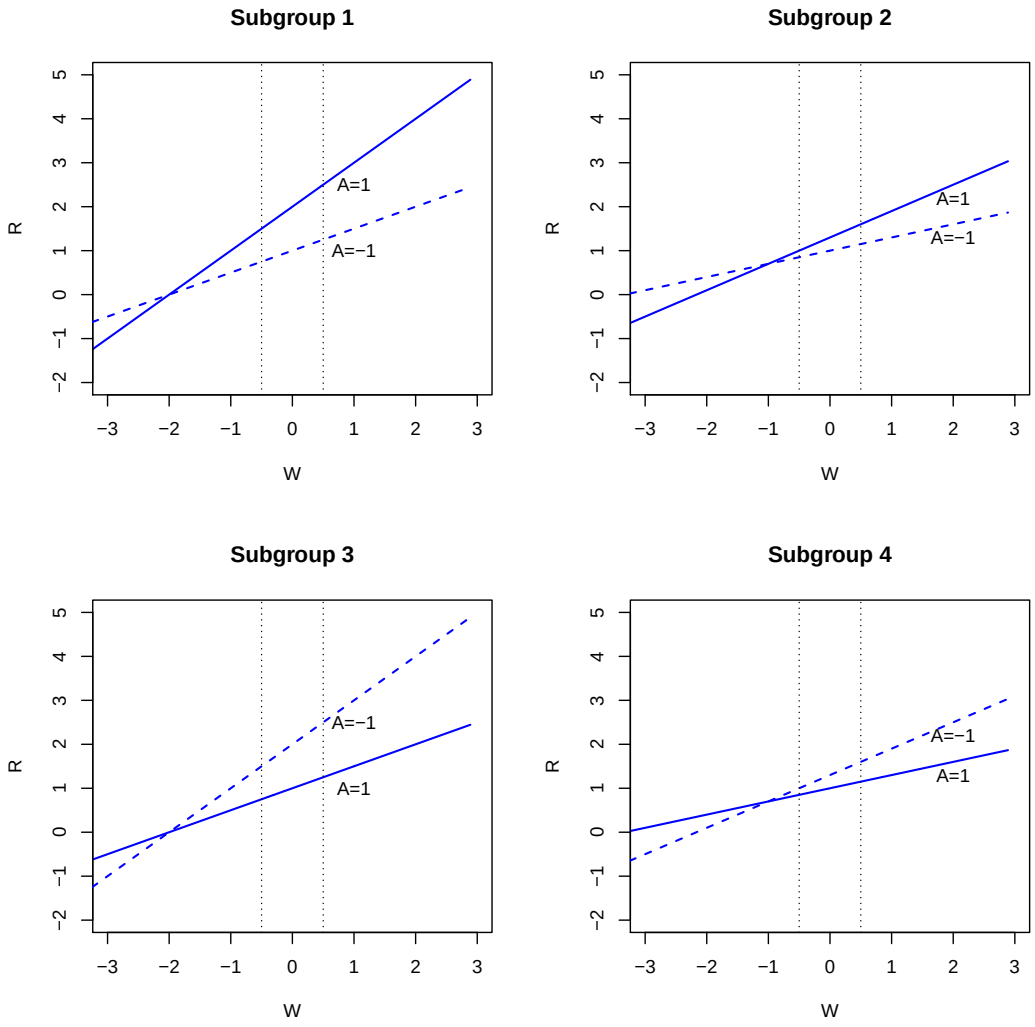

Figure A.2. Clinical outcome (R) versus W with treatment 1 or -1 in each latent group in the simulation setting described in Section 3. Two vertical dotted lines indicate $W = -0.5$ and $W = 0.5$.

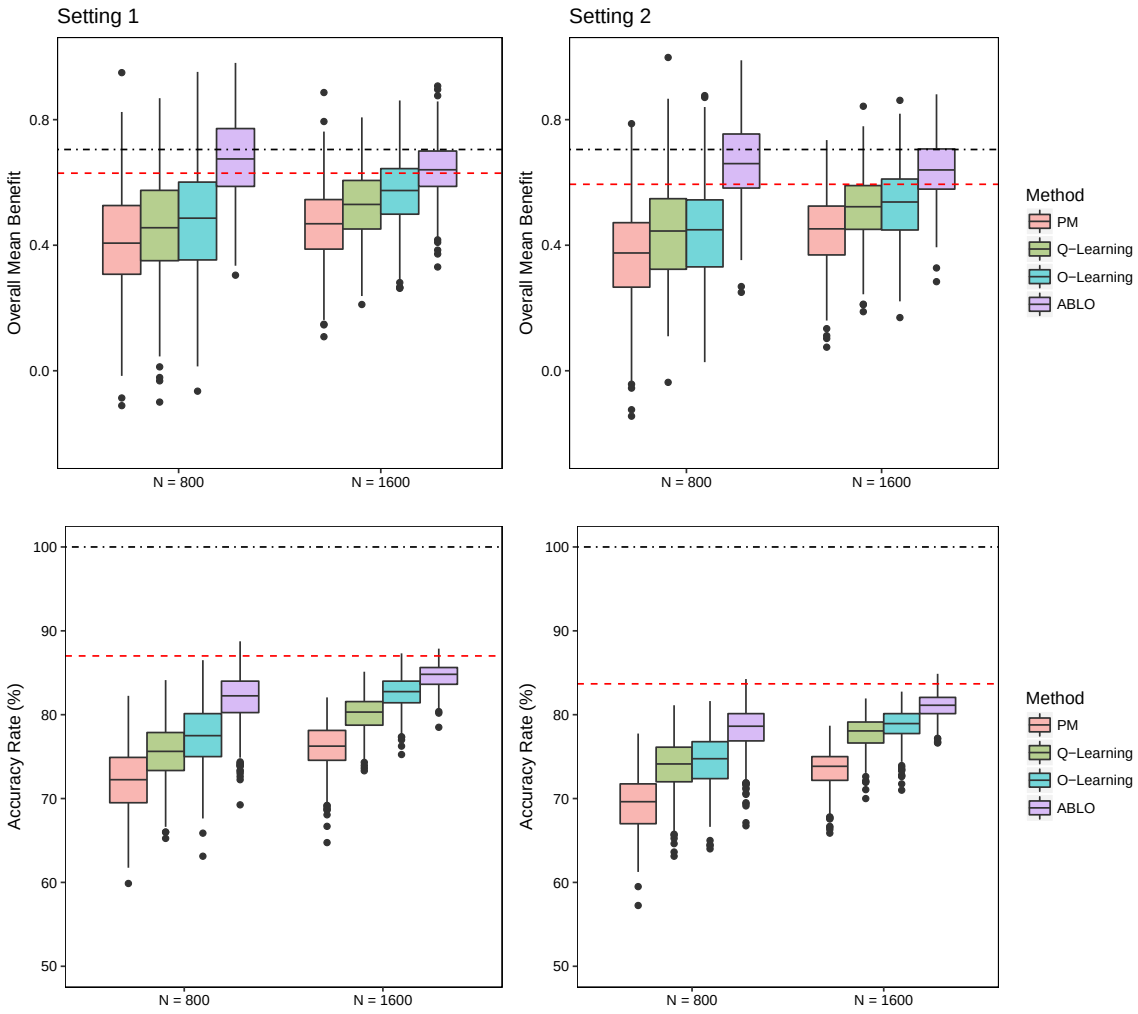

Figure A.3. Simulation results: Overall ITR benefit and accuracy rates for the four methods. Dotted-dashed lines represent the benefit (top panels) and accuracy (bottom panels under the theoretical global optimal treatment f^*. Dashed lines represent the benefit and accuracy under the theoretical optimal linear rule f_L^*. The methods being compared are (from left to right): PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning. This figure appears in color in the electronic version of this article.

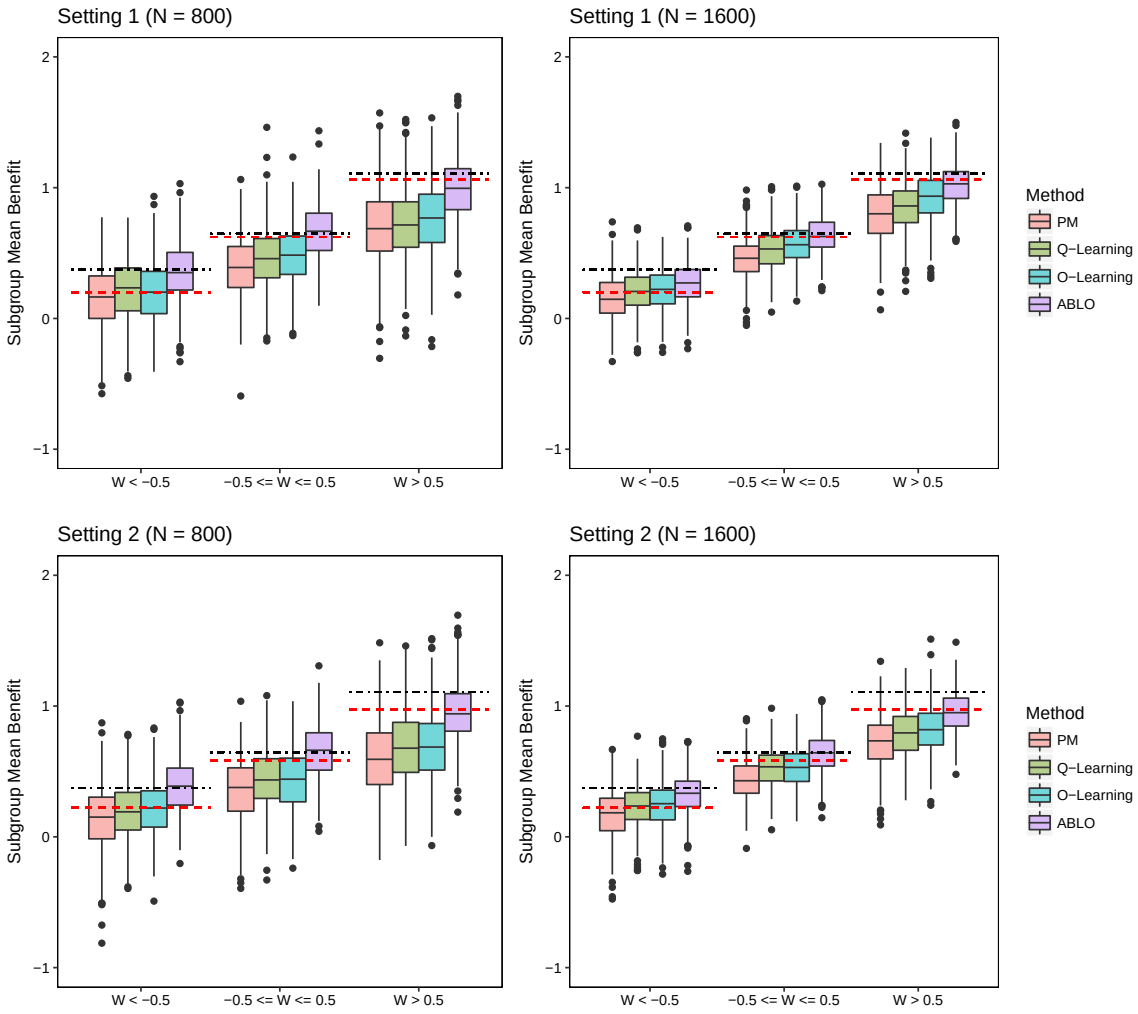

Figure A.4. Simulation results: Subgroup ITR benefit for the four methods. Black dotted-dashed lines represent the benefit under the theoretical global optimal treatment f^*. Red dashed lines represent the benefit under the theoretical optimal linear rule f_L^*. The methods being compared are (from left to right): PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning. This figure appears in color in the electronic version of this article.

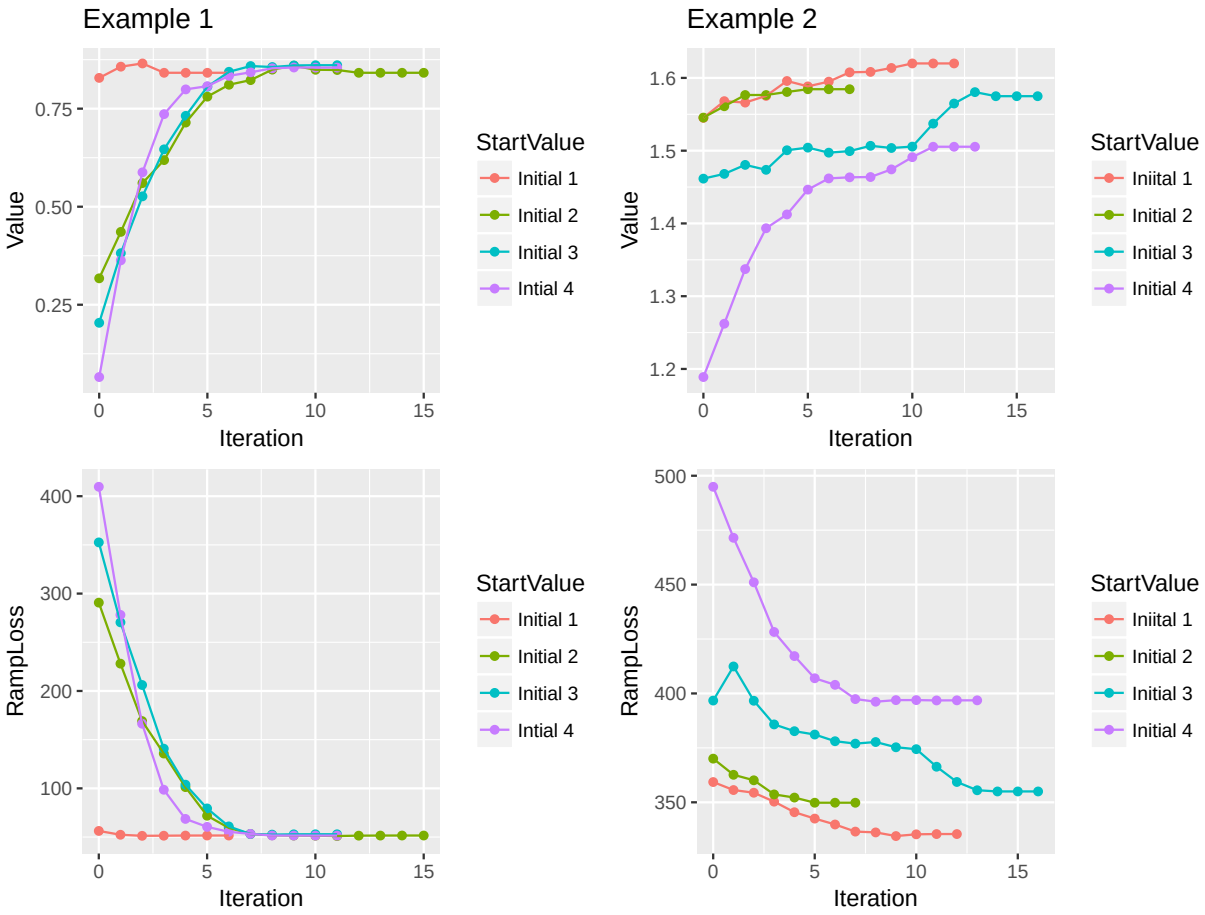

Figure A.5. Performance of the algorithm on two example datasets evaluated by value function and penalized weighted sum of ramp loss. Initial 1 starts $\boldsymbol{\beta}$ from the estimates obtained by O-learning; Initial 2 starts from $\boldsymbol{\beta} = \mathbf{0}$; Initial 3 starts from $\boldsymbol{\beta} = (1, \cdots, 1, -1, \cdots, -1)^T$, where half of the components are 1 and the other half are -1; Initial 4 starts from $\boldsymbol{\beta} = (1, 0, \cdots, 0)^T$. This figure appears in color in the electronic version of this article.

Table A.1

Simulation results: mean and standard deviation of the accuracy rate, mean benefit, and coverage probability for estimation of the benefit of the optimal ITR. PM: predictive modeling by random forest; Q-learning: Q-learning with linear regression; O-learning: improved single stage O-learning (Liu et al., 2014); ABLO: asymptotically best linear O-learning.

Setting 1. Four region means = $(1, 0.5, -1, -0.5)$.									
		Overall Benefit		$W < -0.5$		$W \in [-0.5, 0.5]$		$W > 0.5$	
	Accuracy rate	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage
$N = 800$									
PM	0.72 (0.04)	0.41 (0.17)	0.77	0.16 (0.24)	0.97	0.40 (0.23)	0.87	0.69 (0.30)	0.76
Q-learning	0.75 (0.03)	0.46 (0.17)	0.83	0.21 (0.24)	0.97	0.46 (0.23)	0.89	0.72 (0.27)	0.77
O-learning	0.77 (0.04)	0.48 (0.17)	0.85	0.20 (0.24)	0.97	0.48 (0.23)	0.92	0.76 (0.28)	0.82
ABLO	0.82 (0.03)	0.68 (0.13)	0.94	0.36 (0.22)	0.91	0.66 (0.21)	0.95	0.99 (0.24)	0.93
$N = 1600$									
PM	0.76 (0.03)	0.47 (0.12)	0.72	0.15 (0.17)	0.96	0.46 (0.16)	0.85	0.80 (0.21)	0.70
Q-learning	0.80 (0.02)	0.53 (0.11)	0.86	0.21 (0.16)	0.97	0.53 (0.16)	0.92	0.85 (0.18)	0.80
O-learning	0.83 (0.02)	0.57 (0.10)	0.94	0.23 (0.16)	0.97	0.57 (0.15)	0.95	0.93 (0.19)	0.90
ABLO	0.85 (0.01)	0.64 (0.09)	0.96	0.27 (0.15)	0.93	0.64 (0.14)	0.95	1.02 (0.16)	0.94
Best linear rule*	0.870	$\delta^l_{\mathcal{C}} = 0.630$		$\delta^l_{\mathcal{C}} = 0.200$		$\delta^l_{\mathcal{C}} = 0.624$		$\delta^l_{\mathcal{C}} = 1.063$	
Setting 2. Four region means = $(1, 0.3, -1, -0.3)$.									
		Overall Benefit		$W < -0.5$		$W \in [-0.5, 0.5]$		$W > 0.5$	
	Accuracy rate	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage	Mean (sd)	Coverage
$N = 800$									
PM	0.69 (0.03)	0.37 (0.17)	0.75	0.15 (0.24)	0.97	0.36 (0.23)	0.87	0.60 (0.29)	0.75
Q-learning	0.74 (0.03)	0.44 (0.16)	0.85	0.20 (0.22)	0.98	0.45 (0.23)	0.91	0.68 (0.26)	0.82
O-learning	0.74 (0.03)	0.45 (0.16)	0.86	0.21 (0.22)	0.98	0.44 (0.23)	0.91	0.69 (0.27)	0.84
ABLO	0.78 (0.03)	0.67 (0.13)	0.92	0.39 (0.20)	0.92	0.65 (0.21)	0.95	0.95 (0.24)	0.94
$N = 1600$									
PM	0.73 (0.02)	0.44 (0.11)	0.76	0.17 (0.18)	0.96	0.44 (0.15)	0.89	0.72 (0.19)	0.76
Q-learning	0.78 (0.02)	0.51 (0.11)	0.90	0.23 (0.16)	0.98	0.52 (0.16)	0.94	0.79 (0.19)	0.83
O-learning	0.79 (0.02)	0.53 (0.11)	0.91	0.24 (0.17)	0.96	0.53 (0.16)	0.94	0.82 (0.18)	0.88
ABLO	0.81 (0.01)	0.64 (0.10)	0.92	0.33 (0.16)	0.92	0.63 (0.15)	0.93	0.95 (0.16)	0.96
Best linear rule	0.837	$\delta^l_{\mathcal{C}} = 0.594$		$\delta^l_{\mathcal{C}} = 0.222$		$\delta^l_{\mathcal{C}} = 0.585$		$\delta^l_{\mathcal{C}} = 0.974$	
Best global rule		$\delta_{\mathcal{C}} = 0.705$		$\delta_{\mathcal{C}_1} = 0.373$		$\delta_{\mathcal{C}_2} = 0.647$		$\delta_{\mathcal{C}_3} = 1.109$	

*: The theoretical best linear rule for setting 1 is $\text{sign}(X_s * 0.98 + W * 0.19 - 0.03)$, where $X_s = X^1 + X^2 + \cdots + X^{10}$.
The theoretical best linear rule for setting 2 is $\text{sign}(\ X_s * 0.90 + W * 0.39 - 0.19)$.

Table A.2
*Results of STAR*D Data Analysis*

	QIDS score	ITR benefit	Subgroup ITR benefit by baseline QIDS score		
	Mean(sd)	Mean(sd)	QIDS$\leqslant 10$	QIDS$\in [11, 15]$	QIDS$\geqslant 16$
PM	9.69(0.38)	0.38(0.76)	1.29(0.82)	-0.10(1.02)	0.40(1.67)
Q-learning	9.50(0.35)	0.77(0.70)	2.08(0.68)	-0.17(0.92)	1.09(1.62)
O-learning	9.55(0.41)	0.66(0.82)	1.58(0.92)	-0.23(0.95)	1.20(1.84)
ABLO	9.32(0.23)	1.11(0.46)	2.22(0.45)	-0.18(0.51)	2.02(1.12)

*: lower QIDS score indicates a better outcome; higher benefit indicates a better outcome.