Electronic Journal of Statistics
Vol. 17 (2023) 1344–1370
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2127

Flexible inference of optimal individualized treatment strategy in covariate adjusted randomization with multiple covariates*

Trinetri Ghosh†

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison
e-mail: tghosh3@wisc.edu

Yanyuan Ma

Department of Statistics, Pennsylvania State University
e-mail: yzm63@psu.edu

Rui Song

Department of Statistics, North Carolina State University
e-mail: rsong@ncsu.edu

Pingshou Zhong

Department of Mathematics, Statistics and Computer Science, University of Illinois Chicago
e-mail: pszhong@uic.edu

Abstract: To maximize clinical benefit, clinicians routinely tailor treatment to the individual characteristics of each patient, where individualized treatment rules are needed and are of significant research interest to statisticians. In the covariate-adjusted randomization clinical trial with many covariates, we model the treatment effect with an unspecified function of a single index of the covariates and leave the baseline response completely arbitrary. We devise a class of estimators to consistently estimate the treatment effect function and its associated index while bypassing the estimation of the baseline response, which is subject to the curse of dimensionality. We further develop inference tools to identify predictive covariates and isolate effective treatment region. The usefulness of the methods is demonstrated in both simulations and a clinical data example.

Keywords and phrases: Covariate adjusted randomization, estimating equations, nonparametric regression, robustness, semiparametric methods, single index model, treatment effect.

Received January 2022.

*This work was supported by the *National Science Foundation* and the *National Institute of Health*.

†Corresponding Author.

1. Introduction

Precision medicine, which is defined as treatments targeted to individual patients' needs based on genetics, biomarker, phenotypic, or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations [12], has generated tremendous interest in statistical research. Precision medicine based on individual's health-related metrics and environmental factors is used to discover individualized treatment regimes (ITRs); methodology for such discovery is an expanding field of statistics.

Various methods have been proposed in the statistical literature to estimate the optimal ITRs. Q-learning [39, 40, 21, 46, 4, 48, 25, 9, 32, 5] and A-learning [20, 27, 1, 23, 10, 17, 29] are two backward induction methods for deriving optimal dynamic treatment regimes. Other related approaches include parametric methods [33, 34, 35], model-free or direct value search method by maximizing a nonparametric estimator of value function [43, 42, 7, 49, 50, 30], semiparametric methods [18, 38, 37, 19, 31, 14, 41, 16], studies in inference [2, 3, 15], studies in high dimensional settings [36, 13], and machine learning methods [47, 24, 45, 44, 22]. [47] proposed outcome weighted learning, which can be viewed as a weighted classification problem using the covariate information weighted by the individual response to maximize the overall outcome, and [16], a generalization of [31], where single index in the treatment effect can be replaced by multiple indices and the function form is less restrictive. [5] proposed a robust Q-learning approach for a two-stage dynamic binary treatment regime by adapting the Robinson's transformation [28]. Their estimators are robust to the misspecification of conditional mean outcome models given that the probability of treatment assignment models are consistently estimated at a sufficiently fast rate. [22] proposed a quasi-oracle estimator of heterogeous treatment effect by using penalized kernel regression to minimize the loss function based on Robinson's transformation [28] and named it R-learning, where both the propensity score and the conditional mean outcome are estimated via machine learning methods. In contrast, we consider the case where the probability of treatment assignment is known. Instead of considering the conditional mean outcome models, we directly estimate the treatment effect function. Further, we formulated the treatment effect function by single index model. Finally, our proposed method is easily extended to continuous treatment assignment, which is discussed in Section 4.

The rules based on parametric models of clinical outcomes given treatment and other covariates are simple, yet can be incorrect when parametric models are misspecified in the wide class of rules that can include forms out of the parametric family. The rules based on machine learning techniques to determine the relationship between clinical outcomes and treatment plus covariates, such as in Zhao et al. (2009, 2011), are nonparametric and flexible but are often complex and may have large variability. Existing semiparametric methodologies, such as [20] and [31], flexibly incorporate the relationship between the covariates and the response variables, but are not efficient due to the challenges in estimating the decision rules. Moreover, to the best of our knowledge, these existing semiparametric models do not allow covariate adjusted randomization, where a patient is randomized to one of the treatment arms based on the patient's covariate and a predetermined randomization scheme. Therefore we are motivated to derive semiparametric efficient estimators for flexible models that allow covariate adjusted randomization.

Let $\mathbf{X}_i$ be the p-dimensional covariate vector, Z_i be the treatment indicator, and Y_i be the response, and assume $(\mathbf{X}_i, Z_i, Y_i)$'s are independent and identically distributed (iid) for $i = 1, \ldots, n$. Here, the treatment indicator Z_i is often categorical (i.e. when several treatment arms are considered) but can also be continuous (i.e. when a continuous of dosage range is considered). In the simplest case when Z_i is binary, we denote $Z_i = 1$ if the ith patient is randomized to treatment and $Z_i = 0$ to placebo. When the randomization is covariate adjusted, the probability distribution of Z_i depends on $\mathbf{X}_i$. Our goal is to analyze the data $(\mathbf{X}_i, Z_i, Y_i)$ for $i = 1, \ldots, n$, so that we can develop a treatment rule that is best for each new individual. Without loss of generality, we assume larger value of Y_0 indicates better outcome. Equivalently, we want to identify a deterministic decision rule, $d(\mathbf{x})$, which takes as input a given value $\mathbf{x}$ of $\mathbf{X}_0$ and outputs a treatment choice of Z_0 that will maximize the potential treatment result Y_0. Here $\mathbf{X}_0$ denotes the covariate of the new individual, Z_0 the treatment decision and Y_0 the potential treatment result given Z_0. Let P^d be the distribution of $(\mathbf{X}_0, Z_0, Y_0)$ and E^d be the expectation with respect to this distribution. The value function is defined as $V(d) = E^d(Y_0)$. An optimal individualized treatment rule d_0 maximizes $V(d)$ over all decision rules. That is, $d_0 = \text{argmax}_d V(d) = \text{argmax}_d E_{\mathbf{X}_0}[E\{Y_0|Z_0 = d(\mathbf{X}_0), \mathbf{X}_0\}]$. Furthermore, we refer to $\max_d V(d)$ as the optimal value function. It is important to recognize the difference between the clinical trial data $(\mathbf{X}_i, Z_i, Y_i), i = 1, \ldots, n$ and the potential data associated with the new individual $(\mathbf{X}_0, Z_0, Y_0)$.

In the covariate adjusted randomization, let the probability of assigning to the treatment or placebo arm be a known function $f_{Z|\mathbf{X}}(\mathbf{x}, z)$. Here and throughout, we omit the subindex $_i$ when describing the data from a randomized clinical trial as long as it does not cause confusion. We model the treatment effect using a single index model of the covariates, while leaving the baseline response, $f(\mathbf{X})$, unspecified. This yields the model

$$Y = f(\mathbf{X}) + Zg(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + \epsilon. \tag{1}$$

We assume the regression error ϵ is independent of the covariates $\mathbf{X}$ and the treatment indicator Z, and $\epsilon \sim N(0, \sigma^2)$. We do not assume Z and $\mathbf{X}$ to be independent, hence allow the covariate adjusted randomization procedure. Naturally from (1), $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ is the treatment effect as a function of the covariate value $\mathbf{x}$. We assume g to be a smooth function. We can easily see that different treatment described by Z causes different result, with the difference is reflected in $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. Hence, under model (1), to compare treatment outcome difference and find the optimal individualized treatment rule, we only need to estimate $g(\cdot)$ and $\boldsymbol{\beta}$. In other words, our goal is to estimate $\boldsymbol{\beta}$ and $g(\cdot)$, hence to estimate the treatment effect given any covariate $\mathbf{x}$. To retain the identifiability of

$g(\cdot)$ and $\boldsymbol{\beta}$, we fix the first component of $\boldsymbol{\beta}$ to be 1. Then we denote the lower $(p-1)$-dimensional vector by $\boldsymbol{\beta}_L$, after dividing it by β_1.

The contribution of our proposed method beyond existing literature can be summarized in the following five folds. First, we propose an estimator and show that it is locally semiparametric efficient. Second, our method allows randomization to depend on patients' covariates, which significantly generalizes the practicality of the semiparametric methodology in real studies. Third, although illustrated for binary treatment selection, the proposed method does not restrict to the binary treatment case and is directly applicable even when Z is categorical or continuous. Thus, the method applies to slope estimation while the intercept, i.e. $f(\mathbf{X})$ in (1), is unspecified and inestimable due to the curse of dimensionality caused by the multi-dimension of $\mathbf{X}$. Fourth, different from [31] where B-spline expansion were used, we employ kernel smoothing techniques into the estimation and inference for optimal treatment regimes. Furthermore, in our proposed method we do not need to assume monotonicity of the treatment effect model $g(\cdot)$ as in [31]. Fifth, our method is more flexible in terms of model assumptions compared to [50].

The rest of the article is organized as the following. In Section 2, we devise a class of estimators for $\boldsymbol{\beta}$ and $g(\cdot)$ while bypassing the estimation of $f(\cdot)$. We study the large sample properties of the estimators and carry out inference to detect effective treatment region in Section 3. We also derived a class of estimators for $\boldsymbol{\beta}$ and $g(\cdot)$ for continuous and categorical treatment indicators along with their asymptotic properties in Sections 4 and 5, respectively. Simulation experiments are carried out in Section 6, and the method is implemented on a clinical trial data in Section 7. We conclude the article with a discussion in Section 8 and collect all the technical derivations and proofs in the Supplementary Material.

2. Estimation of β and $g(\cdot)$

2.1. Basic estimation

The estimation of $\boldsymbol{\beta}$ and $g(\cdot)$ is complicated by the presence of the intercept term $f(\mathbf{X})$. When $\mathbf{X}$ is of high or even moderate dimension, $f(\mathbf{X})$ is challenging to estimate due to the well known curse of dimensionality. Thus, a simple treatment is to eliminate the effect of $f(\mathbf{X})$. Following this approach, multiplying Z and $E(Z \mid \mathbf{X})$ on (1), we have

$$\begin{aligned} ZY &= Z^2 g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + Zf(\mathbf{X}) + Z\epsilon, \\ \text{and } E(Z \mid \mathbf{X})Y &= ZE(Z \mid \mathbf{X})g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + E(Z \mid \mathbf{X})f(\mathbf{X}) + E(Z \mid \mathbf{X})\epsilon. \end{aligned}$$

Taking difference of the above two equations yields $\{Z - E(Z \mid \mathbf{X})\}Y = \{Z^2 - ZE(Z \mid \mathbf{X})\}g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + \{Z - E(Z \mid \mathbf{X})\}\{f(\mathbf{X}) + \epsilon\}$. Note that $E[\{Z - E(Z \mid \mathbf{X})\}\{f(\mathbf{X}) + \epsilon\} \mid \mathbf{X}] = E\{Z - E(Z \mid \mathbf{X}) \mid \mathbf{X}\}f(\mathbf{X}) + 0 = 0$ and $E[\{Z^2 - ZE(Z \mid \mathbf{X})\}g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) \mid \mathbf{X}] = \mathrm{var}(Z \mid \mathbf{X})g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$. So we obtain $E[\{Z - E(Z \mid \mathbf{X})\}Y \mid \mathbf{X}] = \mathrm{var}(Z \mid \mathbf{X})g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$. This indicates that if we write $\widetilde{Y} \equiv \{Z - E(Z \mid$

$\mathbf{X})\}Y\{\text{var}(Z \mid \mathbf{X})\}^{-1}$, then $E(\widetilde{Y} \mid \mathbf{X}) = g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$. Viewing $\mathbf{X}$ and $\widetilde{Y}$ as the covariate and response respectively, this is a classical single index model, and we can estimate $\boldsymbol{\beta}$ and g using standard methods [11]. Considering that $\text{var}(Z \mid \mathbf{X})$ might be close to zero, one may want to avoid taking its inverse. In this case, a more stable estimator would be to minimize

$$\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}[\{z_i - E(Z_i \mid \mathbf{x}_i)\}y_i - \text{var}(Z_i \mid \mathbf{x}_i)\{a_j + b_j\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\}]^2$$

with respect to $a_1, \cdots, a_n, b_1, \cdots, b_n$ and parameters in $\boldsymbol{\beta}$, where

$$w_{ij} = \frac{K_h\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\}}{\sum_{i=1}^{n} K_h\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\}},$$

$K_h(\cdot) = h^{-1}K(\cdot/h)$, h is a bandwidth and $K(\cdot)$ is a kernel function. At a fixed $\boldsymbol{\beta}$ value, the minimization with respect to (a_j, b_j) does not rely on other (a_k, b_k) values for $k \neq j$, hence can be done separately, while the optimization with respect to $\boldsymbol{\beta}$ involves all the terms. The resulting $\widehat{\boldsymbol{\beta}}$ estimates $\boldsymbol{\beta}$ and the resulting $\widehat{a}_j$ estimates $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j)$. We show the consistency of $\widehat{\boldsymbol{\beta}}$ in Section A.4 of the supplement. We can further estimate $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0)$ by $\widehat{a}_0$ which is obtained by

$$(\widehat{a}_0, \widehat{b}_0) = \arg\min_{a_0, b_0} \sum_{i=1}^{n} w_{i0}[\{z_i - E(Z_i \mid \mathbf{x}_i)\}y_i - \text{var}(Z_i \mid \mathbf{x}_i)\{a_0 + b_0\widehat{\boldsymbol{\beta}}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_0)\}]^2.$$

2.2. Proposed estimator

Although the above procedure provides one way of estimating $\boldsymbol{\beta}$ and $g(\cdot)$, it is somewhat ad hoc, and it is unclear if other estimators exist to achieve similar or better estimation. To investigate the problem more thoroughly and systematically, we start with the likelihood function of a single observation $(\mathbf{x}, z, y)$,

$$f_{\mathbf{X},Z,Y}(\mathbf{x}, z, y, \boldsymbol{\beta}, \sigma, g, f, \eta) = \eta(\mathbf{x}) f_{Z|\mathbf{X}}(\mathbf{x}, z)\sigma^{-1}\phi[\sigma^{-1}\{y - f(\mathbf{x}) - zg(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}]. \quad (2)$$

Here $\eta(\mathbf{x})$ is the marginal probability density or mass function (pdf or pmf) of $\mathbf{X}$, and $\phi(\cdot)$ is the standard normal pdf. We first focus our attention on estimating $\boldsymbol{\beta}$ alone, thus we view g together with f, η and σ as nuisance parameters. In this case, (2) is a semiparametric model, thus we derive the estimators for $\boldsymbol{\beta}$ through deriving its influence functions and constructing estimating equations. It is easy to obtain the score function with respect to $\boldsymbol{\beta}$ through taking partial derivative of the loglikelihood function with respect to the parameter. In Supplementary Material A.1, we further project the score functions $\mathbf{S}_{\beta}(\mathbf{x}, z, y, \boldsymbol{\beta}, g, f, \sigma)$ onto the nuisance tangent space, a space spanned by the nuisance score functions, and obtain the efficient score function $\mathbf{S}_{\text{eff}}(\mathbf{x}, z, y, \boldsymbol{\beta}, g, f, \sigma) = \sigma^{-2}\epsilon\{z - E(Z \mid \mathbf{x})\}g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\{\mathbf{x}_L - \mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}$, where $g'(\cdot)$ is the first derivative of $g(\cdot)$, $\mathbf{x}_L$ is the lower $(p-1)$-dimensional sub-vector of $\mathbf{x}$, $\epsilon \equiv y - f(\mathbf{x}) - zg(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$

and $\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) \equiv E\left\{\mathrm{var}(Z \mid \mathbf{X})\mathbf{X}_L \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}/E\left\{\mathrm{var}(Z \mid \mathbf{X}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}$. Further denote $\mathbf{S}^*_{\mathrm{eff}}(\mathbf{x}, z, y, \boldsymbol{\beta}, g, f, \sigma) = \sigma^{-2}\{y - f^*(\mathbf{x}) - zg(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\{z - E(Z \mid \mathbf{x})\}h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ $\{\mathbf{x}_L - \mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}$, where the working models for $f(\mathbf{x})$ and $g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ are $f^*(\mathbf{x})$ and $h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, respectively. Even under these working models $f^*(\mathbf{x})$ and $h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, we can verify that $E\{\mathbf{S}^*_{\mathrm{eff}}(\mathbf{x}, z, y, \boldsymbol{\beta}, g, f, \sigma)\} = 0$, which implies that we still get consistent estimator for $\boldsymbol{\beta}$. Inspired by the form of the efficient score function, we propose a general class of consistent estimating equations for $\boldsymbol{\beta}$ as

$$\sum_{i=1}^{n}\{y_i - f^*(\mathbf{x}_i) - z_i g^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}\{z_i - E(Z_i \mid \mathbf{x}_i)\}h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\{\mathbf{x}_{Li} - \mathbf{u}^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\} = \mathbf{0},$$

where f^* is an arbitrary function of $\mathbf{x}$ and h^* is an arbitrary function of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$. Regarding g^* and $\mathbf{u}^*$, we have the freedom of estimating one of the two functions and replacing the other with an arbitrary function of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$, or estimating both.

To explore the various flexibilities suggested above, let $f^*(\mathbf{x})$ be an arbitrary predecided function. For example, $f^*(\mathbf{x}) = 0$. We first examine the choice of approximating both $\mathbf{u}(\cdot)$ and $g(\cdot)$. As a by-product of local linear estimation, we also approximate $g'(\cdot)$. Let $\widehat{\mathbf{u}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ be a nonparametric estimation of $\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. Note that $\widehat{\mathbf{u}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ involves only univariate nonparametric regression. Specifically, using kernel method,

$$\widehat{\mathbf{u}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \frac{\sum_{i=1}^n K_h\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x})\}\mathrm{var}(Z \mid \mathbf{x}_i)\mathbf{x}_{Li}}{\sum_{i=1}^n K_h\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x})\}\mathrm{var}(Z \mid \mathbf{x}_i)}. \tag{3}$$

Let $\boldsymbol{\alpha} = (\alpha_c, \alpha_1)^{\mathrm{T}}$ and for $j = 1, \ldots, n$, let $\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, j)$ solve the estimating equation w.r.t. α_c and α_1 for given $\boldsymbol{\beta}$ and $\mathbf{x}_j$

$$\sum_{i=1}^{n} w_{ij}[y_i - f^*(\mathbf{x}_i) - z_i\{\alpha_c + \alpha_1\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\}]\{z_i - E(Z \mid \mathbf{x}_i)\}\left\{1, \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\right\}^{\mathrm{T}} = \mathbf{0}. \tag{4}$$

Specifically, let $v_{0j} = \sum_{i=1}^n w_{ij}z_i\{z_i - E(Z \mid \mathbf{x}_i)\}$, $v_{1j} = \sum_{i=1}^n w_{ij}z_i\{z_i - E(Z \mid \mathbf{x}_i)\}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)$, $v_{2j} = \sum_{i=1}^n w_{ij}z_i\{z_i - E(Z \mid \mathbf{x}_i)\}\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\}^2$, then

$$\widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, j) = \begin{bmatrix} \left(v_{0j}v_{2j} - v_{1j}^2\right)^{-1}\sum_{i=1}^n w_{ij}\{y_i - f^*(\mathbf{x}_i)\}\{z_i - E(Z \mid \mathbf{x}_i)\}\{v_{2j} - v_{1j}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\} \\ \left(v_{0j}v_{2j} - v_{1j}^2\right)^{-1}\sum_{i=1}^n w_{ij}\{y_i - f^*(\mathbf{x}_i)\}\{z_i - E(Z \mid \mathbf{x}_i)\}\{v_{0j}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j) - v_{1j}\} \end{bmatrix}.$$

In (4), we can replace w_{ij} with $K_h\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\}$ and the resulting estimating equation is identical. Note that the above procedure enables us to obtain $\widehat{\alpha}_c(\boldsymbol{\beta}, j)$ as an approximation of $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j)$ and $\widehat{\alpha}_1(\boldsymbol{\beta}, j)$ as an approximation of $g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j)$. We then plug in the estimated $\widehat{\mathbf{u}}, \widehat{g}$ and $\widehat{g}'$ and solve

$$\sum_{i=1}^{n}\{y_i - f^*(\mathbf{x}_i) - z_i\widehat{\alpha}_c(\boldsymbol{\beta}, i)\}\{z_i - E(Z \mid \mathbf{x}_i)\}\widehat{\alpha}_1(\boldsymbol{\beta}, i)\{\mathbf{x}_{Li} - \widehat{\mathbf{u}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\} = \mathbf{0} \tag{5}$$

to obtain $\widehat{\boldsymbol\beta}$. It is easily seen that the procedure we described above is a type of profiling estimator, where we estimate $g(\cdot), g'(\cdot)$, $\mathbf{u}(\cdot)$ as functions of a given $\boldsymbol\beta$, and then estimate $\boldsymbol\beta$. The idea behind the construction of (4) is similar to the consideration of the efficient score function, and we describe the detailed derivation in Supplementary Material A.2. The estimator based on solving (5) is guaranteed to be consistent, and has the potential to be efficient. In fact, it will be efficient if $f^*(\cdot)$ happens to be the true $f(\cdot)$ function. Of course, this is not very likely to happen in practice. However, this is still the estimator we propose, because $f(\cdot)$ is formidable to estimate for large or even moderate p. Compared to all other estimators that rely on the same $f^*(\cdot)$, the estimator proposed here yields the smallest estimation variability and is the most stable, as we will demonstrate in Section 3. The explicit algorithm of estimating $\boldsymbol\beta$ and g is described below.

1. Select a candidate function $f^*(\mathbf{x})$.
2. Solve the estimating equation (5) to obtain $\widehat{\boldsymbol\beta}$, where $\widehat\alpha_c(\boldsymbol\beta, i)$ and $\widehat\alpha_1(\boldsymbol\beta, i)$ are solutions given right below (4) and $\widehat{\mathbf{u}}(\boldsymbol\beta^{\mathrm{T}}\mathbf{x}_i)$ is given in (3).
3. Solve (4) once more at $\mathbf{x}_j = \mathbf{x}_0$ while fix $\boldsymbol\beta$ at $\widehat{\boldsymbol\beta}$ to obtain $\widehat g(\boldsymbol\beta^{\mathrm{T}}\mathbf{x}_0) = \widehat\alpha_c(\widehat{\boldsymbol\beta}, 0)$.

Remark 1. *We left out the details on how to select the bandwidths in the algorithm. In Section 3, we will show that a large range of bandwidth will yield the same asymptotic results and no under-smoothing is required. In other words, when the sample size is sufficiently large, the estimator of $\boldsymbol\beta$ is very insensitive to the bandwidth. However, for small or moderate sample size and when the estimation of the functional form g is of interest, then different bandwidths may yield different results. In this case, a careful bandwidth selection procedure such as leave-one-out cross-validation needs to be implemented.*

2.3. Alternative simpler estimators

In Sections 2.1 and 2.2, we have derived an ad hoc estimator and the locally semiparametric efficient estimator. They can be used to estimate both $\boldsymbol\beta$ and $g(\cdot)$. If we are interested only in consistent estimator of $\boldsymbol\beta$, then we can use the working model for either $\mathbf{u}(\cdot)$ or $g(\cdot)$, which simplifies (5). We can see that all these methods yield consistent estimators, while the method in Section 2.2 also yields local efficiency. Here, to estimate $\boldsymbol\beta$, instead of estimating both $\mathbf{u}(\cdot)$ and $g(\cdot)$, we can estimate only $\mathbf{u}(\cdot)$, as we now investigate. The basic idea is to replace g and g' using arbitrary functions, say g^* and h^* respectively in the efficient score function construction, and solve

$$\sum_{i=1}^{n}\{y_i - f^*(\mathbf{x}_i) - z_i g^*(\boldsymbol\beta^{\mathrm{T}}\mathbf{x}_i)\}\{z_i - E(Z \mid \mathbf{x}_i)\}h^*(\boldsymbol\beta^{\mathrm{T}}\mathbf{x}_i)\{\mathbf{x}_{Li} - \widehat{\mathbf{u}}(\boldsymbol\beta^{\mathrm{T}}\mathbf{x}_i)\} = \mathbf{0} \quad (6)$$

to obtain an estimator for $\boldsymbol\beta$. The simplest choice will be to set $f^* = g^* = 0, h^* = 1$. Denote the estimator $\widetilde{\boldsymbol\beta}$. To further estimate the function $g(\cdot)$, we can use a simple local constant estimator via solving

$$\sum_{i=1}^{n} K_h\{\widetilde{\boldsymbol\beta}^{\rm T}(\mathbf{x}_i-\mathbf{x}_0)\}\{y_i-f^*(\mathbf{x}_i)-z_i\alpha_c\}\{z_i-E(Z\mid\mathbf{x}_i)\}=0.$$

The resulting solution $\widehat{\alpha}_c(\widetilde{\boldsymbol\beta},0)$ is then our estimate of g at $\boldsymbol\beta^{\rm T}\mathbf{x}_0$, i.e. $\widehat{g}(\boldsymbol\beta^{\rm T}\mathbf{x}_0)=\widehat{\alpha}_c(\widetilde{\boldsymbol\beta},0)$.

Likewise, we can also opt to estimate $g(\cdot)$ instead of $\mathbf{u}(\cdot)$. We can choose either to estimate $g'(\cdot)$ or to make a subjective choice for it. When we estimate $g'(\cdot)$ along with $g(\cdot)$, the procedure is the following. Solve (4) to obtain $\widehat{\boldsymbol\alpha}(\boldsymbol\beta,j)$ for $j=1,\dots,n$. Then solve

$$\sum_{i=1}^{n}\{y_i-f^*(\mathbf{x}_i)-z_i\widehat{\alpha}_c(\boldsymbol\beta,i)\}\{z_i-E(Z\mid\mathbf{x}_i)\}\widehat{\alpha}_1(\boldsymbol\beta,i)\{\mathbf{x}_{Li}-\mathbf{u}^*(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}=\mathbf{0} \quad (7)$$

to obtain $\widehat{\boldsymbol\beta}$. Here $\mathbf{u}^*$ is an arbitrarily chosen function, for example, the simplest is to set $\mathbf{u}^*=\mathbf{0}$.

Because the procedures involved in these two simpler estimators are similar and are both simpler compared with the estimator described in Section 2.2, we omit the details on the computational algorithms.

3. Large sample property and inference

Theorem 3.1. *Assume the regularity conditions listed in Supplementary Material A.3 hold. When $n\to\infty$, the estimator obtained by solving (5) satisfies $\widehat{\boldsymbol\beta}_L-\boldsymbol\beta_L=O_p(h^2+n^{-1/2}h^{-1/2})$. Furthermore, $\sqrt{n}(\widehat{\boldsymbol\beta}_L-\boldsymbol\beta_L)\to N(\mathbf{0},\mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{A}_1^{-1^{\rm T}})$ in distribution when $n\to\infty$, where*

$$\begin{aligned}\mathbf{A}_1 &= E\left[var(Z\mid\mathbf{X})g'^2(\boldsymbol\beta^{\rm T}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol\beta^{\rm T}\mathbf{X})\}^{\otimes 2}\right],\\ \mathbf{B}_1 &= \sigma^2E[var(Z\mid\mathbf{X})g'^2(\boldsymbol\beta^{\rm T}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol\beta^{\rm T}\mathbf{X})\}^{\otimes 2}]\\ &\quad +E[\{f(\mathbf{X})-f^*(\mathbf{X})\}^2var(Z\mid\mathbf{X})g'^2(\boldsymbol\beta^{\rm T}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol\beta^{\rm T}\mathbf{X})\}^{\otimes 2}].\end{aligned}$$

Here and throughout the text, $\mathbf{a}^{\otimes 2}\equiv\mathbf{a}\mathbf{a}^{\rm T}$ for a generic vector or matrix $\mathbf{a}$. When $f^*(\mathbf{X})=f(\mathbf{X})$, the estimator obtains the optimal efficiency bound.

Remark 2. *From Theorem 3.1, we can see that the efficiency loss by using misspecified $f^*(\mathbf{X})$ is $\mathbf{A}_1^{-1}E[\{f(\mathbf{X})-f^*(\mathbf{X})\}^2var(Z\mid\mathbf{X})g'^2(\boldsymbol\beta^{\rm T}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol\beta^{\rm T}\mathbf{X})\}^{\otimes 2}]\mathbf{A}_1^{-1^{\rm T}}$.*

Remark 3. *In estimation, we successfully avoided estimating the baseline response $f(\mathbf{X})$ to avoid the curse of dimensionality. However, the matrix $\mathbf{B}_1$ contains $f(\mathbf{X})$ which brings difficulties in using it for inference. Thus, we propose to use bootstrap method to estimate the asymptotic variance in practice.*

Theorem 3.2. *Assume the regularity conditions listed in Supplementary Material A.3 hold. Let $g_2(\cdot)$ be the second order derivative of $g(\cdot)$. Then the local linear estimator for $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ in (4) satisfies*

$$\begin{aligned}
bias\{\widehat{g}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})\} &= h^2\frac{g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{2}\int u^2K(u)du + o(h^2),\\
var\{\widehat{g}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})\} &= \frac{\sigma^2}{nhE\{var(Z_i\mid\mathbf{X}_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\}f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\int K^2(u)du\\
&\quad+\frac{E\left[\{f(\mathbf{X}_i)-f^*(\mathbf{X}_i)\}^2var(Z_i\mid\mathbf{X}_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right]}{nh\left[E\{var(Z_i\mid\mathbf{X}_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\}\right]^2f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\int K^2(u)du+O(n^{-1}).
\end{aligned}$$

Remark 4. *From Theorem 3.2, we can see that the efficiency loss by using the working model $f^*(\mathbf{X})$ is $\left(nh\left[E\{var(Z_i\mid\mathbf{X}_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\}\right]^2f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right)^{-1}E\left[\{f(\mathbf{X}_i)-f^*(\mathbf{X}_i)\}^2var(Z_i\mid\mathbf{X}_i)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right]\int K^2(u)du$.*

We also have similar large sample properties for the two alternative estimators given in Section 2.3, stated in the following Theorems 3.3 and 3.4. The proofs of Theorems 3.1, 3.3 and 3.4 are given in the Supplementary Material.

Theorem 3.3. *Assume the regularity conditions listed in Supplementary Material A.3 hold. When $n\to\infty$, the estimator obtained from solving (6) satisfies $\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L=O_p(h^2+n^{-1/2}h^{-1/2})$. Furthermore, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L)\to N(\mathbf{0},\mathbf{A}_2^{-1}\mathbf{B}_2\mathbf{A}_2^{-1^{\mathrm{T}}})$ in distribution when $n\to\infty$, where*

$$\begin{aligned}
\mathbf{A}_2 &= E\left[var(Z\mid\mathbf{X})h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})g'^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}\right]\\
&\quad-E\left[var(Z\mid\mathbf{X})\{g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})-g^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}\frac{\partial h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}}{\partial\boldsymbol{\beta}_L^{\mathrm{T}}}\right],
\end{aligned}$$

$$\begin{aligned}
\mathbf{B}_2 &= \sigma^2E\left[var(Z\mid\mathbf{X})h^{*2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\left\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}^{\otimes 2}\right]\\
&\quad+E\left[\{f(\mathbf{X})-f^*(\mathbf{X})\}^2var(Z\mid\mathbf{X})h^{*2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\left\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}^{\otimes 2}\right]\\
&\quad+E\Big[Z^2\{Z-E(Z\mid\mathbf{X})\}^2\{g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\\
&\qquad-g^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^2h^{*2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\left\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}^{\otimes 2}\Big]\\
&\quad+E\left(\frac{\{var(Z\mid\mathbf{X})\}^2\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}}{\left[E\left\{var(Z\mid\mathbf{X})\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right\}\right]^2}\right)
\end{aligned}$$

$$+\,2E\Bigg[Z\{Z-E(Z\mid\mathbf{X})\}^2\{f(\mathbf{X})-f^*(\mathbf{X})\}\{g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})-g^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}\\ \times h^{*2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\left\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}^{\otimes 2}\Bigg]\\ -\,2E\left[\frac{\{var(Z\mid\mathbf{X})\}^2\{g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})-g^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\left\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\right\}^{\otimes 2}}{E\left\{var(Z\mid\mathbf{X})\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right\}}\right]. \tag{8}$$

Here, $g'^*(\cdot)$ is the first derivative of $g^*(\cdot)$.

Interestingly, even when $f^*(\mathbf{X})=f(\mathbf{X})$, $g^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})=g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ and $h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})=g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$, the estimator still does not achieve the optimal efficiency bound. This is in stark contrast with the proposed estimator, which is optimal as long as $f^*(\mathbf{X})=f(\mathbf{X})$.

Theorem 3.4. *Assume the regularity conditions listed in Supplementary Material A.3 hold. When $n\to\infty$, the estimator obtained from solving (7) satisfies $\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L=O_p(h^2+n^{-1/2}h^{-1/2})$. Furthermore, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_L-\boldsymbol{\beta}_L)\to N(\mathbf{0},\mathbf{A}_3^{-1}\mathbf{B}_3\mathbf{A}_3^{-1^{\mathrm{T}}})$ in distribution when $n\to\infty$, where*

$$\begin{aligned}\mathbf{A}_3 &= E\left[var(Z\mid\mathbf{X})g'^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}\mathbf{X}_L^{\mathrm{T}}\right],\\ \mathbf{B}_3 &= \sigma^2E\left[var(Z\mid\mathbf{X})g'^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}\right]\\ &\quad+E\left[\{f(\mathbf{X})-f^*(\mathbf{X})\}^2var(Z\mid\mathbf{X})g'^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\{\mathbf{X}_L-\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}\right].\end{aligned} \tag{9}$$

When $f^*(\mathbf{X})=f(\mathbf{X})$ and $\mathbf{u}^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})=\mathbf{u}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$, the estimator is efficient.

4. Estimation and Inference when Z is continuous

When Z is continuous, typically representing the dosage, the treatment effect of the form $Zg(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ may not be adequate. The only conclusion that can be drawn from such a model is that if $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ is positive, the largest dosage Z should be selected and when it is negative, the smallest dosage should be taken. More useful models in this case include a quadratic treatment effect $Zg_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})+Z^2g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$, a cubic model $Zg_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})+Z^2g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})+Z^3g_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$, or some other nonlinear model of Z which does not have to limit to polynomials. For example, if the quadratic model is used, a new patient with covariate X_0 and $g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_0)<0$ would have the best treatment result if $Z_0=-g_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_0)/\{2g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_0)\}$. In this paper, we consider a general polynomial model of the form

$$Y=f(X)+\sum_{k=1}^{K}Z^kg_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})+\epsilon, \tag{10}$$

while keep all other aspects of the model assumption identical to that of (1). Similar to (2), the pdf in this case is

$$\begin{aligned} & f_{\mathbf{X},Z,Y}(\mathbf{x}, z, y, \boldsymbol{\beta}, \sigma, g, f, \eta) \\ & \quad = \eta(\mathbf{x}) f_{Z|\mathbf{X}}(\mathbf{x}, z)\sigma^{-1}\phi[\sigma^{-1}\{y - f(\mathbf{x}) - \sum_{k=1}^{K} z^k g_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}], \end{aligned} \tag{11}$$

and it has efficient score $\mathbf{S}_{\text{eff}}(\mathbf{x}, z, y, \boldsymbol{\beta}, g, f, \sigma) = \sigma^{-2}\epsilon\Big\{\mathbf{u} - \sum_{r=1}^{K} W_r E \Big(W_r \mathbf{U} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\Big)\Big\}$, where $\mathbf{U} = \sum_{k=1}^{K} g_{k1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\left\{Z^k - E(Z^k \mid \mathbf{X})\right\}\mathbf{x}_L$. Here $g_{k1}(\cdot)$ is the first derivative of $g_k(\cdot)$ for $k = 1, \ldots, K$, $\epsilon \equiv y - f(\mathbf{x}) - \sum_{k=1}^{K} z^k g_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. Furthermore, $W_1, \ldots, W_K$ are linear transformations of $Z - E(Z \mid \mathbf{X}), \ldots, Z^K - E(Z^K \mid \mathbf{X})$, i.e. $\mathbf{A}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})(W_1, \ldots, W_K)^{\mathrm{T}} = \{Z - E(Z \mid \mathbf{X}), \ldots Z^K - E(Z^K \mid \mathbf{X})\}^{\mathrm{T}}$, such that $E(W_j W_k \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) = I(j = k)$, i.e. they form a set of orthonormal bases. Supplementary Material A.9 provides the detailed construction of W_k's and the derivation of the efficient score function. Naturally, based on the form of the efficient score function and our experience in the binary Z case, we propose a general class of consistent estimating equations for $\boldsymbol{\beta}$ as

$$\sum_{i=1}^{n}\{y_i - f^*(\mathbf{x}_i) - \sum_{k=1}^{K} z_i^k \widehat{g}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}\{\widehat{\mathbf{u}}_i - \sum_{r=1}^{K} w_{ri}\widehat{E}(W_r\widehat{\mathbf{U}} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\} = \mathbf{0}, \tag{12}$$

where $\widehat{\mathbf{U}} = \sum_{k=1}^{K} \widehat{g}_{k1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\{Z^k - E(Z^k \mid \mathbf{X})\}\mathbf{x}_L$, f^* is an arbitrary function of $\mathbf{x}$. Here at any $\mathbf{x}_0 = \mathbf{x}_1, \ldots, \mathbf{x}_n$ and any $\boldsymbol{\beta}$, $\widehat{g}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0, \boldsymbol{\beta}), \widehat{g}_{k1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0, \boldsymbol{\beta})$ are nonparametric profiling estimators obtained from

$$\sum_{i=1}^{n} w_{i0}[y_i - f^*(\mathbf{x}_i) - \sum_{k=1}^{K} z_i^k\{\alpha_{kc} + \alpha_{k1}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_0)\}]\mathbf{v}_i = \mathbf{0}, \tag{13}$$

where $\mathbf{V} = \Big[\left\{Z - E(Z \mid \mathbf{X})\right\}(1, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0), \ldots, \left\{Z^K - E(Z^K \mid \mathbf{X})\right\}(1, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0)\Big]^{\mathrm{T}}$.

Similar to the binary Z case, we also have the standard $\sqrt{n}$ consistency and local efficiency of $\widehat{\boldsymbol{\beta}}$ for continuous Z. We state the results in Theorem 4.1 while skipping the proof because it is almost identical to that of Theorems 3.1.

Theorem 4.1. *Assume the regularity conditions listed in Supplementary Material A.3 hold. When $n \to \infty$, the estimator obtained by solving (12) satisfies $\widehat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_L = O_p(h^2 + n^{-1/2}h^{-1/2})$. Furthermore, the estimator described in (12) and (13) satisfies the property that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_L) \to N(\mathbf{0}, \mathbf{A}_4^{-1}\mathbf{B}_4\mathbf{A}_4^{-1})$ in distribution as $n \to \infty$, where*

$$\mathbf{A}_4 = \sigma^2 E\left[\{\mathbf{U} - \sum_{r=1}^{K} W_r E(W_r\mathbf{U} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}\right]$$

$$\mathbf{B}_4 \quad = \quad E\left([\sigma^2 + \{f(\mathbf{X}) - f^*(\mathbf{X})\}^2]\{\mathbf{U} - \sum_{r=1}^{K} W_r E(W_r \mathbf{U} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}\right).$$

Additionally, when $f^*(\mathbf{X}) = f(\mathbf{X})$, the estimator obtains the optimal efficiency bound.

5. Estimation and Inference when Z is categorical

To make the analysis complete, we now consider the case where Z is categorical. This arises naturally in practice when several different treatment arms are compared. Assume we consider K treatment arms in addition to the control arm. Thus, we have K binary variables, $Z_1, \dots, Z_K$, each takes value 0 or 1. Note that at most one of the $Z_k, k = 1, \dots, K$ values can be 1 in each observation. A sensible model in this scenario is

$$Y = f(X) + \sum_{k=1}^{K} Z_k g_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + \epsilon, \tag{14}$$

and the pdf is

$$\begin{aligned} &f_{\mathbf{X},\mathbf{Z},Y}(\mathbf{x}, \mathbf{z}, y, \boldsymbol{\beta}, \sigma, g, f, \eta) \\ &\quad = \eta(\mathbf{x}) f_{\mathbf{Z}|\mathbf{X}}(\mathbf{x}, \mathbf{z}) \sigma^{-1} \phi[\sigma^{-1}\{y - f(\mathbf{x}) - \sum_{k=1}^{K} z_k g_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}], \end{aligned} \tag{15}$$

where $\mathbf{Z} = (Z_1, \dots, Z_k)^{\mathrm{T}}$. The corresponding efficient score is $\mathbf{S}_{\mathrm{eff}}(\mathbf{x}, \mathbf{z}, y, \boldsymbol{\beta}, g, f, \sigma) = \sigma^{-2}\epsilon\left\{\mathbf{u} - \sum_{r=1}^{K} W_r E\left(W_r \mathbf{U} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}$, where $\mathbf{U} = \sum_{k=1}^{K} g_{k1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) \{Z_k - E(Z_k \mid \mathbf{X})\}\mathbf{x}_L$. Here $g_{k1}(\cdot)$ is the first derivative of $g_k(\cdot)$ for $k = 1, \dots, K$, $\epsilon \equiv y - f(\mathbf{x}) - \sum_{k=1}^{K} z_k g_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. Furthermore, $W_1, \dots, W_K$ are linear transformations of $Z_1 - E(Z_1 \mid \mathbf{X}), \dots Z_K - E(Z_K \mid \mathbf{X})$, i.e. $\mathbf{A}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})(W_1, \dots, W_K)^{\mathrm{T}} = \{Z - E(Z \mid \mathbf{X}), \dots Z_K - E(Z_K \mid \mathbf{X})\}^{\mathrm{T}}$, such that $E(W_j W_k \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) = I(j = k)$, i.e. they form a set of orthonormal bases. We can see that the efficient score has much resemblance with the case for the continuous Z case, except that we are now treating a multivariate $\mathbf{Z}$. We show the derivation of the efficient score in Supplementary Material A.10, where the detailed construction of W_k's are also given. Similar to the continuous univariate Z case, we propose a general class of consistent estimating equations for $\boldsymbol{\beta}$ as

$$\sum_{i=1}^{n}\{y_i - f^*(\mathbf{x}_i) - \sum_{k=1}^{K} z_{ki}\widehat{g}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}\{\widehat{\mathbf{u}}_i - \sum_{r=1}^{K} w_{ri}\widehat{E}(W_r\widehat{\mathbf{U}} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\} = \mathbf{0}, \tag{16}$$

where $\widehat{\mathbf{U}} = \sum_{k=1}^{K} \widehat{g}_{k1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\{Z_k - E(Z_k \mid \mathbf{X})\}\mathbf{x}_L$, f^* is an arbitrary function of $\mathbf{x}$. Here at any $\mathbf{x}_0 = \mathbf{x}_1, \dots, \mathbf{x}_n$ and any $\boldsymbol{\beta}$, $\widehat{g}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0, \boldsymbol{\beta}), \widehat{g}_{k1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0, \boldsymbol{\beta})$ are nonparametric profiling estimators obtained from

$$\sum_{i=1}^{n} w_{i0}[y_i - f^*(\mathbf{x}_i) - \sum_{k=1}^{K} z_{ki}\{\alpha_{kc} + \alpha_{k1}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_0)\}]\mathbf{v}_i = \mathbf{0}, \tag{17}$$

where $\mathbf{V} = \Big[\{Z_1 - E(Z_1 \mid \mathbf{X})\}(1, \boldsymbol\beta^{\mathrm{T}}\mathbf{X} - \boldsymbol\beta^{\mathrm{T}}\mathbf{x}_0), \ldots, \{Z_K - E(Z_K \mid \mathbf{X})\}(1, \boldsymbol\beta^{\mathrm{T}}\mathbf{X} - \boldsymbol\beta^{\mathrm{T}}\mathbf{x}_0)\Big]^{\mathrm{T}}$. Similarly, we summarize the asymptotic properties of $\widehat{\boldsymbol\beta}$ in Theorem 5.1 and skip the proof because of its similarity to that of Theorems 3.1.

Theorem 5.1. *Assume the regularity conditions listed in Supplementary Material A.3 hold. When $n \to \infty$, the estimator obtained by solving (16) satisfies $\widehat{\boldsymbol\beta}_L - \boldsymbol\beta_L = O_p(h^2 + n^{-1/2}h^{-1/2})$. Furthermore, the estimator described in (16) and (17) satisfies the property that $\sqrt{n}(\widehat{\boldsymbol\beta}_L - \boldsymbol\beta_L) \to N(\mathbf{0}, \mathbf{A}_5^{-1}\mathbf{B}_5\mathbf{A}_5^{-1^{\mathrm{T}}})$ in distribution as $n \to \infty$, where*

$$\begin{aligned}
\mathbf{A}_5 &= \sigma^2 E\left[\{\mathbf{U} - \sum_{r=1}^{K} W_r E(W_r\mathbf{U} \mid \boldsymbol\beta^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}\right] \\
\mathbf{B}_5 &= E\left([\sigma^2 + \{f(\mathbf{X}) - f^*(\mathbf{X})\}^2]\{\mathbf{U} - \sum_{r=1}^{K} W_r E(W_r\mathbf{U} \mid \boldsymbol\beta^{\mathrm{T}}\mathbf{X})\}^{\otimes 2}\right).
\end{aligned}$$

Additionally, when $f^*(\mathbf{X}) = f(\mathbf{X})$, the estimator obtains the optimal efficiency bound.

6. Simulation studies

We conduct five sets of simulation experiments, covering all the scenarios discussed so far, to investigate the finite sample performance of the methods proposed in Section 2. The results under various scenarios reflect the superior performance of Method I given in (5) and show reasonably accurate inference results as well. We also compute the percentage of making correct decisions based on $g(\cdot)$ functions. In most of the scenarios, we are able to make correct decisions more than 90% of the times. In each experiment, we simulate 500 data sets.

6.1. Simulation 1

In the first experiment, we consider a relatively simple setting. Specifically, the covariate $\mathbf{X}_i = (X_{i1}, X_{i2})^{\mathrm{T}}$ is generated from bivariate normal distribution with zero mean and identity covariance matrix. We set $f(\mathbf{X}_i) = 0.05(X_{i1} + X_{i2})$ and $g(\boldsymbol\beta^{\mathrm{T}}\mathbf{X}_i) = 2\boldsymbol\beta^{\mathrm{T}}\mathbf{X}_i$, where $\boldsymbol\beta = (1, -1)^{\mathrm{T}}$. Note that the function g is monotone. We generate the model error ϵ_i from a centered normal distribution with standard error 0.3, and generate the treatment indicator Z_i from a Bernoulli distribution with probability 0.5.

We implement four methods for estimating the unknown parameter $\boldsymbol\beta$ for comparison. Method I is the one given in (5). While implementing Method I, we consider $f^* = 0$ to avoid estimating $f(\mathbf{X})$ due to curse of dimensionality. Method II is the alternative method proposed in (6), where we let $f^* = g^* = 0, h^* = 1$. Method III corresponds to (7) and Method IV refers to the method proposed in Section 2.1. To ensure identifiability, we fix the first component of $\boldsymbol\beta$ to be 1.

From the results reported in Table 1, we can observe that while all the estimation methods have very small biases and standard deviations, Method I performs significantly better than all other methods in terms of both estimation bias and the standard errors. This is within our expectation because Method I is a locally efficient estimator. In addition, in our experience, the computation time for all four methods are also comparable. Thus we recommend using Method I due to its superior performance. Consequently, we focus on Method I to proceed with the inference of $\boldsymbol{\beta}$ and the estimation and inference of $g(\cdot)$.

Although Theorem 3.1 provides an explicit form of the asymptotic variance, the resulting form contains $f(\mathbf{X})$, whose estimation is subject to the curse of dimensionality and we have successfully avoided so far. Thus, to proceed with inference on $\boldsymbol{\beta}$, we evaluate the coverage probabilities of the 95% confidence regions for $\boldsymbol{\beta}$ resorting to the bootstrap method. The coverage probabilities for β_2 are reported in the upper part of Table 2 under Method I. All results are reasonably close to the nominal level. To evaluate the accuracy of the subsequent treatment assignment rule based on $\text{sign}\{\hat{g}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X})\}$, we calculate the percentage of making correct decisions, defined as $\text{PCD} = 1 - n^{-1}\sum_{i=1}^{n} |I\{\widehat{g}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i) > 0\} - I\{g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i) > 0\}|$. We also evaluate the optimal value function via

$$\widehat{\text{VF}} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i I[Z_i = I\{\widehat{g}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i) > 0\}]}{\text{pr}[Z_i = I\{\widehat{g}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i) > 0\}|\mathbf{X}_i]}.$$

In this particular simulation, due to covariate-independent randomization, $\widehat{\text{VF}} = (2/n)\sum_{i=1}^{n} Y_i I[I\{\widehat{g}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}_i) > 0\}]$. The values of PCD and $\widehat{\text{VF}}$ and their standard errors are reported in the upper part of Table 2 under Method I.

For comparison, we consider the semiparametric single index model (SIM) approach by [31] to estimate the optimal individualized treatment strategy. Results are provided in Table 2. We can clearly see that our proposed Method I performs better in that it yields smaller bias in estimating parameter β_2, PCD and the optimal value function, compared to the SIM approach. The estimation variability of our Method I is also less than SIM method. To further check the performance, we plot the median, pointwise 95% confidence bands for the estimation of $g(\cdot)$ in Figure 1 and the performance is satisfactory.

TABLE 1
Bias and standard deviation (SD) for the estimation of β_2 in Simulation 1 at $n = 600$.

Methods	Bias	SD
I	−0.0015	0.0177
II	−0.0056	0.1046
III	0.0028	0.1340
IV	0.0214	0.1415

TABLE 2

Estimation and inference results for β_2 in Simulation 1 using Method I and SIM approach [31] at $n = 600$. We report biases, empirical standard deviations (SD), mean estimated standard errors (SE) and 95% confidence interval coverage of the single index coefficients, and estimation results for PCD and the optimal value function over 500 replications with their empirical standard deviations (SD). True optimal value function, i.e., $\max_d V(d)$, is 1.1306.

Methods	Bias(β_2)	SD(β_2)	SE(β_2)	CP(β_2)	PCD	SD(PCD)	$\widehat{\text{VF}}$	SD($\widehat{\text{VF}}$)
I	−0.0015	0.0177	0.0194	0.972	0.9953	0.0040	1.1264	0.1071
SIM	−0.0046	0.0940	0.0867	0.93	0.9881	0.0108	0.7978	0.0665

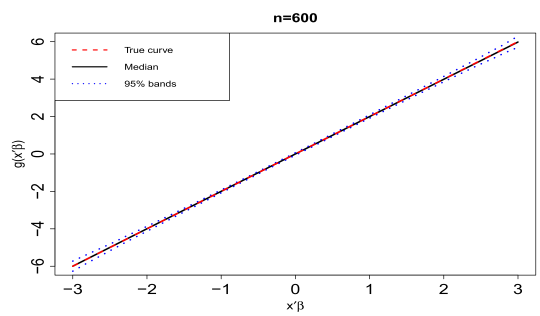

FIG 1. *Estimation of $g(\cdot)$, when $g(\mathbf{x}'\beta) = 2\mathbf{x}'\beta$ in Simulation 1.*

6.2. Simulation 2

In the second experiment, the covariate $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})^{\mathrm{T}}$, where X_{i1}, X_{i2}, X_{i3} are independently generated from a uniform(0,1) distribution. We let $f(\mathbf{X}_i) = 0.05(X_{i1} + X_{i2} + X_{i3})$, $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i) = \sin(\pi\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)$ with $\boldsymbol{\beta}^{\mathrm{T}} = (1, -1, 2)$. Thus, the g function here is both nonlinear and non-monotone. The regression errors ϵ_i and the treatment indicator Z_i are generated identically as in Simulation 1.

Similar to Simulation 1, we implement the four methods to compare the estimation of unknown parameter $\boldsymbol{\beta}$, while fixing the first component of $\boldsymbol{\beta}$ at 1. The estimation bias and standard errors are reported in Table 3. From the results, we can again conclude that Method I performs significantly better than all other methods, even though all methods have acceptable bias and standard errors. We compute the coverage probabilities of the 95% bootstrapped confidence regions for $\boldsymbol{\beta}$ for Method I. The results are reported in Table 4. All results are reasonably close to the nominal level. We also report the mean and standard deviations of $\widehat{\text{VF}}$ and PCD in the lower part of Table 4 under Method I. We can see that even though PCD by the proposed method is higher than the SIM method, SIM estimated a larger optimal value function, much larger than the true optimal value function. This shows the SIM estimate of the optimal value function is severely biased. In contrast, the proposed method results in very small bias in estimating the optimal value function. To further check the performance, we plot the median, pointwise 95% confidence bands for the estimation

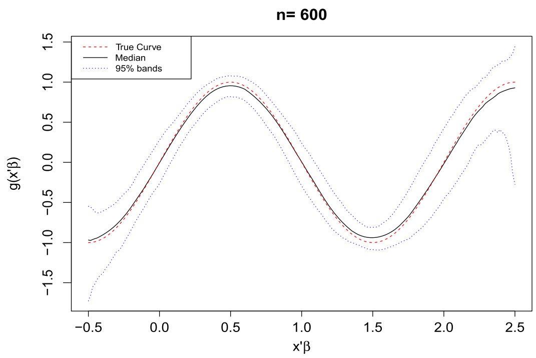

FIG 2. *Estimation of* $g(\cdot)$, $g(\mathbf{x}'\boldsymbol{\beta}) = \sin(\pi\mathbf{x}'\boldsymbol{\beta})$ *in Simulation 2.*

of $g(\cdot)$ in Figure 2 and the performance is satisfactory. We also report the results obtained by the SIM approach [31] in Table 4. As expected, our proposed Method I performs better compared to SIM method, even when the treatment effect function is not monotone. Our proposed Method I has smaller bias and less estimation variability.

TABLE 3
Bias and standard deviation (SD) for the estimation of (β_2, β_3) *in Simulation 2 at* $n = 600$.

Methods	Bias(β_2)	SD(β_2)	Bias(β_3)	SD(β_3)
I	−0.0048	0.0604	0.0059	0.0979
II	−0.0276	0.1831	0.0796	0.2685
III	−0.0413	0.0835	0.0886	0.1438
IV	−0.0512	0.0906	0.0930	0.1442

TABLE 4
Estimation and inference results for β_2 *in Simulation 2 using SIM approach [31] at* $n = 600$. *Other caption is same as Table 2. True optimal value function, i.e.,* $\max_d V(d)$, *is 0.3937.*

Methods	Bias(β_2)	SD(β_2)	SE(β_2)	Bias(β_3)	SD(β_3)	SE(β_3)
I	−0.0048	0.0604	0.0971	0.0059	0.0979	0.1456
SIM	−0.0016	0.2754	0.1565	0.0674	0.5060	0.4961
Methods	CP(β_2)	CP(β_3)	PCD	SD(PCD)	$\widehat{\text{VF}}$	SD($\widehat{\text{VF}}$)
I	0.972	0.976	0.975	0.010	0.3964	0.0328
SIM	0.778	0.934	0.9554	0.0223	0.7215	0.0270

6.3. Simulation 3

In the third simulation, the covariate $\mathbf{X}_i = (X_{i1}, X_{i2})^{\mathrm{T}}$ is generated the same as in Simulation 1. We set $g(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i) = 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i + \sin(2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)$ with $\boldsymbol{\beta} = (1, -1)^{\mathrm{T}}$.

The regression errors ϵ_i and the baseline response model $f(\mathbf{X}_i)$ are generated identically as in Simulation 1. We allow the distribution of treatment indicator Z_i to depend on the covariates. Specifically, Z_i is generated from a Bernoulli distribution with probability of success $\exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i)/\{1+\exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i)\}$, where $\boldsymbol{\gamma}=(0.3,-0.2)^{\mathrm{T}}$. Note that here we consider covariate adjusted randomization.

From the results in Table 5, we can observe that all estimation methods have acceptable biases and standard errors even in covariate adjusted randomization setup. Method I performs significantly better compared to other methods, as expected. In the upper part of Table 6, we summarize the coverage probability of the 95% bootstrap confidence intervals of $\boldsymbol{\beta}$, mean and standard deviations for estimating $\widehat{\mathrm{VF}}$ and PCD for Method I. We also plot the pointwise 95% confidence band and median for $g(\cdot)$ in Figure 3. From the results provided in Table 6, we can clearly see that our proposed Method I performs better in estimating parameter, β_2, PCD and value function, compared to the SIM approach. The estimation variability of our Method I is also less than SIM.

Table 5

Bias and standard deviation (SD) for the estimation of β_2 in Simulation 3 at $n=600$.

Methods	Bias	SD
I	−0.0012	0.0294
II	−0.0044	0.1365
III	0.0038	0.0981
IV	0.0153	0.1069

Table 6

Estimation and inference results for β_2 in Simulation 3 using Method I and SIM approach at $n=600$. Other caption is same as Table 2. True optimal value function, i.e., $\max_d V(d)$, is 0.9027.

Methods	Bias(β_2)	SD(β_2)	SE(β_2)	CP(β_2)	PCD	SD(PCD)	$\widehat{\mathrm{VF}}$	SD($\widehat{\mathrm{VF}}$)
I	−0.0012	0.0294	0.0296	0.946	0.9900	0.0064	0.8977	0.0775
SIM	0.0358	0.0997	0.0823	0.834	0.9857	0.0114	0.8555	0.0766

6.4. Simulation 4

In this simulation setting, we consider the categorical treatment indicator. Our covariate $\mathbf{X}_i=(X_{i1},X_{i2},X_{i3})^{\mathrm{T}}$ is generated from multivariate standard normal distribution. We consider three levels for treatment indicator Z_i, 0, 1 and 2 with probability 0.4, 0.4 and 0.2, respectively. Here, relative to the baseline at $Z=0$, $g_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)=0.5(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)^2-1$ is the treatment effect, when $Z_i=1$ and $g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)=(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\sin(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)-1$ is the treatment effect for $Z_i=2$ with $\boldsymbol{\beta}=(1,-1,1)^{\mathrm{T}}$. Here, $f(\mathbf{X}_i)=0.5(X_{i1}+X_{i2}+X_{i3})$ and the regression error ϵ_i is generated identically as in Simulation 1.

As we have repeatedly observed in previous simulations that Method I is superior to other three methods, we only consider Method I to estimate the

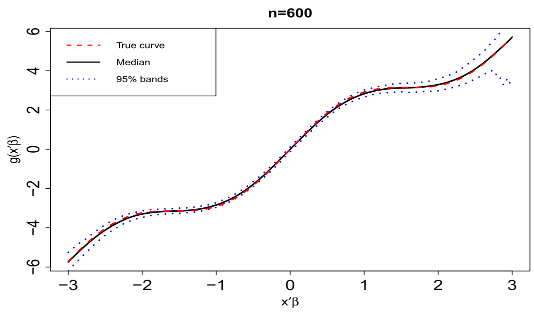

FIG 3. *Estimation of $g(\cdot)$, when $g(\mathbf{x}'\beta) = 2(\mathbf{x}'\boldsymbol{\beta}) + \sin\{2(\mathbf{x}'\boldsymbol{\beta})\}$ in Simulation 3.*

unknown parameter $\boldsymbol{\beta}$ and to proceed with the inference of $\boldsymbol{\beta}$ and the estimation and inference of $g_1(\cdot)$ and $g_2(\cdot)$. The results are summarized in the first parts of Tables 7 and 8. In the second parts of the aforementioned tables, we have provided the results obtained from SIM. From Table 7, we can see that using our proposed Method I, estimation bias is acceptable and the coverage probability obtained from bootstrap confidence interval is close to the nominal level of 0.95. On the other hand, as expected, due to the non-monotone treatment effect models $g_1(\cdot)$ and $g_2(\cdot)$, SIM yields large bias. Also, SIM performs poorly in terms of inference of $\boldsymbol{\beta}$. From Table 8, we can clearly observe that our proposed Method I yields higher PCD values and also small bias in estimating the optimal value function, compared to SIM. In Figure 4, we have computed pointwise 95% confidence band and median for $g_1(\cdot)$ and $g_2(\cdot)$.

TABLE 7
Bias, standard deviation (SD) and coverage probabilities (CP) for the estimation of β_2 and β_3 in Simulation 4 using proposed method and SIM method at $n = 600$.

	β_2				β_3			
Methods	Bias	SD	SE	CP	Bias	SD	SE	CP
I	0.0001	0.0859	0.1068	0.964	−0.0085	0.0983	0.1140	0.954
SIM	0.6166	2.3757	0.4028	0.200	−0.5619	1.5774	0.3716	0.400

TABLE 8
The percentage of correct decisions (PCDs) with standard errors, and the estimated optimal value function ($\widehat{VF}$) with empirical standard deviations in Simulation 4, using proposed method and SIM method at $n = 600$. Here, PCD1 and PCD2 are related to $g_1(\cdot)$ and $g_2(\cdot)$, respectively. True optimal value function, i.e., $\max_d V(d)$, is 0.8056.

Methods	PCD1	SD(PCD1)	PCD2	SD(PCD2)	$\widehat{\text{VF}}$	SD($\widehat{\text{VF}}$)
I	0.9455	0.0295	0.8821	0.0558	0.7884	0.1521
SIM	0.5175	0.0263	0.5000	0.0263	−0.2997	0.0958

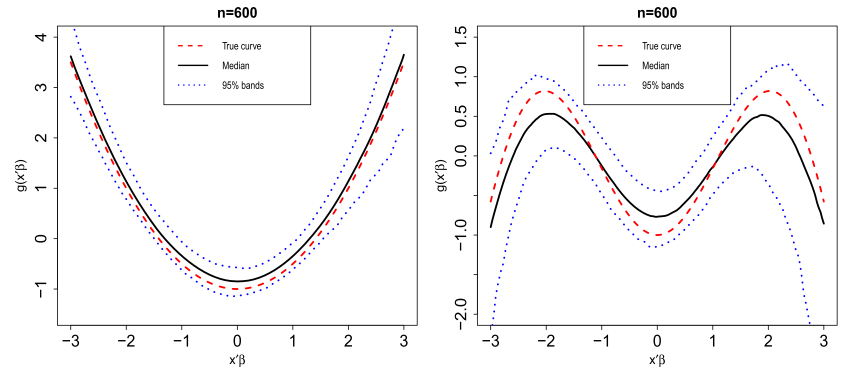

FIG 4. *Estimation of* $g_1(\cdot)$ *(left) and* $g_2(\cdot)$ *(right), where* $g_1(\mathbf{x}'\beta) = 0.5(\mathbf{x}'\boldsymbol{\beta})^2 - 1$ *and* $g_2(\mathbf{x}'\beta) = (\mathbf{x}'\boldsymbol{\beta})\sin(\mathbf{x}'\boldsymbol{\beta}) - 1$ *in Simulation 4.*

6.5. Simulation 5

In the last setting, we consider continuous treatment indicator. We generate covariate $\mathbf{X}_i$ as in Simulation 1, and consider a quadratic treatment effect $Z_i g_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i) + Z_i^2 g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)$. Here Z_i follows the uniform distribution on $[0, 1]$, $g_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i) = (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - 2)^2 - 1$ and $g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i) = \{0.1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i/3)^3 - 1\}\{(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i - 2)^2 - 1\}$ with $\boldsymbol{\beta} = (1, -1)^{\mathrm{T}}$. We let $f(\mathbf{X}_i) = 0.5(X_{i1} + X_{i2})$ and generate the regression error ϵ_i as in Simulation 1.

Similar to Simulation 4, we implement Method I to estimate the unknown parameter $\boldsymbol{\beta}$ and proceed with the inference of $\boldsymbol{\beta}$ and the estimation of value function and PCD for $g_1(\cdot)$ and $g_2(\cdot)$. The results are summarized in the upper parts of Tables 9 and 10. From the upper part of Table 9, we can see that estimation bias is acceptable and the coverage probability obtained from bootstrap confidence interval is close to the nominal level of 0.95 for our proposed method. Also, value function obtained by proposed method is close to true value. In Table 10, we have summarized the mean and standard deviations for PCD for $g_1(\cdot)$ and $g_2(\cdot)$. In Figure 5, we also plot pointwise 95% confidence band and median for $g_1(\cdot)$ and $g_2(\cdot)$. We also compare the performance of our method with the existing kernel assisted learning (KAL) method [50]. From the lower part of Table 9, we clearly see that the our method performs better than KAL method. The estimators obtained by KAL method yield large bias and standard deviation. Indeed, [50] consider the following set up: $E(Y \mid Z, \mathbf{X}) = f(\mathbf{X}) + Q\{Z - g(\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X})\}H(\mathbf{X})$, where $Q(\cdot)$ is a unimodal function which is maximized at 0 and $H(\mathbf{X})$ is a non-negative function. Then $E(Y \mid Z, \mathbf{X})$ is maximized at $Z = g(\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X})$, where $g : \mathbb{R} \to \mathcal{A}$ is a predefined link function which ensures that the suggested dose falls within safe dose range ($\mathcal{A}$). On the other hand, we consider a general polynomial model, $E(Y \mid Z, \mathbf{X}) = f(\mathbf{X}) + \sum_{k=1}^{K} Z^k g_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$, where K is the degree of the polynomial. In this particular simulation setting, we consider K=2. Thus, we try to re-write our model similar to their model so that we can easily implement their method in our simulation study. After re-writing, we get

$$E(Y \mid Z, \mathbf{X}) = \left\{ f(\mathbf{X}) - \frac{g_1^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{4g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} \right\} + \left[- \left\{ Z - \frac{-g_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{2g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})} \right\}^2 \right] \{-g_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\}$$

and our model satisfies the conditions of KAL when $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} \leq 1$ or $3 \leq \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} < 6.4633$. Thus, we use the subset of simulated data that fits into the KAL model requirement to to obtain the KAL estimate of β_2 and the value function.

Remark 5. *For illustration, we performed model selection, where we choose K that minimizes $n^{-1}\sum_{i=1}^{n}\{y_i - \sum_{k=1}^{K} z_i^k \widehat{g}_k(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\}^2 + 2K^2\sqrt{log(n)/n}$. The method is able to select the correct $K(=2)$ 87.6% of the times at $n = 600$. As the sample size n increases, the percentage of detecting the true K also increases. For $n = 800$ and $n = 1000$, the detection rate are 91.6% and 96.6%, respectively. Ideally, the consistency of the model selection procedure should be established and post model selection inference issues need to be studied. These are not straightforward and we leave them to future work.*

Table 9

Estimation and inference results of β_2 and estimation of the optimal value function using our proposed method and KAL method [50] at $n = 600$ in Simulation 5. True optimal value function, i.e., $\max_d V(d)$, is 0.3926.

Methods	Bias(β_2)	SD(β_2)	SE(β_2)	CP(β_2)	$\widehat{\mathrm{VF}}$	SD($\widehat{\mathrm{VF}}$)
I	−0.0399	0.1107	0.1361	0.942	0.3797	0.1497
KAL	0.2855	2.1624	2.7085	0.564	0.0397	13.5458

Table 10

The percentage of correct decisions (PCDs) with standard deviations obtained by proposed method at $n = 600$ in Simulation 5. Here, PCD1 and PCD2 are related to $g_1(\cdot)$ and $g_2(\cdot)$, respectively.

PCD1	SD(PCD1)	PCD2	SD(PCD2)
0.8957	0.0622	0.8953	0.0627

We have also considered the sample sizes, $n = 800$ and $n = 1000$ for each simulation study and summarized the results in Section A.11 in the supplementary material. The conclusions are similar to that under the sample size $n = 600$.

7. Real data example

In this section, we demonstrate our proposed method using the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. The STAR*D study

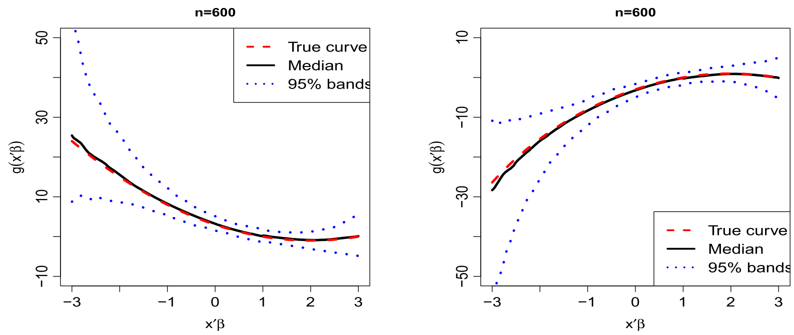

FIG 5. *Estimation of* $g_1(\cdot)$ *(left) and* $g_2(\cdot)$ *(right), where* $g_1(\mathbf{x}'\beta) = (\mathbf{x}'\boldsymbol{\beta} - 2)^2 - 1$ *and* $g_2(\mathbf{x}'\beta) = \{0.1(\mathbf{x}'\boldsymbol{\beta}/3)^3 - 1\}\{(\mathbf{x}'\boldsymbol{\beta} - 2)^2 - 1\}$ *in Simulation 5.*

was a sequential, multiple-assignment, randomized trial (SMART, [21] and [26]) for patients with non-psychotic major depressive disorder. This study aimed to determine what antidepressant medications should be given to patients to yield the optimal treatment effect. For illustration purpose, we focused on a subset of 319 participants who were given treatment bupropion (BUP) or sertraline (SER). Among these participants, 153 patients were randomly assigned to the BUP treatment group and the rest of them were assigned to the SER treatment group. The 16-item Quick Inventory of Depressive Symptomatology-Self-Report scores (QIDS-SR(16)) were recorded at treatment visits for each patient and considered as the outcome variable. We used $R = -$QIDS-SR(16) as the reward to accommodate the assumption that the larger reward is more desirable.

In the original data set, there are a large number of covariates that describe participant features such as age, gender, socioeconomic status, and ethnicity. However, many of them are not significantly related to the QIDS-SR(16). According to the study conducted by [6], we included five important covariates into our study. These five covariates are "fatigue or loss of energy" in baseline protocol eligibility (DSMLE, X_1), patient's age (Age, X_2), "ringing in ears" in patient rated inventory of side effects at Level 2 (EARNG-Level2, X_3), "feeling of worthlessness or guilt" in baseline protocol eligibility (DSMFW, X_4) and "hard to control worrying" in psychiatric diagnostic, screening questionnaire at baseline (WYCRL, X_5).

Let $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5)^{\mathrm{T}}$ and let Z_i be the treatment indicator where $Z_i = 1$ means the participant was assigned to the BUP treatment, $Z_i = 0$ represents the SER group. We fitted the model $Y_i = f(\mathbf{X}_i) + Z_i g(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) + \epsilon_i$. We are interested in finding the optimal treatment assignment to the patients. Since treatment effects are described by $g(\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})$, we first obtained the estimators for $\boldsymbol{\beta}$ and the function $g(\cdot)$, using the proposed locally efficient estimator described in Method I. In the analysis reported here, we did not attempt to approximate $f(\mathbf{X})$. We simply set $f^*(\mathbf{X}) = 0$. In the supplementary file, we further considered a more complex working model $f^*(\mathbf{X})$. The results are summarized in

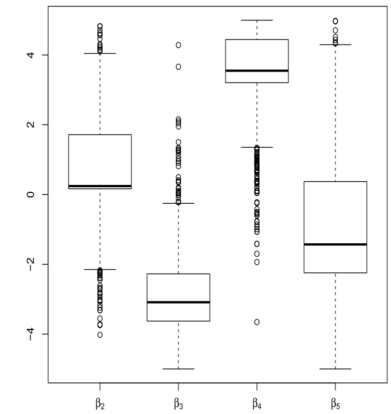

FIG 6. *Box plots of the bootstrap estimators of the coefficients* $\beta_2, \ldots, \beta_4$.

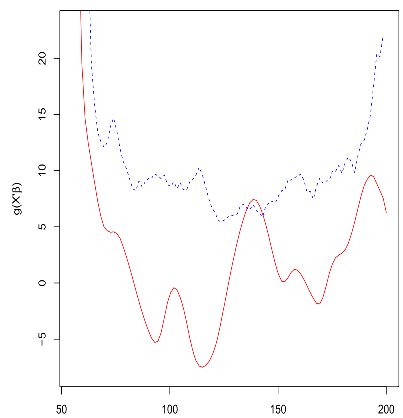

FIG 7. $\widehat{g}(\cdot)$ *and upper 95% pointwise quantile curve based on 1000 permuted data.*

Section A.15 of the supplementary file. We fixed the coefficient of X_1 to be 1 and estimated the remaining four coefficients. The estimator for $(\beta_2, \beta_3, \beta_4, \beta_5)^{\mathrm{T}}$ was $(2.59033, -3.31058, 1.49003, 3.64217)^{\mathrm{T}}$. To check if these coefficients are significant, we used bootstrap to obtain the estimation variance. We provide a box plot of the bootstrap estimators of $\boldsymbol{\beta}$ in Figure 6. We also constructed the confidence intervals for β_2 to β_5 respectively. We found that the effects of β_3 and β_4 are significantly different from zero. The 95% confidence intervals for $\beta_2, \ldots, \beta_5$ are, respectively, $(-2.001, 3.621)$, $(-4.737, -0.299)$, $(0.657, 4.920)$ and $(-4.207, 3.224)$.

We also estimated the function $g(\cdot)$ nonparametrically according to the method introduced in the Section 2.2 to obtain $\widehat{g}$, with its plot in Figure 7. To evaluate the significance of the nonparametric function $\widehat{g}(\cdot)$, we applied a permutation test. Specifically, we randomly permute the treatment label Z_i's and estimate

$g(\cdot)$ based on each permuted data. If the true $g(\cdot)$ function is zero, then the permutation should not alter the true function $g(\cdot)$ and the function estimated from the original real data, i.e. $\widehat{g}$, is not likely to be an extreme case among all the estimated $g(\cdot)$ functions based on the permuted data sets. We plot the pointwise upper 95% quantile curve based on 1000 permuted data sets in Figure 7. It is clear that there is one region where the estimated curve $\widehat{g}$ is above the 95% quantile curve. Therefore, in this region, we have a significantly positive treatment effect using BUP.

Finally, we compare the optimal treatment assignment with the real treatment assignment in this experiment. Under the estimated optimal ITR, we assign the participants according to their corresponding estimation of $g(\cdot)$, i.e. the ith participant is assigned to the treatment BUP if $\widehat{g}(\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) > 0$ and is assigned to SER if $\widehat{g}(\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) \leq 0$. Then we classified the patients according to the optimal treatment assignments and the real treatment assignments. The results are displayed in Table 11. We can see that a total of 179 participants were not assigned to their corresponding optimal treatment group. Therefore, the proposed method could have been used for improving the patient satisfaction in this example.

Table 11
Real treatment assignments versus the optimal treatment assignments.

	Optimal Assignment	
Real Assignment	BUP	SER
BUP	80	73
SER	106	60

8. Conclusion

In our work, we have assumed the regression error ϵ to be normal for the simplicity. When the regression error is not normal, our estimation procedure can be viewed as least square based instead of likelihood based, hence all the estimators are still consistent. However, the efficiency statements derived in Section 3 do not hold anymore.

Instead of using $f^*(\mathbf{x}) = 0$, one can certainly strive to get a better assessment of $f(\mathbf{x})$ and use it as $f^*(\mathbf{x})$, in the hope of gaining efficiency. To this end, based on the fact that $f(\mathbf{x}) = E(Y \mid Z = 0, \mathbf{x}) = f(\mathbf{x})$ and the subset of data with $Z_i = 0$, we can use linear model, additive model, single index model, or in fact any parametric or semipametric model that is estimable to get $f^*(\mathbf{x})$ as an approximation to $f(\mathbf{x})$. If the adopted model is correct or close to be correct, the resulting $f^*(\mathbf{x})$ may be close to $f(\mathbf{x})$ and may subsequently lead to improved efficiency. Similarly, one can choose to use any nonparametric method to estimate the working model for $g'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. For example, we used the local linear kernel method to obtain the working model $h^*(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ to estimate $\boldsymbol{\beta}_L$ by solving (7). We denote this estimated working model in (7) by $\widehat{\alpha}_1(\boldsymbol{\beta}, i)$.

We also note that in the derivations of the proposed estimator, we did not use the fact that $Z^2 = Z$ in the binary data. In fact, the estimation procedure, asymptotic theory and implementation details are directly applicable even when Z is categorical or continuous. A categorical Z corresponds to the case of comparing multiple treatments, while a continuous Z could be used when a range of dosage levels are under examination.

In many clinical studies, the covariate is often of very high dimension. To develop optimal individualized treatment rules in this case, it will be important to develop simultaneous variable selection and estimation of individualized rules. It is also of great interest to extend the current approach to multi-stage randomized clinical studies.

Acknowledgments

This research was supported by the National Science Foundation and the National Institute of Health. The authors would also like to thank the associate editor and reviewers for their helpful comments and suggestions.

Supplementary Material

Supplement to 'Flexible Inference of Optimal Individualized Treatment Strategy in Covariate Adjusted Randomization with Multiple Covariates'
(doi: 10.1214/23-EJS2127SUPP; .pdf). Supplementary Material contains the derivation of efficient score functions of our proposed methods and the proofs of Theorem 3.1–3.4.

References

[1] Blatt, D., Murphy, S. A. and Zhu, J. (2004). A-learning for approximate planning. *Ann Arbor* **1001** 48109–2122.

[2] Chakraborty, B., Laber, E. B. and Zhao, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics* **69** 714–723. MR3106599

[3] Chakraborty, B., Laber, E. B. and Zhao, Y.-Q. (2014). Inference about the expected performance of a data-driven dynamic treatment regime. *Clinical Trials* **11** 408–417.

[4] Chakraborty, B., Murphy, S. and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical methods in medical research* **19** 317–343. MR2757118

[5] Ertefaie, A., McKay, J. R., Oslin, D. and Strawderman, R. L. (2021). Robust Q-learning. *Journal of the American Statistical Association* **116** 368–381. MR4227700

[6] Fan, A., Lu, W. and Song, R. (2016). Sequential advantage selection for optimal treatment regime. *The annals of applied statistics* **10** 32. MR3480486

[7] Fan, C., Lu, W., Song, R. and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society. Series B, Statistical methodology* **79** 1565. MR3731676

[8] Ghosh, T., Ma, Y., Song, R. and Zhong , Y. (2023). Supplement to "Flexible inference of optimal individualized treatment strategy in covariate adjusted randomization with multiple covariates" DOI: 10.1214/23-EJS2127SUPP.

[9] Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *Annals of statistics* **40** 529. MR3014316

[10] Henderson, R., Ansell, P. and Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics* **66** 1192–1201. MR2758507

[11] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of econometrics* **58** 71–120. MR1230981

[12] Jameson, J. L. and Longo, D. L. (2015). Precision Medicine? Personalized, Problematic, and Promising. *New England journal of medicine* **372** 2229–2234.

[13] Jeng, X. J., Lu, W. and Peng, H. (2018). High-dimensional inference for personalized treatment decision. *Electronic journal of statistics* **12** 2074. MR3816967

[14] Kang, S., Lu, W. and Zhang, J. (2018). On estimation of the optimal treatment regime with the additive hazards model. *Statistica Sinica* **28** 1539. MR3821017

[15] Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E. and Murphy, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electronic journal of statistics* **8** 1225. MR3263118

[16] Liang, M. and Yu, M. (2020). A semiparametric approach to model effect modification. *Journal of the American Statistical Association* 1–13. MR4436310

[17] Liang, S., Lu, W. and Song, R. (2018). Deep advantage learning for optimal dynamic treatment regime. *Statistical theory and related fields* **2** 80–88. MR3941700

[18] Lunceford, J., Davidian, M. and Tsiatis, A. (2002). Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics* **58** 48–57. MR1891042

[19] Moodie, E. E. M., Platt, R. W. and Kramer, M. S. (2009). Estimating Response-Maximized Decision Rules With Applications to Breastfeeding. *Journal of the American Statistical Association* **104** 155–165. MR2663039

[20] Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65** 331–355.

MR1983752

[21] Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Statistics in medicine* **24** 1455–1481. MR2137651

[22] Nie, X. and Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133

[23] Orellana, L., Rotnitzky, A. and Robins, J. M. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: main content. *The international journal of biostatistics* **6**. MR2602551

[24] Qi, Z. and Liu, Y. (2018). D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics* **12** 3601–3638. MR3870507

[25] Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of statistics* **39** 1180. MR2816351

[26] Qian, M., Nahum-Shani, I. and Murphy, S. A. (2013). Dynamic Treatment Regimes. In *Modern Clinical Trial Analysis* 127–148. Springer.

[27] Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the second seattle Symposium in Biostatistics* 189–326. Springer. MR2129402

[28] Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society* 931–954. MR0951762

[29] Shi, C., Fan, A., Song, R. and Lu, W. (2018). High-dimensional A-learning for optimal dynamic treatment regimes. *Annals of statistics* **46** 925. MR3797992

[30] Shi, C., Song, R. and Lu, W. (2021). Concordance and value information criteria for optimal treatment decision. *The Annals of Statistics* **49** 49–75. MR4206669

[31] Song, R., Luo, S., Zeng, D., Zhang, H. H., Lu, W. and Li, Z. (2017). Semiparametric single-index model for estimating optimal individualized treatment strategy. *Electronic Journal of Statistics* **11** 364–384. MR3608677

[32] Song, R., Wang, W., Zeng, D. and Kosorok, M. R. (2015). Penalized q-learning for dynamic treatment regimens. *Statistica Sinica* **25** 901. MR3409730

[33] Thall, P., Millikan, R. and Sung, H. (2000). Evaluating multiple treatment courses in clinical trials. *Statistics In Medicine* **19** 1011–1028.

[34] Thall, P., Sung, H. and Estey, E. (2002). Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association* **97** 29–39. MR1947271

[35] Thall, P. F., Wooten, L. H., Logothetis, C. J., Millikan, R. E. and Tannir, N. M. (2007). Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Statistics In Medicine* **26** 4687–4702. MR2413392

[36] Tian, L., Alizadeh, A. A., Gentles, A. J. and Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*

109 1517–1532. MR3293607

[37] Wahed, A. and Tsiatis, A. (2006). Semiparametric efficient estimation of survival distributions in two-stage randomisation designs in clinical trials with censored data. *Biometrika* **93** 163–177. MR2277748

[38] Wahed, A. S. and Tsiatis, A. A. (2004). Optimal estimator forthe survival distribution and relatedquantities for treatment policies in two-stagerandomised designs in clinical trials. *Biometrics* **60** 124–33. MR2043627

[39] Watkins, C. J. (1989). Learning from delayed rewards, PhD thesis, University of Cambridge, England.

[40] Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine Learning* **8** 279–292.

[41] Xiao, W., Zhang, H. H. and Lu, W. (2019). Robust regression for optimal individualized treatment rules. *Statistics in medicine* **38** 2059–2073. MR3937486

[42] Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M. and Laber, E. (2012). Estimating optimal treatment regimes from a classification perspective. *Stat* **1** 103–114. MR4027418

[43] Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2012). A Robust Method for Estimating Optimal Treatment Regimes. *Biometrics* **68** 1010–1018. MR3040007

[44] Zhang, Y., Laber, E. B., Davidian, M. and Tsiatis, A. A. (2018). Interpretable dynamic treatment regimes. *Journal of the American Statistical Association* **113** 1541–1549. MR3902228

[45] Zhang, Y., Laber, E. B., Tsiatis, A. and Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics* **71** 895–904. MR3436715

[46] Zhao, Y., Kosorok, M. R. and Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in medicine* **28** 3294–3315. MR2750277

[47] Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012). Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association* **107** 1106–1118. MR3010898

[48] 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. MR2872393

[49] Zhou, J., Zhang, J., Lu, W. and Li, X. (2019). On restricted optimal treatment regime estimation for competing risks data. *Biostatistics.* MR4246886

[50] Zhu, L., Lu, W., Kosorok, M. R. and Song, R. (2020). Kernel Assisted Learning for Personalized Dose Finding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'20* 56–65. Association for Computing Machinery, New York, NY, USA.