ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 5, *pp.* 1565–1582

Concordance-assisted learning for estimating optimal individualized treatment regimes

Caiyun Fan,

Shanghai University of International Business and Economics, People's Republic of China

Wenbin Lu and Rui Song

North Carolina State University, Raleigh, USA

and Yong Zhou

Shanghai University of Finance and Economics and Chinese Academy of Sciences, Beijing, People's Republic of China

[Received April 2015. Final revision September 2016]

Summary. We propose new concordance-assisted learning for estimating optimal individualized treatment regimes. We first introduce a type of concordance function for prescribing treatment and propose a robust rank regression method for estimating the concordance function. We then find treatment regimes, up to a threshold, to maximize the concordance function, named the prescriptive index. Finally, within the class of treatment regimes that maximize the concordance function, we find the optimal threshold to maximize the value function. We establish the rate of convergence and asymptotic normality of the proposed estimator for parameters in the prescriptive index. An induced smoothing method is developed to estimate the asymptotic variance of the estimator. We also establish the $n^{1/3}$-consistency of the estimated optimal threshold and its limiting distribution. In addition, a doubly robust estimator of parameters in the prescriptive index is developed under a class of monotonic index models. The practical use and effectiveness of the methodology proposed are demonstrated by simulation studies and an application to an acquired immune deficiency syndrome data set.

Keywords: Concordance; Optimal treatment regime; Propensity score; Rank estimation; Value function

1. Introduction

An individualized treatment regime is a deterministic function of predictors, such as clinical and genetic factors of patients, which aims to account for patients' heterogeneity in response to treatment and to maximize an expected clinical outcome of interest. Deriving optimal individualized treatment regimes has recently attracted much attention for treating many complex diseases, such as cancer and acquired immune deficiency syndrome (AIDS).

There is a large and growing body of literature for estimating the optimal individualized treatment regimes with a single decision time point and multiple decision time points. The latter are referred to as optimal dynamic treatment regimes. Two main dynamic learning approaches based on backward induction have been proposed for estimating optimal dynamic treatment

Address for correspondence: Wenbin Lu, Department of Statistics, 5212 SAS Hall, 2311 Stinson Drive, North Carolina State University, Raleigh, NC 27695, USA.
E-mail: lu@stat.ncsu.edu

1369–7412/17/791565

regimes. One is Q-learning (e.g. Watkins (1989), Watkins and Dayan (1992), Zhao *et al.* (2009, 2011) and Qian and Murphy (2011)), which posits regression models for the outcome of interest and so-called Q-functions. The other is A-learning (e.g. Murphy (2003) and Blatt *et al.* (2004)), which directly builds models for the contrast functions and uses doubly robust estimating equations to estimate the contrast functions by incorporating the estimated propensity score functions. Compared with Q-learning, A-learning is more robust to model misspecification. More recently, for deriving the optimal treatment regime with a single decision time point, Zhang *et al.* (2012b) formulated the problem in a missing data framework and proposed inverse propensity score weighted (IPSW) and augmented IPSW (AIPSW) estimators for the mean potential outcome under a given treatment regime, i.e. the value function. The estimated optimal treatment regime is then obtained by maximizing the estimated value function within a class of prespecified regimes, such as linear decision rules. The value-function-based optimization method was extended to estimate the optimal dynamic treatment regime in Zhang, Tsiatis, Laber and Davidian (2013). In addition, Zhao *et al.* (2012) and Zhang *et al.* (2012a) recast the estimation of the optimal treatment regime from a classification perspective and used machine learning tools, such as the outcome-weighted support vector machine, to optimize the estimated value function. Other development includes the subgroup identification method of Foster *et al.* (2011), the target population selection method of Zhao *et al.* (2013) and the marker-guided treatment selection method of Matsouaka *et al.* (2014).

The value-function-based estimation method that was proposed by Zhang *et al.* (2012b) is robust and appealing. It does not require a correct specification of the underlying model for the response and it finds the best treatment regime that maximizes the estimated value function in a class of interested regimes even when the true optimal regime is not contained in this class. However, it also has some limitations. First, the rates of convergence of the estimators for the parameters in the decision rules are slower than the standard $n^{1/2}$-rate, and their asymptotic distributions are not normal. Inference by these estimators was not studied in Zhang *et al.* (2012b). Second, the optimization of the estimated value function can be challenging especially when the dimension of predictors is relatively large since it is very bumpy. Zhang *et al.* (2012b) proposed to use a genetic algorithm to search for the optimum.

In this work, we propose a new criterion for optimal treatment decision. First, we introduce a type of concordance function for prescribing treatment. Here, concordance for treatment prescription means that, if one subject has a bigger benefit of receiving a treatment compared with another subject, he or she is more likely to be assigned to this treatment by the regime, which is a natural requirement for a good treatment regime. Then, we propose a robust rank regression method for estimating the concordance function. The rank estimator proposed is a U-statistic of order 2 similar to the maximum rank correlation estimator that has been widely studied in the econometrics literature (Han, 1987; Sherman, 1993; Cavanagh and Sherman, 1998; Chen, 2002; Abrevaya, 2003). Second, we find treatment regimes, up to a threshold, to maximize the concordance function, named the prescriptive index. The optimization of the estimated concordance function can be done by using a simplex algorithm, e.g. the `optim` function in R, even with a relatively larger number of predictors. Finally, within the class of treatment regimes that maximize the concordance function, we find the optimal threshold to maximize the IPSW estimator of the value function. A similar method was used by Matsouaka *et al.* (2014) to estimate the optimal threshold in marker-guided treatment regimes. We establish the $n^{1/2}$-consistency and asymptotic normality of the proposed estimator for parameters in the prescriptive index. An induced smoothing method is developed to estimate the asymptotic variance of the estimator proposed. We also establish the $n^{1/3}$-consistency of the estimated optimal threshold and its limiting distribution. The asymptotic distribution of the estimated value function for the estimated optimal treatment regimes is also derived. In addition, a doubly robust estimator of parameters in the prescriptive index is developed under a class of monotonic index models.

The rest of the paper is organized as follows. In Section 2, we introduce the notation, the concordance function for treatment prescription and the concordance-assisted estimation methods of the optimal treatment regime. The asymptotic properties of the proposed estimators for the parameters in the optimal regime are also studied. Section 3 presents extensive simulation studies to demonstrate the performance of the methods proposed. An application to AIDS clinical trial data is given in Section 4, followed by a discussion section. All the technical proofs are provided in Appendix A.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from `http://www4.stat.ncsu.edu/~lu/programcodes.html`.

2. Our estimation method

2.1. *Notation and inverse propensity score weighted estimation*

Let Y be the continuous response variable, $\mathbf{X}$ be the p-dimensional vector of covariates and A, taking values in $\mathcal{A}=\{1,0\}$, be the treatment indicator. It is assumed that a larger value of Y implies better response. The observed data are $\{(Y_i,\mathbf{X}_i,A_i), i=1,\ldots,n\}$, which are independently and identically distributed across i. Let $Y^*(a)$ denote the potential outcome that would result if the subject were given treatment $a\in\mathcal{A}$. A treatment regime $d(\mathbf{x})$ is a deterministic function that maps $\mathbf{x}\in\mathcal{X}$ to $a\in\mathcal{A}$. An optimal treatment regime in class $\mathcal{D}$ is defined as $d^{\text{opt}}=\arg\max_{d\in\mathcal{D}} E[Y^*\{d(\mathbf{X})\}]$, where $\mathcal{D}$ is a class of treatment regimes of interest. For example, we may consider a class of linear decision rules $d(\mathbf{x})=I(\boldsymbol{\beta}'\mathbf{x}\geqslant c)$, where $\boldsymbol{\beta}$ is a p-dimensional vector of parameters and c is a scalar. Here $E[Y^*\{d(\mathbf{X})\}]$ is called the value function of a given treatment regime d. To estimate the value function on the basis of observed data, two assumptions are typically made:

(a) (the stable unit treatment value assumption) $Y=Y^*(1)A+Y^*(0)(1-A)$;
(b) (the no-unmeasured-confounders assumption) $A\perp\{Y^*(0),Y^*(1)\}|\mathbf{X}$.

On the basis of these two assumptions, Zhang *et al.* (2012b) proposed an IPSW estimator for the value function, i.e.

$$\hat{V}_n(d)=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i I\{A_i=d(\mathbf{X}_i)\}}{A_i\pi(\mathbf{X}_i)+(1-A_i)\{1-\pi(\mathbf{X}_i)\}},\tag{1}$$

where $\pi(\mathbf{X}_i)=P(A_i=1|\mathbf{X}_i)$ is the propensity score. To search for the optimal treatment regime in a class of linear decision rules, they suggested maximizing $\hat{V}_n(d)\equiv\hat{V}_n(\boldsymbol{\beta},c)$ with respect to $\boldsymbol{\beta}$ and c under the constraint $\|(c,\boldsymbol{\beta}')'\|=1$, where $\|a\|$ is the Euclidean norm of a vector a.

2.2. *Concordance-assisted learning*

From now on, we consider linear decision rules $d(\mathbf{x})=I(\boldsymbol{\beta}'\mathbf{x}\geqslant c)$ for simplicity. The value function $E[Y^*\{d(\mathbf{X})\}]=E[\{Y^*(1)-Y^*(0)\}d(\mathbf{X})]+E\{Y^*(0)\}$. To maximize the value function, it is equivalent to maximize $E[\{Y^*(1)-Y^*(0)\}d(\mathbf{X})]$. Here $Y^*(1)-Y^*(0)$ is the gain of receiving treatment 1 against treatment 0 of a subject. The optimal treatment regime that maximizes the estimated value function tends to assign a subject to treatment 1 if his or her $Y^*(1)-Y^*(0)>0$ and 0 otherwise. For an optimal treatment regime, it is also natural to require that for any two subjects i and j, if $Y_i^*(1) - Y_i^*(0) > Y_j^*(1) - Y_j^*(0)$, the regime should be more likely to assign subject i to treatment 1 compared with subject j, i.e. $\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j$ in terms of linear decision rules. This motivates us to propose concordance-assisted learning (CAL) for estimating the optimal treatment regime in two steps. In the first step, we find $\boldsymbol{\beta}$ to maximize the concordance function, defined as

$$C(\boldsymbol{\beta}) = E([Y_i^*(1) - Y_i^*(0) - \{Y_j^*(1) - Y_j^*(0)\}]I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)),$$

with the constraint $\|\boldsymbol{\beta}\| = (\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2} = 1$. Let $\boldsymbol{\beta}^*$ denote the maximizer of $C(\boldsymbol{\beta})$. In the second step, we find c to maximize the value function $V(\boldsymbol{\beta}^*, c) = E[Y^*\{I(\boldsymbol{\beta}^{*\prime}\mathbf{X} \geqslant c)\}]$. Let c^* denote the maximizer. The optimal linear decision rule under CAL is $d^{*,\text{opt}}(\mathbf{x}) = I(\boldsymbol{\beta}^{*\prime}\mathbf{X} \geqslant c^*)$. Here, the index $\boldsymbol{\beta}^{*\prime}\mathbf{X}$ is named the prescriptive index, i.e. the larger the prescriptive index of a subject, the more benefit he or she tends to gain if assigned to treatment 1.

The optimal treatment regime that is defined under CAL may not maximize the value function in the class of linear decision rules, but it maximizes the concordance function. Moreover, among all linear decision rules that maximize the concordance function, it gives the maximal value function. Under the stable unit treatment value assumption and the no-unmeasured-confounders assumption, it can be shown that $E[\{Y^*(1) - Y^*(0)\}I(\boldsymbol{\beta}'\mathbf{X} \geqslant c)] = E\{D(\mathbf{X})I(\boldsymbol{\beta}'\mathbf{X} \geqslant c)\}$ and

$$E([Y_i^*(1) - Y_i^*(0) - \{Y_j^*(1) - Y_j^*(0)\}]I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)) = E[\{D(\mathbf{X}_i) - D(\mathbf{X}_j)\}I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)],$$

where $D(\mathbf{X}_i) = E(Y_i|A_i = 1, \mathbf{X}_i) - E(Y_i|A_i = 0, \mathbf{X}_i)$. For a class of monotonic index models with $D(\mathbf{X}_i) = Q(\boldsymbol{\beta}_0'\mathbf{X}_i)$, where $Q(\cdot)$ is an unspecified strictly monotone increasing function and $\|\boldsymbol{\beta}_0\| = 1$. Define $c_0 = Q^{-1}(0)$. Since $Q(\cdot)$ is a strictly monotone increasing function, it is easy to show that the maximizer of $E[Q(\boldsymbol{\beta}_0'\mathbf{X}_i)I(\boldsymbol{\beta}'\mathbf{X} \geqslant c)]$ is given by $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $c = c_0$. In addition, following similar arguments by Cavanagh and Sherman (1998) for the maximum rank correlation estimator, the maximizer of the concordance function $E[\{Q(\boldsymbol{\beta}_0'\mathbf{X}_i) - Q(\boldsymbol{\beta}_0'\mathbf{X}_j)\}I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)]$ can also be shown to be $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$. Similarly, we have $c^* = c_0$. Therefore, the optimal treatment regime under CAL coincides with the optimal treatment regime that maximizes the value function.

2.3. Estimation with known propensity score

In this section we assume that the propensity score model $\pi(\mathbf{X})$ is known as in randomized clinical trials. Similarly to the argument of A-learning (Murphy, 2003), we have

$$E\left[\frac{\{Y - \nu(\mathbf{X})\}\{A - \pi(\mathbf{X})\}}{\pi(\mathbf{X})\{1 - \pi(\mathbf{X})\}}\middle|\mathbf{X}\right] = E(Y|\mathbf{X}, A = 1) - E(Y|\mathbf{X}, A = 0) = D(\mathbf{X}),$$

where $\nu(\mathbf{X})$ is an arbitrary function of $\mathbf{X}$. This motivates us to consider the following estimator of the concordance function:

$$\begin{aligned}\hat{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) &= \frac{1}{n(n-1)} \sum_{i \neq j} \left[\frac{\{Y_i - \nu(\mathbf{X}_i, \hat{\boldsymbol{\theta}})\}\{A_i - \pi(\mathbf{X}_i)\}}{\pi(\mathbf{X}_i)\{1 - \pi(\mathbf{X}_i)\}} - \frac{\{Y_j - \nu(\mathbf{X}_j, \hat{\boldsymbol{\theta}})\}\{A_j - \pi(\mathbf{X}_j)\}}{\pi(\mathbf{X}_j)\{1 - \pi(\mathbf{X}_j)\}}\right] \\ &\quad \times I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j) \\ &\equiv \frac{1}{n(n-1)} \sum_{i \neq j} \Lambda_{ij}(\hat{\boldsymbol{\theta}}) I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j), \end{aligned} \qquad (2)$$

where $\nu(\mathbf{X}, \boldsymbol{\theta})$ is a posited parametric model for $\mu(\mathbf{X}) \equiv E(Y|\mathbf{X}, A = 0)$, such as a constant model or linear model, and $\hat{\boldsymbol{\theta}}$ is an estimator of $\boldsymbol{\theta}$. Define $\hat{\boldsymbol{\beta}} = \arg\max_{\|\boldsymbol{\beta}\|=1} \hat{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$. In addition, an estimator of c^* is given by $\hat{c} = \arg\max_c \hat{V}_n(\hat{\boldsymbol{\beta}}, c)$, where $\hat{V}_n(\hat{\boldsymbol{\beta}}, c)$ is the IPSW estimator of the value function for the decision rule $d(\mathbf{x}) = I(\hat{\boldsymbol{\beta}}'\mathbf{x} \geqslant c)$ as defined in equation (1). In our implementation, we consider a linear model for $\nu(\mathbf{X}, \boldsymbol{\theta})$, and $\hat{\boldsymbol{\theta}}$ is the associated least squares estimator based on data from subjects with $A = 0$. In addition, since $\hat{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ is a U-statistic with order of 2, it is much less bumpy than the IPSW estimator of the value function, and its optimization can be directly obtained by using the `optim` function in R even with relatively large p. Finally, since c is a scalar, the optimization of $\hat{V}_n(\hat{\boldsymbol{\beta}}, c)$ can be simply done by using a grid search.

Next we study the asymptotic properties of the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{c}$. For simplicity of presentation, we first introduce some notation. Define

$$\tau(\boldsymbol{\beta}, \mathbf{X}_1, \mathbf{X}_2) = \{D(\mathbf{X}_1) - D(\mathbf{X}_2)\} I(\boldsymbol{\beta}'\mathbf{X}_1 > \boldsymbol{\beta}'\mathbf{X}_2)$$

and

$$\varrho(\boldsymbol{\beta}, \mathbf{x}) = E\{\tau(\boldsymbol{\beta}, \mathbf{x}, \mathbf{X})\} + E\{\tau(\boldsymbol{\beta}, \mathbf{X}, \mathbf{x})\}.$$

Let $\nabla_m \varrho(\boldsymbol{\beta}, \mathbf{x})$ denote the mth partial derivative operator with respect to $\boldsymbol{\beta}$, and define

$$|\nabla_m| \varrho(\boldsymbol{\beta}, \mathbf{x}) = \sum_{i_1+\ldots+i_m=m} \left| \frac{\partial^m \varrho(\boldsymbol{\beta}, \mathbf{x})}{\partial \beta_{i_1} \ldots \partial \beta_{i_m}} \right|.$$

To establish the asymptotic results, we need the following conditions.

Condition 1. The propensity score $\pi(\mathbf{x})$ is known and $0 < \pi(\mathbf{x}) < 1$ for all $\mathbf{x} \in \mathcal{X}$.

Condition 2. The estimator $\hat{\boldsymbol{\theta}}$ converges almost surely to a deterministic vector of parameters $\boldsymbol{\theta}^*$, and $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = O_p(1)$.

Condition 3. The concordance function $C(\boldsymbol{\beta}) = E\{\tau(\boldsymbol{\beta}, \mathbf{X}_1, \mathbf{X}_2)\}$ has a unique maximizer at $\boldsymbol{\beta} = \boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_p^*)'$ with $\|\boldsymbol{\beta}^*\| = 1$.

Condition 4.

(a) The support of $\mathbf{X}$ is not contained in a proper linear subspace of $\mathbb{R}^p$.
(b) The density function of $\boldsymbol{\beta}'\mathbf{X}$ is everywhere positive for $\boldsymbol{\beta} \in \mathcal{B}$, where $\mathcal{B}$ is a neighbourhood of $\boldsymbol{\beta}^*$.
(c) $E\{D(\mathbf{X})^2\} < \infty$ and

$$E\left(\left[\frac{\{Y - \nu(\mathbf{X}, \boldsymbol{\theta}^*)\}\{A - \pi(\mathbf{X})\}}{\pi(\mathbf{X})\{1 - \pi(\mathbf{X})\}}\right]^2\right) < \infty.$$

Condition 5.

(a) The function $\varrho(\boldsymbol{\beta}, \mathbf{x})$ is twice differentiable with respect to $\boldsymbol{\beta}$.
(b) There is an integrable function $\Upsilon(\mathbf{x})$ such that, for any $\mathbf{x} \in \mathcal{X}$ and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ with $\|\boldsymbol{\beta}_1\| = \|\boldsymbol{\beta}_2\| = 1$, $\|\nabla_2 \varrho(\boldsymbol{\beta}_1, \mathbf{x}) - \nabla_2 \varrho(\boldsymbol{\beta}_2, \mathbf{x})\| < \Upsilon(\mathbf{x}) \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|$.
(c) $E\{|\nabla_1 \varrho(\boldsymbol{\beta}^*, \mathbf{X})|^2\} < \infty$ and $E\{|\nabla_2| \varrho(\boldsymbol{\beta}^*, \mathbf{X})\} < \infty$.
(d) $E\{\nabla_2 \varrho(\boldsymbol{\beta}^*, \mathbf{X})\}$ is negative definite.

Condition 6. The value function $V(\boldsymbol{\beta}^*, c) = E[Y^*\{I(\boldsymbol{\beta}^{*\prime}\mathbf{X} \geqslant c)\}]$ has a unique maximizer at $c = c^*$. In addition, there is a neighbourhood of c^* and a constant $M > 0$ such that $V(\boldsymbol{\beta}^*, c) - V(\boldsymbol{\beta}^*, c^*) \leqslant -M(c - c^*)^2$ for every c in this neighbourhood.

Condition 1 is assumed to simplify the theoretical arguments. It can be extended to the situation when the propensity score model is correctly specified, for example, by using a logistic regression. The parameters in the propensity score model can be consistently estimated from data. The variation of the estimators then needs to be taken into account when deriving the asymptotic variance of $\hat{\boldsymbol{\beta}}$. Condition 2 usually holds for the least squares estimator under mild conditions. Conditions 3 and 4 are assumed to establish the consistency of $\hat{\boldsymbol{\beta}}$. In particular, condition 3 assumes the existence and uniqueness of population parameters that maximize the concordance function. For the class of monotonic index models, we have $C(\boldsymbol{\beta}) = E[\{Q(\boldsymbol{\beta}_0'\mathbf{X}_i) - Q(\boldsymbol{\beta}_0'\mathbf{X}_j)\}I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)]$. Following similar arguments by Cavanagh and Sherman (1998), condition 2 can be shown to hold with $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$. Moreover, condition (b) of condition 4 is assumed to show the continuity of $C(\boldsymbol{\beta})$. It holds when there is a component of $\mathbf{X}$, say X_j, that has an everywhere positive density conditional on the rest covariates and $\beta_j^* \neq 0$. Condition (c) of condition 4 is assumed to show the uniform convergence of $\hat{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ to $C(\boldsymbol{\beta})$. Condition 5 is assumed to ensure the asymptotic normality of $\hat{\boldsymbol{\beta}}$. Conditions 4 and 5 are often used to establish the large sample properties of maximum rank correlation estimators (e.g. Sherman (1993) and Cavanagh and Sherman (1998)). Condition 6 is assumed to establish the consistency and convergence rate of $\hat{c}$. For the class of monotonic index models, we have

$$V(\boldsymbol{\beta}^*, c) = V(\boldsymbol{\beta}_0, c) = E\{Y_i^*(0)\} + E\{Q(\boldsymbol{\beta}_0'\mathbf{X}_i)I(\boldsymbol{\beta}_0'\mathbf{X}_i \geqslant c)\} = E\{Y_i^*(0)\} + \int_c^\infty Q(u)\, f(u)\, \mathrm{d}u,$$

where $f(\cdot)$ is the density function of $\boldsymbol{\beta}_0'\mathbf{X}_i$. Then, $\partial V(\boldsymbol{\beta}^*, c)/\partial c = -Q(c)\, f(c)$, which is less than 0 if $c > c_0 = Q^{-1}(0)$ and bigger than 0 otherwise. Therefore, $V(\boldsymbol{\beta}^*, c)$ has a unique maximizer with $c^* = c_0$. In addition, $\partial^2 V(\boldsymbol{\beta}^*, c)/\partial c^2 = -\dot{Q}(c)\, f(c) - Q(c)\, \dot{f}(c)$, where $\dot{a}(c)$ is the first derivative of $a(c)$. Then, $\partial^2 V(\boldsymbol{\beta}^*, c)/\partial c^2|_{c=c^*} = -\dot{Q}(c^*)\, f(c^*) < 0$. This implies that there is a neighbourhood of c^* and a constant $M > 0$ such that $V(\boldsymbol{\beta}^*, c) - V(\boldsymbol{\beta}^*, c^*) \leqslant -M(c - c^*)^2$ for every c in this neighbourhood. Condition 6 holds.

Theorem 1. Under conditions 1–5, we have, as $n \to \infty$:

(a) $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\| \to 0$ almost surely;
(b) $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \to N(0, \boldsymbol{\Sigma})$ in distribution, where $\boldsymbol{\Sigma} = V^{-1}\Delta(V^{-1})'$, $2V = E\{\nabla_2 \varrho(\boldsymbol{\beta}^*, \mathbf{X})\}$ and $\Delta = E\{\nabla_1 \varrho(\boldsymbol{\beta}^*, \mathbf{X})\nabla_1 \varrho(\boldsymbol{\beta}^*, \mathbf{X})'\}$.

Theorem 2. Under conditions 1–6, we have, as $n \to \infty$:

(a) $|\hat{c} - c^*| = O_p(n^{-1/3})$;
(b) $n^{1/3}(\hat{c} - c^*)$ converges in distribution to $\arg\max_h G(h)$, where $G(h)$ is a two-sided Gaussian process defined in Appendix A.

Proofs of theorems 1 and 2 are given in Appendix A. Since $\hat{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ is not a smooth function of $\boldsymbol{\beta}$, to estimate the asymptotic variance matrix of $\hat{\boldsymbol{\beta}}$, we derive an induced smoothing method that is similar to those studied in Brown and Wang (2005, 2007) and Pang *et al.* (2012). Define the smoothed concordance function $\tilde{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \mathbf{H}) = E\{\hat{C}_n(\boldsymbol{\beta} + \mathbf{H}^{1/2}\mathbf{U}, \hat{\boldsymbol{\theta}})\}$, where $\mathbf{U}$ is a standard p-variate normal random vector, $\mathbf{H}$ is a $p \times p$ positive definite matrix with order $O(n^{-1})$ and the expectation is taken with respect to the distribution of $\mathbf{U}$. We have

$$\tilde{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \mathbf{H}) = \frac{1}{n(n-1)} \sum_{i \neq j} \Lambda_{ij}(\hat{\boldsymbol{\theta}}) \left\{ 1 - \Phi\left(\frac{\mathbf{X}_{ij}'\boldsymbol{\beta}}{\sigma_{ij}^{\mathbf{H}}} \right) \right\},$$

where $\mathbf{X}_{ij} = \mathbf{X}_j - \mathbf{X}_i$, $\sigma_{ij}^{\mathbf{H}} = (\mathbf{X}_{ij}'\mathbf{H}\mathbf{X}_{ij})^{1/2}$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Denote the first and second derivatives of $\tilde{C}_n(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \mathbf{H})$ with respect to $\boldsymbol{\beta}$ by

$$\tilde{S}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}},\mathbf{H})=\frac{\partial\tilde{C}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}},\mathbf{H})}{\partial\boldsymbol{\beta}}=\frac{1}{n(n-1)}\sum_{i\neq j}\Lambda_{ij}(\hat{\boldsymbol{\theta}})\phi\left(\frac{\mathbf{X}_{ij}'\boldsymbol{\beta}}{\sigma_{ij}^{\mathbf{H}}}\right)\left(\frac{\mathbf{X}_{ji}}{\sigma_{ij}^{\mathbf{H}}}\right),$$

$$\tilde{V}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}},\mathbf{H})=\frac{\partial^2\tilde{C}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}},\mathbf{H})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}=-\frac{1}{n(n-1)}\sum_{i\neq j}\Lambda_{ij}(\hat{\boldsymbol{\theta}})\dot{\phi}\left(\frac{\mathbf{X}_{ij}'\boldsymbol{\beta}}{\sigma_{ij}^{\mathbf{H}}}\right)\left(\frac{\mathbf{X}_{ij}}{\sigma_{ij}^{\mathbf{H}}}\right)^{\otimes 2},$$

where $\phi(\cdot)$ is the density function of the standard normal distribution, $\dot{\phi}(\cdot)$ is the first derivative of $\phi(\cdot)$ and, for a vector $\mathbf{v}$, $\mathbf{v}^{\otimes 2}=\mathbf{v}\mathbf{v}'$. Then, an estimator of $\boldsymbol{\Sigma}$ is given by

$$\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H})=\tilde{V}_n^{-1}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H})\tilde{\Delta}_n(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H})(\tilde{V}_n^{-1}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H}))',$$

where

$$\begin{aligned}\tilde{\Delta}_n(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H})&=n\widehat{\text{var}}\{\tilde{S}_n(\boldsymbol{\beta}^*,\hat{\boldsymbol{\theta}},\mathbf{H})\}\\&=\frac{4}{n^3}\sum_{i=1}^{n}\left\{\sum_{j}\Lambda_{ij}(\hat{\boldsymbol{\theta}})\phi\left(\frac{\mathbf{X}_{ij}'\hat{\boldsymbol{\beta}}}{\sigma_{ij}^{\mathbf{H}}}\right)\left(\frac{\mathbf{X}_{ji}}{\sigma_{ij}^{\mathbf{H}}}\right)\right\}^{\otimes 2}.\end{aligned}$$

We have the following theorem.

Theorem 3. Under conditions 1–5, we have that, for any positive definite matrix $\mathbf{H}$, $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H})\to\boldsymbol{\Sigma}$ almost surely, as $n\to\infty$.

The proof of theorem 3 can follow similar arguments to those of Zhang, Jin, Shao and Ying (2013) for a self-induced smoothing approach for transformation models. The details are omitted here. In our numerical implementation, we set $\mathbf{H}=n^{-1}I_p$ for simplicity. From our numerical experience, the results are not sensitive to the choice of $\mathbf{H}$ as long as it is of the order of $O(n^{-1})$. In practice, one more iteration may help to improve slightly the accuracy of the estimated variance of $\hat{\boldsymbol{\beta}}$. Specifically, we initially set $\mathbf{H}$ as $\mathbf{H}^{(0)}=n^{-1}I_p$ and compute $\hat{\Sigma}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H}^{(0)})$. Then, we update $\mathbf{H}$ as $\mathbf{H}^{(1)}=n^{-1}\hat{\Sigma}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}},\mathbf{H}^{(0)})$. Moreover, a new estimator $\tilde{\boldsymbol{\beta}}$ can be defined as the maximizer of the smoothed concordance function $\tilde{C}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}},\mathbf{H})$. It can be shown that $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ have the same asymptotic distribution, whereas $\tilde{\boldsymbol{\beta}}$ may have slightly smaller standard deviation in finite samples compared with $\hat{\boldsymbol{\beta}}$.

Next, we establish the asymptotic distribution of the estimated value function, which is given by

$$\hat{V}_n(\boldsymbol{\beta},c)=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i I\{A_i=I(\boldsymbol{\beta}'\mathbf{X}\geqslant c)\}}{A_i\pi(\mathbf{X}_i)+(1-A_i)\{1-\pi(\mathbf{X}_i)\}}.$$

Define $V(\boldsymbol{\beta},c)=E[Y^*\{I(\boldsymbol{\beta}'\mathbf{X}\geqslant c)\}]$.

Theorem 4. Under conditions 1–6, we have, as $n\to\infty$,

$$n^{1/2}\{\hat{V}_n(\hat{\boldsymbol{\beta}},\hat{c})-V(\boldsymbol{\beta}^*,c^*)\}\to N(0,\boldsymbol{\Sigma}_V)$$

in distribution, where $\boldsymbol{\Sigma}_V$ is defined in Appendix A.

2.4. Doubly robust estimation

If the propensity score $\pi(\mathbf{X})$ is unknown as in observational studies, a parametric model, such as a logistic regression model, is usually assumed for the propensity score. Let $\pi(\mathbf{X},\boldsymbol{\alpha})$ denote the posited propensity score model. The parameters $\boldsymbol{\alpha}$ can be estimated from the observed data; let $\hat{\boldsymbol{\alpha}}$ denote the estimator. If the propensity score model is misspecified, although $\pi(\mathbf{X},\hat{\boldsymbol{\alpha}})$ may not consistently estimate the true propensity score $\pi(\mathbf{X})$, it can be shown that $\hat{\boldsymbol{\alpha}}$ converges almost surely to a deterministic vector of parameters $\boldsymbol{\alpha}^*$ under mild conditions. Under such a situation, $\hat{C}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})$ is not a consistent estimator of the concordance function. To improve the robustness of the estimator $\hat{\boldsymbol{\beta}}$, which was proposed in the previous section, we develop a doubly robust estimation method for the class of monotonic index models: $D(\mathbf{X}) = Q(\boldsymbol{\beta}_0'\mathbf{X})$.

Recall that $\mu(\mathbf{X}) = E(Y|\mathbf{X}, A=0)$. We have

$$\begin{aligned}E\left[\frac{\{Y-\mu(\mathbf{X})\}\{A-\pi(\mathbf{X},\boldsymbol{\alpha})\}}{\pi(\mathbf{X},\boldsymbol{\alpha})\{1-\pi(\mathbf{X},\boldsymbol{\alpha})\}}|\mathbf{X}\right] &= E\left[\frac{A\,Q(\boldsymbol{\beta}_0'\mathbf{X})\{A-\pi(\mathbf{X},\boldsymbol{\alpha})\}}{\pi(\mathbf{X},\boldsymbol{\alpha})\{1-\pi(\mathbf{X},\boldsymbol{\alpha})\}}|\mathbf{X}\right]\\ &= Q(\boldsymbol{\beta}_0'\mathbf{X})\frac{\pi(\mathbf{X})}{\pi(\mathbf{X},\boldsymbol{\alpha})}.\end{aligned}$$

Define

$$\Lambda_{ij}^{\mathrm{DR}}(\boldsymbol{\theta},\boldsymbol{\alpha}) = \frac{\{Y_i-\nu(\mathbf{X}_i,\boldsymbol{\theta})\}\{A_i-\pi(\mathbf{X}_i,\boldsymbol{\alpha})\}A_j}{\pi(\mathbf{X}_i,\boldsymbol{\alpha})\{1-\pi(\mathbf{X}_i,\boldsymbol{\alpha})\}\pi(\mathbf{X}_j,\boldsymbol{\alpha})} - \frac{\{Y_j-\nu(\mathbf{X}_j,\boldsymbol{\theta})\}\{A_j-\pi(\mathbf{X}_j,\boldsymbol{\alpha})\}A_i}{\pi(\mathbf{X}_j,\boldsymbol{\alpha})\{1-\pi(\mathbf{X}_j,\boldsymbol{\alpha})\}\pi(\mathbf{X}_i,\boldsymbol{\alpha})},$$

where $\nu(\mathbf{X},\boldsymbol{\theta})$ is a posited parametric model for $\mu(\mathbf{X})$. Then, if the propensity score model is correctly specified,

$$E\{\Lambda_{ij}^{\mathrm{DR}}(\boldsymbol{\theta},\boldsymbol{\alpha})I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)\} = E[\{Q(\boldsymbol{\beta}_0'\mathbf{X}_i) - Q(\boldsymbol{\beta}_0'\mathbf{X}_j)\}I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)];$$

in contrast, if the baseline mean model $\nu(\mathbf{X},\boldsymbol{\theta})$ is correctly specified,

$$E\{\Lambda_{ij}^{\mathrm{DR}}(\boldsymbol{\theta},\boldsymbol{\alpha})I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)\} = E\left[\frac{\pi(\mathbf{X}_i)\,\pi(\mathbf{X}_j)}{\pi(\mathbf{X}_i,\boldsymbol{\alpha})\,\pi(\mathbf{X}_j,\boldsymbol{\alpha})}\{Q(\boldsymbol{\beta}_0'\mathbf{X}_i) - Q(\boldsymbol{\beta}_0'\mathbf{X}_j)\}I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)\right].$$

Under either case, the maximizer of $E\{\Lambda_{ij}^{\mathrm{DR}}(\boldsymbol{\theta},\boldsymbol{\alpha})I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j)\}$ is $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. This motivates us to consider the loss function

$$\begin{aligned}\hat{C}_n^{\mathrm{DR}}(\boldsymbol{\beta},\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) = \frac{1}{n(n-1)}\sum_{i\neq j}&\left[\frac{\{Y_i-\nu(\mathbf{X}_i,\hat{\boldsymbol{\theta}})\}\{A_i-\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}})\}A_j}{\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}})\{1-\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}})\}\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}})}\right.\\ &\left.-\frac{\{Y_j-\nu(\mathbf{X}_j,\hat{\boldsymbol{\theta}})\}\{A_j-\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}})\}A_i}{\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}})\{1-\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}})\}\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}})}\right]I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j). \qquad (3)\end{aligned}$$

Denote $\hat{\boldsymbol{\beta}}^{\mathrm{DR}} = \arg\max_{\|\boldsymbol{\beta}\|=1}\hat{C}_n^{\mathrm{DR}}(\boldsymbol{\beta},\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})$. We have the following theorem.

Theorem 5. Assume that either the propensity score model $\pi(\mathbf{X},\boldsymbol{\alpha})$ or the baseline mean model $\nu(\mathbf{X},\boldsymbol{\theta})$ is correctly specified and that $D(\mathbf{X}) = E\{Y^*(1) - Y^*(0)|\mathbf{X}\} = Q(\boldsymbol{\beta}_0'\mathbf{X})$. Under conditions 1′–4′ given in the on-line appendix, we have, as $n\to\infty$:

(a) $\|\hat{\boldsymbol{\beta}}^{\mathrm{DR}} - \boldsymbol{\beta}_0\| \to 0$ almost surely;
(b) $n^{1/2}(\hat{\boldsymbol{\beta}}^{\mathrm{DR}} - \boldsymbol{\beta}_0) \to N(0,\boldsymbol{\Sigma}^{\mathrm{DR}})$ in distribution, where the asymptotic variance matrix $\boldsymbol{\Sigma}^{\mathrm{DR}}$ is defined in Appendix A.

Since the asymptotic variance matrix $\boldsymbol{\Sigma}^{\mathrm{DR}}$ has a very complicated form, the direct estimation of $\boldsymbol{\Sigma}^{\mathrm{DR}}$ may be difficult. Following similar techniques to those in Jin *et al.* (2001), we derive a resampling method to estimate $\boldsymbol{\Sigma}^{\mathrm{DR}}$. Specifically, we consider the perturbed loss function

$$\begin{aligned}\hat{C}_n^{\mathrm{DR},*}(\boldsymbol{\beta},\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*) = \frac{1}{n(n-1)}\sum_{i\neq j}\xi_i\xi_j&\left[\frac{\{Y_i-\nu(\mathbf{X}_i,\hat{\boldsymbol{\theta}}^*)\}\{A_i-\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}}^*)\}A_j}{\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}}^*)\{1-\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}}^*)\}\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}}^*)}\right.\\ &\left.-\frac{\{Y_j-\nu(\mathbf{X}_j,\hat{\boldsymbol{\theta}}^*)\}\{A_j-\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}}^*)\}A_i}{\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}}^*)\{1-\pi(\mathbf{X}_j,\hat{\boldsymbol{\alpha}}^*)\}\pi(\mathbf{X}_i,\hat{\boldsymbol{\alpha}}^*)}\right]I(\boldsymbol{\beta}'\mathbf{X}_i > \boldsymbol{\beta}'\mathbf{X}_j),\end{aligned}$$

where $\xi_1,\ldots,\xi_n$ are independent and identically distributed exponential variables with mean 1,

$$\hat{\boldsymbol{\theta}}^* = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \xi_i(1-A_i)\{Y_i - \nu(\mathbf{X}_i,\boldsymbol{\theta})\}^2,$$

and $\hat{\boldsymbol{\alpha}}^*$ is the solution to the equation

$$\arg\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \xi_i(1,\mathbf{X}_i')'\{A_i - \pi(\mathbf{X},\boldsymbol{\alpha})\} = 0.$$

Denote $\hat{\boldsymbol{\beta}}^{\text{DR},*} = \arg\max_{\|\boldsymbol{\beta}\|=1} \hat{C}_n^{\text{DR},*}(\boldsymbol{\beta},\hat{\boldsymbol{\theta}}^*,\hat{\boldsymbol{\alpha}}^*)$. We can use the empirical variance matrix of $\hat{\boldsymbol{\beta}}^{\text{DR},*}$ to estimate $\boldsymbol{\Sigma}^{\text{DR}}$.

3. Simulations

3.1. Simulations for monotonic index models

In the first set of simulations, we consider a class of monotonic index models

$$Y = \mu(\mathbf{X}) + AD(\mathbf{X}) + \epsilon, \tag{4}$$

where $\mathbf{X} = (X_1, X_2, X_3, X_4)'$, A is generated from Bernoulli$\{\pi(\mathbf{X})\}$ and ϵ is generated from $N(0, 0.5^2)$. Four cases are studied:

(a) case I, $\mu(\mathbf{X}) = 1 + \boldsymbol{\gamma}_1'\mathbf{X}$ and $D(\mathbf{X}) = 2\boldsymbol{\beta}_0'\mathbf{X}$;
(b) case II, $\mu(\mathbf{X}) = 1 + \boldsymbol{\gamma}_1'\mathbf{X}$ and $D(\mathbf{X}) = \exp(\boldsymbol{\beta}_0'\mathbf{X}) - 1$;
(c) case III, $\mu(\mathbf{X}) = 1 + \sin(\boldsymbol{\gamma}_1'\mathbf{X}) + 0.5(\boldsymbol{\gamma}_2'\mathbf{X})^2$ and $D(\mathbf{X}) = (\boldsymbol{\beta}_0'\mathbf{X})^3$;
(d) case IV, $\mu(\mathbf{X}) = 1 + X_1X_2 + 0.5X_3^2$ and $D(\mathbf{X}) = (\boldsymbol{\beta}_0'\mathbf{X})^3$.

Here X_1, X_2, X_3 and X_4 are independent and identically distributed standard normal random variables. In each case, we set $\boldsymbol{\beta}_0 = (1, 1, -1, 1)'$, $\boldsymbol{\gamma}_1 = (1, -1, 1, 1)'$ and $\boldsymbol{\gamma}_2 = (1, 0, -1, 0)'$. For all the cases, the optimal treatment regime that maximizes the value function and the optimal treatment regime that is defined by our proposed CAL are the same, which is given by $d^{\text{opt}}(\mathbf{x}) = I(\boldsymbol{\beta}_0'\mathbf{x} \geqslant c_0)$ with $c_0 = 0$ and $\boldsymbol{\beta}_0 = (0.5, 0.5, -0.5, 0.5)'$ after imposing the constraint $\|\boldsymbol{\beta}_0\| = 1$. For each case, we carry out 500 simulation runs with sample size $n = 200$. The optimization in the method proposed is done by the `optim` function in R with the default method 'Nelder-Mead' for searching the maximizer.

For the propensity score model, we first consider randomized trials with $\pi(\mathbf{X}) = 0.5$. We compare the IPSW and AIPSW estimators of Zhang *et al.* (2012b), the estimator obtained based on a linear regression (LR) model for both the baseline covariate effect and treatment–covariates interaction, the proposed CAL estimator and its doubly robust variant, denoted by CAL-DR. In all the estimators, we assume that the propensity score is known and is set as 0.5. For the AIPSW estimator, a linear model was fitted for the augmented term as done in Zhang *et al.* (2012b). Under case I, the fitted linear model is correctly specified, whereas, under cases II–IV, it is not. For the CAL and CAL-DR estimators, we consider a linear model for $\nu(\mathbf{X},\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is estimated on the basis of data from subjects with $A = 0$. For the IPSW, AIPSW and LR estimators, we normalize $\hat{\boldsymbol{\beta}}$ to have norm 1 for comparisons with the CAL and CAL-DR estimators and adjust $\hat{c}$ accordingly by $\hat{c}/\|\hat{\boldsymbol{\beta}}\|$. To assess the performance of the estimators, for $\hat{\boldsymbol{\beta}}$, we report the mean and standard deviation of the estimators, the mean of the estimated standard errors and the empirical coverage probability of 95% Wald-type confidence intervals; for $\hat{c}$, we report the mean and standard deviation of the estimators. Here, the standard error of the CAL estimator for $\boldsymbol{\beta}$ is estimated by the induced smoothing method proposed and that of the CAL-DR estimator is estimated by the resampling method proposed. To evaluate the accuracy of the estimated optimal treatment regime $\hat{d}^{\text{opt}}(\mathbf{x}) = I(\hat{\boldsymbol{\beta}}'\mathbf{x} \geqslant \hat{c})$, we report the mean and standard deviation of the percentages of making correct decisions, PCD, defined as $1 - n^{-1}\Sigma_{i=1}^{n}|I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i \geqslant \hat{c}) - I(\boldsymbol{\beta}_0'\mathbf{X}_i \geqslant c_0)|$. Furthermore, we report the mean and standard deviation of the value functions and concordance functions for the estimated optimal treatment regime, which are obtained via simulations. Specifically, we generate data for $N = 10\,000$ subjects from model (4) and obtain the value function and the concordance function for $\hat{d}^{\text{opt}}(\mathbf{x})$ by

$$\hat{V} = \frac{1}{N}\sum_{i=1}^{N}\{\mu(\mathbf{X}_i) + \hat{d}^{\text{opt}}(\mathbf{X}_i)D(\mathbf{X}_i)\}$$

and

$$\hat{C} = \frac{1}{N(N-1)}\sum_{i \neq j}\{D(\mathbf{X}_i) - D(\mathbf{X}_j)\}I(\hat{\boldsymbol{\beta}}'\mathbf{X}_i > \hat{\boldsymbol{\beta}}'\mathbf{X}_j)$$

respectively. Similarly, we can compute the value function and the concordance function for the true optimal treatment regime, $d^{\text{opt}}(\mathbf{x})$, denoted by V_0 and C_0 respectively. In all the tables, we report the mean of the estimators, Est, the standard deviation of the estimators, SD, the mean of the estimated standard errors, SE, and the empirical coverage probability CP% of a Wald-type 95% confidence interval.

The simulation results for cases I and II are summarized in Table 1 and those for cases III and IV are given in the on-line supplementary appendix for brevity. From the simulation results, we make the following observations. First, our proposed CAL and CAL-DR estimators for $\boldsymbol{\beta}$ are nearly unbiased under all cases and have smaller biases than the IPSW and AIPSW estimators. In addition, the standard deviations of the CAL and CAL-DR estimators for $\boldsymbol{\beta}$ are much smaller than those of the IPSW and AIPSW estimators, indicating the big gain in efficiency of CAL for estimating the optimal treatment regime. One possible reason is that the CAL and CAL-DR estimators have a faster rate of convergence than those of the IPSW and AIPSW estimators. Second, the mean of the estimated standard errors of the CAL and CAL-DR estimators for $\boldsymbol{\beta}$ is close to the standard deviation of the estimators and the empirical coverage probability of 95% confidence intervals is close to the nominal level under all cases. Third, the CAL-DR estimator for $\boldsymbol{\beta}$ has smaller standard deviations than the CAL estimator, indicating that the CAL-DR estimator may be more efficient than the CAL estimator. Fourth, the estimated value function and concordance function of the optimal treatment regimes that are obtained by the CAL and CAL-DR estimators are all close to their true values and are larger than those obtained by the IPSW estimator. Fifth, the PCDs of the optimal treatment regimes obtained by the CAL and CAL-DR estimators range from 0.875 to 0.925, which are much higher than those obtained by the IPSW estimator. Sixth, for case I, where the fitted linear model was correctly specified, the AIPSW estimators of $\boldsymbol{\beta}$ have much smaller standard deviations than the IPSW estimators, showing the gain in efficiency of the AIPSW estimators as studied in the literature. However, for cases II–IV, where the fitted linear model was misspecified, the AIPSW and IPSW estimators have comparable performance in terms of the value function and PCD. Finally, the LR estimator has the best performance under case I as expected since the LR model is correctly specified, but it may have worse performance than the CAL and CAL-DR estimators when the linear model is misspecified. For example, in case II, the LR estimators of $\boldsymbol{\beta}$ have larger biases and standard deviations, and the optimal treatment regimes that are obtained by the LR estimators give smaller value function, PCD and concordance function.

For comparison, we also implemented the method of Zhao *et al.* (2013). For brevity, the results are provided in the on-line supplementary appendix. We made the following observations. First,

Table 1. Simulation results for cases I and II†

Method	*Statistic*	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{c}$	$\hat{V}$	*PCD*	$\hat{C}$
Case I ($V_0=2.595$; $C_0=2.258$)									
IPSW	Est	0.456	0.591	−0.561	0.495	0.063	2.473	0.883	2.131
	SD	0.202	0.193	0.153	0.152	0.237	0.104	0.053	0.125
AIPSW	Est	0.575	0.568	−0.574	0.572	0.002	2.585	0.962	2.244
	SD	0.058	0.060	0.062	0.087	0.080	0.034	0.020	0.021
LR	Est	0.500	0.500	−0.500	0.499	0.000	2.597	0.989	2.257
	SD	0.016	0.016	0.016	0.015	0.018	0.032	0.008	0.017
CAL	Est	0.498	0.502	−0.495	0.497	0.040	2.543	0.920	2.249
	SD	0.048	0.043	0.042	0.046	0.234	0.066	0.055	0.019
	SE	0.051	0.051	0.050	0.050	—	—	—	—
	CP (%)	95.0	96.6	96.2	94.8	—	—	—	—
CAL-DR	Est	0.501	0.501	−0.497	0.499	0.031	2.545	0.925	2.256
	SD	0.022	0.021	0.021	0.019	0.222	0.009	0.056	0.017
	SE	0.023	0.023	0.021	0.024	—	—	—	—
	CP (%)	95.4	97.2	93.4	98.0	—	—	—	—
Case II ($V_0=7.723$; $C_0=6.248$)									
IPSW	Est	0.376	0.525	−0.530	0.453	−0.137	7.558	0.818	5.699
	SD	0.311	0.284	0.245	0.215	0.334	0.259	0.088	0.877
AIPSW	Est	0.422	0.403	−0.408	0.432	0.160	7.471	0.773	5.209
	SD	0.291	0.325	0.324	0.392	0.416	0.629	0.127	1.405
LR	Est	0.481	0.479	−0.486	0.480	0.472	7.663	0.811	6.125
	SD	0.135	0.134	0.136	0.136	0.114	0.538	0.042	0.555
CAL	Est	0.498	0.491	−0.503	0.496	−0.097	7.667	0.875	6.226
	SD	0.054	0.057	0.057	0.057	0.456	0.179	0.090	0.526
	SE	0.064	0.062	0.063	0.062	—	—	—	—
	CP (%)	97.2	95.0	95.6	95.6	—	—	—	—
CAL-DR	Est	0.498	0.497	−0.504	0.498	−0.078	7.668	0.875	6.244
	SD	0.025	0.025	0.025	0.023	0.455	0.185	0.094	0.525
	SE	0.026	0.026	0.023	0.026	—	—	—	—
	CP (%)	95.8	96.2	92.4	96.2	—	—	—	—

†The true optimal regime is $d^{\mathrm{opt}}(\mathbf{x}) = I(\boldsymbol{\beta}_0'\mathbf{x} \geqslant c_0)$ with $\boldsymbol{\beta}_0 = (0.5, 0.5, -0.5, 0.5)'$ and $c_0 = 0$. Est, mean of estimators; SD, standard deviation of estimators; SE, mean of estimated standard errors; CP, empirical coverage probability of the 95% confidence interval.

the method of Zhao *et al.* (2013) depends on the choice of ξ. In practice, a data-adaptive way is needed for choosing the optimal ξ. Second, for case I, where the linear models are correctly specified, the method of Zhao *et al.* (2013) with the best choice of ξ and the methods proposed have comparable performance in terms of the value $\hat{V}$ and PCD. However, for cases II–IV, where fitted linear models are misspecified, the estimated optimal treatment rules that are obtained by the methods proposed give larger values and PCD than those of Zhao *et al.* (2013). Third, the estimated optimal treatment rules that are obtained by the methods proposed give larger concordance values $\hat{C}$ with much smaller standard deviations than those of Zhao *et al.* (2013) for all cases. In summary, the methods proposed showed very competitive performance.

In addition, we compare the methods proposed and the method of Matsouaka *et al.* (2014) to examine the inference of the estimated value function. The results are provided in the on-line supplementary appendix. All the methods have proper coverage probabilities that are close to the nominal level, and the estimators of the value functions have comparable standard deviations.

We also conducted simulations with $p = 10$ covariates, generated from the standard normal distribution. The results are provided in the supplementary appendix. It can be seen that the methods proposed give larger concordance, value and PCD compared with the IPSW method and require much shorter computational time on average. This demonstrates the ability of the methods for handling relatively large number of covariates.

Next we consider simulations when the propensity score $\pi(\mathbf{X})$ is unknown as in observation studies and the posited model is misspecified. The simulation results are given in the on-line supplementary appendix. We observe that the IPSW and CAL estimators have relatively large biases under all cases as expected; the CAL-DR estimator is nearly unbiased under cases I and II and has much smaller biases than the IPSW and CAL estimators under cases III and IV; the mean of the estimated standard errors of the CAL-DR estimators is close to the standard deviation of the estimators and the empirical coverage probability of 95% confidence intervals is close to the nominal level. These findings show that the CAL-DR estimator has the double-robustness property under cases I and II when the propensity score model is misspecified and has superior performance compared with the IPSW and CAL estimators under cases III and IV when both the baseline mean model and the propensity score model are misspecified.

3.2. Simulations for general models

In this section, we consider several models where a monotonic index model for $D(\mathbf{X})$ is violated and the true optimal treatment regime may or may not be defined by a linear decision rule. We first consider model (4) with the following setting: case V, $\mu(\mathbf{X}) = 1 + \sin(\gamma_1'\mathbf{X}) + 0.5(\gamma_2'\mathbf{X})^2$ and $D(\mathbf{X}) = |X_2^2 + X_1X_2 + X_3X_4|\mathrm{sgn}(X_2 - X_1^2 + X_3 - X_4^2)$, where γ_1 and γ_2 are defined the same as before. Here, we consider only randomized studies with $\pi(\mathbf{X}) = 0.5$. In this case, the true optimal treatment regime is given by $d^{\mathrm{opt}}(\mathbf{X}) = I(X_2 - X_1^2 + X_3 - X_4^2 > 0)$, which is not a linear decision rule. We compare the mean and standard deviation of the IPSW, AIPSW, LR, CAL and CAL-DR estimators and the value function and concordance function of the estimated optimal treatment regime obtained by the various methods. For all the methods, we search the optimal treatment regime in the class of linear decision rules. The simulation results are summarized in Table 2. From the simulation results, it can be seen that the CAL and CAL-DR methods give similar estimates of $\boldsymbol{\beta}$ and c, which may have a relatively big difference from the estimates that are obtained by the IPSW and AIPSW methods; the concordance function of the estimated optimal treatment regimes obtained by the CAL and CAL-DR methods are larger than those obtained by the IPSW and AIPSW methods; the value function of the estimated optimal treatment regimes obtained by the CAL and CAL-DR methods are comparable or slightly larger than those obtained by the IPSW, AIPSW and LR methods and are closer to the value function of the true optimal treatment regime $d^{\mathrm{opt}}(\mathbf{X}) = I(X_2 - X_1^2 + X_3 - X_4^2 > 0)$.

In addition, we consider the model

$$Y = \exp\{\mu(\mathbf{X}) + A\,Q(\mathbf{X})\} + \epsilon,$$

where $\epsilon \sim N(0, 0.5^2)$, $A \sim \mathrm{Bernoulli}(0.5)$ and $\mathbf{X}$ is generated the same as before. The following case is studied: case VI, $\mu(\mathbf{X}) = 0.1\{1 + (X_1X_2 + 0.5X_3^2)\}$ and $Q(\mathbf{X}) = \boldsymbol{\beta}_0'\mathbf{X} - 1$, where $\boldsymbol{\beta}_0$ is defined the same as before. In this case, the true optimal treatment regime is given by $d^{\mathrm{opt}}(\mathbf{X}) = I(\boldsymbol{\beta}_0'\mathbf{X} > 1)$, which is a linear decision rule. Here, the contrast function $D(\mathbf{X}) = \exp\{\mu(\mathbf{X}) + Q(\mathbf{X})\} - \exp\{\mu(\mathbf{X})\}$, which is not a monotonic index model. For all the methods, we search the optimal treatment regime in the class of linear decision rules. Thus, the IPSW and AIPSW estimators are consistent but the CAL and CAL-DR estimators are not. The simulation results are also summarized in Table 2. From the simulation results, we observe that the IPSW and AIPSW estimators of $\boldsymbol{\beta}$ have relatively small biases under all cases as expected; compared with the IPSW and AIPSW estimators, the CAL and CAL-DR estimators have comparable

Table 2. Simulation results for cases V and VI

Method	*Statistic*	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{c}$	$\hat{V}$	*PCD*	$\hat{C}$
Case V ($V_0=3.297$)									
IPSW	Est	0.088	0.447	0.209	0.023	0.596	2.877	0.712	0.273
	SD	0.500	0.468	0.350	0.446	0.457	0.131	0.100	0.184
AIPSW	Est	0.057	0.516	0.260	0.011	0.640	2.912	0.739	0.310
	SD	0.532	0.461	0.338	0.409	0.413	0.123	0.100	0.179
LR	Est	0.080	0.570	0.254	0.022	−0.641	2.902	0.734	0.331
	SD	0.457	0.315	0.436	0.326	0.536	0.113	0.086	0.156
CAL	Est	0.091	0.575	0.267	0.018	0.836	2.915	0.735	0.342
	SD	0.448	0.282	0.398	0.390	0.759	0.122	0.106	0.135
CAL-DR	Est	0.085	0.585	0.258	0.023	0.848	2.920	0.738	0.342
	SD	0.448	0.287	0.410	0.365	0.753	0.119	0.101	0.140
Case VI ($V_0=4.366$)									
IPSW	Est	0.444	0.486	−0.445	0.468	0.527	4.259	0.890	3.135
	SD	0.198	0.196	0.194	0.194	0.189	0.447	0.051	0.453
AIPSW	Est	0.448	0.426	−0.423	0.434	0.355	4.204	0.845	2.972
	SD	0.412	0.241	0.243	0.287	0.229	0.498	0.124	0.732
LR	Est	0.483	0.485	−0.477	0.461	0.306	4.175	0.685	3.195
	SD	0.151	0.142	0.161	0.154	0.095	0.421	0.040	0.427
CAL	Est	0.498	0.495	−0.508	0.480	0.548	4.338	0.935	3.255
	SD	0.075	0.071	0.064	0.066	0.205	0.420	0.033	0.416
CAL-DR	Est	0.506	0.501	−0.508	0.475	0.531	4.341	0.937	3.262
	SD	0.047	0.045	0.053	0.047	0.217	0.422	0.045	0.416

biases under cases VI but with smaller standard deviations; the concordance function of the estimated optimal treatment regimes obtained by the CAL and CAL-DR methods are larger than those obtained by the IPSW and AIPSW methods; the value function of the estimated optimal treatment regimes obtained by the CAL and CAL-DR methods are slightly larger than those obtained by the IPSW and AIPSW methods and are closer to the value function of the true optimal treatment regime $d^{\mathrm{opt}}(\mathbf{X})=I(\beta_0'\mathbf{X}>1)$. The LR method shows the worst performance under this case, yielding a much smaller value function and PCD than other methods.

In summary, on the basis of the above simulations the CAL and CAL-DR methods show competitive performance even when the monotonic index model for $D(\mathbf{X})$ is violated.

4. Real data analysis

We demonstrate the proposed method by an application to the data from AIDS Clinical Trials Group Protocol 175, which consists of 2139 subjects infected with the human immunodeficiency virus. The enrolled subjects were randomized to four different treatment groups: zidovudine monotherapy, ZDV, ZDV plus didanosine, ddI, ZDV plus zalcitabine and ddI monotherapy. Here, we focus on the subset of patients receiving the treatment ZDV + ddI or ZDV + zalcitabine. Treatment indicator $A=0$ denotes the treatment ZDV + zalcitabine (524 subjects), whereas $A=1$ denotes the treatment ZDV + ddI (522 subjects). The CD4 cell count (cells per cubic millimetre) at 20 ± 5 weeks post baseline is chosen to be the continuous response variable Y. 12 covariates are considered, including five continuous variables, age (years), weight (kilograms), Karnofsky score (a scale of 0–100), CD4 cell count at baseline and CD8 cell count (cells per cubic millimetre) at baseline, and seven binary variables, haemophilia (0, no; 1, yes), homosexual activity (0, no; 1, yes), history of intravenous drug use (0, no; 1, yes), race (0, white; 1, non-white), gender

Table 3. Estimated optimal treatment regimes for the AIDS Clinical Trials Group 175 data

Method	*Statistic*	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$
CAL	Est	0.929	−0.182	−0.160	−0.252	−0.076	−0.057	−0.059	0.001	−0.033	0.037	−0.019	−0.021
	SE	0.470	0.353	0.957	0.471	0.303	0.083	0.044	0.048	0.029	0.052	0.025	0.042
	PV	0.048	0.608	0.868	0.594	0.802	0.494	0.182	0.982	0.253	0.473	0.454	0.614
CAL-DR	Est	0.946	−0.093	−0.143	−0.202	−0.166	−0.051	−0.052	0.001	−0.030	0.021	−0.017	−0.017
	SE	0.111	0.211	0.207	0.258	0.203	0.053	0.028	0.028	0.02	0.032	0.016	0.021
	PV	0.000	0.658	0.490	0.433	0.413	0.330	0.067	0.960	0.130	0.516	0.279	0.410

(0, female; 1, male), antiretroviral history (0, naive; 1, experienced) and symptomatic status (0, asymptomatic; 1, symptomatic). Here, the propensity score is known and set as $\pi(\mathbf{X}) \equiv 0.5$. We consider both the CAL and the CAL-DR estimators. In our analysis, the five continuous covariates are normalized to have mean 0 and norm 1. The CAL and CAL-DR estimators of $\boldsymbol{\beta}$ and their standard errors are given in Table 3. It can be seen that the CAL and CAL-DR methods yield comparable estimates of $\boldsymbol{\beta}$ whereas the CAL-DR estimator generally has smaller standard errors compared with the CAL estimator. In addition, from the p-values that are reported in Table 3, age is the only significant covariate at the level of 0.05 on the basis of both methods. We refit the CAL and CAL-DR estimators with age as the only covariate. The two methods yield the same estimated optimal treatment regime, which is given by $I(\text{age} > 37.5)$. The results suggest that ZDV + zalcitabine ($A = 0$) is more favourable to young patients with AIDS, but ZDV + ddI ($A = 1$) for old patients. A similar finding was also observed in Lu *et al.* (2013).

Furthermore, to compare the CAL and CAL-DR estimators with the IPSW estimator, we randomly divide the data set into half training and half testing sets for 200 times. For each random split, we compute the CAL, CAL-DR and IPSW estimators on the basis of the training set with age as the only covariate. Then, we compute the estimated value function $\hat{V}_n$ defined in equation (1) for the estimated optimal treatment regimes obtained by the CAL, CAL-DR and IPSW estimators. The averages of $\hat{V}_n$ are 395.1 (12.2) for the CAL and CAL-DR estimators, and 386.5 (18.9) for the IPSW estimators respectively, where the quantities in parentheses are the corresponding standard deviations. These results demonstrate that the optimal treatment regimes that are obtained by the CAL and CAL-DR methods give larger value functions on average but with smaller standard deviations than that obtained by the IPSW method.

5. Discussion

We propose new methods for estimating the optimal treatment regime based on CAL. The estimators of the parameters in the prescriptive index for treatment decision have standard $n^{1/2}$-rate, and their asymptotic distributions and inference are studied. It is worth noting that, although the CAL proposed has appealing interpretation and works well under monotonic index models, the procedure will not always recommend patients with positive $D(\mathbf{X}) = E(Y|A = 1, \mathbf{X}) - E(Y|A = 0, \mathbf{X})$ to receive treatment 1.

In this paper, we consider only a class of linear decision rules for simplicity. However, our method can be extended to incorporate a more general class of decision rules such as $d(\mathbf{X}) = I\{g(\mathbf{X}) \geqslant c\}$, where $g(\cdot)$ is a non-parametric function of $\mathbf{X}$. The corresponding concordance function is defined as

$$C(g) = E[Y_i^*(1) - Y_i^*(0) - \{Y_j^*(1) - Y_j^*(0)\} I\{g(\mathbf{X}_i) > g(\mathbf{X}_j)\}],$$

and the associated optimal treatment regime is given by $d^{*,\text{opt}}(\mathbf{X}) = I\{g^*(\mathbf{X}) \geqslant c^*\}$, where $g^* = \arg\max_g C(g)$ and $c^* = \arg\max_c E(Y^*[I\{g^*(\mathbf{X}) \geqslant c\}])$.

In addition, it is interesting to generalize the proposed methods to multiple-treatment set-ups. For example, consider a study with three treatments, denoted as 0, 1 and 2. Assume monotonic index models $D_1(\mathbf{X}) \equiv E\{Y^*(1) - Y^*(0)|\mathbf{X}\} = Q_1(\beta_1'\mathbf{X})$ and $D_2(\mathbf{X}) \equiv E\{Y^*(2) - Y^*(0)|\mathbf{X}\} = Q_2(\beta_2'\mathbf{X})$, where $Q_1(\cdot)$ and $Q_2(\cdot)$ are two strictly increasing functions. We can use the methods proposed to estimate β_1 and β_2 on the basis of a comparison of two treatments at a time: 1 *versus* 0 and 2 *versus* 0. The resulting estimated optimal treatment rules are denoted as $\hat{d}_{10}^{\text{opt}}(\mathbf{x})$ and $\hat{d}_{20}^{\text{opt}}(\mathbf{x})$. Here, if $\hat{d}_{10}^{\text{opt}}(\mathbf{x}) = 1$, a subject with covariates $\mathbf{X} = \mathbf{x}$ is given treatment 1, and 0 otherwise. The rule $\hat{d}_{20}^{\text{opt}}(\mathbf{x})$ is similarly defined. Therefore, when $\hat{d}_{10}^{\text{opt}}(\mathbf{x}) = \hat{d}_{20}^{\text{opt}}(\mathbf{x}) = 0$, treatment 0 is given; when $\hat{d}_{10}^{\text{opt}}(\mathbf{x}) = 1$ and $\hat{d}_{20}^{\text{opt}}(\mathbf{x}) = 0$, treatment 1 is given; when $\hat{d}_{10}^{\text{opt}}(\mathbf{x}) = 0$ and $\hat{d}_{20}^{\text{opt}}(\mathbf{x}) = 1$, treatment 2 is given; when $\hat{d}_{10}^{\text{opt}}(\mathbf{x}) = \hat{d}_{20}^{\text{opt}}(\mathbf{x}) = 1$, we shall apply CAL based on the comparison between treatments 1 and 2, and obtain the resulting estimated optimal treatment rule $\hat{d}_{21}^{\text{opt}}(\mathbf{x})$, which assigns treatment 2 if $\hat{d}_{21}^{\text{opt}}(\mathbf{x}) = 1$ and treatment 1 otherwise. However, generalization to multiple-category treatment is not straightforward other than using the above *ad hoc* pairwise comparison strategy and its properties need to be investigated. Moreover, the methods proposed can also be extended to incorporate multiple decision time points. The related discussions are given in the on-line supplementary appendix.

Acknowledgements

The authors are grateful to two referees, the Associate Editor and the Joint Editor for their thoughtful and suggestive comments, which have helped to improve greatly on an earlier manuscript. The work was partially supported by National Institutes of Health grant P01 CA142538.

Appendix A

Here, we give the proofs for only theorems 1 and 2. Those for theorems 4 and 5 are given in the on-line supplementary appendix.

A.1. Proof of theorem 1

To establish the consistency of $\hat{\beta}$, similarly to Cavanagh and Sherman (1998), we need to show that

(a) $C(\beta)$ has a unique maximizer at β^*,
(b) $\sup_{\|\beta\|=1} |\hat{C}_n(\beta, \hat{\theta}) - C(\beta)| = o_p(1)$ and
(c) $C(\beta)$ is continuous.

Here, property (a) is assumed by condition 3, which holds for the class of monotonic index models. Under condition 1, we have that $E\{\Lambda_{12}(\theta)|\mathbf{X}_1, \mathbf{X}_2\} = D(\mathbf{X}_1) - D(\mathbf{X}_2)$ for any fixed θ. Define $f_{ij}(\beta, \theta) = \Lambda_{ij}(\theta) I(\beta'\mathbf{X}_i > \beta'\mathbf{X}_j) - C(\beta)$. Then, $\hat{C}_n(\beta, \theta) - C(\beta) = U_n f_{ij}(\beta, \theta)$, where U_n denotes the random measure putting mass $1/\{n(n-1)\}$ on each pair of data. Therefore, $U_n f_{ij}(\beta, \theta)$ is a zero-mean U-process of order 2. In addition, by condition (c) of condition 4, it can be shown that the class $\{f_{12}(\beta, \theta) : \|\beta\| = 1, \theta - \theta^* = O_p(n^{-1/2})\}$ is Euclidean with a square integrable envelope. Thus, property (b) holds. Finally, condition (b) of condition 4 implies that $P(\beta'\mathbf{X}_1 = \beta'\mathbf{X}_2) = 0$ for $\beta \in \mathcal{B}$. Then, $\tau(\beta_m, \mathbf{X}_1, \mathbf{X}_2) \to \tau(\beta, \mathbf{X}_1, \mathbf{X}_2)$ in probability, where $\{\beta_m\}$ is a sequence of elements of $\mathcal{B}$ converging to β as $m \to \infty$. Applying the dominated convergence theorem, we have $C(\beta_m) \to C(\beta)$ as $m \to \infty$. Thus, property (c) is proved. The consistency of $\hat{\beta}$ then follows Amemiya (1985), pages 106–107.

Next, we establish the limiting distribution of $\hat{\beta}$. We have

$$\hat{C}_n(\beta, \hat{\theta}) - \hat{C}_n(\beta^*, \hat{\theta}) = \hat{C}_n(\beta, \hat{\theta}) - \hat{C}_n(\beta, \theta^*) - \{\hat{C}_n(\beta^*, \hat{\theta}) - \hat{C}_n(\beta^*, \theta^*)\} + \hat{C}_n(\beta, \theta^*) - \hat{C}_n(\beta^*, \theta^*),$$

where $\hat{C}_n(\beta, \theta^*)$ is a U-statistic of order 2. Following similar arguments to those in Sherman (1993), by condition 5, we have, uniformly over any $o_p(1)$ neighbourhood of β^*,

$$\hat{C}_n(\boldsymbol{\beta},\boldsymbol{\theta}^*)-\hat{C}_n(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*)=\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)'V(\boldsymbol{\beta}-\boldsymbol{\beta}^*)+\frac{1}{\sqrt{n}}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)'W_n+o_p(\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|^2)+o_p\left(\frac{1}{n}\right), \tag{5}$$

where $W_n=n^{-1/2}\Sigma_{i=1}^n\nabla_1\varrho(\boldsymbol{\beta}^*,\mathbf{X}_i)\rightarrow N(0,\Delta)$ in distribution and $2V=E\{\nabla_2\varrho(\boldsymbol{\beta}^*,\mathbf{X})\}$.

In addition, we have $E\{\partial\Lambda_{ij}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|\mathbf{X}_i,\mathbf{X}_j\}=E\{\partial^2\Lambda_{ij}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'|\mathbf{X}_i,\mathbf{X}_j\}=0$ for any $\boldsymbol{\theta}$. First applying a second-order Taylor series expansion with respect to $\boldsymbol{\theta}$ and then following similar arguments to derive equation (5), we have, uniformly over any $o_p(1)$ neighbourhood of $\boldsymbol{\beta}^*$,

$$\begin{aligned}\hat{C}_n(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})-\hat{C}_n(\boldsymbol{\beta},\boldsymbol{\theta}^*)-\{\hat{C}_n(\boldsymbol{\beta}^*,\hat{\boldsymbol{\theta}})-\hat{C}_n(\boldsymbol{\beta}^*,\boldsymbol{\theta}^*)\}\\ =\{o_p(\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|^2)+o_p(1/n)\}O_p(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|+\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|^2)+o_p(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|^2)\\ =o_p(\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|^2)+o_p(1/n).\end{aligned} \tag{6}$$

Combining equations (5) and (6), we have

$$0\leqslant\hat{C}_n(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})-\hat{C}_n(\boldsymbol{\beta}^*,\hat{\boldsymbol{\theta}})=\frac{1}{2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)'V(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)+\frac{1}{\sqrt{n}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)'W_n+o_p(\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|^2)+o_p\left(\frac{1}{n}\right),$$

which implies that $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*=O_p(1/\sqrt{n})$ since V is negative definite. Therefore,

$$\hat{C}_n(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})-\hat{C}_n(\boldsymbol{\beta}^*,\hat{\boldsymbol{\theta}})=\frac{1}{2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)'V(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)+\frac{1}{\sqrt{n}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)'W_n+o_p\left(\frac{1}{n}\right).$$

By theorem 2 of Sherman (1993), we have

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)=V^{-1}W_n+o_p(1)\rightarrow N(0,\boldsymbol{\Sigma}), \tag{7}$$

in distribution, where $\boldsymbol{\Sigma}=V^{-1}\Delta V^{-1}$ with $\Delta=E\{\nabla_1\varrho(\boldsymbol{\beta}^*,\mathbf{X})\nabla_1\varrho(\boldsymbol{\beta}^*,\mathbf{X})'\}$.

A.2. Proof of theorem 2

Define $\tilde{Y}=Y/[\pi(\mathbf{X})A+\{1-\pi(\mathbf{X})\}(1-A)]$. Note that $V(\boldsymbol{\beta},c)=E[\tilde{Y}I\{A=I(\boldsymbol{\beta}'\mathbf{X}\geqslant c)\}]$. By condition 6, we have $V(\boldsymbol{\beta}^*,c)-V(\boldsymbol{\beta}^*,c^*)\leqslant -M(c-c^*)^2$. In addition, define

$$\mathcal{F}_1=\{\tilde{Y}I\{A=I(\boldsymbol{\beta}'\mathbf{X}\geqslant c)\}:|c-c^*|\leqslant\eta,n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)=O_p(1)\},$$

where η is a positive constant. Under condition 1, it can be easily shown that $\mathcal{F}_1$ is a Donsker class. Moreover, by theorem 11.2 of Kosorok (2007), we have

$$E\left[n^{1/2}\sup_{|c-c^*|<\eta}|\hat{V}_n(\hat{\boldsymbol{\beta}},c)-V(\hat{\boldsymbol{\beta}},c)-\{\hat{V}_n(\hat{\boldsymbol{\beta}},c^*)-V(\hat{\boldsymbol{\beta}},c^*)\}|\right]\leqslant\int_0^\eta\sqrt{}\log[N_{[]}\{\varepsilon F,\mathcal{F}_1,L_2(P)\}]\,\mathrm{d}\varepsilon\leqslant K\eta^{1/2}$$

for all n sufficiently large and sufficiently small η, where $F=|\tilde{Y}|$ is a square integrable envelope function, $N_{[]}\{\varepsilon F,\mathcal{F}_1,L_2(P)\}$ is the bracket number of $\mathcal{F}_1$ with $L_2(P)$ norm metric and K is a positive constant. Therefore,

$$\begin{aligned}&E\left[n^{1/2}\sup_{|c-c^*|<\eta}|\hat{V}_n(\hat{\boldsymbol{\beta}},c)-\hat{V}_n(\hat{\boldsymbol{\beta}},c^*)-\{V(\boldsymbol{\beta}^*,c)-V(\boldsymbol{\beta}^*,c^*)\}|\right]\\ &\qquad=E\left[n^{1/2}\sup_{|c-c^*|<\eta}|\hat{V}_n(\hat{\boldsymbol{\beta}},c)-V(\hat{\boldsymbol{\beta}},c)-\{V_n(\hat{\boldsymbol{\beta}},c^*)-V(\hat{\boldsymbol{\beta}},c^*)\}|\right.\\ &\qquad\quad\left.+n^{1/2}\sup_{|c-c^*|<\eta}|V(\hat{\boldsymbol{\beta}},c)-V(\boldsymbol{\beta}^*,c)-\{V(\hat{\boldsymbol{\beta}},c^*)-V(\boldsymbol{\beta}^*,c^*)\}|\right]\\ &\qquad\leqslant K\eta^{1/2}+\eta\,O(1)=O(1)\phi(\eta),\end{aligned}$$

where $\phi(\eta)=\eta^{1/2}+\eta$. Note that $\phi_n(\eta)/\eta^\alpha$ is decreasing in η for any $\alpha\in[1,2)$. Set $r_n=n^{1/3}$; then $r_n^2\phi(1/r_n)=n^{1/2}+n^{1/3}=O(n^{1/2})$. By theorem 14.4 of Kosorok (2007), we obtain $n^{1/3}(\hat{c}-c^*)=O_p(1)$.

Next, we establish the limiting distribution of $\hat{h}_n=n^{1/3}(\hat{c}-c^*)$. We have that h_n is the argmax of the process $n^{2/3}\{\hat{V}_n(\hat{\boldsymbol{\beta}},n^{-1/3}h+c^*)-\hat{V}_n(\hat{\boldsymbol{\beta}},c^*)\}$ that is indexed by h. In addition,

$$\begin{aligned}n^{2/3}\{\hat{V}_n(\hat{\boldsymbol{\beta}},n^{-1/3}h+c^*)-\hat{V}_n(\hat{\boldsymbol{\beta}},c^*)\}=n^{2/3}\{\hat{V}_n(\hat{\boldsymbol{\beta}},n^{-1/3}h+c^*)-V(\hat{\boldsymbol{\beta}},n^{-1/3}h+c^*)\}-n^{2/3}\{\hat{V}_n(\hat{\boldsymbol{\beta}},c^*)\\ -V(\hat{\boldsymbol{\beta}},c^*)\}+n^{2/3}\{V(\hat{\boldsymbol{\beta}},n^{-1/3}h+c^*)-V(\hat{\boldsymbol{\beta}},c^*)\}.\end{aligned}$$

We have

$$\begin{aligned} n^{2/3}\{V(\hat{\boldsymbol{\beta}}, n^{-1/3}h+c^*) - V(\hat{\boldsymbol{\beta}}, c^*)\} &= n^{2/3}\{V(\boldsymbol{\beta}^*, n^{-1/3}h+c^*) - V(\boldsymbol{\beta}^*, c^*)\} \\ &\quad + n^{2/3}[V(\hat{\boldsymbol{\beta}}, n^{-1/3}h+c^*) - V(\boldsymbol{\beta}^*, n^{-1/3}h+c^*) \\ &\quad - \{V(\hat{\boldsymbol{\beta}}, c^*) - V(\boldsymbol{\beta}^*, c^*)\}] \\ &= \tfrac{1}{2}\upsilon h^2 + o_p(1) + O_p(n^{1/3}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|) = \tfrac{1}{2}\upsilon h^2 + o_p(1), \end{aligned}$$

where $\upsilon = \partial^2 V(\boldsymbol{\beta}^*, c^*)/\partial c^2$. In addition, define

$$\mathcal{F}_{2n} = \{n^{1/6}[\tilde{Y} I\{A = I(\boldsymbol{\beta}'\mathbf{X} \geqslant n^{-1/3}h + c^*)\} - \tilde{Y} I\{A = I(\boldsymbol{\beta}'\mathbf{X}_i \geqslant c^*)\}] : n^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}^*) = O_p(1)\},$$

which can be shown to be a Donsker class. In addition, we have

$$\begin{aligned} &E(n^{1/3}[\tilde{Y} I\{A = I(\boldsymbol{\beta}^{*\prime}\mathbf{X} \geqslant n^{-1/3}h + c^*)\} - \tilde{Y} I\{A = I(\boldsymbol{\beta}^{*\prime}\mathbf{X}_i \geqslant c^*)\}]^2) \\ &\quad = E[n^{1/3}\tilde{Y}^2\{I(c^* \leqslant \boldsymbol{\beta}^{*\prime}\mathbf{X} < n^{-1/3}h + c^*)I(h>0) + I(n^{-1/3}h + c^* \leqslant \boldsymbol{\beta}^{*\prime}\mathbf{X} < c^*)I(h<0)\}] \\ &\quad \to q^*|h|, \qquad \text{almost surely,} \end{aligned}$$

as $n \to \infty$, where q^* is a positive constant. Therefore, $n^{2/3}\{\hat{V}_n(\hat{\boldsymbol{\beta}}, n^{-1/3}h + c^*) - V(\hat{\boldsymbol{\beta}}, n^{-1/3}h + c^*)\} - n^{2/3}\{\hat{V}_n(\hat{\boldsymbol{\beta}}, c^*) - V(\hat{\boldsymbol{\beta}}, c^*)\}$ converges weakly to a two-sided Gaussian process $\sqrt{q^*}\mathcal{Z}(h)$, where $\mathcal{Z}(h)$ is a standard two-sided Brownian motion process. By the argmax theorem of Kosorok (2007), $\hat{h}_n$ converges in distribution to $\arg\max_h G(h)$, where $G(h) = \sqrt{q^*}\mathcal{Z}(h) + \upsilon h^2/2$.

References

Abrevaya, J. (2003) Pairwise-difference rank estimation of the transformation model. *J. Bus. Econ. Statist.*, **21**, 437–447.

Amemiya, T. (1985) *Advanced Econometrics*. Boston: Harvard University Press.

Blatt, D., Murphy, S. and Zhu, J. (2004) A-learning for approximate planning. *Technical Report 04-63*. Methodology Center, Pennsylvania State University, State College.

Brown, B. and Wang, Y.-G. (2005) Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, **92**, 149–158.

Brown, B. and Wang, Y.-G. (2007) Induced smoothing for rank regression with censored survival times. *Statist. Med.*, **26**, 828–836.

Cavanagh, C. and Sherman, R. P. (1998) Rank estimators for monotonic index models. *J. Econmetr.*, **84**, 351–381.

Chen, S. (2002) Rank estimation of transformation models. *Econometrica*, **70**, 1683–1697.

Foster, J. C., Taylor, J. M. and Ruberg, S. J. (2011) Subgroup identification from randomized clinical trial data. *Statist. Med.*, **30**, 2867–2880.

Han, A. K. (1987) Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *J. Econmetr.*, **35**, 303–316.

Jin, Z., Ying, Z. and Wei, L. J. (2001) A simple resampling method by perturbing the minimand. *Biometrika*, **88**, 381–390.

Kosorok, M. R. (2007) *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

Lu, W., Zhang, H. and Zeng, D. (2013) Variable selection for optimal treatment decision. *Statist. Meth. Med. Res.*, **22**, 493–504.

Matsouaka, R. A., Li, J. and Cai, T. (2014) Evaluating marker-guided treatment selection strategies. *Biometrics*, **70**, 489–499.

Murphy, S. A. (2003) Optimal dynamic treatment regimes. *J. R. Statist. Soc.* B, **65**, 331–355.

Pang, L., Lu, W. and Wang, H. J. (2012) Variance estimation in censored quantile regression via induced smoothing. *Computnl Statist. Data Anal.*, **56**, 785–796.

Qian, M. and Murphy, S. A. (2011) Performance guarantees for individualized treatment rules. *Ann. Statist.*, **39**, 1180–1210.

Sherman, R. P. (1993) The limiting distribution of the maximum rank correlation estimator. *Econometrica*, **61**, 123–137.

Watkins, C. J. C. H. (1989) Learning from delayed rewards. *PhD Thesis*. University of Cambridge, Cambridge.

Watkins, C. J. and Dayan, P. (1992) Q-learning. *Mach. Learn.*, **8**, 279–292.

Zhang, J., Jin, Z., Shao, Y. and Ying, Z. (2013) Statistical inference on transformation models: a self-induced smoothing approach. *Preprint arXiv:1302.6651*. Department of Statistics, Columbia University, New York.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M. and Laber, E. (2012a) Estimating optimal treatment regimes from a classification perspective. *Stat*, **1**, 103–114.

Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2012b) A robust method for estimating optimal treatment regimes. *Biometrics*, **68**, 1010–1018.

Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2013) Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, **100**, 681–694.

Zhao, Y., Kosorok, M. R. and Zeng, D. (2009) Reinforcement learning design for cancer clinical trials. *Statist. Med.*, **28**, 3294–3315.

Zhao, L., Tian, L., Cai, T., Claggett, B. and Wei, L. J. (2013) Effectively selecting a target population for a future comparative study. *J. Am. Statist. Ass.*, **108**, 527–539.

Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012) Estimating individualized treatment rules using outcome weighted learning. *J. Am. Statist. Ass.*, **107**, 1106–1118.

Zhao, Y., Zeng, D., Socinski, M. A. and Kosorok, M. R. (2011) Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics*, **67**, 1422–1433.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary appendix for: “Concordance-assisted learning for estimating optimal individualized treatment regimes”’.