ABSTRACT

SHI, CHENGCHUN. On Statistical Learning for Individualized Decision Making with Complex Data. (Under the direction of Wenbin Lu and Rui Song).

The motivation behind my dissertation research stems from real world applications. In precision medicine, individualizing the treatment decision rule can capture patients' heterogeneous response towards treatment (Piquette-Miller and Grant 2007). In finance, individualizing the investment decision rule can improve individual's financial well-being (Ding et al. 2019). In a ride-sharing company, individualizing the order dispatching strategy can increase its revenue and customer satisfaction (Xu et al. 2018). With the fast development of new technology, modern datasets often consist of massive observations, high-dimensional covariates and are characterized by some degree of heterogeneity. For example, medical studies are likely to obtain a large number of prognostic factors for each patient. In a ride-sharing company such as Didi Chuxing (one of the world's leading ride-sharing platforms), its app generates more than 70TB of data every day.

In this thesis, we mainly focus on the application of precision medicine and propose new statistical learning methods for estimation and inference of the optimal treatment regime (OTR) with modern complex datasets. This thesis is structured as follows. In Chapter 2-4, we propose methods to estimate OTR for these complex datasets. Specifically, in Chapter 2, we propose a maximin-projection learning method for recommending a reliable treatment regime based on the observed data from different populations with heterogeneity in optimal decision making. The proposed treatment regime has good statistical interpretation in terms of maximizing the minimum groupwise average percentage of making correct decisions and value difference function. It also has a nice geometrical characterization using the notion of optimal equicorrelated point. Moreover, the estimating procedure can be efficiently implemented using quadratic programming. To handle massive data, we propose a divide and conquer method to estimate the OTR in Chapter 3. The proposed estimator achieves a faster convergence rate and is asymptotically normal, which is more tractable in both computation and inference than the original value-search estimator based on the pooled data. In Chapter 4, we propose a penalized A-learning method to estimate the optimal dynamic treatment regime (ODTR) with high-dimensional covariates. The resulting estimator preserves the doubly robustness property of the classical A-learning method even when the number of covariates is much larger than the sample size. We further introduce two doubly robust information criteria (the concordance and value information

criteria) for tuning parameter selection in penalized A-learning.

Despite the popularity of estimating the O(D)TR in the literature, less attention has been devoted to statistical inference regarding the O(D)TR. This is the focus of Chapter 5-7. In Chapter 5, we propose to test if implementing the OTR is equivalent to the "one-size-fits-all" method. The test is constructed based on sparse random projections of high-dimensional covariates into a low-dimensional space. We show the power function of our test is asymptotically the same as the "oracle" test that is constructed based on the optimal projection matrix. In Chapter 6, we further develop a test for assessing the incremental value of a set of new variables in treatment decision making conditional on an existing set of variables.

Prior to adopting any O(D)TR, it is crucial to know the impact of implementing such a treatment regime. However, evaluating the mean outcome in the population under an O(D)TR (the optimal value) is a nonregular inference problem and has been shown to be difficult (Robins 2004). In Chapter 7, we propose to construct the confidence interval for the optimal value, based on subsample aggregating and refitted cross validation. Subagging has been recognized as an effective variance reduction technique in hard decision problems (Bühlmann and Yu 2002). However, it remains unknown whether such a procedure could yield valid inference results. We prove that the proposed CI achieves nominal coverage. In addition, due to the variance reduction property of subagging, our method has certain statistical "optimality". Specifically, We show both numerically and theoretically that our method achieves better performance when compared to the existing state-of-the-art method (Luedtke and van der Laan 2016) and the "oracle" method which works as if the optimal IDR were known in advance.

On Statistical Learning for Individualized Decision Making with Complex Data

by
Chengchun Shi

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina
2019

APPROVED BY:

Wenbin Lu Co-chair of Advisory Committee	Rui Song Co-chair of Advisory Committee
Anastasios A. Tsiatis	Soumendra N. Lahiri

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisors, Dr. Wenbin Lu and Dr. Rui Song. They mentored me for approximately five years and managed to meet with me almost every week in these five years. They raised me up, so I could stand on mountains. They raised me up to walk on stormy seas. Up to date, we have jointly written 14 manuscripts. They constantly encouraged me to pursue my career in academia and helped me a lot especially in my first two years when I did not know how to conduct statistical research at all and this year when I was on the job market looking for a faculty position in statistics/data science. I was very fortunate to work with them during my PhD. Without their support, I cannot achieve what I have done so far.

During my first year, I was very fortunate to take the special topic course “causal inference" taught by Dr. Anastasios Tsiatis. The notion of “potential outcomes" and the philosophy behind the causal inference framework completely changed my thoughts on investigating causal relationships. The course stimulated my interest in estimation and inference of optimal treatment regime, which is the focus of this thesis. In my second year, I was lucky to take the “advanced probability theory" course taught by Dr. Soumendra Lahiri. This course helped me build a solid mathematical foundation. I was very fortunate to have them as my committee members.

It was a great pleasure for me to work with Dr. Runze Li, Dr. Hongtu Zhu and Dr. Bo Fu. Their insightful comments and suggestions enormously helped me in research. I also would like to express my thanks to Dr. Dan Harris, Dr. Justin Post, Dr. Jessie Jeng and Dr. Jung-Ying Tzeng. I worked with them as a teaching assistant and learned a lot from them.

I felt so lucky to have all the talented friends, classmates, officemates and roommates in the past five years. Hengfeng and Rui, my best friends, drove me to school and grocery stores when I didn’t own a car. Brian, went to the gym together with me since his first PhD year. Wenhao, taught me to drive several times. Hao Hu, helped me fill out my tax return forms for consecutive two years. I always enjoyed travelling and playing games with my friends, Qian, Lin Dong, Zhengtao, Lin Su, Lin Yu, Haoyu Wang, Maggie, Wanying, Wei Zhao, Wei Xiao, Wen, Liuyi, Marshall, Anran, Lantian, Yuantong, Yi, Yue, Mengxing, Yan, Ailin, Jiapeng and Jiabin. All of them supported me a lot during my PhD as well. Thank Norb, Marion, Dave and Fernandez for constantly inviting me to their party. Thank Jocelyn Chi for polishing my teaching and research statements. I also would like to express my gratitude to those students in our reading groups, especially Ye, Haoyu Chen, Sheng, Dana, Kevin, Hengrui, Runzhe and Miao who have presented several research papers and book chapters that broadened my horizon.

Finally, I would like to thank my family for their unconditional love. Both my parents have neither respectable employments nor higher education background, even a high school diploma. But they virtually support every decision I have made, even those they could hardly financially support. And they have always been progressive, leaving me to draw up my blueprint independently. Never have them forced me to do something I do not like. Neither is my fiancee, Yuan. She always gives me the strongest support, no matter what happens. I really owe a lot to them, and all I could do in return is to be the best of what I am on the road I have chosen myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

In precision medicine, individualizing the treatment decision rule can capture patients' heterogeneous response towards treatment (Piquette-Miller and Grant 2007). In finance, individualizing the investment decision rule can improve individual's financial well-being (Ding et al. 2019). In a ride-sharing company, individualizing the order dispatching strategy can increase its revenue and customer satisfaction (Xu et al. 2018).

In this thesis, we consider the problem of individualized decision making with a particular focus on the application of precision medicine, a medical paradigm that receives considerable scientific and commercial attention. The goal of precision medicine is to assign the optimal treatment for each individual patient, according to his/her personal information, such as a patient's genetic content, clinical response and demographic characteristics, etc. A treatment regime (TR) is a decision rule that assigns treatments to patients based on their observed covariates. Among the set of all possible treatment regimes, the one that optimize patients' expected outcomes of interest is referred to as the optimal TR (OTR). For chronic diseases such as cancer and diabetes, treatment of patients involves a series of decisions. In these applications, it is of considerable interest to estimate the optimal dynamic TR (ODTR) that consists of a list of decision rules for assigning treatment based on a patient's covariates and treatment history.

In the literature, various methods have been proposed to estimating the OTR (or ODTR), including Q-learning (Watkins and Dayan 1992; Chakraborty et al. 2010; Song et al. 2015) and A-learning (Robins et al. 2000; Murphy 2003). Both Q-learning and A-learning rely on a backward induction algorithm to find the optimal dynamic treatment regime, however, Q-learning models the conditional mean of the outcome given predictors and treatment while A-learning directly models the contrast function that is sufficient for treatment decision. In particular, A-learning has the so-called doubly robust property, i.e. when either the baseline mean function or the propensity score model is correctly specified, the resulting A-learning estimating equation for the contrast function is consistent. Recently, Zhang et al. (2012, 2013) proposed to estimate the O(D)TR by directly maximizing the estimated expected outcome, i.e, the value function. Fan et al. (2017) introduced a type of concordance function for prescribing treatment and proposed a concordance-assisted learning for estimating the O(D)TR. Some other popular methods include outcome weighted learning (Zhao et al. 2012, 2015), tree-based methods (Laber and Zhao 2015; Zhu et al. 2017) and decision list-based methods (Zhang et al. 2015, 2018).

1.1 Motivation: the need for analyzing complex data

With the fast development of new technology, modern datasets often consist of massive observations, high-dimensional covariates and are characterized by some degree of heterogeneity. In a ride-sharing company such as Didi Chuxing (one of the world's leading ride-sharing platforms), its app generates more than 70TB of data every day. In a world of explosively large data, effective estimation procedures are needed to deal with the computational challenge arisen from analysis of massive data.

In medical studies, it is possible to gather an extraordinary large number of prognostic factors for each individual, such as patient's genetic information, demographic characteristics, medical history and clinical measurements over time. For example, in the Sequenced Treatment Alternative to Relieve Depression (STAR*D) study, over 305 covariates are collected from each patient. For such high-dimensional data, it is important to make effective use of information that is relevant to make OTR. This makes variable selection as an emerging need for implementing precision medicine. Although there is a large amount of work on developing variable selection methods for prediction (see discussions in Fan and Lv 2010), variable selection tools for deriving OTR have been less studied, especially when the number of predictors is much larger than the sample size.

Moreover, the OTRs may also vary for patients from different subpopulations. This is typically the case in meta analysis, where we combine the results of multiple studies conducted at different locations or times. One motivating example is from a multi-centre randomised controlled trial as studied in Tarrier et al. (2004). The goal is to examine the effectiveness of cognitive-behavioural therapy for patients with early schizophrenia. Patients can be classified into three groups according to their treatment centres (Manchester, Liverpool and North Nottinghamshire). As we can see in Section 2.6.2, the group-wise OTRs can vary across different centres. Another example is from an observational study for investigating the influence of early disease modifying antirheumatic drug (DMARD) treatment on patients with recent onset inflammatory polyarthritis (Farragher et al. 2010). According to patients' enrollment time, they can be classified into three groups. As studied in Section 2.6.1, the group-wise OTRs can vary across different enrollment periods. The heterogeneity in OTRs may be explained by the differences in characteristics of treatment setting across subgroups. For instance, in the schizophrenia example, the strength of therapeutic alliance between therapist and patient, the adherence to treatment protocols and the quality of treatment provided can vary from one treatment centre to another (Dunn and Bentall 2007); in the inflammatory polyarthritis example, there are more use of hydroxychloroquine for the methotrexate combination strategy in recruitment time group 3 (1997-2000) than in group 1 (1990-1992) or group 2 (1993-1996) as hydroxychloroquine was increasingly used in the UK before anti-tumour necrosis factor therapy was introduced to treat rheumatoid arthritis in 2001. Moreover, these characteristics are often unobserved or partially observed, and they may explain the interaction between subgroups and OTRs.

These applications motivate us to develop statistical learning methods with complex datasets. Before presenting our methodologies, we introduce a causal framework (Rubin 2005) to formulate our problem. In Section 1.2, we introduce the model setup and formally define the OTR in a point treatment study. In Section 1.3, we focus on multiple time point studies and define the corresponding ODTR.

1.2 Point treatment study

We begin by considering a single stage study with two treatments. Let $\boldsymbol{X}_0 \in \mathbb{X}$ be a patient's baseline covariates, $A_0 \in \{0,1\}$ denote the treatment a patient receives, and Y_0 denote a patient's clinical outcome (the larger the better by convention). A treatment regime $d(\cdot)$ is a deterministic function that maps $\mathbb{X}$ to $\{0,1\}$. Let $Y_0^*(0)$ and $Y_0^*(1)$ be a patient's potential outcomes, representing the response he/she would get if treated by treatment 0 and 1, respectively. In addition, define the potential outcome

$$Y_0^*(d) = Y_0^*(0)\{1 - d(\boldsymbol{X}_0)\} + Y_0^*(1)d(\boldsymbol{X}_0),$$

representing the response a patient would have if treated according to a TR d. Let $V(d) = \mathbb{E}\{Y_0^*(d)\}$ be the value function under a TR d. An OTR d^{opt} is defined as the maximizer of the expected potential outcome $V(d)$ among the set of all possible treatment regimes, i.e,

$$d^{opt} \equiv \arg\max_d V(d).$$

However, the OTR may not be unique. Let $\mathscr{D}^{opt}$ denote the set of all OTRs, i.e,

$$\mathscr{D}^{opt} = \{d_0 : V(d_0) = \max_d V(d)\}.$$

Assume the following three assumptions hold.

(C1.) SUTVA: $Y_0 = (1 - A_0)Y_0^*(0) + A_0 Y_0^*(1)$.

(C2.) No unmeasured confounders: $Y_0^*(0), Y_0^*(1) \perp A_0|\boldsymbol{X}_0$.

(C3.) Positivity: $\pi(a, \boldsymbol{x}) \geq \epsilon_0, \forall a \in \{0, 1\}, \boldsymbol{x} \in \mathbb{X}$ for some constant $0 < \epsilon_0 < 1$, where $\pi(a, \boldsymbol{x}) = \mathbb{P}(A_0 = a|\boldsymbol{X}_0 = \boldsymbol{x})$ denotes the propensity score function that characterizes the treatment assignment mechanism. We may sometimes use a shorthand and write $\pi(1, \boldsymbol{x})$ as $\pi(\boldsymbol{x})$.

Define the contrast function

$$\tau(\boldsymbol{x}) \equiv h(1, \boldsymbol{x}) - h(0, \boldsymbol{x}),$$

where $h(a, \boldsymbol{x}) = \mathbb{E}(Y_0|A_0 = a, \boldsymbol{X}_0 = \boldsymbol{x})$, for $a \in \{0, 1\}$. We may sometimes write $h(0, \boldsymbol{x})$ as $h(\boldsymbol{x})$. The following lemma relates OTR to the function $\tau(\cdot)$.

Lemma 1.2.1 *Let* $\mathbb{X}_1 = \{\boldsymbol{x} \in \mathbb{X} : \tau(\boldsymbol{x}) > 0\}$ *and* $\mathbb{X}_2 = \{\boldsymbol{x} \in \mathbb{X} : \tau(\boldsymbol{x}) < 0\}$. *Assume C1-C3 hold, and* $\mathbb{E}|\tau(\boldsymbol{X}_0)| < \infty$. *Then, for any* $d \in \mathscr{D}^{opt}$, *we have*

$$\mathbb{P}(\boldsymbol{X}_0 \in \mathbb{X}_1 \cap \mathbb{X}_{2,d}) = 0 \ \ \textit{and} \ \ \mathbb{P}(\boldsymbol{X}_0 \in \mathbb{X}_2 \cap \mathbb{X}_{1,d}) = 0, \tag{1.1}$$

where $\mathbb{X}_{1,d} = \{\boldsymbol{x} \in \mathbb{X} : d(\boldsymbol{x}) = 1\}$ *and* $\mathbb{X}_{2,d} = \{\boldsymbol{x} \in \mathbb{X} : d(\boldsymbol{x}) = 0\}$. *Conversely, for any treatment regime* d *satisfying* (1.1), *we have* $d \in \mathscr{D}^{opt}$.

Lemma 1.2.1 implies that $d^{opt,0} \in \mathscr{D}^{opt}$ where

$$d^{opt,0}(\boldsymbol{x}) = \mathbb{I}\{\tau(\boldsymbol{x}) > 0\}, \quad \forall \boldsymbol{x} \in \mathbb{X}, \tag{1.2}$$

where $\mathbb{I}(\cdot)$ stands for the indicator function. The proof of Lemma 1.2.1 can be found in Shi et al. (2019d).

The objective is to estimate an OTR $d^{opt} \in \mathscr{D}^{opt}$ based on the observed data $\{(\boldsymbol{X}_i, A_i, Y_i)\}_i$ that are assumed to be i.i.d copies of $(\boldsymbol{X}_0, A_0, Y_0)$. In view of (1.2), one way to derive the OTR is to estimate $\tau(\cdot)$. Q-learning and A-learning fall into this category. Denoted by $\widehat{\tau}(\cdot)$ the estimated contrast function, the estimated OTR can is given by $\mathbb{I}\{\widehat{\tau}(\cdot) > 0\}$. Value-search methods, outcome weighted learning, concordance-assisted learning, tree-based methods and decision list-based methods, on the other hand, directly search d^{opt} among a restricted class of TRs that maximizes some estimated value function.

1.3 Multiple time point study

Consider a multistage study where the treatment decisions are made at a finite number of time points $t_1, \ldots, t_K$. The data for a subject can be summarized as

$$(\boldsymbol{X}_0^{(1)}, A_0^{(1)}, \boldsymbol{X}_0^{(2)}, A_0^{(2)}, \ldots, \boldsymbol{X}_0^{(K)}, A_0^{(K)}, Y_0),$$

where Y_0 denotes the outcome of interest, $\boldsymbol{X}_0^{(1)}$ stands for the set of covariates obtained prior to the time point t_1, $A_0^{(1)}$ denotes the treatment received at t_1. For $k = 2, \ldots, K$, $\boldsymbol{X}_0^{(k)}$ denotes some additional covariates collected between time points t_{k-1} and t_k, and $A_0^{(k)}$ denotes the treatment given at t_k. For simplicity, we assume $A_0^{(1)}, \ldots, A_0^{(K)}$ are all binary treatments. For $k = 1, \ldots, K$, let

$$\bar{\boldsymbol{X}}_0^{(k)} = (\boldsymbol{X}_0^{(1)}, \ldots, \boldsymbol{X}_0^{(k)}) \in \bar{\mathbb{X}}^{(k)} \quad \text{and} \quad \bar{\boldsymbol{A}}_0^{(k)} = (A_0^{(1)}, \ldots, A_0^{(k)}) \in \{0,1\}^k,$$

denote a patient's covariates and treatment history. For any $a_1, \ldots, a_K \in \{0,1\}$, denoted by $\bar{\boldsymbol{a}}_k = (a_1, \ldots, a_k)$ for $k = 1, \ldots, K$. The set of all potential outcomes is given by

$$\boldsymbol{W} = \left\{ \left(\boldsymbol{X}_0^{(2)*}(a_1), \boldsymbol{X}_0^{(3)*}(\bar{\boldsymbol{a}}_2), \ldots, \boldsymbol{X}_0^{(K)*}(\bar{\boldsymbol{a}}_{K-1}), Y_0^*(\bar{\boldsymbol{a}}_K) \right) : \forall \bar{\boldsymbol{a}}_K \in \{0,1\}^K \right\}, \tag{1.3}$$

where $\boldsymbol{X}_0^{(k)*}(\bar{\boldsymbol{a}}_{k-1})$ denotes the potential time-dependent covariates of a patient that would occur between t_{k-1} and t_k assuming he/she receives treatments $(a_1, \ldots, a_{k-1})$ at decision points $(t_1, \ldots, t_{k-1})$ and $Y_0^*(\bar{\boldsymbol{a}}_K)$ denotes the potential outcome that would result assuming he/she receives treatments $(a_1, \ldots, a_K)$.

A dynamic treatment regime $d = \{d_k\}_{k=1}^K$ is a set of decision rules that treats a patient over time. For $k = 1, \ldots, K$, $d_k = d_k(\bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{x}}_k)$ corresponds to the k-th decision rule that takes as input a patient's realized covariate and treatment history and outputs a treatment option $a_k \in \{0,1\}$. Let $\bar{d}_k = \{d_j\}_{j=1}^k$ for $k = 1, \ldots, K-1$, the potential outcome associated with d is given by

$$\left(\boldsymbol{X}_0^{(2)*}(d_1), \boldsymbol{X}_0^{(3)*}(\bar{d}_2), \ldots, \boldsymbol{X}_0^{(K)*}(\bar{d}_{K-1}), Y_0^*(d)\right),$$

where $\boldsymbol{X}_0^{(k)*}(\bar{d}_{k-1})$ stands for the potential covariates of a patient between t_{k-1} and t_k assuming he/she receives the treatments sequentially according to the decision rules $(d_1, \ldots, d_{k-1})$ and $Y_0^*(d)$ stands for the potential outcome assuming the treatments he/she receives are determined by the treatment regime d. An optimal dynamic treatment regime d^{opt} is defined to maximize the average potential outcome, i.e,

$$d^{opt} = \arg\max_d \mathbb{E}\, Y_0^*(d) = \arg\max_d V(d).$$

For any $\bar{\boldsymbol{a}}_K \in \{0,1\}^K$ and $\bar{\boldsymbol{x}}_K \in \bar{\mathbb{X}}^{(K)}$, let $h_K(\bar{\boldsymbol{a}}_K, \bar{\boldsymbol{x}}_K) = \mathbb{E}(Y_0 | \bar{\boldsymbol{X}}_0^{(K)} = \bar{\boldsymbol{x}}_K, \bar{\boldsymbol{A}}_0^{(K)} = \bar{\boldsymbol{a}}_K)$ and $\tau_K(\bar{\boldsymbol{a}}_{K-1}, \bar{\boldsymbol{x}}_K) = h_K\{(\bar{\boldsymbol{a}}_{K-1}, 1), \bar{\boldsymbol{x}}_K\} - h_K\{(\bar{\boldsymbol{a}}_{K-1}, 0), \bar{\boldsymbol{x}}_K\}$. In addition, for $k = K-1, \cdots, 2$, we sequentially define

$$h_k(\bar{\boldsymbol{a}}_k, \bar{\boldsymbol{x}}_k) = \mathbb{E}\left(\arg\max_{a_{k+1} \in \{0,1\}} h_{k+1}\{(\bar{\boldsymbol{a}}_k, a_{k+1}), \bar{\boldsymbol{X}}_0^{(k+1)}\} \,\middle|\, \bar{\boldsymbol{X}}_0^{(k)} = \bar{\boldsymbol{x}}_k, \bar{\boldsymbol{A}}_0^{(k)} = \bar{\boldsymbol{a}}_k\right),$$

and $\tau_k(\bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{x}}_k) = h_k\{(\bar{\boldsymbol{a}}_{k-1}, 1), \bar{\boldsymbol{x}}_k\} - h_k\{(\bar{\boldsymbol{a}}_{k-1}, 0), \bar{\boldsymbol{x}}_k\}$, for any $\bar{\boldsymbol{a}}_k \in \{0,1\}^k$ and $\bar{\boldsymbol{x}}_k \in \bar{\mathbb{X}}^{(k)}$. For $k = 1$, let

$$h_1(a_1, \boldsymbol{x}_1) = \mathbb{E}\left(\arg\max_{a_2 \in \{0,1\}} h_2\{(a_1, a_2), \bar{\boldsymbol{X}}_0^{(2)}\} \,\middle|\, \boldsymbol{X}_0^{(1)} = \boldsymbol{x}_1, A_0^{(1)} = a_1\right)$$

and $\tau_1(\boldsymbol{x}_1) = h_1(1, \boldsymbol{x}_1) - h_1(0, \boldsymbol{x}_1)$ for any $a_1 \in \{0,1\}, \boldsymbol{x}_1 \in \bar{\mathbb{X}}_1$. Define the propensity score function $\pi_k(\bar{\boldsymbol{a}}_k, \bar{\boldsymbol{x}}_k) = \mathbb{P}(A_0^{(k)} = a_k | \bar{\boldsymbol{X}}_0^{(k)} = \bar{\boldsymbol{x}}_k, \bar{\boldsymbol{A}}_0^{(k-1)} = \bar{\boldsymbol{a}}_{k-1})$ for $k = 2, \ldots, K$ and $\pi_1(a_1, \boldsymbol{x}_1) = \mathbb{P}(A_0^{(1)} = a_1 | \boldsymbol{X}_0^{(1)} = \boldsymbol{x}_1)$. Under the following three conditions,

(MC1.) $\boldsymbol{X}_0^{(k)} = \sum_{\bar{\boldsymbol{a}}_{k-1} \in \{0,1\}^{k-1}} \boldsymbol{X}_0^{(k)*}(\bar{\boldsymbol{a}}_{k-1}) \mathbb{I}(\bar{\boldsymbol{A}}_0^{(k-1)} = \bar{\boldsymbol{a}}_{k-1})$ and
$Y_0 = \sum_{\bar{\boldsymbol{a}}_K \in \{0,1\}^K} Y_0^*(\bar{\boldsymbol{a}}_K) \mathbb{I}(\bar{\boldsymbol{A}}_0^{(K)} = \bar{\boldsymbol{a}}_K)$, $\forall k = 2, \ldots, K$ and $\bar{\boldsymbol{a}}_K \in \{0,1\}$,
(MC2.) $A_0^{(k)} \perp \boldsymbol{W} | \bar{\boldsymbol{X}}_0^{(k)}, \bar{\boldsymbol{A}}_0^{(k-1)}, \forall k = 1, \ldots, K$ where $\boldsymbol{W}$ is defined in (1.3),

(MC3.) $\pi_k(\bar{a}_k, \bar{\boldsymbol{x}}_k) \geq \epsilon_0, \forall \bar{\boldsymbol{a}}_k \in \{0,1\}^k, \bar{\boldsymbol{x}}_k \in \bar{\mathbb{X}}_k, k \in \{1,\ldots,K\}$,
we can show

$$h(\bar{\boldsymbol{a}}_K, \bar{\boldsymbol{x}}_K) = \mathbb{E}\{Y_0^*(\bar{\boldsymbol{a}}_K) | \bar{\boldsymbol{X}}_0^{(K)*}(\bar{\boldsymbol{a}}_{K-1}) = \bar{\boldsymbol{x}}_K\}, \tag{1.4}$$

and for $2 \leq k \leq K-1$,

$$h(\bar{\boldsymbol{a}}_k, \bar{\boldsymbol{x}}_k) = \mathbb{E}[V_0^{(k+1)}\{\bar{\boldsymbol{a}}_k, \bar{\boldsymbol{X}}_0^{(k+1)*}(\bar{\boldsymbol{a}}_k)\} | \bar{\boldsymbol{X}}_0^{(k)*}(\bar{\boldsymbol{a}}_{k-1}) = \bar{\boldsymbol{x}}_k], \tag{1.5}$$

and

$$h(a_1, \boldsymbol{x}_1) = \mathbb{E}[V_0^{(2)}\{a_1, \bar{\boldsymbol{X}}_0^{(2)*}(a_1)\} | \boldsymbol{X}_0^{(1)} = \boldsymbol{x}_1], \tag{1.6}$$

where

$$\begin{aligned}
&V_0^{(K)}(\bar{\boldsymbol{a}}_{K-1}, \bar{\boldsymbol{x}}_K) = \max_{a_K \in \{0,1\}} \mathbb{E}\{Y_0^*(\bar{\boldsymbol{a}}_K) | \bar{\boldsymbol{X}}_0^{(K)*}(\bar{\boldsymbol{a}}_{K-1}) = \bar{\boldsymbol{x}}_K\}, \\
&V_0^{(k)}(\bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{x}}_k) = \max_{a_k \in \{0,1\}} \mathbb{E}[V_0^{(k+1)}\{\bar{\boldsymbol{a}}_k, \bar{\boldsymbol{X}}_0^{(k+1)*}(\bar{\boldsymbol{a}}_k)\} | \bar{\boldsymbol{X}}_0^{(k)*}(\bar{\boldsymbol{a}}_{k-1}) = \bar{\boldsymbol{x}}_k], \\
&\bar{\boldsymbol{X}}_0^{(k)*}(\bar{\boldsymbol{a}}_{k-1}) = \{\boldsymbol{X}_0^{(1)}, \boldsymbol{X}_0^{(2)*}(a_1), \ldots, \boldsymbol{X}_0^{(k)*}(\bar{\boldsymbol{a}}_{k-1})\}.
\end{aligned}$$

Here, Condition MC2 and MC3 automatically holds in sequentially randomized studies (Murphy 2005).

Define the set of dynamic treatment regimes $\mathscr{D}^{opt}$ such that any $d = \{d_k\}_{k=1}^K \in \mathscr{D}^{opt}$ shall satisfy

$$\begin{aligned}
d_K(\bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{x}}_k) \in \arg\max_{a \in \{0,1\}} a\tau_k(\bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{x}}_k), k = 2, \ldots, K, \\
\text{and} \quad d_1(\boldsymbol{x}_1) \in \arg\max_{a \in \{0,1\}} a\tau_1(\boldsymbol{x}_1),
\end{aligned} \tag{1.7}$$

for any $\bar{\boldsymbol{x}}_K \in \bar{\mathbb{X}}^{(K)}, \ldots, \bar{\boldsymbol{x}}_2 \in \bar{\mathbb{X}}^{(2)}, \boldsymbol{x}_1 \in \bar{\mathbb{X}}^{(1)}$ and $\bar{\boldsymbol{a}}_{K-1} \in \{0,1\}^{K-1}, \ldots, \bar{\boldsymbol{a}}_2 \in \{0,1\}^2, a_1 \in \{0,1\}$. By (1.4)-(1.6) and backward induction, we can show that

$$\mathscr{D}^{opt} \subseteq \arg\max_d \mathbb{E} Y_0^*(d).$$

Notice that the argmax in (1.7) is not unique when $\tau_k(\bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{x}}_k) = 0$ or $\tau_1(\boldsymbol{x}_1) = 0$. Therefore, the optimal dynamic treatment regime may not be unique. Given the observed data $\{(\boldsymbol{X}_i^{(1)}, A_i^{(1)}, \boldsymbol{X}_i^{(2)}, A_i^{(2)}, \ldots, \boldsymbol{X}_i^{(K)}, A_i^{(K)}, Y_i)\}_i$, the goal is to estimate an ODTR that belongs to

$\mathscr{D}^{opt}$.

1.4 Notations

In this section, we introduce some common notations used in this thesis. Throughout this thesis, we use C_0 and $\bar{C}$ to denote some universal constants, whose values may change from place to place. For any arbitrary matrix $\mathbf{\Phi} \in \mathbb{R}^{M\times M}$ and any arbitrary vector $\phi \in \mathbb{R}^M$, the superscript $\mathbf{\Phi}^{(j)}$ is used to denote the jth column of $\mathbf{\Phi}$, $\phi^{(j)}$ the jth element of ϕ. For subsets $J, J' \subset \{1,\ldots,M\}$, let $|J|$ be the cardinality of J, J^c be the complement of J. We denote by ϕ^J the vector in $\mathbb{R}^{|J|}$ that has the same coordinates as ϕ on J, and Φ^J the submatrix formed by columns in J, $\Phi_J^{J'}$ the submatrix formed by rows in J and columns in J'. Let $\|\phi\|_p$ denote the ℓ_p norm of ϕ and $\|\mathbf{\Phi}\|_p$ denote the operator norm corresponding to the p-norm vector. We use $\boldsymbol{I}_M$ to denote an $M \times M$ identity matrix.

For any two random variables Z_1 and Z_2, $Z_1 \overset{d}{=} Z_2$ means Z_1 and Z_2 have the same distribution function. The notations $\overset{d}{\to}$ and $\overset{P}{\to}$ stand for convergence in distribution and convergence in probability, respectively. Let $\|Z\|_{\psi_p}$ be the Orlicz norm for any random variable Z, defined as

$$\|Z\|_{\psi_p} \equiv \inf_{u>0} \left\{ \mathbb{E} \exp\left(\frac{|Y|}{u}\right)^p \leq 2 \right\},$$

for some $p \geq 1$. We use $\Phi(\cdot)$ to denote the cumulative distribution function of a standard normal random variable. In addition, z_α denotes the αth upper quantile of a standard normal distribution.

For any two positive sequences $\{a_n\}$ and $\{b_n\}$, $a_n \gg b_n$ means $\lim_n b_n/a_n = 0$. The notation $a_n \asymp b_n$ means there exists some universal constant $C_0 \geq 1$ that satisfies $C_0^{-1} b_n \leq a_n \leq C_0 b_n$.

1.5 Outline

The rest of the thesis is organized as follows. In Chapter 2-4, we propose methods to estimate O(D)TR with complex data. Specifically, in Chapter 2, we propose a maximin-projection learning method for recommending a reliable treatment regime based on the observed data from different populations with heterogeneity in optimal decision making. To handle massive data, we propose a divide and conquer method to estimate the OTR in Chapter 3.

In Chapter 4, we propose a penalized A-learning method to estimate the ODTR with high-dimensional covariates. In Section 4.10, we also propose two doubly robust information criteria (concordance and value information criteria) for selecting the tuning parameters in penalized A-learning.

Chapter 5-7 are concerned with statistical inference regarding the O(D)TR. In Chapter 5, we propose to test if implementing the OTR is equivalent to the “one-size-fits-all" method. In Chapter 6, we further develop a test for assessing the incremental value of a set of new variables in treatment decision making conditional on an existing set of variables. In Chapter 7, we proposed to construct the confidence interval for the optimal value, based on subsample aggregating and refitted cross validation.

For each chapter, theoretical properties of the proposed methods are thoroughly investigated. Extensive simulation studies are conducted to back up our theoretical findings. Real data examples are presented to illustrate the usefulness of the proposed methods.

CHAPTER

2

MAXIMIN-PROJECTION LEARNING FOR HETEROGENEOUS DATA

2.1 Introduction

Data from clinical trials and medical studies are often characterized by some degree of inhomogeneity. Optimal treatment regimes have been developed to account for patients' heterogeneity in response to treatment. However, the OTRs may also vary for patients from different subpopulations. This is typically the case in meta analysis, where we combine the results of multiple studies conducted at different locations or times (see two real data examples in Section 2.6 for details).

The aim of this chapter is to propose a reliable OTR for new patients based on the observed data from different groups with heterogeneity in optimal treatment decision. The group of new patients may differ from any of the currently observed groups in terms of optimal treatment decision. For example, in the schizophrenia example (see Section 2.6.2), compared with existing data, the group of new patients, who come from a new treatment centre, may have a different OTR because of different strength of therapeutic alliance or different quality of treatment provided in the new treatment centre. Therefore, the true OTR for the group of new patients is not estimable at all based on the observed data, and any of the group-wise OTRs may not be the best choice. The challenge becomes how to derive a meaningful and reliable treatment regime that can take into account the heterogeneity in optimal treatment decision for different groups of patients. One simple approach is to pool the data of different groups together and obtain the “pooled" OTR based on the pooled data. Another method is to first obtain the OTR for each group, and then aggregate the group-wise OTRs in certain ways. Random effects meta-analysis (DerSimonian and Laird 1986) is commonly used to combine subject-specific studies. Using its multivariate extensions (see for example, Jackson et al. 2010; Chen et al. 2012), we can aggregate the groupwise OTRs based on random effects models. The resulting OTR is similar to the “pooled" OTR when we have large numbers of subgroup patients. These OTRs maybe reasonable choices when the OTRs for different groups do not vary much. However, when there is certain degree of heterogeneity in OTRs across different groups as demonstrated in the toy example given in the next section, these OTRs are uniformly worse than the proposed OTR for any of the groups. One possible reason is that these OTRs for different groups may assign the same patient to different treatments and thus their effects are averaged out when pooling the data from different groups.

Bühlmann and Meinshausen (2016) and Meinshausen and Bühlmann (2015) considered a maximin criteria which has a nice characterization in linear models and proposed to use maximin aggregation (magging) to obtain the maximin estimator. Their proposed estimator is shown to be more robust than the pooled estimator in linear regression. The key idea of the maximin criteria is to find an estimator that works the best under the worst-case scenario. In optimal treatment decision, the percentage of making the correct decision (PCD) and value function are two commonly used measures to evaluate the effectiveness of a treatment regime. A natural maximin criteria for optimal treatment decision is to find an OTR that maximizes the minimum PCD or the minimum value function of all groups. Such a maximin OTR is appealing due to its nice interpretation and robustness. However, it is hard to implement in practice due to the following reasons. First, the PCD of a treatment regime is generally not estimable from data since the true OTR is unknown. Second, the empirical estimator of the value function as studied in Zhang et al. (2012) is non-smooth and non-concave, thus the estimation of the associated maximin OTR is not feasible.

In this chapter, we propose a novel maximin-projection learning (MPL) to aggregate linear OTRs across different groups. Specifically, the proposed MPL finds a linear decision rule that maximizes the minimum “inner product" between the vectors of regression parameters in the linear rule and the group-wise linear OTRs. We show that under certain model assumptions, the OTR obtained by MPL maximizes the minimum percentage of making the correct decision and value function of different groups, i.e. achieve the desired maximin properties. In addition, the corresponding estimation procedure can be represented as a linear programming problem with a quadratic constraint (Lee et al. 2016), which can be efficiently solved in $O(Gp^2 + p^3)$ flops. Here G denotes the number of groups and p the dimension of baseline covariates. Consistency and the asymptotic distribution of the corresponding maximin-projection estimators are established. Such kind of asymptotic results are rarely studied in the literature. To derive such asymptotic properties, we establish a necessary and sufficient condition for the existence and uniqueness of the population maximin-projection parameters and obtain a closed-form expression for the resulting estimator.

The rest of the chapter is organized as follows. In Section 2.2, we introduce the model setup, and provide a heuristic comparison between the maximin OTR, the pooled OTR and the OTR based on random effects models with a toy example. In Section 2.3, we formally introduce the proposed maximin-projection learning including its statistical interpretation and geometrical characterization. Section 2.4 presents the estimating procedure of the maximin-projection estimator and the associated asymptotic properties. Simulation studies to evaluate the empirical performance of the proposed maximin OTR are conducted in Section 6.6. We apply our method to two real examples in Section 2.6. All the proofs and additional numerical studies can be found in Shi et al. (2018c).

2.2 Preliminaries and a toy example

2.2.1 Preliminaries

We consider a single stage study with two treatments. The setting considered in this chapter is slightly different from that in Section 1.2, since we need to take population heterogeneity into account. Specifically, we assume there are G groups of patients. For $g = 1, \ldots, G$, let $Y_{0,g}$, $A_{0,g}$ and $\boldsymbol{X}_{0,g}$ denote the response, the (binary) treatment and the baseline covariates of patients in Group g. In addition, let $Y_{0,g}^*(0)$ and $Y_{0,g}^*(1)$ be the corresponding potential outcomes and define $h_g(a, \boldsymbol{x}) = \mathbb{E}\{Y_{0,g}^*(a)|\boldsymbol{X}_{0,g} = \boldsymbol{x}\}$ for $a \in \{0,1\}$. We may sometimes use a shorthand and write $h_g(0, \boldsymbol{x})$ as $h_g(\boldsymbol{x})$. For any treatment regime $d(\cdot)$ that maps the covariate space to $\{0,1\}$, define $Y_{0,g}^*(d) = Y_{0,g}^*(1)d(\boldsymbol{X}_{0,g}) + Y_{0,g}^*(0)\{1 - d(\boldsymbol{X}_{0,g})\}$.

For simplicity, assume the contrast function is linear, i.e,

$$h_g(1,\boldsymbol{x})-h_g(0,\boldsymbol{x})=\boldsymbol{\beta}_g^T\boldsymbol{x}+c_g, \qquad \forall \boldsymbol{x},$$

for some $\boldsymbol{\beta}_g \in \mathbb{R}^p$ and $c_g \in \mathbb{R}$. Suppose SUTVA, no unmeasured confounders and the positivy assumption hold. It follows that

$$\mathbb{E}(Y_{0,g}|A_{0,g}=1,\boldsymbol{X}_{0,g}=\boldsymbol{x})-\mathbb{E}(Y_{0,g}|A_{0,g}=1,\boldsymbol{X}_{0,g}=\boldsymbol{x})=\boldsymbol{\beta}_g^T\boldsymbol{x}+c_g, \tag{2.1}$$

$$\mathbb{E}(Y_{0,g}|A_{0,g}=1,\boldsymbol{X}_{0,g}=\boldsymbol{x})=h_g(\boldsymbol{x}), \qquad \forall \boldsymbol{x}. \tag{2.2}$$

Without loss of generality, we further assume all covariates $\boldsymbol{X}_{0,g}$s are standardized to have zero mean and identity covariance matrix. Otherwise, we consider variable transformation $\boldsymbol{X}_{0,g}^* = \boldsymbol{\Sigma}_g^{-1/2}(\boldsymbol{X}_{0,g}-\boldsymbol{\mu}_g)$, $\boldsymbol{\beta}_g^* = \boldsymbol{\Sigma}_g^{1/2}\boldsymbol{\beta}_g$, $c_g^* = c_g + \boldsymbol{\mu}_g^T\boldsymbol{\beta}_g$ where $\boldsymbol{\mu}_g = \mathbb{E}(\boldsymbol{X}_{0,g})$ and $\boldsymbol{\Sigma}_g = \mathbb{COV}(\boldsymbol{X}_{0,g})$. Then Model (2.1) can be represented as

$$\mathbb{E}(Y_{0,g}|A_{0,g}=1,\boldsymbol{X}_{0,g}^*=\boldsymbol{x})-\mathbb{E}(Y_{0,g}|A_{0,g}=1,\boldsymbol{X}_{0,g}^*=\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{\beta}_g^*+c_g^*,$$

$$\mathbb{E}(Y_{0,g}|A_{0,g}=0,\boldsymbol{X}_{0,g}^*=\boldsymbol{x})=h_g^*(\boldsymbol{x}), \qquad \forall \boldsymbol{x},$$

for some function h_g^*.

The parameter c_g stands for the marginal treatment effects (average causal effects) after adjusting covariates. Mathematically, we have

$$c_g=\mathbb{E}(\boldsymbol{\beta}_g^T\boldsymbol{X}_{0,g}+c_g)=\mathbb{E}[\mathbb{E}\{Y_{0,g}^*(1)|\boldsymbol{X}_{0,g}\}-\mathbb{E}\{Y_{0,g}^*(0)|\boldsymbol{X}_{0,g}\}].$$

When $c_g > 0$, treatment 1 is generally better for patients in Group g. The vector $\boldsymbol{\beta}_g$ describes individualized treatment effects. For patients in Group g with covariates $\boldsymbol{x}$, the larger $\boldsymbol{\beta}_g^T\boldsymbol{x}$, the more benefits he or she receives if assigned to treatment 1.

Define $\pi_g(a,\boldsymbol{x}) = \mathbb{P}(A_{0,g}=a|\boldsymbol{X}_{0,g}=\boldsymbol{x})$ for $a \in \{0,1\}$ as the propensity score in Group g. We will sometimes write $\pi_g(1,\cdot)$ as $\pi_g(\cdot)$. Model (2.1) allows h_g and π_g to vary across groups, which we refer to baseline effect heterogeneity and treatment assignment heterogeneity respectively. These sources of heterogeneity are not related to treatment decisions since they do not appear in the contrast function. The following sources of groupwise heterogeneity will affect decision making: the marginal treatment effects c_g and the individualized treatment effects $\boldsymbol{\beta}_g$. In this chapter, we mainly focus on heterogeneity caused by different $\boldsymbol{\beta}_g$'s. We assume $c_1 = \cdots = c_G = c_0$ for some c_0, that is, the same marginal treatment effect for all groups.

To introduce the pooled and the maximin optimal treatment regime, we need some optimality criterion. Here, we consider the difference of patient's mean response (value function) between a regime $d(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{\beta}^T \boldsymbol{x} > -c)$ and $d_0(\boldsymbol{x}) = 0$, which assigns all patients to treatment 0. Specifically, the difference of value functions is defined as

$$\mathrm{VD}_g(\boldsymbol{\beta}, c) = \mathbb{E}\{Y^*_{0,g}(d)\} - \mathbb{E}\{Y^*_{0,g}(d_0)\} = \mathbb{E}\{(\boldsymbol{X}^T_{0,g}\boldsymbol{\beta}_g + c_0)\mathbb{I}(\boldsymbol{X}^T_{0,g}\boldsymbol{\beta} > -c)\}.$$

In this section, for illustrative purposes only, we consider a special case with $c_0 = c = 0$. A general discussion will be given in the next section. When the distributions of $\boldsymbol{X}_{0,g}$s are the same across groups, we can represent $\mathrm{VD}_g(\boldsymbol{\beta}, 0)$ as

$$\mathrm{VD}(\boldsymbol{\beta}, \boldsymbol{\beta}_g) = \mathbb{E}\{(\boldsymbol{X}^T_{0,g}\boldsymbol{\beta}_g)\mathbb{I}(\boldsymbol{X}^T_{0,g}\boldsymbol{\beta} > 0)\}.$$

We assume the same number of patients across all groups. Then, the pooled optimal treatment regime is defined as $d_P^{opt,0}(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}^P > 0)$ where

$$\boldsymbol{\beta}^P = \operatorname*{arg\,max}_{\|\boldsymbol{\beta}\|_2=1} \frac{1}{G}\sum_{g=1}^{G} \mathrm{VD}(\boldsymbol{\beta}, \boldsymbol{\beta}_g), \tag{2.3}$$

and the maximin optimal treatment regime is defined as $d_M^{opt,0}(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}^M > 0)$ where

$$\boldsymbol{\beta}^M = \operatorname*{arg\,max}_{\|\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \mathrm{VD}(\boldsymbol{\beta}, \boldsymbol{\beta}_g). \tag{2.4}$$

We add the L_2 constraint on $\boldsymbol{\beta}$ to make $\boldsymbol{\beta}^P$ and $\boldsymbol{\beta}^M$ identifiable. Therefore, the pooled optimal treatment regime aims to maximize the average value difference while the maximin optimal treatment regime aims to maximize the minimum value difference in G groups, i.e. maximize the reward of the worst-case scenario.

The random effects meta-analyses assume the following model for $\boldsymbol{\beta}_g$'s:

$$\boldsymbol{\beta}_g = \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_g,$$

where $\boldsymbol{\varepsilon}_g$'s are independent and satisfy $\mathbb{E}(\boldsymbol{\varepsilon}_g) = 0$, $\mathbb{COV}(\boldsymbol{\varepsilon}_g) = \boldsymbol{\Omega}_0$ for all g. For any subgroup estimators $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_G$ with $\mathbb{COV}(\widehat{\boldsymbol{\beta}}_g) = \boldsymbol{\Omega}_g$, the aggregated estimator is given by

$$\widehat{\boldsymbol{\beta}}^R = \left(\sum_{g=1}^{G}(\widehat{\boldsymbol{\Omega}}_g + \widehat{\boldsymbol{\Omega}}_0)^{-1}\right)^{-1}\left(\sum_{g=1}^{G}(\widehat{\boldsymbol{\Omega}}_g + \widehat{\boldsymbol{\Omega}}_0)^{-1}\widehat{\boldsymbol{\beta}}_g\right),$$

where $\widehat{\boldsymbol{\Omega}}_g$'s and $\widehat{\boldsymbol{\Omega}}_0$ denote some estimators for $\boldsymbol{\Omega}_g$'s and $\boldsymbol{\Omega}_0$. Given sufficiently many observations, we have $\|\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g\|_2 \xrightarrow{P} 0$ and $\|\widehat{\boldsymbol{\Omega}}_g\|_2 \xrightarrow{P} 0$. As a result, we have

$$\widehat{\boldsymbol{\beta}}^R \xrightarrow{P} \left(\sum_{g=1}^{G} (\widehat{\boldsymbol{\Omega}}_0)^{-1}\right)^{-1} \left(\sum_{g=1}^{G} (\widehat{\boldsymbol{\Omega}}_0)^{-1} \boldsymbol{\beta}_g\right) = \frac{1}{G} \sum_g \boldsymbol{\beta}_g \equiv \boldsymbol{\beta}^R. \tag{2.5}$$

The corresponding optimal treatment regime is defined as $d_R^{opt,0}(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T \boldsymbol{\beta}^R > 0)$.

More generally, we can treat the parameters $\boldsymbol{\beta}_g$ in the group-specific contrast function as a multivariate random variable and assume that the parameters $\boldsymbol{\beta}_g$'s of training groups are generated according to some distribution F, either continuous or discrete, and let H denote the support of F. Then, we define $\boldsymbol{\beta}^R$, $\boldsymbol{\beta}^P$ and $\boldsymbol{\beta}^M$ as

$$\begin{aligned} \boldsymbol{\beta}^R &= \mathbb{E}_{\text{train}}(\boldsymbol{b}), \\ \boldsymbol{\beta}^P &= \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \, \mathbb{E}_{\text{train}}\{\text{VD}(\boldsymbol{\beta}, \boldsymbol{b})\}, \\ \boldsymbol{\beta}^M &= \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \min_{\boldsymbol{b} \in H} \text{VD}(\boldsymbol{\beta}, \boldsymbol{b}), \end{aligned}$$

where the expectation $\mathbb{E}_{\text{train}}$ is taken with respect to F. Definitions in (2.3), (2.4) and (2.5) correspond to the special case where F only takes values in $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G\}$ with an equal probability. Our objective is to minimize $\mathbb{E}_{\text{test}}\{\text{VD}(\boldsymbol{\beta})\}$, where $\mathbb{E}_{\text{test}}$ is taken with respect to G, the distribution of $\boldsymbol{\beta}_g$ for future groups of patients.

2.2.2 A toy example

Recall that p is the dimension of $\boldsymbol{X}_{0,g}$. For illustration, we take $p = 2$, and assume that patients' baseline covariates are generated independently from a standard normal distribution. Since $\|\boldsymbol{\beta}\|_2 = 1$, after some calculation, we have

$$\begin{aligned} \text{VD}(\boldsymbol{\beta}, \boldsymbol{\beta}_g) &= \mathbb{E}\{\boldsymbol{X}_{0,g}^T \boldsymbol{\beta}_g \mathbb{I}(\boldsymbol{X}_{0,g}^T \boldsymbol{\beta} > 0)\} = \mathbb{E}\{(\boldsymbol{X}_{0,g}^T \boldsymbol{\beta}_g - \boldsymbol{\beta}_g^T \boldsymbol{\beta} \boldsymbol{X}_{0,g}^T \boldsymbol{\beta} + \boldsymbol{\beta}_g^T \boldsymbol{\beta} \boldsymbol{X}_{0,g}^T \boldsymbol{\beta}) \mathbb{I}(\boldsymbol{X}_{0,g}^T \boldsymbol{\beta} > 0)\} \\ &= \boldsymbol{\beta}_g^T \boldsymbol{\beta} \mathbb{E}\left\{\boldsymbol{X}_{0,g}^T \boldsymbol{\beta} \mathbb{I}(\boldsymbol{X}_{0,g}^T \boldsymbol{\beta} > 0)\right\} = \boldsymbol{\beta}_g^T \boldsymbol{\beta} \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

The first equality in the second line is due to the independence between $\boldsymbol{X}_{0,g}^T \boldsymbol{\beta}_g -$

$\boldsymbol{\beta}_g^T \boldsymbol{\beta} \boldsymbol{X}_{0,g}^T \boldsymbol{\beta}$ and $\boldsymbol{X}_{0,g}^T \boldsymbol{\beta}$. Hence, we obtain

$$\begin{aligned}
\boldsymbol{\beta}^P &= \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \frac{1}{G} \sum_{g=1}^{G} \boldsymbol{\beta}^T \boldsymbol{\beta}_g = \frac{\sum_g \boldsymbol{\beta}_g}{\|\sum_g \boldsymbol{\beta}_g\|_2}, \\
\boldsymbol{\beta}^M &= \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \min_{g\in\{1,\ldots,G\}} \boldsymbol{\beta}^T \boldsymbol{\beta}_g.
\end{aligned} \tag{2.6}$$

Therefore, $\boldsymbol{\beta}^P$ is proportional to $\boldsymbol{\beta}^R$ which equals a simple average of all subgroup parameters, while $\boldsymbol{\beta}^M$ maximizes its minimum inner product across different $\boldsymbol{\beta}_g$'s. When all $\boldsymbol{\beta}_g$'s have the same L_2 norm, $\boldsymbol{\beta}^M$ becomes

$$\boldsymbol{\beta}^M = \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \min_{g\in\{1,\ldots,G\}} \frac{\boldsymbol{\beta}^T \boldsymbol{\beta}_g}{\|\boldsymbol{\beta}_g\|_2}, \tag{2.7}$$

or equivalently

$$\boldsymbol{\beta}^M = \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\min} \max_{g\in\{1,\ldots,G\}} \angle(\boldsymbol{\beta}, \boldsymbol{\beta}_g), \tag{2.8}$$

where $\angle(a, b) = \arccos(a^T b)$ stands for the angle between two vectors. The equivalence between (2.7) and (2.8) is due to the monotonicity of the arccos function. In (2.7) or (2.8), $\boldsymbol{\beta}^M$ is defined to maximize (minimize) the minimum correlation (maximum angle) between all subgroup coefficients. Such formulation is referred to as the maximin correlation approach in the classification literature (c.f, Avi-Itzhak et al. 1995; Lee et al. 2016). In general, we weight the correlation $\boldsymbol{\beta}^T \boldsymbol{\beta}_g / \|\boldsymbol{\beta}_g\|_2$ by the L_2 norm of $\boldsymbol{\beta}_g$. The $\boldsymbol{\beta}^M$ defined in (2.6) is more informative since it not only takes the heterogeneity due to different directions $\boldsymbol{\beta}_g / \|\boldsymbol{\beta}_g\|_2$ into consideration, but different magnitudes $\|\boldsymbol{\beta}_g\|_2$ as well.

Since $\boldsymbol{\beta}^P$ is proportional to $\boldsymbol{\beta}^R$, the VD under $d_P^{opt,0}$ is the same as $d_R^{opt,0}$. Therefore, in the following, we focus on comparing $\boldsymbol{\beta}^P$ with $\boldsymbol{\beta}^M$. We set $G = 4$ and assume $\|\boldsymbol{\beta}_g\|_2 = 1$, $g = 1, 2, 3, 4$. Since $p = 2$, we represent each $\boldsymbol{\beta}_g$ as $\boldsymbol{\beta}_g = \{\cos(\psi_g), \sin(\psi_g)\}^T$ with $\psi_g \in [0, \pi)$. The parameter ψ_g is the angle between $\boldsymbol{\beta}_g$ and the x-axis in a 2-dimensional coordinate system. In this special case, $\boldsymbol{\beta}^M$ lies on the bisector of the largest angles formed by all $\boldsymbol{\beta}_g$'s and it can be shown that $\boldsymbol{\beta}^M = \{\cos(\psi^M), \sin(\psi^M)\}^T$ where

$$\psi^M = \frac{1}{2}\left(\psi_{(1)} + \psi_{(4)}\right),$$

$\psi_{(1)}$ and $\psi_{(4)}$ denote the smallest and largest angles of ψ_g's.

Similarly define $\boldsymbol{\beta}^P = \{\cos(\psi^P), \sin(\psi^P)\}^T$. We set $\psi_1 = 0°$, $\psi_2 = 15°$, $\psi_3 = 70°$ and $\psi_4 =$

90°. Consider the following leave-one-group-out cross validation procedure. For the ith round, we choose the ith group as the testing group, and obtain $\boldsymbol{\beta}^P$ and $\boldsymbol{\beta}^M$ based on the remaining 3 groups. Then we evaluate the value difference of the pooled and maximin OTRs based on the ith group. In other words, we set F to be a discrete distribution that takes value on $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_4\}/\{\boldsymbol{\beta}_i\}$ with equal probability, and G a degenerate distribution that concentrates on $\boldsymbol{\beta}_i$. Table 2.1 summarizes the results.

Table 2.1: Different combinations of training groups and the corresponding ψ^P, ψ^M, and their value differences on the testing group

Training groups	ψ^M (deg)	ψ^P (deg)	ψ_{test} (deg)	VD($\boldsymbol{\beta}^M, \boldsymbol{\beta}_{\text{test}}$)	VD($\boldsymbol{\beta}^P, \boldsymbol{\beta}_{\text{test}}$)
(1,2,3)	35	27.44	90	0.23	0.18
(1,2,4)	45	32.63	70	0.36	0.32
(1,3,4)	45	55.32	15	0.35	0.30
(2,3,4)	52.5	59.25	0	0.24	0.20

From Table 2.1, we can see that for all four cases, the value differences of the maximin optimal treatment regime are uniformly larger than those of the pooled optimal treatment regime on the testing groups. To illustrate the idea graphically, we plot $\boldsymbol{\beta}^P$ (denoted by the snow symbol), $\boldsymbol{\beta}^M$ (denoted by the circle symbol), and $\boldsymbol{\beta}_g$ of the training (denoted by the square symbol) and testing (denoted by the plus symbol) groups for the second and third cases in Figure 6.1, where the left panel is for the second case and the right one is for the third case. For both cases, $\boldsymbol{\beta}^M$ is closer to $\boldsymbol{\beta}_g$ of the testing groups, while $\boldsymbol{\beta}^P$ is pulled towards the area where most $\boldsymbol{\beta}_g$'s of the training groups locate due to the averaging effect.

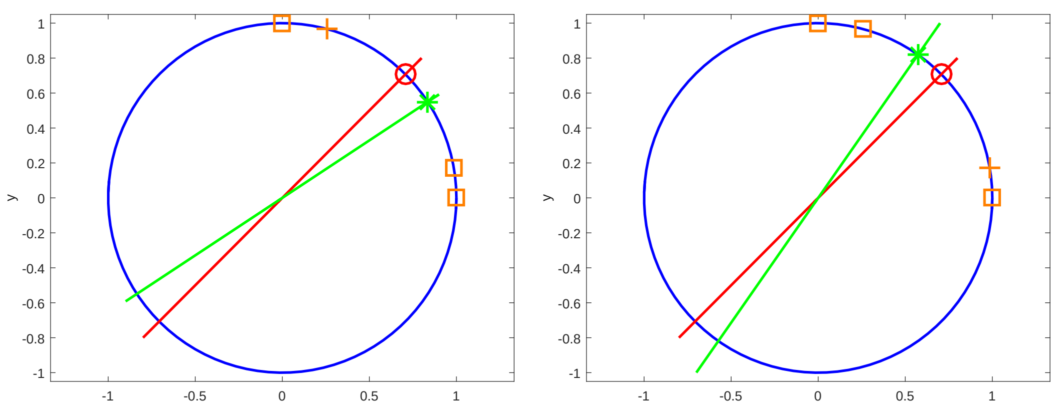

Figure 2.1: Plots of $\boldsymbol{\beta}^P$ (denoted by the snow symbol), $\boldsymbol{\beta}^M$ (denoted by the circle symbol), and $\boldsymbol{\beta}_g$ of the training (denoted by the square symbol) and testing groups (denoted by the plus symbol) for the second (left panel) and third (right panel) cases.

2.3 Maximin-projection learning

We now formally introduce our maximin projection treatment regime. Based on model (2.1) and the common marginal treatment effect assumption, the optimal treatment regime for the gth subgroup is $d_g^{opt,0}(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}_g > -c_0)$. Here, our goal is to find a single treatment regime $d_M^{opt,0}(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}^M > -c^M)$ with $\|\boldsymbol{\beta}^M\|_2 = 1$ that performs uniformly well for heterogeneous data. Motivated by the toy example in the previous section, our proposed maximin-projection learning is aim to find

$$\boldsymbol{\beta}^M = \mathop{\arg\max}_{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \boldsymbol{\beta}^T\boldsymbol{\beta}_g.$$

2.3.1 Statistical interpretation

In this subsection, we show that the maximin projection, represented by $\boldsymbol{\beta}^M$, has two nice statistical interpretations in terms of maximizing the minimum PCD and value difference (VD). Specifically, in group g, the PCD of a treatment regime $d(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta} > -c)$ is defined as

$$\text{PCD}_g(\boldsymbol{\beta}, c) = 1 - \mathbb{E}\{|\mathbb{I}(\boldsymbol{X}_{0,g}^T\boldsymbol{\beta} > -c) - \mathbb{I}(\boldsymbol{X}_{0,g}^T\boldsymbol{\beta}_g > -c_0)|\},$$

and the VD is defined as

$$\text{VD}_g(\boldsymbol{\beta}, c) = \mathbb{E}[Y_{0,g}^*\{\mathbb{I}(\boldsymbol{X}_{0,g}^T\boldsymbol{\beta} > -c)\}] - \mathbb{E}\{Y_{0,g}^*(0)\} = \mathbb{E}\left\{(\boldsymbol{X}_{0,g}^T\boldsymbol{\beta}_g + c_0)\mathbb{I}(\boldsymbol{X}_{0,g}^T\boldsymbol{\beta} > -c)\right\}.$$

Here, the larger PCD and VD values, the better the treatment regime $d(\boldsymbol{x})$ approximates the groupwise optimal treatment regime $d_g^{opt,0}(\boldsymbol{x})$.

Based on the defined PCD and VD, for any fixed constant c, we consider the following maximin treatment regimes: $d_1(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}_{(1)}^M > -c)$ where

$$\boldsymbol{\beta}_{(1)}^M = \mathop{\arg\max}_{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \text{PCD}_g(\boldsymbol{\beta}, c), \tag{2.9}$$

and $d_2(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}_{(2)}^M > -c)$ where

$$\boldsymbol{\beta}_{(2)}^M = \mathop{\arg\max}_{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \text{VD}_g(\boldsymbol{\beta}, c). \tag{2.10}$$

The two maximin treatment regimes, defined by $\boldsymbol{\beta}_{(1)}^M$ and $\boldsymbol{\beta}_{(2)}^M$, are appealing for their nice statistical interpretations. However, we note that the definition of $\boldsymbol{\beta}_{(1)}^M$ involves unknown parameters. The empirical estimators of VD are of non-smooth and non-concave functional forms of the corresponding estimators. Therefore, their estimations are not feasible and they may not be practically useful.

It is worth noting that $\boldsymbol{\beta}_{(1)}^M$ would be meaningless when not all $\|\boldsymbol{\beta}_g\|_2$'s are the same. This is because PCD only measures the similarity between the overall and groupwise optimal treatment decisions, but does not account for the magnitude of groupwise contrast function. When $\|\boldsymbol{\beta}_g\|_2$'s are not the same, the L_2 norm of groupwise contrast function $\{\mathbb{E}(\boldsymbol{X}_{0,g}^T\boldsymbol{\beta}_g + c_0)^2\}^{1/2}$ would be different. This implies that PCDs are not comparable across different groups. In comparison, VD is a better criterion since it takes both the sign and magnitude of contrast function into consideration. Below, under some conditions, we establish the equivalence between these two maximin treatment regimes and our proposed maximin-projection treatment regime.

Theorem 2.3.1 (Equivalence of $\boldsymbol{\beta}^M$ and $\boldsymbol{\beta}_{(1)}^M$) *Assume that $\boldsymbol{X}_{0,g}$'s are i.i.d. spherically distributed, and all $\|\boldsymbol{\beta}_g\|_2$'s are the same. Then, for any fixed c,*

$$\boldsymbol{\beta}^M = \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \min_{g\in\{1,\dots,G\}} PCD_g(\boldsymbol{\beta}, c).$$

Theorem 2.3.2 (Equivalence of $\boldsymbol{\beta}^M$ and $\boldsymbol{\beta}_{(2)}^M$) *Assume $\boldsymbol{X}_{0,g}$'s are i.i.d. spherically distributed. Then, for any fixed c,*

$$\boldsymbol{\beta}^M = \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \min_{g\in\{1,\dots,G\}} VD_g(\boldsymbol{\beta}, c).$$

Theorems 2.3.1 and 2.3.2 require $\boldsymbol{X}_{0,g}$ to have a spherical distribution which requires $\boldsymbol{X}_{0,g} \overset{d}{=} \boldsymbol{U}\boldsymbol{X}_{0,g}$ for any orthogonal $p \times p$ matrix $\boldsymbol{U}$. This includes a rich class of symmetric multivariate distributions (see Fang et al. 1990).

The definition of $\boldsymbol{\beta}^M$ has nice statistical interpretations. However, it has two drawbacks. First, when $F_0 \equiv \max_{\|\boldsymbol{\beta}\|_2=1} \min_g \boldsymbol{\beta}^T\boldsymbol{\beta}_g < 0$, the uniqueness of $\boldsymbol{\beta}^M$ is not guaranteed. This may cause identifiability issues when we establish properties of the corresponding estimators. In addition, the optimization problem in (2.6) is not concave. This can make the implementation of the estimating procedure infeasible.

To address these concerns, we define

$$\boldsymbol{\beta}_{(0)}^M = \underset{\|\boldsymbol{\beta}\|_2\le 1}{\arg\max} \min_{g\in\{1,\dots,G\}} \boldsymbol{\beta}^T\boldsymbol{\beta}_g. \tag{2.11}$$

Compared to $\boldsymbol{\beta}^M$, it replaces the feasible set $\|\beta\|_2 = 1$ with a closed convex set $\|\beta\|_2 \leq 1$. Lemma 2.3.1 below states that $\beta^M_{(0)}$ is well defined, when $F_0 \neq 0$. Moreover, the optimization problem (2.11) is concave, which can be easily implemented.

Lemma 2.3.1 *The maximin-projection estimator $\boldsymbol{\beta}^M_{(0)}$ always exists. Moreover, when $F_0 \neq 0$, $\boldsymbol{\beta}^M_{(0)}$ is unique.*

The existence of $\boldsymbol{\beta}^M_{(0)}$ is guaranteed by the continuity of the objective function $F(\boldsymbol{\beta}) = \min_{g\in\{1,\ldots,G\}} \boldsymbol{\beta}^T \boldsymbol{\beta}_g$, boundedness and closeness of the feasible set $\boldsymbol{\beta} : \|\beta\|_2 \leq 1$. Its uniqueness is a byproduct of lemma 2.3.3, which is stated in the next subsection. When $F_0 = 0$, $\beta^M_{(0)}$ is not unique and the set of solutions is given by

$$\{a\boldsymbol{\beta} : a \in [0,1], \|\beta\|_2 = 1, \max_{\|\beta\|_2=1} \min_{g\in\{1,\ldots,G\}} \boldsymbol{\beta}^T \boldsymbol{\beta}_g = 0\}.$$

The problem of estimating $\beta^M_{(0)}$ then becomes non-regular and all the large sample theories about the maximin estimator fail (see Section 2.4).

Define $G_0 = \max_{\|\beta\|_2\leq 1} \min_g \boldsymbol{\beta}^T \boldsymbol{\beta}_g$. It is obvious that $G_0 \geq 0$. In addition, $G_0 > 0$ if and only if $F_0 > 0$. When $G_0 = 0$, we can set $\boldsymbol{\beta}^M_{(0)} = 0$, which leads to a trivial regime by assigning the same treatment to all patients. From now on, we focus on the situation when $G_0 > 0$. In this case, we have $\boldsymbol{\beta}^M = \beta^M_{(0)}$. Define

$$c^M_{(0)} = c_0/G_0.$$

Note that $c^M_{(0)}$ and c_0 are sign equivalent. Our maximin-projection OTR is given by

$$d_M^{opt,0}(\boldsymbol{x}) = \mathbb{I}(\boldsymbol{x}^T \boldsymbol{\beta}^M_{(0)} > -c^M_{(0)}).$$

Theorem 2.3.3 *Under conditions of theorem 2.3.1, if $G_0 > 0$, we have*

$$c^M_{(0)} = \arg\max_c \min_{g\in\{1,\ldots,G\}} PCD_g(\beta^M_{(0)}, c).$$

Theorem 2.3.4 *Under conditions of theorem 2.3.2, if $G_0 > 0$, we have*

$$c^M_{(0)} = \arg\max_c \min_{g\in\{1,\ldots,G\}} VD_g(\boldsymbol{\beta}^M_{(0)}, c).$$

Together with theorems 2.3.1 and 2.3.2, theorems 2.3.3 and 2.3.4 suggest that the treatment regime $d_M^{opt,0}(x)$ maximizes the minimum PCD and the minimum VD among different groups.

2.3.2 Geometrical characterization

In this subsection we give a geometrical view of $\boldsymbol{\beta}_{(0)}^M$ when $G_0 > 0$. Findings in this subsection are similar in rationale with the results in Avi-Itzhak et al. (1995). However, we generalize their results by getting rid of the unit L_2-norm condition $\|\boldsymbol{\beta}_g\|_2 = 1$ and allowing the set of vectors $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G\}$ to be linear dependent, which is the case when $p \geq G$.

We first introduce some notation. For an arbitrary $p \times G$ matrix $\boldsymbol{\Psi}$ and a set $K \subseteq \{1, \ldots, G\}$, let $\boldsymbol{\Psi}_K$ denote the submatrice of $\boldsymbol{\Psi}$ formed by columns in K. Define the equicorrelated points set

$$\mathrm{E}_K(\boldsymbol{\Psi}) = \left\{\boldsymbol{t} \in \mathbb{R}^p | \boldsymbol{t}^T \boldsymbol{\Psi}_j = \boldsymbol{t}^T \boldsymbol{\Psi}_i, \forall i, j \in K\right\},$$

and the optimal equicorrelated point

$$\mathrm{E}_K^\star(\boldsymbol{\Psi}) = \underset{\substack{\boldsymbol{t} \in E_K(\boldsymbol{\Psi}) \\ \|\boldsymbol{t}\|_2 = 1}}{\arg\max} \left\{\boldsymbol{t}^T \boldsymbol{\Psi}_i, \forall i \in K\right\},$$

where $\boldsymbol{\Psi}_i$ refers to the ith column vector of matrix $\boldsymbol{\Psi}$. When $|K| = 1$ and $\boldsymbol{\Psi}_K = \boldsymbol{\psi}$, $\mathrm{E}_K^\star(\boldsymbol{\Psi}) = \boldsymbol{\psi}/\|\boldsymbol{\psi}\|_2$. Readers can refer to Shi et al. (2018c) for a detailed discussion on the equicorrelated points set and the optimal equicorrelated point.

For any matrix $\boldsymbol{\Omega}$, Let $\boldsymbol{\Omega}^+$ denote the Moore-Penrose matrix inverse of $\boldsymbol{\Omega}$ and $C(\boldsymbol{\Omega})$ the column space of $\boldsymbol{\Omega}$. Let $\boldsymbol{e}$ denote a vector of ones. We have the following result.

Lemma 2.3.2 *For any $\boldsymbol{\Psi}$ and $K \subseteq \{1, \ldots, n\}$, when $\boldsymbol{e} \in C(\boldsymbol{\Psi}_K^T)$, the optimal equicorrelated point of $\boldsymbol{\Psi}_K$ exists and is unique. Moreover, it takes the form*

$$E_K^\star(\boldsymbol{\Psi}) = \{\boldsymbol{e}^T (\boldsymbol{\Psi}_K^T \boldsymbol{\Psi}_K)^+ \boldsymbol{e}\}^{-1/2} \boldsymbol{\Psi}_K (\boldsymbol{\Psi}_K^T \boldsymbol{\Psi}_K)^+ \boldsymbol{e}. \tag{2.12}$$

Define matrix $\boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_G)$ whose gth column is the subgroup parameter $\boldsymbol{\beta}_g$.

Lemma 2.3.3 *Assume $G_0 > 0$. Then there exists a unique nonempty set $K_0 \subseteq \{1, \ldots, G\}$ such that $\boldsymbol{\beta}_{(0)}^M = E_{K_0}^\star(\boldsymbol{B})$ and $\min_{g \in K_0^c} \boldsymbol{\beta}_{(0)}^{M\,T} \boldsymbol{\beta}_g > G_0$, where $K_0^c = \{1, \ldots, G\} - K_0$. Moreover, if the set of vectors $\boldsymbol{\beta}_g, g \in K_0$ are linearly independent, then a necessary and sufficient condition for $\boldsymbol{\beta}_{(0)}^M = E_{K_0}^\star(\boldsymbol{B})$ is that each element in the vector $(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1} \boldsymbol{e}$ is nonnegative.*

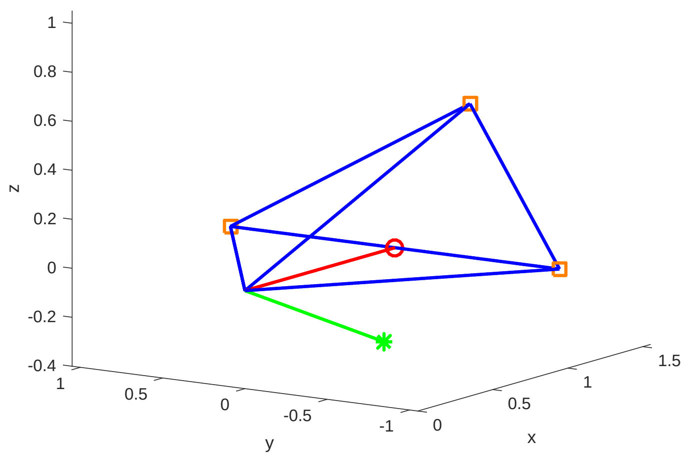

Figure 2.2: Plots of β_g (denoted by the square symbol), $\mathrm{E}^\star_{\{1,2,3\}}(\boldsymbol{B})$ (denoted by the snow symbol) and $\mathrm{E}^\star_{\{1,2\}}(\boldsymbol{B})$ (denoted by the circle symbol)

We denote K_0 as the maximin optimal equicorrelated points set when $G_0 > 0$. In lemma 2.3.2, the condition $\boldsymbol{e} \in C(\boldsymbol{\Psi}_K^T)$ automatically holds when $\boldsymbol{\Psi}_K^T$ has full row rank. In lemma 2.3.3, we assume the set of vectors $\beta_g, g \in K_0$ are linearly independent. This implies the matrix $\boldsymbol{B}_{K_0}^T$ has full row rank. As a result, we have $e \in C(\boldsymbol{B}_{K_0}^T)$.

In lemma 2.3.3, the non-negativity of $(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1}\boldsymbol{e}$ is sufficient and necessary for $\beta_{(0)}^M = \mathrm{E}_{K_0}(\boldsymbol{B})$. Together with lemma 2.3.2, lemma 2.3.3 implies that $\beta_{(0)}^M$ is uniquely defined by

$$\beta_{(0)}^M = \mathbb{E}^\star_{K_0}(\boldsymbol{B}) = \{\boldsymbol{e}^T(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1}\boldsymbol{e}\}^{-1/2}\boldsymbol{B}_{K_0}(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1}\boldsymbol{e}.$$

This implies $\mathrm{E}^\star_{K_0}(\boldsymbol{B})$ is proportional to $\boldsymbol{B}_{K_0}(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1}\boldsymbol{e}$ and can be represented as a linear combination of the column vectors in $\boldsymbol{B}_{K_0}$. Geometrically, the non-negativity of $(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1}\boldsymbol{e}$ requires $\mathrm{E}^\star_{K_0}(\boldsymbol{B})$ to lie in the convex cone of $\beta_g, g \in K_0$, i.e, $\{\sum_{g \in K_0} a_g \beta_g : a_g \geq 0, \forall g \in K_0\}$. To better understand lemma 2.3.3, in Figure 2.2, we take $s = 3$, $G = 3$ and $\boldsymbol{B} = (\beta_1, \beta_2, \beta_3)$ where $\beta_1 = (1,1,0)$, $\beta_2 = (1,-1,0)$ and $\beta_3 = (1.2,0,0.5)$. Both $\mathrm{E}^\star_{\{1,2,3\}}(\boldsymbol{B})$ and $\mathrm{E}^\star_{\{1,2\}}(\boldsymbol{B})$ satisfy the necessary conditions of lemma 2.3.3. While $\mathrm{E}^\star_{\{1,2\}}(\boldsymbol{B})$ lies in the convex cone of β_1 and β_2, $\mathrm{E}^\star_{\{1,2,3\}}(\boldsymbol{B})$ appears outside the convex cone of β_1, β_2 and β_3. Therefore, $\mathrm{E}^\star_{\{1,2\}}(\boldsymbol{B})$ satisfies the sufficient conditions of lemma 2.3.3 and $\mathrm{E}^\star_{\{1,2,3\}}(\boldsymbol{B})$ does not. As a result, we have $\beta_{(0)}^M = \mathrm{E}^\star_{\{1,2\}}(\boldsymbol{B})$.

2.4 Estimation procedure

The data are summarized as $(Y_{j,g}, A_{j,g}, \boldsymbol{X}_{j,g})$, for $g = 1, \ldots, G$, $j = 1, \ldots, m_g$, where m_g is the number of patients in Group g. We assume that the data are independent across $g = 1, \ldots, G$ and $j = 1, \ldots, m_g$. Based on the data, parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G$ and c_0 in model (2.1) can be estimated with existing methods. In this chapter, we implement with the popular Q-learning and A-learning and give a brief discussion on estimating these parameters in Section 2.4.2. Let $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_G$ and $\widehat{c}_0$ be the corresponding estimators. We propose to estimate $\boldsymbol{\beta}_{(0)}^M$ by solving the following optimization problem:

$$\widehat{\boldsymbol{\beta}}^M = \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2 \le 1}{\arg\max} \min_{g \in \{1,\ldots,G\}} \boldsymbol{\beta}^T \widehat{\boldsymbol{\beta}}_g. \tag{2.13}$$

Note that the objective function $\min_g \boldsymbol{\beta}^T \widehat{\boldsymbol{\beta}}_g$ is concave in $\boldsymbol{\beta}$ and the region $\|\boldsymbol{\beta}\|_2 \le 1$ is convex. Therefore, (2.13) is a tractable convex optimization problem. It can be further casted as a quadratic constraint linear programming (QCLP) problem, specifically, $\widehat{\boldsymbol{\beta}}^M$ is equivalent to the solution of

$$\begin{aligned} &\text{maximize} && t \in \mathbb{R} \\ &\text{subject to} && \boldsymbol{\beta}^T \widehat{\boldsymbol{\beta}}_g \ge t, g = 1, \ldots, G \\ &&& \boldsymbol{\beta}^T \boldsymbol{\beta} \le 1. \end{aligned}$$

The above optimization problem can be efficiently computed using existing softwares. Define $\widehat{c}^M = \widehat{c}_0 / \widehat{G_0}$, where $\widehat{G_0} = \min_g \widehat{\boldsymbol{\beta}}_g^T \widehat{\boldsymbol{\beta}}^M$.

Given a group of future patients, denoted by $\{\boldsymbol{X}_{j,G+1}\}_{j=1}^n$ their baseline covariates. We calculate $\widehat{\boldsymbol{\mu}}_{G+1} = \sum_{j=1}^n \boldsymbol{X}_{j,G+1}/n$ and $\widehat{\boldsymbol{\Sigma}}_{G+1} = \sum_{j=1}^n (\boldsymbol{X}_{j,G+1} - \widehat{\boldsymbol{\mu}}_{G+1})(\boldsymbol{X}_{j,G+1} - \widehat{\boldsymbol{\mu}}_{G+1})^T/(n-1)$. The recommend treatment for the jth patient is given by

$$\mathbb{I}\{(X_{j,G+1} - \widehat{\boldsymbol{\mu}}_{G+1})^T \widehat{\boldsymbol{\Sigma}}_{G+1}^{-1/2} \widehat{\boldsymbol{\beta}}^M > -\widehat{c}^M\}.$$

2.4.1 Statistical properties

In this subsection we investigate the asymptotic properties of the maximin-projection estimator $\widehat{\boldsymbol{\beta}}^M$ obtained by solving the optimization problem (2.13). We first study the consistency of the estimator by assuming the following two conditions.

(S1.) Assume that $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_G$ and $\widehat{c}_0$ converge in probability to $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G$ and c_0, respectively.

(S2.) Assume that $F_0 \neq 0$. When $F_0 > 0$, assume that the column vectors in $\boldsymbol{B}_{K_0}$ are linearly independent and all elements in the vector $(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1}\boldsymbol{e}$ are nonzero, where K_0 is the maximin optimal equicorrelated points set as defined previously.

Condition S1 requires each subgroup estimator to be consistent. The condition $F_0 \neq 0$ in S2 ensures the existence and uniqueness of $\boldsymbol{\beta}_{(0)}^M$. Apparently, $\boldsymbol{\beta}_{(0)}^M$ is not stable when F_0 approaches to 0, since its L_2 norm will change from 1 to 0. To ensure the stability of $\boldsymbol{\beta}_{(0)}^M$ in the sense that it will not deviate too much when there are minor changes in the set of vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G$, we would expect

$$\|(\tilde{\boldsymbol{B}}_{K_0}^T \tilde{\boldsymbol{B}}_{K_0})^+ - (\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^+\|_2 \to 0, \tag{2.14}$$

as $\tilde{\boldsymbol{B}}_{K_0} \to \boldsymbol{B}_{K_0}$, where $\tilde{\boldsymbol{B}} = (\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_G)$ represents the coefficient matrix with some disturbance. A sufficient condition to establish (2.14) is that $\boldsymbol{B}_{K_0}$ is of full column rank, as assumed in Condition S2. Lemma 2.3.2 suggests $\boldsymbol{\beta}_{(0)}^M$ can be represented as $\boldsymbol{\omega}_0^T \boldsymbol{B}_{K_0}$, for some weight vector $\boldsymbol{\omega}_0$ proportional to $(\boldsymbol{B}_{K_0}^T \boldsymbol{B}_{K_0})^{-1}\boldsymbol{e}$. Condition S2 further assumes the weights are nonzero. Such a condition guarantees that for any coefficient matrix $\tilde{\boldsymbol{B}} \to \boldsymbol{B}$, K_0 is the optimal equicorrelated points set of $\tilde{\boldsymbol{B}}$ as well.

Theorem 2.4.1 (Consistency) *Define $\widehat{\boldsymbol{B}} = (\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_G)$. Assume Conditions S1 and S2 are satisfied. Then with probability tending to 1, the estimator $\widehat{\boldsymbol{\beta}}^M$ is equal to*

$$\begin{cases} \{\boldsymbol{e}^T(\widehat{\boldsymbol{B}}_{K_0}^T \widehat{\boldsymbol{B}}_{K_0})^{-1}\boldsymbol{e}\}^{-1/2} \widehat{\boldsymbol{B}}_{K_0} (\widehat{\boldsymbol{B}}_{K_0}^T \widehat{\boldsymbol{B}}_{K_0})^{-1}\boldsymbol{e} & \text{if } F_0 > 0, \\ 0 & \text{if } F_0 < 0. \end{cases}$$

In addition, assume there exist some $r_n^{(1)}, r_n^{(2)} \to 0$ such that $\max_{g \in K_0} \|\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g\|_2 = O_p(r_n^{(1)})$ and $\widehat{c}_0 = c_0 + O_p(r_n^{(2)})$. When $F_0 > 0$, we have $\|\widehat{\boldsymbol{\beta}}^M - \boldsymbol{\beta}_{(0)}^M\|_2 = O_p(r_n^{(1)})$, $\widehat{c}^M = c_{(0)}^M + O_p(r_n^{(1)} + r_n^{(2)})$.

Theorem 2.4.1 implies that $(\widehat{\boldsymbol{\beta}}^M, \widehat{c}^M)$ is consistent as long as each subgroup estimator is consistent. The first part of the theorem follows as a consequence of lemma 2.3.3.

Next, we study the asymptotic normality of the estimator. For notational simplicity, we assume $m_1 = \cdots = m_G = m$ and posit the following condition.

(S3.) Assume that for all $g \in K_0$, $\sqrt{m}(\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g)$ and $\sqrt{m}(\widehat{c}_0 - c_0)$ are jointly asymptotically normal with mean zero.

Theorem 2.4.2 (Asymptotic normality) *Assume that Conditions S1–S3 hold, and that $F_0 > 0$. We have that $\sqrt{m}(\widehat{\boldsymbol{\beta}}^M - \boldsymbol{\beta}_{(0)}^M)$ and $\sqrt{m}(\widehat{c}^M - c_{(0)}^M)$ are jointly asymptotically normal with* *mean zero and some covariance matrix $\boldsymbol{V}^M$. The expression of $\boldsymbol{V}^M$ is given in Appendix C of Shi et al. (2018c).*

Since the expression of the asymptotic covariance matrix $\boldsymbol{V}^M$ is quite complicated, we propose to estimate it using a bootstrap method. Here, the bootstrap sampling is done within each subgroup. Specifically, we independently generate B bootstrap samples for each group $g = 1, \ldots, G$,

$$\left\{(Y_{1,g}^{(j)}, A_{1,g}^{(j)}, \boldsymbol{X}_{1,g}^{(j)}), \ldots, (Y_{m,g}^{(j)}, A_{m,g}^{(j)}, \boldsymbol{X}_{m,g}^{(j)})\right\},$$

$j = 1, \ldots, B$. For each j, we obtain estimators $\widehat{\boldsymbol{\beta}}^{(j)}$ and $\widehat{c}^{(j)}$ based on the data

$$\left\{(Y_{1,1}^{(j)}, A_{1,1}^{(j)}, \boldsymbol{X}_{1,1}^{(j)}), \ldots, (Y_{m,1}^{(j)}, A_{m,1}^{(j)}, \boldsymbol{X}_{m,1}^{(j)})\right\}, \ldots, \left\{(Y_{1,G}^{(j)}, A_{1,G}^{(j)}, \boldsymbol{X}_{1,G}^{(j)}), \ldots, (Y_{m,G}^{(j)}, A_{m,G}^{(j)}, \boldsymbol{X}_{m,G}^{(j)})\right\}.$$

Confidence intervals of $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{c}^M$ are calculated based on quantiles of $(\widehat{\boldsymbol{\beta}}^{(1)}, \ldots, \widehat{\boldsymbol{\beta}}^{(B)})$ and $(\widehat{c}^{(1)}, \ldots, \widehat{c}^{(M)})$.

2.4.2 Estimation of group-specific regimes

In this subsection we discuss two popular approaches to obtain subgroup estimators $\widehat{\boldsymbol{\beta}}_g$ and $\widehat{c}_0$.

Example 2.4.1 (Q-learning) *We estimate $\boldsymbol{\beta}_g$ and c_0 by modeling the Q-functions, which represent the conditional mean of the response given the covariates and the treatment. Specifically, the baseline function is assumed to have some parametric form $h_g(\boldsymbol{x}, \boldsymbol{\eta}_g)$ with parameter $\boldsymbol{\eta}_g$. Then,*

$$Q_g(\boldsymbol{X}_{0,g}, A_g; \boldsymbol{\beta}_g, c_0, \boldsymbol{\eta}_g) \equiv \mathbb{E}(Y_g | A_g, \boldsymbol{X}_{0,g}) = h_g(\boldsymbol{X}_{0,g}, \boldsymbol{\eta}_g) + A_g(\boldsymbol{X}_{0,g}^T \boldsymbol{\beta}_g + c_0), \quad g = 1, \ldots, G.$$

Since c_0 is common across all subgroups, we propose to estimate $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G$ and c_0 by jointly solving the following set of estimating equations:

$$\sum_j \frac{\partial h_g(\boldsymbol{X}_{j,g}, \boldsymbol{\eta}_g)}{\partial \boldsymbol{\eta}_g} \{Y_{j,g} - Q_g(\boldsymbol{X}_{j,g}, A_{j,g}; \boldsymbol{\beta}_g, c_0, \boldsymbol{\theta}_g)\} = 0, \quad g = 1, \ldots, G,$$
$$\sum_j A_{j,g} \boldsymbol{X}_{j,g} \{Y_{j,g} - Q_g(\boldsymbol{X}_{j,g}, A_{j,g}; \boldsymbol{\beta}_g, c_0, \boldsymbol{\theta}_g)\} = 0, \quad g = 1, \ldots, G,$$
$$\sum_g \sum_j A_{j,g} \{Y_{j,g} - Q_g(\boldsymbol{X}_{j,g}, A_{j,g}; \boldsymbol{\beta}_g, c_0, \boldsymbol{\theta}_g)\} = 0.$$

When the parametric models $h_g(\boldsymbol{x}, \boldsymbol{\eta}_g)$'s are correctly specified, the resulting estimators $\widehat{\boldsymbol{\beta}}_g$'s and $\widehat{c}_0$ are consistent and jointly asymptotically normal.

Example 2.4.2 (A-learning) *Here, we posit some parametric model $\pi_g(\boldsymbol{x}, \boldsymbol{\alpha}_g)$ for the propensity score and $h_g(\boldsymbol{x}, \boldsymbol{\eta}_g)$ for the baseline function. The parameters $\boldsymbol{\alpha}_g$'s, $\boldsymbol{\eta}_g$'s, $\boldsymbol{\beta}_g$'s and c_0 are estimated by solving the following set of estimating equations:*

$$\sum_j \frac{1}{\pi_g(\boldsymbol{X}_{j,g}, \boldsymbol{\alpha}_g)\{1-\pi_g(\boldsymbol{X}_{j,g}, \boldsymbol{\alpha}_g)\}} \frac{\partial \pi_g(\boldsymbol{X}_{j,g}, \boldsymbol{\alpha}_g)}{\partial \boldsymbol{\alpha}_g}\{A_{j,g} - \pi_g(\boldsymbol{X}_{j,g}, \boldsymbol{\alpha}_g)\} = 0, \quad g = 1, \dots, G,$$
$$\sum_j \frac{\partial h_g(\boldsymbol{X}_{j,g}, \boldsymbol{\eta}_g)}{\partial \boldsymbol{\eta}_g}\{Y_{j,g} - h_g(\boldsymbol{X}_{j,g}, \boldsymbol{\eta}_g) - A_{j,g}(\boldsymbol{X}_{j,g}^T \boldsymbol{\beta}_g + c_0)\} = 0, \quad g = 1, \dots, G,$$
$$\sum_j \boldsymbol{X}_{j,g}\{A_{j,g} - \pi_g(\boldsymbol{X}_{j,g}, \boldsymbol{\alpha}_g)\}\{Y_{j,g} - h_g(\boldsymbol{X}_{j,g}, \boldsymbol{\eta}_g) - A_{j,g}(X_{j,g}^T \boldsymbol{\beta}_g + c_0)\} = 0, \quad g = 1, \dots, G,$$
$$\sum_g \sum_j \{A_{j,g} - \pi_g(\boldsymbol{X}_{j,g}, \boldsymbol{\alpha}_g)\}\{Y_{j,g} - h_g(\boldsymbol{X}_{j,g}, \boldsymbol{\eta}_g) - A_{j,g}(\boldsymbol{X}_{j,g}^T \boldsymbol{\beta}_g + c_0)\} = 0.$$

It can be shown that when either the propensity score or the baseline function for each group is correctly specified, the resulting estimators $\widehat{\boldsymbol{\beta}}_g$'s and $\widehat{c}_0$ are consistent and jointly asymptotically normal. This is the so-called doubly robust property of the A-learning estimation.

2.5 Simulation studies

We consider four groups of patients. In each group, we generate 200 samples according to the following model

$$Y_{j,g} = h(\boldsymbol{X}_{j,g}) + A_{j,g}\boldsymbol{X}_{j,g}^T \boldsymbol{\beta}_g + \boldsymbol{\varepsilon}_{j,g},$$

where $\boldsymbol{X}_{j,g} = (X_{j,g}^{(1)}, X_{j,g}^{(2)})^T \overset{iid}{\sim} N(0, I_2)$ and $\boldsymbol{\varepsilon}_{j,g} \overset{iid}{\sim} N(0, 0.25)$. Two baseline models are considered for h, including a linear model $h(X_{j,g}) = 1 + 0.5X_{j,g}^{(1)} + 0.5X_{j,g}^{(2)}$ and a nonlinear model $h(X_{j,g}) = 1 + \sin(0.5\pi X_{j,g}^{(1)} + 0.5\pi X_{j,g}^{(2)})$. We generate treatments from two propensity score models, a constant model, $\mathbb{P}(A_{j,g} = 1) = 0.5$ and a probit model, $\mathbb{P}(A_{j,g} = 1|X_{j,g}) = \Phi(X_{j,g}^{(1)} - X_{j,g}^{(2)})$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. This yields four simulation settings.

We further consider two scenarios for the subgroup parameters to exhibit different degrees of heterogeneity. In the first scenario, we set $\boldsymbol{\beta}_1^T = (2, 0), \boldsymbol{\beta}_2^T = (2\cos(15°), 2\sin(15°))$, $\boldsymbol{\beta}_3^T = (2\cos(70°), 2\sin(70°))$, $\boldsymbol{\beta}_4^T = (0, 2)$. Hence, all $\boldsymbol{\beta}_g$'s have the same L_2 norm and their directions $\boldsymbol{\beta}_g/\|\boldsymbol{\beta}_g\|_2$ differ. For the second scenario, we choose subgroup parameters to have similar directions but allow their L_2 norms to vary. Specifically, $\boldsymbol{\beta}_1^T = (2.2\cos(30°), 2.2\sin(30°))$, $\boldsymbol{\beta}_2^T = (1.5\cos(45°), 1.5\sin(45°))$, $\boldsymbol{\beta}_3^T = (2.2\cos(54°), 2.2\sin(54°))$, $\boldsymbol{\beta}_4^T = (2\cos(60°), 2\sin(60°))$. It can be shown that $\boldsymbol{\beta}_{(0)}^M = (\cos(45°), \sin(45°))$ and $c_{(0)}^M = 0$ for all scenarios.

We first obtain the subgroup estimators of $\boldsymbol{\beta}_g$ and c_0 using the A-learning estimating equations discussed in Section 4.2. Here, a logistic regression model is fitted for the propensity score and a linear model for the baseline function. As a result, both the propensity score model and the baseline model are correctly specified in the first setting; either of them is misspecified in the second and the third setting; while both are misspecified in the last setting. We then obtain the estimators $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{c}^M$ using the proposed maximin-projection learning. Confidence intervals for the resulting estimators are obtained based on 600 bootstrap samples.

Table 2.2: Biases, standard deviations (in parenthesis) of $\widehat{\boldsymbol{\beta}}^M$, $\widehat{c}^M$ and coverage probabilities (CP) of 95% Wald-type confidence intervals for $\boldsymbol{\beta}_{(0)}^M$ and $c_{(0)}^M$.

Scenario 1	$\widehat{\beta}_1^M$	$\widehat{\beta}_2^M$	$\widehat{c}^M$	CP for $\widehat{\beta}_1^M$	CP for $\widehat{\beta}_2^M$	CP for $\widehat{c}^M$
Setting 1	−0.002(0.027)	0.001(0.027)	0.0003(0.024)	96.0%	96.0%	95.3%
Setting 2	−0.003(0.053)	−0.001(0.052)	0.001(0.045)	94.7%	94.7%	93.8%
Setting 3	−0.003(0.036)	0.001(0.035)	−0.0005(0.035)	96.2%	96.2%	94.5%
Setting 4	−0.003(0.068)	−0.004(0.068)	0.002(0.068)	96.0%	96.0%	95.0%
Scenario 2	$\widehat{\beta}_1^M$	$\widehat{\beta}_2^M$	$\widehat{c}^M$	CP for $\widehat{\beta}_1^M$	CP for $\widehat{\beta}_2^M$	CP for $\widehat{c}^M$
Setting 1	−0.002(0.036)	0.0002(0.036)	0.0002(0.023)	95.5%	95.5%	95.3%
Setting 2	−0.009(0.061)	0.003(0.060)	−0.001(0.043)	96.0%	96.0%	93.8%
Setting 3	−0.010(0.091)	−0.002(0.089)	−0.001(0.033)	93.7%	93.7%	94.5%
Setting 4	−0.029(0.136)	0.034(0.130)	−0.002(0.056)	98.3%	98.3%	95.0%

For each setting, we conduct 600 simulations. The biases, standard deviations (SD) of $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{c}^M$, and coverage probabilities (CP) of 95% Wald-type confidence intervals for $\boldsymbol{\beta}_{(0)}^M$ and $c_{(0)}^M$ are reported in Tables 2.2. In all scenarios, the proposed estimators achieve the smallest biases and standard deviations in Setting 1, where the baseline function and the propensity score are both correctly specified. In Settings 2 and 3, the proposed estimators are nearly unbiased, showing the doubly robust property of the subgroup estimators obtained using the A-learning estimating equations. In Setting 4, where the baseline function and the propensity score are both misspecified, biases and standard deviations of the estimators tend to be larger, however, the biases are still reasonably small. In addition, the coverage probabilities of 95% Wald-type confidence intervals are close to the nominal level for all cases.

To further assess the performance of the proposed maximin OTRs, we compare it with the estimated pooled OTR, $\widehat{d}^P(x) = \mathbb{I}(x^T \widehat{\boldsymbol{\beta}}^P > -\widehat{c}^P)$ and the OTR based on random effects models, $\widehat{d}^R(x) = \mathbb{I}(x^T \widehat{\boldsymbol{\beta}}^R > -\widehat{c}^R)$. Here, $\widehat{\boldsymbol{\beta}}^P$ and $\widehat{c}^P$ are obtained based on pooled data by solving a single A-learning estimating equation. To obtain $\widehat{\boldsymbol{\beta}}^R$ and $\widehat{c}^R$, we first obtain $\widehat{\boldsymbol{\beta}}_g$, $\widehat{c}_g$ by solving A-learning estimating equations, based on $\{X_{j,g}, A_{j,g}, Y_{j,g}\}_{j=1}^m$. The covariance of $(\widehat{\boldsymbol{\beta}}_g^T, \widehat{c}_g)^T$ is estimated by the sandwich estimator. Based on these estimators, we calculate $\widehat{\boldsymbol{\beta}}^R$ and $\widehat{c}^R$ using the R package `mvmeta`. The between-group covariance matrix is estimated by the method of moments. For both scenarios, we consider the following leave-one-group-out cross-validation procedure for evaluation. We first obtain estimators $\widehat{\boldsymbol{\beta}}^M$, $\widehat{c}^M$, $\widehat{\boldsymbol{\beta}}^P$, $\widehat{c}^P$, $\widehat{\boldsymbol{\beta}}^R$ and $\widehat{c}^R$ based on pooled samples of any three groups. Then, we evaluate the PCD and the VD as defined in Section 3.1 under the obtained maximin OTR and the pooled OTR for the remaining testing group, using Monte Carlo simulations based on the true model for the testing group.

Table 2.3 and 2.4 summarize the results of the VD for Scenario 1 and Scenario 2. The OTR obtained by random effects meta-analyses is close to the estimated pooled OTR in both scenarios. The results of the PCD are given in Table 2.5 and 2.6. In Scenario 1, both the PCD and the VD under our maximin OTR are much higher than those under the other two OTRs for all the testing groups. Taking PCD as an example, on average, the PCD under the maximin OTR is approximately $5 \sim 6\%$ higher than those under the other OTRs. This demonstrates the advantages of the proposed maximin-projection learning when there is relatively large heterogeneity in optimal treatment decision-making across subgroups. In Scenario 2, since the groupwise optimal treatment regimes are “close" to each other in “angles", all the estimated OTRs do not differ much. From Table 2.4, it can be seen that our maximin OTR performs better than the other OTRs when the first group is taken as the testing group, while it has comparable performance with the other OTRs for other groups as testing groups.

Some additional simulation experiments with non-normal covariates can be found in Shi et al. (2018c). Findings are similar to those with normal covariates.

Table 2.3: VD results (with standard errors in parenthesis) for Scenario 1 under the estimated maximin OTR $\widehat{d}_M$, the pooled OTR $\widehat{d}_P$ and the OTR obtained by random effects meta-analyses $\widehat{d}_R$.

Testing group		First group	Second group	Third group	Fourth group
Setting 1	$\widehat{d}_P$	0.407(0.002)	0.606(0.001)	0.632(0.002)	0.368(0.002)
	$\widehat{d}_R$	0.408(0.001)	0.608(0.001)	0.633(0.001)	0.367(0.001)
	$\widehat{d}_M$	0.486(0.001)	0.690(0.001)	0.723(0.001)	0.458(0.001)
Setting 2	$\widehat{d}_P$	0.406(0.002)	0.606(0.002)	0.630(0.002)	0.366(0.002)
	$\widehat{d}_R$	0.407(0.001)	0.608(0.001)	0.633(0.001)	0.366(0.001)
	$\widehat{d}_M$	0.483(0.002)	0.689(0.001)	0.719(0.001)	0.452(0.002)
Setting 3	$\widehat{d}_P$	0.407(0.003)	0.604(0.002)	0.630(0.002)	0.367(0.003)
	$\widehat{d}_R$	0.405(0.002)	0.606(0.001)	0.632(0.001)	0.367(0.002)
	$\widehat{d}_M$	0.483(0.002)	0.688(0.001)	0.723(0.001)	0.454(0.002)
Setting 4	$\widehat{d}_P$	0.406(0.003)	0.602(0.003)	0.628(0.003)	0.365(0.003)
	$\widehat{d}_R$	0.406(0.002)	0.606(0.001)	0.632(0.001)	0.366(0.002)
	$\widehat{d}_M$	0.473(0.003)	0.686(0.002)	0.716(0.001)	0.439(0.004)

Table 2.4: VD results (with standard errors in parenthesis) for Scenario 2 under the estimated maximin OTR $\widehat{d}_M$, the pooled OTR $\widehat{d}_P$ and the OTR obtained by random effects meta-analyses $\widehat{d}_R$.

Testing group		First group	Second group	Third group	Fourth group
Setting 1	$\widehat{d}_P$	0.803(<0.001)	0.597(<0.001)	0.865(<0.001)	0.762(<0.001)
	$\widehat{d}_R$	0.803(<0.001)	0.598(<0.001)	0.865(<0.001)	0.761(<0.001)
	$\widehat{d}_M$	0.847(<0.001)	0.588(<0.001)	0.865(<0.001)	0.769(<0.001)
Setting 2	$\widehat{d}_P$	0.802(0.001)	0.597(<0.001)	0.864(<0.001)	0.761(<0.001)
	$\widehat{d}_R$	0.803(<0.001)	0.598(<0.001)	0.865(<0.001)	0.762(<0.001)
	$\widehat{d}_M$	0.843(0.001)	0.587(<0.001)	0.863(<0.001)	0.767(0.001)
Setting 3	$\widehat{d}_P$	0.801(0.001)	0.597(<0.001)	0.863(<0.001)	0.760(0.001)
	$\widehat{d}_R$	0.801(0.001)	0.597(<0.001)	0.864(<0.001)	0.761(0.001)
	$\widehat{d}_M$	0.841(0.001)	0.588(<0.001)	0.861(0.001)	0.765(0.001)
Setting 4	$\widehat{d}_P$	0.799(0.001)	0.595(<0.001)	0.861(0.001)	0.758(0.001)
	$\widehat{d}_R$	0.804(0.001)	0.597(<0.001)	0.863(<0.001)	0.759(0.001)
	$\widehat{d}_M$	0.826(0.002)	0.587(0.001)	0.853(0.001)	0.756(0.002)

Table 2.5: The PCD results (%, with standard errors in parenthesis) for Scenario 1 under the estimated maximin OTR $\widehat{d}_M$, the pooled OTR $\widehat{d}_P$ and the OTR estimated by random effects meta-analyses $\widehat{d}_R$.

Testing group		First group	Second group	Third group	Fourth group
Setting 1	$\widehat{d}_P$	67.1(0.1)	77.6(0.1)	79.2(0.1)	65.3(0.1)
	$\widehat{d}_R$	67.1(<0.1)	77.6(<0.1)	79.2(<0.1)	65.2(<0.1)
	$\widehat{d}_M$	70.9(0.1)	83.3(0.1)	86.1(0.1)	69.5(0.1)
Setting 2	$\widehat{d}_P$	67.1(0.1)	77.6(0.1)	79.1(0.1)	65.2(0.1)
	$\widehat{d}_R$	67.0(<0.1)	77.6(<0.1)	79.2(<0.1)	65.2(<0.1)
	$\widehat{d}_M$	70.8(0.1)	83.3(0.1)	85.9(0.1)	69.3(0.1)
Setting 3	$\widehat{d}_P$	67.1(0.1)	77.6(0.1)	79.2(0.1)	65.3(0.1)
	$\widehat{d}_R$	67.0(0.1)	77.5(0.1)	79.2(0.1)	65.2(0.1)
	$\widehat{d}_M$	70.8(0.1)	83.2(0.1)	86.2(0.1)	69.3(0.1)
Setting 4	$\widehat{d}_P$	67.1(0.2)	77.6(0.2)	79.3(0.2)	65.3(0.2)
	$\widehat{d}_R$	67.1(0.1)	77.5(0.1)	79.2(0.1)	65.2(0.1)
	$\widehat{d}_M$	70.4(0.2)	83.2(0.1)	85.7(0.1)	68.7(0.2)

Table 2.6: The PCD results (%, with standard errors in parenthesis) for Scenario 2 under the estimated maximin OTR $\widehat{d}_M$, the pooled OTR $\widehat{d}_P$ and the OTR estimated by random effects meta-analyses $\widehat{d}_R$.

Testing group		First group	Second group	Third group	Fourth group
Setting 1	$\widehat{d}_P$	86.8(<0.1)	98.2(<0.1)	94.6(<0.1)	90.4(<0.1)
	$\widehat{d}_R$	86.8(<0.1)	98.4(<0.1)	94.6(<0.1)	90.4(<0.1)
	$\widehat{d}_M$	91.7(0.1)	94.6(0.1)	94.9(0.1)	91.6(0.1)
Setting 2	$\widehat{d}_P$	86.8(0.1)	97.9(<0.1)	94.5(0.1)	90.3(0.1)
	$\widehat{d}_R$	86.8(<0.1)	98.3(<0.1)	94.6(<0.1)	90.4(<0.1)
	$\widehat{d}_M$	91.5(0.1)	94.6(0.1)	94.8(0.1)	91.5(0.1)
Setting 3	$\widehat{d}_P$	86.8(0.1)	97.8(<0.1)	94.6(0.1)	90.4(0.1)
	$\widehat{d}_R$	86.7(0.1)	97.8(<0.1)	94.6(0.1)	90.4(0.1)
	$\widehat{d}_M$	91.5(0.1)	94.8(0.1)	94.5(0.1)	91.6(0.1)
Setting 4	$\widehat{d}_P$	86.7(0.1)	96.9(0.1)	94.3(0.1)	90.3(0.1)
	$\widehat{d}_R$	87.0(0.1)	97.8(<0.1)	94.4(0.1)	90.2(0.1)
	$\widehat{d}_M$	90.3(0.2)	94.7(0.1)	93.7(0.2)	91.0(0.2)

Although our maximin estimators have better performance for treatment decision making in the above simulation examples, they can have larger variances compared with the random effects models. This is a potential disadvantage of our method.

2.6 Real data applications

2.6.1 Health assessment questionnaire (HAQ) progration data

The health assessment questionnaire progression data comes from an observational study to investigate the influence of early disease modifying antirheumatic drug (DMARD) treatment and its duration for patients with recent onset inflammatory polyarthritis (Farragher et al. 2010). Early DMARDs treatment was routinely used in the management of rheumatoid arthritis (RA). Among conventional DMARDs, Methotrexate is the most widely used one and is now considered a benchmark against new treatments to be used. Previous studies showed that RA patients who have failed to respond to methotrexate may have clinically important improvements if treated with combination DMARDs, such as methotrexate-sulfasalazine-hydroxychloroquine, methotrexate-sulfasalazine-steroids or other Methotrexate combinations (Boers et al. 1997). However, Methotrexate combinations did not work for all RA patients, and they may not add benefits in some patients who were stable on DMARD monotherapy (Symmons et al. 2005). It is of clinical interest to develop individualized OTRs and to know which patients will benefit from treating with Methotrexate combinations. The study sample include 420 patients who were recruited to the study from 1990 to 2000 and were treated with either methotrexate monotherapy or methotrexate combinations. Age, gender, duration of disease, HAQ score, number of swollen joints and number of tender joints were recorded at baseline. We standardize all six covariates such that their sample covariance matrix equals the identity matrix within each group. We compare methotrexate combinations ($A_{0,g} = 1$) with methotrexate monotherapy ($A_{0,g} = 0$). The difference HAQ scores between baseline and 5-year is set to be the response. Here, we classify 420 patients into three groups according to their recruitment time. Specifically, group 1 includes patients enrolled from 1990 to 1992; group 2 includes those enrolled from 1993 to 1996; and group 3 includes those enrolled from 1997 to 2000. Sample sizes of the three groups are 265, 78 and 77, respectively.

In our analysis, we use the last two standardized covariates to fit the contrast function, since the regression coefficients of other variables are not significant. Denoted these two covariates by $X_{j,g}^{(1)}$ and $X_{j,g}^{(2)}$, respectively. For each group g, we consider the following model

$$\mathbb{E}(Y_{j,g}|X_{j,g}, A_{j,g}) = h_g(X_{j,g}) + A_{j,g}(c_0 + \boldsymbol{\beta}_{g1}X_{j,g}^{(1)} + \boldsymbol{\beta}_{g2}X_{j,g}^{(2)}).$$

The parameters $c_0, \boldsymbol{\beta}_{g1}, \boldsymbol{\beta}_{g2}$ are estimated using the A-learning estimating equations as discussed in Section 4.2. Here, a linear model is fitted for the baseline function and a logistic regression model is fitted for the propensity score. When fitting the propensity score model, all six covariates are included. Table 2.7 reports the group-wise estimators obtained using the A-learning estimating equations, suggesting there is some heterogeneity in optimal treatment regimens across three groups.

Table 2.7: Estimators of groupwise OTR (standard errors in paranthesis) for the HAQ data.

	Group 1	Group 2	Group 3
$\widehat{\boldsymbol{\beta}}_{g1}$	0.05(0.11)	−0.40(0.17)	0.07(0.21)
$\widehat{\boldsymbol{\beta}}_{g2}$	0.07(0.11)	0.06(0.21)	0.32(0.16)

Table 2.8: $\widehat{d}_M$, $\widehat{d}_P$, $\widehat{d}_R$ and their value functions

Testing group	Group 1			Group 2			Group 3		
	$\widehat{d}_M$	$\widehat{d}_P$	$\widehat{d}_R$	$\widehat{d}_M$	$\widehat{d}_P$	$\widehat{d}_R$	$\widehat{d}_M$	$\widehat{d}_P$	$\widehat{d}_R$
$\widehat{c}$	−0.87	−0.14	−0.12	−2.38	−0.21	−0.11	−3.08	−0.31	−0.32
$\widehat{\beta}_1$	−0.48	−0.02	−0.00	0.61	0.16	0.16	−0.02	0.06	−0.01
$\widehat{\beta}_2$	0.88	0.25	0.23	0.79	0.10	0.14	1.00	0.06	0.10
$\widehat{\mathbb{E}}Y_g^\star(d)$	−0.08	−0.09	−0.09	−0.09	−0.19	−0.22	−0.12	−0.13	−0.12

We use the same leave-one-group-out cross validation procedure as done in simulations to evaluate the performance of the proposed method. We calculate the maximin OTR $\widehat{d}_M$, the pooled OTR $\widehat{d}_P$, and the OTR obtained by random effects meta-analyses $\widehat{d}_R$ based on every two groups of patients, and evaluate them on the remaining group based on the estimated value function. For a given treatment regime d and group g, the estimated value function is given by

$$\widehat{\mathbb{E}}Y_g^\star(d) = \frac{1}{m_g}\sum_{j=1}^{m_g}\left[Y_{j,g} + \left(\widehat{c}_0 + \widehat{\boldsymbol{\beta}}_{g1}X_{j,g}^{(1)} + \widehat{\boldsymbol{\beta}}_{g2}X_{j,g}^{(2)}\right)\{d(X_{j,g}) - A_{j,g}\}\right],$$

which is computed based on the advantage function as introduced in Murphy (2003). Results are given in Table 2.8. Value under the maximin OTR are uniformly better than those under other OTRs across all three groups, showing a big improvement for group 2. Besides, the estimators involved in the regimes $\widehat{d}_P$ and $\widehat{d}_R$ are very close.

2.6.2 The schizophrenia study

Tarrier et al. (2004) conducted a multi-center, randomized controlled trial with 18-month follow-up , to examine the effects of cognitive-behavioral therapy (CBT) and supportive counselling (SC) on the outcomes of an early episode of schizophrenia. Patients were randomized to three treatment options, including the cognitive-behavioural therapy plus treatment as usual (CBT), supportive counselling plus treatment as usual (SC) and treatment as usual (TAU). The primary outcome, the Positive and Negative Syndromes Schedule (PANSS, Kay et al. 1987), was measured at baseline and the end of follow-up. Patients' durations of untreated psychosis, years of education and social functioning scores were also recorded at baseline.

As previous studies showed that both psychological treatment groups (CBT and SC) had a superior treatment effect compared to the control group (TAU), we focus on comparing two treatment arms: CBT ($A_{0,g} = 1$) and SC ($A_{0,g} = 0$) to determine individual OTRs. The reduction of PANSS score at the 18th month's visit is set as a patient's response $Y_{0,g}$. We consider two covariates: PANSS score at baseline ($X_{0,g}^{(1)}$) and log duration of untreated psychosis ($X_{0,g}^{(2)}$). Over 400 patients were initially enrolled in 3 treatment centres. Among them, only 165 finished the follow-up study and had completed records of the final response and baseline information. 85 of them received CBT or SC. As in Tarrier (2004), we classify 85 patients into 3 groups according to their treatment centres (Manchester, Liverpool and North Nottinghamshire). We first standardize the two covariates such that their sampling covariance matrix equals the identity matrix within each group and then jointly estimate c_0, $\boldsymbol{\beta}_{g1}$, $\boldsymbol{\beta}_{g2}$ by the A-learning estimating equations as discussed previously. Estimators for $\boldsymbol{\beta}_{g1}$ and $\boldsymbol{\beta}_{g2}$ are given in Table 2.9.

Table 2.9: Estimators of groupwise OTR (standard errors in paranthesis) for the CBT study.

	Group 1	Group 2	Group 3
$\widehat{\boldsymbol{\beta}}_{g1}$	1.35(10.21)	1.17(11.52)	−20.71(13.27)
$\widehat{\boldsymbol{\beta}}_{g2}$	7.87(10.39)	−10.56(8.84)	3.45(9.14)

Differences of $\boldsymbol{\beta}_{gi}$ between different groups are not statistically significant. The large standard errors are due to the small sample size of each group. However, some of the estimated coefficients $\widehat{\boldsymbol{\beta}}_{gi}$'s among different groups are not even sign consistent, indicating potential existence of heterogeneity in optimal treatment regimes across different groups.

We adopt the leave-one-group-out cross validation procedure as done in the previous example. We report the estimated maximin OTR $\widehat{d}_M$, the estimated pooled OTR $\widehat{d}_P$, the OTR obtained based on random effects models $\widehat{d}_R$, as well as the corresponding estimated value functions in Table 2.10. All three OTRs have similar value functions for Groups 1 and 2. However, for Group 3, value function under the maximin OTR is much higher than those under other OTRs.

Table 2.10: $\widehat{d}_M$, $\widehat{d}_P$, $\widehat{d}_R$ and their estimated value functions.

Testing group	Group 1			Group 2			Group 3		
	$\widehat{d}_M$	$\widehat{d}_P$	$\widehat{d}_R$	$\widehat{d}_M$	$\widehat{d}_P$	$\widehat{d}_R$	$\widehat{d}_M$	$\widehat{d}_P$	$\widehat{d}_R$
$\widehat{c}$	0.33	−1.13	−1.25	4.25	4.52	3.26	2.72	1.83	1.05
$\widehat{\beta}_1$	0.11	−0.45	−2.79	1.00	−2.42	−2.89	1.00	0.04	0.11
$\widehat{\beta}_2$	−0.99	−0.45	−3.15	−0.07	−3.23	−5.06	−0.01	3.86	4.62
$\widehat{\mathbb{E}}Y_g^\star(d)$	26.25	25.66	25.32	29.92	30.81	32.04	24.01	16.29	14.36

2.7 Discussion

In this chapter, we propose a maximin-projection learning to aggregate OTRs for patients from different populations with heterogeneity. It has appealing statistical interpretations in the sense of maximizing the minimum PCD and the minimum value difference across subgroups. The corresponding estimation procedure is easy to implement via quadratically constrained linear programming, and the asymptotic properties of the resulting estimators are studied.

2.7.1 Alternative maximin formulation

Our procedure requires to scale the baseline covariates $\boldsymbol{X}_{0,g}$ to mean zero and identity covariance matrix for $g = 1, 2, \ldots, G, G+1$. Let $\boldsymbol{X}_{0,g}^{\text{raw}}$ be the original variable prior to transformation and $\boldsymbol{\beta}_g^{\text{raw}}$, c_g^{raw} the corresponding individualized and marginal treatment effects, respectively. The proposed maximin OTR is constructed based on

$$\boldsymbol{\beta}^M = \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\max} \min_{g\in\{1,\ldots,G\}} \boldsymbol{\beta}^T \boldsymbol{\beta}_g,$$

or equivalently,

$$\boldsymbol{\beta}^{M*} = \arg \max_{\|\boldsymbol{\Sigma}_{G+1}^{1/2}\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{G+1}^{1/2} \boldsymbol{\Sigma}_g^{1/2} \boldsymbol{\beta}_g^{\text{raw}},$$

where $\boldsymbol{\Sigma}_g$ is the covariance matrix of $\boldsymbol{X}_{0,g}^{\text{raw}}$ for $g = 1, \ldots, G+1$.

Alternatively, we can also consider the maximin OTR based on $\boldsymbol{\beta}^{M**}$ where

$$\boldsymbol{\beta}^{M**} = \arg \max_{\|\boldsymbol{\Sigma}_{G+1}^{1/2}\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{G+1} \boldsymbol{\beta}_g^{\text{raw}}.$$

Assuming $\mathbb{E}\boldsymbol{X}_{0,1}^{\text{raw}} = \mathbb{E}\boldsymbol{X}_{0,2}^{\text{raw}} = \cdots = \mathbb{E}\boldsymbol{X}_{0,G}^{\text{raw}} = \mathbb{E}\boldsymbol{X}_{G+1,0}^{\text{raw}} = 0$, $c_1^{\text{raw}} = c_2^{\text{raw}} = \cdots = c_G^{\text{raw}}$ and $\boldsymbol{X}_{0,G+1}$ is spherically distributed, we can show

$$\boldsymbol{\beta}^{M**} = \arg \max_{\|\boldsymbol{\Sigma}_{G+1}^{1/2}\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \mathbb{E}(\boldsymbol{X}_{0,G+1}^{\text{raw}T} \boldsymbol{\beta}_g^{\text{raw}} + c_g^{\text{raw}}) \mathbb{I}(X_{G+1}^{\text{raw}T} \boldsymbol{\beta} + c > 0),$$

for any c. This implies that $\boldsymbol{\beta}^{M**}$ maximizes the minimum groupwise value difference function under the new distribution $\boldsymbol{X}_{0,G+1}^{\text{raw}}$.

Below, we briefly compare the proposed maximin OTR with the maximin OTR based on $\boldsymbol{\beta}^{M**}$ and discuss their connections. First, $\boldsymbol{\beta}^{M**}$ maximizes the minimum groupwise value difference function under the new distribution $\boldsymbol{X}_{0,G+1}^{\text{raw}}$ while $\boldsymbol{\beta}^{M}$ maximizes the minimum groupwise value difference function under the new distribution $\boldsymbol{X}_{0,G+1}$ after scaling. To see this, note that when $c_1 = \cdots = c_G$ and $\boldsymbol{X}_{0,G+1}$ is spherically distributed, we have

$$\boldsymbol{\beta}^{M} = \arg\max_{\|\boldsymbol{\beta}\|_2=1} \min_{g\in\{1,\ldots,G\}} \mathbb{E}(\boldsymbol{X}_{0,G+1}^{T} \boldsymbol{\beta}_g + c_g) \mathbb{I}(X_{0,G+1}^{T} \boldsymbol{\beta} + c > 0),$$

for any c. Second, $\boldsymbol{\beta}^{M**}$ usually doesn't coincide with $\boldsymbol{\beta}^{M*}$. A sufficient condition for $\boldsymbol{\beta}^{M*} = \boldsymbol{\beta}^{M**}$ is that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_G = \boldsymbol{\Sigma}_{G+1}$. Lastly, estimating $\boldsymbol{\beta}^{M**}$ might exhibit less variances than $\boldsymbol{\beta}^{M*}$, since it doesn't require the estimation of $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_G$. However, the OTR based on $\boldsymbol{\beta}^{M**}$ is not scale invariant. To see this, let $X_{0,G+1}^{**} = \boldsymbol{C} X_{0,G+1}^{\text{raw}}$ for some invertible matrix $\boldsymbol{C}$. The covariance matrix of $X_{0,G+1}^{\text{raw}}$ is equal to $\boldsymbol{C}\boldsymbol{\Sigma}_{G+1}\boldsymbol{C}^T$. Let

$$\boldsymbol{\beta}^{M***} = \arg \max_{\|\Sigma_{G+1}^{1/2} C^T \boldsymbol{\beta}\|_2=1} \min_{g\in 1,\ldots,G} \boldsymbol{\beta}^T \boldsymbol{C} \boldsymbol{\Sigma}_{G+1} \boldsymbol{C}^T \boldsymbol{\beta}_g^{\text{raw}},$$

there's no guarantee that $\boldsymbol{\beta}^{M***} = (\boldsymbol{C}^T)^{-1} \boldsymbol{\beta}^{M**}$.

2.7.2 Extensions

In current work, we mainly deal with heterogeneity caused by groupwise individualized treatment effects $\boldsymbol{\beta}_g$'s, and assume the same marginal treatment effects c_g for all groups. It is possible to extend our proposed maximin projection learning to the case when c_g's vary across different groups as well. Specifically, consider

$$(\widehat{\boldsymbol{\beta}}^M, \widehat{c}^M) = \arg \max_{\|\beta\|_2^2+c^2\le 1} \min_{g\in\{1,\ldots,G\}} \left(\widehat{\boldsymbol{\beta}}_g^T \boldsymbol{\beta} + \widehat{c}_g c\right),$$

where $\widehat{\boldsymbol{\beta}}_g$ and $\widehat{c}_g$ are subgroup estimators. Statistical properties of $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{c}^M$ can be similarly established. For example, $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{c}^M$ can be shown to converge almost surely to some $\boldsymbol{\beta}_{(0)}^M$ and $c_{(0)}^M$, respectively. However, the defined $\boldsymbol{\beta}_{(0)}^M$ and $c_{(0)}^M$ can no longer preserve the interpretation of maximizing the minimum PCD and the minimum VD, due to the fact that the PCD and the VD are complicated functions of $(\boldsymbol{\beta}_g, c_g)$ and $(\boldsymbol{\beta}, c)$ when c_gs vary across groups. Consequently, the angle interpretation as demonstrated by the toy example given in Section 2.2.2 does not hold.

To establish the consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{c}^M$, we require $\boldsymbol{\beta}_g, g \in K_0$ to be linearly independent. In Shi et al. (2018c), we conduct some additional simulation studies to examine our methods under settings where some of the $\boldsymbol{\beta}_g$'s are the same. Results suggest that $\widehat{\boldsymbol{\beta}}^M$ and $\widehat{c}^M$ are still consistent to $\boldsymbol{\beta}_{(0)}^M$ and $c_{(0)}^M$, in these settings. We further evaluate the VD and the PCD under the estimated maximin OTR and compare them with those under the estimated pooled OTR. Findings are similar to those in Section 6.6.

In addition, in our current work, we assume a linear interaction between treatment and baseline covariates. It is interesting to consider a more general model as follows:

$$h_g(1, \boldsymbol{X}_{0,g}) - h_g(0, \boldsymbol{X}_{0,g}) = Q(\boldsymbol{\beta}_g^T \boldsymbol{X}_{0,g} + c_g) + e_g, \qquad g = 1, \ldots, G, \tag{2.15}$$

where Q is a strictly monotone increasing function with $Q(0) = 0$. The parameters $\boldsymbol{\beta}_g$ in each group can be consistently estimated using the concordance-assisted learning method by Fan et al. (2017). The properties of the corresponding maximin-projection estimator warrant further investigation.

CHAPTER

3

DIVIDE AND CONQUER FOR MASSIVE DATA

3.1 Introduction

The divide and conquer method is a commonly used approach for handling massive data, which divides data into several groups and aggregate all subgroup estimators by a simple average to lessen the computational burden. A number of problems have been studied for the divide and conquer method, including variable selection (Chen and Xie 2014), nonparametric regression (Zhang et al. 2013; Zhao et al. 2016) and bootstrap inference (Kleiner et al. 2014), to mention a few. Most papers establish that the aggregated estimators achieve the oracle result, in the sense that they possess the same nonasymptotic error bounds or limiting distributions as the pooled estimators, which are obtained by fitting all the data in a single model. This implies that the divide and conquer scheme can not only maintain efficiency, but also obtain a feasible solution for analyzing massive data.

Our objective is to estimate the OTR in a data-rich environment. The value-search estimator proposed by Zhang et al. (2012) is robust and appealing. In addition to the computational advantages for handling massive data, the divide and conquer method, somewhat surprisingly, can lead to aggregated estimators with improve efficiency over the original value-search estimator with $n^{1/3}$ convergence rate. This phenomenon is expected to hold under many other cube-root estimation problems. For example, Chernoff (1964) studied a cubic-rate estimator for estimating the mode. It was shown therein that the estimator converges in distribution to the argmax of a Brownian motion minus a quadratic drift. Kim and Pollard (1990) systematically studied a class of cubic-rate M-estimators and established their limiting distributions as the argmax of a general Gaussian process minus a quadratic form. These results were extended to a more general class of M-estimators using modern empirical process results (van der Vaart and Wellner 1996; Kosorok 2008).

In this chapter, we generalize our setup by studying a class of M-estimators with cubic-rate and develop a general inference framework for the aggregated estimators obtained by the divide and conquer method. In particular, we will cover the value-search estimator as a special example. Our theory states that the aggregated estimators can achieve a faster convergence rate than the pooled estimators and have asymptotic normal distributions when the number of groups diverges at a proper rate as the sample size of each group grows. This enables a simple way for estimating the covariance matrix of the aggregated estimators.

When establishing the asymptotic properties of the aggregated estimators, a major technical challenge is to quantify the accumulated bias. Different from estimators with standard $n^{1/2}$ convergence rate, M-estimators with $n^{1/3}$ convergence rate generally do not have a nice linearization representation and the magnitude of the associated biases is difficult to quantify. One way to obtain the magnitude of the bias is by establishing a coupling inequality for the cubic-rate estimator. For example, Banerjee et al. (2019) derived a nonasymptotic bound for the biases of the isotonic estimator in a monotone regression model and its inverse, based on the coupling inequality of the isotonic estimator (see Lemma 8.10 in that chapter, and also Equation (29) in Durot (2002)). Groeneboom et al. (1999) provided a coupling inequality for the inverse process of the Grenander estimator. Their results can be used to establish the bias of the Grenander estimator. While such strategy is useful for studying the bias of some one-dimensional cubic-rate estimators, it is not suitable for multi-dimensional estimators. On one hand, these coupling inequalities are all based on Komos-Major-Tusnady (KMT) approximation (Komlós et al. 1975) and its extensions (Csörgő et al. 1985; Sakhanenko 2006) that only apply to the empirical distribution or the quantile process. There are extensions of the KMT approximation for more general empirical process (Rio 1994; Koltchinskii 1994). However, the rate of the approximation will depend on the dimension of the parameter and decays fast as the dimension increases. On the other hand, proofs of these coupling inequalities all rely on the properties of the argmax of a Brownian motion process with a parabolic drift (see Proposition 1 in Durot (2002) and the discussions therein), and are not applicable to cubic-rate estimators that converge to the argmax of a more general Gaussian process minus a quadratic term. Here, we propose a novel approach to derive an upper bound for the bias, without establishing the coupling inequalities. To the best of our knowledge, this is the first time that a nonasymptotic error bound for the bias of a general cubic-rate estimator is provided.

A key innovation in our analysis is to introduce a linear perturbation in the empirical objective function. In that way, we transform the problem of quantifying the bias into comparison of the expected supremum of the empirical objective function and that of its limiting Gaussian process. To bound the difference of these expected suprema, we adopt similar techniques that have been recently studied by Chernozhukov et al. (2013) and Chernozhukov et al. (2014). Specifically, they compared a function of the maximum for sum of mean-zero Gaussian random vectors with that of multivariate mean-zero random vectors with the same covariance function, and provided an associated coupling inequality. We improve their arguments by providing more accurate approximation results (Lemma 3.6.3) for the identity function of maximums as needed in our applications.

We also provide a tail inequality for cubic-rate M-estimators (Theorem 3.5.1). This helps us to construct a truncated estimator with bounded second moment, which is essential to apply Lyapunov's central limit theorem for establishing the normality of the aggregated estimator. Under some additional tail assumptions on the underlying empirical process, our results can be viewed as a generalization of empirical process theories that establish consistency and $n^{1/3}$ convergence rate for the M-estimators. Based on the results, we show that the asymptotic variance of the aggregated estimator can be consistently estimated by the sample variance of individual M-estimators in each group, which largely simplifies the inference procedure for M-estimators.

The rest of the chapter is organized as follows. We show the value-search estimator belongs to the general class of cubic-rate M-estimators in Section 3.2. We also describe the divide and conquer method for M-estimators and state the major central limit theorem (Theorem 3.2.1). We revisit the value-search estimator in Section 3.3 to illustrate the application of Theorem 3.2.1. In Section 3.4, we demonstrate the empirical performance of the aggregated estimators over the orignal value-search estimator using both simulation studies and an application to the Yahoo! Front Page Today Module user click log dataset. Section 3.5 studies a tail inequality and Section 3.6 provides the analysis of bias of M-estimators that are needed to prove Theorem 3.2.1, followed by a Discussion Section. All other technical proofs can be found in Shi et al. (2018b).

3.2 Method

3.2.1 From value-search estimators to general cubic-rate M-estimators

The value-search estimator was introduced by Zhang et al. (2012) for estimating the OTR. We focus on the setting with a single decision point given in Section 1.2.

The true contrast function $\tau(\cdot)$ can be complex. As suggested by Zhang et al. (2012), in practice we can find the restricted optimal regimen within a class of decision rules, such as linear treatment decision rules $d^{\text{lin}}(\boldsymbol{x}, \boldsymbol{\beta}, c) = \mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta} + c > 0)$ indexed by some $\boldsymbol{\beta} \in \mathbb{R}^p$ and $c \in \mathbb{R}$. Let $(\boldsymbol{\beta}^*, c^*) = \arg\max_{(\boldsymbol{\beta},c)} V(d^{\text{lin}}(\cdot, \boldsymbol{\beta}, c))$ where $V(\cdot)$ stands for the value function. To make $\boldsymbol{\beta}^*$ and c^* identifiable, we assume $c^* \in \{-1, 0, 1\}$. Zhang et al. (2012) proposed an inverse propensity score weighted estimator of the value function $V(d^{\text{lin}}(\cdot, \boldsymbol{\beta}, c))$ and the associated value search estimator by maximizing the estimated value function. Specifically, the value search estimator is defined as

$$(\widehat{\boldsymbol{\beta}}, \widehat{c}) = \arg\max_{\substack{\boldsymbol{\beta}\in\mathbb{R}^p \\ c\in\{-1,0,1\}}} \sum_{i=1}^{n} \frac{\mathbb{I}(A_i = d^{\text{lin}}(\boldsymbol{X}_i, \boldsymbol{\beta}, c))}{n\pi(A_i, \boldsymbol{X}_i)} Y_i.$$

Suppose c^* is uniquely defined,

$$V(d^{\text{lin}}(\cdot, \boldsymbol{\beta}^*, c^*)) > \sup_{\substack{\boldsymbol{\beta}\in\mathbb{R}^p \\ c\in\{-1,0,1\}-\{c^*\}}} V(d^{\text{lin}}(\cdot, \boldsymbol{\beta}, c)),$$

and

$$\sup_{\substack{\boldsymbol{\beta}\in\mathbb{R}^p \\ c\in\{-1,0,1\}}} \left| \sum_{i=1}^{n} \frac{\mathbb{I}(A_i = d^{\text{lin}}(\boldsymbol{X}_i, \boldsymbol{\beta}, c))}{n\pi(A_i, \boldsymbol{X}_i)} Y_i - V(d^{\text{lin}}(\cdot, \boldsymbol{\beta}, c)) \right| = o_p(1). \tag{3.1}$$

Then we have $\widehat{c} = c^*$ and

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n} \frac{\mathbb{I}(A_i = d^{\text{lin}}(\boldsymbol{X}_i, \boldsymbol{\beta}, c^*))}{n\pi(A_i, \boldsymbol{X}_i)} Y_i, \tag{3.2}$$

with probability tending to 1. Here, Condition (3.1) implies that the class of functions

$\{\pi^{-1}(A_0, \boldsymbol{X}_0)\mathbb{I}(A_0 = d^{\text{lin}}(\boldsymbol{X}_0, \beta, c))Y_0 : \boldsymbol{\beta} \in \mathbb{R}^p, c \in \{-1, 0, 1\}\}$ belongs to the class of Glivenko-Cantelli class.

The optimization in (3.2) can be further generalized to the following M-estimation problems

$$\widetilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} m(\boldsymbol{O}_i, \boldsymbol{\theta}), \tag{3.3}$$

where $\boldsymbol{O}_i$s are some i.i.d random vectors, $m(\cdot, \cdot)$ is some non-smooth objective function that involves indicators and $\boldsymbol{\theta}$ is a d-dimensional vector of parameters that belongs to a compact parameter space $\boldsymbol{\theta}$.

3.2.2 Divide and conquer

The divide and conquer scheme for M-estimators is described as follows. In the first step, the data $\{\boldsymbol{O}_i\}_{i=1}^N$ are randomly divided into several groups $\{\boldsymbol{O}_{i,j}\}_{i,j}$. For the jth group, consider the following M-estimator

$$\widehat{\boldsymbol{\theta}}_j = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\theta}} \frac{1}{n_j} \sum_{i=1}^{n_j} m(\boldsymbol{O}_{i,j}, \boldsymbol{\theta}), \qquad j = 1, \ldots, S,$$

where $(\boldsymbol{O}_{1,j}, \ldots, \boldsymbol{O}_{n_j,j})$ denote the data for the jth group, n_j is the number of observations in the jth group, S is the number of groups. In the second step, the aggregated estimator $\widehat{\boldsymbol{\theta}}$ is obtained as a weighted average of all subgroup estimators,

$$\widehat{\boldsymbol{\theta}} = \sum_{j=1}^{S} \omega_j \widehat{\boldsymbol{\theta}}_j = \frac{\sum_{j=1}^{S} n_j^{2/3} \widehat{\boldsymbol{\theta}}_j}{\sum_{j=1}^{S} n_j^{2/3}}. \tag{3.4}$$

The weights ω_j's are chosen such that $\widehat{\boldsymbol{\theta}}$ achieves the smallest asymptotic covariance matrix among the class of linearly aggregated estimators $\{\boldsymbol{\theta}_\omega = \sum_j \omega_j \widehat{\boldsymbol{\theta}}_j | \sum_j \omega_j = 1, \omega_j \geq 0, \forall j = 1, \ldots, S\}$ (see Section F of Shi et al. (2018b) for detailed illustrations). When $n_1 = n_2 = \cdots = n_S$, $\widehat{\boldsymbol{\theta}}$ reduces to a simple average of all $\widehat{\boldsymbol{\theta}}_j$'s.

We assume that all the $\boldsymbol{O}_{i,j}$'s are independent and identically distributed across i and j. Here, we only consider M-estimation with non-smooth functions $m(\cdot, \boldsymbol{\theta})$ of $\boldsymbol{\theta}$, and the resulting M-estimators $\widehat{\boldsymbol{\theta}}_j$'s have a convergence rate of $O_p(n_j^{-1/3})$. Such cubic-rate M-estimators have been widely studied in the literature, for example, the location estimator and maximum score estimator as demonstrated in the next section. Let $N = \sum_j n_j$ and $n = N/S$.

The main goal of this chapter is to establish the convergence rate and asymptotic normality of $\widehat{\boldsymbol{\theta}}$ under suitable conditions for S and n_j's.

Before introducing our main results, we first provide an intuitive explanation here why the divide and conquer method can improve the efficiency in cubic-rate M-estimation problems. Assume for now, $n_1 = n_2 = \cdots = n_S = n$ and S is fixed. Following Kim and Pollard (1990), we can show that

$$n^{1/3}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0) \xrightarrow{d} h_0,$$
$$N^{1/3}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} h_0,$$

where $\widetilde{\boldsymbol{\theta}}$ is the pooled estimator defined in (3.3), $\boldsymbol{\theta}_0$ is the unique maximizer of $\mathbb{E}\{m(\cdot,\boldsymbol{\theta})\}$ and $\boldsymbol{h}_0 = \arg\max_{\boldsymbol{h}} Z(\boldsymbol{h})$ with

$$Z(\boldsymbol{h}) = G(\boldsymbol{h}) - \frac{1}{2}\boldsymbol{h}^T \boldsymbol{V} \boldsymbol{h}. \tag{3.5}$$

Here G is a mean-zero Gaussian process and $\boldsymbol{V} = \partial^2 \mathbb{E}\{m(\cdot,\boldsymbol{\theta})\}/\partial \boldsymbol{\theta}\boldsymbol{\theta}^T|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is a positive definite matrix.

Assume $\|N^{1/3}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|_2^2$ and $\|n^{1/3}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)\|_2^2$ are uniformly integrable. Then, we have

$$N^{2/3}\mathbb{E}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \to \mathbb{COV}(h_0), \text{ as } N \to \infty \tag{3.6}$$
$$n^{2/3}\mathbb{E}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)^T \to \mathbb{COV}(h_0), \text{ as } N \to \infty.$$

Under equal allocation, $\widehat{\boldsymbol{\theta}}_j$'s are independent and identical. We have

$$\begin{aligned} & N^{2/3}\mathbb{E}\{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\} \\ = \; & N^{2/3}\frac{1}{S^2}\sum_{j=1}^{S}\mathbb{E}\{(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)^T\} + N^{2/3}\frac{1}{S^2}\sum_{j\neq k}\mathbb{E}\{(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)\mathbb{E}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0)^T\} \\ = \; & \frac{n^{2/3}}{S^{1/3}}\mathbb{E}\{(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T\} + \boldsymbol{b}_n\boldsymbol{b}_n^T S^{2/3}(S-1)/S \to S^{-1/3}\mathbb{COV}(h_0), \end{aligned} \tag{3.7}$$

where $\boldsymbol{b}_n = n^{1/3}\mathbb{E}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0) = o(1)$ is the bias of $n^{1/3}\widehat{\boldsymbol{\theta}}_j$. Comparing (7.12) with (3.7), we can see that the aggregated estimator is more efficient than the pooled estimator in the fixed S scenario.

Now let S grow with N. As long as S satisfies $S = O(1/(\|\boldsymbol{b}_n\|_2^2))$, we have

$$\boldsymbol{b}_n\boldsymbol{b}_n^T S^{2/3}(S-1)/S = O(S^{-1/3}),$$

and hence $N^{2/3}\mathbb{E}\{(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\} = O(S^{-1/3})$. In view of (7.12), this implies that the aggregated estimator can have a faster convergence rate than the pooled estimator.

3.2.3 Main results

We assume the dimension d is fixed, while the number of groups $S \to \infty$ as $N \to \infty$. Let $\|\cdot\|_2$ denote the Euclidean norm for vectors or induced matrix L_2 norm for matrices. We first introduce some conditions.

(A1.) There exists a small neighborhood $N_\delta = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2 \le \delta\}$ in which $\mathbb{E}m\{(\cdot,\boldsymbol{\theta})\}$ is twice continuously differentiable with the Hessian matrix $-V(\boldsymbol{\theta})$, where $V(\boldsymbol{\theta})$ is positive definite in N_δ. Moreover, assume $\mathbb{E}\{m(\cdot,\boldsymbol{\theta}_0)\} > \sup_{\boldsymbol{\theta}\in N_\delta^c} \mathbb{E}\{m(\cdot,\boldsymbol{\theta})\}$.
(A2.) For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in N_\delta$, we have $\mathbb{E}\{|m(\cdot,\boldsymbol{\theta}_1)-m(\cdot,\boldsymbol{\theta}_2)|^2\} \le K\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2$ for a constant K that is independent of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$.
(A3.) There exists some positive constant ω such that $|m(x,\boldsymbol{\theta})| \le \omega$ for all x and $\boldsymbol{\theta}$.
(A4.) The envelope function $M_R(\cdot) \equiv \sup_{\boldsymbol{\theta}}\{|m(\cdot,\boldsymbol{\theta})| : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2 \le R\}$ satisfies $\mathbb{E}M_R^2 = O(R)$ when $R \le \delta$.
(A5.) The set of functions $\{m(\cdot,\boldsymbol{\theta}) | \boldsymbol{\theta} \in \boldsymbol{\theta}\}$ has Vapnik-Chervonenkis (VC) index $1 \le v < \infty$.
(A6.) For any $\boldsymbol{\theta} \in N_\delta$, $\|V(\boldsymbol{\theta}) - V\|_2 = O(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2)$, where $V = V(\boldsymbol{\theta}_0)$.
(A7.) Let $L(\cdot)$ denote the variance process of $G(\cdot)$ satisfying $L(\boldsymbol{h}) > 0$ whenever $\boldsymbol{h} \ne 0$. (i) The function $L(\cdot)$ is symmetric and continuous, and has the rescaling property: $L(k\boldsymbol{h}) = kL(\boldsymbol{h})$ for $k > 0$. (ii) For any $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^d$ satisfying $\|\boldsymbol{h}_1\|_2 \le n^{1/3}\delta$ and $\|\boldsymbol{h}_2\|_2 \le n^{1/3}\delta$, we have

$$\left| L(\boldsymbol{h}_1 - \boldsymbol{h}_2) - n^{1/3}\mathbb{E}\left\{m(\cdot,\boldsymbol{\theta}_0 + n^{-1/3}\boldsymbol{h}_1) - m(\cdot,\boldsymbol{\theta}_0 + n^{-1/3}h_2)\right\}^2 \right| = O\left(\frac{(\|\boldsymbol{h}_1\| + \|\boldsymbol{h}_2\|)^2}{n^{1/3}}\right).$$

(A8.) Let $c_j = n_j/n$. Assume there exists some constant $\bar{c} > 1$ such that $1/\bar{c} \le c_j \le \bar{c}$ for all j.

Theorem 3.2.1 *Under Conditions A1-A8, if $S = o(n^{1/6}/\log^{5/6} n)$ and $S \to \infty$ as $n \to \infty$, we have*

$$\sqrt{c_1^{2/3} + \cdots + c_S^{2/3}}\, n^{1/3}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{A}), \tag{3.8}$$

for some positive definite matrix $\boldsymbol{A}$.

Under Condition A8, Theorem 3.2.1 suggests that $\widehat{\boldsymbol{\theta}}$ converges at a rate of $O_p(S^{-1/2}n^{-1/3})$. In contrast, the original M-estimator obtained based on pooled data has a convergence rate of $O_p(S^{-1/3}n^{-1/3})$. This implies that we can gain efficiency by adopting the split and conquer scheme for cubic-rate M-estimators. Such result is interesting as most aggregated estimators in the divide and conquer literature share the same convergence rates as the original estimators based on pooled data.

The constraints on S suggest that the number of group cannot diverge too fast. A main reason as we showed in the proof of Theorem 3.2.1 is that if S grows too fast, the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ will fail due to accumulation of bias in the aggregation of subgroup estimators. Given a data of size N, we can take $S \approx N^l$, $n = N/S \approx N^{1-l}$ with $l < 1/7$ to fulfill this requirement. It turns out that this requirement on S can be relaxed under some special cases. In particular, when $d = 1$, i.e, $\boldsymbol{\theta}_0$ is a scalar, the aggregated estimator is asymptotically normal as long as $S \le N^l$ with $l < 4/13$. See Section A.5 of Shi et al. (2018b) for details.

Conditions A1 - A5 and A7 (i) are similar to those in Kim and Pollard (1990) and are used to establish the cubic-rate convergence of the M-estimator in each group. Conditions A6 and A7 (ii) are used to establish the normality of the aggregated estimator. In particular, Condition A7 (ii) implies that the Gaussian process $G(\cdot)$ has stationary increments, i.e. $\mathbb{E}[\{G(\boldsymbol{h}_1)-G(\boldsymbol{h}_2)\}^2] = L(\boldsymbol{h}_1-\boldsymbol{h}_2)$ for any $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^d$, which is used to control the bias of the aggregated estimator. Condition A8 automatically holds when $n_1 = \cdots = n_S$.

In the rest of this section, we give a sketch for the proof of Theorem 3.2.1. The details of the proof are given in Section 3.5 and Section 3.6. Let $\widehat{\boldsymbol{h}}_j = n_j^{1/3}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)$. By definition, it is equivalent to show

$$\frac{1}{\sqrt{c_1^{2/3} + \cdots + c_S^{2/3}}} \sum_{j=1}^{S} c_j^{1/3} \widehat{\boldsymbol{h}}_j \stackrel{d}{\to} N(0, \boldsymbol{A}). \tag{3.9}$$

When S diverges, intuitively, (3.9) follows by a direct application of central limit theorem for triangular arrays (see for example, Theorem 11.1.1, Athreya and Lahiri 2006). However, a few challenges remain. First, the estimator $\widehat{\boldsymbol{h}}_j$ may not possess finite second moment. Analogous to Kolmogorov's 3-series theorem (see for example, Theorem 8.3.5, Athreya and Lahiri 2006), we handle this by first defining $\widetilde{\boldsymbol{h}}_j$, which is a truncated version of $\widehat{\boldsymbol{h}}_j$ with $\|\widetilde{\boldsymbol{h}}_j\|_2 \le \delta_{n_j}$ for some $\delta_{n_j} > 0$, such that $\sum_j \widehat{\boldsymbol{h}}_j$ and $\sum_j \widetilde{\boldsymbol{h}}_j$ are tail equivalent, i.e.

$$\lim_k \mathbb{P}\left(\bigcap_{n \ge k} \left\{ \sum_{j=1}^{S(n)} c_j^{1/3} \widehat{\boldsymbol{h}}_j = \sum_{j=1}^{S(n)} c_j^{1/3} \widetilde{\boldsymbol{h}}_j \right\} \right) = 1.$$

Using Borel-Cantelli lemma, it suffices to show

$$\sum_{n} \mathbb{P}\left(\sum_{j=1}^{S(n)} c_j^{1/3} \widehat{\boldsymbol{h}}_j \neq \sum_{j=1}^{S(n)} c_j^{1/3} \widetilde{\boldsymbol{h}}_j \right) < \infty. \tag{3.10}$$

Now it remains to show

$$\frac{1}{\sqrt{\sum_j c_j^{2/3}}} \sum_{j=1}^{S} c_j^{1/3} \widetilde{\boldsymbol{h}}_j = \frac{1}{\sqrt{\sum_j c_j^{2/3}}} \sum_{j=1}^{S} \{ \widetilde{\boldsymbol{h}}_j - \mathbb{E}(\widetilde{\boldsymbol{h}}_j) \} + \frac{1}{\sqrt{\sum_j c_j^{2/3}}} \sum_{j} \mathbb{E} c_j^{1/3} \widetilde{\boldsymbol{h}}_j \xrightarrow{d} N(0, \boldsymbol{A}).$$

The second challenge is to control the accumulated bias in the aggregated estimator, i.e. showing

$$\frac{1}{\sqrt{\sum_j c_j^{2/3}}} \sum_{j} c_j^{1/3} \mathbb{E}(\widetilde{\boldsymbol{h}}_j) \to 0,$$

or

$$\sqrt{S} \sup_j |\mathbb{E}(\widetilde{\boldsymbol{h}}_j)| \to 0, \tag{3.11}$$

by Assumption A8. Finally, it remains to show that the second and third moments of $\widetilde{\boldsymbol{h}}_j$ satisfies

$$\sup_j |\mathbb{E}(\boldsymbol{a}^T \widetilde{\boldsymbol{h}}_j)^2 - \boldsymbol{a}^T \boldsymbol{A} \boldsymbol{a}| \to 0, \tag{3.12}$$

$$\sup_j \mathbb{E} \|\widetilde{\boldsymbol{h}}_j\|_2^3 < \infty, \tag{3.13}$$

for any $a \in \mathbb{R}^d$. When (3.10), (3.11), (3.12) and (3.13) are established, Theorem 3.2.1 follows by Lyapunov's central limit theorem (see Corollary 11.1.4 Athreya and Lahiri 2006). Section 3.5 is devoted to verifying (3.10), (3.12) and (3.13), while Section 3.6 is devoted to proving (3.11).

3.3 Application to estimating the OTR

In this section, we revisit the value-search estimator defined in (3.2). Applications to other cubic-rate M-estimators, including simple one-dimensional location estimator and more complicated maximum score estimator can be found in Shi et al. (2018b). We assume

$n_1 = n_2 = \cdots = n_S = n$ so that A8 automatically holds. Suppose the observed data for the jth group is summarized as $\{(\boldsymbol{X}_{i,j}, A_{i,j}, Y_{i,j})\}_{i=1,\ldots,n}$. Without loss of generality, assume $c^* = -1$. Define $\boldsymbol{\theta}^* = \boldsymbol{\beta}^*$ and

$$\widehat{\boldsymbol{\theta}}_j = \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^p} \sum_{i=1}^{n} \frac{\mathbb{I}(A_{i,j} = d^{\text{lin}}(\boldsymbol{X}_{i,j}, \boldsymbol{\theta}, -1))}{n\pi(A_{i,j}, \boldsymbol{X}_{i,j})} Y_{i,j}. \tag{3.14}$$

Let $\pi_{i,j} = \mathbb{P}(A_{i,j} = 1 | \boldsymbol{X}_{i,j})$. Define $m(O_{i,j}, \boldsymbol{\theta}) = \xi_{i,j} d^{\text{lin}}(\boldsymbol{X}_{i,j}, \boldsymbol{\theta}, -1)$, where

$$\xi_{i,j} = \left(\frac{A_{i,j}}{\pi_{i,j}} - \frac{1 - A_{i,j}}{1 - \pi_{i,j}}\right) Y_{i,j}.$$

With some algebra, we can show that $\widehat{\boldsymbol{\theta}}_j$ also maximizes $\sum_{i=1}^n m(O_{i,j}, \boldsymbol{\theta})/n$. To fulfill (A3), we need $0 < \gamma_1 < \pi_{i,j} < 1 - \gamma_1 < 1$ and $|Y_{i,j}| \le \gamma_2$ for some constants $\gamma_1, \gamma_2 > 0$.

To show (A1) and (A6), we evaluate the integral

$$\Gamma(\boldsymbol{\theta}) = \mathbb{E}\{\xi_0 d^{\text{lin}}(\boldsymbol{X}_0, \boldsymbol{\theta})\} = \mathbb{E}\{\tau(\boldsymbol{X}_0) d^{\text{lin}}(\boldsymbol{X}_0, \boldsymbol{\theta})\} = \int_{\boldsymbol{x}^T\boldsymbol{\theta}>1} \tau(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}, \tag{3.15}$$

where $\xi_0 = \{\pi^{-1}(1, \boldsymbol{X}_0)A_0 - \pi^{-1}(0, \boldsymbol{X}_0)(1 - A_0)\} Y_0$ and $p(\cdot)$ is the density function of $\boldsymbol{X}_0$. Consider the transformation

$$T_{\boldsymbol{\theta}} = (\boldsymbol{I} - \|\boldsymbol{\theta}\|_2^{-2} \boldsymbol{\theta}\boldsymbol{\theta}^T) + \|\boldsymbol{\theta}\|_2^{-2} \boldsymbol{\theta}(\boldsymbol{\theta}^*)^T,$$

which maps the region $\{\boldsymbol{x}^T\boldsymbol{\theta}^* > 1\}$ onto $\{\boldsymbol{x}^T\boldsymbol{\theta} > 1\}$, and $\{\boldsymbol{x}^T\boldsymbol{\theta}^* = 1\}$ onto $\{\boldsymbol{x}^T\boldsymbol{\theta} = 1\}$. We exclude the trivial case with $\boldsymbol{\theta}^* = 0$. The above definition is meaningful when $\boldsymbol{\theta}$ is taken over a small neighborhood N_δ of $\boldsymbol{\theta}^*$. We assume that functions p and τ are continuously differentiable. Note that

$$\frac{\partial T_{\boldsymbol{\theta}} \boldsymbol{x}}{\partial \boldsymbol{\theta}} = -\frac{\{\boldsymbol{\theta}^T \boldsymbol{x} - (\boldsymbol{\theta}^*)^T \boldsymbol{x}\}}{\|\boldsymbol{\theta}\|_2^2} I - \frac{\boldsymbol{\theta}\boldsymbol{x}^T}{\|\boldsymbol{\theta}\|_2^2} + \frac{2\boldsymbol{\theta}\boldsymbol{\theta}^T(\boldsymbol{x}^T\boldsymbol{\theta} - \boldsymbol{x}^T\boldsymbol{\theta}^*)}{\|\boldsymbol{\theta}\|_2^4}.$$

Using some differential geometry arguments similarly as in Section 5 of Kim and Pollard (1990), we can show that the integral (3.15) can be represented as

$$\Gamma(\boldsymbol{\theta}) = \int_{\boldsymbol{x}^T\boldsymbol{\theta}^*>1} \left(-\frac{1}{\|\boldsymbol{\theta}\|_2^2} \boldsymbol{\theta}^T \frac{\partial \tau(\boldsymbol{x}) p(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{x} + \frac{\{\boldsymbol{\theta}^T \boldsymbol{x} - (\boldsymbol{\theta}^*)^T \boldsymbol{x}\}}{\|\boldsymbol{\theta}\|_2^4} \boldsymbol{\theta}^T \frac{\partial \tau(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{\theta} \right.$$
$$\left. - \frac{\boldsymbol{\theta}^T \boldsymbol{x} - (\boldsymbol{\theta}^*)^T \boldsymbol{x}}{\|\boldsymbol{\theta}\|_2^2} \frac{\partial \tau(\boldsymbol{x}) p(\boldsymbol{x})}{\partial \boldsymbol{x}} \right) d\boldsymbol{x},$$

which is thrice differentiable under certain conditions on $\tau(\boldsymbol{x})$, $p(\boldsymbol{x})$ and their derivatives.

To show A7, we assume that the conditional density $p(\boldsymbol{x}|z)$ of $\boldsymbol{x}$ given $Z_0 = 1-\boldsymbol{X}_0^T\boldsymbol{\theta}^*$ exists and is continuously differentiable with respect to z. Similarly assume that the density $q(z)$ of Z_0 exists and is continuously differentiable. Let $g(\boldsymbol{X}_0) = \mathbb{E}(\xi_0^2|\boldsymbol{X}_0)$. For any $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^d$, we have

$$\begin{aligned} & n^{1/3}\mathbb{E}\left\{\xi_0^2\left|\mathbb{I}(\boldsymbol{X}_0^T\boldsymbol{\theta}^* + n^{-1/3}\boldsymbol{X}_0^T\boldsymbol{h}_1 > 1) - \mathbb{I}(\boldsymbol{X}_0^T\boldsymbol{\theta}^* + n^{-1/3}\boldsymbol{X}_0^T\boldsymbol{h}_2 > 1)\right|^2\right\} \\ = \; & n^{1/3}\int g(\boldsymbol{x})\left|\mathbb{I}(n^{-1/3}\boldsymbol{x}^T h_1 > z) - \mathbb{I}(n^{-1/3}\boldsymbol{x}^T\boldsymbol{h}_2 > z)\right| p(\boldsymbol{x}|z)q(z)d\boldsymbol{x}dz. \end{aligned}$$

Let $w = n^{1/3}z$. The last expression in the above equation can be written as

$$\begin{aligned} & \int g(\boldsymbol{x})\left|\mathbb{I}(\boldsymbol{x}^T\boldsymbol{h}_1 > w) - \mathbb{I}(\boldsymbol{x}^T\boldsymbol{h}_2 > w)\right| p(x|0)q(0)d\boldsymbol{x}dw + R \\ = \; & \int g(\boldsymbol{x})|\boldsymbol{x}^T(\boldsymbol{h}_1 - \boldsymbol{h}_2)|p(\boldsymbol{x}|0)q(0)d\boldsymbol{x} + R, \end{aligned}$$

with the remainder term

$$R = \int g(\boldsymbol{x})\left|\mathbb{I}(\boldsymbol{x}^T\boldsymbol{h}_1 > w) - \mathbb{I}(\boldsymbol{x}^T\boldsymbol{h}_2 > w)\right| \{p(\boldsymbol{x}|n^{-1/3}w)q(n^{-1/3}w) - p(x|0)q(0)\}d\boldsymbol{x}dw,$$

which is $O(n^{-1/3}(\|h_1\|_2^2 + \|h_2\|_2^2))$ under certain conditions on $q(\boldsymbol{x})$ and $p(\boldsymbol{x}|\cdot)$. Conditions A2 and A4 can be similarly verified. Since the class of functions $\{g(\boldsymbol{x})\mathbb{I}(\boldsymbol{x}^T\boldsymbol{\theta} > 1) : \boldsymbol{\theta} \in \mathbb{R}^d\}$ has finite VC index, Condition A5 also holds. Theorem 3.2.1 then follows.

3.4 Numerical studies

In this section, we examine the numerical performance of the aggregated value-search estimator and compare it with the original value-search estimator based on pooled data, denoted as the pooled estimator. Some additional simulation studies regarding other types of cubic-rate M-estimation applications can be found in Shi et al. (2018b).

3.4.1 Simulations

Consider the model $Y_{i,j} = 1 + A_{i,j}(2X_{i,j} - 1) + e_{i,j}$, $i = 1, \cdots, n$, $j = 1, \ldots, S$, where $X_{i,j} \sim N(0,1)$, $e_{i,j} \sim N(0, 0.25)$, and $\mathbb{P}(A_{i,j} = 1) = 0.5$. Under this model assumption, the optimal treatment rule takes the form,

$$d^{opt}(x) = \mathbb{I}(2x > 1),$$

and hence $\theta^* = 2$.

We take $N = 2^{24}, 2^{25}, 2^{26}$ and 2^{27}. When $N = 2^{24}$ and 2^{25}, we choose $S = 2^j$ for $j = 4,5,6,7$. When $N = 2^{26}$ and 2^{27}, we choose $S = 2^j$ for $j = 5,6,7,8$. This gives a total of 16 scenarios. For each combination of N and S, we estimated the standard error of $\widehat{\theta}$ by

$$\widehat{SE}(\widehat{\theta}) = \frac{1}{\sqrt{S}} \left\{ \frac{1}{S-1} \sum_{l=1}^{S} \left(\widehat{\theta}_l - \widehat{\theta} \right)^2 \right\}^{1/2}.$$

We plot the coverage probabilities (denoted by CP) of 95% confidence intervals for $\boldsymbol{\theta}_0$ in Figure 3.1, with these combinations of S and N. When $S \le S^* \approx N^{0.27}$, the coverage probabilities are close to 95%. This verifies our theoretical findings which state that with properly chosen S, the estimated standard error of $\widehat{\theta}$ is close to its standard deviation and the coverage probability is close to the nominal level.

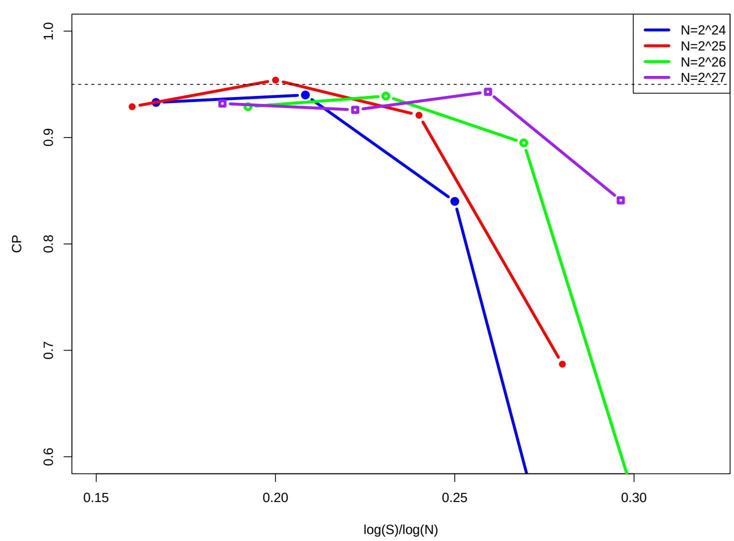

Figure 3.1: Coverage probability of 95% predictive interval with different choices of N and S, for the value search estimator

3.4.2 Yahoo! Today Module user click log dataset

Online content recommendation services have received extensive attention both in the machine learning and statistics literature. These online services strive to make recommendations of advertisements or news articles to individual users by making use of both the content and user information. In this subsection, we apply the proposed method to a Yahoo! Today Module user click log dataset, which contains 45,811,883 user visits to the Today Module, during the first ten days in May 2009. Given such a large number of observations, it is extremely difficult to analyze the entire data on a single computer. This makes the divide and conquer method as an emerging need to deal with such large datasets.

For the ith visit, the dataset contains a binary response variable Y_i, an ID of the recommended article and a 6 dimensional feature vector of the user. Due to sensitivity and privacy concerns, feature definitions and article names were not included in the data. Here, $Y_i = 1$ means the user clicked the recommended article and $Y_i = 0$ means the user didn't click. The last element in the feature vector is always 1, and the first five sums to 1. Therefore, we took the first three and the fifth elements in the feature vector to form the covariates $\boldsymbol{X}_i$. For illustration, we only consider a subset of data that contains visits on May 1st where the recommended article ID is either 109510 or 109520. There were a total of 50 candidate articles on May 1st. We chose these two articles since they were being recommended most on that day. This gives us a total of 405888 visits. On the reduced dataset, define $A_i = 1$ if the recommended article is 109510 and $A_i = 0$ otherwise. In this example, the online recommendation problem can be formulated as follows. Denoted by $\mathscr{D}$ a given set of functions that maps the covariate space to the space of article ID's. Our aim is to find the optimal recommendation strategy to maximize user's click through rate. We consider estimating the optimal recommendation rule among the set of linear decision functions $\mathscr{D} = \{\mathbb{I}(x^T\boldsymbol{\theta} > 1) : \forall \boldsymbol{\theta} \in \mathbb{R}^4\}$. Hence, estimating the optimal recommendation strategy is similar to the problem of estimating the optimal treatment regime as described in Section 3.3. Specifically, we divide the data randomly into S pieces: $\{(\boldsymbol{X}_{i,j}, A_{i,j}, Y_{i,j}) : i = 1, \ldots, n_j\}_{j=1,\ldots,S}$ and obtain

$$\widehat{\boldsymbol{\theta}}_j = \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^4} \frac{1}{n_j}\sum_{i=1}^{n_j}\left\{\left(\frac{A_{i,j}}{\widehat{\pi}_{i,j}}\mathbb{I}(\boldsymbol{\theta}^T\boldsymbol{X}_{i,j} > 1) + \frac{1-A_{i,j}}{1-\widehat{\pi}_{i,j}}\mathbb{I}(\boldsymbol{\theta}^T\boldsymbol{X}_{i,j} \le 1)\right)Y_{i,j} + \right. \tag{3.16}$$

$$\left.\left(\frac{A_{i,j}}{\widehat{\pi}_{i,j}}\mathbb{I}(\boldsymbol{\theta}^T\boldsymbol{X}_{i,j} > 1) + \frac{1-A_{i,j}}{1-\widehat{\pi}_{i,j}}\mathbb{I}(\boldsymbol{\theta}^T\boldsymbol{X}_{i,j} \le 1) - 1\right)\{\widehat{\boldsymbol{h}}_{0i,j}\mathbb{I}(\boldsymbol{\theta}^T\boldsymbol{X}_{i,j} \le 1) + \widehat{\boldsymbol{h}}_{1i,j}\mathbb{I}(\boldsymbol{\theta}^T\boldsymbol{X}_{i,j} > 1)\}\right\}$$

as the subgroup estimator, where $\widehat{\pi}_{i,j}$, $\widehat{\boldsymbol{h}}_{0i,j}$, $\widehat{\boldsymbol{h}}_{1i,j}$ are estimators of $\mathbb{P}(A_{i,j} = 1|\boldsymbol{X}_{i,j})$, $\mathbb{P}(Y_{i,j} =$

$1|A_{i,j} = 0, \boldsymbol{X}_{i,j})$ and $\mathbb{P}(Y_{i,j} = 1|A_{i,j} = 1, \boldsymbol{X}_{i,j})$ respectively. The estimators $\widehat{\pi}_{i,j}, \widehat{\boldsymbol{h}}_{0i,j}, \widehat{\boldsymbol{h}}_{1i,j}$ are obtained by logistic regressions. We chose n_j such that $\max_j n_j - \min_j n_j \le 1$. The estimated optimal recommendation strategy is given as $\mathbb{I}(x^T\widehat{\boldsymbol{\theta}} > 1)$ where $\widehat{\boldsymbol{\theta}} = \sum_j \widehat{\boldsymbol{\theta}}_j/S$. Compared to the value search estimator defined in (3.14), here we obtain the subgroup estimator by maximizing an augmented version of the inverse propensity score weighted estimator. The resulting estimator also converges at a rate of $n^{-1/3}$ but is more efficient than the original one in (3.14).

Due to data confidentiality agreement, we are not able to use the raw data. Here, we generate pseudo responses $\widetilde{Y}_{i,j}$ given $\boldsymbol{X}_{i,j}$ and $A_{i,j}$ from the Yahoo data, and use the dataset $\{(\boldsymbol{X}_{i,j}, A_{i,j}, \widetilde{Y}_{i,j}) : i = 1, \dots, n_j, j = 1, \dots, S\}$ in our application. The generated variables $\widetilde{Y}_{i,j}$'s are similar to the original responses $Y_{i,j}$'s. For example, we have $\sum_{i,j} Y_{i,j}/\sum_j n_i \approx 4.71\%$ while $\sum_{i,j} \widetilde{Y}_{i,j}/\sum_j n_i \approx 4.73\%$. Besides, under our data generating process, the population limit of $\widehat{\boldsymbol{\theta}}_j$ in (3.16) can be explicitly calculated as $\boldsymbol{\theta}_0 = (\theta_{0,1}, \theta_{0,2}, \theta_{0,3}, \theta_{0,4})^T = (2.534, 2.881, 2.796, 3.200)^T$ for any j. Hence, $\boldsymbol{\theta}_0$ is also the population limit of $\widehat{\boldsymbol{\theta}}$ when S does not diverge too fast. Detailed descriptions of generating $\widetilde{Y}_{i,j}$'s are given in Section I of Shi et al. (2018b).

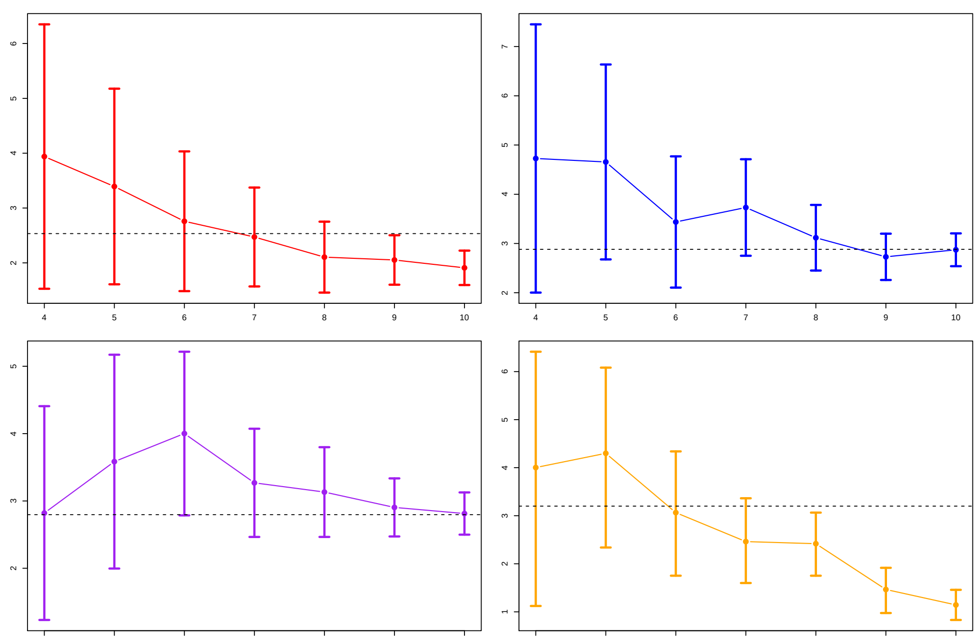

Figure 3.2: 95% confidence intervals of $\theta_0^{(1)}$, $\theta_0^{(2)}$, $\theta_0^{(3)}$ and $\theta_0^{(4)}$ from top to bottom and from left to right, against $\log(S)/\log(2)$. Dash lines are the corresponding $\theta_0^{(i)}$'s.

We choose $S = 2^j$ for $j = 4, 5, \dots, 10$. Under a given S, denoted by $\widehat{\boldsymbol{\theta}}_{(S)} = (\widehat{\theta}_{(S)}^{(1)}, \widehat{\theta}_{(S)}^{(2)}, \widehat{\theta}_{(S)}^{(3)}, \widehat{\theta}_{(S)}^{(4)})^T$ the corresponding aggregated estimator. For each S, we use sample variance to estimate the variance of the aggregated estimator. Based on these estimates, we plot the estimators $\widehat{\boldsymbol{\theta}}_{(S)}^{(i)}$ and the Wald-type 95% confidence intervals of $\theta_0^{(i)}$ in Figure 3.2, for $i = 1, \ldots, 4$ with different choices of S.

It is clear from Figure 3.2 that the variance of $\widehat{\boldsymbol{\theta}}_{(S)}$ decreases as S increases, since the width of confidence intervals decreases as S increases. Moreover, when S is extremely large, some of the parameters are not covered in the 95% confidence intervals. For example, from the top left plot in Figure 3.2, $\theta_0^{(1)}$ is not covered in the confidence intervals based on $\widehat{\theta}_{(S)}^{(1)}$ when $S = 2^9$ and 2^{10}. Such phenomenon is due to the large bias of $\widehat{\boldsymbol{\theta}}_{(S)}$. These empirical results demonstrate the bias-variance trade off for the aggregated estimator, and are consistent with our theoretical findings.

3.5 Tail inequality for $\widehat{\boldsymbol{h}}_j$

In this section, we establish tail inequalities for $\widehat{\boldsymbol{\theta}}_j$ and $\widehat{\boldsymbol{h}}_j$, which are used to construct $\widetilde{\boldsymbol{h}}_j$, a truncated version of $\widehat{\boldsymbol{h}}_j$ with tail equivalence.

Theorem 3.5.1 *Under Conditions (A1)-(A5), for sufficiently large n_j, there exists some constant C_0, such that*

$$\mathbb{P}(\widehat{\boldsymbol{\theta}}_j \notin N_\delta) \leq 2\exp(-C_0 n_j). \tag{3.17}$$

Moreover, for sufficiently large n_j, there exist some constants $C_1, C_2 > 0$ and $N_0 \geq 2$, such that

$$\mathbb{P}(\|\widehat{\boldsymbol{h}}_j\|_2 \geq x | \widehat{\boldsymbol{\theta}}_j \in N_\delta) \leq C_2 \exp(-C_1 x^3), \tag{3.18}$$

for any $N_0 \leq x \leq n_j^{1/3}\delta$.

Inequalities (3.17) and (3.18) can be viewed as generalization of the consistency and rate of convergence results established for cube root estimators (cf. Corollary 4.2 in Kim and Pollard 1990).

We represent $\widehat{\boldsymbol{h}}_j$ as

$$\widehat{\boldsymbol{h}}_j = \arg\max_{\boldsymbol{h} \in H_{n_j}} M_{n_j,j}(\boldsymbol{h}) \equiv \arg\max_{\boldsymbol{h} \in H_{n_j}} \left\{ n_j^{1/6} \mathbb{G}_{n_j}^{(j)}(m_{\boldsymbol{h}}^{(j)}) + n_j^{2/3} \mathbb{E}(m_{\boldsymbol{h}}^{(j)}) \right\},$$

where $H_{n_j} = \{\boldsymbol{h} \in \mathbb{R}^d : n_j^{-1/3}\boldsymbol{h} + \boldsymbol{\theta}_0 \in \boldsymbol{\theta}\}$, $\mathbb{G}_{n_j}^{(j)} = n_j{}^{1/2}(\mathbb{P}_{n_j}^{(j)} - \mathbb{E})$ and $m_{\boldsymbol{h}}^{(j)}(\cdot) = m(\cdot, \boldsymbol{\theta}_0 + n_j^{-1/3}\boldsymbol{h}) -$

$m(\cdot,\boldsymbol{\theta}_0)$. Similarly define

$$\widetilde{\boldsymbol{h}}_j = \arg\max_{h\in H_{n_j}\cap H_{\delta_n}} M_{n_j,j}(h) = \arg\max_{h\in H_{n_j}\cap H_{\delta_n}} \left\{ n_j^{1/6}\mathbb{G}_n^{(j)}(m_{\boldsymbol{h}}^{(j)}) + n_j^{2/3}\mathbb{E}(m_{\boldsymbol{h}}^{(j)}) \right\},$$

where $H_{\delta_n} = \{h : \|h\|_2 \le \delta_n\}$. By its definition, we have $\|\widetilde{\boldsymbol{h}}_j\|_2 \le \delta_n$. The following Corollaries are immediate applications of Theorem 3.5.1.

Corollary 3.5.1 *Assume $\delta_n \le n_j^{1/3}\delta$. Under Conditions (A1)-(A5), for sufficiently large n_j, there exist some constants $N_0 \ge 2$, C_4 and C_5, such that*

$$\mathbb{P}(\|\widetilde{\boldsymbol{h}}_j\|_2 > x) \le C_5 \exp(-C_4 x^3), \qquad \forall x \ge N_0. \tag{3.19}$$

The proof is straightforward by noting that for any $x \le n_j^{1/3}\delta$,

$$\begin{aligned}\mathbb{P}(\|\widetilde{\boldsymbol{h}}_j\|_2 > x) &\le \mathbb{P}(\|\widetilde{\boldsymbol{h}}_j\|_2 > x|\widehat{\boldsymbol{\theta}}_j \in N_\delta)\mathbb{P}(\widehat{\boldsymbol{\theta}}_j \in N_\delta) + \mathbb{P}(\widehat{\boldsymbol{\theta}}_j \notin N_\delta)\\ &\le C_2\exp(-C_1x^3) + 2\exp(-C_0 n_j) \le C_5\exp(-C_4x^3).\end{aligned}$$

Corollary 3.5.1 suggests that $\widetilde{\boldsymbol{h}}_j$ has finite moments of all orders. For any $\boldsymbol{a} \in \mathbb{R}^d$ and positive integer k, this implies that the sequence of random variables $|\boldsymbol{a}^T\widetilde{\boldsymbol{h}}_j|^k$ are uniformly integrable. This result is useful in establishing the convergence for moments of $\widetilde{\boldsymbol{h}}_j$ (see Corollary 3.5.3).

Corollary 3.5.2 *Under Conditions (A1)-(A5) and (A8), set $\delta_n = \max(3^{1/3}, 3^{1/3}/C_1^{1/3})\log^{1/3} n_j$ where C_1 is defined in Theorem 3.5.1, then $\widetilde{\boldsymbol{h}}_j$ and $\widehat{\boldsymbol{h}}_j$ are tail equivalent. If $S = o(n^3)$, then $\sum_{j=1}^S \widetilde{\boldsymbol{h}}_j$ and $\sum_{j=1}^S \widehat{\boldsymbol{h}}_j$ are also tail equivalent.*

Tail equivalence of $\widetilde{\boldsymbol{h}}_j$ and $\widehat{\boldsymbol{h}}_j$ follows by

$$\mathbb{P}\left(\widetilde{\boldsymbol{h}}_j \neq \widehat{\boldsymbol{h}}_j\right) = \mathbb{P}\left(\|\widehat{\boldsymbol{h}}_j\|_2 > \delta_n\right) \le \frac{C_2}{n_j^3} + 2\exp(-C_0 n_j) \le \frac{C_2\bar{c}^3}{n^3} + 2\exp\left(-\frac{C_0 n}{\bar{c}}\right), \tag{3.20}$$

where the first inequality is implied by Theorem 3.5.1 and the last inequality is due to Condition A8. The second assertion follows by an application of Bonferroni's inequality.

Corollary 3.5.2 proves (3.10). From now on, we take $\delta_{n_j} = \max(3^{1/3}, 3^{1/3}/C_1^{1/3})\log^{1/3} n_j$. By (3.20), Slutsky's Theorem implies $\widetilde{\boldsymbol{h}}_j \xrightarrow{d} \boldsymbol{h}_0$. Applying Skorohod's representation Theorem (cf. Section 9.4 in Athreya and Lahiri 2006), we have that there exist random vectors $\widetilde{\boldsymbol{h}}_j^\star \overset{d}{=} \widetilde{\boldsymbol{h}}_j$ and $\boldsymbol{h}_0^\star \overset{d}{=} \boldsymbol{h}_0$ such that $\widetilde{\boldsymbol{h}}_j^\star \to \boldsymbol{h}_0^\star$, almost surely. This together with the uniform integrability of $\|\widetilde{\boldsymbol{h}}_j\|_2^k$ gives the following Corollary.

Corollary 3.5.3 *Under Conditions (A1)-(A5), for any* $\boldsymbol{a} \in \mathbb{R}^d$ *and integer* $k \geq 1$, *we have* $\mathbb{E}\{(\boldsymbol{a}^T \widetilde{\boldsymbol{h}}_j)^k\} \to \mathbb{E}\{(\boldsymbol{a}^T \boldsymbol{h}_0)^k\}$ *as* $n_j \to \infty$.

Due to the i.i.d assumption of $\boldsymbol{X}_{i,j}$, $\mathbb{E}\{(\boldsymbol{a}^T \widetilde{\boldsymbol{h}}_j)^k\}$ is a function of n_j only. Under Condition A8, Corollary (3.5.3) implies

$$\sup_j |\mathbb{E}\{(\boldsymbol{a}^T \widetilde{\boldsymbol{h}}_j)^k\} - \mathbb{E}\{(\boldsymbol{a}^T \boldsymbol{h}_0)^k\}| \to 0, \text{ as } n \to \infty.$$

Taking $k = 2$, it proves (3.12). Taking $k = 3$, it proves (3.13). Moreover, Corollary 3.5.3 suggests a simple scheme for estimating the covariance matrix $A \equiv \mathbb{COV}(\boldsymbol{h}_0)$ given in (3.8). For any vector $\boldsymbol{a}$, by law of large numbers, we obtain

$$\frac{1}{S}\sum_{j=1}^{S}(\boldsymbol{a}^T \widetilde{\boldsymbol{h}}_j)^2 - \frac{1}{S}\sum_{j=1}^{S}\mathbb{E}(\boldsymbol{a}^T \widetilde{\boldsymbol{h}}_j)^2 \stackrel{a.s.}{\to} 0.$$

This together with tail equivalence between $\widetilde{\boldsymbol{h}}_j$ and $\widehat{\boldsymbol{h}}_j$, and (3.12) implies that $\sum_j (\boldsymbol{a}^T \widehat{\boldsymbol{h}}_j)^2/S$ converges to $\boldsymbol{a}^T \boldsymbol{A} \boldsymbol{a}$.

3.6 Analysis of the bias

In this section, we control the accumulated bias in the aggregated estimator as in (6). Our method is inspired by the work of Pimentel (2014), which bounds the expectation of the argmax of a stochastic process by the difference of the expected suprema of the stochastic processes with and without a linear perturbation. To illustrate our idea, we first consider a trivial case by analyzing the bias $\mathbb{E}(\boldsymbol{h}_0)$, where the definition of $\boldsymbol{h}_0$ is given in equation (2) in the main chapter.

3.6.1 Stochastic process with a linear perturbation

Recall $\boldsymbol{h}_0 = \arg\max_{h \in \mathbb{R}^d} Z(\boldsymbol{h})$, where $Z(\boldsymbol{h}) = G(\boldsymbol{h}) - 1/2 h^T V h$. Under Condition A7, the covariance function $\Omega(\boldsymbol{h}_1, \boldsymbol{h}_2)$ of G is equal to $\{L(\boldsymbol{h}_1) + L(\boldsymbol{h}_2) - L(\boldsymbol{h}_1 - \boldsymbol{h}_2)\}/2$. Symmetry of $L(\cdot)$ implies $G(\cdot) \stackrel{d}{=} G(-\cdot)$ and $Z(\cdot) \stackrel{d}{=} Z(-\cdot)$. Hence,

$$\mathbb{E}(\boldsymbol{h}_0) = \mathbb{E}\{\arg\max Z(\boldsymbol{h})\} = \frac{1}{2}\left[\mathbb{E}\{\arg\max Z(\boldsymbol{h})\} + \mathbb{E}\{\arg\max Z(-\boldsymbol{h})\}\right] = 0.$$

Here we provide an alternative but superfluous proof for this trivial case. Define the stochastic process with a linear perturbation

$$Z^{\varepsilon,a}(\boldsymbol{h}) = Z(\boldsymbol{h}) + \varepsilon \boldsymbol{a}^T \boldsymbol{h},$$

for any $\varepsilon \in \mathbb{R}$ and $\boldsymbol{a} \in \mathbb{R}^d$. We have

$$\varepsilon \boldsymbol{a}^T \boldsymbol{h}_0 + \sup_{\boldsymbol{h}} Z(\boldsymbol{h}) \leq \sup_{\boldsymbol{h}} Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h}).$$

Therefore, for any $\varepsilon > 0$,

$$a^T \boldsymbol{h}_0 \leq \frac{1}{\varepsilon}\left(\sup_{\boldsymbol{h}} Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h}) - \sup_{\boldsymbol{h}} Z(\boldsymbol{h})\right), \tag{3.21}$$

and

$$a^T \boldsymbol{h}_0 \geq \frac{1}{-\varepsilon}\left(\sup_{\boldsymbol{h}} Z^{-\varepsilon,\boldsymbol{a}}(\boldsymbol{h}) - \sup_{\boldsymbol{h}} Z(\boldsymbol{h})\right). \tag{3.22}$$

It follows from (3.21) and (3.22) that

$$|\mathbb{E}(\boldsymbol{a}^T \boldsymbol{h}_0)| \leq \frac{1}{\varepsilon} \max\left(|\mathbb{E}\{\sup_{\boldsymbol{h}} Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h}} Z(\boldsymbol{h})\}|, |\mathbb{E}\{\sup_{\boldsymbol{h}} Z^{-\varepsilon,\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h}} Z(\boldsymbol{h})|\} \right) \tag{3.23}$$

In the next Lemma we show that the right-hand side of (3.23) is of the order $O(\varepsilon)$ for any $\boldsymbol{a} \in \mathbb{R}^d$ with $\|\boldsymbol{a}\|_2 = 1$. Taking $\varepsilon \to 0$, we obtain $\mathbb{E}(\boldsymbol{a}^T \boldsymbol{h}_0) = 0$, which implies $\mathbb{E}(\boldsymbol{h}_0) = 0$.

Lemma 3.6.1 *Let $X(\boldsymbol{h}) = B(\boldsymbol{h}) - \boldsymbol{h}^T \boldsymbol{W} \boldsymbol{h}/2$, where $B(\boldsymbol{h})$ is a mean zero process with stationary increments and W is a positive definite matrix. Assume $\mathbb{E}\{\sup_{\boldsymbol{h} \in \mathbb{R}^d} X^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h})\} < \infty$, where $X^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h}) = X(\boldsymbol{h}) + \varepsilon \boldsymbol{a}^T \boldsymbol{h}$. Then, we have*

$$\sup_{\|\boldsymbol{a}\|_2 = 1} \left| \mathbb{E}\{\sup_{\boldsymbol{h} \in \mathbb{R}^d} X^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h} \in \mathbb{R}^d} X(\boldsymbol{h})\} \right| = O(\varepsilon^2).$$

As a result, we have $\mathbb{E}\{\arg\max_{\boldsymbol{h}} X(\boldsymbol{h})\} = 0$.

Lemma 3.6.1 can be viewed as a generalization of Theorem 4 in Pimentel (2014). Here we only require the underlying process to have stationary increments. In addition, we allow the underlying process to be indexed by multi-dimensional parameters.

The proof of Lemma 3.6.1 relies on the stationary increments property of B, which implies

$$\sup_{\boldsymbol{h}} X^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h}) \stackrel{d}{=} \sup_{\boldsymbol{h}} X(\boldsymbol{h}) + B(\varepsilon \boldsymbol{W}^{-1}\boldsymbol{a}) + \varepsilon^2 \boldsymbol{a}^T \boldsymbol{W}^{-1}\boldsymbol{a}/2.$$

In the following lemma, we prove the finiteness of $\mathbb{E}\{\sup_{\boldsymbol{h}} Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h})\}$.

Lemma 3.6.2 *Under Conditions A1 - A7, there exist some positive constants C_6, C_7, C_8 and K_0 such that*

$$\sup_{\|\boldsymbol{a}\|_2=1, |\varepsilon|\le 1} \mathbb{P}\left(\sup_{\boldsymbol{h}} Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h}) > x\right) \le C_6 \exp(-C_7 x^2) + C_8 \exp(-x),$$

for any $x \ge K_0$. As a result, for any integer $m > 0$, we have

$$\sup_{\|\boldsymbol{a}\|_2=1} \sup_{|\varepsilon|\le 1} \mathbb{E}\left[\{\sup_{\boldsymbol{h}} Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h})\}^m\right] < \infty.$$

Lemma 3.6.2 shows that not only $\sup_{\boldsymbol{h}} Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h})$ possesses finite moments of all orders, but also has a subexponential tail. This is quite surprising since the supremum is taken on $\mathbb{R}^d$. This result is due to the rescaling property of $L(\cdot)$.

3.6.2 Nonasymptotic bound for the bias

We now establish the order of $|\mathbb{E}(\boldsymbol{a}^T \widetilde{\boldsymbol{h}}_j)|$. Define the following process with a linear drift

$$M_{n_j,j}^{\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h}) = M_{n_j,j}(\boldsymbol{h}) + \varepsilon_{n_j} \boldsymbol{a}^T \boldsymbol{h},$$

for some sequence ε_{n_j}. By Condition A3, it is immediate to see $\mathbb{E}\{\sup_{\boldsymbol{h}\in H_{n_j}\cap H_{\delta_{n_j}}} M_{n_j,j}^{\varepsilon_n,\boldsymbol{a}}(\boldsymbol{h})\} < \infty$ for any ε_{n_j} and a. Similar to (3.21), (3.22) and (3.23), we can show

$$|\sqrt{S}\mathbb{E}(\boldsymbol{a}^T \widetilde{\boldsymbol{h}})| \le \frac{\sqrt{S}}{\varepsilon_{n_j}} \max\Bigg(|\mathbb{E}\{\sup_{\boldsymbol{h}\in H_{n_j}\cap H_{\delta_{n_j}}} M_{n_j,j}^{\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h}\in H_{n_j}\cap H_{\delta_{n_j}}} M_{n_j,j}(\boldsymbol{h})\}|,$$
$$|\mathbb{E}\{\sup_{\boldsymbol{h}\in H_{n_j}\cap H_{\delta_{n_j}}} M_{n_j,j}^{-\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h}\in H_{n_j}\cap H_{\delta_{n_j}}} M_{n_j,j}(\boldsymbol{h})\}| \Bigg), \tag{3.24}$$

for any positive sequence ε_{n_j}.

Since $M_{n_j,j}(\boldsymbol{h})$ converges weakly to $Z(\boldsymbol{h})$, the expected supremum of $M_{n_j,j}(\boldsymbol{h})$ and

$M_{n_j,j}^{\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h})$ should be close to those of $Z(\boldsymbol{h})$ and $Z^{\varepsilon,\boldsymbol{a}}(\boldsymbol{h})$, respectively. Define

$$\Delta_{n_j} = \sup_{|\varepsilon_{n_j}|\le 1, \|\boldsymbol{a}\|_2=1} |\mathbb{E}\{\sup_{\boldsymbol{h}\in H_{n_j}\cap H_{\delta_{n_j}}} M_{n_j,j}^{\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h}\in\mathbb{R}^d} Z^{\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h})\}|. \tag{3.25}$$

It follows from (3.23) that

$$|\sqrt{S}\mathbb{E}(\boldsymbol{a}^T\widetilde{\boldsymbol{h}}_j)| \le 2\frac{\sqrt{S}}{\varepsilon_{n_j}}\Delta_{n_j} + \frac{\sqrt{S}}{\varepsilon_{n_j}}\max\Bigg(|\mathbb{E}\{\sup_{\boldsymbol{h}\in\mathbb{R}^d} Z^{\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h}\in\mathbb{R}^d} Z(\boldsymbol{h})\}|,$$
$$|\mathbb{E}\{\sup_{\boldsymbol{h}\in\mathbb{R}^d} Z^{-\varepsilon_{n_j},\boldsymbol{a}}(\boldsymbol{h})\} - \mathbb{E}\{\sup_{\boldsymbol{h}\in\mathbb{R}^d} Z(\boldsymbol{h})\}|\Bigg). \tag{3.26}$$

The second term in (3.26) is $O(\sqrt{S}\varepsilon_{n_j})$ by Lemma 3.6.1. The first term in (3.26) represents the approximation error of the expected supremum of Gaussian processes, whose order will be studied in the next section. If we take $\varepsilon_{n_j} = \sqrt{\Delta_{n_j}}$, the right-hand side of (3.26) is $O(\sqrt{S\Delta_{n_j}})$. This suggests that the asymptotic normality holds as long as $S = o(\min_j \Delta_{n_j}^{-1})$. Intuitively, this implies that the number of slices S cannot diverge too fast, otherwise the bias will accumulate.

3.6.3 Bound for the approximation error Δ_{n_j}

To establish an upper bound for Δ_{n_j}, we adopt the techniques in Chernozhukov et al. (2013) and Chernozhukov et al. (2014). Specifically, they established the nonasymptotic bound for the following difference (see P.1590 in Chernozhukov et al., 2014 or Theorem 2.1 in Chernozhukov et al, 2013):

$$\left|\mathbb{E}\{g(\max_{j=1}^{p} S_{n,j})\} - \mathbb{E}\{g(\max_{j=1}^{p} T_{n,j})\}\right|,$$

where g is a smooth function with third order derivatives, $S_{n,j} = \sum_{i=1}^n X_{i,j}$ for some mean zero random vectors $\boldsymbol{X}_i = (X_{i,1},\cdots,X_{i,p})^T \in \mathbb{R}^p$, and $T_{n,j} = \sum_{i=1}^n Y_{i,j}$ for some mean zero Gaussian vectors $Y_i = (Y_{i,1},\cdots,Y_{i,p})^T \in \mathbb{R}^p$ with the same covariance matrix as $\boldsymbol{X}_i$.

Here, we improve the result in two ways. First, it's not necessary to assume $\boldsymbol{X}_i$ and Y_i to be mean zero. The same conclusion holds as long as $\mathbb{E}\boldsymbol{X}_i = \mathbb{E}\boldsymbol{Y}_i = \boldsymbol{\mu}_i$ for any $\mu_i < \infty$. Second, we improve the result by taking g to be the identity function as needed in our application. We summarize our result in the following lemma.

Lemma 3.6.3 *Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be independent random vectors in $\mathbb{R}^p$ with finite fourth absolute moments. Define $\mathbb{E}(\boldsymbol{X}_i) = \boldsymbol{\mu}_i = (\mu_{i1}, \cdots, \mu_{ip})^T$, $Z = \max_{1 \le j \le p} \sum_{i=1}^n X_{ij}$, and $\widetilde{Z} = \max_{1 \le j \le p} \sum Y_{ij}$, where $\mathbf{Y}_i = (Y_{i,1}, \cdots, Y_{i,p})^T$ is distributed as $N\{\boldsymbol{\mu}_i, \mathbb{E}(\boldsymbol{X}_i \boldsymbol{X}_i^T)\}$. Then, we have for any $\beta > 0$,*

$$|\mathbb{E}Z - \mathbb{E}\widetilde{Z}| \le 2\beta^{-1} \log p + C_0 \beta \{B_1 + \beta(B_2 + B_3) + \beta^2(B_4 + B_5)\},$$

where C_0 is a constant independent of $\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_n$,

$$B_1 = \mathbb{E}\left\{\max_{1 \le j,k \le p} |\sum_{i=1}^n \widetilde{X}_{ij}\widetilde{X}_{ik} - n\mathbb{E}(\widetilde{X}_{ij}\widetilde{X}_{ik})|\right\},$$
$$B_2 = \mathbb{E}\left\{\max_{1 \le j,k,l \le p} |\sum_{i=1}^n \widetilde{X}_{ij}\widetilde{X}_{ik}\widetilde{X}_{il} - n\mathbb{E}(\widetilde{X}_{ij}\widetilde{X}_{ik}\widetilde{X}_{il})|\right\},$$
$$B_3 = \max_{j,k} \mathbb{E}\left\{\max_{1 \le l \le p} |\sum_{i=1}^n \mathbb{E}(\widetilde{X}_{ij}\widetilde{X}_{ik})\widetilde{X}_{il}|\right\}, \quad B_4 = \mathbb{E}\left(\max_{1 \le j \le p} \sum_{i=1}^n |\widetilde{X}_{ij}|^4\right),$$
$$B_5 = n\mathbb{E}\left\{\max_{1 \le j \le p} |\widetilde{X}_{1j}|^4 \mathbb{I}(\max_{1 \le j \le p} |\widetilde{X}_{1j}| > \frac{1}{2\beta})\right\},$$

$\widetilde{X}_{ij} = X_{ij} - \mu_{ij}$ and $\widetilde{\boldsymbol{X}}_i = \boldsymbol{X}_i - \boldsymbol{\mu}_i$, $j = 1, \cdots, p$, $i = 1, \cdots, n$.

Lemma 3.6.4 *Under Conditions A1-A7, we have $\Delta_{n_j} = O(n_j^{-1/6} \log^{5/6} n_j)$ for sufficiently large n_j.*

Combining Lemma 3.6.4 together with (3.26) suggests that

$$||\mathbb{E}\widetilde{\boldsymbol{h}}_j||_2 = O(\sqrt{\Delta_{n_j}}) = O(n_j^{-1/12} \log^{5/12} n_j),$$

where the bound is uniform in all j. This in turn implies the bias of a cubic-rate estimator can be bounded by $O((n/\log n)^{-5/12})$. The proof of Theorem 2.1 is hence completed.

Finally, we would like to point out that our method of analyzing bias is not specific to cubic-rate M-estimators. In fact, as long as the limiting Gaussian process of an M-estimator has stationary increments, the bias of the aggregated estimator can be similarly bounded using our method.

3.7 Discussion

In this chapter, we provide a general inference framework for aggregated M-estimators with cubic rates obtained by the divide and conquer method. Our results demonstrate that the aggregated estimators have faster convergence rate than the original M-estimators based on pooled data and achieve the asymptotic normality when the number of groups S does not grow too fast with respect to n, the average sample size of each group. The proposed methodology is further applied to estimating the OTR based on the value-search method.

3.7.1 Rate of the bias

For a general cubic-rate estimator with sample size n, we showed its bias can be bounded by $O((n/\log n)^{-5/12})$. In comparison, Banerjee et al. (2019) obtained a sharper bound in the specific setting of monotone regression and showed that the bias of the isotonic estimator can be bounded by $O(n^{-7/15+\zeta})$ for any $\zeta > 0$ and the bias of its inverse bounded by $o(n^{-1/2})$ (see Theorem 4.3 and 4.4 in that paper). As commented before, this is because we work on a more general setting and their techniques cannot be easily generalized to other cubic-rate M-estimation problems.

However, it is possible to sharpen the bound for some special cases. In particular, when the parameter is one-dimensional, the bias of the estimator can be bounded by $O(n^{-5/9}\log^{9/14} n)$ based on the KMT approximation (see Theorem A.1 and Corollary A.1 of Shi et al. (2018b)). Under Assumption A8, the asymptotic normality holds for the aggregated estimator as long as the number of machines satisfies $S = O(N^l)$ for some $l < 4/13$, where N is the total number of observations. Again this upper bound on S may still be conservative, however it improves a lot compared to Theorem 2.1. Readers can refer to Section A.5 of Shi et al. (2018b) for more details.

For the bias of a general cubic-rate M-estimator, our proof relies on the Gaussian approximation of the suprema of empirical processes (cf. Chernozhukov et al. 2013, 2014) and the Sudakov-Fernique type error bound (Chatterjee 2005). The proofs for these theorems are based on smooth approximation of the supremum function. It remains unknown whether the rates of these error bounds are optimal and whether they can be improved using other techniques. This is an interesting problem that needs further investigation.

3.7.2 The super-efficiency phenomenon

In the context of isotonic regression, Banerjee et al. (2019) showed that the faster convergence rate of the aggregated estimator of the inverse function for a fixed model comes at a price, that is, the maximal risk over a class of models in a neighborhood of the given model remains bounded for the pooled estimator but diverges to infinity for the aggregated estimator (see Theorem 6.1 in Banerjee et al. 2019). This is referred to as the super-efficiency phenomenon, which is seen in nonparametric function estimation as well (cf. Brown et al. 1997).

We believe such super-efficiency phenomenon holds for many other cubic-rate M-estimation problems as well. Shi et al. (2018b) mathematically formalize the notion of the super-efficiency phenomenon for general M-estimation problems, and establish such phenomenon for the location estimator (see Section B.1) and the value search estimator (see Section B.2). The super-efficiency phenomenon is essentially due to that the maximal bias over a large class of models for the aggregated estimator will diverge to infinity. This is because the condition on the Lipschitz continuity of the Hessian matrix (Assumption A6) cannot hold uniformly for all models in such a class.

CHAPTER

4

PENALIZED A-LEARNING FOR HIGH-DIMENSIONAL DATA

4.1 Introduction

With the fast development of new technology, it becomes possible to gather an extradinary large number of prognostic factors for each individual. This makes variable selection as an emerging need for estimating the O(D)TR. In the literature, variable selection tools for deriving O(D)TR haven't been fully explored. Qian and Murphy (2011) proposed to estimate the conditional mean response using an ℓ_1-penalized regression and studied the error bound of the value function for the estimated treatment regime. When the number of covariates is fixed, Lu et al. (2013) introduced a new penalized least squared regression framework and established the oracle property of the estimator, which is robust against the misspecification of the conditional mean function. Shi et al. (2016) extended this result to the setting allowing NP-dimensionality of covariates. However, all these works only consider studies with a single treatment decision. When moving to multiple-stage studies, the asymptotic properties of the estimated optimal dynamic treatment regime are much harder to derive since it needs to handle model misspecification of the contrast functions in the presence of NP-dimensionality of covariates. Moreover, these methods are not doubly robust.

In this chapter, we propose a penalized A-learning method for deriving the ODTR when the number of covariates is of NP-order of the sample size. To preserve the doubly robust property of the A-learning method, we adopt the Dantzig selector (Candès and Tao 2007) which directly penalizes the A-leaning estimating equations. The technical challenges and advances of the proposed estimators are described as follows.

First, to prove the theoretical properties of the Dantzig estimator in linear regression setting, the uniform uncertainty principle (UUP, Candès and Tao 2007) or restricted eigenvalue condition (RE, Bickel et al. 2009) is required on the Gram matrix $\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T$, where $\boldsymbol{X}_i$ stands for the ith covariates. The UUP condition essentially requires that every principle submatrix with the number of rows or columns less than some specified s behaves like an orthonormal system. The RE condition is the weakest and hence the most general condition in the literature to ensure the theoretical properties of Lasso and Dantzig estimators. A close connection between these two conditions are discussed in Bickel et al. (2009). In a random design case, Candès and Tao (2007) studied the UUP condition for Gaussian, Bernoulli and Fourier ensembles. Mendelson et al. (2007, 2008) obtained a similar result for a more general class of design matrices, the isotropic subgaussian matrices, based on some empirical process results. These results were further extended by Zhou (2009), where the UUP and RE conditions are developed for subgaussian ensembles with a correlated covariance structure. In the proposed penalized A-learning method, however, such conditions are required on matrices involving estimates, such as

$$\sum_{i=1}^{n} A_i(1-\widehat{\pi}_i)\boldsymbol{X}_i \boldsymbol{X}_i^T, \tag{4.1}$$

where A_i denotes the treatment received by the ith subject, and $\widehat{\pi}_i$ denotes the corresponding estimated propensity scores. The presence of the estimated propensity score in (4.1) adds extraordinary difficulties in establishing theoretical properties of such a random matrix. We establish the UUP and RE conditions under a proper convergence rate of the estimated propensity score, which provides a new theoretical framework for studying random matrices that involve estimates of unknown parameters.

Second, in the proposed penalized A-learning method, we need to estimate the baseline mean function and the propensity score model with NP-dimensionality of covariates. We adopt the penalized regressions with the folded-concave penalties, for example, a linear regression for the baseline mean function and a logistic regression for the propensity score model, with the SCAD penalties. Several difficulties need to be addressed for deriving the theoretical properties of the resulting penalized estimators. First, to our knowledge, penalized regressions with folded-concave penalties have seldom been studied in a random design setting. A major difficulty of adapting the existing results for the fixed design case to the random design case is to control the maximum eigenvalues of some random matrices,

$$\max_j \lambda_{\max}\left(\sum_{i=1}^{n}(\boldsymbol{X}_i^{\mathscr{M}})^T|X_i^{(j)}|\boldsymbol{X}_i^{\mathscr{M}}\right),$$

where $\lambda_{\max}(\boldsymbol{M}_0)$ denotes the maximum eigenvalue of a matrix $\boldsymbol{M}_0$, $\mathscr{M}$ is a given subset of $\{1,\cdots,p\}$, $X_i^{(j)}$ denotes the jth element of $\boldsymbol{X}_i$, and $\boldsymbol{X}_i^{\mathscr{M}}$ the submatrix formed by columns in $\mathscr{M}$. Such a problem is not standard since matrix $\sum_i(\boldsymbol{X}_i^{\mathscr{M}})^T|\boldsymbol{X}_i^j|\boldsymbol{X}_i^{\mathscr{M}}$ does not possess sub-exponential tail even when each $\boldsymbol{X}_i$ is a sub-Gaussian vector. We derive some concentration inequalities for such random matrices and for summations of subexponential and subgaussian random variables. Based on these results, we establish the weak oracle (Lv and Fan 2009) properties, i.e, sign consistency and L_∞ convergence rate of the estimators under subgaussian ensembles. Moreover, the posited models for the baseline mean function or the propensity score may be misspecified. Therefore, the derivation of the asymptotic properties needs to take into account model misspecification with NP-dimensionality of covariates, which is challenging.

Third, a challenge for extending the results for a single treatment decision to sequential treatment decisions is that the contrast functions are likely to be misspecified in the backward induction algorithm, such as A-learning. This together with the NP-dimensionality of covariates make it extremely hard to study theoretical properties of the value function under the estimated optimal dynamic treatment regime. We overcome this difficulty by first defining population-level least favorable parameters in the misspecified contrast functions. Moreover, we derive the error bounds for the corresponding estimates under the model misspecification, which in turn leads to an error bound for the difference between the value functions of the estimated optimal dynamic treatment regime and the underlying true optimal dynamic treatment regime.

The remainder of the paper is organized as follows. We introduce the proposed penalized A-learning method in Section 4.2. Some implementation issues are addressed in Section 4.3, followed by simulation results in Section 4.4. We apply our method to a data from the STAR*D study in Section 4.5. Section 4.6 studies the error bounds of the penalized A-learning estimator and the difference between the value functions of the estimated optimal regime and the true optimal regime, at the second stage. Section 4.7 characterizes such results for the estimates at the first stage. Section 4.8 presents the weak oracle properties of the penalized estimators in the propensity score and baseline mean models under a random design setting. Section 4.9 discusses the UUP and RE condition in the context of A-learning. In Section 4.10, we propose concordance and value information criteria that help select the tuning parameters in penalized A-learning. In Section 4.11, we briefly discuss the issue of statistical inference of the optimal value function in a high-dimensional setting. All technical lemmas and proofs can be found in Shi et al. (2018a).

4.2 Penalized A-Learning

We focus on the multiple time point study (see Section 1.3 for the notations and the model setup) with the total number of decision points $K = 2$. Extension to a three-stage study can be found in Section 11 of Shi et al. (2018a). The observed data from n subjects can then be summarized as

$$O_i = (\boldsymbol{X}_i^{(1)}, A_i^{(1)}, \boldsymbol{X}_i^{(2)}, A_i^{(2)}, Y_i), i = 1, \ldots, n,$$

which are assumed to be independently and identically distributed copies of

$$O_0 = (\boldsymbol{X}_0^{(1)}, A_0^{(1)}, \boldsymbol{X}_0^{(2)}, A_0^{(2)}, Y_0).$$

For $i = 0, 1, 2, \ldots, n$, let $\boldsymbol{S}_i = ((\boldsymbol{X}_i^{(1)})^T, A_i^{(1)}, (\boldsymbol{X}_i^{(2)})^T)^T$ be the p-dimensional vector of covariates for the ith patient observed prior to the second-stage treatment. We will write $\tau_2(a_1, \bar{\boldsymbol{x}}_2)$, $h_2\{(a_1, 0), \bar{\boldsymbol{x}}_2\}$ and $\pi_2\{(a_1, 1), \bar{\boldsymbol{x}}_2\}$ as $\tau_2(\boldsymbol{s})$, $h_2(\boldsymbol{s})$ and $\pi_2(\boldsymbol{s})$ when $\boldsymbol{s} = (\boldsymbol{x}_1^T, a_1, \boldsymbol{x}_2^T)^T$, and write $h_1(0, \boldsymbol{x}_1)$, $\pi_1(1, \boldsymbol{x}_1)$ as $h_1(\boldsymbol{x}_1)$, $\pi_1(\boldsymbol{x}_1)$. In addition, we will omit the superscript (1) and write $\boldsymbol{X}_i^{(1)}$ as $\boldsymbol{X}_i$ for $i = 0, 1, \ldots, n$. The ODTR is given by $d^{opt,0} = (d_1^{opt,0}, d_2^{opt,0})$ where

$$d_1^{opt,0}(\boldsymbol{x}) = \mathbb{I}(\tau_1(\boldsymbol{x}) > 0) \quad \text{and} \quad d_2^{opt,0}(\boldsymbol{s}) = \mathbb{I}(\tau_2(\boldsymbol{s}) > 0).$$

To estimate $d_1^{opt,0}$ and $d_2^{opt,0}$, we posit the following models for $\tau_1(\cdot)$, $\tau_2(\cdot)$, $h_1(\cdot)$, $h_2(\cdot)$,

$\pi_1(\cdot)$, and $\pi_2(\cdot)$:

$$\pi_1(\boldsymbol{x}) = \exp(\boldsymbol{x}^T\boldsymbol{\alpha}_1)/\{1+\exp(\boldsymbol{x}^T\boldsymbol{\alpha}_1)\}, \tag{4.2}$$

$$\pi_2(\boldsymbol{s}) = \exp(\boldsymbol{s}^T\boldsymbol{\alpha}_2)/\{1+\exp(\boldsymbol{s}^T\boldsymbol{\alpha}_2)\}, \tag{4.3}$$

$$h_1(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\theta}_1,\ h_2(\boldsymbol{s}) = \boldsymbol{s}^T\boldsymbol{\theta}_2,\ \tau_1(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}_1,\ \tau_2(\boldsymbol{s}) = \boldsymbol{s}^T\boldsymbol{\beta}_2, \tag{4.4}$$

for some high-dimensional vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$.

Models in (4.2)-(4.4) can be misspecified, however, we require that either h_j or π_j is correct for $j = 1, 2$. For simplicity, we require τ_2 to be correctly specified. The general case when τ_2 is misspecified can be similarly discussed. We use backward induction to estimate the ODTR. At the second decision point, we first estimate the parameters in the posited propensity score and baseline mean models using penalized regressions. Specifically, define

$$\widehat{\boldsymbol{\alpha}}_2 = \operatorname*{arg\,min}_{\boldsymbol{\alpha}_2\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}\left(\log\{1+\exp(\boldsymbol{S}_i^T\alpha_2)\} - A_i^{(2)}\boldsymbol{S}_i^T\alpha_2\right) + \sum_{j=1}^{p}\lambda_{1n}^{(2)}\rho_1^{(2)}(|\alpha_2^{(j)}|, \lambda_{1n}^{(2)}),$$

and

$$\widehat{\boldsymbol{\theta}}_2 = \operatorname*{arg\,min}_{\boldsymbol{\theta}_2\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}(1-A_i^{(2)})(Y_i - \boldsymbol{S}_i^T\boldsymbol{\theta}_2)^2 + \sum_{j=1}^{p}\lambda_{2n}^{(2)}\rho_2^{(2)}(|\theta_2^{(j)}|, \lambda_{2n}^{(2)}),$$

where $\boldsymbol{\alpha}_2 = (\alpha_2^{(1)}, \cdots, \alpha_2^{(p)})^T$, $\boldsymbol{\theta}_2 = (\theta_2^{(1)}, \cdots, \theta_2^{(p)})^T$, $\rho_1^{(2)}$ and $\rho_2^{(2)}$ belong to the class of folded-concave penalty functions (Lv and Fan 2009), such as SCAD (Fan and Li 2001), and $\lambda_{1n}^{(2)}$, $\lambda_{2n}^{(2)}$ the associated regularization parameters.

Next, we estimate $\boldsymbol{\beta}_2$ in (4.4) using the Dantzig selector based on A-learning estimating function (Murphy 2003), defined by

$$\widehat{\boldsymbol{\beta}}_2 = \arg\min_{\boldsymbol{\beta}_2\in\Lambda^{(2)}} \|\boldsymbol{\beta}_2\|_1, \tag{4.5}$$

where

$$\Lambda^{(2)} = \left\{\boldsymbol{\beta}_2 \in \mathbb{R}^p : \left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{S}_i(A_i^{(2)} - \widehat{\pi}_{2,i})(Y_i - \boldsymbol{S}_i^T\widehat{\boldsymbol{\theta}}_2 - A_i^{(2)}\boldsymbol{S}_i^T\boldsymbol{\beta}_2)\right\|_\infty \le \lambda_{3n}^{(2)}\right\},$$

$$\widehat{\pi}_{2,i} = \exp(\boldsymbol{S}_i^T\widehat{\boldsymbol{\alpha}}_2)/\{1+\exp(\boldsymbol{S}_i^T\widehat{\boldsymbol{\alpha}}_2)\}, \quad i = 1, \ldots, n,$$

and $\lambda_{3n}^{(2)}$ the regularization parameter.

To estimate the regime at the first decision point, we define the pseudo-outcome $\widehat{V}_i$ using the advantage function (Murphy 2003) by

$$\widehat{V}_i = Y_i + \boldsymbol{S}_i^T \widehat{\boldsymbol{\beta}}_2 \{\mathbb{I}(\boldsymbol{S}_i^T \widehat{\boldsymbol{\beta}}_2 > 0) - A_i^{(2)}\}. \tag{4.6}$$

Similarly, define

$$\widehat{\boldsymbol{\alpha}}_1 = \underset{\boldsymbol{\alpha}_1 \in \mathbb{R}^q}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left(\log\{1+\exp(\boldsymbol{X}_i^T \boldsymbol{\alpha}_1)\} - A_i^{(1)} \boldsymbol{X}_i^T \boldsymbol{\alpha}_1\right) + \sum_{j=1}^{q} \rho_1^{(1)}(|\boldsymbol{\alpha}_1^{(j)}|, \lambda_{1n}^{(1)}),$$

and

$$\widehat{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\theta}_1 \in \mathbb{R}^q}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (1 - A_i^{(1)})(\widehat{V}_i - \boldsymbol{X}_i^T \boldsymbol{\theta}_1)^2 + \sum_{j=1}^{q} \rho_2^{(1)}(|\boldsymbol{\theta}_1^{(j)}|, \lambda_{2n}^{(1)}),$$

where q is the dimension of $\boldsymbol{X}_0$, $\boldsymbol{\alpha}_1 = (\boldsymbol{\alpha}_1^{(1)}, \cdots, \boldsymbol{\alpha}_1^{(q)})^T$, $\boldsymbol{\theta}_1 = (\boldsymbol{\theta}_1^{(1)}, \cdots, \boldsymbol{\theta}_1^{(q)})^T$, and $\rho_1^{(1)}$ and $\rho_2^{(1)}$ are folded-concave penalty functions. Then, we estimate $\boldsymbol{\beta}_1$ by

$$\widehat{\boldsymbol{\beta}}_1 = \underset{\boldsymbol{\beta}_1 \in \Lambda^{(1)}}{\arg\min} \|\boldsymbol{\beta}_1\|_1, \tag{4.7}$$

where

$$\Lambda^{(1)} = \left\{ \boldsymbol{\beta}_1 \in \mathbb{R}^q : \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i (A_i^{(1)} - \widehat{\pi}_{1,i})(\widehat{V}_i - \boldsymbol{X}_i^T \widehat{\boldsymbol{\theta}}_1 - A_i^{(1)} \boldsymbol{X}_i^T \boldsymbol{\beta}_1) \right\|_\infty \le \lambda_{3n}^{(1)} \right\},$$

$$\widehat{\pi}_{1,i} = \exp(\boldsymbol{X}_i^T \widehat{\boldsymbol{\alpha}}_1)/\{1+\exp(\boldsymbol{X}_i^T \widehat{\boldsymbol{\alpha}}_1)\}, \quad i = 1, \cdots, n.$$

The estimated ODTR is given by

$$\widehat{d}_1(\boldsymbol{x}) = \mathbb{I}(\widehat{\boldsymbol{\beta}}_1^T \boldsymbol{x} > 0) \quad \text{and} \quad \widehat{d}_2(\boldsymbol{s}) = \mathbb{I}(\widehat{\boldsymbol{\beta}}_2^T \boldsymbol{s} > 0). \tag{4.8}$$

4.3 Some Implementation Issues

When the tuning parameters in optimization problems (4.5) and (4.7) are fixed, the Dantzig selector can be solved by a standard linear programming algorithm. One issue for implementing Dantzig selector is the choice of the tuning parameters. We use a BIC criterion for selecting tuning parameters. For Dantzig selector (4.5), $\lambda_{3n}^{(2)}$ is chosen as the minimizer of

$$\mathrm{BIC}(\lambda) = n\log(RSS(\lambda)/n) + d(\lambda)\{\log(n) + \log(p+1)\}, \tag{4.9}$$

where $RSS(\lambda) = \sum_{i=1}^{n}\{(A_i^{(2)} - \widehat{\pi}_{2,i})(Y_i^{(2)} - \boldsymbol{S}_i^T\widehat{\boldsymbol{\theta}}_2 - A_i^{(2)}\boldsymbol{S}_i^T\widehat{\beta}_2)\}^2$, and $d(\lambda)$ is the number of nonzero components in $\widehat{\boldsymbol{\beta}}_2$. A similar BIC criterion was proposed by Chen and Chen (2008). We use a similar criterion for choosing $\lambda_{3n}^{(1)}$.

It was observed that the Dantzig estimators may underestimate the true values of parameters due to the shrinkage estimation (Candès and Tao 2007). Therefore, we use a two-step procedure for practical implementation, which is referred as Gauss-Dantzig selector in Candès and Tao (2007). Specifically, in the first step, we apply the proposed penalized A-learning to select important variables for making an optimal decision, i.e. those variables with non-zero estimated coefficients. Then, in the second step, their corresponding coefficients are re-calculated by solving the unpenalized A-learning estimating equations with important variables only.

4.4 Simulation Studies

4.4.1 Settings

To evaluate the numerical performance of the proposed penalized A-learning method, we consider simulation studies with two treatment decision points, based on the following model:

$$Y_0 = A_0^{(1)}A_0^{(2)} + A_0^{(2)}(\boldsymbol{\nu}_2^T\boldsymbol{X}_0 + X_0^{(2)}) + A^{(1)}(\boldsymbol{\nu}_1^T\boldsymbol{X}_0) + \varepsilon_0, \tag{4.10}$$

where the random error ε_0 follows a normal distribution with mean 0 and variance 0.25. Here, the baseline covariates $\boldsymbol{X}_0$ follow a multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{I}_q$. In addition, the intermediate covariate $X_0^{(2)}$ is a scalar and generated as $X_0^{(2)} = (1, \underbrace{0,\cdots,0}_{q-1})^T\boldsymbol{X}_0 + A_0^{(1)} + A_0^{(1)}(1, \underbrace{0,\cdots,0}_{q-1})^T\boldsymbol{X}_0 + e_0$, where e_0 follows a normal distribution with mean 0 and variance 0.25.

Based on Model (4.10), the optimal treatment regime at stage 2 is $\mathbb{I}(A_0^{(1)} + \boldsymbol{\nu}_2^T\boldsymbol{X}_0 + X_0^{(2)} > 0)$.

Following this optimal treatment regime at stage 2, we have

$$\begin{aligned} h_1(A_0^{(1)}, \boldsymbol{X}_0) &= \mathbb{E}\left((A_0^{(1)} + \boldsymbol{\nu}_2^T \boldsymbol{X}_0 + X_0^{(2)})_+ \middle| \boldsymbol{X}_0, A_0^{(1)}\right) + A_0^{(1)}(\boldsymbol{\nu}_1^T \boldsymbol{X}_0) \\ &= \frac{\boldsymbol{\beta}_2}{\sqrt{8\pi}} \exp\left(-2\mu_0^2\right) + \mu_0\{1 - \Phi(-2\mu_0)\} + A_0^{(1)}(\boldsymbol{\nu}_1^T \boldsymbol{X}_0), \end{aligned}$$

where $\mu_0 = A_0^{(1)} + \boldsymbol{\nu}_2^T \boldsymbol{X}_0 + (1, \underbrace{0, \cdots, 0}_{q-1})\boldsymbol{X}_0 + A_0^{(1)} + A_0^{(1)}(1, \underbrace{0, \cdots, 0}_{q-1})\boldsymbol{X}_0$ and $a_+ = \max(a, 0)$.

To evaluate the double robustness of the proposed method, we consider a variety of scenarios with correctly specified and misspecified baseline mean and/or propensity score models. At stage 2, a linear model with covariates $\boldsymbol{X}_i$, $X_i^{(2)}$ and $A_i^{(1)}$ is fitted for the baseline function, while the true baseline is $A_i^{(1)}(\boldsymbol{\nu}_1^T \boldsymbol{S}_i^{(1)})$. We choose $\boldsymbol{\nu}_1 = \mathbf{0}_q$, for which the baseline mean function is correctly specified, and $\boldsymbol{\nu}_1 = (\mathbf{0}_4, 1, -1, \mathbf{0}_{q-6})^T$, for which the baseline mean function is misspecified. At stage 1, a linear model with covariates $\boldsymbol{X}_i$ is fitted for the baseline mean function, which is always misspecified. Logistic models are used for estimating the propensity scores, which are correctly specified for the constant model but misspecified for the probit model. The following four settings are considered:

Setting 1: $\boldsymbol{\nu}_1 = \mathbf{0}_q$, $P(A_0^{(2)} = 1) = 0.5$;

Setting 2: $\boldsymbol{\nu}_1 = (\mathbf{0}_4, 1, -1, \mathbf{0}_{q-6})^T$, $P(A^{(2)} = 1) = 0.5$;

Setting 3: $\boldsymbol{\nu}_1 = \mathbf{0}_q$, $P(A_0^{(2)} = 1) = \mathbb{P}(N(0,1) \leq \boldsymbol{X}_0^T \gamma)$;

Setting 4: $\boldsymbol{\nu}_1 = (\mathbf{0}_4, 1, -1, \mathbf{0}_{q-6})^T$, $P(A_0^{(2)} = 1) = \mathbb{P}(N(0,1) \leq \boldsymbol{X}^T \gamma)$,

where $N(0,1)$ denotes a standard normal random variable. For other parameters, we choose $P(A_0^{(1)} = 1 | \boldsymbol{X}_0) = 0.5$, $\boldsymbol{\nu}_2 = (0, 0, 1, -1, \mathbf{0}_{q-4})^T$, and $\boldsymbol{\gamma} = (\mathbf{0}_{q-2}, 1, -1, 1)^T$. Table 4.1 summarizes the information of model misspecification for the baseline mean and propensity score models and associated important variables under different settings. In next section, we show simulation results of the four settings with $q = 1000$ and sample size $n = 150/300$ over 500 replications.

4.4.2 Competing methods

We further compare our method with outcome weighted learning (OWL, Zhao et al. 2012), which is a robust method which estimates individualized treatment rule by directly maximizing the estimated value function. Zhao et al. (2015) further introduced backward outcome

Table 4.1: Simulation Settings

	Stage	Baseline	Propensity Score	Important Variables
Setting 1	Stage 2	right	right	$(S^{(2)}, A_1, S_3^{(1)}, S_4^{(1)})$
	Stage 1	wrong	right	$(S_1^{(1)}, S_3^{(1)}, S_4^{(1)})$
Setting 2	Stage 2	wrong	right	$(S^{(2)}, A_1, S_3^{(1)}, S_4^{(1)})$
	Stage 1	wrong	right	$(S_1^{(1)}, S_3^{(1)}, S_4^{(1)}, S_5^{(1)}, S_6^{(1)})$
Setting 3	Stage 2	right	wrong	$(S^{(2)}, A_1, S_3^{(1)}, S_4^{(1)})$
	Stage 1	wrong	right	$(S_1^{(1)}, S_3^{(1)}, S_4^{(1)})$
Setting 4	Stage 2	wrong	wrong	$(S^{(2)}, A_1, S_3^{(1)}, S_4^{(1)})$
	Stage 1	wrong	right	$(S_1^{(1)}, S_3^{(1)}, S_4^{(1)}, S_5^{(1)}, S_6^{(1)})$

weighted learning (BOWL) and simultaneous outcome weighted learning (SOWL) to extend their methods to multiple stage studies. Here, we consider a double robust version of BOWL (DR-BOWL) for comparison. For a single stage study, the developed DR-BOWL method is similar to the residual weighted learning method (Zhou et al. 2015).

Specifically, we first estimate the propensity score $\widehat{\pi}_{2,i}$ and baseline $\widehat{h}_{2,i} = \boldsymbol{S}_i^T \widehat{\boldsymbol{\theta}}_2$ for $i = 1, \ldots, n$ as in Section 2. We consider the linear decision rule $\mathbb{I}(\boldsymbol{s}^T \boldsymbol{\beta}_{2,0} > 0)$ and estimate $\boldsymbol{\beta}_{2,0}$ by minimizing the following loss function:

$$\widetilde{\boldsymbol{\beta}}_2 = \arg\min_{\boldsymbol{\beta}_2} \frac{1}{n} \sum_i \frac{(Y_i - \widehat{h}_{2,i})\{1 - (2A_i^{(2)} - 1)\boldsymbol{S}_i^T \boldsymbol{\beta}_2\}_+}{A_i^{(2)} \widehat{\pi}_{2,i} + (1 - A_i^{(2)})(1 - \widehat{\pi}_{2,i})} + \lambda_{3n}^{(2)} \|\boldsymbol{\beta}_2\|_1.$$

The penalty term in original OWL is $\lambda_{3n}^{(2)} \|\boldsymbol{\beta}_2\|_2^2$. We replace it with the L_1 norm here to simultaneously select variables. Then we construct the pseudo outcome $\widetilde{V}_i$ using augmented inverse propensity score estimator (AIPWE Zhang et al. 2012),

$$\widetilde{V}_i = \frac{A_i^{(2)} \widetilde{d}_2(\boldsymbol{S}_i) + (1 - A_i^{(2)})\{1 - \widetilde{d}_2(\boldsymbol{S}_i)\}}{A_i^{(2)} \widehat{\pi}_{2,i} + (1 - A_i^{(2)})(1 - \widehat{\pi}_{2,i})} Y_i - \left(\frac{A_i^{(2)} \widetilde{d}_2(\boldsymbol{S}_i) + (1 - A_i^{(2)})\{1 - \widetilde{d}_2(\boldsymbol{S}_i)\}}{A_i^{(2)} \widehat{\pi}_{2,i} + (1 - A_i^{(2)})(1 - \widehat{\pi}_{2,i})} - 1 \right) [\widehat{h}_{2,i}\{1 - \widetilde{d}_2(\boldsymbol{S}_i)\} + \widehat{\Phi}_i^{(2)} \widetilde{d}_2(\boldsymbol{S}_i)]$$

where $\widetilde{d}_2(\boldsymbol{S}_i) = \mathbb{I}(\boldsymbol{S}_i^T \widetilde{\boldsymbol{\beta}}_2 > 0)$, and $\widehat{\Phi}_i^{(2)}$ is an estimate of $\Phi_i^{(2)} = \mathbb{E}(Y_0 | A_0^{(2)} = 1, \boldsymbol{S}_0 = \boldsymbol{S}_i)$. Here, we fit a linear model for $\mathbb{E}(Y_0 | A_0^{(2)} = 1, \boldsymbol{S}_0)$ and use nonconcave penalized regression with SCAD penalty to obtain $\widehat{\Phi}_i^{(2)}$. Denoted by $\widehat{\pi}_{1,i}$ and $\widehat{h}_{1,i} = \boldsymbol{X}_i^T \widehat{\boldsymbol{\theta}}^{(1)}$ the estimated propensity score and baseline at the first stage, we consider linear treatment regime of the form $\mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}_1^* > 0)$ and estimate $\boldsymbol{\beta}_1^*$ by

$$\widetilde{\boldsymbol{\beta}}_1 = \arg\min_{\boldsymbol{\beta}_1} \frac{1}{n}\sum_i \frac{(\widetilde{V}_i - \widehat{h}_{1,i})\{1-(2A_i^{(1)}-1)\boldsymbol{X}_i^T\boldsymbol{\beta}_1\}_+}{A_i^{(1)}\widehat{\pi}_{1,i} + (1-A_i^{(1)})(1-\widehat{\pi}_{1,i})} + \lambda_{3n}^{(1)}\|\boldsymbol{\beta}_1\|_1.$$

Tuning parameters $\lambda_{3n}^{(2)}$ and $\lambda_{3n}^{(1)}$ are obtained by minimizing a value-based BIC criterion.

4.4.3 Results

Table 4.2 summarizes variable selection results for optimal treatment decisions and the empirical performance of the estimated optimal treatment regime compared with the true optimal regime, using our penalized A-learning method (denoted as PAL) and DR-BOWL, respectively. Specifically, it reports the false negative (FN) rate (the percentage of important variables that are missed) and false positives (FP) rate (the percentage of unimportant variables that are selected), the ratio of value functions (denoted by VR) calculated using the value function of the estimated optimal treatment regime divided by that of the true optimal regime, and the error rates (ER) of the estimated optimal treatment regimes for treatment decision making, in both stages. Here, the ER at stage 2 is calculated as the mean of $n^{-1}\sum_{i=1}|\mathbb{I}(\widehat{\boldsymbol{\beta}}_2^T\boldsymbol{S}_i > 0) - \mathbb{I}(\boldsymbol{\beta}_{2,0}^T\boldsymbol{S}_i > 0)|$ and at stage 1 as the mean of $n^{-1}\sum_{i=1}|\mathbb{I}(\widehat{\boldsymbol{\beta}}_1^T\boldsymbol{X}_i > 0) - \mathbb{I}(\tau_1(\boldsymbol{X}_i) > 0)|$ where $\boldsymbol{\beta}_{2,0}$ is the true regression coefficients. The value function of a given treatment regime is calculated using Monte Carlo simulations based on 10,000 replications. The VR at stage 2 (devoted by VR*) is to compare the estimated optimal treatment regime at stage 2 and a randomly assigned treatment at stage 1 as in simulated data with the true optimal dynamic treatment regime for both stages. The VR at stage 1 is to compare the estimated optimal dynamic treatment regime with the true optimal dynamic treatment regime for both stages.

The DR-BOWL methods fail in all settings. Take Setting 1, $n = 300$ as an example, FN = 78.1% for the second stage where the baseline, propensity score and contrast functions are all correctly specified. It missed approximately 3/4 of the important variables. Besides, VR = 50.2, indicating the poor performance of the estimated treatment rules.

On the other hand, the overall performance of our penalized A-learning method is good. We make the following observations. First, the FN rates are much higher than the FP rates. This suggests that the Dantzig selector tends to have conservative variable selection results, which is commonly seen in the literature. Second, the variable selection results and the error rates of the estimated optimal treatment regime at stage 2 are generally much better than those at stage 1, which is expected since the optimal linear treatment decision rule is correctly specified at stage 2 but not at stage 1. At stage 2, for $n = 150$, over 55% important variables are not selected for all 4 settings. Thirdly, our method requires correct specification of either the propensity score or the baseline model, especially when the sample size is small. This is implied by comparing results in Setting 4 with other three settings. For example, when $n = 150$, the false negative at second stage reaches 55.7%, which is much higher than those FN's in other three settings. Besides, our estimator is very efficient in Setting 1 where both models are correctly specified. Even when $n = 150$, the ratio of the value functions reaches 98.7%, and all error rates are abound 6-7%. These results are even comparable with those under Setting 2 and 3 when $n = 300$. Lastly, the estimation and variable selection performance of the estimated optimal dynamic treatment regimes improves as the sample size increases. In particular, in Setting 1-3 when $n = 300$, the VR's are all above 97.9% and ER's are all below 8%, which implies that the estimated optimal treatment regimes nearly maximize the value functions.

4.4.4 Nonregularity

In this section, we examine our methods under settings with different degrees of nonregularity. Specifically, we consider the setting where all covariates in $\boldsymbol{X}_0^{(1)}$ are independent Rademacher random variables. We set $X_0^{(2)}$ to be another Rademacher random variable independent of $\boldsymbol{X}_0^{(1)}$ and $A_0^{(1)}$.

Denoted by $A_0^{(1)*} = 2A_0^{(1)} - 1$, the response Y_0 is generated as follows,

$$Y_0 = 2A_0^{(2)}\{A_0^{(1)*} + \delta_1 X_0^{(1),(1)} + \boldsymbol{X}_0^{(2)} - \delta_2\} + A_0^{(1)}(\boldsymbol{\delta}_0^T \boldsymbol{X}_0^{(1)}) + \varepsilon_0, \tag{4.11}$$

where $X_0^{(1),(1)}$ denotes the first element of $\boldsymbol{X}_0^{(1)}$, $\varepsilon_0 \sim N(0, 0.25)$ and $\mathbb{P}(A_0^{(1)} = 1|\boldsymbol{X}_0^{(1)}) = 0.5$.

For each stage, we fit linear models for the baseline and contrast function, and a logistic regression model for the propensity score. The parameter $\boldsymbol{\delta}_0$ in (4.11) determines the baseline function on the second stage. Similar to the regular case discussed in Section 4.4.1, we also consider four Settings here:

Setting 1: $\boldsymbol{\delta}_0 = \mathbf{0}_q$, $\mathbb{P}(A^{(2)} = 1|\boldsymbol{S}_0) = 0.5$;

Setting 2: $\boldsymbol{\delta}_0 = (\mathbf{0}_4, 1, -1, \mathbf{0}_{q-6})^T$, $\mathbb{P}(A^{(2)} = 1|\boldsymbol{S}_0) = 0.5$;

Setting 3: $\boldsymbol{\delta}_0 = \mathbf{0}_q$, $\mathbb{P}(A^{(2)} = 1|\boldsymbol{S}_0) = \mathbb{P}(N(0,1) \le \boldsymbol{S}_0^T \gamma)$;

Table 4.2: Variable Selection Simulation Results (%).

		Stage 2				Stage 1			
n	method	FN	FP	VR*	ER	FN	FP	VR	ER
Setting 1									
150	PAL	12.6	0.1	64.7	6.1	63.8	0.1	98.3	7.0
	DR-BOWL	85.7	0.1	39.0	34.7	99.5	0.1	39.1	48.3
300	PAL	1.1	0.1	65.4	2.6	41.9	0.1	99.7	6.2
	DR-BOWL	78.1	0.1	49.2	27.5	98.0	0.2	50.2	48.3
Setting 2									
150	PAL	25.9	0.1	57.8	10.4	56.2	0.2	90.8	15.7
	DR-BOWL	86.3	0.1	35.1	35.6	99.0	0.2	35.9	47.2
300	PAL	11.0	0.1	59.6	6.2	32.5	<0.05	97.9	8.0
	DR-BOWL	79.8	0.1	42.4	29.9	97.2	0.2	44.6	47.1
Setting 3									
150	PAL	33.7	0.3	59.9	13.5	64.5	0.1	93.0	9.1
	DR-BOWL	18.8	1.3	60.2	7.5	72.3	0.5	92.4	24.4
300	PAL	12.3	0.3	64.2	7.2	52.7	<0.05	98.3	6.9
	DR-BOWL	74.9	0.2	55.3	23.2	97.8	<0.01	56.4	48.4
Setting 4									
150	PAL	55.7	0.2	48.2	22.4	62.2	0.1	79.4	17.7
	DR-BOWL	75.0	0.1	51.0	23.4	99.0	<0.01	51.7	47.2
300	PAL	26.4	0.3	56.2	13.2	36.4	<0.05	94.3	8.4
	DR-BOWL	74.9	0.2	50.9	23.1	97.4	<0.01	52.8	47.0

FN: proportion of related variables with zero coefficients
FP: proportion of unrelated variables with nonzero coefficients
VR: value ratio between estimated and true treatment regimes
ER: error rate of estimated treatment regimes

Setting 4: $\boldsymbol{\delta}_0 = (\mathbf{0}_4, 1, -1, \mathbf{0}_{q-6})^T$, $\mathbb{P}(A^{(2)} = 1|\boldsymbol{S}_0) = \mathbb{P}(N(0,1) \leq \boldsymbol{S}_0^T\gamma)$,

where $\boldsymbol{\gamma} = (\mathbf{0}_{q-2}, 1, -1, 1)^T$.

Parameters δ_1 and δ_2 in (4.11) controls the degree of nonregularity on the second stage. We consider three choices of δ_1 and δ_2. Set $\delta_1 = 1$, $\delta_2 = 1$, we obtain

$$\mathbb{P}(\tau_2(\boldsymbol{S}_0) = 0) = \mathbb{P}(A^{(1)*} + X_0^{(1),(1)} + X_0^{(2)} = 1) = 0.375.$$

Table 4.3: Simulation for Nonregular Settings

	Stage	Baseline	Propensity Score	Important Variables
Setting 1	Stage 2	right	right	$(S^{(2)}, A_1, S_1^{(1)})$
	Stage 1	right	right	$(S_1^{(1)})$
Setting 2	Stage 2	wrong	right	$(S^{(2)}, A_1, S_1^{(1)})$
	Stage 1	right	right	$(S_1^{(1)}, S_5^{(1)}, S_6^{(1)})$
Setting 3	Stage 2	right	wrong	$(S^{(2)}, A_1, S_1^{(1)})$
	Stage 1	right	right	$(S_1^{(1)})$
Setting 4	Stage 2	wrong	wrong	$(S^{(2)}, A_1, S_1^{(1)})$
	Stage 1	right	right	$(S_1^{(1)}, S_5^{(1)}, S_6^{(1)})$

Set $\delta_1 = 1.1$, $\delta_2 = 1.1$, we have

$$\mathbb{P}(\tau_2(\boldsymbol{S}_0) = 0) = \mathbb{P}(A^{(1)*} + X_0^{(2)} = 0, X_0^{(1),(1)} = 1) = 0.25.$$

Set $\delta_1 = 1$, $\delta_2 = 1.1$, we have

$$\mathbb{P}(\tau_2(\boldsymbol{S}_0) = 0) = 0.$$

With some calculation, we can show the Q-function on the first stage takes the following form:

$$h_1(A_0^{(1)}, \boldsymbol{X}_0^{(1)}) = A_0^{(1)}(\boldsymbol{\delta}_0^T \boldsymbol{X}_0^{(1)} + f_1 X_0^{(1),(1)} + f_2).$$

Hence, the contrast function is correctly specified on the first stage. When $\delta_1 = 1, \delta_2 = 1$ or $\delta_1 = 1.1, \delta_2 = 1.1$, we have $f_1 = f_2 = 1$. When $\delta_1 = 1$, $\delta_2 = 1.1$, we have $f_1 = f_2 = 0.95$. Information about model specification and important variables in the contrast function are given in Table 4.3.

We also consider two choices of sample size, $n = 150$ and $n = 300$, respectively. This gives us a total of 24 scenarios. For each scenario, we report FN, FP, VR and ER as Section 4.3. ER for the first and second stage are calculated as

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} |\mathbb{I}(\widehat{\boldsymbol{\beta}}_1^T \boldsymbol{X}_i > 0) - \mathbb{I}(\tau_1(\boldsymbol{X}_i) > 0)| \mathbb{I}(\tau_1(\boldsymbol{X}_i) \neq 0) \right\} / \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\tau_1(\boldsymbol{X}_i) \neq 0) \right\}$$

and

$$\left\{\frac{1}{n}\sum_{i=1}^{n}|\mathbb{I}(\widehat{\boldsymbol{\beta}}_2^T\boldsymbol{S}_i>0)-\mathbb{I}(\boldsymbol{\beta}_{2,0}^T\boldsymbol{S}_i>0)|\mathbb{I}(\boldsymbol{\beta}_{2,0}^T\boldsymbol{S}_i\neq 0)\right\}/\left\{\frac{1}{n}\sum_{i=1}^{n}I(\boldsymbol{\beta}_{2,0}^T\boldsymbol{S}_i\neq 0)\right\},$$

where $\boldsymbol{\beta}_{2,0}$ corresponds to the true regression coefficients.

Compared to definitions in Section 4.3, error rates here are calculated with respect to those patients with nonzero contrast functions. Such definitions are more meaningful since both two treatments are optimal for these patients. We simulate over 200 replications. Results are reported in Table 4.4.

Within each setting, most results are similar across different choices of δ_1 and δ_2. This suggests the nonregularity issues don't have a big impact on the variable selection results. Apart from results in Setting 4, false negatives and false positives are all very small. When the sample size increases to 300, false negatives for most scenarios are exactly equal to 0 while false positives for all settings are below 0.4%, demonstrating perfect variables selections performance of our methods. In Settings 1-3, most error rates are below 7% while the ratios of value function are all above 85%, indicating our estimated optimal treatment regimes are very close to the truth in these scenarios.

4.5 Application to STAR*D Study

We applied the proposed method to a dataset from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, which was conducted to compare different treatments for patients with major depressive disorder (MDD). There were 4041 participants (age 18-75) with nonpsychotic MDD enrolled in this study. At first level, all participants were treated with citalopram (CIT) up to 14 weeks. Subsequently, 3 more levels of treatments were provided for participants without a satisfactory response to CIT. At each level, participants were randomly assigned to treatment options acceptable to them. At Level 2, participants were eligible for seven treatment options: sertraline (SER), venlafaxine (VEN), bupropion (BUP), cognitive therapy (CT), and augmenting CIT with bupropion (CIT+BUP), buspirone (CIT+BUS) or cognitive therapy (CIT+CT). Participants without a satisfactory response to CT were proceeded to Level 2A for additional medication treatments. All participants who did not respond satisfactorily at Level 2 or 2A were eligible for four treatments at Level 3: medication switch to mirtazapine (MIRT) or nortriptyline (NTP), and medication augmentation with either lithium (Li) or thyroid hormone (THY). Participants without

Table 4.4: Variable Selection Simulation Results for Non-regular Settings (%).

n	Nonregularity	Stage 2 FN	Stage 2 FP	Stage 2 VR*	Stage 2 ER	Stage 1 FN	Stage 1 FP	Stage 1 VR	Stage 1 ER
Setting 1									
150	$\delta_1 = 1, \delta_2 = 1$	0	< 0.01	53.6	0	4.0	0.3	93.9	5.0
	$\delta_1 = 1.1, \delta_2 = 1.1$	0	< 0.01	53.5	0.1	0.5	0.3	95.1	3.2
	$\delta_1 = 1, \delta_2 = 1.1$	0	0.01	52.9	5.0	1.0	0.3	93.8	4.2
300	$\delta_1 = 1, \delta_2 = 1$	0	< 0.01	53.3	0	0	0.4	97.3	0.9
	$\delta_1 = 1.1, \delta_2 = 1.1$	0	< 0.01	53.9	0	0	0.3	97.9	0.9
	$\delta_1 = 1, \delta_2 = 1.1$	0	< 0.01	52.8	2.0	0	0.3	97.0	0.9
Setting 2									
150	$\delta_1 = 1, \delta_2 = 1$	0	< 0.05	46.1	0	14.7	0.3	90.7	5.6
	$\delta_1 = 1.1, \delta_2 = 1.1$	0	< 0.05	45.9	2.0	14	0.3	89.5	5.7
	$\delta_1 = 1, \delta_2 = 1.1$	0	< 0.05	44.4	11.6	9.7	0.3	89.7	11.5
300	$\delta_1 = 1, \delta_2 = 1$	0	< 0.01	45.8	0	0	0.2	97.1	0.4
	$\delta_1 = 1.1, \delta_2 = 1.1$	0	< 0.01	45.6	0.5	0	0.2	96.4	0.4
	$\delta_1 = 1, \delta_2 = 1.1$	0	0.01	45.1	6.8	0	0.2	98.1	7.4
Setting 3									
150	$\delta_1 = 1, \delta_2 = 1$	5.7	0.6	45.0	2.9	19.0	0.2	85.6	4.1
	$\delta_1 = 1.1, \delta_2 = 1.1$	8.2	0.5	45.1	6.6	17.0	0.2	87.3	3.4
	$\delta_1 = 1, \delta_2 = 1.1$	6.3	0.6	44.6	14.6	18.5	0.2	85.8	4.1
300	$\delta_1 = 1, \delta_2 = 1$	0	0.1	53.1	0	0	0.3	97.1	1.0
	$\delta_1 = 1.1, \delta_2 = 1.1$	0	0.1	53.8	1.4	0	0.3	98.0	0.6
	$\delta_1 = 1, \delta_2 = 1.1$	0	0.1	52.9	8.4	0	0.3	97.9	0.8
Setting 4									
150	$\delta_1 = 1, \delta_2 = 1$	20.7	0.5	25.4	8.6	52	0.2	66.6	14.0
	$\delta_1 = 1.1, \delta_2 = 1.1$	20.8	0.5	25.3	12.4	54.2	0.2	62.5	14.9
	$\delta_1 = 1, \delta_2 = 1.1$	21.5	0.6	23.7	22.6	51.7	0.2	61.7	22.1
300	$\delta_1 = 1, \delta_2 = 1$	0.3	0.2	44.9	0.2	3.3	0.2	95.8	0.7
	$\delta_1 = 1.1, \delta_2 = 1.1$	0	0.2	44.8	3.9	0.2	0.2	97.5	0.4
	$\delta_1 = 1, \delta_2 = 1.1$	0	0.2	43.8	13.2	0.3	0.2	97.1	8.3

satisfactory response to Level 3 were re-randomized at Level 4 to either tranylcypromine (TCP) or a combination of mirtazapine and venlafaxine (MIRT+VEN). See Fava et al. (2003) and Rush et al. (2004) for more details of the STAR*D study. One goal of the study is to determine which treatment strategies, in what order or sequence, provide the optimal treatment effect.

As an illustration, we focused on a subset of participants who were given treatment BUP or SER at Level 2 and did not receive satisfactory responses, and then were randomized to treatment MIRT or NTP at Level 3. For this study, we considered 381 covariates collected at baseline and intermediate levels as possible relevant predictors. For treatment regime at Level 3, all the 381 covariates as well as the assigned treatment at Level 2 were considered as possible predictors for making optimal treatment decision. For treatment regime at Level 2, 305 covariates that were collected before giving treatment at Level 2 were considered for making optimal treatment decision. Negative 16-item Quick Inventory of Depressive Symptomatology-Clinician-Rated (QIDS-C_{16}) was used as the final response, which is a measurement of symptomatic status of depression. There are 73 participants who had complete records in the subset of data we are interested in. Among these participants, 36 were treated with BUP and 37 were treated with SER at Level 2, and 33 were treated with NTP and 40 were treated with MIRT at Level 3.

The selection and estimation results are summarized as follows. At Level 3, our method selected two covariates: "age" in baseline demographics (AGE), and the suicide risk of the patient (SUICD). The estimated optimal treatment regime is $\mathbb{I}(1.459 - 0.091 \times \text{AGE} + 0.158 \times \text{SUICD} \geq 0)$, where 1 represents treatment NTP and 0 represents treatment MIRT. This optimal treatment regime assigns 27 participants to NTP and the rest 46 participants to MIRT. At Level 2, our method also selected two covariates: age and QIDS-C percent improvement" in clinic visit clinical record form at Level 1 (QCIMP). The estimated optimal treatment regime is $\mathbb{I}(-8.600+0.145\times\text{AGE}+0.125\times\text{QCIMP} \geq 0)$, where 1 stands for treatment BUP and 0 stands for treatment SER. This optimal treatment regime assigns 37 participants to BUP and the rest 36 participants to SER.

To further examine the estimated optimal dynamic treatment regime, we compare the estimated value function of our estimated optimal treatment regime with values under those four non-dynamic treatment regimes, BUP+NTP, BUP+MIRT, SER+NTP and SER+MIRT. For a given dynamic treatment regime $d = (d^{(1)}, d^{(2)})$, we evaluate its average value function using AIPWE (Zhang et al. 2013),

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}_{A_i}^{(1)}}{\widehat{\pi}_{A_i}^{(1)}}\left(\frac{\mathrm{d}_{A_i}^{(2)}}{\widehat{\pi}_{A_i}^{(2)}}Y_i - \frac{\mathrm{d}_{A_i}^{(2)} - \widehat{\pi}_{A_i}^{(2)}}{\widehat{\pi}_{A_i}^{(2)}}\{d_i^{(2)}(\widehat{h}_{2,i} + \boldsymbol{S}_i^T\widehat{\boldsymbol{\beta}}_2) + (1 - d_i^{(2)})\widehat{h}_{2,i}\}\right)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}_{A_i}^{(1)} - \widehat{\pi}_{A_i}^{(1)}}{\widehat{\pi}_{A_i}^{(1)}}\{d_i^{(1)}(\widehat{h}_{1,i} + \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_1) + (1 - d_i^{(1)})\widehat{h}_{1,i}\},$$

where $d_{A_i}^{(2)} = A_i^{(2)}d_i^{(2)} + (1 - A_i^{(2)})$, $d_{A_i}^{(1)} = A_i^{(1)}d_i^{(1)} + (1 - A_i^{(1)})(1 - d_i^{(1)})$, $\widehat{\pi}_{A_i}^{(2)} = A_i^{(2)}\widehat{\pi}_{2,i} + (1 - A_i^{(2)})(1 - \widehat{\pi}_{2,i})$, $\widehat{\pi}_{A_i}^{(1)} = A_i^{(1)}\widehat{\pi}_{1,i} + (1 - A_i^{(1)})(1 - \widehat{\pi}_{1,i})$, $d_i^{(2)}$ and $d_i^{(1)}$ the assigned treatment for the ith patient, according to $d^{(2)}$ and $d^{(1)}$. Based on this formula, we report the estimated value functions of the four non-dynamic treatment regimes in Table 4.5.

Estimating the value of the optimal treatment regime is well-known to be a non-regular problem when there's nonzero probability that the contrast function (either at the second or the first stage) is equal to zero. To evaluate the value function under our estimated optimal treatment regime, we consider the online estimator proposed by Luedtke and van der Laan (2016). Specifically, for $i = l_n + 1, l_n + 2, \dots, n$, we obtain the estimated optimal dynamic treatment regime $\widehat{d}^{opt(i)} = (\widehat{d}^{opt(i)(1)}, \widehat{d}^{opt(i)(2)})$ and its associated parameters $\widehat{\beta}_2^{(i)}$, $\widehat{\beta}_1^{(i)}$, propensity score function $\widehat{\pi}^{(i)(2)}$, $\widehat{\pi}^{(i)(1)}$, baseline function $\widehat{h}^{(i)(2)}$, $\widehat{h}^{(i)(1)}$ based on data from patients 1 to $i-1$, using penalized A-learning. Then we evaluate the value of $\widehat{d}^{opt(i)} = (\widehat{d}^{opt(i)(1)}, \widehat{d}^{opt(i)(2)})$ on the ith patient using (AIPWE, Zhang et al. 2013)

$$\begin{aligned}
\widehat{V}_i(i) &= \frac{\widehat{d}_{A_i}^{opt(i)(1)}}{\widehat{\pi}_{A_i}^{(i)(1)}} \left(\frac{\widehat{d}_{A_i}^{opt(i)(2)}}{\widehat{\pi}_{A_i}^{(i)(2)}} Y_i - \frac{\widehat{d}_{A_i}^{opt(i)(2)} - \widehat{\pi}_{A_i}^{(i)(2)}}{\widehat{\pi}_{A_i}^{(i)(2)}} \{ \widehat{d}_i^{opt(i)(2)} (\widehat{h}_i^{(i)(2)} \right. \\
&+ \left. \boldsymbol{S}_i^T \widehat{\beta}_2^{(i)}) + (1 - \widehat{d}_i^{opt(i)(2)}) \widehat{h}_i^{(i)(2)} \} \right) - \frac{\widehat{d}_{A_i}^{opt(i)(1)} - \widehat{\pi}_{A_i}^{(i)(1)}}{\widehat{\pi}_{A_i}^{(i)(1)}} \{ \widehat{d}_i^{opt(i)(1)} (\widehat{h}_i^{(i)(1)} \\
&+ \boldsymbol{X}_i^T \widehat{\beta}_1^{(i)}) + (1 - \widehat{d}_i^{opt(i)(1)}) \widehat{h}_i^{(i)(1)} \}.
\end{aligned}$$

The variance of $\widehat{V}_i(i)$ conditional on data from patients 1 to $i-1$ is evaluated by

$$\widetilde{\sigma}_i^2 = \frac{1}{i-1} \sum_{j=1}^{i-1} \widehat{V}_i^2(j) - \left(\frac{1}{i-1} \sum_{j=1}^{i-1} \widehat{V}_i(j) \right)^2,$$

where $\widehat{V}_i(j)$ is the estimated value of $\widehat{d}^{opt(i)}$ on the jth patient.

The final estimator is given by

$$\widehat{V} = \frac{\sum_{j=l_n+1}^{n} \widetilde{\sigma}_j^{-1} \widehat{V}_j(j)}{\sum_{j=l_n+1}^{n} \widetilde{\sigma}_j^{-1}},$$

with the estimated standard error

$$\widehat{\sigma} = \frac{\sqrt{n - l_n}}{\sum_{j=l_n+1}^{n} \widetilde{\sigma}_j^{-1}}.$$

Since the sample size of our dataset is small, we choose $l_n \approx 2n/3$, i.e, $l_n = 49$. The estimated value $\widehat{V}$ is equal to −9.02 with an estimated standard error $\widehat{\sigma} = 1.66$. From Table 4.5, we can

Table 4.5: Estimated Values of Different Treatment Regimes

Treatment Regime	Estimated Value
estimated optimal regime	-9.02
BUP + NTP	-12.86
BUP + MIRT	-12.57
SER + NTP	-12.57
SER + MIRT	-12.28

see the value under our estimated treatment regime is much larger than those under four nondynamic treatment regime.

4.6 Oracle inequalities for $\widehat{\beta}_2$ and the value function of the estimated regime at the second stage

We first introduce some notation. For an arbitrary vector $\phi \in \mathbb{R}^M$, the support of ϕ is defined by $\text{supp}(\phi) = \{j \in \{1, \ldots, M\} : \phi^{(j)} \neq 0\}$. For any positive definite matrix $\mathbf{\Phi}$, define

$$\rho^s_{\min}(\mathbf{\Phi}) = \min_{\substack{\|\boldsymbol{y}\|_2=1 \\ |supp(\boldsymbol{y})| \le s}} \|\mathbf{\Phi}^{1/2}\boldsymbol{y}\|_2 \quad \text{and} \quad \rho^s_{\max}(\mathbf{\Phi}) = \max_{\substack{\|\boldsymbol{y}\|_2=1 \\ |supp(\boldsymbol{y})| \le s}} \|\mathbf{\Phi}^{1/2}\boldsymbol{y}\|_2.$$

To simplify the analysis, we assume the residual $e_0 = Y_0 - h_2(\boldsymbol{S}_0) - A_0^{(2)}\boldsymbol{S}_0^T\boldsymbol{\beta}_{2,0}$ is independent of $A_0^{(2)}$ and $\boldsymbol{S}_0$.

4.6.1 Oracle inequality for $\widehat{\beta}_2$

Recall $\tau_2(\boldsymbol{s}) = \boldsymbol{s}^T\boldsymbol{\beta}_2$, according to our assumption. Let $\boldsymbol{\beta}_{2,0}$ denote the true values of $\boldsymbol{\beta}_2$. Define $s_{\boldsymbol{\beta}_2} = |\mathcal{M}_{\boldsymbol{\beta}_2}| = O(n^{l_6})$ for some $0 \le l_6 < 1$, the nonsparsity size of $\boldsymbol{\beta}_{2,0}$, $\mathcal{M}_{\boldsymbol{\beta}_2}$ the support of $\boldsymbol{\beta}_{2,0}$. We allow the number of covariates p to grow exponentially fast with respect to the sample size n, i.e, $\log p = O(n^{a_2})$ for some $0 < a_2 < 1$. To deal with such NP-dimensionality, following Zhou (2009), we assume

$$\boldsymbol{S}_0 = \mathbf{\Sigma}^{1/2}\boldsymbol{U}_0, \tag{4.12}$$

where all the diagonal elements of $\boldsymbol{\Sigma}$ equal 1, and $\boldsymbol{U}_0$ is some p-dimensional isotropic random vector. More specifically, we require that for any vector $\boldsymbol{a} \in \mathbb{R}^p$,

$$\mathbb{E}(\boldsymbol{a}^T \boldsymbol{U}_0)^2 = \boldsymbol{a}^T \boldsymbol{a} \text{ and } \|\boldsymbol{a}^T \boldsymbol{U}_0\|_{\psi_2} \leq \omega \|\boldsymbol{a}\|_2, \tag{4.13}$$

for some isotropic constants ω. Thus, each $\boldsymbol{S}_i$ can be represented as $\boldsymbol{S}_i = \boldsymbol{\Sigma}^{1/2} \boldsymbol{U}_i$ for some isotrpoic random vector $\boldsymbol{U}_i \overset{d}{=} \boldsymbol{U}_0$.

The definition of the isotropic random vector was firstly introduced by Milman and Pajor (2003). Independent normal and independent Rademacher random variables are two most important examples of isotropic random vectors. More generally, coordinates of the isotropic random vector do not need to be independent. They can be distributed uniformly on various convex and symmetric bodies, for example, an appropriate multiple of the unit ball in $\mathbb{R}^p$ equipped with the ℓ_k-norm for any $1 \leq k \leq \infty$. For these distributions, we denote ω_k as their isotropic constants. It is further shown in (Milman and Pajor 1989) that ω_ks are uniformly bounded for $k \geq 1$. However, it remains unknown whether the isotropic property holds for all uniform distributions on arbitrary symmetric convex bodies with Lebesgue measure 1.

The isotropic formulation requires covariates in $\boldsymbol{U}_0$ to be uncorrelated, and hence does not allow for correlated Bernoullis. However, according to our definition $\boldsymbol{S}_0 = \boldsymbol{\Sigma}^{1/2} \boldsymbol{U}_0$, different covariates in the design matrix $\boldsymbol{S}_0$ can be correlated when $\boldsymbol{\Sigma}$ is not a diagonal matrix. Such formulations allows us to impose conditions on the tail of $\boldsymbol{U}_0$ and the covariance matrix $\boldsymbol{\Sigma}$ separately.

Since the A-learning estimating equation involves the plug-in estimators $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$, we need some conditions on these two estimators to establish oracle inequalities for $\widehat{\boldsymbol{\beta}}_2$. More precisely, we assume that $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$ converge to some $\boldsymbol{\alpha}_{\boldsymbol{2}}^*$ and $\boldsymbol{\theta}_{\boldsymbol{2}}^*$, respectively. When the propensity score model and the baseline model are correctly specified, $\boldsymbol{\alpha}_{\boldsymbol{2}}^*$ and $\boldsymbol{\theta}_{\boldsymbol{2}}^*$ represent the true coefficients in these two models. When the models are misspecified, $\boldsymbol{\alpha}_{\boldsymbol{2}}^*$ and $\boldsymbol{\theta}_{\boldsymbol{2}}^*$ correspond to the population-level least favorable parameters. Denote $\mathcal{M}_{\alpha_2}$ and $\mathcal{M}_{\theta_2}$ the support of $\boldsymbol{\alpha}_{\boldsymbol{2}}^*$ and $\boldsymbol{\theta}_{\boldsymbol{2}}^*$, respectively. Let $s_{\alpha_2} = |\mathcal{M}_{\alpha_2}|$ and $s_{\theta_2} = |\mathcal{M}_{\theta_2}|$, the number of nonzero elements. We assume $s_{\alpha_2} = O(n^{l_4})$ and $s_{\theta_2} = O(n^{l_5})$ for some $0 \leq l_4, l_5 < 1/2$.

(B1.) Assume that there exist some positive constants γ_{α_2} and γ_{θ_2}, such that with probability at least $1 - O(n^{-1})$,

$$\widehat{\boldsymbol{\alpha}}_2^{\mathcal{M}_{\alpha_2}^c} = 0, \ \ \|\widehat{\boldsymbol{\alpha}}_2^{\mathcal{M}_{\alpha_2}} - \boldsymbol{\alpha}_{\boldsymbol{2}}^{*\,\mathcal{M}_{\alpha_2}}\|_\infty \leq O(1) n^{-\gamma_{\alpha_2}} \log n, \tag{4.14}$$

$$\widehat{\boldsymbol{\theta}}_2^{\mathcal{M}_{\theta_2}^c} = 0, \ \ \|\widehat{\boldsymbol{\theta}}_2^{\mathcal{M}_{\theta_2}} - \boldsymbol{\theta}_{\boldsymbol{2}}^{*\,\mathcal{M}_{\theta_2}}\|_\infty \leq O(1) n^{-\gamma_{\theta_2}} \log n, \tag{4.15}$$

where $O(1)$ denotes some positive universal constant. Moreover, assume $d_{\alpha_2} \gg n^{-\gamma_{\alpha_2}} \log n$ and $d_{\theta_2} \gg n^{-\gamma_{\theta_2}} \log n$, where $d_{\alpha_2} = \min_{j \in \mathscr{M}_{\alpha_2}} |\boldsymbol{\alpha}_2^{*(j)}|/2$ and $d_{\theta_2} = \min_{j \in \mathscr{M}_{\theta_2}} |\boldsymbol{\theta}_2^{*(j)}|/2$.

Condition B1 assumes the weak oracle properties of $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$, i.e., selection consistency and consistency under L_∞ norm. The weak oracle properties of $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$ are established in Theorems 4.8.1 and 4.8.2 of Section 4.8, respectively.

Define

$$\boldsymbol{C}^{(2)} = \mathbb{E}\{\boldsymbol{S}_i \pi_{2,i}^* (1-\pi_{2,i}^*) \boldsymbol{S}_i^T\}, \;\; \boldsymbol{D}^{(2)} = \mathbb{E}\{\boldsymbol{S}_i \boldsymbol{S}_i^T (1 - A_i^{(2)})\},$$

and $\pi_{2,i}^* \equiv \pi^{(2)}(\boldsymbol{X}_i, \boldsymbol{\alpha}_2^*)$.

(B2.) Assume that matrices $\boldsymbol{D}^{(2)}$, $\boldsymbol{C}^{(2)}$ and $\boldsymbol{\Sigma}$ satisfy

$$\lambda_{\max}(\boldsymbol{\Sigma}_{\mathscr{M}_{\alpha_2}}^{\mathscr{M}_{\alpha_2}}) = O(1), \qquad \lambda_{\max}(\boldsymbol{\Sigma}_{\mathscr{M}_{\theta_2}}^{\mathscr{M}_{\theta_2}}) = O(1),$$
$$\liminf_n \lambda_{\min}(\boldsymbol{D}_{\mathscr{M}_{\theta_2}}^{(2)\mathscr{M}_{\theta_2}}) > 0, \qquad \liminf_n \lambda_{\min}(\boldsymbol{C}_{\mathscr{M}_{\alpha_2}}^{(2)\mathscr{M}_{\alpha_2}}) > 0.$$

Define $\boldsymbol{\Omega}^{(2)}(\alpha_2) = \mathbb{E}[\boldsymbol{S}_i \boldsymbol{S}_i^T A_i^{(2)}\{1-\pi_2(\boldsymbol{S}_i, \alpha_2)\}]$ and $\boldsymbol{\Omega}_n^{(2)} = n^{-1}\sum_i \boldsymbol{S}_i \boldsymbol{S}_i^T A_i^{(2)}(1-\widehat{\pi}_{2,i})$. For any positive semidefinite matrix $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$, integer s and positive number c, define function $K(s, c, \boldsymbol{\Psi})$ as follows,

$$K(s, c, \boldsymbol{\Psi}) = \min_{\substack{J \subset \{1,\ldots,p\} \\ |J| \le s}} \; \min_{\substack{\boldsymbol{y} \neq 0 \\ \|\boldsymbol{y}^{J^c}\|_1 \le c \|\boldsymbol{y}^J\|_1}} \frac{\|\boldsymbol{\Psi}^{1/2}\boldsymbol{y}\|_2}{\|\boldsymbol{y}^J\|_2} > 0.$$

The following condition ensures that the RE condition holds for the matrix $\boldsymbol{\Omega}_n^{(2)}$.

(B3.) Assume that for any $0 < \theta < 1$, we have with probability at least $1 - O(n^{-1})$ that

$$K(s_{\boldsymbol{\beta}_2}, 1, \boldsymbol{\Omega}_n^{(2)}) > (1-\theta) \inf_{\boldsymbol{\alpha}_2 \in H_{\alpha_2}} K(s_{\boldsymbol{\beta}_2}, 1, \boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2)) > 0, \tag{4.16}$$

where H_{α_2} denotes the set of vectors $\boldsymbol{\alpha}_2$ that satisfies the weak oracle property (4.14).

It is tedious to verify (4.16) due to the plug-in estimator $\widehat{\pi}_{2,i}$. The key to prove such a result is that the estimator $\widehat{\boldsymbol{\alpha}}_2$ in $\widehat{\pi}_{2,i}$ should be sparse. That is the reason we use penalized regression with a folded-concave penalty to obtain $\widehat{\boldsymbol{\alpha}}_2$, since it can ensure selection consistency of the estimator. We provide a general result characterizing the UUP and RE conditions for the random matrix $\boldsymbol{\Omega}_n^{(2)}$ in Lemmas 4.9.1 and 4.9.2 of Section 4.9.

To establish the oracle inequality for $\widehat{\boldsymbol{\beta}}_2$, we first provide an upper bound for

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{S}_i (A_i^{(2)} - \widehat{\pi}_{2,i})(Y_i - \boldsymbol{S}_i^T \widehat{\boldsymbol{\theta}}_2 - A_i^{(2)} \boldsymbol{S}_i^T \boldsymbol{\beta}_{2,0}) \right\|_\infty ,$$

which is given in the following Lemma.

Lemma 4.6.1 *Assume that Condition (B1) and (B2) hold,* $\|h_2(\boldsymbol{S}_0) - \boldsymbol{S}_0^T \boldsymbol{\theta}_2^*\|_{\psi_1} < \infty$, $\|e_0\|_{\psi_2} < \infty$, $a_2 + l_4 < 1$, *and that either* π_2 *or* h_2 *is correctly specified. Then, there exists some constant* $c^{(2)} > 0$ *such that we have with probability at least* $1 - O(n^{-1})$ *that*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{S}_i (A_i^{(2)} - \widehat{\pi}_{2,i})(Y_i - \boldsymbol{S}_i^T \widehat{\boldsymbol{\theta}}_2 - A_i^{(2)} \boldsymbol{S}_i^T \boldsymbol{\beta}_{2,0}) \right\|_\infty \le c^{(2)}(E_1 + E_2 + E_3 + E_4),$$

where

$$\begin{aligned} E_1 &= \sqrt{\log p / n}, \quad E_2 = s_{\alpha_2} n^{-2\gamma_{\alpha_2}} \log^2 n + s_{\theta_2} n^{-2\gamma_{\theta_2}} \log^2 n, \\ E_3 &= \sigma_3 \{ \sqrt{s_{\alpha_2} \log n / n} + \sqrt{s_{\alpha_2}} \lambda_{1n}^{(2)} \rho_2^{(1)}(d_{n\alpha_2}) \}, \\ E_4 &= \sigma_4 \{ \sqrt{s_{\theta_2} \log n / n} + \sqrt{s_{\theta_2}} \lambda_{2n}^{(2)} \rho_2^{(2)}(d_{n\theta_2}) \}, \end{aligned}$$

$\sigma_3^2 = \mathbb{E}\{h_2(\boldsymbol{S}_0) - \boldsymbol{S}_0^T \boldsymbol{\theta}_2^*\}^2$ *and* $\sigma_4^2 = \mathbb{E}[\pi_2(\boldsymbol{S}_0) - \{1 + \exp(-\boldsymbol{S}_0^T \boldsymbol{\alpha}_2^*)\}^{-1}]^2$.

Recall that $\log p = O(n^{a_2})$, $s_{\alpha_2} = O(n^{l_4})$ for some $0 \le a_2, l_4 < 1$. The condition $a_2 + l_4 < 1$ implies $n \gg s_{\alpha_2} \log p$. Here, E_1 describes how the curse of dimensionality takes effect, E_2 is due to estimation errors of $\widehat{\boldsymbol{\alpha}}^{(2)}$ and $\widehat{\boldsymbol{\theta}}^{(2)}$, E_3 and E_4 are due to model misspecification. Since we assume that at least one of h_2 and π_2 is correctly specified, either E_3 or E_4 is zero.

Theorem 4.6.1 *Assume that conditions in Lemma 4.6.1 and Condition B3 hold, and* $\lambda_{3n}^{(2)} \ge c^{(2)}(E_1 + E_2 + E_3 + E_4)$ *where the constant* $c^{(2)}$ *is defined in Lemma 4.6.1. Then, for some fixed* $0 < \theta < 1$, *the following two inequalities hold with probability at least* $1 - O(n^{-1})$:

$$\|\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}\|_2 \le \frac{12 \lambda_{3n}^{(2)} \sqrt{s_{\boldsymbol{\beta}_2}}}{(1-\theta)^2 \inf_{\boldsymbol{\alpha}_2 \in H_{\alpha_2}} K^2(s_{\boldsymbol{\beta}_2}, 1, \boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2))}, \tag{4.17}$$

$$\|\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}\|_1 \le \frac{8 \lambda_{3n}^{(2)} s_{\boldsymbol{\beta}_2}}{(1-\theta)^2 \inf_{\boldsymbol{\alpha}_2 \in H_{\alpha_2}} K^2(s_{\boldsymbol{\beta}_2}, 1, \boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2))}. \tag{4.18}$$

Moreover, we have $\|\widehat{\boldsymbol{\beta}}_2^{\mathscr{M}_{\beta_2}^c}\|_1 \le \|\widehat{\boldsymbol{\beta}}_2^{\mathscr{M}_{\beta_2}} - \boldsymbol{\beta}_{2,0}^{\mathscr{M}_{\beta_2}}\|_1$.

From (4.17), it is immediate to see that $\|\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}\|_2 \xrightarrow{P} 0$ as long as

$$\frac{\sqrt{s_{\boldsymbol{\beta}_2}}(E_1 + E_2 + E_3 + E_4)}{\inf_{\alpha_2 \in H_{\alpha_2}} K^2(s_{\boldsymbol{\beta}_2}, 1, \boldsymbol{\Omega}^{(2)}(\alpha_2))} \to 0, \tag{4.19}$$

which implies the doubly robust property of $\widehat{\boldsymbol{\beta}}_2$. We provide a sufficient condition for (4.19) in the following Corollary.

Corollary 4.6.1 (Double robustness of $\widehat{\boldsymbol{\beta}}_2$) *Assume that conditions in Theorem 4.6.1 and the following conditions hold:*

$$l_6 < \min(4\gamma_{\theta_2} - 2l_5, 4\gamma_{\alpha_2} - 2l_4), \tag{4.20}$$

$$\lambda_{2n}^{(2)} \rho_2^{(2)}(d_{n\theta_2}) = O(n^{-1/2}) \quad \textit{and} \quad \lambda_{1n}^{(2)} \rho_1^{(2)}(d_{n\alpha_2}) = O(n^{-1/2}). \tag{4.21}$$

$$\liminf \inf_{\boldsymbol{\alpha}_2 \in H_{\alpha_2}} K(s_{\boldsymbol{\beta}_2}, 1, \boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2)) > 0. \tag{4.22}$$

If either the baseline h_2 or the propensity score model π_2 is correctly specified, then $\|\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}\|_2 \xrightarrow{P} 0$.

Condition (4.20) imposes a constraint between the sparsity of population parameters and the convergence rates of $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$. When $s_{\boldsymbol{\beta}_2} = O(1)$, it requires $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$ to be consistent under L_2 norm. Condition (4.21) automatically holds for SCAD penalty function when $d_{n\theta_2} \gg \lambda_{2n}^{(2)}$ and $d_{n\alpha_2} \gg \lambda_{1n}^{(2)}$.

4.6.2 Oracle inequality for the value function of the estimated regime at the second stage

Now we establish the error bound for the difference between the mean responses (i.e. the value functions) of the estimated optimal regime at the second stage $\widehat{d}_2(\boldsymbol{S}_0) = \mathbb{I}(\boldsymbol{S}_0^T \widehat{\boldsymbol{\beta}}_2 > 0)$ and the true optimal one $d_2^{opt,0}(\boldsymbol{S}_0) = \mathbb{I}(\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0} > 0)$ for an individual with covariate $\boldsymbol{S}_0$, distributed according to (4.12), independent of $\boldsymbol{S}_i, i = 1, \dots, n$. In addition, the regime at the first stage is chosen the same as the actually received treatment $A_0^{(1)}$ at the first stage. Notice that the difference of the corresponding value functions is given by

$$\begin{aligned} &\mathbb{E}\{Y_0^*(A_0^{(1)}, d_2^{opt,0})\} - \mathbb{E}\{Y_0^*(A_0^{(1)}, \widehat{d}_2)\} \\ &= \mathbb{E}[\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0}\{\mathbb{I}(\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0} > 0) - \mathbb{I}(\boldsymbol{S}_0^T \widehat{\boldsymbol{\beta}}_2 > 0)\}]. \end{aligned} \tag{4.23}$$

Since (4.23) is nonnegative, it suffices to provide an upper bound. Here, we impose the following condition.

(B4.) The probability density function $g^{(2)}(\cdot)$ of $\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0}$ exists and is bounded.

Condition B4 is a mild condition on the true optimal decision function, which holds in most cases when at least one of the important covariates (the corresponding component of $\boldsymbol{\beta}_{2,0}$ is nonzero) is continuous.

Theorem 4.6.2 *Assume that conditions in Theorem 4.6.1 and Condition B4 hold. Assume* $\mathbb{E}(\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0})^2 = O(1)$*. Then, for any fixed* $0 < \theta < 1$*,*

$$\mathbb{E}\{Y_0^*(A_0^{(1)}, d_2^{opt})\} - \mathbb{E}\{Y_0^*(A_0^{(1)}, \widehat{d}_2)\} \leq O(1)\frac{\omega}{n} + O(1)\frac{\omega^2 \rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma})\left(\lambda_{3n}^{(2)}\right)^2 s_{\beta_2} \log^2 n}{(1-\theta)^4 \inf_{\alpha_2 \in H_{\alpha_2}} K^4(s_{\beta_2}, 1, \boldsymbol{\Omega}^{(2)}(\alpha_2))},$$

where $O(1)$ *denotes some universal constant.*

Error bound for the difference of the value functions follows from the error bound on $\widehat{\boldsymbol{\beta}}_2$ and Condition B4. Since the first term in the upper bound is small, the difference of the value functions is mainly characterized by the second term.

4.7 Error bounds for $\widehat{\boldsymbol{\beta}}_1$ and the value function of the estimated dynamic treatment regime

4.7.1 Misspecified contrast function

In the context of A-learning, a major challenge arising in multi-stage studies is that the contrast functions are likely to be misspecified in backward induction. In order to study the finite sample bounds of $\widehat{\boldsymbol{\beta}}_1$, we need to first define least favorable parameters under the misspecification of the contrast function.

Recall that $\tau_1(\cdot)$ is the true contrast function for the ith patient, which can be a very complex function due to the backward induction. For notational convenience, we use a shorthand $\tau(\boldsymbol{x})$ for $\tau_1(\boldsymbol{x})$. We posit a linear model $\boldsymbol{x}^T \boldsymbol{\beta}_1$ for $\tau(\boldsymbol{x})$, which is often misspecified. When either the propensity score model π_1 or the baseline mean function h_1 is correctly specified, the associated least favorable parameters $\boldsymbol{\beta}_1^*$ is defined as follows:

$$\boldsymbol{\beta}_1^* = \arg\min_{\boldsymbol{\beta}_1 \in \Lambda^*} \|\boldsymbol{\beta}_1\|_1, \tag{4.24}$$

where

$$\Lambda^* = \left\{\boldsymbol{\beta}_1 \in \mathbb{R}^q : \|\mathbb{E}[\boldsymbol{X}_0 A_0^{(1)}(1-\pi_{1,0}^*)\{\tau(\boldsymbol{X}_0) - \boldsymbol{X}_i^T \boldsymbol{\beta}_1\}]\|_\infty \le \kappa_0\right\},$$

$\pi_{1,i}^* = \pi_1(\boldsymbol{X}_i, \boldsymbol{\alpha}_1^*)$ for $i = 0, 1, \ldots, n$ and κ_0 is a nonnegative constant. Define

$$\kappa_0^* = \|\mathbb{E}[\boldsymbol{X}_0 A_0^{(1)}(1-\pi_{1,0}^*)\{\tau(\boldsymbol{X}_i) - S_i^T \boldsymbol{\beta}_1^*\}]\|_\infty.$$

By simple algebra, we can show $\kappa_0^* \le \min\{\kappa_0, O(\sigma_0)\}$, where $\sigma_0^2 = \mathbb{E}[\{\tau(\boldsymbol{X}_0) - \boldsymbol{X}_0^T \boldsymbol{\beta}_1^*\}^2]$, describing the degree of misspecification of the contrast function. Define $s_{\boldsymbol{\beta}_1} = |M_{\boldsymbol{\beta}_1}| = O(n^{l_3})$ for some $0 \le l_3 < 1/2$, where $\mathscr{M}_{\boldsymbol{\beta}_1}$ corresponds to the support of $\boldsymbol{\beta}_1^*$.

4.7.2 Error bound for $\widehat{\boldsymbol{\beta}}_1$

Assume that $\log q = O(n^{a_1})$ for some $0 < a_1 < 1$ and

$$\boldsymbol{X}_0 \overset{d}{=} \boldsymbol{\Psi}^{1/2} \boldsymbol{V}_0, \tag{4.25}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{q \times q}$ is some positive definite matrix with diagonal elements equal to 1, and $\boldsymbol{V}_0$ is a q-dimensional isotropic random vector with some isotropic constants ζ. As in the second stage, we first give conditions on $\widehat{\boldsymbol{\alpha}}_1$ and $\widehat{\boldsymbol{\theta}}_1$. Assume that these two estimators converge to some $\boldsymbol{\alpha}_1^\star$ and $\boldsymbol{\theta}_1^\star$, respectively, under possible model misspecification. Denote $\mathscr{M}_{\boldsymbol{\alpha}_1} = \text{supp}(\boldsymbol{\alpha}_1^\star)$, $\mathscr{M}_{\boldsymbol{\theta}_1} = \text{supp}(\boldsymbol{\theta}_1^\star)$, $s_{\boldsymbol{\alpha}_1} = |\mathscr{M}_{\boldsymbol{\alpha}_1}| = O(n^{l_1})$, and $s_{\boldsymbol{\theta}_1} = |\mathscr{M}_{\boldsymbol{\theta}_1}| = O(n^{l_2})$ for some $0 \le l_1, l_2 < 1/2$.

(B5.) Assume that there exist some positive constants $\gamma_{\boldsymbol{\alpha}_1}$ and $\gamma_{\boldsymbol{\theta}_1}$, with probability at least $1 - O(n^{-1})$, the following holds:

$$\widehat{\boldsymbol{\alpha}}_1^{\mathscr{M}_{\boldsymbol{\alpha}_1}^c} = 0, \;\; \|\widehat{\boldsymbol{\alpha}}_1^{\mathscr{M}_{\boldsymbol{\alpha}_1}} - \boldsymbol{\alpha}_1^{\star \mathscr{M}_{\boldsymbol{\alpha}_1}}\|_\infty \le O(1) n^{-\gamma_{\boldsymbol{\alpha}_1}} \log n, \tag{4.26}$$

$$\widehat{\boldsymbol{\theta}}_1^{\mathscr{M}_{\boldsymbol{\theta}_1}^c} = 0, \;\; \|\widehat{\boldsymbol{\theta}}_1^{\mathscr{M}_{\boldsymbol{\theta}_1}} - \boldsymbol{\theta}_1^{\star \mathscr{M}_{\boldsymbol{\theta}_1}}\|_\infty \le O(1) n^{-\gamma_{\boldsymbol{\theta}_1}} \log n, \tag{4.27}$$

where $O(1)$ denotes some positive constant. Moreover, assume $d_{\boldsymbol{\alpha}_1} \gg n^{-\gamma_{\boldsymbol{\alpha}_1}} \log n$ and $d_{\boldsymbol{\theta}_1} \gg n^{-\gamma_{\boldsymbol{\theta}_1}} \log n$, where $d_{\boldsymbol{\alpha}_1} = \min_{j \in \mathscr{M}_{\alpha_1}} |\alpha_1^{*(j)}|/2$ and $d_{\boldsymbol{\theta}_1} = \min_{j \in \mathscr{M}_{\boldsymbol{\theta}_1}} |\theta_1^{*(j)}|/2$.
(B6.) Assume that $\boldsymbol{D}^{(1)}$, $\boldsymbol{C}^{(1)}$ and $\boldsymbol{\Psi}$ satisfy

$$\lambda_{\max}(\boldsymbol{\Psi}_{\mathscr{M}_{\boldsymbol{\alpha}_1}}^{\mathscr{M}_{\boldsymbol{\alpha}_1}}) = O(1), \qquad \lambda_{\max}(\boldsymbol{\Psi}_{\mathscr{M}_{\boldsymbol{\theta}_1}}^{\mathscr{M}_{\boldsymbol{\theta}_1}}) = O(1),$$
$$\liminf_n \lambda_{\min}(\boldsymbol{D}_{\mathscr{M}_{\boldsymbol{\theta}_1}}^{(1)M_{\boldsymbol{\theta}_1}}) > 0, \qquad \liminf_n \lambda_{\min}(\boldsymbol{C}_{\mathscr{M}_{\boldsymbol{\alpha}_1}}^{(1)M_{\boldsymbol{\alpha}_1}}) > 0,$$

where

$$\boldsymbol{D}^{(1)} = \mathbb{E}\{\boldsymbol{X}_0\boldsymbol{X}_0^T(1-A_0^{(1)})\},\ \ \boldsymbol{C}^{(1)} = \mathbb{E}\{\boldsymbol{X}_0\boldsymbol{X}_0^T\pi_{1,0}^*(1-\pi_{1,0}^*)\},$$

and $\pi_{1,0}^* = \pi_1(\boldsymbol{X}_0, \boldsymbol{\alpha}_1^*)$.

Since both the propensity score model and the contrast function at the first stage can be misspecified, we need the following condition to control their effect on estimation of β_1^*.
(B7.) Assume that

$$\tau_0 \equiv \|\boldsymbol{F}^{\mathcal{M}_{\alpha_1}}[\boldsymbol{C}^{(1)\mathcal{M}_{\alpha_1}}_{\mathcal{M}_{\alpha_1}}]^{-1}\boldsymbol{b}^{(1)\mathcal{M}_{\alpha_1}}\|_\infty < \infty, \tag{4.28}$$

where $\boldsymbol{b}^{(1)} = \mathbb{E}\{\boldsymbol{X}_0(A_0^{(1)} - \pi_{1,0}^*)\}$ and

$$\boldsymbol{F} = \mathbb{E}[\boldsymbol{X}_0 A_0^{(1)}\pi_{1,0}^*(1-\pi_{1,0}^*)\{\tau(\boldsymbol{X}_0) - \boldsymbol{X}_0^T\boldsymbol{\beta}_1^*\}\boldsymbol{X}_0^T].$$

It is immediate to see $\tau_0 = 0$ when either the contrast function or the propensity score model is correctly specified.

When going back to the first stage, the error bound of $\widehat{\boldsymbol{\beta}}_1$ is directly affected by that of $\widehat{\boldsymbol{\beta}}_2$. This is because the estimated response $\widehat{V}_i$ in the first stage is obtained based on $\widehat{\boldsymbol{\beta}}_2$ using the advantage function. To simplify presentation, we introduce the following condition.
(B8.) Assume that with probability at least $1 - O(n^{-1})$, there exists some constant $\mu_1 > 0$ such that

$$\sqrt{\rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma})}\|\widehat{\boldsymbol{\beta}}_2 - \beta_{2,0}\|_2 \leq O(1)n^{-\mu_1}\log n), \tag{4.29}$$

and $\|\widehat{\boldsymbol{\beta}}_2^{\mathcal{M}_{\beta_2}^c}\|_1 \leq \|\widehat{\boldsymbol{\beta}}_2^{\mathcal{M}_{\beta_2}} - \beta_{2,0}^{\mathcal{M}_{\beta_2}}\|_1$, where $O(1)$ denotes some positive constant.

A more explicit form of the error bound for (4.29) is given in Theorem 4.6.1. In the next Lemma, we provide an upper bound for the term:

$$\frac{1}{n}\|\sum_{i=1}^n \boldsymbol{X}_i^T(A_i^{(1)} - \widehat{\pi}_{1,i})(\widehat{V}_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\theta}}_1 - A_i^{(1)}\boldsymbol{X}_i^T\boldsymbol{\beta}_1^*)\|_\infty. \tag{4.30}$$

Lemma 4.7.1 *Assume that Conditions B5-B8 and those in Theorem 4.6.1 hold,* $\|\tau(\boldsymbol{X}_0) - \boldsymbol{X}_0^T\boldsymbol{\beta}_1^*\|_{\psi_1} < \infty$, $\|V_0^{(2)}(A_0^{(1)}, \bar{\boldsymbol{X}}_0^{(2)*}(A_0^{(1)}) - \mathbb{E}(V_0^{(2)}(A_0^{(1)}, \bar{\boldsymbol{X}}_0^{(2)*}(A_0^{(1)}))|\boldsymbol{X}_0, A_0^{(1)})\|_{\psi_2} < \infty$, $a_1 + l_1 < 1$, $n \gg s_{\beta_2}\log p\,\rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma})^2/\rho_{\min}^{s_{\beta_2}}(\boldsymbol{\Sigma})$, *and either* $\pi^{(1)}$ *or* $h^{(1)}$ *is correctly specified. Then, with probability at least* $1 - O(n^{-1})$, (4.30) *can be bounded from above by* $\boldsymbol{C}^{(1)}(E_5 + E_6 + E_7 + E_8 + E_9 + E_{10})$

for some constant $\boldsymbol{C}^{(1)} > 0$*, where*

$$
\begin{aligned}
E_5 &= \sqrt{\log q/n}\log^2 n, \quad E_6 = s_{\boldsymbol{\alpha}_1} n^{-2\gamma_{\alpha_1}} \log^2 n + s_{\boldsymbol{\theta}_1} n^{-2\gamma_{\theta_1}} \log^2 n,\\
E_7 &= \sigma_1\{\sqrt{s_{\boldsymbol{\alpha}_1}\log n/n} + \sqrt{s_{\boldsymbol{\alpha}_1}}\lambda_{1n}^{(1)}\rho_1^{(1)}(d_{n\boldsymbol{\alpha}_1})\},\\
E_8 &= \sigma_2\{\sqrt{s_{\boldsymbol{\theta}_1}\log n/n} + \sqrt{s_{\boldsymbol{\theta}_1}}\lambda_{2n}^{(1)}\rho_2^{(1)}(d_{n\boldsymbol{\theta}_1})\},\\
E_9 &= \sigma_0\{\sqrt{s_{\boldsymbol{\alpha}_1}\log n/n} + \sqrt{s_{\boldsymbol{\alpha}_1}}\lambda_{1n}^{(1)}\rho_1^{(1)}(d_{n\boldsymbol{\alpha}_1}) + \tau_0 + \kappa_0^*\},
\end{aligned}
$$

$$E_{10} = n^{-\mu_1}\log n,\ \sigma_0^2 = \mathbb{E}\{\tau(\boldsymbol{X}_0) - \boldsymbol{X}_i^T\boldsymbol{\beta}_1^*\}^2,\ \sigma_1^2 = \mathbb{E}\{h_1(\boldsymbol{X}_0) - \boldsymbol{X}_0^T\boldsymbol{\theta}_1^*\}^2,\ \sigma_2^2 = \mathbb{E}\{\pi_{1,0}^* - \pi_1(\boldsymbol{X}_0)\}^2.$$

The terms $E_5 - E_8$ have similar interpretations as $E_1 - E_4$ in Lemma 4.6.1, respectively. The additional term E_{10} is due to the error bound of $\widehat{\boldsymbol{\beta}}_2$ in the backward induction, while E_9 is due to the misspecification of the contrast function.

Define $\boldsymbol{\Omega}^{(1)}(\boldsymbol{\alpha}_1) = \mathbb{E}[\boldsymbol{X}_0\boldsymbol{X}_0^T A_0^{(1)}\{1+\exp(\boldsymbol{X}_0^T\boldsymbol{\alpha}_1)\}^{-1}]$ and $\boldsymbol{\Omega}_n^{(1)} = n^{-1}\sum_i \boldsymbol{X}_i\boldsymbol{X}_i^T A_i^{(1)}(1-\widehat{\pi}_{1,i})$. Similar as in stage 2, we need the following condition to ensure the RE condition for the matrix $\boldsymbol{\Omega}_n^{(1)}$.

(B9.) Assume that for any $0 < \theta < 1$, we have with probability at least $1 - O(n^{-1})$ that

$$K(s_{\boldsymbol{\beta}_1}, 1, \boldsymbol{\Omega}_n^{(1)}) > (1-\theta) \inf_{\boldsymbol{\alpha}_1 \in H_{\boldsymbol{\alpha}_1}} K(s_{\boldsymbol{\beta}_1}, 1, \boldsymbol{\Omega}^{(1)}(\boldsymbol{\alpha}_1)) > 0, \tag{4.31}$$

where $H_{\boldsymbol{\alpha}_1}$ denotes the set of vectors $\boldsymbol{\alpha}_1$ that satisfies the weak oracle property (4.26).

Theorem 4.7.1 *Assume that Condition B9 and those conditions in Lemma 4.7.1 hold, and* $\lambda_{3n}^{(1)} \geq \boldsymbol{C}^{(1)}\sum_{k=5}^{10} E_k$*. The constant* $\boldsymbol{C}^{(1)}$ *is defined in Lemma 4.7.1. Then, for any fixed* $0 < \theta < 1$*, with probability at least* $1 - O(n^{-1})$*, the error bounds for* $\widehat{\boldsymbol{\beta}}_1$ *are given by*

$$
\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*\|_2 &\leq \frac{12\lambda_{3n}^{(1)}\sqrt{s_{\boldsymbol{\beta}_1}}}{(1-\theta)^2 \inf_{\boldsymbol{\alpha}_1 \in H_{\boldsymbol{\alpha}_1}} K^2(s_{\boldsymbol{\beta}_1}, 1, \boldsymbol{\Omega}^{(1)}(\boldsymbol{\alpha}_1))},\\
\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*\|_1 &\leq \frac{8\lambda_{3n}^{(1)} s_{\boldsymbol{\beta}_1}}{(1-\theta)^2 \inf_{\boldsymbol{\alpha}_1 \in H_{\boldsymbol{\alpha}_1}} K^2(s_{\boldsymbol{\beta}_1}, 1, \boldsymbol{\Omega}^{(1)}(\boldsymbol{\alpha}_1))}.
\end{aligned}
$$

4.7.3 Error bound for the value function of the estimated dynamic treatment regime

The difference of the value functions under the estimated optimal dynamic treatment regime (1.2) and the true optimal regime (d_1^{opt}, d_2^{opt}) is given by

$$\begin{aligned} & \mathbb{E}\{Y_0^*(d_1^{opt,0}, d_2^{opt,0})\} - \mathbb{E}\{Y_0^*(\widehat{d}_1, \widehat{d}_2)\} \\ = \; & \mathbb{E}\big[\tau(\boldsymbol{X}_0)\{\mathbb{I}(\tau(\boldsymbol{X}_0) > 0) - \mathbb{I}(\boldsymbol{X}_0^T \widehat{\boldsymbol{\beta}}_1 > 0)\}\big] + \mathbb{E}\big[\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0}\{\mathbb{I}(\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0} > 0) - \mathbb{I}(\boldsymbol{S}_0^T \widehat{\boldsymbol{\beta}}_2 > 0)\}\big]. \end{aligned}$$

Similar to Condition B4, we impose the following condition.

(B10.) Assume that the probability density function $g^{(1)}(\cdot)$ of $\boldsymbol{X}_0^T \boldsymbol{\beta}_1^*$ exists and is bounded.

Theorem 4.7.2 *Assume that conditions in Theorem 4.7.1 and Condition B10 hold. Assume* $\mathbb{E}(\boldsymbol{S}_0^T \boldsymbol{\beta}_{2,0})^2 = O(1), \mathbb{E}(\boldsymbol{X}_0^T \boldsymbol{\beta}_1^*)^2 = O(1)$. *Then, for some fixed* $0 < \theta < 1$, *we have with probability at least* $1 - O(n^{-1})$ *that,*

$$\begin{aligned} 0 \le \mathbb{E}\{Y_0^\star(d_1^{opt}, d_2^{opt})\} - \mathbb{E}\{Y_0^\star(\widehat{d}_1, \widehat{d}_2)\} \le O(1)\frac{(\omega+\zeta)}{n} + O(1)\sigma_0^{4/3} \\ + O(1)\frac{\omega^2 \rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma}) {\lambda_{3n}^{(2)}}^2 s_{\beta_2} \log^2 n}{(1-\theta)^4 \inf\limits_{\boldsymbol{\alpha}_2 \in H_{\alpha_2}} K^4(s_{\beta_2}, 1, \boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2))} + O(1)\frac{\zeta^2 \rho_{\max}^{s_{\beta_1}}(\boldsymbol{\Psi}) {\lambda_{3n}^{(1)}}^2 s_{\beta_1} \log^2 n}{(1-\theta)^4 \inf\limits_{\boldsymbol{\alpha}_1 \in H_{\alpha_1}} K^4(s_{\beta_1}, 1, \boldsymbol{\Omega}^{(1)}(\boldsymbol{\alpha}_1))}, \end{aligned}$$

where $O(1)$ *denotes some positive constant.*

Theorem 4.7.2 suggests that the upper bound for the difference of the value functions come from three major components: the misspecification of the contrast function, described by σ_0^2, and estimation errors of $\widehat{\boldsymbol{\beta}}_2$ and $\widehat{\boldsymbol{\beta}}_1$.

4.8 Weak oracle properties of $\widehat{\boldsymbol{\alpha}}_j$'s and $\widehat{\boldsymbol{\theta}}_j$'s

In order to prove the error bounds of $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$ and the value functions of the estimated treatment regimes presented in Sections 4.6 and 4.7, we need to establish the weak oracle properties of $\widehat{\boldsymbol{\alpha}}_j$ and $\widehat{\boldsymbol{\theta}}_j$ $(j = 1, 2)$ in the posited models for the propensity score and baseline mean functions. Here, we prove the results based on a posited logistic regression model for the propensity score and a linear model for the baseline mean function under a random design setting. However, these results can be extended to generalized linear models (McCullagh and Nelder 1989).

4.8.1 Weak oracle properties of $\widehat{\alpha}_2$ and $\widehat{\theta}_2$

We assume that $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$ converge to some population parameters $\boldsymbol{\alpha}_{\mathbf{2}}^*$ and $\boldsymbol{\theta}_{\mathbf{2}}^*$, respectively. Under Conditions B11 - B16 given in Section 4.8.3, we establish the weak oracle properties of $\widehat{\boldsymbol{\alpha}}_2$ and $\widehat{\boldsymbol{\theta}}_2$ in the following two Theorems. Recall that $s_{\alpha_2} = |\mathscr{M}_{\alpha_2}| = O(n^{l_4})$ for some $0 \leq l_4 < 1/2$.

Theorem 4.8.1 *Assume that Conditions B11-B13 hold, $l_4 + a_2 < 1$ and $\lambda_{\max}(\boldsymbol{\Sigma}_{\mathscr{M}_{\alpha_2}}^{\mathscr{M}_{\alpha_2}}) = O(1)$. Then, there exists some constant $\gamma_{\alpha_2} > 0$, such that with probability at least $1 - O(n^{-1})$,*

a. $\widehat{\boldsymbol{\alpha}}_2^{\mathscr{M}_{\alpha_2^c}} = 0$.

b. $\|\widehat{\boldsymbol{\alpha}}_2^{\mathscr{M}_{\alpha_2}} - \boldsymbol{\alpha}_{\mathbf{2}}^{*\mathscr{M}_{\alpha_2}}\|_\infty \leq O(1) n^{-\gamma_{\alpha_2}} \log n$,

where $O(1)$ denotes some positive constant.

Theorem 4.8.2 *Assume that Conditions B14-B16 hold, $\lambda_{\max}(\boldsymbol{\Sigma}_{\mathscr{M}_{\theta_2}}^{\mathscr{M}_{\theta_2}}) = O(1)$, and $\|e_0\|_{\psi_2} < \infty$. Then, there exists some constant $\gamma_{\theta_2} > 0$, such that with probability at least $1 - O(n^{-1})$,*

a. $\widehat{\boldsymbol{\theta}}_2^{\mathscr{M}_{\theta_2}^c} = 0$.

b. $\|\widehat{\boldsymbol{\theta}}_2^{\mathscr{M}_{\theta_2}} - \boldsymbol{\theta}_{\mathbf{2}}^{*\mathscr{M}_{\theta_2}}\|_\infty \leq O(1) n^{-\gamma_{\theta_2}} \log n$,

where $O(1)$ denotes some positive constant.

Theorem 1 in Shi, Song and Lu (2015) established weak oracle results of the penalized estimators for a fixed design setting. This is mainly for technical convenience. Its proofs can be obtained using similar arguments as in Fan and Lv (2011). In this paper, we focus on a random design setting, which is more practical in medical studies. To the best of our knowledge, the weak oracle properties of penalized estimators have not been studied in a random design setting with the NP dimensionality. The major difficulty lies in developing some random matrix theories, such as controlling the maximum eigenvalue of some random matrices. Such results are established in Theorem 8.1 and 8.2.

The condition $l_4 + a_2 < 1$ ensures that

$$\max_{j=1}^{p} \lambda_{\max}\left(\sum_{i=1}^{n} \boldsymbol{S}_i^{\mathscr{M}_{\alpha_2}} \boldsymbol{S}_i^{\mathscr{M}_{\alpha_2} T} |S_i^{(j)}| \right) = O(n), \tag{4.32}$$

with probability approaching 1. A major technical difficulty in deriving (4.32) is that the matrix $\sum_{i=1}^n \boldsymbol{S}_i^{\mathscr{M}_{\alpha_2}} \boldsymbol{S}_i^{\mathscr{M}_{\alpha_2} T} |S_i^{(j)}|$ does not have the subexponential tail (see Definition G.2 in

Shi et al. (2018a)). When $s_{\alpha_2} \le n$, we can bound $\max_{j\in\mathcal{M}_{\alpha_2}} |S_i^{(j)}|$ from above by $\sqrt{2}\omega \log n$ with probability at least $1-2/n$, which ensures the subexponential tail of the truncated matrix. Lemma B.2 in Shi et al. (2018a) proves such a result for a more general case.

4.8.2 Weak oracle properties of $\widehat{\alpha}_1$ and $\widehat{\theta}_1$

The weak oracle properties of $\widehat{\boldsymbol{\alpha}}_1$ can be similarly derived as for $\widehat{\boldsymbol{\alpha}}_2$. However, unlike the results for $\widehat{\boldsymbol{\theta}}_2$, the weak oracle properties of $\widehat{\boldsymbol{\theta}}_1$ depend on $\widehat{\boldsymbol{\beta}}_2$ even when the baseline mean function $h^{(1)}$ is correctly specified. This is because the estimated response $\widehat{V}_i$ is obtained based on $\widehat{\boldsymbol{\beta}}_2$. A necessary condition to ensure $\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^\star\|_\infty \xrightarrow{P} 0$ is that $\|\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_{2,0}\|_2 \xrightarrow{P} 0$, which is established in Corollary 4.6.1.

Theorem 4.8.3 *Assume that Condition B8 and Conditions B17-B22 in Section 4.8.3 hold. Further assume that $\lambda_{\max}(\boldsymbol{\Psi}_{\mathcal{M}_{\alpha_1}}^{\mathcal{M}_{\alpha_1}}) = O(1)$, $\lambda_{\max}(\boldsymbol{\Psi}_{\mathcal{M}_{\theta_1}}^{\mathcal{M}_{\theta_1}}) = O(1)$, $n \gg s_{\beta_2} \log p \{\rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma})\}^2/\rho_{\min}^{s_{\beta_2}}(\boldsymbol{\Sigma})$, $a_1 + l_1 < 1$, $\|e_0\|_{\psi_2} < \infty$ and $\|V_0^{(2)}(A_0^{(1)}, \bar{\boldsymbol{X}}_0^{(2)}(A_0^{(1)})) - \mathbb{E}(V_0^{(2)}(A_0^{(1)}, \bar{\boldsymbol{X}}_0^{(2)}(A_0^{(1)}))|\boldsymbol{X}_0, A_0^{(1)})\|_{\psi_2} < \infty$. Then, there exist some constants $\gamma_{\alpha_1} > 0$ and $\gamma_{\theta_1} > 0$, with probability at least $1 - O(n^{-1})$, such that the estimators $\widehat{\boldsymbol{\alpha}}_1$ and $\widehat{\boldsymbol{\theta}}_1$ must satisfy*

a. $\widehat{\boldsymbol{\alpha}}_1^{\mathcal{M}_{\alpha_1}^c} = 0$, $\widehat{\boldsymbol{\theta}}_1^{\mathcal{M}_{\theta_1}^c} = 0$,

b. $\|\widehat{\boldsymbol{\alpha}}_1^{\mathcal{M}_{\alpha_1}} - \boldsymbol{\alpha}_1^{*\mathcal{M}_{\alpha_1}}\|_\infty \le O(1) n^{-\gamma_{\alpha_1}} \log n$, $\|\widehat{\boldsymbol{\theta}}_1^{\mathcal{M}_{\theta_1}} - \boldsymbol{\theta}_1^{*\mathcal{M}_{\theta_1}}\|_\infty \le O(1) n^{-\gamma_{\theta_1}} \log n$,

where $O(1)$ denotes some positive constant.

4.8.3 Technical conditions

(B11.) Assume

$$\|(\boldsymbol{C}_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}})^{-1} b^{(2)\mathcal{M}_{\alpha_2}}\|_\infty = O(n^{-\gamma_{\alpha_2}} \log n), \tag{4.33}$$

$$\|\boldsymbol{b}^{(2)\mathcal{M}_{\alpha_2}^c} - C_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}^c} (C_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}})^{-1} b^{(2)\mathcal{M}_{\alpha_2}}\|_\infty \ll \lambda_{1n}^{(2)}, \tag{4.34}$$

where $b^{(2)} = \mathbb{E}[\boldsymbol{S}_0\{A_0^{(2)} - \pi_{2,0}^*\}]$.

The left hand-side of (4.33) is the limit of

$$\left(\sum_{i=1}^n \boldsymbol{S}_i^{\mathcal{M}_{\alpha_2}} \pi_{2,i}^* (1-\pi_{2,i}^*) \boldsymbol{S}_i^{\mathcal{M}_{\alpha_2} T}\right)^{-1} \sum_{i=1}^n \boldsymbol{S}_i^{\mathcal{M}_{\alpha_2}} (A_i^{(2)} - \pi_{2,i}^*),$$

when s_{α_2} is fixed. When choosing $\boldsymbol{\alpha}_2^*$ to be the argmin of

$$\mathbb{E}[\log\{1+\exp(\boldsymbol{S}_i^T\boldsymbol{\alpha}_2)\}-A_i\boldsymbol{S}_i^T\boldsymbol{\alpha}_2],$$

subjected to the constraint $|\text{supp}(\boldsymbol{\alpha}_2)| \leq s_{\alpha_2}$, (4.33) automatically holds. When the model is misspecified, this condition mainly determines how fast $\widehat{\boldsymbol{\alpha}}_2$ converges to $\boldsymbol{\alpha}_2^*$. Condition (4.34) assumes a weak correlation between $\boldsymbol{S}_0^{\mathcal{M}_{\alpha_2}^c}$ and $A_0^{(2)}-\pi_{2,0}^*$, conditional on $\boldsymbol{S}_0^{\mathcal{M}_{\alpha_2}}$. When the propensity model is correctly specified, these two conditions are simultaneously satisfied.

(B12.) Assume that $\boldsymbol{C}^{(2)}$ satisfies

$$\|(\boldsymbol{C}_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}})^{-1}\|_\infty = O(b_{\alpha_2}), \tag{4.35}$$

$$\|\boldsymbol{C}_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}^c}(C_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}})^{-1}\|_\infty \leq \min\left\{\bar{C}_1\frac{\rho_1^{'(2)}(0+)}{\rho_1^{'(2)}(d_{n\alpha_2})}, O(n^{t_5})\right\}, \tag{4.36}$$

for some constant $0<\bar{C}_1<1$, $t_5\in[0,1/2]$ and $d_{n\alpha_2}=\frac{1}{2}\min_{j\in\mathcal{M}_{\alpha_2}}|\alpha_2^{*(j)}|$.

Condition B12 is the regularity condition posited on the covariance matrix $\boldsymbol{C}^{(2)}(\alpha_2)$. Empirical version of such a condition for the design matrix in a fix design study can be found in Fan and Lv (2011). Condition (4.36) is referred as the strong irrepresentable condition when Lasso penalty is employed and it requires the left hand-side of (4.36) to be uniformly smaller than 1.

(B13.) Assume that $\rho_1^{(2)}$, $\lambda_{1n}^{(2)}$, s_{α_2} and $d_{n\alpha_2}$ satisfy

$$\lambda_{1n}^{(2)} \gg n^{\nu_3}\log^2 n,\ \lambda_{1n}^{(2)}\rho_{1n}^{\prime(2)}(d_{n\alpha_2}) = o(n^{-\gamma_{\alpha_2}}/b_{\alpha_2}),\ d_{n\alpha_2}\gg n^{-\gamma_{\alpha_2}}\log n,$$
$$b_{\alpha_2} = o[\min(n^{1/2-\gamma_{\alpha_2}}\sqrt{\log n},\ n^{\gamma_{\alpha_2}}/s_{\alpha_2}\log n)],$$
$$\lambda_{\min}(C_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}}) \gg \max(\sqrt{s_{\alpha_2}}n^{-\gamma_{\alpha_2}}\log n, \lambda_{1n}^{(2)}\kappa_{\alpha_2})$$

where $\nu_3=\max(a_2/2-1/2, t_5-1/2, t_5-2\gamma_{\alpha_2}+l_4)$, $\kappa_{\alpha_2}=\max_{\boldsymbol{\delta}\in H_{\alpha_2}}\kappa(\rho_1^{(2)},\boldsymbol{\delta})$ with $H_{\alpha_2}=\{\boldsymbol{\delta}\in\mathbb{R}^p:$ $\boldsymbol{\delta}^{\mathcal{M}_{\alpha_2}^c}=0, \|\boldsymbol{\delta}-\boldsymbol{\alpha}_2^*\|_\infty = O(n^{-\alpha_{\theta_2}}\log n)\}$ and the function $\kappa(\rho,\boldsymbol{v})$ is defined as

$$\kappa(\rho,\boldsymbol{v}) = \lim_{\epsilon\to 0^+}\max_{1\leq j\leq k}\sup_{t_1<t_2\in(|v_j|-\epsilon,|v_j|+\epsilon)} -\frac{\rho^{'}(t_2)-\rho^{'}(t_1)}{t_2-t_1}$$

with $\boldsymbol{v}=(v_1,\ldots,v_k)^T$ and $\|\boldsymbol{v}\|_0=k$.

Condition B13 adds constraints on the penalty function, nonsparsity size and strength of the signal.

(B14.) Define $\boldsymbol{f}^{(2)} = \mathbb{E}[\boldsymbol{S}_0\{h_2(\boldsymbol{S}_0) - \boldsymbol{S}_0^T\boldsymbol{\theta}^*\}(1-A_0^{(2)})]$, assume

$$\|(\boldsymbol{D}_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}})^{-1}\boldsymbol{f}_{\mathcal{M}_{\theta_2}}^{(2)}\|_\infty = O(n^{-\gamma_{\theta_2}}\log n),$$
$$\|\boldsymbol{f}^{(2)\mathcal{M}_{\theta_2}^c} - \boldsymbol{D}_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}^c}(\boldsymbol{D}_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}})^{-1}\boldsymbol{f}^{(2)\mathcal{M}_{\theta_2}}\|_\infty \ll \lambda_{2n}^{(2)},$$
$$\|h_2(\boldsymbol{S}_0) - \boldsymbol{\theta}_2^{*T}\boldsymbol{S}_0^{\mathcal{M}_{\theta_2}}\|_{\psi_1} < \infty,$$

(B15.) Assume that $\boldsymbol{D}^{(2)}$ satisfies

$$\|(\boldsymbol{D}_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}})^{-1}\|_\infty = O(b_{\theta_2}),$$
$$\|\boldsymbol{D}_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}^c}(\boldsymbol{D}_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}})^{-1}\|_\infty \leq \min\left\{\bar{C}_2\frac{\rho_2^{'(2)}(0+)}{\rho_2^{'(2)}(d_{n\theta_2})}, O(n^{t_6})\right\},$$

where b_{θ_2} is allowed to diverge, $d_{n\theta_2} = \frac{1}{2}\min_{j\in\mathcal{M}_{\theta_2}}|\boldsymbol{\theta}_2^{*(j)}|$, $\bar{C}_2 \in (0,1)$ and $t_6 \in [0,1/2]$.

(B16.) Assume that s_{θ_2}, λ_{2n}, $\rho_2^{'}$ and $d_{n\theta_2}$ satisfy

$$\lambda_{2n} \gg n^{\nu_4}\log^2 n,\ \lambda_{2n}\rho_2^{'}(d_{n\theta_2}) = o(n^{-\gamma_{\theta_2}}\log n/b_{\theta_2}), d_{n\theta_2} \gg n^{-\gamma_{\theta_2}}\log n,$$
$$b_{\theta_2} \leq n^{\frac{1}{2}-\gamma_{\theta_2}}\sqrt{\log n}, \lambda_{\min}(D_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}}) \gg \max(\lambda_{2n}\kappa_{\theta_2}, \sqrt{s_{\theta_2}\log n}/\sqrt{n}),$$

where $s_{\theta_2} = O(n^{l_5})$, $\nu_4 = \max(a_2/2 - 1/2, t_6 - 1/2)$, $\kappa_{\theta_2} = \max_{\delta\in H_{\theta_2}}\kappa(\rho_2, \boldsymbol{\delta})$ with $H_{\theta_2} = \{\boldsymbol{\delta}\in\mathbb{R}^p : \boldsymbol{\delta}^{\mathcal{M}_{\theta_2}^c} = 0, \|\boldsymbol{\delta} - \boldsymbol{\theta}_2^*\|_\infty = O(n^{-\gamma_{\theta_2}}\log n)\}$.

Condition on the minimum eigenvalue of $\boldsymbol{D}_{\mathcal{M}_{\theta_2}}^{(2)\mathcal{M}_{\theta_2}}$ is less restrictive compared to that for $\boldsymbol{C}_{\mathcal{M}_{\alpha_2}}^{(2)\mathcal{M}_{\alpha_2}}$ in Condition B13. This is because we choose a linear model for $\boldsymbol{h}_2$. Let

$$\boldsymbol{f}^{(1)} = \mathbb{E}[\boldsymbol{X}_0\{h_1(\boldsymbol{X}_0) - \boldsymbol{X}_0^T\boldsymbol{\theta}_1^*\}(1-A_i^{(1)})].$$

(B17.) There are constants $a_0 > 0$ and $r_2 < 0$ such that the following hold:

$$\|(\boldsymbol{D}_{\mathcal{M}_{\theta_1}}^{(1)\mathcal{M}_{\theta_1}})^{-1}\boldsymbol{f}^{(1)\mathcal{M}_{\theta_1}}\|_\infty = O(n^{-a_0}\log n), \tag{4.37}$$
$$\|\boldsymbol{f}^{(1)\mathcal{M}_{\theta_1}^c} - \boldsymbol{D}_{\mathcal{M}_{\theta_1}}^{(1)\mathcal{M}_{\theta_1}^c}(\boldsymbol{D}_{\mathcal{M}_{\theta_1}}^{(1)\mathcal{M}_{\theta_1}})^{-1}\boldsymbol{f}^{(1)\mathcal{M}_{\theta_1}}\|_\infty \ll \lambda_{2n}^{(1)}, \tag{4.38}$$
$$\|\boldsymbol{h}_1(\boldsymbol{S}_0) - \boldsymbol{\theta}_1^{*T}\boldsymbol{S}_0^{\mathcal{M}_{\theta_1}}\|_{\psi_1} < \infty. \tag{4.39}$$

Similar to Condition B11, Condition B17 is used to describe the degree of model misspecification.

(B18.) Assume that $\boldsymbol{D}^{(1)}$ satisfies

$$\|\boldsymbol{D}_{\mathcal{M}_{\theta_1}}^{(1)\mathcal{M}_{\theta_1}^c}(\boldsymbol{D}_{\mathcal{M}_{\theta_1}}^{(1)\mathcal{M}_{\theta_1}})^{-1}\|_\infty \le \bar{C}_3 \frac{\rho_2^{'(1)}(0+)}{\rho_2^{'(1)}(d_{n\theta_1})}, \tag{4.40}$$

$$\liminf \lambda_{\min}(\boldsymbol{D}_{\mathcal{M}_{\theta_1}}^{(1)\mathcal{M}_{\theta_1}}) > 0, \tag{4.41}$$

for some constant $\bar{C}_3 \in (0,1)$ and $d_{n\theta_1} = 1/2 \min\{|\boldsymbol{\theta}_1^{*(j)} : \boldsymbol{\theta}_1^{*(j)} \neq 0|\}$.

A major difference between Condition B18 and Condition B15 is that we impose the eigenvalue condition (4.41) on $\boldsymbol{D}^{(1)}$, which aims to control the estimation loss of $\widehat{\boldsymbol{\theta}}_1$ due to the estimated response $\widehat{V_i}$.

Assume

$$\sqrt{s_{\theta_1}}\lambda_{2n}^{(1)}\rho_2^{\prime(1)}(d_{n\theta_1}) = O(n^{-\mu_2}\log n),$$

for some $\mu_2 < 0$. Define

$$\gamma_{\theta_1} = \min(a_0, \mu_1, \mu_2), \tag{4.42}$$

where μ_1 is defined in (7.6). The L_∞ convergence rate of $\widehat{\boldsymbol{\theta}}_1$ is given as $O(n^{-\gamma_{\theta_1}}\log n)$. Based on (4.42), we can see that the convergence rate is affected by the degree of model misspecification (a_0), the bias produced by the penalty function (μ_2) and the estimation loss of $\widehat{\beta}_2$ (μ_1).

(B19.) Assume that $\lambda_{2n}^{(1)}$, $\rho_2^{(1)\prime}$ and $d_{n\theta_1}$ satisfy

$$\lambda_{2n}^{(1)} \gg n^{\nu_2}\log^2 n, \quad \lambda_{2n}^{(1)}\kappa_{\theta_1} = o(1), \quad d_{n\theta_1} \gg n^{-\gamma_{\theta_1}}\log n,$$

where $\nu_2 = \max(a_1 - 1/2, (a_1 + l_2 - 1 - a_0)/2, l_2 - 1/2, -\mu_1)$ with $s_{\theta_1} = O(n^{l_2})$, $\kappa_{\theta_1} = \max_{\delta \in H_{\theta_1}} \kappa(\rho_2, \delta)$ with $H_{\theta_1} = \{\delta \in \mathbb{R}^q : \delta^{\mathcal{M}_{\theta_1}^c} = 0, \|\delta - \theta_1^\star\|_\infty = O(n^{-\gamma_{\theta_1}}\log n)\}$.

(B20.) Assume

$$\|(\boldsymbol{C}_{\mathcal{M}_{\alpha_1}}^{(1)\mathcal{M}_{\alpha_1}})^{-1}\boldsymbol{b}^{(1)\mathcal{M}_{\alpha_1}}\|_\infty = O(n^{-\gamma_{\alpha_1}}\log n),$$

$$\|\boldsymbol{b}^{(1)\mathcal{M}_{\alpha_1}^c} - \boldsymbol{C}_{\mathcal{M}_{\alpha_1}}^{(1)\mathcal{M}_{\alpha_1}^c}(\boldsymbol{C}_{\mathcal{M}_{\alpha_1}}^{(1)\mathcal{M}_{\alpha_1}})^{-1}\boldsymbol{b}^{(1)\mathcal{M}_{\alpha_1}}\|_\infty \ll \lambda_{1n}^{(1)}.$$

(B21.) Assume that $\boldsymbol{C}^{(1)}$ satisfies

$$\|(\boldsymbol{C}^{(1)\mathcal{M}_{\alpha_1}}_{\mathcal{M}_{\alpha_1}})^{-1}\|_\infty = O(b^{(1)}_{\alpha_1}),$$
$$\|\boldsymbol{C}^{(1)\mathcal{M}^c_{\alpha_1}}_{\mathcal{M}_{\alpha_1}}(\boldsymbol{C}^{(1)\mathcal{M}_{\alpha_1}}_{\mathcal{M}_{\alpha_1}})^{-1}\|_\infty \le \min\left\{\bar{C}_4\frac{\rho_1^{'(1)}(0+)}{\rho_1^{'(1)}(d_{n\alpha_1})}, O(n^{t_1})\right\},$$

for some constant $\bar{C}_4 \in (0,1)$, $t_1 \in [0,1/2]$, $b^{(1)}_{\alpha_1}$ is allowed to diverge and $d_{n\alpha_1} = \frac{1}{2}\min_j |\boldsymbol{\alpha}_1^{*(j)}|$.
(B22.) Assume that $\rho_1^{(1)}$, $\lambda_{1n}^{(1)}$, s_{α_1} and $d_{n\alpha_1}$ satisfy

$$\lambda_{1n}^{(1)} \gg n^{\nu_1}\log^2 n, \lambda_{1n}^{(1)}\rho_{1n}^{'}(d_{n\alpha_1}) = o(n^{-\gamma_{\alpha_1}}/b^{(1)}_{\alpha_1}), d_{n\alpha_1} \gg n^{-\gamma_{\alpha_1}}\log n,$$
$$b^{(1)}_{\alpha_1} = o[\min(n^{1/2-\gamma_{\alpha_1}}\sqrt{\log n}, n^{\gamma_{\alpha_1}}/s_{\alpha_1}\log n)],$$
$$\lambda_{\min}(\boldsymbol{C}^{(1)\mathcal{M}_{\alpha_1}}_{\mathcal{M}_{\alpha_1}}) \gg \max(\sqrt{s_{\alpha_1}}n^{-\gamma_{\alpha_1}}\log n, \lambda_{1n}^{(1)}\kappa_{\alpha_1})$$

where $\nu_1 = \max(a_1/2 - 1/2, t_1 - 1/2, t_1 - 2\gamma_{\alpha_1} + l_1)$, $\kappa_{\alpha_1} = \max_{\boldsymbol{\delta}\in H_{\alpha_1}} \kappa(\rho_1, \boldsymbol{\delta})$, with $H_{\alpha_1} = \{\boldsymbol{\delta} \in \mathbb{R}^q : \boldsymbol{\delta}^{\mathcal{M}^c_{\alpha_1}} = 0, \|\boldsymbol{\delta} - \boldsymbol{\alpha}_1^*\|_\infty = O(n^{-\gamma_{\alpha_1}}\log n)\}$.

Condition B20-B22 are similar to B21-B23.

4.9 Uniform uncertainty principle and restricted eigenvalue conditions in A-learning

In this section, we establish the UUP and RE conditions in the context of A-learning. In our setting, these two conditions are needed on random matrices $\boldsymbol{\Omega}_n^{(2)}$ and $\boldsymbol{\Omega}_n^{(1)}$.

For brevity, we only study the UUP and RE conditions for the random matrix $\boldsymbol{\Omega}_n^{(2)}$. Those for $\boldsymbol{\Omega}_n^{(1)}$ can be similarly derived. Recall that $\mathcal{M}_{\alpha_2}$ refers to the support of $\boldsymbol{\alpha}_2^*$, $\mathcal{M}_{\beta_2} = \mathrm{supp}(\boldsymbol{\beta}_{2,0})$, and $s_{\beta_2} = |\mathcal{M}_{\beta_2}|$. We assume that the weak oracle properties of $\widehat{\boldsymbol{\alpha}}_2$ are achieved such that with probability at least $1 - O(n^{-1})$,

$$\widehat{\boldsymbol{\alpha}}_2^{\mathcal{M}^c_{\alpha_2}} = 0 \text{ and } \|\widehat{\boldsymbol{\alpha}}_2 - \alpha_2^\star\|_\infty = O(n^{-\gamma_{\alpha_2}}\log n), \tag{4.43}$$

for some $\gamma_{\alpha_2} > 0$. The following Lemma establishes the UUP condition for $\boldsymbol{\Omega}_n^{(2)}$.

Lemma 4.9.1 *Assume $\widehat{\boldsymbol{\alpha}}_2$ satisfies*

$$\|\widehat{\boldsymbol{\alpha}}_2 - \boldsymbol{\alpha}_2^*\|_2 = O(\sqrt{s_{\alpha_2}}n^{-\gamma_{\alpha_2}}\log n) = O(1),$$

with probability at least $1-O(n^{-1})$*, and the sample size satisfies*

$$n \gg \frac{\{\rho_{\max}^{s_{\beta_2}}(\Sigma)\}^2 (s_{\beta_2}\log p + s_{\alpha_2}^2)}{\inf\limits_{\alpha_2 \in H_{\alpha_2}} \rho_{\min}^{s_{\beta_2}}(\Omega^{(2)}(\alpha_2))}. \tag{4.44}$$

Then for any $0<\theta<1$*, with probability at least* $1-O(n^{-1})$*, we have*

$$\begin{aligned} &\|\frac{1}{n}\boldsymbol{y}^T\boldsymbol{\Omega}_n^{(2)}\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{\Omega}^{(2)}\boldsymbol{y}\|_2 \\ \leq\ & \left\{\theta + \frac{4\omega^2}{n} + \sqrt{2}\omega^2\|\widehat{\boldsymbol{\alpha}}_2 - \boldsymbol{\alpha}_2^*\|_2\sqrt{\lambda_{\max}(\Sigma_{\mathcal{M}_{\alpha_2}}^{\mathcal{M}_{\alpha_2}})}\right\}\rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma})\|y\|_2^2, \end{aligned} \tag{4.45}$$

for any $\boldsymbol{y}\in\mathbb{R}^p$ *and* $|supp(\boldsymbol{y})|\leq s_{\beta_2}$.

In our setting, if the following regularity conditions hold

$$\liminf_{\boldsymbol{\alpha}_2\in H_{\alpha_2}} \rho_{\min}^{s_{\alpha_2}}(\boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2)) > 0 \text{ and } \rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma}) = O(1),$$

the requirement on the sample size (4.44) reduces to $n \gg s_{\beta_2}\log p$ since $s_{\alpha_2}^2 = O(n^{2l_4}) \ll n$.

The second term on the right-hand side of (4.45) represents the difference between $\boldsymbol{y}^T\widetilde{\boldsymbol{\Omega}}_n^{(2)}\boldsymbol{y}$ and $\boldsymbol{y}^T\boldsymbol{\Omega}_n^{(2)}\boldsymbol{y}$, where $\widetilde{\boldsymbol{\Omega}}_n^{(2)}$ is defined as the expectation of the truncated random matrix

$$\frac{1}{n}\sum_i \boldsymbol{S}_i\boldsymbol{S}_i^T A_i^{(2)}\{1-\pi_2(\boldsymbol{S}_i,\widehat{\boldsymbol{\alpha}}_2)\}\mathbb{I}(\|\boldsymbol{S}_i^{\mathcal{M}_{\alpha_2}}\|_\infty \leq \sqrt{2}\omega\log n). \tag{4.46}$$

This term will vanish as $n\to\infty$. The third term represents the estimation error of $\widehat{\boldsymbol{\alpha}}_2$. When $\rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma}) < 2$ and $\sqrt{\lambda_{\max}(\boldsymbol{\Sigma}_{\mathcal{M}_{\alpha_2}}^{\mathcal{M}_{\alpha_2}})}\|\widehat{\boldsymbol{\alpha}}_2 - \boldsymbol{\alpha}_{\boldsymbol{2}}^*\|_2 \to 0$, (4.45) proves the UUP condition for $\boldsymbol{\Omega}_n^{(2)}$.

A key assumption in Lemma 4.9.1 is the sparsity of $\boldsymbol{\alpha}_2^*$, which is needed to bound the infinity norm in the indicator function of (4.46). This extra requirement comes from the involvement of the estimated propensity scores in $\boldsymbol{\Omega}_n^{(2)}$, which adds significant difficulties in proving Lemma 4.9.1.

After some algebra, the RE condition for $\boldsymbol{\Omega}_n^{(2)}$ follows similarly from Lemma 4.9.1, which is presented below.

Lemma 4.9.2 *For any integer* c_0*, assume that* $\|\widehat{\boldsymbol{\alpha}}_2 - \boldsymbol{\alpha}_2^*\|_2 = O(1)$ *with probability at least*

$1 - O(n^{-1})$*, and the sample size satisfies*

$$n \gg \frac{\{\rho_{\max}^{s_{\beta_2}}(\boldsymbol{\Sigma})\}^2 (s_{\beta_2} \log p + s_{\alpha_2}^2)}{\inf\limits_{\boldsymbol{\alpha}_2 \in H_{\alpha_2}} K^2(s_\beta, c_0, \boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2))}. \tag{4.47}$$

Then, for any $0 < \theta < 1$ *and sufficiently large* n*, with probability at least* $1 - O(n^{-1})$*,we have*

$$K(s_{\beta_2}, c_0, \boldsymbol{\Omega}_n^{(2)}) > (1-\theta) \inf_{\boldsymbol{\alpha}_2 \in H_{\alpha_2}} K(s_{\beta_2}, c_0, \boldsymbol{\Omega}^{(2)}(\boldsymbol{\alpha}_2)).$$

The sample size requirement (4.47) is stronger than (4.44). To see this, for any positive semidefinite matrix $\boldsymbol{\Psi}$, and positive integers s and c_0, we have

$$K^2(s, c_0, \boldsymbol{\Psi}) \le K^2(s, 0, \boldsymbol{\Psi}) = \rho_{\min}^s(\boldsymbol{\Psi}).$$

4.10 Concordance and value information criteria

Bayesian information criteria (BIC) is used in our numerical studies to tune the penalty functions. BIC has been widely used in model selection for selecting the tuning parameter when the goal is prediction. In high dimensional regressions, Chen and Chen (2008) proposed an extended BIC for model selection, and showed their BIC is consistent when the number of predictors grows polynomially in sample size. Fan and Tang (2013) proposed a similar criterion and showed its consistency when the number of predictors is in the non-polynomial order of the sample size.

However, the proposed BIC-type criterion in Section 5.2.3 is not doubly robust. In this section, we introduce two information criteria: the concordance and value information criteria that are selection consistent when either the propensity score model or the baseline model is correctly specified. It is worth mentioning that these information criteria are generally applicable and are not specifically tailored to certain estimating procedures. Specifically, they can be applied to robust learning (Zhang et al. 2012), concordance-assisted learning (Fan et al. 2017), penalized A-learning introduced in this chapter and sparse concordance assisted learning (Liang et al. 2018). In this section, we will focus on the use of penalized A-learning. Detailed implementation based on other procedures can be found in Shi et al. (2019b).

For simplicity, we focus on a single stage study. Extensions to multi-stage settings can be found in Section 11 of Shi et al. (2019b). The concordance function stands for the average difference of the benefit in receiving a treatment for two patients, if one is more likely to be assigned to this treatment compared to another under a given regime. Specifically, for any linear treatment regime $\mathbb{I}(\boldsymbol{x}^T\boldsymbol{\beta}+c>0)$, the concordance function $C(\boldsymbol{\beta})$ is defined as

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

for two subjects $i\neq j$. The rationale behind their method is that if $Y_i^*(1)-Y_i^*(0)>Y_j^*(1)-Y_j^*(0)$, the optimal treatment regime should be more likely to assign treatment 1 to subject i compared with subject j.

We focus on the setting where $\tau(\boldsymbol{x})=Q(\boldsymbol{x}^T\boldsymbol{\beta}_0)$ for some monotonically increasing function Q, we have by Condition C1 and C2 that

$$C(\boldsymbol{\beta})=\mathbb{E}\{Q(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)-Q(\boldsymbol{X}_j^T\boldsymbol{\beta}_0)\}\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>\boldsymbol{X}_j^T\boldsymbol{\beta}).$$

It follows that

$$C(\boldsymbol{\beta}_0)-C(\boldsymbol{\beta})=\mathbb{E}\{Q(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)-Q(\boldsymbol{X}_j^T\boldsymbol{\beta}_0)\}\{\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}_0>X_j^T\boldsymbol{\beta}_0)-\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>\boldsymbol{X}_j^T\boldsymbol{\beta})\}.$$

When $\boldsymbol{X}_i^T\boldsymbol{\beta}_0>\boldsymbol{X}_j^T\boldsymbol{\beta}_0$, it follows from the monotonicity of Q that $Q(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)>Q(\boldsymbol{X}_j^T\boldsymbol{\beta}_0)$. Therefore, we have for any $\boldsymbol{\beta}\in\mathbb{R}^p$,

$$\{Q(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)-Q(\boldsymbol{X}_j^T\boldsymbol{\beta}_0)\}\{\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}_0>\boldsymbol{X}_j^T\boldsymbol{\beta}_0)-\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>\boldsymbol{X}_j^T\boldsymbol{\beta})\}I(\boldsymbol{X}_i^T\boldsymbol{\beta}_0>X_j^T\boldsymbol{\beta}_0)\geq 0.$$

One can similarly show

$$\{Q(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)-Q(\boldsymbol{X}_j^T\boldsymbol{\beta}_0)\}\{\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}_0>\boldsymbol{X}_j^T\boldsymbol{\beta}_0)-\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>\boldsymbol{X}_j^T\boldsymbol{\beta})\}\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}_0\leq\boldsymbol{X}_j^T\boldsymbol{\beta}_0)\geq 0.$$

It follows that $C(\boldsymbol{\beta}_0)\geq C(\boldsymbol{\beta}),\forall\boldsymbol{\beta}\in\mathbb{R}^p$. Hence, we have

$$\boldsymbol{\beta}_0=\arg\max_{\boldsymbol{\beta}\in\mathbb{R}^p} C(\boldsymbol{\beta}).$$

In an observational study, the propensity score is unknown and needs to be estimated from data. Usually, a parametric model $\pi(\boldsymbol{x},\boldsymbol{\alpha})$ is assumed for the propensity score $\pi(\boldsymbol{x})$. To calculate our doubly-robust information criteria, we also fit a parametric model $h(\boldsymbol{x},\boldsymbol{\eta})$ for the baseline function $h(\boldsymbol{x})$. We assume estimators $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\eta}}$ converge to some $\boldsymbol{\alpha}^*\in\mathbb{R}^{q_1}$ and $\boldsymbol{\eta}^*\in\mathbb{R}^{q_2}$. When the models are correct, $\boldsymbol{\alpha}^*$ and $\boldsymbol{\eta}^*$ correspond to the true parameters in the model, i.e, $\pi(\boldsymbol{x})=\pi(\boldsymbol{x},\boldsymbol{\alpha}^*)$, $h(\boldsymbol{x})=h(\boldsymbol{x},\boldsymbol{\eta}^*)$. Otherwise, these parameters stand for some population-level least false parameters. Let $\boldsymbol{\theta}=(c,\boldsymbol{\beta}^T)^T$. Define

$$\begin{aligned} V^{DR}(\boldsymbol{\theta}) &= \mathbb{E}\left\{\frac{A_0\mathbb{I}(\boldsymbol{X}_0^T\boldsymbol{\beta}>-c)}{\pi(\boldsymbol{X}_0,\boldsymbol{\alpha}^*)}+\frac{(1-A_0)\mathbb{I}(\boldsymbol{X}_0^T\boldsymbol{\beta}\le -c)}{1-\pi(\boldsymbol{X}_0,\boldsymbol{\alpha}^*)}\right\}Y_0 \\ &- \mathbb{E}\left\{\frac{A_0\mathbb{I}(\boldsymbol{X}_0^T\boldsymbol{\beta}>-c)}{\pi(\boldsymbol{X}_0,\boldsymbol{\alpha}^*)}+\frac{(1-A_0)\mathbb{I}(\boldsymbol{X}_0^T\boldsymbol{\beta}\le -c)}{1-\pi(\boldsymbol{X}_0,\boldsymbol{\alpha}^*)}-1\right\}h(\boldsymbol{X}_0,\boldsymbol{\eta}^*), \end{aligned}$$

and

$$\begin{aligned} C^{DR}(\boldsymbol{\beta}) &= \mathbb{E}\left\{\frac{\{A_i-\pi(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}\{Y_i-h(\boldsymbol{X}_i,\boldsymbol{\eta}^*)\}A_j}{\pi(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\{1-\pi(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}\pi(\boldsymbol{X}_j,\boldsymbol{\alpha}^*)}\right. \\ &- \left.\frac{\{A_j-\pi(\boldsymbol{X}_j,\boldsymbol{\alpha}^*)\}\{Y_j-h(\boldsymbol{X}_j,\boldsymbol{\eta}^*)\}A_i}{\pi(\boldsymbol{X}_j,\boldsymbol{\alpha}^*)\{1-\pi(\boldsymbol{X}_j,\boldsymbol{\alpha}^*)\}\pi(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)}\right\}\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>\boldsymbol{X}_j^T\boldsymbol{\beta}). \end{aligned}$$

Under Assumptions C1-C3, when either the propensity score model or the baseline model is correct, we can show

$$\begin{aligned} V^{DR}(\boldsymbol{\theta}) &= \mathbb{E}\left\{h(\boldsymbol{X}_0)+\frac{\pi(\boldsymbol{X}_0)}{\pi(\boldsymbol{X}_0,\boldsymbol{\alpha}^*)}\tau(\boldsymbol{X}_0)\mathbb{I}(\boldsymbol{X}_0^T\boldsymbol{\beta}>-c)\right\}, \\ C^{DR}(\boldsymbol{\beta}) &= \mathbb{E}\left\{\frac{\pi(\boldsymbol{X}_i)\pi(\boldsymbol{X}_j)}{\pi(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\pi(\boldsymbol{X}_j,\boldsymbol{\alpha}^*)}\{\tau(\boldsymbol{X}_i)-\tau(\boldsymbol{X}_j)\}\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>\boldsymbol{X}_j^T\boldsymbol{\beta})\right\}. \end{aligned}$$

Suppose $\tau(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{\beta}_0+c_0$ for some $\boldsymbol{\beta}_0\in\mathbb{R}^p$, $c_0\in\mathbb{R}$. Let $\boldsymbol{\theta}_0=(c_0,\boldsymbol{\beta}_0^T)^T$. When the propensity score model is correct, we have $V^{DR}(\boldsymbol{\theta}_0)=V(d^{opt})$ for any $d^{opt}\in\mathscr{D}^{opt}$ and $C^{DR}(\boldsymbol{\beta}_0)=C(\boldsymbol{\beta}_0)$. This result generally does not hold when the propensity score model is not correct. However, $\theta_0(\beta_0)$ still maximizes $V^{DR}(C^{DR})$ as long as either of the models is correct. This suggests V^{DR} and C^{DR} can be used to construct information criteria. Define

$$\text{VIC}^{DR}(\theta)=n\widehat{V}^{DR}(\theta)-\kappa_n\|\beta\|_0,\ \ \text{CIC}^{DR}(\beta)=n\widehat{V}^{DR}(\beta)-\kappa_n\|\beta\|_0,$$

where $\widehat{V}^{DR}$ and $\widehat{C}^{DR}$ are empirical estimators for V^{DR} and C^{DR},

$$\begin{aligned} \widehat{V}^{DR}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_i\left\{\frac{A_i\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>-c)}{\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})}+\frac{(1-A_i)I(\boldsymbol{X}_i^T\boldsymbol{\beta}\le -c)}{1-\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})}\right\}Y_i \\ &- \left\{\frac{A_i\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>-c)}{\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})}+\frac{(1-A_i)I(\boldsymbol{X}_i^T\boldsymbol{\beta}\le -c)}{1-\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})}-1\right\}h(\boldsymbol{X}_i,\widehat{\eta}), \end{aligned}$$

$$\widehat{C}^{DR}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i\neq j}\left\{\frac{\{A_i-\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})\}\{Y_i-h(\boldsymbol{X}_i,\widehat{\eta})\}A_j}{\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})\{1-\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})\}\pi(X_j,\widehat{\boldsymbol{\alpha}})}\right.$$
$$\left. - \frac{\{A_j-\pi(\boldsymbol{X}_j,\widehat{\boldsymbol{\alpha}})\}\{Y_j-h(\boldsymbol{X}_j,\widehat{\eta})\}A_i}{\pi(\boldsymbol{X}_j,\widehat{\boldsymbol{\alpha}})\{1-\pi(\boldsymbol{X}_j,\widehat{\boldsymbol{\alpha}})\}\pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})}\right\}\mathbb{I}(\boldsymbol{X}_i^T\boldsymbol{\beta}>\boldsymbol{X}_j^T\boldsymbol{\beta}).$$

To apply the proposed information criteria, we need some penalization methods (such as penalized A-learning) to simultaneously select and estimate $\boldsymbol{\theta}_0$, with some tuning parameter λ. For each $\lambda\in[\lambda_{\min},\lambda_{\max}]$ where $\lambda_{\min}$ and $\lambda_{\max}$ are allowed to vary with n, denote $\widehat{\mathscr{M}}(\lambda)$ as the nonzero entries selected by our estimating procedure and $\widehat{\theta}_{\widehat{\mathscr{M}}(\lambda)}=(\widehat{c}_{\widehat{\mathscr{M}}(\lambda)},\widehat{\beta}^T_{\widehat{\mathscr{M}}(\lambda)})^T\in\mathbb{R}^{p+1}$ the corresponding estimator. We define

$$\widehat{\mathscr{M}_V}=\operatorname*{arg\,max}_{\substack{|\widehat{\mathscr{M}}(\lambda)|\le s_n\\ \lambda\in[\lambda_{\min},\lambda_{\max}]}}\ \mathrm{VIC}(\widehat{\theta}_{\widehat{\mathscr{M}}(\lambda)})\text{ and }\widehat{\mathscr{M}_C}=\operatorname*{arg\,max}_{\substack{|\widehat{\mathscr{M}}(\lambda)|\le s_n\\ \lambda\in[\lambda_{\min},\lambda_{\max}]}}\ \mathrm{CIC}(\widehat{\beta}_{\widehat{\mathscr{M}}(\lambda)}),$$

as the selected model for $\boldsymbol{\theta}_0$ (or equivalently, $\boldsymbol{\beta}_0$).

Shi et al. (2019b) derived the consistencies of VIC^{DR} and CIC^{DR} under the fixed-p scenario. Comparatively speaking, CIC^{DR} is more reliable than VIC^{DR} in model selection, although both criteria are consistent. In our numerical experiments given below, CIC^{DR} achieves smaller false negative and false positive when compared with VIC^{DR}. In our theoretical results, conditions to ensure model selection consistency for VIC^{DR} are more restrictive than those for CIC^{DR}. This is because the estimated concordance function in CIC^{DR} is a U-process of order two, and is more "smooth" than the estimated value function in VIC^{DR}, which is an empirical process that involves summation of indicator functions.

When p is comparable or much larger than n, we can fit the baseline or propensity score models via penalized regression as discussed in this chapter. Under certain conditions on the penalized regression estimators, consistencies of VIC^{DR} and CIC^{DR} can be similarly proven. We omit the technical details to save space.

In the rest of the section, we conduct a simulation study to investigate the numerical performance of the proposed information criteria. Specifically, we generate the response from the following model:

$$Y_0=h_0(X_0^{(1)},X_0^{(3)})+A_i(X_0^{(1)}+X_0^{(2)})+\varepsilon_0,$$

where $\boldsymbol{X}_0\sim N(0,\boldsymbol{I}_p)$, $A_0\sim\text{Bernoulli}(\pi_0(X_i))$, $\varepsilon_0\sim N(0,0.5^2)$. The contrast function takes the linear form, $\tau(\boldsymbol{x}) = x^{(1)} + x^{(2)}$ and the optimal treatment regime is

$$d^{opt,0}(\boldsymbol{x}) = \mathbb{I}(x^{(1)} + x^{(2)} > 0).$$

Table 4.6: Simulation settings in Section 6.2

	S1	S2	S3	S4
$h_0(x,y)$	$1+x-y$	$1+xy$	$1+x-y$	$1+xy$
$\pi_0(x)$	0.5	0.5	$\Phi(x_{p-1}-x_p)$	$\Phi(x_{p-1}-x_p)$

We design four settings by considering two choices of the baseline function, and two choices of the propensity score function. Table 4.6 gives the propensity and baseline function in each setting. We fit a penalized linear regression model for the baseline and a penalized logistic regression model for the propensity score, and choose SCAD as the penalty function. Hence, both the propensity score and baseline models are correctly specified in Setting 1. One of them is misspecified in Settings 2 and 3. In Setting 4, both models are misspecified. We use penalized A-learning to estimate $\boldsymbol{\theta}_0$, based on a series of tuning parameters λ. The estimating procedure is similar to (4.5). In implementation, we compute

$$\bar{\boldsymbol{\theta}} = (\bar{c}, \bar{\boldsymbol{\beta}}^T) = \underset{(c,\boldsymbol{\beta}^T)}{\arg\min} \left\| \sum_i \overline{\boldsymbol{X}}_i (A_i - \widehat{\pi}_i)(Y_i - \widehat{h}_i - A_i c - A_i \boldsymbol{X}_i^T \boldsymbol{\beta}) \right\|_\infty,$$
$$\text{subject to } \|\boldsymbol{\beta}\|_1 \le \lambda,$$

where $\overline{\boldsymbol{X}}_i = (1, \boldsymbol{X}_i^T)^T$, $\widehat{\pi}_i$ and $\widehat{h}_i$ denote the estimated propensity score and baseline for the ith individual, based on penalized logistic regression and penalized linear regression, with SCAD penalty function.

Let $\widehat{\mathscr{M}}(\lambda)$ be the support of $\bar{\boldsymbol{\beta}}$. We then compute $\widehat{\boldsymbol{\beta}}_{\widehat{\mathscr{M}}(\lambda)}$ and $\widehat{c}_{\widehat{\mathscr{M}}(\lambda)}$ by solving

$$\sum_i (A_i - \widehat{\pi}_i)(Y_i - \widehat{h}_i - A_i \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{\widehat{\mathscr{M}}(\lambda)} - A_i \widehat{c}_{\widehat{\mathscr{M}}(\lambda)}) = 0,$$
$$\sum_i \boldsymbol{X}_i^{\widehat{\mathscr{M}}(\lambda)} (A_i - \widehat{\pi}_i)(Y_i - \widehat{h}_i - A_i \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{\widehat{\mathscr{M}}(\lambda)} - A_i \widehat{c}_{\widehat{\mathscr{M}}(\lambda)}) = 0,$$

with $\widehat{\boldsymbol{\beta}}_{\widehat{\mathscr{M}}(\lambda)}^{\widehat{\mathscr{M}^c}(\lambda)} = 0$.

Table 4.7: Simulation results for Setting 1 and 2 (%, standard deviations in parenthesis)

		S1		S2	
	n	200	300	200	300
CIC^{DR}	TP	100.00(0.00)	100.00(0.00)	77.00(4.23)	90.00(3.02)
	FN	0.00(0.00)	0.00(0.00)	7.00(1.88)	0.50(0.50)
	FP	0.00(0.00)	0.00(0.00)	0.03(0.01)	0.01(0.00)
	ER	1.18(0.08)	1.17(0.09)	8.34(1.09)	4.08(0.43)
	VR	99.96(0.00)	99.97(0.00)	97.02(0.66)	99.39(0.15)
BIC	TP	95.00(2.19)	94.00(2.39)	69.00(4.65)	89.00(3.14)
	FN	0.00(0.00)	0.00(0.00)	4.50(1.60)	0.00(0.00)
	FP	0.01(0.00)	0.01(0.00)	0.03(0.01)	0.02(0.00)
	ER	1.44(0.12)	1.34(0.13)	7.98(0.85)	4.00(0.41)
	VR	99.94(0.01)	99.94(0.02)	97.77(0.47)	99.44(0.11)
VIC_3^{DR}	TP	100.00(0.00)	100.00(0.00)	61.00(4.90)	86.00(3.49)
	FN	0.00(0.00)	0.00(0.00)	13.50(2.34)	2.00(0.98)
	FP	0.00(0.00)	0.00(0.00)	0.02(0.01)	0.01(0.00)
	ER	1.16(0.08)	1.08(0.08)	10.64(1.08)	4.67(0.56)
	VR	99.97(0.00)	99.97(0.00)	96.13(0.61)	99.12(0.21)
VIC_4^{DR}	TP	100.00(0.00)	100.00(0.00)	56.00(4.99)	80.00(4.02)
	FN	0.00(0.00)	0.00(0.00)	11.00(2.20)	2.00(0.98)
	FP	0.00(0.00)	0.00(0.00)	0.05(0.01)	0.02(0.00)
	ER	1.08(0.08)	1.11(0.09)	11.05(1.13)	5.09(0.54)
	VR	99.97(0.00)	99.97(0.00)	95.98(0.66)	99.02(0.21)
VIC_5^{DR}	TP	99.00(1.00)	100.00(0.00)	54.00(5.01)	72.00(4.51)
	FN	0.00(0.00)	0.00(0.00)	9.00(2.06)	2.00(0.98)
	FP	0.00(0.00)	0.00(0.00)	0.07(0.01)	0.03(0.01)
	ER	1.18(0.09)	1.1(0.08)	11.20(1.14)	5.80(0.59)
	VR	99.96(0.01)	99.97(0.00)	95.92(0.72)	98.81(0.22)

We compute $\bar{\boldsymbol{\theta}}$ for a series of log-spaced values $\exp(-3) = \lambda_0, \lambda_1, \ldots, \lambda_{100} = \exp(2)$. Tuning parameters are selected by CIC^{DR} and VIC^{DR}. In CIC^{DR}, we set

$$\kappa_n = \log(p)\log_{10}(n)\log(\log_{10}(n)).$$

In VIC^{DR}, we set

$$\kappa_n = n^{1/3} \log^{2/3}(p) \log(\log(n))/\kappa,$$

where κ is a constant from a set $\{3, 4, 5\}$. For each κ, we denote the corresponding information criterion as VIC$_\kappa^{DR}$.

Table 4.8: Simulation results for Setting 3 and 4 (%, standard deviations in parenthesis)

		S3		S4	
	n	200	300	200	300
CIC^{DR}	TP	91.00(2.88)	99.00(1.00)	48.00(5.02)	66.00(4.76)
	FN	3.00(1.19)	0.00(0.00)	27.00(3.21)	14.00(2.47)
	FP	0.00(0.00)	0.00(0.00)	0.03(0.01)	0.03(0.01)
	ER	3.15(0.61)	1.42(0.11)	16.50(1.50)	10.90(1.27)
	VR	99.23(0.28)	99.95(0.01)	92.09(1.03)	95.63(0.75)
BIC	TP	55.00(5)	61.00(4.9)	32.00(4.69)	40.00(4.92)
	FN	0.00(0.00)	0.00(0.00)	18(3.06)	8.00(2.10)
	FP	0.08(0.01)	0.11(0.02)	0.14(0.02)	0.15(0.02)
	ER	4.23(0.33)	3.69(0.35)	17.4(1.35)	13.99(1.2)
	VR	99.48(0.06)	99.57(0.06)	92.27(0.95)	94.52(0.74)
VIC$_3^{DR}$	TP	82.00(3.86)	98.00(1.41)	39.00(4.90)	59.00(4.94)
	FN	7.00(1.74)	0.00(0.00)	29.50(3.26)	17.50(2.60)
	FP	0.01(0.00)	0.00(0.00)	0.05(0.01)	0.03(0.01)
	ER	5.16(0.85)	1.38(0.12)	18.02(1.46)	12.39(1.29)
	VR	98.4(0.38)	99.94(0.01)	91.4(1.01)	94.79(0.77)
VIC$_4^{DR}$	TP	89.00(3.14)	98.00(1.41)	43.00(4.98)	59.00(4.94)
	FN	2.50(1.10)	0.00(0.00)	26.00(3.29)	14.50(2.49)
	FP	0.01(0.00)	0.00(0.00)	0.06(0.01)	0.05(0.01)
	ER	2.97(0.54)	1.42(0.13)	17.41(1.53)	12.2(1.28)
	VR	99.35(0.24)	99.94(0.01)	91.61(1.06)	94.88(0.78)
VIC$_5^{DR}$	TP	90.00(3.02)	97.00(1.71)	43.00(4.98)	54.00(5.01)
	FN	2.00(0.98)	0.00(0.00)	23.00(3.21)	13.50(2.45)
	FP	0.01(0.00)	0.00(0.00)	0.07(0.01)	0.07(0.01)
	ER	2.85(0.52)	1.43(0.13)	16.95(1.48)	12.52(1.28)
	VR	99.4(0.25)	99.94(0.01)	91.98(1.02)	94.73(0.78)

We further compare our information criteria with the BIC-type criterion (see Section

5.2.3). For any $\boldsymbol{\theta} = (c, \boldsymbol{\beta}^T)^T$, define

$$\text{BIC}(\theta) = n\log(\text{RSS}(\theta)/n) + \|\beta\|_0\{\log(n) + \log(p+1)\},$$

where

$$\text{RSS}(\theta) = \sum_{i=1}^{n}(A_i - \widehat{\pi}_i)^2(Y_i - \widehat{h}_i - A_i c - A_i \boldsymbol{X}_i^T \boldsymbol{\beta})^2.$$

We report the false positives (FP) rate (the percentage of unimportant variables that are selected),

$$\text{FP} = \frac{1}{L}\sum_{l=1}^{L}\frac{|\mathcal{M}_\beta^c \cap \widehat{\mathcal{M}^{(l)}}|}{|\mathcal{M}_\beta^c|},$$

the false negatives (FN) rate (the percentage of important variables that are missed),

$$\text{FN} = \frac{1}{L}\sum_{l=1}^{L}\frac{|\mathcal{M}_\beta \cap (\widehat{\mathcal{M}^{(l)}})^c|}{|\mathcal{M}_\beta|},$$

the percentage of selecting the true models (TP),

$$\text{TP} = \frac{1}{L}\sum_{l=1}^{L} I(\mathcal{M}_\beta = \widehat{\mathcal{M}^{(l)}}),$$

the average error rate (ER) and average ratio of value (VR) of the estimated optimal treatment regime,

$$\text{ER} = \frac{1}{L}\sum_{l=1}^{L}\mathbb{E}|\hat{d}^{(l)}(X_0) - d^{opt}(X_0)|, \quad \text{VR} = \frac{1}{L}\sum_{l=1}^{L}\frac{\mathbb{E}Y_0^*(\hat{d}^{(l)})}{\mathbb{E}Y_0^*(d^{opt})}, \tag{4.48}$$

where $\hat{d}^{(l)}(\boldsymbol{x}) = I(\widehat{c}_{\widehat{\mathcal{M}^{(l)}}} + \boldsymbol{x}^T\widehat{\boldsymbol{\beta}}_{\widehat{\mathcal{M}^{(l)}}} > 0)$, $\widehat{\mathcal{M}^{(l)}}$ is the set of important variables selected by a particular information criteria in the l-th simulation and L is the total number of simulations.

Tables 4.7 and 4.8 summaize the results with sample size $n = 200/300$ and 100 simulation replications. CIC^{DR} outperform VIC^{DR} and BIC in all settings, in terms of TP. For example, in Setting 2 with $n = 200$, CIC^{DR} correctly recover 77% of the models, while TP's for other criteria are smaller than 70%. In addition, except for Setting 2, VIC^{DR} outperforms BIC in all other settings. Take Setting 3 with $n = 300$ as an example, TP's for VIC$_3^{DR}$, VIC$_4^{DR}$, VIC$_5^{DR}$

are all very close to 1 while BIC only correctly recovers 61% of the models. False positives of BIC are much higher compared to our information criteria in Setting 3. Moreover, all the information criteria work extremely well in Setting 1 where both the propensity score and baseline models are correctly specified, and perform much worse in Setting 4 where both models are misspecified. Except for BIC and VIC_5^{DR}, all other criteria always select the true model in Setting 1. In Setting 4 with $n = 200$, however, TP's of all criteria are below 50%.

4.11 Post selection inference

The main goal of constructing optimal DTRs is to find treatments that are significantly superior to other treatment options. This requires addressing a post selection inference issue, i.e, the problem of influencing the estimated optimal value function (or the difference between the estimated value and the value function under other treatment options). In the fixed dimension setting, we can use either the empirical average of the advantage function (Murphy, 2003) or the augmented inverse propensity score type estimates (AIPWE, Zhang et. al, 2012) to estimate the optimal value function. Both type of estimators are asymptotically normally distributed. However, the inference based on the advantage function may not be valid in high dimensions. This is because when the number of predictors is large, the parameter estimates in the contrast function may not have oracle property (i.e, model selection consistency and asymptotic normality).

For a single stage study, assuming a linear interaction form $\boldsymbol{X}_0^T\boldsymbol{\beta}_0$ for the contrast function. Under certain conditions, Shi et al. (2019e) showed that AIPWE is asymptotically normal even for NP-dimensionality if (i) $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2 = o_p(n^{-1/4})$, (ii) with probability going to 1, $\widehat{\boldsymbol{\beta}}^{\mathscr{M}_\beta^c} = 0$ where $\mathscr{M}_\beta$ is the support of $\boldsymbol{\beta}_0$. For our penalized A-learning estimator, Assumption (i) can be achieved assuming certain conditions on the dimension of covariates, sample size and the sparsity of parameters in the contrast, the baseline and propensity score function. Assumption (ii) is typically satisfied by thresholding the Lasso, Dantzig or folded-concave type estimators. Alternatively, we can use the one step online estimator as in Luedtke and van der Laan (2016). However, the asymptotic variance will be larger since it does not use all the data to construct the estimator.

CHAPTER

5

TESTING QUALITATIVE TREATMENT EFFECTS

5.1 Introduction

In many medical studies, patients may differ significantly in the way they respond to the treatment. There have been increasing interests in estimating the OTR. All those existing estimation methods described in Chapter 1 implicitly assume that patients' covariates have qualitative interactions with treatment, which means that there exists a subset of patients whose "best" treatments assigned according to the OTR are different from others.

We consider testing the existence of OTR due to the following reasons. First, the OTR may not always exist in practice, see the data from the Nefazodone-CBASP clinical trial study in Section 5.4 for an example. In this case, one treatment is better than the other for all patients and there is no need of estimating the OITR. Second, we note that implementing the OTR requires future patients' covariates which can be expensive to collect in some cases (Baker et al. 2009; Gail 2009; Huang et al. 2015). In these cases, we recommend to adopt the "one-size-fits-all" paradigm when the null hypothesis of no OTR is not rejected. Third, our test is constructed based on estimated value functions' difference comparing the OTR and a fixed regime (i.e. assign all to the best treatment). The test is not significant implies that the value functions' difference is not significant. Under such a situation, although we can still estimate the OTR, the gain of the obtained OTR over the fixed regime in terms of the improvement of value is not significant. Thus, the obtained OTR under such a situation may not be of practical interest. Therefore, it is essential to test the overall qualitative treatment effects of the prognostic covariates to determine whether we need to implement the OTR for future patients. Gunter et al. (2011) developed an S-score to quantify the magnitude of the marginal qualitative treatment effects of a single covariate. However, the S-score doesn't characterize the overall qualitative treatment effects of all covariates. Besides, no theoretical guarantees were provided for the S-score.

For binary treatments, testing qualitative treatment effects is equivalent to testing whether the interaction between treatment and covariates (i.e, the contrast function) is almost surely positive or negative. To test such hypothesis, Chang et al. (2015) proposed a test based on a L_1-type functional of kernel smoothing estimators of conditional treatment effects. Hsu (2017) proposed a Kolmogorov-Smirnov type test statistic based on nonparametric estimation of conditional treatment effects with a hypercube kernel. It is well known that kernel smoothing estimators are undesirable in practice due to the curse of dimensionality. As a result, these test statistics are not reliable when the dimension of the covariates is relatively large. However, in modern biomedical applications, it is likely to obtain a large number of prognostic factors for each individual patient. To the best of our knowledge, there are lack of methods for testing the overall qualitative treatment effects in high-dimensional settings.

In this chapter, we aim to test the overall qualitative treatment effects in a high dimensional setting. This is a very challenging task due to the curse of dimensionality. To better illustrate this point, consider a simple situation where patients' covariates, x, consist of p independent Rademacher variables. Then, it is equivalent to test whether the contrast as a function of the covariates is always positive or negative for any $x \in \{-1,1\}^p$. Therefore, we need to test 2^p moment inequalities even in this very simplified situation. However, for each $x \in \{-1,1\}^p$, we have on average $N/2^P$ observations with covariates equal to x, where N is the total number of observations. When $N = O(2^p)$, this seems impossible without additional assumptions. We show in Lemma 6.2.1 that covariates have the overall qualitative treatment effects if and only if the value function under the OTR is strictly larger than those under fixed treatment regimes. This motivates us to construct test statistics based on the difference between the optimal value function and the value function under fixed treatment regimes. However, inference for such value difference is extremely difficult in the nonregular cases, that is, there is a positive probability that the contrast function is equal to zero. We use a sample-splitting method to construct the test statistic, based on a nonparametric estimator of the contrast function. As long as the estimated contrast function satisfies certain convergence rates, we show our test statistic is consistent.

When the dimension of covariates is large, we construct the test based on sparse random projections of covariates into a low-dimensional space. Random projections have been a powerful method for dimension reduction in the computer science literature. The key idea behind is given in the Johnson-Lindenstrauss Lemma (Johnson and Lindenstrauss 1984), which states that a set of high dimensional vectors can be projected into a suitable lower-dimensional space while approximately preserve their pairwise distances. In the statistics literature, Lopes et al. (2011) proposed a high-dimensional two-sample test which integrates a random projection with the Hotelling T^2 statistic. Recently, Cannings and Samworth (2017) proposed a random projection-based method for the high-dimensional classification.

In this chapter, we propose the use of random projections with sparse matrix. In contrast to the dense sketching matrix used in Lopes et al. (2011) and Cannings and Samworth (2017), only a small proportion of elements in the sparse sketching matrix are nonzero. References on sparse random projections include Omidiran and Wainwright (2010); Li et al. (2006); Nelson and Nguyên (2013). In our simulation studies, we show that our sparse random projection-based test statistics are more powerful compared to those based on dense random projection matrix, when the OTR is “sparse". Besides, we advocate using data-dependent algorithms to generate sparse sketching matrix, since most random projections will be weakly correlated with the contrast function. In theory, we prove the consistency of our sparse random projection-based test. Moreover, in the regular cases, we show that the power function of our test statistic is asymptotically the same as the “oracle" test statistic which is constructed based on the “optimal" projection matrix.

The rest of the paper is organized as follows. In Section 5.2, we present the definition of the overall qualitative treatment effects, introduce our test statistic and study its asymptotic properties under the null and local alternative. Simulation studies and real data applications are conducted in Section 5.3 and Section 5.4 respectively, to examine the empirical performance of the proposed testing procedure. Section 5.5 concludes with a summary and discussions of possible extensions. All the technical proofs can be found in Shi et al. (2019a).

5.2 Proposed tests

We use the setting described in Section 1.2. Covariates $\boldsymbol{X}_0$ are said to have the overall qualitative treatment effects (OQTE) if

$$\mathbb{P}\{\tau(\boldsymbol{X}_0) > 0\} > 0 \text{ and } \mathbb{P}\{\tau(\boldsymbol{X}_0) < 0\} > 0.$$

In this chapter, we consider testing the following hypothesis:

$$H_0: \boldsymbol{X}_0 \text{ dosen't have OQTE} \text{ versus } H_1: \boldsymbol{X}_0 \text{ has OQTE.} \tag{5.1}$$

Assume C1-C3 hold. Under H_0, the optimal treatment regime assigns the same treatment to all patients. Therefore, testing OQTE is equivalent to testing the existence of OTR.

5.2.1 A simple value-based test statistic in fixed p case

Assume the observed data are summarized as $\{O_i = (\boldsymbol{X}_i, A_i, Y_i), i = 1, \dots, N\}$, where O_i's are i.i.d. copies of $O_0 = (\boldsymbol{X}_0, A_0, Y_0)$. The distribution of O_0 is allowed to vary with N. To illustrate the idea, we first assume p is small and fixed, and present here a value-based test statistic for the null hypothesis (5.1). Later in this section, we will consider the more challenging high dimensional setting. Let $V(0) = \mathbb{E}\{Y^*(0)\}$ and $V(1) = \mathbb{E}\{Y^*(1)\}$. The following lemma relates OQTE to the difference between the optimal value function and the value functions under fixed treatment regimes.

Lemma 5.2.1 *Assume $\mathbb{E}|\tau(\boldsymbol{X}_0)| < \infty$, and conditions C1-C3 hold. Then the followings are equivalent: (i) $\boldsymbol{X}_0$ doesn't have OQTE; (ii) $V(d^{opt,0}) = \max\{V(0), V(1)\}$.*

By definition, we have $V(d^{opt,0}) \geq \max\{V(0), V(1)\}$. Under H_1, Lemma 6.2.1 implies $V(d^{opt,0}) > \max(V(0), V(1))$. Therefore, it suffices to test

$$H_0: V(d^{opt,0}) = \max\{V(0), V(1)\} \text{ versus } H_1: V(d^{opt,0}) > \max\{V(0), V(1)\}.$$

For simplicity, we assume $V(1) \geq V(0)$. This implies that the new treatment is at least as good as the standard one on average. The hypothesis $V(1) \geq V(0)$ can be tested using historical data or data from a pilot study. When $V(0) \geq V(1)$, the test statistic can be similarly constructed.

Lemma 6.2.1 motivates us to consider test statistics based on some estimators for the value difference $\text{VD}(d^{opt,0}) = V(d^{opt,0}) - V(1)$. For any treatment regime d, Zhang et al.

(2012) proposed an inverse propensity score weighted estimator (IPSWE) for $V(d)$:

$$\widehat{V}(d) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{A_i d(\boldsymbol{X}_i)}{\pi_i} Y_i + \frac{(1-A_i)\{1-d(\boldsymbol{X}_i)\}}{1-\pi_i} Y_i\right], \tag{5.2}$$

where π_i is a shorthand for $\pi(\boldsymbol{X}_i)$. Plugging $d \equiv 1$, we obtain $\widehat{V}(1) = N^{-1}\sum_i A_i Y_i/\pi_i$. For any fixed d, $\sqrt{N}\widehat{\text{VD}}(d) = \sqrt{N}\{\widehat{V}(d) - \widehat{V}(1)\}$ corresponds to a sum of i.i.d random variables. Therefore, its asymptotic variance can be consistently estimated by the sample variance estimator,

$$\widehat{\sigma}^2(d) = \frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\frac{1-A_i}{1-\pi_i} - \frac{A_i}{\pi_i}\right) Y_i\{1-d(\boldsymbol{X}_i)\} - \widehat{\text{VD}}(d)\right]^2. \tag{5.3}$$

Suppose $\widehat{\tau}(\cdot)$ is an estimate of $\tau(\cdot)$. Based on (5.2) and (5.3), it is natural to use

$$\widehat{T} = \sqrt{N}\widehat{\text{VD}}(\widehat{d})/\widehat{\sigma}(\widehat{d})$$

as the test statistic, where $\widehat{d}(\boldsymbol{x}) = \mathbb{I}\{\widehat{\tau}(\boldsymbol{x}) > 0\}$, and reject H_0 when $\widehat{T} > z_\alpha$ at a given significance level α.

Consistency of such a naive test requires $\mathbb{E}|\widehat{d}(\boldsymbol{X}_0) - d^{opt,0}(\boldsymbol{X}_0)|^2 \to 0$. However, as commented by Luedtke and van der Laan (2016), this assumption is typically violated in the non-regular cases where $\mathbb{P}\{\tau(\boldsymbol{X}_0) = 0\} > 0$, even when $\widehat{\tau}$ is consistent to τ. To solve this problem, we consider a modified version of $\widehat{T}$ based on sample splitting and cross-validation. Let $\mathscr{I}_1$ and $\mathscr{I}_2$ be a random partition of $\{1, \ldots, N\}$ into 2 disjoint subsets of equal sizes $n = N/2$. For any $\mathscr{I} \subseteq \{1, \ldots, N\}$ and treatment regime d, define

$$\begin{aligned}
\widehat{\text{VD}}_{\mathscr{I}}(d) &= \frac{1}{|\mathscr{I}|}\sum_{i\in\mathscr{I}}\left[\left(\frac{1-A_i}{1-\pi_i} - \frac{A_i}{\pi_i}\right) Y_i\{1-d(\boldsymbol{X}_i)\}\right],\\
\widehat{\sigma}^2_{\mathscr{I}}(d) &= \frac{1}{|\mathscr{I}|-1}\sum_{i\in\mathscr{I}}\left[\left(\frac{1-A_i}{1-\pi_i} - \frac{A_i}{\pi_i}\right) Y_i\{1-d(\boldsymbol{X}_i)\} - \widehat{\text{VD}}_{\mathscr{I}}(d)\right]^2,
\end{aligned}$$

where $|\mathscr{I}|$ stands for the number of elements in $\mathscr{I}$. Let $\widehat{\tau}_{\mathscr{I}}$ be the corresponding estimator of τ based on observations in $\mathscr{I}$ and $\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) = I\{\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}) > 0\}$. We define our test statistic by

$$\widehat{T}_{CV} = \max\left(\frac{\sqrt{n}\widehat{\text{VD}}_{\mathscr{I}_1}(\widehat{d}_{\mathscr{I}_2})}{\max\{\widehat{\sigma}_{\mathscr{I}_1}(\widehat{d}_{\mathscr{I}_2}), \delta_n\}}, \frac{\sqrt{n}\widehat{\text{VD}}_{\mathscr{I}_2}(\widehat{d}_{\mathscr{I}_1})}{\max\{\widehat{\sigma}_{\mathscr{I}_2}(\widehat{d}_{\mathscr{I}_1}), \delta_n\}}\right), \tag{5.4}$$

for some positive sequence $\delta_n \to 0$, and reject H_0 when $\widehat{T}_{CV} > z_{\alpha/2}$. The sequence δ_n

guarantees that the denominators in $\widehat{T}_{CV}$ are strictly greater than 0.

Alternative to the sample splitting method, one can consider a Wald-type test statistic based on the online one-step estimator proposed by Luedtke and van der Laan (2016). However, calculating such test statistic is more computationally expensive than ours. Besides, the asymptotic normality of such test statistic requires the class of functions $\Big\{[(1-A)/\{1-\pi(\boldsymbol{x})\}-A/\pi(\boldsymbol{x})]Y\{1-d(\boldsymbol{x})\}:d\Big\}$ to be Glivenko-Cantelli, where d varies over the range of estimators $\widehat{d}$ (see Section 7.3 in Luedtke and van der Laan 2016). In contrast, our testing procedure is valid under H_0 for any $\widehat{d}$.

Theorem 5.2.1 *Assume Conditions C1-C3 hold,* $\mathbb{E}|Y_0|^3 = O(1)$ *and* $\delta_n \gg n^{-1/6}$. *Then under* H_0, *for any* $0<\alpha<1$, *we have*

$$\limsup_n \mathbb{P}(\widehat{T}_{CV} > z_{\alpha/2}) \le \alpha.$$

Moreover, assume that

$$\mathbb{VAR}\left\{\left(\frac{A_0}{\pi(\boldsymbol{X}_0)}-\frac{1-A_0}{1-\pi(\boldsymbol{X}_0)}\right)Y_0\{1-\widehat{d}_{\mathcal{I}_j}(\boldsymbol{X}_0)\}\mid\{O_i\}_{i\in\mathcal{I}_j}\right\} = o_p(\delta_n), \tag{5.5}$$

for $j=1,2$, *where* $\mathbb{VAR}(V_1 \mid V_2)$ *denotes the variance of* V_1 *conditional on* V_2. *Then, we have* $\mathbb{P}(\widehat{T}_{CV} > z_{\alpha/2}) \to 0$.

The following theorem states the consistency of our proposed test statistic. It relies on Conditions C4 and C5. We provide these conditions in Section 5.2.4.

Theorem 5.2.2 *Assume Conditions C1-C5 hold,* $\mathbb{E}|Y_0|^3 = O(1)$ *and* $\delta_n \to 0$. *Under* $H_1: V(d^{opt,0}) = V(1)+h_n$, *if* $h_n \gg n^{-1/2}$, *then we have* $\mathbb{P}(\widehat{T}_{CV} > z_{\alpha/2}) \to 1$. *Moreover, assume* $\mathbb{P}\{\tau(\boldsymbol{X}_0)=0\}=0$ *and* $\liminf_n \sigma_0^2 > 0$ *where*

$$\sigma_0^2 = \mathbb{VAR}\left\{\left(\frac{A}{\pi(\boldsymbol{X}_0)}-\frac{1-A}{1-\pi(\boldsymbol{X}_0)}\right)Y\{1-d^{opt,0}(\boldsymbol{X}_0)\}\right\}.$$

If $\sqrt{n}h_n = O(1)$, *then we have*

$$\mathbb{P}\left(\widehat{T}_{CV} > z_{\alpha/2}\right) = 2\bar{\Phi}\left(z_{\frac{\alpha}{2}}-\frac{\sqrt{n}h_n}{\sigma_0}\right)-\bar{\Phi}^2\left(z_{\frac{\alpha}{2}}-\frac{\sqrt{n}h_n}{\sigma_0}\right)+o(1),$$

where $\bar{\Phi}(z) = 1-\Phi(z)$.

Theorem 5.2.1 and 5.2.2 show the consistency of our testing procedure. Note that Conditions C4 and C5 are not required to ensure Theorem 5.2.1. This suggests the type-I error is well controlled regardless of any estimating procedure. On the other hand, conditions on δ_n in Theorem 5.2.1 are stronger than those in Theorem 5.2.2. In the regular cases when $\mathbb{P}\{\tau(\boldsymbol{X}_0)=0\}=0$, Theorem 5.2.2 provides the asymptotic power function of our test. Notice that h_n is equal to $\mathbb{E}[-\tau(\boldsymbol{X}_0)\{\tau(\boldsymbol{X}_0)<0\}]$ which relies on the dependence structure of the covariates. As a result, the power of our test depends crucially on the underlying data-generating process.

In this paper, $\widehat{d}$ is obtained by a plug-in estimator based on some nonparametric estimation of the contrast function. Alternatively, one can directly estimate d^{opt} using OWL. Theorem 5.2.2 holds as long as the estimated decision function $\widehat{d}$ satisfies $V(\widehat{d}_{\mathscr{I}})=V(d^{opt,0})+o_p(|\mathscr{I}|^{-1/2})$.

Since we assume $V(1)\ge V(0)$, under H_0, we have $\mathbb{P}\{\tau(\boldsymbol{X}_0)\ge 0\}=1$. In the regular cases where $\mathbb{P}\{\tau(\boldsymbol{X}_0)=0\}=0$, we have $\mathbb{P}\{d^{opt,0}(\boldsymbol{x})=1\}=1$ and hence

$$\mathbb{VAR}\left\{\left(\frac{A_0}{\pi(\boldsymbol{X}_0)}-\frac{1-A_0}{1-\pi(\boldsymbol{X}_0)}\right)Y_0\{1-d^{opt,0}(\boldsymbol{X}_0)\}\right\}=0.$$

Besides, in the regular cases, $d^{opt,0}$ can be consistently estimated by $\widehat{d}_{\mathscr{I}_j}$ (see Equation (S.18) in Shi et al. (2019a)). Assume conditions C4 and C5 hold with $\gamma\ge 1$. Then we can show (5.5) holds. Hence, the type-I error of our test will go to 0.

5.2.2 A sparse random projection-based test statistic

When p is large, it is far more challenging to estimate the contrast function $\tau(\boldsymbol{x})$ due to the curse of dimensionality. To handle high-dimensional covariates, we project the covariate space into a low dimensional vector space to construct our test statistic. Throughout this paper, we assume the dimension of the projected space, q is fixed. For a given matrix $\boldsymbol{S}\in\mathbb{R}^{q\times p}$ and any $\boldsymbol{\omega}\in\mathbb{R}^q$, define

$$\tau^{\boldsymbol{S}}(\omega)=\mathbb{E}\{\tau(\boldsymbol{X}_0)|\boldsymbol{S}\boldsymbol{X}_0=\omega\}.$$

Under (A1)-(A3), the treatment regime $d_S^{opt,0}(\boldsymbol{x})=\mathbb{I}\{\tau^{\boldsymbol{S}}(\boldsymbol{S}\boldsymbol{x})>0\}$ is optimal in the sense that it maximizes the value function among the class of treatment regimes based only on the projected covariates $\boldsymbol{S}\boldsymbol{X}_0$.

Since q is small, $\tau^{\boldsymbol{S}}$ can be consistently estimated. We can construct a value-based test statistic as discussed in Section 5.2.1 based on the projected data $\{O_i^{\boldsymbol{S}}\}_{i\in\{1,\dots,N\}}$ where $O_i^{\boldsymbol{S}} = (\boldsymbol{S}\boldsymbol{X}_i, A_i, Y_i)$. The power of such test statistic depends crucially on the sketching matrix $\boldsymbol{S}$. To better understand this, consider the following example:

$$\tau(\boldsymbol{X}_0) = \left\{\left(\frac{X_0^{(1)}+X_0^{(2)}}{\sqrt{2}}\right)^2 - \delta\right\}\left(\frac{X_0^{(3)}+X_0^{(4)}+X_0^{(5)}+X_0^{(6)}+X_0^{(7)}}{\sqrt{5}}\right)^2, \tag{5.6}$$

for some $\delta > 0$.

Apparently, we have $\tau(\boldsymbol{X}_0) > 0$ if $|X_0^{(1)} + X_0^{(2)}| > \sqrt{2\delta}$ and $\tau(\boldsymbol{X}_0) < 0$ if $|X_0^{(1)} + X_0^{(2)}| < \sqrt{2\delta}$. Assume $\boldsymbol{X}_0 \sim N(0, \boldsymbol{I}_p)$. Then $\boldsymbol{X}_0$ have the OQTE. Let $q = 1$, the "optimal" sketching matrix $\boldsymbol{S}^*$ is equal to

$$\boldsymbol{S}^* = c_0(1, 1, 0, 0, \dots, 0),$$

for any $c_0 \neq 0$. For any $\boldsymbol{S} \in \mathbb{R}^{1\times p}$ such that $\boldsymbol{S}^*\boldsymbol{S}^T = 0$, $\boldsymbol{S}\boldsymbol{X}_0$ is independent of $X^{(1)}+X^{(2)}$. Then, we have

$$\begin{aligned}
& \tau^{\boldsymbol{S}}(\omega) = \mathbb{E}\{\tau(\boldsymbol{x})|\boldsymbol{S}\boldsymbol{X}_0 = \omega\} \\
= \; & \mathbb{E}\left[\left|\left\{\left(\frac{X_0^{(1)}+X_0^{(2)}}{\sqrt{2}}\right)^2 - \delta\right\}\left(\frac{X_0^{(3)}+X_0^{(4)}+X_0^{(5)}+X_0^{(6)}+X_0^{(7)}}{\sqrt{5}}\right)^2\right| \boldsymbol{S}\boldsymbol{X}_0 = \omega\right] \\
= \; & (1-\delta)\mathbb{E}\left\{\left(\frac{X_0^{(3)}+X_0^{(4)}+X_0^{(5)}+X_0^{(6)}+X_0^{(7)}}{\sqrt{5}}\right)^2 \middle| \boldsymbol{S}\boldsymbol{X}_0 = \omega\right\}.
\end{aligned}$$

Hence, $\tau^{\boldsymbol{S}}(\omega)$ is always nonnegative or nonpositive as a function of ω. As a result, the test statistic based on $\{O_i^{\boldsymbol{S}}\}_i$ doesn't have any power to detect the OQTE. The challenge here lies in finding a projection matrix S that is highly correlated with $\boldsymbol{S}^*$.

Below, we propose a data-dependent algorithm to generate $\boldsymbol{S}$ and introduce our test statistic. Our theory shows that our test statistic works as if the optimal sketching matrix $\boldsymbol{S}^*$ were known. Statistical properties of our testing procedure are formally studied in Section 5.2.2.

Test statistic

Assume for now, we have an estimator $\widehat{\tau}_{\mathcal{I}}^{\boldsymbol{S}}$ for $\tau^{\boldsymbol{S}}$ based on any subset of the projected data $\{O_i^{\boldsymbol{S}}\}_{i\in\mathcal{I}}$ and an algorithm to sample sparse sketching matrices whose distribution $G(\boldsymbol{S}, \{O_i\}_{i\in\mathcal{I}})$ is allowed to depend on $\{O_i\}_{i\in\mathcal{I}}$. We describe the whole testing procedure in

Algorithm 1.

1. Input observations $\{O_i\}_{i=1,\ldots,N}$, δ_n, α and a sampling distribution G.
2. Randomly partition the data into two subsets $\{O_i\}_{i\in\mathcal{I}_1}$ and $\{O_i\}_{i\in\mathcal{I}_2}$.
3. For $j=1,2$,
   (i) Independently sample a sparse sketching matrix $\boldsymbol{S}_{\mathcal{I}_j} \sim G(\boldsymbol{S}, \{O_i\}_{i\in\mathcal{I}_j})$;
   (ii) Obtain estimators $\widehat{\tau}_{\mathcal{I}_j}^{\boldsymbol{S}_{\mathcal{I}_j}}$ and $\widehat{d}_{\mathcal{I}_j}^{\boldsymbol{S}_{\mathcal{I}_j}}(\boldsymbol{x}) = I\{\widehat{\tau}_{\mathcal{I}_j}^{\boldsymbol{S}_{\mathcal{I}_j}}(\boldsymbol{S}_{\mathcal{I}_j}\boldsymbol{x}) > 0\}$;
   (iii) Calculate $\widehat{T}^{\boldsymbol{S}_{\mathcal{I}_j}} = \sqrt{n}\widehat{\text{VD}}_{\mathcal{I}_j^c}(\widehat{d}_{\mathcal{I}_j}^{\boldsymbol{S}_{\mathcal{I}_j}})/\max\{\widehat{\sigma}_{\mathcal{I}_j^c}(\widehat{d}_{\mathcal{I}_j}^{\boldsymbol{S}_{\mathcal{I}_j}}), \delta_n\}$.
4. Reject H_0 if $\widehat{T}_{SRP} = \max(\widehat{T}^{\boldsymbol{S}_{\mathcal{I}_1}}, \widehat{T}^{\boldsymbol{S}_{\mathcal{I}_2}}) > z_{\alpha/2}$.

Algorithm 1: Calculate the random projection-based test statistic.

Now we present our algorithm for generating sparse sketching matrix. Let $\mathcal{S}$ denote the space of sparse sketching matrices:

$$\mathcal{S} = \{\boldsymbol{S} \in \mathbb{R}^{q\times p} : \|\boldsymbol{S}_{(i)}\|_0 \le s, \|\boldsymbol{S}_{(i)}\|_2 = 1, \forall i = 1, \ldots, q\},$$

for some fixed integer s that satisfies $2 \le s \le p$, where the notation $\boldsymbol{\Phi}_{(i)}$ stands for the ith row of an arbitrary matrix $\boldsymbol{\Phi}$.

It remains to generate $\boldsymbol{S}_{\mathcal{I}_j}$ based on the sub-dataset $\{O_i\}_{i\in\mathcal{I}_j}$. We first sample many sparse sketching matrices from $\mathcal{S}$. Each row of the sketching matrix is independently and uniformly distributed on the space $\{\boldsymbol{S} \in \mathbb{R}^p : \|\boldsymbol{S}\|_0 = s, \|\boldsymbol{S}\|_2 = 1\}$. This corresponds to Step 2 in our proposed algorithm below. Then we output the sparse sketching matrix that maximizes the estimated value difference function. Specifically, we propose using data-splitting strategy for evaluation of the value difference function. That is, for each sketching matrix, we randomly divide $\{O_i\}_{i\in\mathcal{I}_j}$ into $\mathbb{K}$ folds, use any of the $\mathbb{K}-1$ subsamples to estimate the OTR based on projected covariates, use the remaining subsamples to evaluate the corresponding value difference function, and aggregate these value difference functions over different subsamples. This corresponds to Step 3-5 in our proposed algorithm below. We summarize our procedure in Algorithm 2.

Asymptotic properties under the null and local alternative

We first show the validity of the proposed test, which applies to any estimator $\widehat{\tau}_{\mathcal{I}}^{S}$.

1. Input observations $\{O_i\}_{i\in\mathscr{I}}$, integers B, s, q and $\mathbb{K}\geq 2$.
2. Generate i.i.d matrices $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_B$ according as $\boldsymbol{S}_0$ whose distribution is described as follows. For $j=1,\ldots,q$,
   (i) Independently select a simple random sample $\mathscr{M}_j$ of size s from $\{1,\ldots,p\}$;
   (ii) Independently generate a Gaussian random vector $\boldsymbol{g}_j \sim N(0, \boldsymbol{I}_s)$;
   (iii) Set $\boldsymbol{S}_{0,(j)}^{\mathscr{M}_j^c}=0$ and $\boldsymbol{S}_{0,(j)}^{\mathscr{M}_j}=\boldsymbol{g}_j/\|\boldsymbol{g}_j\|_2$.
3. Randomly divide $\mathscr{I}$ into $\mathbb{K}$ subsets $\{\mathscr{I}^{(k)}\}_{k=1}^{\mathbb{K}}$ of equal sizes. Let $\mathscr{I}^{(k)-}=\mathscr{I}\cap(\mathscr{I}^{(k)})^c$.
4. For $b=1,\ldots,B$,
   (i) For $k=1,\ldots,\mathbb{K}$,
      (i.1) Obtain the estimator $\widehat{\tau}_{\mathscr{I}^{(k)-}}^{\boldsymbol{S}_b}$ and $\widehat{d}_{\mathscr{I}^{(k)-}}^{\boldsymbol{S}_b}(\boldsymbol{x})=I\{\widehat{\tau}_{\mathscr{I}^{(k)-}}^{\boldsymbol{S}_b}(\boldsymbol{S}_b\boldsymbol{x})>0\}$;
      (i.2) Evaluate the value difference $\widehat{\mathrm{VD}}_{\mathscr{I}^{(k)}}(\widehat{d}_{\mathscr{I}^{(k)-}}^{\boldsymbol{S}_b})$.
   (ii) Obtain the cross-validated estimator $\widehat{\mathrm{VD}}_{CV}^{\boldsymbol{S}_b}=\sum_k \widehat{\mathrm{VD}}_{\mathscr{I}^{(k)}}(\widehat{d}_{\mathscr{I}^{(k)-}}^{\boldsymbol{S}_b})/\mathbb{K}$.
5. Output $\boldsymbol{S}_{\widehat{b}}$, where $\widehat{b}=\arg\max_{b=1}^{B}\widehat{\mathrm{VD}}_{CV}^{\boldsymbol{S}_b}$.

Algorithm 2: Generate data-dependent sparse random sketching matrix.

Theorem 5.2.3 *Assume C1-C3 hold,* $\mathbb{E}|Y_0|^3=O(1)$ *and* $\delta_n \gg n^{-1/6}$*. Then under* H_0*, we have*

$$\limsup_n \mathbb{P}\left(\widehat{T}_{SRP}>z_{\alpha/2}\right)\leq \alpha.$$

Moreover, assume that

$$\mathbb{VAR}\left\{\left(\frac{A_0}{\pi(\boldsymbol{X}_0)}-\frac{1-A_0}{1-\pi(\boldsymbol{X}_0)}\right)Y_0\{1-\widehat{d}_{\mathscr{I}_j}^{\boldsymbol{S}_{\mathscr{I}_j}}(\boldsymbol{X}_0)\}\,|\,\{O_i\}_{i\in\mathscr{I}_j},\boldsymbol{S}_{\mathscr{I}_j},\mathscr{I}_j\right\}=o_p(\delta_n),$$

for $j=1,2$*. Then we have* $\mathbb{P}(\widehat{T}_{SRP}>z_{\alpha/2})\to 0$.

Let $\boldsymbol{S}^*=\arg\max_{\boldsymbol{S}\in\mathscr{S}} V(d_{\boldsymbol{S}}^{opt,0})$ be the optimal sketching matrix. The optimal sketching matrix $\boldsymbol{S}^*$ may not be unique. To see this, for any sketching matrix $\boldsymbol{S}^*\in\mathscr{S}$ that maximizes $V(d_{\boldsymbol{S}}^{opt,0})$, $-\boldsymbol{S}^*$ also maximizes $V(d_{\boldsymbol{S}}^{opt,0})$ and we have $-\boldsymbol{S}^*\in\mathscr{S}$. Moreover, when $q\geq 2$, there may exist infinitely many maximizers.

Our theoretical studies are mostly concerned with the “oracle" test statistic. The oracle knew the set $\mathscr{S}^*$ ahead of time. In Algorithm 1: Step 3(i), instead of using Algorithm 2 to sample $\boldsymbol{S}^{\mathscr{I}_1}$ and $\boldsymbol{S}^{\mathscr{I}_2}$, we use the oracle set $\boldsymbol{S}^{\mathscr{I}_1}=\boldsymbol{S}^{\mathscr{I}_2}=\boldsymbol{S}^*$ for an arbitrary $\boldsymbol{S}^*\in\mathscr{S}^*$. Denoted by $\widehat{T}_{oracle}$ the resulting oracle test statistic. Let $h_n^*=\arg\max_{\boldsymbol{S}^*\in\mathscr{S}^*} V(d_{\boldsymbol{S}^*}^{opt,0})-V(1)$. Similar to Theorem 5.2.2, under H_1, if $h_n^*\gg n^{-1/2}$, then we can show

$$\mathbb{P}\left(\widehat{T}_{oracle}>z_{\alpha/2}\right)\to 1.$$

Assume

$$V(d^{opt,0}_{\boldsymbol{S}^*}) = V(d^{opt,0}), \;\; \forall \boldsymbol{S}^* \in \mathscr{S}^*. \tag{5.7}$$

This condition means the optimal decision rule depends on the set of projected covariates $\boldsymbol{S}^*\boldsymbol{X}_0$ only. It holds when $\tau(\boldsymbol{x}) = \phi(\boldsymbol{S}^*\boldsymbol{x})\mathrm{g}(\boldsymbol{x})$ for some function $\phi(\cdot)$ and some nonnegative function $\mathrm{g}(\cdot)$. In the regular cases where $\Pr(\tau(\boldsymbol{X}_0) = 0) = 0$, (5.7) implies that $\Pr(d^{opt,0}(\boldsymbol{X}_0) = d^{opt,0}_{\boldsymbol{S}^*}(\boldsymbol{X}_0)) = 1, \forall \boldsymbol{S}^* \in \mathscr{S}^*$. Thus, the class of optimal treatment regimes $\{d^{opt,0}_{\boldsymbol{S}^*} : \boldsymbol{S}^* \in \mathscr{S}^*\}$ will almost surely recommend the same treatment to any given patient. Assume (5.7) holds and $\Pr(\tau(\boldsymbol{X}_0) = 0) = 0$. Similar to Theorem 5.2.2, the asymptotic power of $\widehat{T}_{oracle}$ can be derived as

$$\mathbb{P}\left(\widehat{T}_{oracle} > z_{\alpha/2}\right) = 2\bar{\Phi}\left(z_{\frac{\alpha}{2}} - \frac{\sqrt{n}h_n}{\sigma_0}\right) - \bar{\Phi}^2\left(z_{\frac{\alpha}{2}} - \frac{\sqrt{n}h_n}{\sigma_0}\right) + o(1), \tag{5.8}$$

where h_n and σ_0 are defined in Theorem 5.2.2.

In the following, we prove the consistency of our proposed testing procedure when using Algorithm 2 to generate the sparse sketching matrix. Moreover, we show our test statistic possesses the oracle property. This means the power function of $\widehat{T}_{SRP}$ is asymptotically the same as the oracle test statistic $\widehat{T}_{oracle}$.

Define the semimetric

$$d^{\tau}(\boldsymbol{S}_1, \boldsymbol{S}_2) = \sqrt{\mathbb{E}|\tau^{\boldsymbol{S}_1}(\boldsymbol{S}_1\boldsymbol{X}_0) - \tau^{\boldsymbol{S}_2}(\boldsymbol{S}_2\boldsymbol{X}_0)|^2}, \quad \forall \boldsymbol{S}_1, \boldsymbol{S}_2 \in \mathscr{S}.$$

We make the following assumptions.

(C6.) For any sketching matrices $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_B \in \mathscr{S}$ and any $\mathscr{I} \subseteq \{1, 2, \ldots, N\}$ with $|\mathscr{I}| \geq n/2$, assume the following event holds with probability tending to 1,

$$\max_{b=1}^{B} \mathbb{E}^{\boldsymbol{X}_0} |\widehat{\tau}^{\boldsymbol{S}_j}_{\mathscr{I}}(\boldsymbol{S}_j\boldsymbol{X}_0) - \tau^{\boldsymbol{S}_j}(\boldsymbol{S}_j\boldsymbol{X}_0)|^2 = O\left(n^{-r_0}\log n\right),$$

where the expectation $\mathbb{E}^{\boldsymbol{X}_0}$ is taken with respect to $\boldsymbol{X}_0$, and the little-o term is uniform in $\mathscr{I}$ and $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_B$.

(C7.) Assume $B \gg (p\sqrt{n})^{(s-1)q}$. In addition, assume there exist some constant $\bar{C} > 0$ and some sketching matrix $\boldsymbol{S}^* \in \mathscr{S}^*$ such that

$$d^{\tau}(\boldsymbol{S}, \boldsymbol{S}^*) \leq \bar{C}\left(\sum_{j=1}^{q} \|\boldsymbol{S}^{(j)} - \boldsymbol{S}^{*(j)}\|_2^2\right)^{1/2}, \quad \forall \boldsymbol{S} \in \mathscr{S}. \tag{5.9}$$

(C8.) Assume there exist some constants γ, ε_0, $\delta_0 > 0$ such that for any sketching matrix $\boldsymbol{S}$ satisfying $V(d_{\boldsymbol{S}}^{opt}) \geq V(d_{\boldsymbol{S}^*}^{opt}) - \varepsilon_0$, we have $\mathbb{P}\{0 < |\tau^{\boldsymbol{S}}(\boldsymbol{S}\boldsymbol{X})| \leq t\} = O(t^{\gamma})$, where the big-$O$ term is uniform in $0 < t < \delta_0$ and $\boldsymbol{S}$.

Condition C6 assumes the uniform convergence rate of $\widehat{\tau}_{\mathscr{I}}^{\boldsymbol{S}_b}$ for $b = 1, \ldots, B$. Since the uniform convergence rate increases as B increases, Condition C6 gives the upper bound for B. On the contrary, Condition C7 gives the lower bound for B. It requires B to diverge at a proper rate, to give us a good chance for finding a random projection with a large value function. More specifically, under C7, we can show that

$$\mathbb{P}\left\{\max_{b=1}^{B} V(d_{\boldsymbol{S}_b}^{opt}) = V(d_{\boldsymbol{S}^*}^{opt}) + o(n^{-1/2})\right\} \to 1.$$

Shi et al. (2019a) showed C7 holds when $\tau(\boldsymbol{x}) = \phi(\boldsymbol{S}^*\boldsymbol{x})$ for some sketching matrix $\boldsymbol{S}^* \in \mathscr{S}^*$ and some Lipschitz continuous function $\phi(\cdot)$.

Condition C8 holds with $\gamma = 1$ when $\tau^{\boldsymbol{S}}(\boldsymbol{S}\boldsymbol{X}_0)$ has a uniformly bounded density function near 0 for any sketching matrix $\boldsymbol{S}$ that nearly maximizes the value function (see Shi et al. 2019a, for detailed discussion). Assume $\tau(\boldsymbol{x}) \geq \delta_0$ almost surely or $\tau(\boldsymbol{x}) \leq -\delta_0$ almost surely. Then for any sketching matrix $\boldsymbol{S}$, we have $\tau^{\boldsymbol{S}}(\boldsymbol{S}\boldsymbol{X}) \geq \delta_0$ almost surely or $\tau^{\boldsymbol{S}}(\boldsymbol{S}\boldsymbol{X}) \leq -\delta_0$. As a result, C8 automatically holds for any $\gamma > 0$.

In Section C.2 of Shi et al. (2019a), they considered a simple model and showed C6-C8 hold.

Theorem 5.2.4 *Assume Conditions C1-C3, C6, C7 hold,* $\mathbb{E}|Y_0|^3 = O(1)$, $\log B = o(n^{1/3})$ *and* $\delta_n \to 0$. *If* $h_n^* \gg \max(\sqrt{\log B}/\sqrt{n}, n^{-r_0/2}\sqrt{\log n})$*, then we have*

$$\mathbb{P}\left(\widehat{T}_{SRP} > z_{\alpha/2}\right) \to 1.$$

Moreover, assume (5.7) *and C8 hold,* $\mathbb{P}\{\tau(\boldsymbol{x}) = 0\} = 0$, $\sqrt{n}h_n = O(1)$, $B = O(n^{\kappa_B})$ *for some* $\kappa_B > 0$, $r_0 > \frac{\gamma+2}{2\gamma+2}$ *and* $\liminf_n \sigma_0 > 0$. *Then we have*

$$\mathbb{P}\left(\widehat{T}_{SRP} > z_{\alpha/2}\right) = 2\bar{\Phi}\left(z_{\frac{\alpha}{2}} - \frac{\sqrt{n}h_n}{\sigma_0}\right) - \bar{\Phi}^2\left(z_{\frac{\alpha}{2}} - \frac{\sqrt{n}h_n}{\sigma_0}\right) + o(1).$$

Assume $p = O(n)$ and we set $B = c_* n^{\{3q(s-1)+\epsilon\}/2}$ for any $c_*, \epsilon > 0$. Then the conditions $B \gg (p\sqrt{n})^{(s-1)q}$ in C7 and $B = O(n^{\kappa_B})$ in Theorem 7.3.2 automatically hold. It is worth mentioning that when h_n and σ_0 don't depend on p, Theorem 7.3.2 implies that the asymptotic power of our test is independent of p.

5.2.3 Some implementation issues

Doubly-robust test statistics

So far we have assumed that the propensity scores are known for all patients. In the following, we introduce a doubly-robust test statistic to deal with data from an observational study. We begin by introducing a doubly-robust value difference estimator, which requires the estimation of the propensity score $\pi(a,\boldsymbol{x})$ and the conditional mean functions $h(a,\boldsymbol{x})$. Denoted by $\widehat{\pi}(a,\cdot)$ and $\widehat{h}(a,\cdot)$ the corresponding estimators. Zhang et al. (2012) proposed a doubly-robust estimator for the value function under a given treatment regime d,

$$\begin{aligned}\widehat{V}^{dr}(d) &= \frac{1}{N}\sum_{i=1}^{N}\left\{\left(\frac{A_i}{\widehat{\pi}(1,\boldsymbol{X}_i)}d_i+\frac{1-A_i}{\widehat{\pi}(0,\boldsymbol{X}_i)}(1-d_i)\right)Y_i\right.\\ &- \left.\left(\frac{A_i}{\widehat{\pi}(1,\boldsymbol{X}_i)}d_i+\frac{1-A_i}{\widehat{\pi}(0,\boldsymbol{X}_i)}(1-d_i)-1\right)\{\widehat{h}(0,\boldsymbol{X}_i)(1-d_i)+\widehat{h}(1,\boldsymbol{X}_i)d_i\}\right\},\end{aligned}$$

where d_i is a shorthand for $d(X_i)$. When either the propensity score or the conditional mean models are correctly specified, $\widehat{V}^{dr}(d)$ is consistent to $V(d)$ (Zhang et al. 2012). Based on $\widehat{V}^{dr}$, for any $\mathscr{I}\subset\{1,\ldots,N\}$ and a given treatment regime d, we define our doubly-robust value difference estimator as

$$\widehat{\mathrm{VD}}_{\mathscr{I}}^{dr}(d)=\frac{1}{|\mathscr{I}|}\sum_{i\in\mathscr{I}}\left\{\left(\frac{1-A_i}{\widehat{\pi}_{0,i}^{\mathscr{I}}}-\frac{A_i}{\widehat{\pi}_{1,i}^{\mathscr{I}}}\right)Y_i-\left(\frac{1-A_i}{\widehat{\pi}_{0,i}^{\mathscr{I}}}-1\right)\widehat{h}_{0,i}^{\mathscr{I}}+\left(\frac{A_i}{\widehat{\pi}_{1,i}^{\mathscr{I}}}-1\right)\widehat{h}_{1,i}^{\mathscr{I}}\right\}(1-d_i),$$

where $\widehat{\pi}_{a,i}^{\mathscr{I}}=\widehat{\pi}^{\mathscr{I}}(a,\boldsymbol{X}_i)$, $\widehat{h}_{a,i}^{\mathscr{I}}=\widehat{h}_0^{\mathscr{I}}(a,\boldsymbol{X}_i)$ and $\widehat{\pi}_{a,i}^{\mathscr{I}}$s, $\widehat{h}_{a,i}^{\mathscr{I}}$s are obtained based on $\{O_i\}_{\mathscr{I}}$. When p is large, we recommend to estimate π, h_0 and h_1 via penalized regression. The asymptotic variance of $\sqrt{|\mathscr{I}|}\widehat{\mathrm{VD}}^{dr}(d)$ can be consistently estimated by $\widehat{\sigma}_{\mathscr{I}}^{dr}(d)$ whose exact form is given in Section 5.2.5.

We briefly summarize our test procedures. Similar to Algorithm 1, we first randomly partition the data into two halves $\{O_i\}_{\mathscr{I}_1}$ and $\{O_i\}_{\mathscr{I}_2}$, and obtain the estimators $\widehat{\pi}^{\mathscr{I}_j}$, $\widehat{h}_0^{\mathscr{I}_j}$, $\widehat{h}_1^{\mathscr{I}_j}$ based on $\{O_i\}_{i\in\mathscr{I}_j}$ for $j=1,2$. Then we independently sample the sparse sketching matrices $\boldsymbol{S}_{\mathscr{I}_1}$ and $\boldsymbol{S}_{\mathscr{I}_2}$. The sampling algorithm is similar to Algorithm 2. Specifically, for $j=1,2$, we randomly divide $\mathscr{I}_j$ into $\{\mathscr{I}_j^{(k)}\}_{k=1}^{\mathbb{K}}$ and independently sample $\boldsymbol{S}_1,\ldots,\boldsymbol{S}_B$ as Steps 2 and 3 of Algorithm 2. Then we calculate the doubly-robust value difference estimator,

$$\widehat{\mathrm{VD}}_{CV}^{dr,\boldsymbol{S}_b}=\mathbb{K}^{-1}\sum_{k}\widehat{\mathrm{VD}}_{\mathscr{I}_j^{(k)}}^{dr}(\widehat{d}_{\mathscr{I}_j^{(k)-}}^{\boldsymbol{S}_b}), \tag{5.10}$$

for each $\boldsymbol{S}_b$ where $\mathscr{I}_j^{(k)-} = \mathscr{I}_j \cap (\mathscr{I}_j^{(k)})^c$, and set $\boldsymbol{S}_{\mathscr{I}_j} = \boldsymbol{S}_{\widehat{b}}$ where $\widehat{b} = \arg\max_{b=1}^{B} \widehat{\text{VD}}_{CV}^{dr,\boldsymbol{S}_b}$. Finally, we define our test statistic by

$$\widehat{T}_{SRP}^{dr} = \max\left(\frac{\sqrt{n}\widehat{\text{VD}}_{\mathscr{I}_2}^{dr}(\widehat{d}_{\mathscr{I}_1}^{\boldsymbol{S}_{\mathscr{I}_1}})}{\max\{\widehat{\sigma}_{\mathscr{I}_2}^{dr}(\widehat{d}_{\mathscr{I}_1}^{\boldsymbol{S}_{\mathscr{I}_1}}), \delta_n\}}, \frac{\sqrt{n}\widehat{\text{VD}}_{\mathscr{I}_1}^{dr}(\widehat{d}_{\mathscr{I}_2}^{\boldsymbol{S}_{\mathscr{I}_2}})}{\max\{\widehat{\sigma}_{\mathscr{I}_1}^{dr}(\widehat{d}_{\mathscr{I}_2}^{\boldsymbol{S}_{\mathscr{I}_2}}), \delta_n\}} \right), \tag{5.11}$$

and reject the test if $\widehat{T}_{SRP}^{dr} > z_{\alpha/2}$ for a given significance level $\alpha > 0$. Statistical properties of $\widehat{T}_{SRP}^{dr}$ can be similarly established.

Estimation of $\tau^{\boldsymbol{S}}$

The projected contrast function $\tau^{\boldsymbol{S}}$ can be estimated by any machine learning or statistical nonparametric methods. In our implementation, we estimate $\tau^{\boldsymbol{S}}$ using cubic B-splines. Let $\mathscr{I}$ be an arbitrary subset of $\{1, \ldots, N\}$. Based on the dataset $\{O_i\}_{i \in \mathscr{I}}$, we first estimate $\pi(1, \cdot)$ using the penalized logistic regression with SCAD penalty functions (Fan and Li 2001). Denoted by $\widehat{\pi}_{1,i}$ the corresponding estimator for $\pi(1, \boldsymbol{X}_i)$, we set $\widehat{\pi}_{0,i} = 1 - \widehat{\pi}_{1,i}$. We then estimate $h(0, \cdot)$, $h(1, \cdot)$ using the penalized linear regression with SCAD penalty functions. Let $\widehat{h}_{a,i}^{\mathscr{I}}$ be the corresponding estimator for $h_a(\boldsymbol{X}_i)$. These penalized regressions are implemented by the R package `ncvreg` and the tuning parameters are selected via 10-folded cross-validation. Recall that $\boldsymbol{S}_{(j)} \in \mathbb{R}^{1 \times p}$ is the jth row of sketching matrix $\boldsymbol{S}$. We define the pseudo outcome

$$\widehat{\tau}_i^{\mathscr{I}} = \left(\frac{A_i}{\widehat{\pi}_{1,i}^{\mathscr{I}}} - \frac{1 - A_i}{\widehat{\pi}_{0,i}^{\mathscr{I}}} \right) Y_i - \left(\frac{A_i}{\widehat{\pi}_{1,i}^{\mathscr{I}}} - 1 \right) \widehat{h}_{1,i}^{\mathscr{I}} + \left(\frac{1 - A_i}{\widehat{\pi}_{0,i}^{\mathscr{I}}} - 1 \right) \widehat{h}_{0,i}^{\mathscr{I}}, \tag{5.12}$$

and minimize

$$(\widehat{\xi}_1^{\mathscr{I}}, \cdots, \widehat{\xi}_q^{\mathscr{I}}) = \arg\min_{\xi_1, \cdots, \xi_q \in \mathbb{R}^k} \frac{1}{|\mathscr{I}|} \sum_{i \in \mathscr{I}} \left(\widehat{\tau}_i^{\mathscr{I}} - \sum_{j=1}^{q} \sum_{k=1}^{K_0+4} N_k^{\boldsymbol{S}^{(j)}} (\boldsymbol{S}^{(j)} \boldsymbol{X}_i) \xi_{j,k} \right)^2, \tag{5.13}$$

where $N_1^{\boldsymbol{S}^{(j)}}(\cdot), \ldots, N_{K_0+4}^{\boldsymbol{S}^{(j)}}(\cdot)$ are cubic B-spline bases of $\boldsymbol{S}^{(j)} \boldsymbol{X}_i$ and K_0 is the number of interior knots. Given K_0, we place the interior knots at equally-spaced sample quantiles of the projected covariates $\{\boldsymbol{S}\boldsymbol{X}_i\}_{i \in \mathscr{I}}$. After solving (5.13), we set $\widehat{\tau}_{\mathscr{I}}^{\boldsymbol{S}}(\boldsymbol{S}\boldsymbol{x}) = \sum_{j=1}^{q} \sum_{k=1}^{K_0+4} N_k^{\boldsymbol{S}^{(j)}} (\boldsymbol{S}^{(j)} \boldsymbol{x}) \widehat{\xi}_{j,k}^{\mathscr{I}}$.

Based on the B-spline methods, Shi et al. (2019a) showed that C6 holds with $r_0 = 4/5$ when $q = 1$ and $B = O(n^{\kappa_B})$ for any $\kappa_B > 0$. Assume C8 holds with $\gamma > 2/3$. The condition $r_0 > (\gamma + 2)/(2\gamma + 2)$ in Theorem 7.3.2 is thus satisfied. More generally, we may use series estimator (Belloni et al. 2015) to estimate $\tau^{\boldsymbol{S}}$. Then the rate r_0 in C6 will decrease as the number of projected dimension q increases.

Choice of s

Our testing procedure requires specification of s, which determines the number of nonzero elements in each row of the sketching matrix. Ideally, one could treat s as a tuning parameter and choose s to maximize the estimated value difference defined in (5.10). However, this approach would be time-consuming. In our implementation, we set s as a discrete random variable when sampling $\boldsymbol{S}_1,\ldots,\boldsymbol{S}_B$. More specifically, for $b=1,\ldots,B$, we first independently sample s according as some random variable s_0, and then sample $\boldsymbol{S}_b$ according to Step 3 of Algorithm 2.

We recommend to set $s_0 = 2+\text{Binom}(p-2, p_0)$, where $\text{Binom}(m, p_0)$ is a binomial random variable with the total number of trials equal to m and the probability of success equal to p_0. In our simulation study, we set $p_0 = 2/(p-2)$.

Choice of q

The choice of the projection dimension q involves a trade-off. If q is too large, then the curse of dimensionality will affect the uniform convergence rates of $\widehat{\tau}_{\mathscr{I}}^{\boldsymbol{S}_j}$ in (A8), resulting in decreased power of the corresponding test. If q is too small, then the OTR is not well approximated. In our numerical experiments, we set $q=1$. In the supplementary article, we examine the performance of the proposed test with difference choices of q. Results show that the optimal choice of q depends on the number of covariates involved in the OTR and varies across different simulation settings. We further propose a method that adaptively determines q. Detailed algorithm can be found in Shi et al. (2019a). In our simulations, we find such adaptive method is no worse than any fixed choice of q and has nearly optimal performance in some cases.

Choices of other hyperparameters

We recommend to set the number of folds $\mathbb{K}$ in Algorithm 2 to be 5 or 10. The number of sketching matrices B shall diverge as $N, p \to \infty$. In practice, we recommend to set $B = N^{\kappa_N} p^{\kappa_p}$ for some $\kappa_N, \kappa_p \geq 1$.

5.2.4 Technical conditions

(C4.) Assume there exist some positive constants γ and δ_0 such that

$$\mathbb{P}\{0 < |\tau(\boldsymbol{X}_0)| \le t\} = O(t^{\gamma}),$$

where the big-O term is uniform in $0 < t < \delta_0$.
(C5.) Assume $\widehat{\tau}$ satisfies

$$\mathbb{E}|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_0) - \tau(\boldsymbol{X}_0)|^2 = o\left(|\mathscr{I}|^{-(2+\gamma)/(2+2\gamma)}\right) \quad \text{as } |\mathscr{I}| \to \infty,$$

where the little-o term is uniform in the training samples $\mathscr{I}$.

Condition C4 is closely related to the margin assumption (Tsybakov 2004; Audibert and Tsybakov 2007) in the classification literature. It is often used to obtain sharp upper bounds on the difference between the value function under $d^{opt,0}$ and that under an estimated OTR (Qian and Murphy 2011; Luedtke and van der Laan 2016). The larger the structure parameter γ in C4, the sharper the upper bounds. When $\tau(\boldsymbol{X}_0)$ has a bounded density function near 0, C4 holds with $\gamma = 1$. If there exists some $\delta_0 > 0$ such that $|\tau(\boldsymbol{X}_0)| \ge \delta_0$ almost surely, then C4 holds with $\gamma = +\infty$.

Condition C5 depends on the "structural" parameter γ in C4 and the convergence rates of the estimated contrast function. The larger the γ, the more likely C5 holds. When $\gamma = 1$, C5 requires $\mathbb{E}|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_0) - \tau(\boldsymbol{X}_0)|^2 = o(|\mathscr{I}|^{-3/4})$. The rates of convergence of the estimated contrast function are available for most often used machine learning or statistical methods, such as spline methods (Zhou et al. 1998), kernel ridge regression (Steinwart and Christmann 2008; Zhang et al. 2013) and random forests (Biau 2012). Shi et al. (2019a) showed C5 holds when $\widehat{\tau}$ is computed by some of the aforementioned methods. Combining C4 together with C5 gives $V(\widehat{d}_{\mathscr{I}}) = V(d^{opt,0}) + o_p(|\mathscr{I}|^{-1/2})$.

5.2.5 Variance estimator in Section 5.2.3

Define $\widehat{\boldsymbol{\alpha}}_{\mathscr{I}}$ to be the penalized logistic regression estimator for $\pi(1,\cdot)$ based on $\{(\boldsymbol{X}_i, A_i)\}_{i\in\mathscr{I}}$, $\widehat{\boldsymbol{\theta}}_{0,\mathscr{I}}$ and $\widehat{\boldsymbol{\theta}}_{1,\mathscr{I}}$ to be the penalized linear regression estimators for $h(0,\cdot)$ and $h(1,\cdot)$, based on $\{(\boldsymbol{X}_i, Y_i)\}_{i\in\mathscr{I},A_i=0}$ and $\{(\boldsymbol{X}_i, Y_i)\}_{i\in\mathscr{I},A_i=1}$ respectively. Denoted by $\mathscr{M}_{\alpha,\mathscr{I}}$ the support of $\widehat{\boldsymbol{\alpha}}_{\mathscr{I}}$, i.e, $\mathscr{M}_{\alpha,\mathscr{I}} = \{j = 1,\dots,p : \widehat{\boldsymbol{\alpha}}_{\mathscr{I},j} \ne 0\}$. Similarly define $\mathscr{M}_{\theta_0,\mathscr{I}}$ and $\mathscr{M}_{\theta_1,\mathscr{I}}$ to be the supports of $\widehat{\boldsymbol{\theta}}_{0,\mathscr{I}}$

and $\widehat{\boldsymbol{\theta}}_{1,\mathscr{I}}$ respectively. Let

$$\widehat{\pi}_{1,i} = \frac{\exp(X_i^T \widehat{\boldsymbol{\alpha}}_{\mathscr{I}})}{1+\exp(X_i^T \widehat{\boldsymbol{\alpha}}_{\mathscr{I}})} \text{ and } \widehat{\pi}_{0,i} = 1 - \widehat{\pi}_{1,i}.$$

For any treatment regime d, we define

$$\widehat{\sigma}^2_{DR,\mathscr{I}}(d) = \frac{1}{|\mathscr{I}|-1}\sum_{i\in\mathscr{I}}\kappa_i^2 - \frac{1}{|\mathscr{I}|(|\mathscr{I}|-1)}\left(\sum_{i\in\mathscr{I}}\kappa_i\right)^2,$$

where

$$\begin{aligned}
\kappa_i &= \left\{\left(\frac{1-A_i}{\widehat{\pi}_{0,i}} - \frac{A_i}{\widehat{\pi}_{1,i}}\right)Y_i - \left(\frac{1-A_i}{\widehat{\pi}_{0,i}} - 1\right)\boldsymbol{X}_i^T\widehat{\boldsymbol{\theta}}_{0,\mathscr{I}} + \left(\frac{A_i}{\widehat{\pi}_{1,i}} - 1\right)\boldsymbol{X}_i^T\widehat{\boldsymbol{\theta}}_{1,\mathscr{I}}\right\}\{1-d(\boldsymbol{X}_i)\} \\
&+ \bar{I}_1^T\left(\frac{1}{|\mathscr{I}|}\sum_{i\in\mathscr{I}}\boldsymbol{X}_{i\mathscr{M}_{\alpha,\mathscr{I}}}^T\widehat{\pi}_{1,i}\widehat{\pi}_{0,i}\boldsymbol{X}_{i\mathscr{M}_{\alpha,\mathscr{I}}}\right)^{-1}\boldsymbol{X}_{i\mathscr{M}_{\alpha,\mathscr{I}}}(A_i - \widehat{\pi}_{1,i}) \\
&- \bar{I}_2^T\left(\frac{1}{|\mathscr{I}|}\sum_{i\in\mathscr{I}}(1-A_i)\boldsymbol{X}_{i\mathscr{M}_{\theta_0,\mathscr{I}}}^T\boldsymbol{X}_{i\mathscr{M}_{\theta_0,\mathscr{I}}}\right)^{-1}\boldsymbol{X}_{i\mathscr{M}_{\theta_0,\mathscr{I}}}(1-A_i)(Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\theta}}_0) \\
&+ \bar{I}_3^T\left(\frac{1}{|\mathscr{I}|}\sum_{i\in\mathscr{I}}A_i\boldsymbol{X}_{i\mathscr{M}_{\theta_0,\mathscr{I}}}^T\boldsymbol{X}_{i\mathscr{M}_{\theta_0,\mathscr{I}}}\right)^{-1}\boldsymbol{X}_{i\mathscr{M}_{\theta_0,\mathscr{I}}}A_i(Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\theta}}_1),
\end{aligned}$$

and $\bar{I}_j = \sum_{i\in\mathscr{I}} I_{i,j}/n$ where

$$\begin{aligned}
I_{i,1} &= \left\{\frac{\widehat{\pi}_{1,i}(1-A_i)}{\widehat{\pi}_{0,i}}\{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\theta}}_{0,\mathscr{I}}\} + \frac{A_i\widehat{\pi}_{0,i}}{\widehat{\pi}_{1,i}}\{Y_i - \widehat{\boldsymbol{\theta}}_{1,\mathscr{I}}\}\right\}\boldsymbol{X}_{i\mathscr{M}_{\alpha,\mathscr{I}}}\{1-d(\boldsymbol{X}_i)\}, \\
I_{i,2} &= \left(\frac{1-A_i}{\widehat{\pi}_{0,i}} - 1\right)\boldsymbol{X}_{i\mathscr{M}_{\theta_0,\mathscr{I}}}\{1-d(\boldsymbol{X}_i)\}, \qquad I_{i,3} = \left(\frac{A_i}{\widehat{\pi}_{1,i}} - 1\right)\boldsymbol{X}_{i\mathscr{M}_{\theta_1,\mathscr{I}}}\{1-d(\boldsymbol{X}_i)\}.
\end{aligned}$$

5.3 Simulations

5.3.1 Settings

We examine the finite sample performance of the proposed tests via Monte Carlo simulations. Simulated data with sample size N were generated from

$$Y_0 = 1 + (X_0^{(1)} - X_0^{(2)})/2 + A_0\tau(\boldsymbol{X}_0) + e_0,$$

where $X_0 \sim N(0, \boldsymbol{I}_p)$, $A \sim \text{Binom}(1, 0.5)$ and $e \sim N(0, 0.5^2)$. Here, we set $p = 50$ or 100.

We consider four scenarios. In the first three scenarios, we set

$$\tau(\boldsymbol{x}) = \phi_\delta\{(x^{(1)}+x^{(2)})/\sqrt{2}\}(x^{(3)}+x^{(4)}+x^{(5)}+x^{(6)}+x^{(7)})^2/5,$$

for some function ϕ_δ parameterized by some $\delta > 0$. More specifically, we set $\phi_\delta(z) = z^2 - \delta$ in Scenario 1, $\phi_\delta(z) = \delta\cos(\pi z)$ in Scenario 2, and $\phi_\delta(z) = \delta\sqrt{2\pi}z$ in Scenario 3.

In Scenario 4, we set

$$\tau(\boldsymbol{x}) = \delta\left\{\left(\sum_{j=1}^{2}\frac{x^{(j)}}{\sqrt{2}}\right)^2 - \left(\sum_{j=3}^{20}\frac{x^{(j)}}{\sqrt{18}}\right)^2\right\}(x^{(21)}+x^{(22)}+x^{(23)}+x^{(24)}+x^{(25)})^2/5.$$

It is immediate to see that the OTR is sparse and is a function of $x^{(1)}$ and $x^{(2)}$ in the first three scenarios. In Scenario 4, however, a total of 20 variables are involved in the OTR. In addition, the true OTR is linear in $\boldsymbol{X}_0$ under Scenario 3, but non-linear under Scenarios 1, 2 and 4. We set $N = 500$ in Scenarios 1, 2 and 3, and $N = 1000$ in Scenario 4.

For all scenarios, the parameter δ controls the degree of overall qualitative treatment effects. Specifically, H_0 holds if $\delta = 0$ and H_1 holds if $\delta > 0$. For each scenario, we further consider four cases by setting $\text{VD}(d^{opt,0}) = V(d^{opt,0}) - V(1) = 0, 0.2, 0.35$ and 0.5. Note that in Scenarios 2, 3 and 4, the settings for $\text{VD}(d^{opt,0}) = 0$ are the same. Hence, in Scenarios 3 and 4, we only report the simulation results for $\text{VD}(d^{opt,0}) = 0.2, 0.35$ and 0.5.

We set $q = 1$ and calculate $\widehat{T}^{dr}_{SRP}$ as described in Section 5.2.3. The number of interior knots K in the cubic B-spline bases is specified in the following fashion. When generating $\boldsymbol{S}_{\mathscr{I}_1}$ or $\boldsymbol{S}_{\mathscr{I}_2}$, we fix $K_0 = 3$ when estimating τ^{S_b} for $b = 1, \ldots, B$. After obtaining $\boldsymbol{S}_{\mathscr{I}_1}$ and $\boldsymbol{S}_{\mathscr{I}_2}$, K_0 is tuned with cross-validation when estimating $\tau^{\boldsymbol{S}_{\mathscr{I}_1}}$ and $\tau^{\boldsymbol{S}_{\mathscr{I}_2}}$. We set $B = 10^5$ for $p = 50$ and $B = 4\times 10^5$ for $p = 100$.

The whole simulation program is implemented in R. Some subroutines, including sampling data-dependent sketching matrices $\boldsymbol{S}_{\mathscr{I}_1}$ and $\boldsymbol{S}_{\mathscr{I}_2}$ and estimating $\tau^{\boldsymbol{S}_{\mathscr{I}_1}}$ and $\tau^{\boldsymbol{S}_{\mathscr{I}_2}}$, are written in C with the GNU Scientific Library (GSL, Galassi et al. 2015).

5.3.2 Competing methods

Comparison is made among the following five test statistics:
(i) The proposed sparse random projection-based test statistic $\widehat{T}^{dr}_{SRP}$.
(ii) The dense random projection-based test statistic, denoted by $\widehat{T}^{dr}_{RP}$.
(iii) The cross-validated test statistic with the OTR estimated by the penalized least square method developed in Shi et al. (2016), denoted by $\widehat{T}_{PLS}$.

(iv) The cross-validated test statistic based on step-wise variable selection, denoted by $\widehat{T}_{VS}$.
(v) The supremum-type test statistic $\widehat{T}_{DL}$ based on the desparsified Lasso estimator (Zhang and Zhang 2014; van de Geer et al. 2014).

$\widehat{T}_{RP}^{dr}$ is computed in a similar fashion as $\widehat{T}_{SRP}^{dr}$. We randomly partition $\{1,\ldots,N\}$ into $\mathscr{I}_1 \cup \mathscr{I}_2$ of equal size, generate some data dependent sketching matrices $\boldsymbol{S}_{\mathscr{I}_1}$ and $\boldsymbol{S}_{\mathscr{I}_2}$, and construct the test statistic as in (5.11). When generating $\boldsymbol{S}_{\mathscr{I}_1}$ or $\boldsymbol{S}_{\mathscr{I}_2}$, instead of sampling B sparse sketching matrices as described in Step 3 of Algorithm 2, we generate B dense sketching matrices $\boldsymbol{S}_1,\ldots,\boldsymbol{S}_B$ according to $\boldsymbol{Z}_0/\|\boldsymbol{Z}_0\|_2$, where $\boldsymbol{Z}_0 \in \mathbb{R}^p$ is a Gaussian random vector with mean zero and identity covariance matrix, and set $\boldsymbol{S}_{\mathscr{I}_1}$ or $\boldsymbol{S}_{\mathscr{I}_2}$ to be the one that gives the largest cross-validated value difference as in (5.10). Similar to $\widehat{T}_{SRP}^{dr}$, we set $B = 10^5$ for $p = 50$ and set $B = 4\times 10^5$ for $p = 100$, and use cubic B-splines to estimate $\tau^{\boldsymbol{S}}$ for any sketching matrix S.

To calculate $\widehat{T}_{PLS}$, we first partition the data into two halves $\{O_i\}_{i\in\mathscr{I}_1}$ and $\{O_i\}_{i\in\mathscr{I}_2}$. Then for $j = 1,2$, we set $\widehat{d}_{\mathscr{I}_j}(\boldsymbol{x}) = I(\bar{\boldsymbol{x}}^T\widehat{\beta}^{\mathscr{I}_j} > 0)$ where $\bar{\boldsymbol{x}} = (1,\boldsymbol{x}^T)^T$, $\widehat{\beta}^{\mathscr{I}_j}$ is computed by

$$\widehat{\beta}^{\mathscr{I}_j} = \operatorname*{arg\,min}_{\beta\in\mathbb{R}^{p+1}} \sum_{i\in\mathscr{I}_j} \frac{1}{|\mathscr{I}_j|}\Big(Y_i - \overline{\boldsymbol{X}}_i^T\widehat{\boldsymbol{\theta}}^{\mathscr{I}_j} - (A_i - \widehat{\pi}_{1,i}^{\mathscr{I}_j})\overline{\boldsymbol{X}}_i^T\boldsymbol{\beta}\Big)^2 + \sum_{j=2}^{p+1} p_{\lambda_{n,1}}(|\boldsymbol{\beta}^{(j)}|), \tag{5.14}$$

for some penalty functions p_λ, where $\overline{\boldsymbol{X}}_i = (1,\boldsymbol{X}_i^T)^T$, $\widehat{\pi}_i^{\mathscr{I}_j}$ is the estimated propensity score for the ith patient based on a penalized logistic regression with SCAD penalty function, and $\widehat{\boldsymbol{\theta}}^{\mathscr{I}_j}$ is calculated by

$$\widehat{\boldsymbol{\theta}}^{\mathscr{I}_j} = \operatorname*{arg\,min}_{\theta\in\mathbb{R}^{p+1}} \sum_{i\in\mathscr{I}_j} \frac{1}{|\mathscr{I}_j|}(Y_i - \overline{\boldsymbol{X}}_i^T\theta)^2 + \sum_{j=2}^{p+1} p_{\lambda_{n,2}}(|\theta_j|). \tag{5.15}$$

We use the SCAD penalty in both (5.14) and (5.15). The tuning parameters $\lambda_{n,1}$ and $\lambda_{n,2}$ were selected via 10-folded cross-validation. Finally, define $\widehat{T}_{PLS}$ by

$$\widehat{T}_{PLS} = \max\left(\frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathscr{I}_2}^{dr}(\widehat{d}_{\mathscr{I}_1})}{\max\{\widehat{\sigma}_{\mathscr{I}_2}^{dr}(\widehat{d}_{\mathscr{I}_1}),\delta_n\}}, \frac{\sqrt{n}\widehat{\mathrm{VD}}_{\mathscr{I}_1}^{dr}(\widehat{d}_{\mathscr{I}_2})}{\max\{\widehat{\sigma}_{\mathscr{I}_1}^{dr}(\widehat{d}_{\mathscr{I}_2}),\delta_n\}}\right). \tag{5.16}$$

To compute $\widehat{T}_{VS}$, we similarly split the observations into two sub-datasets $\{O_i\}_{i\in\mathscr{I}_1}$ and $\{O_i\}_{i\in\mathscr{I}_2}$. For each sub-dataset, we apply the sequential advantage selection (SAS, Fan et al. 2016) to select variables with a qualitative interaction with the treatment. SAS is a greedy stepwise selection procedure and uses a BIC-type criterion to choose the best candidate subset of variables. Denoted by $\widehat{\mathscr{M}_{\mathscr{I}_1}}, \widehat{\mathscr{M}_{\mathscr{I}_2}} \subseteq \{1,\dots,p\}$ the corresponding sets of selected variables. Then for each $j=1,2$, we calculate the pseudo responses $\widehat{\tau}_i^{\mathscr{I}_j}, \forall i \in \mathscr{I}_j$ (see the definition in (5.12)) and compute

$$\widehat{\tau}_{\mathscr{I}_j} = \arg\min_{f \in \mathbb{H}_j} \frac{1}{n} \sum_{i \in \mathscr{I}_j} \{\widehat{\tau}_i^{\mathscr{I}_j} - f(X_{i,\widehat{\mathscr{M}_{\mathscr{I}_j}}})\}^2 + \lambda_j \|f\|_{\mathbb{H}_j}^2,$$

where $\lambda_j > 0$ is a tuning parameter, $\mathbb{H}_j$ is the reproducing kernel Hilbert space with the reproducing kernel $K_j(X_{i,\widehat{\mathscr{M}_{\mathscr{I}_j}}}, X_{k,\widehat{\mathscr{M}_{\mathscr{I}_j}}}) = \exp\{-\sum_{l \in \widehat{\mathscr{M}_{\mathscr{I}_j}}} \eta_{j,l} (X_i^{(l)} - X_k^{(l)})^2\}$ where $X_i^{(l)}, X_k^{(l)}$ denote the l-th element in X_i, X_k and $\eta_{j,l} > 0, \forall l \in \widehat{\mathscr{M}_{\mathscr{I}_j}}$ are tuning parameters. The estimating procedure is implemented by the R package `listdtr` and the tuning parameters are selected via leave-one-out cross validation. Then we define $\widehat{d}_{\mathscr{I}_j}(\boldsymbol{x}) = I\{\widehat{\tau}_{\mathscr{I}_j}(x_{\widehat{\mathscr{M}_{\mathscr{I}_j}}}) > 0\}$ and set

$$\widehat{T}_{VS} = \max\left(\frac{\sqrt{n}\widehat{\text{VD}}_{\mathscr{I}_2}^{dr}(\widehat{d}_{\mathscr{I}_1})}{\max\{\widehat{\sigma}_{\mathscr{I}_2}^{dr}(\widehat{d}_{\mathscr{I}_1}), \delta_n\}}, \frac{\sqrt{n}\widehat{\text{VD}}_{\mathscr{I}_1}^{dr}(\widehat{d}_{\mathscr{I}_2})}{\max\{\widehat{\sigma}_{\mathscr{I}_1}^{dr}(\widehat{d}_{\mathscr{I}_2}), \delta_n\}} \right). \tag{5.17}$$

We set $\delta_n = \log(\log_{10}(2n))/(2n)^{1/6}$ in (5.11), (5.16) and (5.17), where $\log_{10}$ denotes the logarithm with base 10.

$\widehat{T}_{DL}$ tests the overall treatment effects by fitting the following linear regression model for the response:

$$\mathbb{E}(Y_0|A_0, \boldsymbol{X}_0) \approx \beta_0 + \boldsymbol{X}_0^T \boldsymbol{\beta}_x + A_0 \beta_a + A \boldsymbol{X}_0^T \boldsymbol{\beta}_{ax}.$$

Based on this model, testing the overall treatment effects is equivalent to test $H_0^*: \boldsymbol{\beta}_{ax} = 0$. Denoted by $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_x^T, \beta_a, \boldsymbol{\beta}_{ax}^T)^T$. To deal with high dimensionality, we estimate $\boldsymbol{\beta}$ by the desparsified Lasso estimator $\widehat{\boldsymbol{\beta}}^{DL}$ and test H_0^* based on the following supremum-type test statistic, $\max_{j \in \mathscr{M}_{ax}} \sqrt{n} |\widehat{\beta}_j^{DL}|$, where $\mathscr{M}_{ax} = \{p+3, \dots, 2p+2\}$ and $\widehat{\boldsymbol{\beta}}^{DL,(j)}$ is the j-th element of $\widehat{\boldsymbol{\beta}}^{DL}$. The critical value of $\widehat{T}^{DL}$ is approximated via bootstrap. Detailed implementation of the test can be found in Zhang and Cheng (2017).

5.3.3 Results

We conduct 500 simulations for each setting and report the proportions of rejecting the null hypothesis (%) in Table 6.1 and Table 5.2, with standard errors in parenthesis (%). Under H_0, the type-I errors of our test statistic is well controlled. Specifically, in Scenario 1 when $\text{VD} = 0$, the rejection probability of $\widehat{T}_{SRP}^{dr}$ is exactly zero. This is in line with our theory which suggests that the type-I error of our test statistics will converge to 0 in the regular cases where $\mathbb{P}\{\tau(\boldsymbol{X}_0)=0\}=0$. In Scenario 2 when VD = 0, the rejection probability of $\widehat{T}^{dr}_{SRP}$ is close to the nominal level.

Table 5.1: Rejection probabilities (%) of the sparse random projection-based test, dense random projection-based test, penalized least square-based test, step-wise selection-based test and the supremum-type test based on the desparsified Lasso estimator, with standard errors in parenthesis (%), under Scenarios 1 and 2 where $\boldsymbol{X}_0 \sim N(0, \boldsymbol{I}_p)$.

Scenario 1		VD = 0		VD = 20%		VD = 35%		VD = 50%	
		α level		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}^{dr}_{SRP}$	50	0(0)	0(0)	24(1.9)	39.6(2.2)	71(2.0)	81(1.8)	90.8(1.3)	95.2(1.0)
	100	0(0)	0(0)	17.4(1.7)	29.6(2.0)	60.8(2.2)	73.8(2.0)	86.6(1.5)	92.4(1.2)
$\widehat{T}^{dr}_{RP}$	50	0(0)	0(0)	0.2(0.2)	0.6(0.4)	0.8(0.4)	3.2(0.8)	7.2(1.2)	18.6(1.7)
	100	0(0)	0(0)	0.4(0.3)	0.4(0.3)	0.4(0.3)	4(0.9)	6.8(1.1)	19(1.8)
$\widehat{T}_{PLS}$	50	0(0)	0(0)	0(0)	0(0)	0.4(0.3)	0.8(0.4)	6(1.1)	17.6(1.7)
	100	0(0)	0(0)	0(0)	0(0)	0.8(0.4)	2.4(0.7)	8.6(1.3)	19.8(1.8)
$\widehat{T}_{VS}$	50	0(0)	0(0)	1.2(0.5)	3.8(0.9)	16 (1.6)	29.4 (2.0)	36.6(2.2)	50.8(2.2)
	100	0(0)	0(0)	0(0)	0.6(0.3)	8.4 (1.2)	17.4 (1.7)	23.8(1.9)	36.4(2.2)
$\widehat{T}_{DL}$	50	10.2(1.4)	22.4(1.9)	11.2(1.4)	22.8(1.9)	10.8 (1.4)	21.8 (1.9)	9.8(1.3)	22.4(1.9)
	100	7.6(1.2)	20.0(1.8)	7.8(1.2)	21.4(1.8)	7.6 (1.2)	22.0 (1.9)	6.8(1.1)	21.6(1.8)
Scenario 2		VD = 0		VD = 20%		VD = 35%		VD = 50%	
		α level		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}^{dr}_{SRP}$	50	1.2(0.5)	5.4(1)	24(1.9)	35.8(2.1)	76.4(1.9)	84.6(1.6)	90.2(1.3)	94(1.1)
	100	0.6(0.3)	5.2(1)	15.2(1.6)	28.2(2)	67(2.1)	78.8(1.8)	84.2(1.6)	90.4(1.3)
$\widehat{T}^{dr}_{RP}$	50	1.8(0.6)	4.6(0.9)	2(0.6)	4.8(1)	1.6(0.6)	5.4(1)	1(0.4)	6(1.1)
	100	1.2(0.5)	4.2(0.9)	1.2(0.5)	5.4(1)	0.6(0.3)	4.8(1)	0.8(0.4)	4.4(0.9)
$\widehat{T}_{PLS}$	50	1.8(0.6)	6(1.1)	1.2(0.5)	4.4(0.9)	1(0.4)	4.2(0.9)	0.8(0.4)	3.8(0.9)
	100	1.2(0.5)	4.2(0.9)	0.8(0.4)	4.6(0.9)	0.6(0.3)	5.6(1)	0.6(0.3)	5(1)
$\widehat{T}_{VS}$	50	1.2(0.5)	6.4(1.1)	0.6(0.3)	4(0.9)	1(0.4)	6.6(1.1)	1(0.4)	5(1)
	100	1.4(0.5)	5(1.0)	1.0(0.4)	5(1.0)	1.4(0.5)	6.4(1.1)	0.6(0.3)	4.6(0.9)
$\widehat{T}_{DL}$	50	1.6(0.6)	6.4(1.1)	2.8(0.7)	11.8(1.4)	4.4 (0.9)	15.4 (1.6)	5.4 (1.0)	17(1.7)
	100	1.2(0.5)	3.6(0.8)	2.8(0.7)	11.8(1.4)	5.2(1.0)	17.6(1.7)	7.2(1.2)	19.8(1.8)

Table 5.2: Rejection probabilities (%) of the sparse random projection-based test, dense random projection-based test, penalized least square-based test, step-wise selection-based test and the supremum-type test based on the desparsified Lasso estimator, with standard errors in parenthesis (%), under Scenarios 3 and 4 where $\boldsymbol{X}_0 \sim N(0, \boldsymbol{I}_p)$.

Scenario 3		VD = 20%		VD = 35%		VD = 50%	
		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}^{dr}_{SRP}$	50	47.2(2.2)	71.8(2)	92.4(1.2)	97.8(0.7)	99(0.4)	100(0)
	100	42.4(2.2)	61.2(2.2)	89.8(1.4)	96.2(0.9)	97.2(0.7)	99.4(0.3)
$\widehat{T}^{dr}_{RP}$	50	4.4(0.9)	16.2(1.6)	13.4(1.5)	35.8(2.1)	22(1.9)	49.4(2.2)
	100	3(0.8)	8.4(1.2)	4(0.9)	14.2(1.6)	5.4(1)	19.6(1.8)
$\widehat{T}_{PLS}$	50	76.4(1.9)	92(1.2)	97.8(0.7)	99.4(0.3)	99.4(0.3)	100(0)
	100	64.8(2.1)	87(1.5)	97(0.8)	99.4(0.3)	98.6(0.5)	99.8(0.2)
$\widehat{T}_{VS}$	50	55.6(2.2)	81.8(1.7)	93(1.1)	99(0.4)	97.8(0.7)	100(0)
	100	49.8(2.2)	74.2(2.0)	90(1.3)	98.6(0.5)	99(0.4)	100(0)
$\widehat{T}_{DL}$	50	99.8(0.7)	100(0)	100(0)	100(0)	100(0)	100(0)
	100	99.2(0.4)	100(0)	100(0)	100(0)	100(0)	100(0)
Scenario 4		**VD = 20%**		**VD = 35%**		**VD = 50%**	
		α level		α level		α level	
	p	0.01	0.05	0.01	0.05	0.01	0.05
$\widehat{T}^{dr}_{SRP}$	50	22.4(1.9)	41.8(2.2)	60.4(2.2)	76.6(1.9)	72.4(2)	87.2(1.5)
	100	15.2(1.6)	28(2)	49.6(2.2)	70.2(2)	70(2)	84(1.6)
$\widehat{T}^{dr}_{RP}$	50	0.4(0.3)	6.2(1.1)	0.6(0.3)	5.4(1)	0.2(0.2)	5.4(1)
	100	1.2(0.5)	6(1.1)	0.8(0.4)	3.8(0.9)	1.2(0.5)	5.2(1)
$\widehat{T}^{dr}_{PLS}$	50	1.2(0.5)	5.4(1)	1.2(0.5)	6(1.1)	1.4(0.5)	4.8(1)
	100	1.6(0.6)	5.8(1)	1.8(0.6)	6(1.1)	1.4(0.5)	5.2(1)
$\widehat{T}^{dr}_{VS}$	50	10.4(1.4)	24.2(1.9)	13.6(1.5)	30.6(2.1)	13.2(1.5)	29.4(2)
	100	5(1)	15.6(1.6)	4.6(0.9)	20(1.8)	8.2(1.2)	18.4(1.7)
$\widehat{T}^{dr}_{DL}$	50	4.2(0.9)	11.4(1.4)	5.4(1)	14.2(1.6)	6.4(1.1)	15.8(1.6)
	100	6.2(1.1)	16(1.6)	6.4(1.1)	18.2(1.7)	6.8(1.1)	19.6(1.8)

Under H_1, we can see that our test statistic is much more powerful compared to other competing test statistics in Scenarios 1, 2 and 4. For example, when VD $= 0.35$ and $\alpha = 0.05$, the rejection probabilities of our test are around 75% in Scenario 1. On the other hand, $\widehat{T}^{dr}_{RP}$, $\widehat{T}_{PLS}$ and $\widehat{T}_{VS}$ fail in Scenario 2. Specifically, the rejection probabilities of these three tests are no more than 6% in all settings. The rejection probabilities of $\widehat{T}_{DL}$ are around 10%-20% in Scenario 2 under H_1. However, $\widehat{T}_{DL}$ doesn't have valid type-I error rates under H_0. Here, the test statistics $\widehat{T}_{PLS}$ and $\widehat{T}_{DL}$ fail mainly due to the fact that the true OTR is not linear, while $\widehat{T}^{dr}_{RP}$ and $\widehat{T}_{VS}$ fail partly because the dense projection and greedy stepwise variable selection cannot correctly identify the variables with qualitative interactions.

In Scenario 3, $\widehat{T}_{DL}$ and $\widehat{T}_{PLS}$ achieve the greatest power in all settings as expected since the true OTR is linear in this scenario. Notice that $X_0^{(1)}, X_0^{(2)}, \cdots, X_0^{(7)}$ are independent. Although the contrast function is not linear, the estimated contrast functions via the penalized least squares (see (5.14) and (5.15)) will converge to $\mathbb{E}\{\tau(\boldsymbol{X}_0)|X_0^{(1)}, X_0^{(2)}\}$. As a result, the estimated OTR is consistent. When VD $= 0.35$ and 0.5, the rejection probabilities of $\widehat{T}^{dr}_{SRP}$ are slightly smaller when compared to $\widehat{T}_{PLS}$, $\widehat{T}_{DL}$ and $\widehat{T}_{VS}$, but are much larger than those of $\widehat{T}_{RP}^{dr}$.

Shi et al. (2019a) conducted some additional simulation studies to examine the numerical performance of $\widehat{T}_{SRP}^{dr}$, $\widehat{T}_{RP}^{dr}$, $\widehat{T}_{PLS}$, $\widehat{T}_{VS}$ and $\widehat{T}_{DL}$ under the scenario where $\boldsymbol{X}_0 \sim N(0, \{0.5^{|i-j|}\}_{i,j=1,\dots,p})$. Results are similar to those presented in Table 6.1 and 5.2.

5.3.4 Computation time

Our tests are computed on a 32 core 2.2GHz machine with 512GB RAM. Fixing $B = 10^5$, it took approximately 3 minutes to implement the test in Scenarios 1-3 where $N = 500$ and 5 minutes in Scenario 4 where $N = 1000$. The computation time can be largely reduced if we use a much smaller B. For example, if we set $B = 10^4$ in some simulation settings, the computation is 10 times faster and the test performance is still satisfactory. Moreover, since our testing procedure independently generates many sketching matrices and retains the one that maximizes the estimated value function, it can be naturally implemented in parallel. This scalability can further effectively reduce the computational cost.

5.4 Real data

We apply our proposed test to the data from the Nefazodone-CBASP clinical trial study (Keller et al. 2000), which enrolled 681 patients with nonpsychotic chronic major depressive disorder (MDD). Patients were randomized to three treatments, including Nafazodone (coded as 0), Cognitive Behavioral-Analysis System of Psychotherapy (CBASP, coded as 1), and the combination of Nefazodone and CBASP (2). The outcome of interests were patients' scores on the 24-item Hamilton Rating Scale for Depression (HRSD). The maximum value of HRSD was 43 and we set $Y_0 = 43 - \text{HRSD}$ as our response. Larger value of Y_0 indicates better clinical outcome. Similarly as in Zhao et al. (2012), we use a subset of 647 patients that have complete records of 50 baseline covariates for analysis. Among them, 216 were treated with Nafazodone, 220 with CBASP and 211 with the combination.

Our objective was to test whether the baseline covariates $\boldsymbol{X}_0$ have overall qualitative treatment effects. This is equivalent to test $H_0 : V(d^{opt,0}) = \max\{V(0), V(1), V(2)\}$, where $V(d^{opt,0})$ is the optimal value function, and $V(j)$ denotes the value function under the fixed treatment regimes by assigning all patients to treatment j, for $j = 0, 1, 2$. Patients' average responses under treatment $0, 1, 2$ are 27.14, 27.27 and 32.13, respectively. Besides, pairwise t tests show that $V(2)$ is significantly larger than $V(0)$ and $V(1)$. Therefore, it suffices to test $H_0 : V(d^{opt,0}) = V(2)$. This is equivalent to test the intersection of the following two hypotheses:

$$H_0^{(j)}: V(d^{opt,(j)}) = \max_{k\in\{0,1,2\},k\neq j} V(k),$$

for $j = 0,1$, where $d^{opt,(j)}$ is the optimal treatment regime comparing Treatment 2 with Treatment j. For testing $H_0^{(j)}$, we computed the test statistic $\widehat{T}_{SRP}^{dr,j}$ as described in Section 5.2.3 and 5.3.1. We set $B = 100000$ and $\delta_n = \log(\log_{10}(2n))/(2n)^{1/6}$. For a given $0 < \alpha < 1$, we reject H_0 if

$$\max_{j=0,1} \widehat{T}_{SRP}^{dr,j} > z_{\alpha/4}.$$

By Bonferroni's inequality, the type-I error is well-controlled.

The two test statistics are equal to −0.67 and 0.31, respectively. We fail to reject H_0 at a significance level of 0.1. Therefore, we suspect that the prognostic covariates in this study might not have qualitative treatment effects. Zhao et al. (2012) performed pairwise comparisons between the combination treatment and any single treatment, and estimate the OTR by the outcome weighted learning. Their estimated optimal treatment regime recommended the combination treatment to all the patients. Our tests formally verify their findings.

5.5 Discussion

In this chapter, we develop tests for overall qualitative treatment effects. The test statistics are constructed by a sample-splitting method. In the high-dimensional setting, we use sparse random projections of the covariate space to construct the test statistic and introduce a data-dependent way to sample sparse projection matrices. In theory, we show the consistency of the proposed test statistic and prove its "oracle" property in the regular cases.

5.5.1 Nonnegative average treatment effects

We assume $V(1) \geq V(0)$ (the new treatment is on average better than the standard control) and consider the test statistic based on estimators for the value difference $V(d^{opt,0}) - V(1)$. When such prior information is not available, let $\widehat{a}_{\mathscr{I}_j} = \arg\max_{a\in\{0,1\}} \widehat{V}_{\mathscr{I}_j}(a)$ for $j = 0,1$ where $\mathscr{I}_1$ and $\mathscr{I}_2$ stands for a random partition of the dataset, $\widehat{V}_{\mathscr{I}_j}(a)$ denotes the estimated value function based on observations in $\mathcal{I}_j$ under the decision rule $d(\boldsymbol{x}) = a, \forall \boldsymbol{x}$. We can consider the following test statistic,

$$\widehat{T}_{CV} = \max\left\{ \frac{\sqrt{|\mathcal{I}_2|}\{\widehat{V}_{\mathcal{I}_2}(\widehat{d}_{\mathcal{I}_1}) - \widehat{V}_{\mathcal{I}_2}(\widehat{a}_{\mathcal{I}_1})\}}{\widehat{\sigma}_{\mathcal{I}_2}(\widehat{d}_{\mathcal{I}_1}, \widehat{a}_{\mathcal{I}_1})}, \frac{\sqrt{|\mathcal{I}_1|}\{\widehat{V}_{\mathcal{I}_1}(\widehat{d}_{\mathcal{I}_2}) - \widehat{V}_{\mathcal{I}_1}(\widehat{a}_{\mathcal{I}_2})\}}{\widehat{\sigma}_{\mathcal{I}_1}(\widehat{d}_{\mathcal{I}_2}, \widehat{a}_{\mathcal{I}_2})} \right\},$$

where $\widehat{\sigma}^2_{\mathcal{I}}(d, a)$ denotes some consistent estimator for the asymptotic variance of $\sqrt{|\mathcal{I}|}\{\widehat{V}_{\mathcal{I}}(d) - \widehat{V}_{\mathcal{I}}(a)\}$ for a given regime d and $a \in \{0, 1\}$. The null is rejected if $\widehat{T}_{CV} > z_{\alpha/2}$ for a given significance level α. Using similar arguments in Theorem 5.2.1 and Theorem 5.2.2, we can show that such a testing procedure is consistent.

5.5.2 Multi-stage studies

Currently, we only consider a single stage study. For multiple-stage studies, it suffices to test whether the value function under the optimal dynamic treatment regime is strictly larger than those under nondynamic treatment regimes. Zhang et al. (2013) proposed an inverse propensity-score weighted estimator for the value function under an arbitrary dynamic treatment regime. Denoted by $\widehat{\text{VD}}_{\mathcal{I}}(d_1, d_2)$ the corresponding estimator for the value difference between two dynamic treatment regimes d_1 and d_2, and $\widehat{d}_{\mathcal{I}}$ the estimated optimal dynamic treatment regime, based on the sub-dataset $\mathcal{I}$. Consider the following test statistic:

$$\widehat{T}_{CV} = \max\left\{ \min_{d \in \mathcal{D}_{nd}} \frac{\sqrt{|\mathcal{I}_2|}\widehat{\text{VD}}_{\mathcal{I}_2}(\widehat{d}_{\mathcal{I}_1}, d)}{\widehat{\sigma}_{\mathcal{I}_2}(\widehat{d}_{\mathcal{I}_1}, d)}, \min_{d \in \mathcal{D}_{nd}} \frac{\sqrt{|\mathcal{I}_1|}\widehat{\text{VD}}_{\mathcal{I}_1}(\widehat{d}_{\mathcal{I}_2}, d)}{\widehat{\sigma}_{\mathcal{I}_1}(\widehat{d}_{\mathcal{I}_2}, d)} \right\},$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ stand for a random partition of the dataset, $\widehat{\sigma}^2_{\mathcal{I}}(d_1, d_2)$ some consistent estimator for the asymptotic variance of $\sqrt{|\mathcal{I}|}\widehat{\text{VD}}_{\mathcal{I}}(d_1, d_2)$ and $\mathcal{D}_{nd}$ denotes the set of non-dynamic treatment regimes.

Note that for $j = 1, 2$, we have that under the null,

$$\min_{d \in \mathcal{D}_{nd}} \frac{\sqrt{|\mathcal{I}_j|}\widehat{\text{VD}}_{\mathcal{I}_j}(\widehat{d}_{\mathcal{I}_j^c}, d)}{\widehat{\sigma}_{\mathcal{I}_j}(\widehat{d}_{\mathcal{I}_j^c}, d)} \le \min_{d \in \mathcal{D}_{nd}} \frac{\sqrt{|\mathcal{I}_j|}\{\widehat{\text{VD}}_{\mathcal{I}_j}(\widehat{d}_{\mathcal{I}_j^c}, d) - \text{VD}(\widehat{d}_{\mathcal{I}_j^c}, d)\}}{\widehat{\sigma}_{\mathcal{I}_j}(\widehat{d}_{\mathcal{I}_j^c}, d)} \xrightarrow{L} \min_{d \in \mathcal{D}_{nd}} Z_d, \quad (5.18)$$

where $\text{VD}(d_1, d_2) = \mathbb{E}\widehat{\text{VD}}_{\mathcal{I}}(d_1, d_2)$ and $\{Z_d\}_{d \in \mathcal{D}_{nd}}$ is a set of mean zero Gaussian random variables whose covariance matrix can be consistently estimated from data. For a given significance level α, we reject the null if $\widehat{T}_{CV} > \widehat{c}_{\alpha/2}$ where $\widehat{c}_\alpha$ corresponds to some consistent estimator for $\mathbb{P}(\min_{d \in \mathcal{D}_{nd}} Z_d > z_\alpha)$. It follows from the Bonferroni's inequality and (5.18) that the type-I error of $\widehat{T}_{CV}$ is well-controlled. In the high-dimensional setting, we can calculate $\widehat{T}_{CV}$ based on sparse random projections of the covariate space. Details are omitted for brevity.

CHAPTER

6

TESTING CONDITIONAL QUALITATIVE TREATMENT EFFECTS

6.1 Introduction

In Chapter 5, we focus on testing the overall qualitative treatment effects of patients' baseline covariates. In this chapter, we develop a novel testing procedure for conditional qualitative treatment effects (CQTE) of a set of variables given another set of variables. The contributions of this chapter can be summarized as three folds. First, we mathematically formalize the notion of CQTE without assuming any parametric form of the treatment-covariates interactions and systematically characterize several equivalent representations of no CQTE. Informally speaking, a variable is said to have no qualitative treatment effects conditional on other variables if including it in treatment decision can not lead to a treatment regime that increases the value function. It naturally generalizes the definition for the qualitative interaction of a single covariate and treatment given in Gunter et al. (2011) and the definition for the overall qualitative treatment effects.

Our second contribution is to propose robust test statistics based on a kernel estimator for the conditional treatment effects for testing the existence of CQTE, which do not require the specification of the outcome model and the parametric form of treatment decision rules. To the best of our knowledge, this is the first time that such hypothesis testing problems are formally studied. Compared with the global tests in Chang et al. (2015) and Hsu (2017), our proposed tests for the CQTE can offer a new and important tool for assessing the incremental value of a set of new variables in optimal treatment decision making conditional on an existing set of qualitative covariates. Take the AIDS Clinical Trials Group Protocol 175 (ACTG175) study as an example. Many works in the literature have found that the age variable has significant qualitative interaction with the treatment (Lu et al. 2013; Fan et al. 2017). It is therefore of great importance to explore the CQTE of a new variable or a set of new variables given the age variable. The proposed tests can also help to construct the optimal treatment regime. When the null hypothesis of no CQTE is rejected, we conclude that including the new variables in treatment decision can increase the value function. Therefore, it is more desirable to construct the optimal treatment regime based on both the new and existing sets of variables.

Using the Poissonization technique (Giné et al. 2003), we show that our test statistic has correct size under the null and non-negligible powers against some nonstandard local alternatives. To deal with data from observational studies, we further introduce a doubly-robust test statistic that is consistent when either the propensity score model or the conditional mean models for the response are correctly specified.

Thirdly, the proposed test can help to discover new variables with the CQTE. Specifically, we develop a procedure for selecting qualitative variables in a sequential order based on the p-values of the proposed CQTE test. For simplicity, we only consider forward selection in this paper. Backward or stepwise selection procedure can be similarly developed.

The rest of the paper is organized as follows. We present the definition of CQTE and a class of equivalent representations for the null hypothesis of no CQTE in Section 6.2. Our proposed testing statistic and its asymptotic properties under the null, fixed alternative and nonstandard local alternatives are given in Section 6.3. In Section 6.4, we extend our testing procedure to the case where the propensity score is unknown and needs to be estimated from data, and introduce a doubly-robust version of the test statistic. Some implementation issues are discussed in Section 6.5. Simulations studies are conducted to evaluate the empirical performance of the proposed test in Section 6.6, followed by an application to an AIDS clinical trial data in Section 6.7. Here, variables with qualitative treatment effects are selected in a forward selection procedure based on the proposed test. A discussion is given in Section 6.8 and all technical proofs can be found in Shi et al. (2019c).

6.2 Conditional qualitative treatment effects

We focus on the setting with a single decision point (see Section 1.2 for the notations and model setup). In treatment decision making, Gunter et al. (2011) made a distinction between predictive and prescriptive variables. In particular, the prescriptive variables have qualitative interaction with treatment, which are important for treatment prescription. They gave a formal definition of the qualitative interaction between a single covariate and treatment. We first extend the definition by introducing the notion of conditional qualitative treatment effects (CQTE). Let B and C be two disjoint subsets of $I_0 \equiv \{1, 2, \dots, p\}$. Denoted by p_B and p_C the number of elements in B and C, respectively.

Definition 6.2.1 (CQTE) *Variables in C have qualitative treatment effect conditional on variables in B if there exist some nonempty sets $\mathscr{C}_1, \mathscr{C}_2 \subseteq \mathbb{R}^{p_C}$, $\mathscr{B} \subseteq \mathbb{R}^{p_B}$, such that (i)*

$$\mathbb{P}\left\{(\boldsymbol{X}_0^B, \boldsymbol{X}_0^C) \in \mathscr{B} \times \mathscr{C}_1\right\} > 0, \ \ and \ \mathbb{P}\left\{(\boldsymbol{X}_0^B, \boldsymbol{X}_0^C) \in \mathscr{B} \times \mathscr{C}_2\right\} > 0;$$

and (ii) for any $\boldsymbol{x}_{C_1} \in \mathscr{C}_1$, $\boldsymbol{x}_{C_2} \in \mathscr{C}_2$ and $\boldsymbol{x}_B \in \mathscr{B}$, we have

$$\begin{aligned} &\arg\max_a \mathbb{E}\left\{ Y_0^*(a) | \boldsymbol{X}_0^B = \boldsymbol{x}_B, \boldsymbol{X}_0^C = \boldsymbol{x}_{C_1} \right\} \\ \neq &\arg\max_a \mathbb{E}\left\{ Y_0^*(a) | \boldsymbol{X}_0^B = \boldsymbol{x}_B, \boldsymbol{X}_0^C = \boldsymbol{x}_{C_2} \right\}. \end{aligned} \tag{6.1}$$

For any $j = 1, 2$, when

$$\mathbb{E}\left\{ Y_0^*(1) | \boldsymbol{X}_0^B = \boldsymbol{x}_B, \boldsymbol{X}_0^C = \boldsymbol{x}_{C_j} \right\} = \mathbb{E}\left\{ Y^*(0) | \boldsymbol{X}_0^B = \boldsymbol{x}_B, \boldsymbol{X}_0^C = \boldsymbol{x}_{C_j} \right\},$$

the argmax in (6.1) is not unique. For any two functions $\psi_1(a)$ and $\psi_2(a)$, we define

$$\arg\max_a \psi_1(a) \neq \arg\max_a \psi_2(a)$$

if any maximizer of ψ_1 is not the maximizer of ψ_2 or vice versa.

Restricting $B = \emptyset$ and $p_C = 1$, we obtain a similar definition of the qualitative interaction between a single covariate and treatment as in Gunter et al. (2011). For an arbitrary subset $D \subseteq I_0$, let

$$\tau^D(\boldsymbol{x}_D) = \mathbb{E}\{\tau(\boldsymbol{X}_0) | \boldsymbol{X}_0^D = \boldsymbol{x}_D\}.$$

We now introduce the optimal treatment regime $d^{opt,D}$ based on covariates in a subset

$D \subseteq I_0$. Similar to the definition of d^{opt}, we define $d^{opt,D}$ to be the treatment regime that maximizes the value function among the class of treatment regimes based only on covariates $\boldsymbol{X}_0^D$. Specifically,

$$d^{opt,D} = \arg\max_{d^D} \mathbb{E}\{Y_0^*(d^D)\},$$

where the maximum is taken over all possible maps $d^D : \boldsymbol{X}_0^D \to \{0,1\}$.

Under C1-C3, we have for a given treatment regime d^D that

$$\begin{aligned} & \mathbb{E}\{Y_0^*(d^D)\} = \mathbb{E}\{Y_0^*(0) + \tau(\boldsymbol{X}_0) d^D(\boldsymbol{X}_0^D)\} \qquad (6.2) \\ = \ & \mathbb{E}\{Y_0^*(0)\} + \mathbb{E}[\mathbb{E}\{\tau(\boldsymbol{X}_0)|\boldsymbol{X}_0^D\} d^D(\boldsymbol{X}_0^D)] \\ = \ & \mathbb{E}\{Y_0^*(0)\} + \mathbb{E}\{\tau^D(\boldsymbol{X}_0^D) d^D(\boldsymbol{X}_0^D)\}. \end{aligned}$$

It follows from (6.2) that

$$d^{opt,D}(\boldsymbol{x}_D) = \mathbb{I}\{\tau^D(\boldsymbol{x}_D) \geq 0\}.$$

The aim of this paper is to test the following null hypothesis:

$$H_0 : \quad \boldsymbol{X}_0^C \text{ does not have CQTE given } \boldsymbol{X}_0^B,$$

against the alternative

$$H_1 : \boldsymbol{X}_0^C \text{ has CQTE given } \boldsymbol{X}_0^B.$$

Let $W = B \cup C$. To better understand the null, we introduce some examples below.

Example 6.2.1 (Testing unconditional qualitative treatment effects) *Let $B = \emptyset$. Then for any set $C \subseteq I$ and we are testing whether $\boldsymbol{X}_0^C$ has qualitative treatment effects. When it does, we can find two nonempty sets Ω_1 and Ω_2 such that $\mathbb{P}(\boldsymbol{X}_0^C \in \Omega_1) > 0, \mathbb{P}(\boldsymbol{X}_0^C \in \Omega_2) > 0$, and $\tau^C(\cdot) > 0$ on Ω_1 while $\tau^C(\cdot) < 0$ on Ω_2. Hence, it is equivalent to test the null hypothesis*

$$\tau^C(\boldsymbol{X}_0^C) \geq 0, a.s, \text{ or } \tau^C(\boldsymbol{X}_0^C) \leq 0, a.s.$$

This reduces to the problem of testing overall qualitative treatment effects discussed in Chapter 5.

Example 6.2.2 (Testing conditional qualitative treatment effects) *Assume we know covariates $\boldsymbol{X}_0^B$ have qualitative treatment effects. Our focus is to test whether some additional variables $\boldsymbol{X}_0^C$ are "important" in decision making given $\boldsymbol{X}_0^B$. Here, the "importance" is measured by the difference of the value functions under regimes $d^{opt,W}$ and $d^{opt,B}$. As we will see below, this definition is equivalent to the conditional qualitative treatment effects of $\boldsymbol{X}_0^C$ given $\boldsymbol{X}_0^B$.*

Define the error rate

$$\mathrm{ER}^{W,B} = \begin{cases} 0, & \text{if } \tau^W(\boldsymbol{X}_0^W) = 0, a.s. \\ \dfrac{\mathbb{E}[|d^{opt,W}(\boldsymbol{X}_0^W) - d^{opt,B}(\boldsymbol{X}_0^B)| I\{\tau^W(\boldsymbol{X}_0^W) \neq 0\}]}{\mathbb{P}\{\tau^W(\boldsymbol{X}_0^W) \neq 0\}}, & \text{otherwise,} \end{cases}$$

and the difference of the value function

$$\begin{aligned} \mathrm{VD}^{W,B} &= \mathbb{E}\{Y_0^*(d^{opt,W})\} - \mathbb{E}\{Y_0^*(d^{opt,B})\} \\ &= \mathbb{E}[\tau(\boldsymbol{X}_0)\{d^{opt,W}(\boldsymbol{X}_0^W) - d^{opt,B}(\boldsymbol{X}_0^B)\}] \\ &= \mathbb{E}[\tau^W(\boldsymbol{X}_0^W)\{d^{opt,W}(\boldsymbol{X}_0^W) - d^{opt,B}(\boldsymbol{X}_0^B)\}]. \end{aligned} \tag{6.3}$$

The error rate measures the proportion that the treatment regime $d^{opt,B}$ makes a different decision compared with $d^{opt,W}$. When $\tau^W(\boldsymbol{X}_0^W) \neq 0, a.s$, $\mathrm{ER}^{W,B}$ is equal to

$$\mathrm{ER}_*^{W,B} = \mathbb{E}|d^{opt,W}(\boldsymbol{X}_0^W) - d^{opt,B}(\boldsymbol{X}_0^B)|. \tag{6.4}$$

For the value difference, it follows from (6.3) that $\mathrm{VD}^{W,B} \geq 0$.

Denoted by $\Omega^W = \Omega^B \times \Omega^C$ the support of $\boldsymbol{X}_0^W$. We assume Ω^B and Ω^C are open subsets in $\mathbb{R}^{p_B}$ and $\mathbb{R}^{p_C}$, respectively. In addition, the density f^W of $\boldsymbol{X}_0^W$ is absolutely continuous with respect to the Lebesgue measure ν. We use subscripts and write $\boldsymbol{x}_W$ (or $\boldsymbol{x}_B$, $\boldsymbol{x}_C$) to refer to an arbitrary $|W|$-dimensional (or $|B|$, $|C|$-dimensional) vector. For any $\boldsymbol{x}_W \in \Omega^W$, we write $\boldsymbol{x}_W^B$ and $\boldsymbol{x}_W^C$ to denote the corresponding sub-vectors of x_W formed by elements in B and C. If B (or C) is a single-element set, i.e, $B = \{j_0\}$, we write $\boldsymbol{x}^B$, $\boldsymbol{x}_W^B$ as $x^{(j_0)}$ and $x_W^{(j_0)}$. When $W = I_0$, we omit the subscript W and write $\boldsymbol{x}_W^B$ as $\boldsymbol{x}^B$. For notational convenience, we write $\tau^W(\boldsymbol{x}_W) = \tau^W(\boldsymbol{x}_W^B, \boldsymbol{x}_W^C)$ for any $\boldsymbol{x}_W \in \Omega^W$.

Theorem 6.2.1 (Characterization of the null) *Assume that $\tau^W(\cdot)$ and $\tau^B(\cdot)$ are continuous, and $\mathbb{E}\{\tau^W(\boldsymbol{X}_0^W)\}^2 < \infty$. Then, the followings are equivalent:*

(i) H_0 holds.

(ii) $VD^{W,B} = 0$.

(iii) $ER^{W,B} = 0$.

(iv) *For any* $\boldsymbol{x}_W$ *such that* $\tau^W(\boldsymbol{x}_W) \neq 0$, *we have* $d^{opt,W}(\boldsymbol{x}_W) = d^{opt,B}(\boldsymbol{x}_W^B)$.

(v) *For any* $\boldsymbol{x}_B \in \Omega^B$, *we have* $\tau^W(\boldsymbol{x}_W) \geq 0$ *for all* $\boldsymbol{x}_W \in \Omega^W$ *such that* $\boldsymbol{x}_W^B = x_B$ *or* $\tau^W(\boldsymbol{x}_W) \leq 0$ *for all* $\boldsymbol{x}_W \in \Omega^W$ *such that* $\boldsymbol{x}_W^B = \boldsymbol{x}^B$.

Theorem 6.2.1 provides the sufficient and necessary conditions for CQTE. Results in (iv) and (v) hold for any $\boldsymbol{x}$, instead of almost surely. This is due to the continuity of $\tau^W(\cdot)$ and $f^W(\cdot)$. Result (ii) suggests $\text{VD}^{W,B} > 0$ if H_1 holds. By definition, this means that variables in $\boldsymbol{X}_0^C$ have CQTE given $\boldsymbol{X}_0^B$ if and only if the optimal regime obtained based on $\boldsymbol{X}_0^B$ and $\boldsymbol{X}_0^C$ together can yield a larger value function than that based on $\boldsymbol{X}_0^B$ only.

Result (iii) implies when H_0 holds, we have $\text{ER}^{W,B} = 0$. However, it can not guarantee that $\text{ER}_*^{W,B}$ defined in (6.4) is equal to 0. We provide a counter example below. Let $p = 2$, $B = \{2\}$, $C = \{1\}$ and hence $W = I_0 = \{1,2\}$. Let $\tau(\boldsymbol{x}) = \tau^W(\boldsymbol{x}_W) = [x^{(1)}]_+(x^{(2)} - 1)$, where $[y]_+ = \max(0, y)$ for any $y \in \mathbb{R}$. Apparently, H_0 holds under this setting. When $X^{(1)}$ and $X^{(2)}$ are independent, we obtain $\tau^B(x^{(2)}) = (x^{(2)} - 1)\mathbb{E}[X_0^{(1)}]_+$. Suppose $X_0^{(2)} < 1$ a.s. and $\mathbb{E}[X_0^{(1)}]_+ > 0$. If $\mathbb{P}(X_0^{(1)} \leq 0) > 0$, we have

$$\begin{aligned} \mathbb{P}(X_0^{(1)} \leq 0) &= \mathbb{P}\{\tau_0^W(\boldsymbol{X}_0^W) \geq 0\} = \mathbb{E}d^{opt,W}(\boldsymbol{X}_0^W) \\ &\neq \mathbb{E}d^{opt,B}(X_0^B) = \mathbb{P}\{\tau^B(X_0^{(2)}) \geq 0\} = 0. \end{aligned}$$

Thus, $\text{ER}_*^{W,B} \neq 0$ if $\mathbb{P}(X_0^{(1)} \leq 0) > 0$.

Assertion (iv) motivates us to consider the following test statistic for H_0

$$S^{W,B} = \int_{\boldsymbol{x}_W \in \Omega^W} \phi\{\tau^W(\boldsymbol{x}_W)\}\{d^{opt,W}(\boldsymbol{x}_W) - d^{opt,B}(\boldsymbol{x}_W^B)\}\omega_0(\boldsymbol{x}_W) d\boldsymbol{x}_W,$$

where $\phi(\cdot)$ is a monotonically increasing function with $\phi(0) = 0$ and $\omega_0(\boldsymbol{x}_W)$ is a nonnegative weight function. Obviously we have $S^{W,B} \geq 0$. When H_0 holds, by Theorem 6.2.1, we obtain $S^{W,B} = 0$. Taking ϕ to be the identity function and $\omega_0(\boldsymbol{x}_W) = f^W(\boldsymbol{x}_W)$, we obtain $S^{W,B} = \text{VD}^{W,B}$. When $\omega_0(\boldsymbol{x}_W) = f^W(\boldsymbol{x}_W)/\mathbb{P}\{\tau^W(\boldsymbol{X}_0^W) \neq 0\}$ and $\phi(z) = \text{sgn}(z)$ where

$$\text{sgn}(z) = \begin{cases} 1, & z > 0, \\ 0, & z = 0, \\ -1, & z < 0, \end{cases}$$

we have $S^{W,B} = \text{ER}^{W,B}$. More generally, we can let $\phi(z) = \text{sgn}(z)|z|^q$ for some $q \geq 0$. The defined $S^{W,B}$ then becomes an L_{q+1} type functional. Alternatively, we can consider the following supremum-type test statistic

$$\sup_{\boldsymbol{x}_W \in \Omega^W} \phi\{\tau^W(\boldsymbol{x}_W)\}\{d^{opt,W}(\boldsymbol{x}_W) - d^{opt,B}(\boldsymbol{x}_W^B)\}\omega_0(\boldsymbol{x}_W). \tag{6.5}$$

Shi et al. (2019c) developed a consistent testing procedure based on (6.5). In these statistics, function $\phi\{\tau^W(\boldsymbol{x}_W)\}$ represents the magnitude of treatment effects, while the difference of two indicators characterizes the discrepancy between the regimes $d^{opt,B}$ and $d^{opt,W}$. We formally introduce our test statistic in the next section.

6.3 Testing procedure

We first introduce nonparametric estimators of τ^W and τ^B. In this Section, we assume the propensity score is correctly specified. In the next Section, we propose a doubly robust test, which allows the misspecification of the propensity score. Consider the following nonparametric estimator of $\tau^W(\boldsymbol{x}_W) f^W(\boldsymbol{x}_W)$:

$$\tau_n^W(\boldsymbol{x}_W) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{\pi(1,\boldsymbol{X}_i)} - \frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\right) Y_i K_{h_W}^W(\boldsymbol{x}_W - \boldsymbol{X}_i^W),$$

where $K_{h_W}^W(\cdot)$ is a multivariate kernel function. In general, $K_{h_W}^W(\cdot)$ can be taken as a p_W-variate density function with $p_W = p_B + p_C$ and h_W being a symmetric positive definite matrix as discussed in Wand and Jones (1993). In practice, for simplicity, we may take $K_{h_W}^W(\cdot)$ as a product of component-wise kernel functions, i.e., $K_{h_W}^W(\boldsymbol{x}_W - \boldsymbol{X}_i^W) = \prod_{j \in W}(h_{W,j})^{-1} K(\frac{x_W^{(j)} - X_i^{(j)}}{h_j})$, where $K(\cdot)$ is a symmetric density function. For notational convenience, we set $h_{W,1} = \cdots h_{W,p} = h_W$, and write $K_{h_W}^W(\boldsymbol{x}_W - \boldsymbol{X}_i^W) = (h_W)^{-p_W} K^W(\frac{\boldsymbol{x}_W - \boldsymbol{X}_i^W}{h_W})$. Note that the propensity score is a function of $\boldsymbol{X}_i$ not just X_i^W. Under C1-C3, we can show that $\tau_n^W(\boldsymbol{x}_W)$ is a consistent estimator of $\tau^W(\boldsymbol{x}_W) f^W(\boldsymbol{x}_W)$.

Let $f^B(\cdot)$ denote the density function of $\boldsymbol{X}_0^B$. Similarly, a nonparametric estimator of $\tau^B(\boldsymbol{x}_B) f^B(\boldsymbol{x}_B)$ is given by

$$\tau_n^B(\boldsymbol{x}_B) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{\pi(1,\boldsymbol{X}_i)} - \frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\right) Y_i K_{h_B}^B(\boldsymbol{x}_B - \boldsymbol{X}_i^B).$$

Based on Remark 6.2, it's natural to consider the test statistic based on

$$S_n^{W,B} = \int_{\boldsymbol{x}_W \in \Omega^W} \tau_n^W(\boldsymbol{x}_W)\{d_n^W(\boldsymbol{x}_W) - d_n^B(\boldsymbol{x}_W^B)\} d\boldsymbol{x}_W, \tag{6.6}$$

where $d_n^W(\boldsymbol{x}_W) = \mathbb{I}\{\tau_n^W(\boldsymbol{x}_W) \geq 0\}$ and $d_n^B(\boldsymbol{x}_W^B) = \mathbb{I}\{\tau_n^B(\boldsymbol{x}_W^B) \geq 0\}$, are corresponding estimators for $d^{opt,W}(\boldsymbol{x}_W)$ and $d^{opt,B}(\boldsymbol{x}_W^B)$ respectively.

When some of the covariates are discrete, we need to modify the integral in (6.6) by some product measure of Lebesgue and counting measures. For notational convenience, in Sections 6.3 and 6.4, we assume $\boldsymbol{X}_0^W$ is continuous. In numerical studies, we allow some covariates to be discrete when implementing our test. Details about the test statistic with discrete covariates can be found in Section 6.5.

Under certain regularity conditions, we will show that there exist some positive sequences $\{a_n\}$ and $\{\sigma_n\}$ such that

$$\frac{\sqrt{n} S_n^{W,B} - a_n}{\sigma_n} \xrightarrow{d} N(0,1),$$

under the null. To construct the test statistic, we replace a_n and σ_n by some appropriate estimators $\bar{a}_n$ and $\bar{\sigma}_n$, and reject the null when $T_n^{W,B} = (\sqrt{n} S_n^{W,B} - \bar{a}_n)/\bar{\sigma}_n > z_\alpha$. Below we introduce our test statistic which is a slightly modified version of $T_n^{W,B}$.

6.3.1 Test statistic

Consider the following test statistic

$$\widetilde{S}_n^{W,B} = \int_{\boldsymbol{x}_W \in \Omega^W} \tau_n^W(\boldsymbol{x}_W)\{d_n^W(\boldsymbol{x}_W) - d_n^B(\boldsymbol{x}_W^B)\} \mathbb{I}(\boldsymbol{x}_W \notin \widehat{E}) d\boldsymbol{x}_W, \tag{6.7}$$

where

$$\widehat{E} = \left\{ \boldsymbol{x}_W \in \Omega^W : \left| \frac{\tau_n^W(\boldsymbol{x}_W)}{\widehat{f}^W(x_W)} \right| \leq \eta_n, \left| \frac{\tau_n^B(\boldsymbol{x}_W^B)}{\widehat{f}^B(\boldsymbol{x}_W^B)} \right| \leq \eta_n \right\},$$

for some sequence $\eta_n \to 0$. Here, $\widehat{f}^W$ and $\widehat{f}^B$ are the kernel density estimators of f^W and f^B, respectively. Specifically,

$$\widehat{f}^W(\boldsymbol{x}_W) = \frac{1}{n} \sum_{i=1}^n K_{h_W}^W(\boldsymbol{x}_W - \boldsymbol{X}_i^W) \text{ and } \widehat{f}^B(\boldsymbol{x}_B) = \frac{1}{n} \sum_{i=1}^n K_{h_B}^B(\boldsymbol{x}_B - \boldsymbol{X}_i^B).$$

Estimators $\tau_n^W(\boldsymbol{x}_W)/\widehat{f}^W(\boldsymbol{x}_W)$ and $\tau_n^B(\boldsymbol{X}_0^B)/\widehat{f}^B(\boldsymbol{X}_0^B)$ are referred to as the Nadaraya-Watson estimators for $\tau^W(\boldsymbol{x}_W)$ and $\tau^B(\boldsymbol{X}_0^B)$.

Similar to $S_n^{W,B}$, we can show $(\sqrt{n}\widetilde{S}_n^{W,B}-\widetilde{a}_n)/\widetilde{\sigma}_n \xrightarrow{d} N(0,1)$, for some $\widetilde{a}_n$ and $\widetilde{\sigma}_n$. The tests based on $S_n^{W,B}$ and $\widetilde{S}_n^{W,B}$ have nontrivial power against certain local alternative as defined later. However, the one based on $\widetilde{S}_n^{W,B}$ is more powerful. To see this, note that

$$\begin{aligned} & \sqrt{n}(S_n^{W,B}-\widetilde{S}_n^{W,B}) \\ = \; & \sqrt{n}\int_{\boldsymbol{x}_W\in\Omega^W}\tau_n^W(\boldsymbol{x}_W)\{d_n^W(\boldsymbol{x}_W)-d_n^B(\boldsymbol{x}_W^B)\}\mathbb{I}(\boldsymbol{x}_W\in\widehat{E})d\boldsymbol{x}_W. \end{aligned} \tag{6.8}$$

With proper choice of η_n, the right-hand side (RHS) of (6.8) is equivalent to

$$\sqrt{n}\int_{\boldsymbol{x}_W\in\Omega^W}\tau_n^W(\boldsymbol{x}_W)\{d_n^W(\boldsymbol{x}_W)-d_n^B(\boldsymbol{x}_W^B)\}\mathbb{I}(\boldsymbol{x}_W\in E_0)d\boldsymbol{x}_W, \tag{6.9}$$

where $E_0=\{\boldsymbol{x}_W:\tau^W(\boldsymbol{x}_W)=0,\tau^B(\boldsymbol{x}_W^B)=0\}$.

The asymptotic mean of (6.9) remains the same under the null and local alternative. However, it has non-degenerate variance and is asymptotically independent of $\widetilde{S}_n^{W,B}$. This implies that $\sqrt{n}S_n^{W,B}-a_n$ and $\sqrt{n}\widetilde{S}_n^{W,B}-\widetilde{a}_n$ have the same shifted mean under the local alternative, but the variance of $\widetilde{S}_n^{W,B}$ is smaller than $S_n^{W,B}$ when the set E_0 has nonzero measure. From now on, we focus on the test statistic $\widetilde{S}_n^{W,B}$.

6.3.2 Consistency of the test

Define

$$\mu^W(\boldsymbol{x}_W)=\mathbb{E}\left[\left\{\frac{A_0}{\pi(1,\boldsymbol{X}_0)}-\frac{1-A_0}{\pi(0,\boldsymbol{X}_0)}\right\}^2 Y_0^2|\boldsymbol{X}_0^W=\boldsymbol{x}_W\right]f^W(\boldsymbol{x}_W)K_*^W(0),$$

where

$$K_*^W(\boldsymbol{t})=\int_{\boldsymbol{x}_W}K^W(\boldsymbol{x}_W)K^W(\boldsymbol{x}_W+\boldsymbol{t})d\boldsymbol{x}_W.$$

For each fixed $\boldsymbol{x}_W$, $\mu^W(\boldsymbol{x}_W)$ is the asymptotic variance of $\sqrt{n(h_W)^{p_W}}\tau_n^W(\boldsymbol{x}_W)$.

Define $F_0=\{\boldsymbol{x}_W\in\Omega^W:\tau^W(\boldsymbol{x}_W)=0,\tau^B(\boldsymbol{x}_W^B)\neq 0\}$. The asymptotic mean and variance of

$\sqrt{n}\widetilde{S}_n^{W,B}$ are given by

$$\begin{aligned}
\widetilde{a}_n &= \frac{1}{\sqrt{2\pi(h_W)^{p_W}}}\int_{\boldsymbol{x}_W\in F_0}\sqrt{\mu_n^W(\boldsymbol{x}_W)}d\boldsymbol{x}_W,\\
\widetilde{\sigma}^2 &= \int_{\substack{\boldsymbol{x}_W\in F_0\\ \boldsymbol{t}\in[-1,1]^{p_W}}}\mu^W(\boldsymbol{x}_W)\mathbb{COV}(\max\{\sqrt{1-\rho^2(\boldsymbol{t})}\mathbb{Z}_1+\rho(\boldsymbol{t})\mathbb{Z}_2,0\},\max\{\mathbb{Z}_2,0\})d\boldsymbol{x}_W d\boldsymbol{t},
\end{aligned}$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are independent standard normal random variables, $\rho(\boldsymbol{t})=K_*^W(\boldsymbol{t})/K_*^W(0)$, and

$$\mu_n^W(\boldsymbol{x}_W)=\frac{1}{(h_W)^{p_W}K_*^W(0)}\mathbb{E}\left[\frac{\mu^W(\boldsymbol{X}_0^W)}{f^W(\boldsymbol{X}_0^W)}\left\{K^W\left(\frac{\boldsymbol{x}_W-\boldsymbol{X}_0^W}{h_W}\right)\right\}^2\right].$$

To estimate $\widetilde{a}_n$ and $\widetilde{\sigma}^2$, we first provide nonparametric estimators for $\mu^W(\boldsymbol{x}_W)$ and F_0. Define

$$\begin{aligned}
&\widehat{\mu}_n^W(\boldsymbol{x}_W)=\frac{1}{n(h_W)^{p_W}}\sum_{i=1}^n\left\{\left(\frac{A_i}{\pi(1,\boldsymbol{X}_i)}-\frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\right)Y_i\right\}^2\left\{K^W\left(\frac{\boldsymbol{x}_W-\boldsymbol{X}_i^W}{h_W}\right)\right\}^2,\\
&\widehat{F}=\{\boldsymbol{x}_W\in\Omega^W:|\tau_n^W(\boldsymbol{x}_W)/\widehat{f}^W(\boldsymbol{x}_W)|\le\eta_n,|\tau_n^B(\boldsymbol{x}_W^B)/\widehat{f}^B(\boldsymbol{x}_W^B)|>\eta_n\},
\end{aligned}$$

where η_n is defined in (6.7). For any set $F\subseteq\Omega$, define $\widehat{a}_n(F)$ and $\widehat{\sigma}_n^2(F)$ as

$$\begin{aligned}
\widehat{a}_n(F) &= \frac{1}{\sqrt{2\pi(h_W)^{p_W}}}\int_{\boldsymbol{x}_W\in F}\sqrt{\widehat{\mu}_n^W(\boldsymbol{x}_W)}d\boldsymbol{x}_W,\\
\widehat{\sigma}_n^2(F) &= \int_{\substack{\boldsymbol{x}_W\in F\\ \boldsymbol{t}\in[-1,1]^{p_W}}}\widehat{\mu}_n^W(\boldsymbol{x}_W)\mathbb{COV}(\max\{\sqrt{1-\rho^2(\boldsymbol{t})}\mathbb{Z}_1+\rho(\boldsymbol{t})\mathbb{Z}_2,0\},\max\{\mathbb{Z}_2,0\})\\
&\quad\times d\boldsymbol{x}_W d\boldsymbol{t}.
\end{aligned}$$

We estimate $\widetilde{a}_n$ and $\widetilde{\sigma}_n^2$ by $\widehat{a}_n(\widehat{F})$ and $\widehat{\sigma}_n^2(\widehat{F})$, respectively.

Let $\nu(\cdot)$ be the Lebesgue measure. Define the test statistic

$$\widetilde{T}_n^{W,B}=\begin{cases}\{\sqrt{n}\widetilde{S}_n^{W,B}-\widehat{a}_n(\widehat{F})\}/\widehat{\sigma}_n(\widehat{F}), & \text{if } \nu(\widehat{F})\neq 0,\\ \{\sqrt{n}\widetilde{S}_n^{W,B}-\widehat{a}_n(\Omega^W)\}/\widehat{\sigma}_n(\Omega^W), & \text{otherwise.}\end{cases}$$

We reject the null when $\widetilde{T}_n^{W,B}>z_\alpha$.

When $\nu(\widehat{F})=0$, $\widehat{\sigma}_n(\widehat{F})=0$ and hence the test statistic $\{\sqrt{n}\widetilde{S}_n^{W,B}-\widehat{a}_n(\widehat{F})\}/\widehat{\sigma}_n(\widehat{F})$ is not well defined. Therefore, in this case we consider $\{\sqrt{n}\widetilde{S}_n^{W,B}-\widehat{a}_n(\Omega^W)\}/\widehat{\sigma}_n(\Omega^W)$ instead. When

F_0 is a strict subset of Ω, the test statistic based on $\{\sqrt{n}\widetilde{S}_n^{W,B} - \widehat{a}_n(\Omega^W)\}/\widehat{\sigma}_n(\Omega^W)$ will be conservative. To study the theoretical properties of the test, we first introduce some conditions.

(D1.) Assume that Ω^W is a bounded subset in $\mathbb{R}^{p_W}$. Assume f^W is continuous and satisfies $\inf_{\boldsymbol{x}_W \in \Omega^W} f^W(\boldsymbol{x}_W) > 0$, and $\sup_{\boldsymbol{x}_W \in \Omega^W} f^W(\boldsymbol{x}_W) < \infty$. Assume τ^W and τ^B are continuous. Moreover, f^W, τ^W, f^B and τ^B are s-times differentiable almost everywhere with uniformly bounded derivatives, for some integer $s > 0$.

(D2.) Assume $K^W(x_W) = \prod_{j=1}^{p_W} K_j(x_W^{(j)})$, and $K^B(x_B) = \prod_{j=1}^{p_B} K_{j+p_W}(x_B^{(j)})$, where each K_j is an s-order kernel function with support $\{\mu \in \mathbb{R} : |\mu| \le 1/2\}$ and bounded, and is of bounded variation and integrates to 1.

(D3.) Assume $\mathbb{E}\exp(t|Y_0|) < \infty$ for some $t > 0$, and $\sup_{\boldsymbol{x}_W \in \Omega^W} \mathbb{E}(Y_0^4 | \boldsymbol{X}_0^W = \boldsymbol{x}_W, A_0 = a) < \infty$ for $a = 0, 1$.

(D4.) Assume there exist some constants c_0 and c_1 that $0 < c_0 \le \pi(1, \boldsymbol{x}) \le c_1 < 1, \forall \boldsymbol{x}$.

(D5.) Assume that $\mu^W(\boldsymbol{x}_W)$ is uniformly continuous and bounded on Ω^W, and that $\inf_{\boldsymbol{x}_W \in \Omega^W} \mu^W(\boldsymbol{x}_W) > 0$.

(D6.) Assume $n(h_W)^{2p_W}/\log n \to \infty$, $n(h_W)^{2s} \to 0$, $h_B^{p_B} \asymp h_W^{p_W}$.

(D7.) Assume $\nu(\partial F_0) = 0$, $\nu(\Omega^W \cap F_0^c) > 0$. Assume there exist some constants ξ_0, $\bar{c}_0 > 0$ such that for any sufficiently small t, $\varepsilon > 0$,

$$\nu(\{\boldsymbol{x}_W : 0 < |\tau^W(\boldsymbol{x}_W)| \le t\}) = O(t^{\xi_0}), \nu(\{\boldsymbol{x}_B : 0 < |\tau^B(\boldsymbol{x}_B)| \le t\}) = O(t^{\xi_0}),$$
$$\nu(\{\boldsymbol{x}_W : 0 < |\tau^W(\boldsymbol{x}_W)| \le t, |\tau^B(\boldsymbol{x}_W^B)| > (1+\varepsilon)t\}) \ge \bar{c}_0 t^{\xi_0}.$$

(D8.) Assume η_n satisfies $\eta_n^{2\xi_0} \gg \log^{\xi_0+1} n / \{n(h_W)^{p_W}\}^{\xi_0}$ and $n\eta_n^{2\xi_0+2} \to 0$.

Condition D1 requires Ω^W to be bounded. In practice, if it's unbounded, we can perform monotone transformations on each component of X to make the support of the transformed variables bounded. Otherwise, we need to focus on a bounded subset $\Omega_0^W = \Omega_0^B \times \Omega_0^C \subseteq \Omega^W$, and write $\widetilde{S}_n^{W,B}$ as

$$\int_{\boldsymbol{x}_W \in \Omega_0^W} \tau_n^W(\boldsymbol{x}_W)\{d_n^W(\boldsymbol{x}_W) - d_n^B(\boldsymbol{x}_W^B)\}\mathbb{I}(\boldsymbol{x}_W \notin \widehat{E}) d\boldsymbol{x}_W.$$

In addition, we modify H_0 as "For any fixed $\boldsymbol{x}_B \in \Omega_0^B$, $\tau(\boldsymbol{x}_B, \boldsymbol{x}_C) \ge 0, \forall \boldsymbol{x}_C \in \Omega_0^C$, or $\tau(\boldsymbol{x}_B, \boldsymbol{x}_C) \ge 0, \forall \boldsymbol{x}_C \in \Omega_0^C$."

Condition D2 requires each K_j to be of order s. The order of the kernel is defined as the first nonzero moment. Condition D6 requires $nh^2 \to \infty$ and $nh^{2s/p_W} \to 0$. This implies $s > p_W$. When $p_W > 2$, this condition requires each kernel K_j to be of high orders.

Such kernels are typically referred to as the bias-reducing kernels. Unlike standard kernel functions, these kernels allow $K_j(z)$ to be negative for some $z \in \mathbb{R}$. Moreover, we assume $h_W^{p_W} \asymp h_B^{p_B}$ in D6. This guarantees $\tau_n^W(\boldsymbol{x}_W)$ and $\tau_n^B(\boldsymbol{x}_W^B)$ converge at the same rate.

Conditions D7 is not restrictive. Obviously, this condition holds when $\inf_{\boldsymbol{x}_W \in \Omega^W} |\tau^W(\boldsymbol{x}_W)| > 0$. In that case, we can set the constants ξ_0 and $\bar{c}_0$ to be any positive constants. Moreover, these conditions are satisfied in many other cases. For example, let $p = 2$, $B = \{2\}$, $C = \{1\}$. Consider

$$\tau(x^{(1)}, x^{(2)}) = \begin{cases} -(x^{(1)} + x^{(2)})^{1/\xi_0}, & \text{if } x^{(1)}, x^{(2)} > 0, \\ -(x^{(1)})^{1/\xi_0}, & \text{if } x^{(1)} > 0, x^{(2)} \le 0, \\ -(x^{(2)})^{1/\xi_0}, & \text{if } x^{(1)} \le 0, x^{(2)} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then, with some calculation, we can show

$$\begin{aligned} & \nu(\{x : 0 < |\tau(x^{(1)}, x^{(2)})| \le t\}) = \nu(\{x : x^{(1)}, x^{(2)} > 0, x^{(1)} + x^{(2)} \le t\}) \\ + \quad & \nu(\{x : 0 < x^{(1)} \le t, x^{(2)} \le 0\}) + \nu(\{x : x^{(1)} \le 0, 0 < x^{(2)} \le t\}) = c_1 t^{2\xi_0} + c_2 t^{\xi_0}, \end{aligned}$$

for some constants c_1, $c_2 > 0$.

Note that $|\tau^{\{2\}}(x^{(2)})| \ge \min(x^{(2)})^{1/\xi_0}$ when $x^{(2)} > 0$, and $\tau^{\{2\}}(x^{(2)})$ is a nonzero constant $c_3 < 0$ for all $x^{(2)} \le 0$. For sufficiently small $t > 0$, we obtain

$$\nu\big(\{x : 0 < |\tau^{\{2\}}(x^{(2)})| \le t\}\big) \le \nu\big(\{x : (x^{(2)})^{1/\xi_0} \le t, x^{(2)} > 0\}\big) = O(t^{\xi_0}).$$

Besides, for any small $\varepsilon_0 > 0$, we have

$$\begin{aligned} & \nu\big(\{x : 0 < |\tau(x^{(1)}, x^{(2)})| \le t, |\tau^{\{2\}}(x^{(2)})| > (1 + \varepsilon_0)t\}\big) \\ \ge \quad & \nu\big(\{x : 0 < |\tau(x^{(1)}, x^{(2)})| \le t, \tau^{\{2\}}(x^{(2)}) = -c_3\}\big) \\ = \quad & \nu\big(\{x : 0 < (x^{(1)})^{1/\xi_0} \le t, x^{(2)} \le 0\}\big) = c_4 t^{\xi_0}, \end{aligned}$$

for some constant $c_4 > 0$. This verifies D7.

Theorem 6.3.1 *Assume Conditions D1-D8 hold. Then, under H_0, we have*

$$\lim_n \mathbb{P}(\widetilde{T}_n^{W,B} > z_\alpha) \le \alpha,$$

for $0 < \alpha \le 0.5$, where the equality holds when $\nu(F_0) > 0$.

Theorem 6.3.1 shows $\widetilde{T}_n^{W,B}$ has correct size under H_0. When $\nu(F_0)=0$, we can show with probability tending to 1, $\sqrt{n}\widetilde{S}_n^{W,B} \le \widehat{a}_n$, and hence

$$\lim_n \mathbb{P}(\widetilde{T}_n^{W,B} > z_\alpha) = 0.$$

When $\nu(F_0) \neq 0$, we will show that $\widetilde{T}_n^{W,B}$ is asymptotically normal. The proof is based on the well-known Poissoinization technique which introduces a Poissonized version of $\widetilde{S}_n^{W,B}$ and transforms the integral into summation of mean zero 1-dependent random fields (see for example Giné et al. 2003; Mason and Polonik 2009; Chang et al. 2015). The asymptotic normality thus follows by standard central limit theorem for m-dependent random fields (Shergin 1990). The details are given in Shi et al. (2019c).

Theorem 6.3.2 *Assume Conditions D1-D8 hold. Then, under H_1, we have*

$$\lim_n \mathbb{P}(\widetilde{T}_n^{W,B} > z_\alpha) \to 1.$$

Theorem 6.3.2 shows $\widetilde{T}_n^{W,B}$ having power going to 1 against fixed alternatives. Together with Theorem 6.3.1, Theorem 6.3.2 suggests that our testing procedure is consistent.

6.3.3 Local alternatives

In this subsection, we investigate the power of the proposed test under local alternatives. We write $\tau_n(\boldsymbol{x})$ as the contrast function and $\tau_n^D(\boldsymbol{x}_D) = \mathbb{E}\{\tau_n(\boldsymbol{X}_0)|\boldsymbol{X}_0^D = \boldsymbol{x}_D\}$ for a given subset $D \subseteq I_0$, with the intention that these functions are allowed to vary with n. Consider the following sequence of local alternatives:

$$H_a : \tau_n^W(\boldsymbol{x}_W) = \tau^W(\boldsymbol{x}_W) + n^{-1/2}\boldsymbol{\delta}_0^W(\boldsymbol{x}_W),$$

for some continuous functions τ^W and $\boldsymbol{\delta}_0^W$ on Ω^W, where for any fixed $\boldsymbol{x}_B \in \Omega^B$,

$$\tau^W(\boldsymbol{x}_B, \boldsymbol{x}_C) \le 0 \text{ for any } \boldsymbol{x}_C \in \Omega^C, \text{ or } \tau^W(\boldsymbol{x}_B, \boldsymbol{x}_C) \ge 0 \text{ for any } \boldsymbol{x}_C \in \Omega^C,$$

and

$$\boldsymbol{\delta}_0^W(\boldsymbol{x}_B, \boldsymbol{x}_C) \le 0 \text{ for any } \boldsymbol{x}_C \in \Omega^C, \text{ or } \boldsymbol{\delta}_0^W(\boldsymbol{x}_B, \boldsymbol{x}_C) \ge 0 \text{ for any } \boldsymbol{x}_C \in \Omega^C.$$

In addition,

$$\delta_0^W(\boldsymbol{x}_B,\boldsymbol{x}_C)\tau^W(\boldsymbol{x}_B,\boldsymbol{x}_C) \le 0, \quad \forall \boldsymbol{x}_B \in \Omega^B, \boldsymbol{x}_C \in \Omega^C.$$

Recall that $F_0 = \{\boldsymbol{x}_W \in \Omega^W : \tau^W(\boldsymbol{x}_W) = 0, \tau^B(\boldsymbol{x}_W^B) \neq 0\}$. Let $\dot{F}_0$ and ∂F_0 denote its interior and boundary, respectively. Since the contrast function $\tau_{n,0}^W$ varies with n, we state a more precise definition of conditional qualitative treatment effects below.

Definition 6.3.1 (CQTE, continued) *Variables in C have qualitative treatment effects conditional on variables in B if there exists some nonempty sets $\mathscr{C}_1$, $\mathscr{C}_2 \in \mathbb{R}^{p_C}$, $\mathscr{B} \in \mathbb{R}^{p_B}$, such that (i)*

$$\mathbb{P}\left\{(\boldsymbol{X}_0^B, \boldsymbol{X}_0^C)^T \in \mathscr{B} \times \mathscr{C}_1\right\} > 0, \ \textit{and}\ \mathbb{P}\left\{(\boldsymbol{X}_0^B, \boldsymbol{X}_0^C)^T \in \mathscr{B} \times \mathscr{C}_2\right\} > 0;$$

and (ii) for any $\boldsymbol{x}_{C_1} \in \mathscr{C}_1$, $\boldsymbol{x}_{C_2} \in \mathscr{C}_2$ and $\boldsymbol{x}_B \in \mathscr{B}$, there exists a sequence $n_k \to \infty$ as $k \to \infty$, such that

$$\arg\max_{a=0,1}\left\{a\tau_{n_k,0}^W(\boldsymbol{x}_B,\boldsymbol{x}_{C_1})\right\} \neq \arg\max_{a=0,1}\left\{a\tau_{n_k,0}^W(\boldsymbol{x}_B,\boldsymbol{x}_{C_2})\right\}. \tag{6.10}$$

It's immediate to see that (6.10) is a modified version of (6.1) where we allow the conditional expectation $\mathbb{E}\{Y_0^*(a)|\boldsymbol{X}_0^B, \boldsymbol{X}_0^C\}$ to vary with n. By the definition of τ^W and δ_0^W, we can see that the magnitude of δ_0^W affects CQTE. We provide a theorem which formally characterizes such results below.

Theorem 6.3.3 *Assume δ_0^W is continuous and bounded on Ω^W. Assume $\nu(\partial F_0) = 0$. Under conditions in Theorem 6.2.1, the following statements are equivalent:*

(i) $\boldsymbol{X}_0^C$ *doesn't have QTE conditional on* $\boldsymbol{X}_0^B$.

(ii) For any $\varepsilon > 0$, there exist a set N_ε and a positive integer n_ε such that $\nu(N_\varepsilon) \le \varepsilon$, and for all $n \ge n_\varepsilon$, the following holds: for any fixed $\boldsymbol{x}_B$, we have $\tau_n(\boldsymbol{x}_W) \ge 0$ for any $\boldsymbol{x}_W \notin N_\varepsilon$ such that $\boldsymbol{x}_W^B = \boldsymbol{x}_B$ or $\tau_n(\boldsymbol{x}_W) \le 0$ for any $\boldsymbol{x}_W \notin N_\varepsilon$ such that $\boldsymbol{x}_W^B = \boldsymbol{x}_B$.

(iii) For all $\boldsymbol{x}_W \in \dot{F}_0$, $\delta_0^W(\boldsymbol{x}_W) = 0$.

(iv) $\displaystyle\int_{\boldsymbol{x}_W \in F_0} |\delta_0^W(\boldsymbol{x}_W)| f^W(\boldsymbol{x}_W) d\boldsymbol{x}_W = 0.$

Result (iv) implies H_0 holds when $\mathbb{P}(\boldsymbol{X}_0^W \in F_0) = 0$, or $\mathbb{P}\{\tau_0^W(\boldsymbol{X}_0^W) = 0\} = 0$. This implies that the local alternatives are nonstandard and only exist in the nonregular cases, i.e., there is a positive probability such that the optimal treatment decision based on $\boldsymbol{X}_0^W$ is not defined.

Theorem 6.3.3 suggests the quantity

$$\int_{\boldsymbol{x}_W \in F_0} |\delta_0^W(\boldsymbol{x}_W)| f^W(\boldsymbol{x}_W) d\boldsymbol{x}_W$$

plays a role in determining CQTE of $\boldsymbol{X}_0^C$ conditional on $\boldsymbol{X}_0^B$. In the theorem below, we establish the power of our test statistic $\widetilde{T}_n^{W,B}$ under the local alternatives. It can be seen that this quantity is closely related to the power of our test.

Theorem 6.3.4 *Assume Conditions D1-D8 hold and that δ_0^W is bounded on Ω^W. Then, under H_a with $\int_{\boldsymbol{x}_W \in F_0} |\delta_0^W(\boldsymbol{x}_W)| f^W(\boldsymbol{x}_W) d\boldsymbol{x}_W > 0$, we have*

$$\lim_n \mathbb{P}(\widetilde{T}_n^{W,B} > z_\alpha) = 1 - \Phi\left(z_\alpha - \frac{1}{2\widetilde{\sigma}} \int_{\boldsymbol{x}_W \in F_0} |\delta_0^W(\boldsymbol{x}_W)| f^W(\boldsymbol{x}_W) d\boldsymbol{x}_W \right).$$

6.4 Doubly robust test statistic

In an observational study, the propensity score $\pi(1,\cdot)$ is usually unknown. In practice, we posit a parametric model indexed by $\boldsymbol{\alpha}$ for the propensity score, for example, a logistic regression model $\pi(1,\boldsymbol{x}) = \pi_1(\boldsymbol{x},\boldsymbol{\alpha}) = \exp(\boldsymbol{x}^T\boldsymbol{\alpha})/\{1+\exp(\boldsymbol{x}^T\boldsymbol{\alpha})\}$. We can obtain an estimator $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ based on data $\{(A_i, X_i), i = 1,\ldots,n\}$, by either maximizing the likelihood function or solving estimating equations. The estimator $\widehat{\boldsymbol{\alpha}}$ will converge to some population-level parameters $\boldsymbol{\alpha}_0$. When the propensity score model is correctly specified, $\boldsymbol{\alpha}_0$ is the true parameter in the model. When the model is wrong, $\boldsymbol{\alpha}_0$ corresponds to some least false parameters that have been widely studied in the literature (see for example, White 1982; Li and Duan 1989).

We also posit some parametric models $\Phi_0(\boldsymbol{x},\boldsymbol{\theta})$ and $\boldsymbol{\Phi}_1(\boldsymbol{x},\boldsymbol{\zeta})$ for $h(0,\boldsymbol{x})$ and $h(1,\boldsymbol{x})$, respectively. Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\zeta}}$ denote the estimator of $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, respectively, which converge to some parameters $\boldsymbol{\theta}^*$ and $\boldsymbol{\zeta}^*$, under potential model misspecification. Let $\widehat{\pi}_{1,i} = \pi_1(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})$, $\widehat{\pi}_{0,i} = 1-\pi_1(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})$, $\widehat{\Phi}_{0,i} = \Phi_0(\boldsymbol{X}_i,\widehat{\boldsymbol{\theta}})$ and $\widehat{\Phi}_{1,i} = \Phi_1(\boldsymbol{X}_i,\widehat{\boldsymbol{\zeta}})$. Define the following doubly robust estimators for $\tau_n^W(\boldsymbol{x}_W)f^W(\boldsymbol{x}_W)$ and $\tau_n^B(\boldsymbol{x}_B)f^B(\boldsymbol{x}_B)$:

$$\begin{aligned}\tau_{n,DR}^W(\boldsymbol{x}_W) &= \frac{1}{n}\sum_{i=1}^n\Bigg[\left\{\frac{A_i}{\widehat{\pi}_{1,i}}Y_i-\left(\frac{A_i}{\widehat{\pi}_{1,i}}-1\right)\widehat{\Phi}_{1,i}\right\}\\ &\quad -\left\{\frac{1-A_i}{\widehat{\pi}_{0,i}}Y_i-\left(\frac{1-A_i}{\widehat{\pi}_{0,i}}-1\right)\widehat{\Phi}_{0,i}\right\}\Bigg]K_{h_W}^W(\boldsymbol{x}_W-\boldsymbol{X}_i^W),\\ \tau_{n,DR}^B(\boldsymbol{x}_B) &= \frac{1}{n}\sum_{i=1}^n\Bigg[\left\{\frac{A_i}{\widehat{\pi}_{1,i}}Y_i-\left(\frac{A_i}{\widehat{\pi}_{1,i}}-1\right)\widehat{\Phi}_{1,i}\right\}\\ &\quad -\left\{\frac{1-A_i}{\widehat{\pi}_{0,i}}Y_i-\left(\frac{1-A_i}{\widehat{\pi}_{0,i}}-1\right)\widehat{\Phi}_{0,i}\right\}\Bigg]K_{h_B}^B(\boldsymbol{x}_B-\boldsymbol{X}_i^B).\end{aligned}$$

We can show that estimators $\tau_{n,DR}^W(\boldsymbol{x}_W)$ and $\tau_{n,DR}^B(\boldsymbol{x}_B)$ are consistent when either $\pi_1(\boldsymbol{x},\boldsymbol{\alpha})$ or $\Phi_0(\boldsymbol{x},\boldsymbol{\theta})$ and $\Phi_1(\boldsymbol{x},\boldsymbol{\zeta})$ are correctly specified. Let $d_{n,DR}^W(\boldsymbol{x}_W)=\mathbb{I}\{\tau_{n,DR}^W(\boldsymbol{x}_W)\geq 0\}$ and $d_{n,DR}^B(\boldsymbol{x}_B)=\mathbb{I}\{\tau_{n,DR}^B(\boldsymbol{x}_B)\geq 0\}$. Consider

$$\widetilde{S}_{n,DR}^{W,B}=\int_{\boldsymbol{x}_W\in\Omega^W}\tau_{n,DR}^W(\boldsymbol{x}_W)\{d_{n,DR}^W(x_W)-d_{n,DR}^B(\boldsymbol{x}_W^B)\}\mathbb{I}(\boldsymbol{x}_W\notin\widehat{E}_{DR})d\boldsymbol{x}_W,$$

where

$$\widehat{E}_{DR}=\left\{\boldsymbol{x}_W\in\Omega^W:\left|\frac{\tau_{n,DR}^W(\boldsymbol{x}_W)}{\widehat{f}^W(\boldsymbol{x}_W)}\right|\leq\eta_n,\left|\frac{\tau_{n,DR}^B(\boldsymbol{x}_W^B)}{\widehat{f}^B(\boldsymbol{x}_W^B)}\right|\leq\eta_n\right\}.$$

For any set F, define

$$\begin{aligned}\widehat{a}_{n,DR}(F) &= \frac{1}{\sqrt{2\pi(h_W)^{p_W}}}\int_{\boldsymbol{x}_W\in F_0}\sqrt{\widehat{\mu}_{n,DR}^W(\boldsymbol{x}_W)}d\boldsymbol{x}_W,\\ \widehat{\boldsymbol{\sigma}}_{n,DR}^2(F) &= \int_{\boldsymbol{x}_W\in F}\int_{\boldsymbol{t}\in[-1,1]^{p_W}}\widehat{\mu}_{n,DR}^W(\boldsymbol{x}_W)\mathbb{COV}(\max\{\sqrt{1-\rho^2(\boldsymbol{t})}\mathbb{Z}_1+\rho(\boldsymbol{t})\mathbb{Z}_2,0\},\\ &\quad \max\{\mathbb{Z}_2,0\})d\boldsymbol{x}_W d\boldsymbol{t},\end{aligned}$$

where

$$\begin{aligned}\widehat{\mu}_{n,DR}^W(\boldsymbol{x}_W) &= \frac{1}{n(h_W)^{p_W}}\sum_{i=1}^n\Bigg[\left\{\frac{A_i}{\widehat{\pi}_{1,i}}Y_i-\left(\frac{A_i}{\widehat{\pi}_{1,i}}-1\right)\widehat{\Phi}_{1,i}\right\}\\ &- \left\{\frac{1-A_i}{\widehat{\pi}_{0,i}}Y_i-\left(\frac{1-A_i}{\widehat{\pi}_{0,i}}-1\right)\widehat{\Phi}_{0,i}\right\}\Bigg]^2\left\{K^W\left(\frac{\boldsymbol{x}_W-\boldsymbol{X}_i^W}{h_W}\right)\right\}^2.\end{aligned}$$

We estimate the asymptotic mean and variance of $\sqrt{n}\widetilde{S}_{n,DR}^{W,B}$ by $\widehat{a}_{n,DR}(\widehat{F}_{DR})$ and $\widehat{\sigma}^2_{n,DR}(\widehat{F}_{DR})$, respectively, with

$$\widehat{F}_{DR} = \{\boldsymbol{x}_W \in \Omega^W : |\tau^W_{n,DR}(\boldsymbol{x}_W)/\widehat{f}^W(\boldsymbol{x}_W)| \le \eta_n, |\tau^B_{n,DR}(\boldsymbol{x}^B_W)/\widehat{f}^B(\boldsymbol{x}^B_W)| > \eta_n\}.$$

Define

$$\widetilde{T}^{W,B}_{n,DR} = \begin{cases} \{\sqrt{n}\widetilde{S}^{W,B}_{n,DR} - \widehat{a}_{n,DR}(\widehat{F}_{DR})\}/\widehat{\sigma}_{n,DR}(\widehat{F}_{DR}), & \text{if } \nu(\widehat{F}_{DR}) = 0, \\ \{\sqrt{n}\widetilde{S}^{W,B}_{n,DR} - \widehat{a}_{n,DR}(\Omega^W)\}/\widehat{\sigma}_{n,DR}(\Omega^W), & \text{otherwise.} \end{cases}$$

We reject the null when $\widetilde{T}^{W,B}_{n,DR} > z_\alpha$.

To establish the asymptotic distributions of $\widetilde{T}^{W,B}_{n,DR}$ under the null and local alternative, we impose the following conditions.

(D4'.) Assume there exist some constants c_0' and c_1' such that $0 < c_0' \le \pi_1(x, \boldsymbol{\alpha}^*) \le c_1' < 1$ for all $\boldsymbol{x} \in \Omega$.

(D5'.) Assume that $\mu^W_{DR}(\boldsymbol{x}_W)$ is uniformly continuous and bounded on Ω^W, and that $\inf_{\boldsymbol{x}_W \in \Omega^W} \mu^W_{DR}(\boldsymbol{x}_W) > 0$, where

$$\begin{aligned} \mu^W_{DR}(\boldsymbol{x}_W) &= \mathbb{E}\Bigg[\Bigg\{\left(\frac{A_0}{\pi_1(\boldsymbol{X}_0, \boldsymbol{\alpha}^*)} - \frac{1-A_0}{\pi_0(\boldsymbol{X}_0, \boldsymbol{\alpha}^*)}\right) Y - \left(\frac{A}{\pi_1(\boldsymbol{X}_0, \boldsymbol{\alpha}^*)} - 1\right)\Phi_1(\boldsymbol{X}_0, \boldsymbol{\theta}^*) \\ &\quad + \left(\frac{1-A_0}{\pi_0(\boldsymbol{X}_0, \boldsymbol{\alpha}^*)} - 1\right)\Phi_0(\boldsymbol{X}_0, \zeta_0)\Bigg\}^2 |\boldsymbol{X}^W_0 = \boldsymbol{x}_W\Bigg] f^W(\boldsymbol{x}_W) K^W_*(0). \end{aligned}$$

(D9.) Assume that $\pi_1(\boldsymbol{x}, \boldsymbol{\alpha})$ is twice continuously differentiable with respect to $\boldsymbol{\alpha}$; $\|\partial \pi_1(\boldsymbol{x}, \boldsymbol{\alpha}^*)/\partial \boldsymbol{\alpha}\|_2$ is uniformly bounded for all $\boldsymbol{x} \in \Omega$; and the elements in the matrix $\partial^2 \pi_1(\boldsymbol{x}, \boldsymbol{\alpha}^*)/\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T$ are uniformly bounded for all $\boldsymbol{x} \in \Omega$ and $\boldsymbol{\alpha}$ in a small neighborhood of $\boldsymbol{\alpha}^*$.

(D10.) Assume that $\Phi_0(\boldsymbol{x}, \boldsymbol{\theta})$ and $\Phi_1(\boldsymbol{x}, \boldsymbol{\zeta})$ are twice continuously differentiable with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, respectively; $\Phi_0(\boldsymbol{x}, \boldsymbol{\theta}^*)$, $\Phi_1(\boldsymbol{x}, \boldsymbol{\zeta}^*)$, $\|\partial \Phi_0(\boldsymbol{x}, \boldsymbol{\theta}^*)/\partial \boldsymbol{\theta}\|_2$ and $\|\partial \Phi_1(\boldsymbol{x}, \boldsymbol{\zeta}^*)/\partial \boldsymbol{\zeta}\|_2$ are uniformly bounded for all $\boldsymbol{x} \in \Omega$; and the elements in the matrices $\partial^2 \Phi_0(\boldsymbol{x}, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T$ and $\partial^2 \Phi_1(\boldsymbol{x}, \boldsymbol{\zeta})/\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}^T$ are uniformly bounded for all $\boldsymbol{x} \in \Omega$ and $\boldsymbol{\theta}$, $\boldsymbol{\zeta}$ in small neighborhoods of $\boldsymbol{\theta}^*$ and $\boldsymbol{\zeta}^*$, respectively.

(D11.) Assume that the estimators $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\zeta}}$ have the following linear representations

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^* = \frac{1}{n}\sum_i \xi_1(O_i) + o_p\left(\frac{1}{\sqrt{n}}\right),$$
$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \frac{1}{n}\sum_i \xi_2(O_i) + o_p\left(\frac{1}{\sqrt{n}}\right),$$
$$\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}^* = \frac{1}{n}\sum_i \xi_3(O_i) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

for some functions ξ_1, ξ_2 and ξ_3 with $\mathbb{E}\{\xi_j(O_i)\} = 0$ and $\mathbb{E}\{\xi_j(O_i)\xi_j^T(O_i)\} < \infty$ for $j = 1, 2, 3$.

Conditions D4' and D5' are similar to D4 and D5. Conditions D9-D11 are required for establishing the asymptotic normality of the estimators for misspecified models (White 1982).

Theorem 6.4.1 (Double robustness of $\widetilde{T}_{n,DR}^{W,B}$) *Assume Conditions D1-D3, D4', D5' and D6-D11 hold. In addition, assume either $\pi_1(\boldsymbol{x}, \boldsymbol{\alpha})$ or $\Phi_0(\boldsymbol{x}, \boldsymbol{\theta})$ and $\Phi_1(\boldsymbol{x}, \boldsymbol{\zeta})$ are correctly specified. Then, under H_0, for any $0 < \alpha \leq 0.5$, we have*

$$\lim_n \mathbb{P}(\widetilde{T}_{n,DR}^{W,B} > z_\alpha) \leq \alpha,$$

where the equality holds when $\nu(F_0) > 0$. In addition, under H_1, we have

$$\lim_n \mathbb{P}(\widetilde{T}_{n,DR}^{W,B} > z_\alpha) \to 1.$$

Theorem 6.4.1 establishes the consistency of the proposed doubly robust test statistic $\widetilde{T}_{n,DR}^{W,B}$. Next, we establish the power of the test under the local alternative.

Theorem 6.4.2 *Assume Conditions in Theorem 6.4.1 hold. Under H_a, assume that δ_0^W is continuous and bounded on Ω^W, and*

$$\int_{\boldsymbol{x}_W \in F} |\delta_0^W(x_W)| f^W(\boldsymbol{x}_W) d\boldsymbol{x}_W > 0.$$

Then, we have

$$\lim_n \mathbb{P}(\widetilde{T}_{n,DR}^{W,B} > z_\alpha) \geq 1 - \Phi\left(z_\alpha - \frac{1}{2\widetilde{\sigma}_{DR}} \int_{\boldsymbol{x}_W \in F} |\delta_0^W(x_W)| f^W(\boldsymbol{x}_W) d\boldsymbol{x}_W\right).$$

For a given function δ_0, the power of $\widetilde{T}_{n,DR}^{W,B}$ increases as $\widetilde{\sigma}_{DR}$ decreases. When the propensity score model is correctly specified, it can be shown that for each $\boldsymbol{x}_W \in \Omega^W$, $\mu^W_{DR}(x_W)$ achieves its minimum when

$$\Phi_0(\boldsymbol{x},\boldsymbol{\theta}^*) = h(0,\boldsymbol{x}), \;\; \Phi_1(\boldsymbol{x},\boldsymbol{\zeta}^*) = h(1,\boldsymbol{x}). \tag{6.11}$$

Therefore, $\widetilde{\sigma}_{DR}$ achieves its minimum if (6.11) holds. This suggests $\widetilde{T}^{W,B}_{n,DR}$ has the greatest power when the posited models for the propensity score and conditional means functions are correctly specified.

6.5 Implementation details

In Sections 6.3 and 6.4, we only consider continuous covariates for notational convenience. In this section, we present a more general testing framework allowing both continuous and discrete covariates, and provide some implementation details. Specifically, we consider the following two cases: (i) all covariates are discrete; and (ii) at least one covariate is continuous. The test statistics are different in these two cases. We focus on randomized studies and assume the propensity score is known. A doubly-robust version of the test statistic can be similarly derived as in Section 6.4 to deal with data from observational studies. We omit the details to save space.

6.5.1 All covariates are discrete

When all covariates are discrete, for each $\boldsymbol{x}$, we calculate

$$\begin{aligned}
\tau^W_n(\boldsymbol{x}_W) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{A_i}{\pi(1,\boldsymbol{X}_i)} - \frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\right) Y_i \mathbb{I}(\boldsymbol{X}^W_i = \boldsymbol{x}_W),\\
\tau^B_n(\boldsymbol{x}_B) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{A_i}{\pi(1,\boldsymbol{X}_i)} - \frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\right) Y_i \mathbb{I}(\boldsymbol{X}^B_i = \boldsymbol{x}_B),\\
\widehat{f}^W(\boldsymbol{x}_W) &= \frac{1}{n}\sum_{i=1}^n \mathbb{I}(\boldsymbol{X}^W_i = \boldsymbol{x}_W), \qquad \widehat{f}^B(\boldsymbol{X}^B_0) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(\boldsymbol{X}^B_i = \boldsymbol{x}_B),\\
\widehat{\mu}^W_n(\boldsymbol{x}_W) &= \frac{1}{n}\sum_{i=1}^n \left\{\left(\frac{A_i}{\pi(1,\boldsymbol{X}_i)} - \frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\right) Y_i \mathbb{I}(\boldsymbol{X}^W_i = \boldsymbol{x}_W) - \tau^W_n(\boldsymbol{x}_W)\right\}^2,\\
\widehat{\mu}^B_n(\boldsymbol{x}_B) &= \frac{1}{n}\sum_{i=1}^n \left\{\left(\frac{A_i}{\pi(1,\boldsymbol{X}_i)} - \frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\right) Y_i \mathbb{I}(\boldsymbol{X}^B_i = \boldsymbol{x}_B) - \tau^B_n(\boldsymbol{x}_B)\right\}^2.
\end{aligned}$$

Define

$$\widehat{E} = \left\{ \boldsymbol{x}_W \in \Omega^W : \left| \frac{\tau_n^W(\boldsymbol{x}_W)}{\widehat{f}^W(\boldsymbol{x}_W)} \right| \le C_1 \eta_n, \left| \frac{\tau_n^B(\boldsymbol{x}_W^B)}{\widehat{f}^B(\boldsymbol{x}_W^B)} \right| \le C_2 \eta_n \right\}, \tag{6.12}$$

$$\widehat{F} = \left\{ \boldsymbol{x}_W \in \Omega^W : \left| \frac{\tau_n^W(\boldsymbol{x}_W)}{\widehat{f}^W(\boldsymbol{x}_W)} \right| \le C_1 \eta_n, \left| \frac{\tau_n^B(\boldsymbol{x}_W^B)}{\widehat{f}^B(\boldsymbol{x}_W^B)} \right| > C_2 \eta_n \right\}. \tag{6.13}$$

Compute

$$\widetilde{S}_n^{W,B} = \sum_{\boldsymbol{x}_W \notin \widehat{E}} \tau_n^W(\boldsymbol{x}_W)\{d_n^W(\boldsymbol{x}_W) - d_n^B(\boldsymbol{x}_W^B)\}.$$

Unlike results in Section 6.3 and 6.4, the limiting distribution of $\widetilde{S}_n^{W,B}$ is not normal. If $\widehat{F} \neq \emptyset$, we reject the null when $\sqrt{n}\widetilde{S}_n^{W,B} > \widehat{c}_\alpha(\widehat{F})$ where $\widehat{c}_\alpha(F)$ is the upper α-quantile of the random variable $\sum_{\boldsymbol{x}_W \in F} \sqrt{\widehat{\mu}_n^W(\boldsymbol{x}_W)} \max(\mathbb{Z}_{\boldsymbol{x}_W}, 0)$ conditional on $\{\widehat{\mu}_n^W(x_W)\}_{\boldsymbol{x}_W \in \Omega^W}$, where $\{\mathbb{Z}_{x_W}\}_{\boldsymbol{x}_W \in \Omega^W}$ are independent standard normal random variables. Otherwise, we reject the null when $\sqrt{n}\widetilde{S}_n^{W,B} > \widehat{c}_\alpha(\Omega^W)$. A formal justification of the aforementioned testing procedure is given in Section 14 of Shi et al. (2019c).

6.5.2 Not all covariates are discrete

Assume $W = W_C \cup W_D$ and $B = B_C \cup B_D$, where W_C, B_C are the sets of continuous variables and W_D, B_D are the sets of discrete covariates. Denoted by $p_{W_C}, p_{W_D}, p_{B_C}$ and p_{B_D} the numbers of elements in these sets. When $p_{B_C} > 0$, define $\omega_i = \{A_i/\pi(1, \boldsymbol{X}_i) - (1 - A_i)/\pi(0, \boldsymbol{X}_i)\} Y_i$ and

$$\tau_n^W(\boldsymbol{x}_W) = \frac{1}{n \prod_{j \in W_C}(\widehat{s}_j h_W)} \sum_{i=1}^n \omega_i \prod_{j \in W_C} K\left(\frac{x_W^{(j)} - X_i^{(j)}}{\widehat{s}_j h_W} \right) \mathbb{I}(x_{W_D} = \boldsymbol{X}_i^{W_D}),$$

$$\tau_n^B(\boldsymbol{x}_B) = \frac{1}{n \prod_{j \in B_C}(\widehat{s}_j h_B)} \sum_{i=1}^n \omega_i \prod_{j \in B_C} K\left(\frac{x_W^{(j)} - X_i^{(j)}}{\widehat{s}_j h_B} \right) \mathbb{I}(x_{B_D} = \boldsymbol{X}_i^{B_D}),$$

$$\widehat{\mu}_n^W(\boldsymbol{x}_W) = \frac{1}{n \prod_{j \in W_C}(\widehat{s}_j h_W)} \sum_{i=1}^n \left\{ \omega_i \prod_{j \in W_C} K\left(\frac{x_W^{(j)} - X_i^{(j)}}{\widehat{s}_j h_W} \right) \mathbb{I}(\boldsymbol{x}_{W_D} = \boldsymbol{X}_i^{W_D}) \right\}^2,$$

$$\widehat{\mu}_n^B(\boldsymbol{x}_B) = \frac{1}{n \prod_{j \in B_C}(\widehat{s}_j h_B)} \sum_{i=1}^n \left\{ \omega_i \prod_{j \in B_C} K\left(\frac{x_W^{(j)} - X_i^{(j)}}{\widehat{s}_j h_B} \right) \mathbb{I}(\boldsymbol{x}_{B_D} = \boldsymbol{X}_i^{B_D}) \right\}^2,$$

$$\widehat{f}^W(x_W) = \frac{1}{n\prod_{j\in W_C}(\widehat{s}_j h_W)}\sum_{i=1}^{n}\prod_{j\in W_C} K\left(\frac{x_W^{(j)}-X_i^{(j)}}{\widehat{s}_j h_W}\right)\mathbb{I}(\boldsymbol{x}_{W_D}=\boldsymbol{X}_i^{W_D}),$$
$$\widehat{f}^B(x_B) = \frac{1}{n\prod_{j\in B_C}(\widehat{s}_j h_B)}\sum_{i=1}^{n}\prod_{j\in B_C} K\left(\frac{x_W^{(j)}-X_i^{(j)}}{\widehat{s}_j h_B}\right)\mathbb{I}(\boldsymbol{x}_{B_D}=\boldsymbol{X}_i^{B_D}),$$

where $\widehat{s}_j$ denotes the sampling variance of the jth covariate. In our numerical studies, we use a fourth-order Epanechnikov kernel for K, i.e.

$$K(u)=\frac{45}{16}\left(1-\frac{28}{3}u^2\right)(1-4u^2).$$

It can be shown that $\int_u K(u)^j du = 0$ for $j=1,2,3$. Then we calculate

$$\widetilde{S}_n^{W,B}=\sum_{\boldsymbol{x}_W^{W_D}}\int_{x_W^{W_C}} \tau_n^W(\boldsymbol{x}_W)\{d_n^W(\boldsymbol{x}_W)-d_n^B(\boldsymbol{x}_W^B)\}\mathbb{I}(\boldsymbol{x}_W\notin\widehat{E})d\boldsymbol{x}_W^{W_C}.$$

When $p_{W_C}\le 2$, the integral in $\widetilde{S}_n^{W,B}$ is computed via a midpoint rule with a uniform grid. Specifically, for each $j\in W$, denoted by m_j and M_j the minimum and maximum value of $x^{(j)}$. We divide the interval $[m_j,M_j]$ into $L=200$ subintervals of equal width. Let $z_k^{(j)}, k=1,\dots,L$, denote the midpoints for these intervals, $\boldsymbol{z}_{(\bar{k})}=(z_{k_1}^{(1)},\dots,z_{k_{W_C}}^{(p_{W_C})})^T$ for $\bar{k}=(k_1,\dots,k_{W_C})^T$, and $\boldsymbol{z}_{W(\bar{k})}$ and $\boldsymbol{z}_{B(\bar{k})}$ the sub-vectors formed by elements in W_C and B_C respectively. We approximate $\widetilde{S}_n^{W,B}$ by

$$I^*\sum_{\boldsymbol{x}_{W_D},\bar{k}}\tau_n^W(\boldsymbol{z}_{W(\bar{k})},\boldsymbol{x}_{W_D})\{d_n^W(\boldsymbol{z}_{W(\bar{k})},\boldsymbol{x}_{W_D})-d_n^B(\boldsymbol{z}_{B(\bar{k})},\boldsymbol{x}_{W_D}^{B_D})\}\mathbb{I}\{(\boldsymbol{z}_{W(\bar{k})},\boldsymbol{x}_{W_D})\notin\widehat{E}\},$$

where $I^*=\prod_{j\in W^C}(M_j-m_j)/L^{p_{W_C}}$, and $\tau_n^W(\boldsymbol{x}_W^{W_C},\boldsymbol{x}_W^{W_D})$ and $\tau_n^B(\boldsymbol{x}_B^{B_C},\boldsymbol{x}_B^{B_D})$ are shorthands for $\tau_n^W(\boldsymbol{x}_W)$ and $\tau_n^B(\boldsymbol{x}_B)$ respectively, $d_n^W(\boldsymbol{z}_{W(k)},\boldsymbol{x}_{W_D})=\mathbb{I}(\tau_n^W(\boldsymbol{z}_{W(k)},\boldsymbol{x}_{W_D})\ge 0)$ and $d_n^B(\boldsymbol{z}_{B(k)},\boldsymbol{x}_{B_D})=\mathbb{I}(\tau_n^B(\boldsymbol{z}_{B(k)},\boldsymbol{x}_{B_D})\ge 0)$.

If $p_{W_C}>2$, we approximate the integral using Monte Carlo methods. Specifically, we generate $N=5000$ random vectors $\boldsymbol{Z}_{(k)}$, uniformly distributed in $\prod_j[m_j,M_j]$, and calculate

$$I'\sum_{\boldsymbol{x}_{W_D}}\sum_{k=1}^{N}\tau_n^W(\boldsymbol{Z}_{W(k)},\boldsymbol{x}_{W_D})\{d_n^W(\boldsymbol{Z}_{W(k)},\boldsymbol{x}_{W_D})-d_n^B(\boldsymbol{Z}_{B(k)},\boldsymbol{x}_{W_D}^{B_D})\}\mathbb{I}\{(\boldsymbol{Z}_{W(k)},\boldsymbol{x}_{W_D})\notin\widehat{E}\},$$

where $I'=\prod_{j\in W^C}(M_j-m_j)/N$, $\boldsymbol{Z}_{W(k)}$ and $\boldsymbol{Z}_{B(k)}$ are the sub-vectors of $\boldsymbol{Z}_{(k)}$ formed by elements in W_C and B_C.

When $\widehat{F} \neq \emptyset$, we calculated $\widehat{a}_n$ and $\widehat{\sigma}_n^2$ by

$$\begin{aligned}\widehat{a}_n &= \frac{1}{\sqrt{2\pi(h_W)^{p_W}}} \sum_{\boldsymbol{x}_W^{W_D}} \int_{\boldsymbol{x}_W^{W_C}} \sqrt{\widehat{\mu}_n^W(\boldsymbol{x}_W)} \mathbb{I}(\boldsymbol{x}_W \in \widehat{F}) d\boldsymbol{x}_W^{W_C}, \\ \widehat{\sigma}_n^2 &= \sum_{\boldsymbol{x}_W^{W_D}} \int_{\substack{x_W^{W_C} \\ \boldsymbol{t}\in[-1,1]^{p_{W_C}}}} \widehat{\mu}_n^W(\boldsymbol{x}_W)\mathbb{I}(\boldsymbol{x}_W \in \widehat{F})\mathbb{COV}(\max\{\sqrt{1-\rho^2(\boldsymbol{t})}\mathbb{Z}_1 + \rho(\boldsymbol{t})\mathbb{Z}_2, 0\}, \\ &\quad \max\{\mathbb{Z}_2, 0\}) d\boldsymbol{x}_W^{W_C} d\boldsymbol{t}.\end{aligned}$$

Definitions of $\widehat{E}$ and $\widehat{F}$ are given in (6.12) and (6.13). When $\widehat{F} = \emptyset$, we replace $\widehat{F}$ by Ω in the integral. The above integrals are calculated similarly as for $\widetilde{S}_n^{W,B}$. We reject the null when $\sqrt{n}\widetilde{S}_n^{W,B} \geq \widehat{a}_n + \widehat{\sigma}_n z_\alpha$.

6.6 Simulations

To evaluate the numerical performance of the proposed testing procedure, we consider simulation studies based on the following model:

$$Y_0 = h_0(X_0^{(1)}, X_0^{(2)}) + A\tau(X_0^{(1)}, X_0^{(2)}) + e_0,$$

where h_0 denotes the baseline, τ denotes the contrast, and $e_0 \sim N(0, 0.25)$ is independent of A_0 and $\boldsymbol{X}_0 = (X_0^{(1)}, X_0^{(2)})^T$. The objective is to test the CQTE of variable $X_0^{(2)}$ conditional on $X_0^{(1)}$. Treatment A_0 was generated from a Bernoulli distribution with probability 0.5, independent of $\boldsymbol{X}_0$. The baseline function h_0 was set to be

$$h_0(x^{(1)}, x^{(2)}) = 1 - \frac{x^{(1)} - x^{(2)}}{2}. \tag{6.14}$$

The contrast function takes the form

$$\tau(x^{(1)}, x^{(2)}) = \varphi_1(x^{(1)})\varphi_2(x^{(2)}), \tag{6.15}$$

for some continuous functions φ_1 and φ_2.

Variables $X_0^{(1)}$ and $X_0^{(2)}$ are independently generated. It follows from Theorem 6.3.1 that the null (no CQTE) holds if and only if $\varphi_2(x^{(2)}) \geq 0, \forall x^{(2)}$ or $\varphi_2(x^{(2)}) \leq 0, \forall x^{(2)}$. We consider five scenarios. In the first four scenarios, $X_0^{(1)}$ and $X_0^{(2)}$ are generated from Unif$[-2,2]$, where Unif$[a,b]$ stands for the uniform distribution on the interval $[a,b]$. We set $\varphi_1(z) = z$ in the

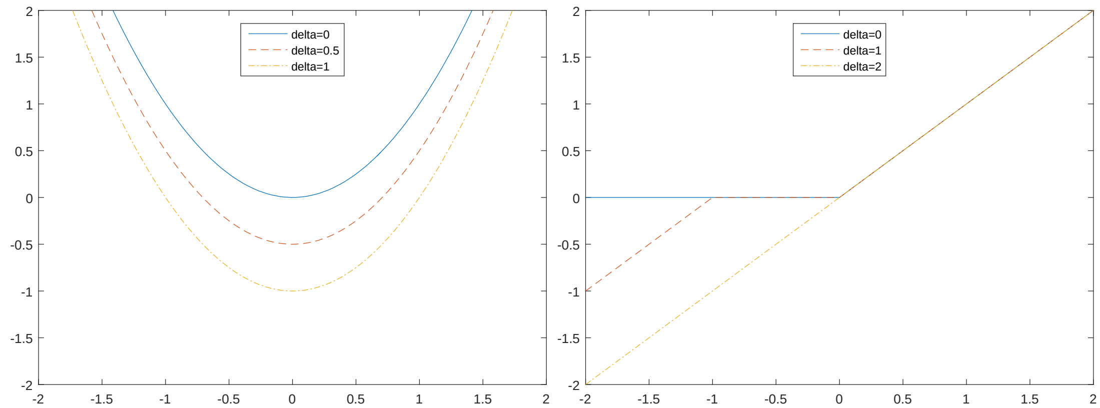

Figure 6.1: Plots of function φ_2 for Scenario 1 and Scenario 2, from left to right, with different choices of δ.

first two scenarios and $\varphi_1(z) = \max(z, 0)$ in the last two scenarios. As for φ_2, in Scenarios 1 and 3,

$$\varphi_2(z) = z^2 - \delta,$$

for some $\delta \geq 0$. In Scenarios 2 and 4,

$$\varphi_2(z) = \begin{cases} z, & 0 \leq z \leq 2, \\ 0, & \delta - 2 \leq z < 0, \\ 2 + z - \delta, & -2 \leq z < \delta - 2, \end{cases}$$

for some $\delta \geq 0$. In figure 6.1, we plot functions φ_2 with different δ.

In the last scenario, $X_0^{(1)}$ is generated from Unif$[-2, 2]$ while $X_0^{(2)}$ is from a uniform discrete distribution. Specifically, $X_0^{(2)}$ has the following probability mass function

$$\mathbb{P}(X^{(2)} = a) = \frac{1}{2}, \quad a = 0, 2.$$

The contrast function is set to be

$$\tau(x^{(1)}, x^{(2)}) = \varphi_1(x^{(1)})\varphi_2(x^{(2)}) = x^{(1)}(x^{(2)} - \delta).$$

In all scenarios, the parameter δ controls the degree of CQTE. When $\delta = 0$, H_0 holds;

Table 6.1: Simulation results.

		VD = 0		VD = 4%		VD = 8%		VD = 12%	
		α level		α level		α level		α level	
	n	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Scenario 1	300	4.3%	6.0%	24.0%	34.0%	58.7%	68.3%	82.2%	87.5%
	600	1.5%	3.3%	36.7%	45.5%	75.8%	83.3%	95.7%	97.3%
Scenario 2	300	7.0%	11.1%	23.8%	32.7%	60.5%	69.3%	88.2%	92.5%
	600	3.7%	7.8%	31.0%	41.8%	83.0%	90.5%	98.3%	99.5%
Scenario 3	300	3.8%	6.5%	37.5%	48.7%	76.5%	79.8%	93.5%	95.5%
	600	2.7%	6.7%	52.5%	61.8%	99.1%	100%	99.8%	99.8%
Scenario 4	300	6.2%	10.2%	39.8%	47.7%	79.2%	87.3%	96.0%	97.8%
	600	5.2%	8.8%	59.3%	68.2%	96.8%	98.3%	100.0%	100.0%
Scenario 5	300	5.2%	9.7%	29.3%	40.5%	68.0%	76.3%	94.0%	96.8%
	600	5.3%	9.5%	46.2%	57.5%	92.2%	95.5%	100.0%	100.0%

Otherwise, H_1 holds. Moreover, it can be calculated that the value differences

$$\mathrm{VD} = \mathbb{E}\left[\tau(X_0^{(1)}, X_0^{(2)})\left\{d^{opt}(\boldsymbol{X}_0) - d^{opt,\{1\}}(X_0^{(1)})\right\}\right],$$

for Scenarios 1-5 are equal to $\delta^{3/2}/3$, $\delta^2/8$, $\delta^{3/2}/6$, $\delta^2/16$ and $\delta/3$ for all $\delta \leq 1$, respectively. In each scenario, we consider four settings by setting $\mathrm{VD} = 0, 0.04, 0.08$ and 0.12. Hence, the null holds in the first setting and the alternative holds in other settings. We also consider two different sample sizes, $n = 300$ and $n = 600$.

When implementing our testing procedure, we first fit a logistic regression model for the propensity score and linear models for the conditional means of Y_0 given A_0 and $\boldsymbol{X}_0$. The test statistics are constructed as discussed in Section 6.5. Based on (6.14) and (6.15), the model for $\mathbb{E}(Y_0|\boldsymbol{X}_0, A = 1)$ is always misspecified, however, the propensity score model is correctly specified. Hence, our test statistics are consistent. In Scenario 1-4, we set the smoothing parameters as $h_W = c_W n^{-1/7}$ and $h_B = c_B n^{-2/7}$ for some constants c_W and c_B. Condition D6 holds for such a choice of the bandwidth. In our implementation, we have tried a few values of c_W and c_B, and find $c_W = 2\sqrt{3}$ and $c_B = 6$ working well for all scenarios. In Scenario 5, we set $h_W = 6n^{-2/7}$. In (6.12) and (6.13), we set $\eta_n = n^{-2/7}$, $C_1 = 3$ and $C_2 = 1$. Such a choice of η_n satisfies Conditions D8-D10 in our simulation settings. We conduct 600 simulations for each setting and report the proportions of rejecting the null hypothesis of the proposed test statistics in Table 6.1.

Under H_0 (i.e. the cases with $\mathrm{VD} = 0$), the empirical type-I error rates in Scenarios 2, 4 and 5 are close to the nominal level. In Scenarios 1 and 3, we have $\nu(F_0) = 0$. The empirical type I error rates in Scenarios 1 and 3 are well below the nominal level. This is in line with our theory which suggests the type-I error rate should go to 0 in these settings. Under H_1, the power increases as the value difference or sample size increases, showing the consistency of our test statistics.

6.7 Application with ACTG175 dataset

We apply our proposed method to a data from AIDS Clinical Trials Group Protocol 175 (ACTG175) study. This is a randomized trial where patients were randomly assigned to the following four treatments, including zidovudine (ZDV) monotherapy, ZDV + didanosine (ddI), ZDV + zalcitabine (zal) and ddI monotherapy. We focus on patients receiving treatments: ZDV+ddI (denoted as 1) and ZDV+zal (denoted as 0). Among them, there are 522 receiving treatment 1 and 524 receiving treatment 0. We choose the CD4 count (cells/mm^3) at 20±5 weeks after receiving the treatment as the response. The baseline covariates include patient's age and weight at baseline, the CD4 and CD8 counts (coded as `CD40` and `CD80` respectively) at baseline, hemophilia (`hemo`, $0 =$ no, $1 =$ yes), homosexual activity (`homo`, $0 =$ no, $1 =$ yes), history of intravenous drug use (`drug`, $0 =$ no, $1 =$ yes), race ($0 =$ white, $1 =$ non-white), gender ($0 =$ female, $1 =$ male), antiretroviral history (`str2`, $0 =$ naive, $1 =$ experienced), and symptomatic status (`sympton`, $0 =$ asymptomatic, $1 =$ symptomatic). The first four variables are continuous while others are binary. Our objective is to select those variables that have qualitative treatment effects in a sequential order. Since the propensity score is known, we consider the statistic $\widetilde{T}^{W,B}$ proposed in Section 6.3. Our procedure proceeds as follows:

1. Set $\widehat{D} = \emptyset$. In the first step, for each variable i, define the set $W_i = \{i\}$ and calculate the p-value p_i for each test statistic $\widetilde{T}^{W_i,\widehat{D}}$ as described in Section 6.5. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p-value in the set $\widehat{D}$, i.e,

$$\widehat{D} \leftarrow \{\arg\min_i p_i\}.$$

2. In the second step, for each variable $i \notin \widehat{D}$, define $W_i = \widehat{D} \cup \{i\}$ and calculate the p-value p_i for each test statistic $\widetilde{T}^{W_i,\widehat{D}}$. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p-value,

$$\widehat{D} \leftarrow \widehat{D} \cup \{\arg\min_i p_i\}.$$

Table 6.2: P-values of each test statistic in all iterations.

age	weight	hemo	homo	drug	race	gender	str2	sympton	CD40	CD80
0.022	0.087	0.793	0.827	0.817	0.831	0.808	0.825	0.825	0.823	0.772
NA	0.986	1.2e-8	0.028	0.288	0.308	0.175	0.257	0.191	0.982	0.975
NA	0.996	NA	0.033	0.067	0.447	0.091	0.155	0.196	0.999	0.998
NA	0.999	NA	NA	0.118	0.116	0.405	0.533	0.066	0.999	0.999

3. Continue the second step until it stops. Output $\widehat{D}$.

It is immediate to see that the above algorithm uses a forward selection procedure. Backward or stepwise selection can be similarly considered. The threshold α determines the significance level for each test statistic. In our implementation, we set $\alpha = 1-\Phi(n^{1/6}/2) \approx 0.056$. Such a choice of α meets the conditions in Theorem 7.3.2 to achieve selection consistency of the forward selection algorithm. As in simulations, we choose the bandwidth $h = 6n^{-2/7}$ when there's only one continuous variable in the kernel estimation. Otherwise, we set $h = 2\sqrt{3}n^{-1/7}$. Sets $\widehat{E}$ and $\widehat{F}$ are estimated by

$$\begin{aligned} \widehat{E} &= \left\{ \boldsymbol{x}_W \in \Omega^W : \left| \frac{\tau_n^W(\boldsymbol{x}_W)}{\sqrt{\widehat{\mu}_n^W(x_W)}} \right| \leq C_0\eta_n, \left| \frac{\tau_n^B(\boldsymbol{x}_W^B)}{\sqrt{\widehat{\mu}_n^B(\boldsymbol{x}_W^B)}} \right| \leq C_0\eta_n \right\}, \\ \widehat{F} &= \left\{ \boldsymbol{x}_W \in \Omega^W : \left| \frac{\tau_n^W(\boldsymbol{x}_W)}{\sqrt{\widehat{\mu}_n^W(x_W)}} \right| \leq C_0\eta_n, \left| \frac{\tau_n^B(\boldsymbol{x}_W^B)}{\sqrt{\widehat{\mu}_n^B(\boldsymbol{x}_W^B)}} \right| > C_0\eta_n \right\}, \end{aligned}$$

where the constant C_0 is set to be 0.03 in the implementation.

For the ACTG175 dataset, our algorithm stops after fourth iteration. At the first iteration, only the variable `age` is significant and is thus selected. At the second iteration, we find out that both `hemo` and `homo` have qualitative effects conditional on `age` and variable `hemo` is chosen. At the third iteration, only `homo` is significant given previously included variables. The algorithm stops at the fourth iteration. We report all the p-values in each iteration in Table 6.2.

Our results indicate that variables `age`, `hemo` and `homo` have qualitative treatment effects and are important for optimal treatment prescription. Denoted by D_{FS} the set of these three variables. We compare our algorithm with the sequential advantage selection (SAS, Fan et al. 2016). SAS uses a forward selection procedure based on a sequential S-score and selects the best candidate subset of variables via a BIC-type criterion. For the ACTG175 dataset, SAS selects a total of 10 variables including `age`, `hemo` and `homo`. Denoted by D_{SAS} the set of these 10 variables.

To further examine the variable selection results, we evaluate the value functions under the optimal treatment regimes based on the set of variables selected by the proposed forward selection algorithm and SAS. For a given set $D \subseteq I_0 = \{1, 2, \ldots, 11\}$, we estimate the optimal value function

$$V^D = \mathbb{E}\{Y_0^*(d_{opt,D})\}$$

via the online estimator proposed by Luedtke and van der Laan (2016). More specifically, for $i = l_n + 1, l_n + 2, \ldots, n$, we first compute the estimated optimal treatment regime $\widehat{d}^D_{(i-1)}(\boldsymbol{x}_D) = \mathbb{I}\{\widehat{h}^D_{1,(i-1)}(\boldsymbol{x}_D) > \widehat{h}^D_{0,(i-1)}(\boldsymbol{x}_D)\}$ and the estimated conditional mean functions $\widehat{\Phi}_{0,(i-1)}(\boldsymbol{x}) = \boldsymbol{x}^T \widehat{\boldsymbol{\theta}}_{0,(i-1)}$ and $\widehat{\Phi}_{1,(i-1)}(\boldsymbol{x}) = \boldsymbol{x}^T \widehat{\boldsymbol{\theta}}_{1,(i-1)}$ based on data from patients 1 to $i-1$.

For any $j = 0, 1$ and $i = l_n + 1, l_n + 2, \ldots, n$, $\widehat{h}^D_{j,(i-1)}$ is calculated via kernel ridge regression, based on the dataset $\{(\boldsymbol{X}_k^D, Y_k)\}_{k \le i-1, A_k = j}$. We use the Gaussian radial basis function kernel. The estimating procedure is implemented by the R package CVST. The tuning parameters in the kernel functions are selected via 5-folded cross-validation. Estimator $\widehat{\boldsymbol{\theta}}_{j,(i-1)}$ is computed via a penalized regression with the SCAD penalty function (Fan and Li 2001), based on the dataset $\{(\boldsymbol{X}_k^D, Y_k)\}_{k \le i-1, A_k = j}$. The penalized regression is implemented by the R package ncvreg, and the tuning parameters are selected via 10-folded cross-validation. Let $\pi_0 = 0.5$, we define for $i = l_n + 1, l_n + 2, \ldots, n,\ j = 1, \ldots, n$,

$$\widehat{V}^D_{(i)}(j) = \frac{\widehat{d}^D_{(i-1),A_j,\boldsymbol{X}_j}}{\pi_0} Y_j - \left(\frac{\widehat{d}^D_{(i-1),A_j,\boldsymbol{X}_j}}{\pi_0} - 1 \right)$$
$$\times \quad \left(\widehat{\Phi}_{1,(i-1)}(\boldsymbol{X}_j^D) \widehat{d}^D_{(i-1)}(\boldsymbol{X}_j^D) + \widehat{\Phi}_{0,(i-1)}(\boldsymbol{X}_j^D)\{1 - \widehat{d}^D_{(i-1)}(\boldsymbol{X}_j^D)\} \right),$$

where $\widehat{d}^D_{(i-1),A_j,\boldsymbol{X}_j} = A_j \widehat{d}^D_{(i-1)}(\boldsymbol{X}_j^D) + (1 - A_j)\{1 - \widehat{d}^D_{(i-1)}(\boldsymbol{X}_j^D)\}$.

The final estimator is given by

$$\widehat{V}^D = \frac{\sum_{i=l_n+1}^n \{\widehat{\sigma}^D(i)\}^{-1} \widehat{V}^D_{(i)}(i)}{\sum_{i=l_n+1}^n \{\widehat{\sigma}^D(i)\}^{-1}},$$

with the estimated standard error

$$\widehat{\sigma}^D = \frac{\sqrt{n - l_n}}{\sum_{i=l_n+1}^n \{\widehat{\sigma}^D(i)\}^{-1}},$$

where

$$\{\widehat{\sigma}^D(i)\}^2 = \frac{1}{i-2}\sum_{j=1}^{i}\{\widehat{V}^D_{(i-2)}(j)\}^2 - \left(\frac{1}{i-1}\sum_{j=1}^{i-1}\widehat{V}^D_{(i)}(j)\right)^2.$$

Under certain conditions, we have

$$\frac{\widehat{V}^D - V^D}{\widehat{\sigma}^D} \xrightarrow{d} N(0,1).$$

Set $l_n = 200$. The estimated value functions $\widehat{V}^{D_{FS}}$ and $\widehat{V}^{D_{SAS}}$ are equal to 401.88 and 402.35 respectively, with estimated standard errors $\widehat{\sigma}^D_{FS} = 7.50$ and $\widehat{\sigma}^D_{SAS} = 7.19$. Since $D_{FS} \subseteq D_{SAS}$, we have $V^{D_{SAS}} \geq V^{D_{FS}}$. However, the difference $V^{D_{SAS}} - V^{D_{FS}}$ is not significant. This implies the proposed forward selection algorithm selects less variables than SAS, while achieves approximately the same value function in optimal treatment decision.

6.8 Discussion

In this chapter, we introduce the notion of conditional qualitative treatment effects (CQTE) and present several equivalent definitions. We also propose a consistent testing procedure for the existence of CQTE. Our test has correct size under the null hypothesis and non-negligible power against some nonstandard local alternatives.

6.8.1 More on the forward selection algorithm

The forward selection algorithm introduced in Section 6.7 is a byproduct of the proposed testing procedure for the existence of CQTE. While it is worthwhile to investigate its statistical properties, this is a very challenging task. In the literature, few works have studied the asymptotic properties of a forward selection procedure. Wang (2009) established the “sure screening property" of the classical forward linear regression in a high dimensional setting. However, the proofs of the major theorems in that paper (Theorem 1 and 2) rely heavily on the specific structure of linear regression and it remains unknown whether the “sure screening property" holds for general forward selection algorithms.

Our forward selection algorithm aims to identify a subset $D_0 \subseteq \{1,\dots,p\}$ with minimum cardinality such that the optimal value function based on variables in $\boldsymbol{X}_0^{D_0}$ is the same as that based on $\boldsymbol{X}_0$. Shi et al. (2019c) established the “sure screening property" and selection consistency of the considered forward selection algorithm based on the p-values of the CQTE tests. Moreover, they conducted some simulation studies to examine the empirical performance of the proposed algorithm and compare it with SAS (Fan et al. 2016). The forward selection algorithm achieves better model selection results when compared to SAS in all considered simulation scenarios. More details can be found in Section 9 of Shi et al. (2019c).

6.8.2 Fully nonparametric implementation

The proposed test statistic in Section 6.3 requires the propensity score function to be correctly specified. In Section 6.4, we introduce a doubly robust test statistic and posit some parametric models for the propensity score and conditional mean functions. Shi et al. (2019c) considered a fully nonparametric procedure based on some nonparametric estimators of the propensity score and the conditional mean functions.

Shi et al. (2019c) further conducted some simulation studies to examine the empirical performance of the nonparametric testing procedure and compare it with the doubly robust test describe in Section 6.4. We briefly summarize the results here: (i) The nonparametric test statistic is more powerful than the doubly robust test statistic. (ii) When the sample size is small, the empirical type I error rates of the nonparametric test statistic are slightly larger than the nominal level in some cases. More details can be found in Section 11 of Shi et al. (2019c).

6.8.3 Extensions to L_p-type and supremum-type functionals

As commented before, the test statistic for no CQTE can be constructed based on

$$S^{W,B} = \int_{\boldsymbol{x}_W \in \Omega^W} \phi\{\tau^W(\boldsymbol{x}_W)\}\{d^{opt,W}(x_W) - d^{opt,B}(\boldsymbol{x}_W^B)\}\omega_0(x_W) d\boldsymbol{x}_W.$$

In the current paper, we set $\phi(\cdot)$ to be the identity function. More generally, we can take $\phi(\cdot)$ to be any monotonically increasing function with $\phi(0) = 0$. In Section 12 of Shi et al. (2019c), they considered the following class of functions $\phi(z) = \text{sgn}(z)|z|^q$, and derive the corresponding test statistic $\widetilde{T}_{n,q}^{W,B}$ for any $q \geq 1$.

Shi et al. (2019c) showed that $\widetilde{T}_{n,q}^{W,B}$ have asymptotically correct size under H_0 and provide its asymptotic power function under H_a. For different q, the asymptotic power function increases as

$$\int_{\boldsymbol{x}_W \in F_0} \frac{2^{(q-3)/2} q \Gamma(q/2)}{\sqrt{\pi}\widetilde{\sigma}_q} \{\mu^W(\boldsymbol{x}_W)\}^{(q-1)/2} \delta_0^W(x_W) f^W(\boldsymbol{x}_W) d\boldsymbol{x}_W$$

increases, where

$$\widetilde{\sigma}_q^2 = \int_{\substack{\boldsymbol{x}_W \in F_0 \\ t \in [-1,1]^{p_W}}} \mu^W(\boldsymbol{x}_W) \mathbb{COV}(\max\{\sqrt{1-\rho^2(\boldsymbol{t})}\mathbb{Z}_1 + \rho(\boldsymbol{t})\mathbb{Z}_2, 0\}^q, \max\{\mathbb{Z}_2, 0\}^q) d\boldsymbol{x}_W d\boldsymbol{t},$$

and $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$.

Besides, when $q > 1$, the assumptions on η_n and the moments of Y_0 conditional on $\boldsymbol{X}_0$ and A_0 are slightly different compared to those in D3 and D8. More details can be found in Section 12 of Shi et al. (2019c).

In addition, Shi et al. (2019c) also developed a supremum-type test based on studentized kernel estimators of the contrast function, with many different bandwidth values. They showed that the test is valid and has nontrivial power against $\sqrt{\log n}/\sqrt{n h_{\max}^p}$-local alternatives, where $h_{\max}$ denotes the maximum of the kernel bandwidth parameter.

Therefore, when compared to the supremum-type test, the L_p-type test is more powerful since it allows for nontrivial testing against $n^{-1/2}$-local alternatives. However, the L_p-type test only uses one bandwidth value for the kernel estimates. As a result, it might be sensitive to the choice of the bandwidth parameter.

6.8.4 Other issues

For simplicity, we only consider a single decision stage and focus on binary treatments. It will be useful in practice to extend CQTE and its associated testing procedure to multi-stages with multiple treatment options. Moreover, our test statistic relies on the kernel-based estimators of the contrast function. It is well known that the kernel-based estimations will behave poorly when the dimension of the covariates is large. How to adapt our test statistics to handle high-dimensional covariates remains challenging.

Our testing procedure requires the specification of the tuning parameters h_W, h_B and η_n (see Section 6.5). In general, one can set $h_W = c_W n^{-\kappa_W}$, $h_B = c_B n^{-\kappa_B}$ and $\eta_n = n^{-\kappa_0}$ for some $c_W, c_B, \kappa_W, \kappa_B, \kappa_0 > 0$. In practice, we recommend to set $c_W = 2\sqrt{3}$, $c_B = 6$, $\kappa_W = 1/7$, $\kappa_B = 2/7$ if $p_W \geq 2$, $p_B = 1$ and $c_W = c_B = 2\sqrt{3}$, $\kappa_W = \kappa_B = 1/7$ if $p_W, p_B \geq 2$, and $\kappa_0 = 2/7$. We have tried various values of tuning parameters in our simulation studies and find such a choice works well in all scenarios. Shi et al. (2019c) examined the performance of our test under other choices of tuning parameters. The simulation results are very similar to those in Section 6.6.

CHAPTER

7

INFERENCE OF THE OPTIMAL VALUE FUNCTION

7.1 Introduction

Prior to adopting any OTR in clinical practice, it is crucial to know the impact of implementing such a policy. This requires to evaluate the mean outcome in the population under an OTR, i.e, the optimal value function. The inference of the optimal value function helps us to evaluate whether the OTR can lead to a clinically meaningful increment value compared to fixed treatment regimes. However, statistical inference of the optimal value function is an extremely challenging task in the nonregular cases where there is a positive probability that the interaction between treatment and covariates (i.e, the contrast function) is equal to zero. The main challenge lies in that the OTR is unknown and needs to be estimated from the data. Consider the following naive method that first estimates the OTR and then evaluates its mean outcome based on the augmented inverse propensity-score weighted estimator (AIPWE) for the value function (Zhang et al. 2012, 2013). The validity of such a procedure relies on the estimated treatment regime being consistent to the OTR. However, this condition is typically violated in the nonregular cases (see Section 4.1 in Luedtke and van der Laan 2016).

Chakraborty et al. (2014) considered inference for the value of an estimated OTR using the m-out-of-n bootstrap. The CI based on this method is valid in the nonregular cases when m grows to infinity at a rate slower than n. However, the length of the CI shrinks at a rate of $m^{-1/2}$. As a result, such CI will be much wider than the CI of our proposed procedure which shrinks at a rate of $n^{-1/2}$. Luedtke and van der Laan (2016) proposed an online one-step estimator that is $n^{-1/2}$-consistent to the optimal value function. Their method mimics the online prediction algorithms and recursively updates the initial estimated OTR and value function using new observations. Based on the online one-step estimator, they developed a valid inference procedure. However, their procedure relies on a data ordering. The online one-step estimator can be sensitive to the order of the data, especially when the sample size is small.

In this chapter, we develop a novel inference method for the optimal value function based on subsample aggregating (subagging) and refitted cross-validation. Specifically, we estimate the OTR based on a random subsample of the data and evaluate its value based on the remaining data using AIPWE. We then iterate this procedure multiple times. Our final estimator is defined as an average of all value estimators. Bootstrap aggregating (bagging) and subagging have been recognized as effective variance reduction techniques in hard decision problems (Bühlmann and Yu 2002). However, it remains unknown whether these procedures can yield valid inference results. We show the proposed estimator is asymptotically normal even at nonregular cases. We further provide a consistent estimator for its asymptotic variance and derive a Wald-type CI for the optimal value function.

At nonregular cases, the estimated OTR might fluctuate randomly and does not converge to a fixed function, even as the sample size grows to infinity. Subagging averages over estimated OTRs computed based on different subsamples, resulting in a smoothed decision rule, yielding smaller variance and mean squared error. Due to the variance reduction effect of subagging, our method enjoys certain statistical optimality. More specifically, we prove that the squared length of the proposed CI is on average shorter than the CI constructed based on the online one-step method (Luedtke and van der Laan 2016) in the nonregular cases. In addition, under certain conditions on the propensity score function, our proposed CI is asymptotically narrower than the CI of the "oracle" method which works as well as if the OTR were known. These theoretical findings are further supported by extensive numerical studies. Moreover, the proposed method can be applied to multi-stage studies to evaluate the mean outcome under an ODTR.

The rest of the chapter is organized as follows. In Section 7.2, we introduce our inference procedure in a point treatment study. In Section 7.3, we discuss the asymptotic optimality of our proposed method. Section 7.4 contains the extension to multi-stage studies. Simulation studies are conducted in Section 7.5. In Section 7.6, we apply the proposed method to a real dataset. The proof of Theorem 7.2.1 is given in Section 7.7, followed by a discussion section. Other proofs are provided in Shi et al. (2019d).

7.2 Point treatment study

7.2.1 Subsample aggregation and sample-split estimation

We begin by considering a single stage study with two treatments (see Section 1.2 for notations and model setup). Let $V_0 = \max_d V(d) = V(d^{opt,0})$. Our objective is to construct confidence intervals (CIs) for V_0. To estimate the optimal value function, we need to estimate the OTR first. We consider the class of plug-in classifiers. More specifically, for any $\mathscr{I} \subseteq \{1,\ldots,n\}$, let $\widehat{\tau}_{\mathscr{I}}(\cdot)$ denote the estimated contrast function based on the sub-dataset $\{O_i\}_{i\in\mathscr{I}}$ and define $\widehat{d}_{\mathscr{I}}(\cdot) = \mathbb{I}\{\widehat{\tau}_{\mathscr{I}}(\cdot) > 0\}$. We may set $\widehat{\tau}_{\mathscr{I}}(\cdot) \equiv 0$ if either $\sum_{i\in\mathscr{I}} A_i = 0$ or $\sum_{i\in\mathscr{I}}\{1-A_i\} = 0$.

For any $\boldsymbol{x}$ that satisfies $\tau(\boldsymbol{x}) = 0$, the estimated OTR $\widehat{d}_{\mathscr{I}}(\boldsymbol{x})$ is "unstable" in the sense that it might recommend different treatments due to sufficiently small changes in the data. As a result, $\widehat{d}_{\mathscr{I}}(\boldsymbol{x})$ might not converge to a deterministic quantity even though $V(\widehat{d}_{\mathscr{I}})$ is consistent to V_0. Subagging remedies this issue by averaging across estimated OTRs from subsamples as illustrated below.

For $a = 0,1$ and any $\boldsymbol{x} \in \mathbb{X}$, define the propensity score function $\pi(a,\boldsymbol{x}) = \mathbb{P}(A_0 = a|\boldsymbol{X}_0 = \boldsymbol{x})$ and the conditional mean function $h(a,\boldsymbol{x}) = \mathbb{E}(Y_0|A_0 = a, \boldsymbol{X}_0 = \boldsymbol{x})$. The observed data can be summarized as $\{O_i = (\boldsymbol{X}_i, A_i, Y_i), i = 1,\ldots,n\}$ where O_i's are i.i.d copies of $O_0 = (\boldsymbol{X}_0, A_0, Y_0)$. Let $\pi^*(\cdot,\cdot)$ and $h^*(\cdot,\cdot)$ denote some estimators for $\pi(\cdot,\cdot)$ and $h(\cdot,\cdot)$. For any $\mathscr{I} \subseteq \{1,2,\ldots,n\}$ and any treatment regime $d(\cdot)$, consider the following augmented inverse propensity-score weighted estimator (AIPWE, Zhang et al. 2012) for $V(d)$,

$$\widehat{V}_{\mathscr{I}}(d;\pi^*,h^*) = \frac{1}{|\mathscr{I}|}\sum_{i\in\mathscr{I}}\left(\frac{\mathrm{g}\{A_i, d(\boldsymbol{X}_i)\}}{\pi^*(A_i,\boldsymbol{X}_i)}\{Y_i - h^*(A_i,\boldsymbol{X}_i)\} + h^*\{d(\boldsymbol{X}_i),\boldsymbol{X}_i\}\right),$$

where $|\mathscr{I}|$ denotes the number of elements in $\mathscr{I}$, $\mathrm{g}(y,z) = yz + (1-y)(1-z)$ for all $y,z \in \mathbb{R}$, and π_i is a shorthand for $\pi(\boldsymbol{X}_i)$. The above estimator is consistent when either π^* or h^* is consistent. To better illustrate our method, in the following, we assume functions π and h are known. In Section 7.2.2, we allow these functions to be estimated from the observed dataset. We use a shorthand and write $\widehat{V}_{\mathscr{I}}(d)=\widehat{V}_{\mathscr{I}}(d;\pi,h)$ for any $\mathscr{I}$ and $d(\cdot)$.

Let $\mathscr{I}_0=\{1,2,\ldots,n\}$. To obtain valid CI for the optimal value function, we apply sample-split estimation with subagging. More specifically, we estimate V_0 by

$$\widehat{V}^*_{\infty}=\frac{1}{\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\widehat{V}_{\mathscr{I}^c}(\widehat{d}_{\mathscr{I}}),$$

where s_n is some diverging sequence, and $\mathscr{I}^c$ denotes the complement of $\mathscr{I}$.

For any integer $j>0$, define $p_j(\boldsymbol{x})=\mathbb{P}(\widehat{\tau}_{\{1,2,\ldots,j\}}(\boldsymbol{x})>0)$. For any $\mathscr{I}\subseteq\mathscr{I}_0$ with $|\mathscr{I}|=s_n$, let $R_{\mathscr{I}}(\boldsymbol{x})=\widehat{d}_{\mathscr{I}}(\boldsymbol{x})-p_{s_n}(\boldsymbol{x})$. It is immediate to see that $\mathbb{E}R_{\mathscr{I}}(\boldsymbol{x})=0$ for all $\boldsymbol{x}\in\mathbb{X}$. For any $y\in\mathbb{R}$, $\boldsymbol{x}\in\mathbb{X}$, let $h(y,\boldsymbol{x})=y\,h(1,\boldsymbol{x})+(1-y)h(0,\boldsymbol{x})$. By definition, we have

$$\begin{aligned}
&\widehat{V}^*_{\infty}=\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\left(\frac{\mathrm{g}\{A_i,\widehat{d}_{\mathscr{I}}(\boldsymbol{X}_i)\}}{\pi(A_i,\boldsymbol{X}_i)}\{Y_i-h(A_i,\boldsymbol{X}_i)\}+h\{\widehat{d}_{\mathscr{I}}(\boldsymbol{X}_i),\boldsymbol{X}_i\}\right)\\
=\;&\underbrace{\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\left(\frac{\mathrm{g}\{A_i,p_{s_n}(\boldsymbol{X}_i)\}}{\pi(A_i,\boldsymbol{X}_i)}\{Y_i-h(A_i,\boldsymbol{X}_i)\}+h\{p_{s_n}(\boldsymbol{X}_i),\boldsymbol{X}_i\}\right)}_{\eta_1}\\
+\;&\underbrace{\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\left(\frac{(2A_i-1)R_{\mathscr{I}}(\boldsymbol{X}_i)}{\pi(A_i,\boldsymbol{X}_i)}\{Y_i-h(A_i,\boldsymbol{X}_i)\}+\tau(\boldsymbol{X}_i)R_{\mathscr{I}}(\boldsymbol{X}_i)\right)}_{\eta_2}.
\end{aligned}$$

When $s_n=o(n)$, we can show $\eta_2=o_p(n^{-1/2})$. Let $\mathscr{I}_{(-i)}=\mathscr{I}_0-\{i\}$, we have

$$\begin{aligned}
\eta_1 &= \frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{i=1}^{n}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_{(-i)}\\|\mathscr{I}|=s_n}}\left(\frac{\mathrm{g}\{A_i,p_{s_n}(\boldsymbol{X}_i)\}}{\pi(A_i,\boldsymbol{X}_i)}\{Y_i-h(A_i,\boldsymbol{X}_i)\}+h\{p_{s_n}(\boldsymbol{X}_i),\boldsymbol{X}_i\}\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathrm{g}\{A_i,p_{s_n}(\boldsymbol{X}_i)\}}{\pi(A_i,\boldsymbol{X}_i)}\{Y_i-h(A_i,\boldsymbol{X}_i)\}+h\{p_{s_n}(\boldsymbol{X}_i),\boldsymbol{X}_i\}\right).
\end{aligned}$$

Notice that η_1 corresponds to the estimated value function under the smoothed decision rule $p_{s_n}(\cdot)$ which is a deterministic function of $\boldsymbol{x}$. As a result, $\sqrt{n}\widehat{V}^*_{\infty}$ is asymptotically equivalent to a sum of i.i.d mean zero random variables and can be used to construct the CI for V_0. Below, we formally establish these results. We need the following conditions.

(C5) Assume $\sup_{\boldsymbol{x}\in\mathbb{X},a=0,1}\mathbb{E}[\{Y_0^*(a)\}^2|\boldsymbol{X}_0=\boldsymbol{x}]=O(1)$.

(C6) Assume there exists some constant $\kappa_0 > (\alpha+2)/(2\alpha+2)$ such that

$$\mathbb{E}|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_0) - \tau(\boldsymbol{X}_0)|^2 = O(|\mathscr{I}|^{-\kappa_0}),$$

for any $\mathscr{I} \subseteq \mathscr{I}_0$.

Condition C5 automatically holds when the potential outcomes $Y_0^*(0)$ and $Y_0^*(1)$ are bounded. In C6, we assume the estimated contrast function shall satisfy certain convergence rates. These rates are available for most often used nonparametric approaches including spline methods (Zhou et al. 1998), kernel ridge regression (Steinwart and Christmann 2008; Zhang et al. 2013), tree-based methods (Zhu et al. 2017) and random forests (Biau 2012). When C4 (defined in Chapter 5) and C6 hold, we can show that $V(\widehat{d}_{\mathscr{I}}) = V_0 + o_p(|\mathscr{I}|^{-1/2})$ for any $\mathscr{I} \subseteq \mathscr{I}_0$. We present our main results below.

Theorem 7.2.1 *Assume C1-C6 hold, and s_n satisfies $s_n \asymp n^{\beta_0}$ for some $(2+\alpha)/\{\kappa_0(2+2\alpha)\} < \beta_0 < 1$. Then, we have*

$$\widehat{V}_\infty^* = \eta_1 + o_p(n^{-1/2}) \text{ and } V_0 = \mathbb{E}\eta_1 + o(n^{-1/2}).$$

Theorem 7.2.1 implies that $\sqrt{n}(\widehat{V}_\infty^* - V_0) = \sqrt{n}\{\eta_1 - \mathbb{E}\eta_1\} + o_p(1)$. For $j = 1, 2, \ldots$, let

$$\sigma_j^2 = \mathbb{VAR}\left(\frac{\text{g}\{A_0, p_j(\boldsymbol{X}_0)\}}{\pi(A_0, \boldsymbol{X}_0)} \{Y_0 - h(A_0, \boldsymbol{X}_0)\} + h\{p_j(\boldsymbol{X}_0), \boldsymbol{X}_0\} \right).$$

By C3 and C5, we have $\sup_{j\geq 1} \sigma_j = O(1)$. Assume $\liminf_n \sigma_n > 0$. By central limit theorem, we have

$$\frac{\sqrt{n}(\widehat{V}_\infty^* - V_0)}{\sigma_{s_n}} \xrightarrow{d} N(0,1).$$

For any $z_1, \ldots, z_n \in \mathbb{R}$, let $\widehat{s.e}^2(\{z_i\}_{i=1}^n)$ denote the sample variance estimator, i.e, $\widehat{s.e}^2(\{z_i\}_{i=1}^n) = \sum_{i=1}^n (z_i - \bar{z})^2/(n-1)$ where $\bar{z} = \sum_{i=1}^n z_i/n$. The asymptotic variance $\sigma_{s_n}^2$ can be consistently estimated by

$$\widehat{\sigma}_\infty^{*2} = \widehat{s.e}^2\left(\left\{ \frac{\text{g}\{A_i, \widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i)\}}{\pi(A_i, \boldsymbol{X}_i)} \{Y_i - h(A_i, \boldsymbol{X}_i)\} + h\{\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i), \boldsymbol{X}_i\} \right\}_{i=1}^n \right),$$

where

$$\widehat{d}_{s_n}^{(-i)}(\boldsymbol{x}) = \frac{1}{\binom{n-1}{s_n}} \sum_{\substack{\mathscr{I} \subseteq \mathscr{I}_{(-i)} \\ |\mathscr{I}|=s_n}} \widehat{d}_{\mathscr{I}}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{X}.$$

Notice that it is intractable to compute $\widehat{d}_{\mathscr{I}}$ over all possible size s_n subsamples of the training data. In practice, we can estimate $\widehat{V}_{\infty}^{*}$ based on Monte Carlo approximations. More specifically, for a sufficiently large integer B, set

$$\widehat{V}_B^* = \frac{1}{B} \sum_{b=1}^{B} \widehat{V}_{\mathscr{I}_b^c}(\widehat{d}_{\mathscr{I}_b}), \tag{7.1}$$

where the subsets $\mathscr{I}_1, \ldots, \mathscr{I}_B$ are drawn uniformly from the set

$$\mathscr{S}_{N_0,s_n} = \left\{ \mathscr{I} \subseteq \mathscr{I}_0 : |\mathscr{I}| = s_n, N_0 \le \sum_{i \in \mathscr{I}} A_i \le s_n - N_0 \right\},$$

for some positive integer N_0. Here, the constraints $N_0 \le \sum_{i \in \mathscr{I}} A_i \le s_n - N_0$ guarantee that the function $\tau(\cdot)$ is estimable based on the sub-dataset $\{O_i\}_{i \in \mathscr{I}}$. Define

$$\widehat{\sigma}_B^{*2} = \widehat{s.e}^2 \left(\left\{ \frac{\mathrm{g}\{A_i, \widehat{d}_{s_n,B}^{(-i)}(\boldsymbol{X}_i)\}}{\pi(A_i, \boldsymbol{X}_i)} \{Y_i - h(A_i, \boldsymbol{X}_i)\} + h(\widehat{d}_{s_n,B}^{(-i)}(\boldsymbol{X}_i), \boldsymbol{X}_i) \right\}_{i=1}^{n} \right), \tag{7.2}$$

where

$$\widehat{d}_{s_n,B}^{(-i)}(\boldsymbol{X}_i) = \frac{1}{n^{(i)}} \sum_{b:\{i \notin \mathscr{I}_b\}} \widehat{d}_{\mathscr{I}_b}(\boldsymbol{X}_i),$$

and $n^{(i)} = \sum_{b=1}^{B} \mathbb{I}(i \notin \mathscr{I}_b)$. The corresponding two-sided CI for V_0 is given by

$$\left[\widehat{V}_B^* - \frac{2 z_{\alpha/2} \widehat{\sigma}_B^*}{\sqrt{n}}, \widehat{V}_B^* + \frac{2 z_{\alpha/2} \widehat{\sigma}_B^*}{\sqrt{n}} \right].$$

7.2.2 Unknown propensity score and conditional mean functions

In practice, the conditional mean function $h(\cdot,\cdot)$ is unknown to us. In observational studies, the propensity score function $\pi(\cdot,\cdot)$ also needs to be estimated from the data. Parametric models are commonly used to estimate these functions. The resulting value estimator is consistent when either $h(\cdot,\cdot)$ or $\pi(\cdot,\cdot)$ is correctly specified. To avoid model misspecification and gain efficiency, we propose to estimate these function nonparametrically and use a sample-splitting method to construct AIPWE. The use of sample-splitting helps reduce the bias of AIPWE resulting from the biases of the estimated propensity score and conditional mean functions.

For any $\mathscr{I} \subseteq \mathscr{I}_0$, denoted by $\widehat{h}_{\mathscr{I}}$ and $\widehat{\pi}_{\mathscr{I}}$ the corresponding estimators for h and π, based on the sub-dataset $\{O_i\}_{i\in\mathscr{I}}$. For simplicity, assume $n - s_n = 2t_n$ for some integer $t_n > 0$. We detail our procedure in the following algorithm.

Step 1. Input observations $\{O_i\}_{i=1,\dots,n}$, $0 < \alpha < 1$, and integers s_n, N_0, B.

Step 2. For $b = 1, \dots, B$,

(i) Draw a subset $\mathscr{I}_b$ from $\mathscr{S}_{N_0,s_n}$ uniformly at random.
(ii) Randomly partition $\mathscr{I}_b^c$ into 2 disjoint subsets $\mathscr{I}_b^{c(1)}$ and $\mathscr{I}_b^{c(2)}$ of equal sizes t_n.
(iii) For $j = 1,2$, let $\mathscr{I}_b^{(j)} = \mathscr{I}_b \cup \mathscr{I}_b^{c(j)}$. Obtain the estimators $\widehat{d}_{\mathscr{I}_b}$, $\widehat{\pi}_{\mathscr{I}_b^{(1)}}$, $\widehat{\pi}_{\mathscr{I}_b^{(2)}}$, $\widehat{h}_{\mathscr{I}_b^{(1)}}$ and $\widehat{h}_{\mathscr{I}_b^{(2)}}$.

Step 3. Compute

$$\widehat{V}_B = \frac{1}{2B}\sum_{b=1}^{B}\Big(\widehat{V}_{\mathscr{I}_b^{c(2)}}(\widehat{d}_{\mathscr{I}_b};\widehat{\pi}_{\mathscr{I}_b^{(1)}},\widehat{h}_{\mathscr{I}_b^{(1)}}) + \widehat{V}_{\mathscr{I}_b^{c(1)}}(\widehat{d}_{\mathscr{I}_b};\widehat{\pi}_{\mathscr{I}_b^{(2)}},\widehat{h}_{\mathscr{I}_b^{(2)}})\Big),$$

and $\widehat{\sigma}_B^2 = \widehat{s.e}^2(\{\widehat{V}^{(i)}\}_{i=1}^n)$ where

$$\widehat{V}^{(i)} = \frac{1}{n^{(i)}}\sum_{b=1}^{B}\Big(\widehat{V}_{\{i\}}(\widehat{d}_{\mathscr{I}_b};\widehat{\pi}_{\mathscr{I}_b^{(1)}},\widehat{h}_{\mathscr{I}_b^{(1)}})\mathbb{I}(i\notin\mathscr{I}_b^{(1)}) + \widehat{V}_{\{i\}}(\widehat{d}_{\mathscr{I}_b};\widehat{\pi}_{\mathscr{I}_b^{(2)}},\widehat{h}_{\mathscr{I}_b^{(2)}})\mathbb{I}(i\notin\mathscr{I}_b^{(2)})\Big),$$

and $n^{(i)} = \sum_{b=1}^{B}\mathbb{I}(i \notin \mathscr{I}_b)$.

Step 4. Output

$$\left[\widehat{V}_B - \frac{z_{\alpha/2}\widehat{\sigma}_B}{\sqrt{n}}, \widehat{V}_B + \frac{z_{\alpha/2}\widehat{\sigma}_B}{\sqrt{n}}\right]. \tag{7.3}$$

Notice that we apply a two-fold cross-validation procedure in Step 2 and 3. More generally, one can use K-fold cross-validation to construct the estimators $\widehat{V}_B$ and $\widehat{\sigma}_B^2$, for any fixed integer $K \geq 2$. The following theorem proves the validity of the CI in (7.3).

Theorem 7.2.2 *Assume* $B \gg n$, $\liminf_n \sigma_n > 0$,

$$\mathbb{P}\left(\inf_{\mathscr{I}\in\mathscr{I}_0, \boldsymbol{x}\in\mathbb{X}, a=0,1} \widehat{\pi}_{\mathscr{I}}(a,\boldsymbol{x}) \geq c^*\right) = 1, \tag{7.4}$$

for some constant $c^* > 0$. *In addition, assume*

$$\max_{a=0,1} \mathbb{E}|\widehat{\pi}_{\mathscr{I}}(a,\boldsymbol{X}_0) - \pi(a,\boldsymbol{X}_0)|^2 = o(|\mathscr{I}|^{-1/2}), \tag{7.5}$$

$$\max_{a=0,1} \mathbb{E}|\widehat{h}_{\mathscr{I}}(a,\boldsymbol{X}_0) - h(a,\boldsymbol{X}_0)|^2 = o(|\mathscr{I}|^{-1/2}), \tag{7.6}$$

for any $\mathscr{I} \subseteq \mathscr{I}_0$. *Then, under the conditions in Theorem 7.2.1, we have*

$$\frac{\sqrt{n}(\widehat{V}_B - V_0)}{\widehat{\sigma}_B} \xrightarrow{d} N(0,1).$$

In (7.5) and (7.6), we require the estimated propensity score and conditional mean functions to satisfy certain convergence rates. These conditions guarantee that $\widehat{V}_B$ and $\widehat{\sigma}_B^2$ are asymptotically equivalent to $\widehat{V}_B^*$ and $\widehat{\sigma}_B^{*2}$, defined in (7.1) and (7.2), respectively.

Theorem 7.2.2 shows the asymptotic normality of $\sqrt{n}(\widehat{V}_B - V_0)/\widehat{\sigma}_B$. As a result, the two-sided CI defined in (7.3) has asymptotically nominal coverage probabilities. Moreover, it also implies that $\widehat{V}_B - z_\alpha \widehat{\sigma}_B/\sqrt{n}$ is an asymptotic $1-\alpha$ lower confidence bound for V_0.

Our proposed method depends on the size of subsamples s_n. A larger s_n helps reduce the bias of the proposed value estimator. However, we require $s_n = o(n)$ to guarantee its asymptotic normality. In practice, we recommend to set $s_n = n^{1-\epsilon}$ for some small $\epsilon > 0$.

7.3 Asymptotic optimality

This section discusses the optimality of the proposed method. The length of the proposed CI (see (7.3)) is given by $L(\widehat{V}_B, \alpha) = 2z_{\alpha/2}\widehat{\sigma}_B/\sqrt{n}$. Under the given conditions in Theorem 7.2.2, the estimator $\widehat{\sigma}_B$ is consistent to σ_{s_n} and we can show

$$\sqrt{n} L(\widehat{V}_B, \alpha) = 2z_{\alpha/2}\sigma_{s_n} + o_p(1), \tag{7.7}$$

and

$$n\mathbb{E}L^2(\widehat{V}_B, \alpha) = 4z_{\alpha/2}^2\sigma_{s_n}^2 + o(1). \tag{7.8}$$

Before presenting our main results, we introduce the following condition.

(C7) For any $\boldsymbol{x} \in \mathbb{X}$ with $\tau(\boldsymbol{x}) = 0$, assume the following holds:

$$\mathbb{P}(\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}) > 0) \to 1/2, \qquad \text{as } |\mathscr{I}| \to \infty.$$

Assume $\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}) - \tau(\boldsymbol{x})$ is asymptotically normal, i.e,

$$\frac{\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}) - \tau(\boldsymbol{x})}{\sigma^*_{|\mathscr{I}|}} \xrightarrow{d} N(0,1), \tag{7.9}$$

for some sequence $\sigma^*_n \to 0$. Then for any $\boldsymbol{x}$ that satisfies $\tau(\boldsymbol{x}) = 0$, it follows from the definition of weak convergence that

$$\mathbb{P}(\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}) > 0) = \mathbb{P}(\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}) - \tau(\boldsymbol{x}) > 0) = \mathbb{P}\left(\frac{\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}) - \tau(\boldsymbol{x})}{\sigma^*_{|\mathscr{I}|}} > 0 \right) \to \frac{1}{2}.$$

Notice that the condition (7.9) holds for a wide variety of nonparametric estimators $\widehat{\tau}_{\mathscr{I}}$ computed by kernel smoothing methods (Härdle 1990), spline methods (Zhou et al. 1998), kernel ridge regression (Zhao et al. 2016), random forests (Wager and Athey 2018), etc.

For simplicity, throughout this section, we assume the following semiparametric regression model for Y_0:

$$Y_0 = h(0, \boldsymbol{X}_0) + A_0 \tau(\boldsymbol{X}_0) + e_0, \tag{7.10}$$

where e_0 is a mean zero random error term independent of $\boldsymbol{X}_0, A_0$. Let $\sigma_0^2 = \mathbb{VAR}(e_0) > 0$.

7.3.1 Comparison with the online one-step estimator

For $j = 1, \ldots, n$, let $\mathscr{I}_{(j)} = \{1, 2, \ldots, j\}$. Let $\{l_n\}_n$ be a sequence of nonnegative integer with $l_n < n$. The online one-step estimator is defined as

$$\widehat{V}^{on} = \left(\sum_{j=l_n}^{n-1} \widetilde{\sigma}^{-1}_{\mathscr{I}_{(j)}} \right)^{-1} \left(\sum_{j=l_n}^{n-1} \widetilde{\sigma}^{-1}_{\mathscr{I}_{(j)}} \widehat{V}_{\{j+1\}}(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}}) \right),$$

where $\widetilde{\sigma}^2_{\mathscr{I}_{(j)}}$ stands for some consistent estimator of

$$\widetilde{\sigma}^2_0(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}}) = \mathbb{VAR}\left(\widehat{V}_{\{j+1\}}(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}}) \middle| \{O_i\}_{i \in \mathscr{I}_{(j)}} \right),$$

computed based on the observations $\{O_i\}_{i\in\mathscr{I}_{(j)}}$.

Under the conditions in Theorem 2 of Luedtke and van der Laan (2016), it follows from martingale central limit theorem that

$$\frac{\sqrt{n-l_n}(\widehat{V}^{on}-V_0)}{\widehat{\sigma}^{on}} \xrightarrow{d} N(0,1),$$

where $\widehat{\sigma}^{on}=\{\sum_{j=l_n}^{n-1}\widetilde{\sigma}_{\mathscr{I}_{(j)}}^{-1}/(n-l_n)\}^{-1}$. The corresponding two-sided CI for V_0 is given by

$$\left[\widehat{V}^{on}-z_{\alpha/2}\frac{\widehat{\sigma}^{on}}{\sqrt{n-l_n}}, \widehat{V}^{on}+z_{\alpha/2}\frac{\widehat{\sigma}^{on}}{\sqrt{n-l_n}}\right]. \tag{7.11}$$

Assume $l_n \to \infty$, under the same conditions as Theorem 7.2.2, we can show that

$$\widehat{\sigma}^{on}=\left(\frac{\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)}{n-l_n}\right)^{-1}+o_p(1). \tag{7.12}$$

The first term on the RHS of the above expression is a random variable depending on $\{O_i\}_{i\in\mathscr{I}_{(n-1)}}$. In the nonregular cases, the conditional variance $\widetilde{\sigma}_0^2(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)$ may not converge to a deterministic quantity. As a result, the length of (7.11) will fluctuate randomly. Therefore, we focus on comparing the average squared length of (7.11) with that of our proposed CI.

When $\{\widetilde{\sigma}_{\mathscr{I}_{(j)}}\}_{j=l_n,\dots,n-1}$ and $\{\widetilde{\sigma}_0(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)\}_{j=l_n,\dots,n-1}$ are uniformly bounded from above, it follows from (7.12) that

$$\mathbb{E}(\widehat{\sigma}^{on})^2=\mathbb{E}\left(\frac{\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)}{n-l_n}\right)^{-2}+o(1).$$

When $l_n=o(n)$, the length of (7.11) satisfies

$$n\mathbb{E}L^2(\widehat{V}^{on},\alpha)=4z_{\alpha/2}^2\mathbb{E}\left(\frac{\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)}{n-l_n}\right)^{-2}+o(1). \tag{7.13}$$

Theorem 7.3.1 *Assume* (7.8), (7.10), (7.13), *C1-C7 hold,* $s_n, l_n \to \infty$. *Then, we have*

$$n\mathbb{E}L^2(\widehat{V}^{on},\alpha)-n\mathbb{E}L^2(\widehat{V}_B,\alpha)\geq\frac{z_{\alpha/2}^2\sigma_0^2}{2\epsilon_0^2}\mathbb{P}\{\tau(\boldsymbol{X}_0)=0\}+o(1),$$

where ϵ_0 *is defined in Condition C3.*

Theorem 7.3.1 implies that the expected squared length of (7.11) is asymptotically larger than that of the proposed CI in the nonregular cases. The difference depends on $\mathbb{P}\{\tau(\boldsymbol{X}_0)=0\}$, which measures the degree of nonregularity.

In the following, we sketch a few lines to see why our proposed CI achieves smaller squared length on average. By the delta method, we have

$$\mathbb{E}\left(\frac{\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)}{n-l_n}\right)^{-2}\approx\left(\frac{\sum_{j=l_n}^{n-1}\{\mathbb{E}\widetilde{\sigma}_0^2(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)\}^{-1/2}}{n-l_n}\right)^{-2}.$$

Under the given conditions, $\mathbb{E}\widetilde{\sigma}_0^2(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)$ converges to a fixed function as $j\to\infty$. Since $s_n,l_n\to\infty$, we have

$$\begin{aligned}\left(\frac{\sum_{j=l_n}^{n-1}\{\mathbb{E}\widetilde{\sigma}_0^2(\widehat{d}_{\mathscr{I}_{(j)}};\pi,h)\}^{-1/2}}{n-l_n}\right)^{-2} &= \left(\frac{\sum_{j=l_n}^{n-1}\{\mathbb{E}\widetilde{\sigma}_0^2(\widehat{d}_{\mathscr{I}_{(s_n)}};\pi,h)\}^{-1/2}}{n-l_n}\right)^{-2}+o(1)\\ &= \mathbb{E}\widetilde{\sigma}_0^2(\widehat{d}_{\mathscr{I}_{(s_n)}};\pi,h)+o(1).\end{aligned}$$

By definition, we have $\sigma_{s_n}^2=\widetilde{\sigma}_0^2(p_{s_n};\pi,h)$ and $p_{s_n}(\boldsymbol{x})=\mathbb{E}\widehat{d}_{\mathscr{I}_{(s_n)}}(\boldsymbol{x})$. The function $\widetilde{\sigma}_0^2(d;\pi,h)$ is convex in d. Therefore, it follows from Jensen's inequality that

$$\mathbb{E}\widetilde{\sigma}_0^2(\widehat{d}_{\mathscr{I}_{(s_n)}};\pi,h)\geq\widetilde{\sigma}_0^2(\mathbb{E}\widehat{d}_{\mathscr{I}_{(s_n)}};\pi,h).$$

This together with (7.8) and (7.13) yields that $n\mathbb{E}L^2(\widehat{V}^{on},\alpha)\geq n\mathbb{E}L^2(\widehat{V}_B,\alpha)+o(1)$.

7.3.2 Beyond oracle property

In this section, we compare the proposed CI with the CI based on the oracle method. The oracle knew the set of optimal treatment regimes $\mathscr{D}^{opt}$ ahead of time. When functions π and h are known, we can estimate V_0 by $\widehat{V}_{\mathscr{I}_0}(d^{opt};\pi,h)$ for an arbitrary $d^{opt}\in\mathscr{D}^{opt}$. To deal with unknown propensity score and conditional mean functions, we can construct our estimator based on the following cross-validation procedure:

Step 1 Input observations $\{O_i\}_{i\in\mathscr{I}_0}$, $0<\alpha<1$.

Step 2 Randomly partition $\mathscr{I}_0$ into 2 disjoint subsets $\mathscr{I}_1$ and $\mathscr{I}_2$ of equal sizes, assuming the sample size n is an even integer.

Step 3 Obtain the estimators $\widehat{\pi}_{\mathscr{I}_j}$ and $\widehat{h}_{\mathscr{I}_j}$ for $j = 1,2$. For any $d^{opt} \in \mathscr{D}^{opt}$, compute

$$\widehat{V}^{or}(d^{opt}) = \frac{1}{2}\left(\widehat{V}_{\mathscr{I}_1}(d^{opt}; \widehat{\pi}_{\mathscr{I}_2}, \widehat{h}_{\mathscr{I}_2}) + \widehat{V}_{\mathscr{I}_2}(d^{opt}; \widehat{\pi}_{\mathscr{I}_1}, \widehat{h}_{\mathscr{I}_1})\right),$$

$$\widehat{\sigma}^{or}(d^{opt}) = \left\{\frac{1}{n-1}\sum_{j=1}^{2}\sum_{i\in\mathscr{I}_j}\left(\widehat{V}_{\{i\}}(d^{opt}; \widehat{\pi}_{\mathscr{I}_j^c}, \widehat{h}_{\mathscr{I}_j^c}) - \widehat{V}(d^{opt})\right)^2\right\}^{1/2}.$$

Step 4 Output

$$\left[\widehat{V}^{or}(d^{opt}) - \frac{z_{\alpha/2}\widehat{\sigma}^{or}(d^{opt})}{\sqrt{n}}, \widehat{V}^{or}(d^{opt}) + \frac{z_{\alpha/2}\widehat{\sigma}^{or}(d^{opt})}{\sqrt{n}}\right]. \tag{7.14}$$

The CI in (7.14) is valid. Under conditions (7.4), (7.5) and (7.6), we can show that

$$\widehat{\sigma}^{or2}(d^{opt}) = \underbrace{\mathbb{VAR}\left(\frac{\mathrm{g}\{A_0, d^{opt}(\boldsymbol{X}_0)\}}{\pi(A_0, \boldsymbol{X}_0)}\{Y_0 - h(A_0, \boldsymbol{X}_0)\} + h\{d^{opt}(\boldsymbol{X}_0), \boldsymbol{X}_0\}\right)}_{\widetilde{\sigma}_0^2(d^{opt};\pi,h)} + o_p(1).$$

Thus, the length of (7.14) satisfies

$$\sqrt{n}L\{\widehat{V}^{or}(d^{opt}), \alpha\} = 2z_{\alpha/2}\widetilde{\sigma}_0(d^{opt}; \pi, h) + o_p(1). \tag{7.15}$$

Theorem 7.3.2 *Assume* (7.7), (7.10), (7.15), *C1-C7 hold and* $s_n \to \infty$. *Assume that* $\min_{a=0,1}\pi(a, \boldsymbol{x}) \ge 1/4$, $\forall \boldsymbol{x} \in \mathbb{X}$ *with* $\tau(\boldsymbol{x}) = 0$. *Then, we have*

$$\inf_{d^{opt}\in\mathscr{D}^{opt}} nL^2\{\widehat{V}^{or}(d^{opt}), \alpha\} - nL^2(\widehat{V}_B, \alpha) \ge c^{**}z_{\alpha/2}^2\sigma_0^2\mathbb{P}\{\tau(\boldsymbol{X}_0) = 0\} + o(1),$$

where

$$c^{**} = \inf_{\substack{a=0,1\\ \boldsymbol{x}\in\mathbb{X}:\tau(\boldsymbol{x})=0}}\left(\frac{3}{\pi(a,\boldsymbol{x})} - \frac{1}{\pi(1-a,\boldsymbol{x})}\right) \ge 0.$$

In randomized studies, we usually have $\pi(1,\boldsymbol{x}) = 1 - \pi(0,\boldsymbol{x}) = \pi^*, \forall \boldsymbol{x} \in \mathbb{X}$ for some constant $\pi^* > 0$. The condition $\min_{a=0,1}\pi(a,\boldsymbol{x}) \ge 1/4$ thus holds if $1/4 \le \pi^* \le 3/4$. Theorem 7.3.2 implies that the proposed CI is asymptotically narrower than (7.14) in the nonregular cases. As discussed in the introduction, this is due to the subagging procedure, which averages over estimated OTRs in the nonregular cases, resulting in a smoothed treatment regime $p_{s_n}(\cdot)$. To give a more formal explanation, let's assume $\tau(\boldsymbol{x}) = 0$ and $\pi(1,\boldsymbol{x}) = \pi^*$, $\forall \boldsymbol{x}$.

In addition, we assume we know the true propensity score and conditional mean functions and set $\widehat{\pi}_{\mathscr{I}} = \pi^*$ and $\widehat{h}_{\mathscr{I}} = h$. Then it follows from Lemma 1.2.1 that $\mathbb{E}\widehat{V}_{\mathscr{I}_0}(d) = V_0$ for any regime d. By (7.10), we have

$$\begin{aligned} n\mathbb{VAR}\{\widehat{V}^{or}(d)\} = n\mathbb{VAR}\{\widehat{V}_{\mathscr{I}_0}(d)\} = \sigma_0^2\mathbb{E}\left(\frac{d^2(\boldsymbol{X}_0)}{\pi^*} + \frac{\{1-d(\boldsymbol{X}_0)\}^2}{1-\pi^*}\right) \\ + \quad \mathbb{VAR}\{h(0,\boldsymbol{X}_0)\} = \sigma_0^2\mathbb{E}\left(\frac{d(\boldsymbol{X}_0)}{\pi^*} + \frac{1-d(\boldsymbol{X}_0)}{1-\pi^*}\right) + \mathbb{VAR}\{h(0,\boldsymbol{X}_0)\} \\ \geq \quad \sigma_0^2 \min\left(\frac{1}{\pi^*}, \frac{1}{1-\pi^*}\right) + \mathbb{VAR}\{h(0,\boldsymbol{X}_0)\}. \end{aligned}$$

As for our proposed value estimator, it follows from Condition (A7) that

$$\begin{aligned} n\mathbb{VAR}\{\widehat{V}_B\} \approx \sigma_{s_n}^2 = \sigma_0^2\mathbb{E}\left(\frac{p_{s_n}^2(\boldsymbol{X}_0)}{\pi^*} + \frac{\{1-p_{s_n}(\boldsymbol{X}_0)\}^2}{1-\pi^*}\right) + \mathbb{VAR}\{h(0,\boldsymbol{X}_0)\} \\ \approx \sigma_0^2\left(\frac{1}{4\pi^*} + \frac{1}{4(1-\pi^*)}\right) + \mathbb{VAR}\{h(0,\boldsymbol{X}_0)\}. \end{aligned}$$

When $1/4 \leq \pi^* \leq 3/4$, we have $1/(4\pi^*) + 1/\{4(1-\pi^*)\} \leq \min\{1/\pi^*, 1/(1-\pi^*)\}$. This implies that our proposed estimator is more efficient than the oracle estimator.

In the regular cases where $\mathbb{P}\{\tau(\boldsymbol{X}_0) = 0\} = 0$, we can show $\widehat{V}_B = \widehat{V}^{or}(d^{opt}) + o_p(n^{-1/2})$ and $\widehat{\sigma}_B = \widehat{\sigma}^{or}(d^{opt}) + o_p(1)$ for any $d^{opt} \in \mathscr{D}^{opt}$. This means the proposed CI is asymptotically equivalent to the CI of the oracle method in the regular cases.

7.4 Multiple time point study

We now consider the setting described in (1.3) and focus on constructing CIs for the optimal value function $V_0 = \max_d \mathbb{E}Y_0^*(d)$, based on the observed dataset:

$$\left\{O_i = \left(\boldsymbol{X}_i^{(1)}, A_i^{(1)}, \boldsymbol{X}_i^{(2)}, A_i^{(2)}, \ldots, \boldsymbol{X}_i^{(K)}, A_i^{(K)}, Y_i\right) : i = 1, \ldots, n\right\}.$$

For $k = 1, \ldots, K$, $i = 0, 1, \ldots, n$, let

$$\bar{\boldsymbol{X}}_i^{(k)} = (\boldsymbol{X}_i^{(1)}, \ldots, \boldsymbol{X}_i^{(k)}) \quad \text{and} \quad \bar{\boldsymbol{A}}_i^{(k)} = (A_i^{(1)}, \ldots, A_i^{(k)}).$$

Define the propensity score function $\pi_k(\bar{\boldsymbol{a}}_k, \bar{\boldsymbol{x}}_k) = \mathbb{P}(A_0^{(k)} = a_k | \bar{\boldsymbol{X}}_0^{(k)} = \bar{\boldsymbol{x}}_k, \bar{\boldsymbol{A}}_0^{(k-1)} = \bar{\boldsymbol{a}}_{k-1})$ for $k = 2, \ldots, K$ and $\pi_1(a_1, \boldsymbol{x}_1) = \mathbb{P}(A_0^{(1)} = a_k | \boldsymbol{X}_0^{(1)} = \boldsymbol{x}_1)$. For any dynamic treatment regime

$d=\{d_k\}_{k=1}^K$, define

$$\begin{aligned}\widehat{V}_i^{(K)}(d;\pi^*,h^*) &= \frac{\mathrm{g}\{A_i^{(K)}, d_K(\bar{\boldsymbol{A}}_i^{(K-1)}, \bar{\boldsymbol{X}}_i^{(K)})\}}{\pi_K^*(\bar{\boldsymbol{A}}_i^{(K)}, \bar{\boldsymbol{X}}_i^{(K)})}\{Y_i - h_K^*(\bar{\boldsymbol{A}}_i^{(K)}, \bar{\boldsymbol{X}}_i^{(K)})\} \\ &+ h_K^*[\{\bar{\boldsymbol{A}}_i^{(K-1)}, d_K(\bar{\boldsymbol{A}}_i^{(K-1)}, \bar{\boldsymbol{X}}_i^{(K)})\}, \bar{\boldsymbol{X}}_i^{(K)}],\end{aligned}$$

and

$$\begin{aligned}\widehat{V}_i^{(k)}(d;\pi^*,h^*) &= \frac{\mathrm{g}\{A_i^{(k)}, d_k(\bar{\boldsymbol{A}}_i^{(k-1)}, \bar{\boldsymbol{X}}_i^{(k)})\}}{\pi_k^*(\bar{\boldsymbol{A}}_i^{(k)}, \bar{\boldsymbol{X}}_i^{(k)})}\{\widehat{V}_i^{(k+1)}(d;\pi^*,h^*) - h_k^*(\bar{\boldsymbol{A}}_i^{(k)}, \bar{\boldsymbol{X}}_i^{(k)})\} \\ &+ h_k^*[\{\bar{\boldsymbol{A}}_i^{(k-1)}, d_k(\bar{\boldsymbol{A}}_i^{(k-1)}, \bar{\boldsymbol{X}}_i^{(k)})\}, \bar{\boldsymbol{X}}_i^{(k)}],\end{aligned}$$

for $k=K-1,\ldots,2$, $i=0,1,\ldots,n$, where $\pi^*\equiv\{\pi_k^*\}_{k=1}^K$ and $h^*\equiv\{h_k^*\}_{k=1}^K$ denote the estimated propensity score and conditional mean functions. For any $\mathscr{I}\subseteq\mathscr{I}_0$, define the following augmented propensity-score weighted estimator,

$$\widehat{V}_{\mathscr{I}}(d;\pi^*,h^*) = \sum_{i\in\mathscr{I}}\frac{1}{|\mathscr{I}|}\Bigg(\frac{\mathrm{g}\{A_i^{(1)}, d_1(\bar{\boldsymbol{X}}_i^{(1)})\}}{\pi_1^*(A_i^{(1)}, \bar{\boldsymbol{X}}_i^{(1)})}\{\widehat{V}_i^{(2)}(d;\pi^*,h^*) - h_1^*(A_i^{(1)}, \bar{\boldsymbol{X}}_i^{(1)})\} \\ + h_1^*\{d_1(\bar{\boldsymbol{X}}_i^{(1)}), \bar{\boldsymbol{X}}_i^{(1)}\}\Bigg).$$

Notice that $\mathbb{E}\widehat{V}_{\mathscr{I}}(d^{opt};\pi^*,h^*)=V_0$ when either $\pi^*=\pi$ or $h^*=h$.

For any $\mathscr{I}\subseteq\mathscr{I}_0$, let $\{\widehat{\tau}_{\mathscr{I},k}\}_{k=1}^K$, $\{\widehat{h}_{\mathscr{I},k}\}_{k=1}^K$, $\{\widehat{\pi}_{\mathscr{I},k}\}_{k=1}^K$ denote some consistent estimators for $\{\tau_k\}_{k=1}^K$, $\{h_k\}_{k=1}^K$ and $\{\pi_k\}_{k=1}^K$, computed based on the sub-dataset $\{O_i\}_{i\in\mathscr{I}}$. Define the estimated treatment regime $\widehat{d}_{\mathscr{I},k}(\cdot,\cdot)=\mathbb{I}\{\widehat{\tau}_{\mathscr{I},k}(\cdot,\cdot)>0\}$, $k=2,\ldots,K$ and $\widehat{d}_{\mathscr{I},1}(\cdot)=\mathbb{I}\{\widehat{\tau}_{\mathscr{I},1}(\cdot)>0\}$. Define the set

$$\mathscr{S}_{N_0,s_n} = \left\{\mathscr{I}\subseteq\mathscr{I}_0 : |\mathscr{I}| = s_n, \min_{a_1,\ldots,a_K\in\{0,1\}}\sum_{i\in\mathscr{I}}\mathbb{I}(A_i^{(1)}=a_1,\ldots,A_i^{(K)}=a_K)\geq N_0\right\},$$

for some integers $s_n>N_0>0$. We summarize our procedure in the following algorithm.

Step 1 Input observations $\{O_i\}_{i\in\mathscr{I}_0}$, $0<\alpha<1$ and integers s_n, N_0 and B.

Step 2 For $b=1,\ldots,B$,

(i) Draw a subset $\mathscr{I}_b$ uniformly from $\mathscr{S}_{N_0,s_n}$.

(ii) Randomly partition $\mathscr{I}_b^c$ into 2 disjoint subsets $\mathscr{I}_b^{c(1)}$ and $\mathscr{I}_b^{c(2)}$ of equal sizes t_n.

(iii) For $j=1,2$, let $\mathscr{I}_b^{(j)}=\mathscr{I}_b\cup\mathscr{I}_b^{c(j)}$. Obtain the estimators $\widehat{d}_{\mathscr{I}_b}=\{\widehat{d}_{\mathscr{I}_b,k}\}_{k=1}^K$, $\widehat{\pi}_{\mathscr{I}_b^{(1)}}=\{\widehat{\pi}_{\mathscr{I}_b^{(1)},k}\}_{k=1}^K$, $\widehat{\pi}_{\mathscr{I}_b^{(2)}}=\{\widehat{\pi}_{\mathscr{I}_b^{(2)},k}\}_{k=1}^K$, $\widehat{h}_{\mathscr{I}_b^{(1)}}=\{\widehat{h}_{\mathscr{I}_b^{(1)},k}\}_{k=1}^K$ and $\widehat{h}_{\mathscr{I}_b^{(2)}}=\{\widehat{\pi}_{\mathscr{I}_b^{(2)},k}\}_{k=1}^K$.

Step 3 Compute

$$\widehat{V}_B = \frac{1}{2B}\sum_{b=1}^{B}\Big(\widehat{V}_{\mathscr{I}_b^{c(2)}}(\widehat{d}_{\mathscr{I}_b};\widehat{\pi}_{\mathscr{I}_b^{(1)}},\widehat{h}_{\mathscr{I}_b^{(1)}}) + \widehat{V}_{\mathscr{I}_b^{c(1)}}(\widehat{d}_{\mathscr{I}_b};\widehat{\pi}_{\mathscr{I}_b^{(2)}},\widehat{h}_{\mathscr{I}_b^{(2)}})\Big),$$

and $\widehat{\sigma}_B^2 = \widehat{s.e}^2(\{\widehat{V}^{(i)}\}_{i=1}^n)$ where

$$\widehat{V}^{(i)} = \frac{1}{n^{(i)}}\sum_{j=1}^{2}\sum_{b=1}^{B}\widehat{V}_{\{i\}}(\widehat{d}_{\mathscr{I}_b};\widehat{\pi}_{\mathscr{I}_b^{(j)}},\widehat{h}_{\mathscr{I}_b^{(j)}})\mathbb{I}(i\notin\mathscr{I}_b^{(j)}),$$

and $n^{(i)} = \sum_{b=1}^{B}\mathbb{I}(i\notin\mathscr{I}_b)$.

Step 4 Output

$$\left[\widehat{V}_B - \frac{z_{\alpha/2}\widehat{\sigma}_B}{\sqrt{n}}, \widehat{V}_B + \frac{z_{\alpha/2}\widehat{\sigma}_B}{\sqrt{n}}\right]. \tag{7.16}$$

Shi et al. (2019d) proved the CI in (7.16) achieves nominal coverage. The technical conditions imposed are very similar to C1-C6 in single-stage studies.

7.5 Simulations

7.5.1 Point treatment study

We consider simulation studies based on the following model:

$$Y_0 = \Phi(X_0^{(1)}, X_0^{(2)}) + A_0\tau(X_0^{(1)}, X_0^{(2)}) + e_0,$$

where the covariate $X_0^{(1)}$ and the treatment A_0 are generated from Ber(0.5) and Ber($0.5+0.1X_0^{(1)}$), respectively, where Ber(p_0) stands for the Bernoulli distribution with probability of success p_0. The random error term e_0 satisfies $\mathbb{E}(e_0|A_0, X_0^{(1)}, X_0^{(2)}) = 0$. We consider six scenarios. In Scenario (A) and (B), $X_0^{(2)}$ is generated from Ber(0.5), and

$$e_0 \sim \text{Ber}\{\Phi(X_0^{(1)}, X_0^{(2)}) + A_0\tau(X_0^{(1)}, X_0^{(2)})\} - \Phi(X_0^{(1)}, X_0^{(2)}) - A_0\tau(X_0^{(1)}, X_0^{(2)}).$$

In Scenario (C)-(F), $X_0^{(2)}$ follows a uniform distribution on the interval $[-2,2]$, and $e_0 \sim N(0, 0.25)$ is independent of A_0, $X_0^{(1)}$ and $X_0^{(2)}$. In addition, $X_0^{(1)}$ and $X_0^{(2)}$ are independently

Table 7.1: Simulation setting

	(A)	(B)	(C)	(D)	(E)	(F)
$\Phi(x_1, x_2)$	0.3	0.3	x_2^2	x_2^2	x_2^2	x_2^2
$\tau(x_1, x_2)$	$0.4\mathbb{I}(x_1=0)$	0.4	$x_1 x_2^2$	$x_2^2-4/3$	$2x_1\cos(\pi x_2/4)$	$2\cos(\pi x_2/4)-4/\pi$
V_0	0.5	0.7	2	1.85	1.97	1.60

generated in all scenarios. Table 7.1 summarizes the information of the baseline function, the contrast function and the optimal value V_0 under different scenarios. In all scenarios, V_0 can be explicitly calculated. The OTR is not uniquely defined in Scenario (A), (C) and (E), since the contrast functions in these scenarios satisfy

$$\mathbb{P}\{\tau(X_0^{(1)}, X_0^{(2)})=0\} = \mathbb{P}(X_0^{(1)}=0) = \frac{1}{2}.$$

On the contrary, we have $\mathbb{P}\{\tau(X_0^{(1)}, X_0^{(2)})=0\}=0$ in the remaining three scenarios. For each scenario, we further consider two different sample sizes, $n=500$ and $n=1000$. This yields a total of 12 settings.

Comparison is made among the following three methods:

(i) The proposed CI in (7.3).

(ii) The CI constructed by the online one-step method (7.11).

(iii) The CI constructed by the oracle method (7.14) with $d^{opt}=d^{opt,0}$ (see (1.2)). (Notice that $d^{opt,0}$ is unknown in practice, we implement this method for comparison purposes only.)

All three methods require estimation of the propensity score and conditional mean functions. For scenario (A) and (B), we use the nonparametric maximum likelihood estimator to estimate these functions. For scenario (C)-(F), we estimate these functions using cubic B-splines. More specifically, for $a=0,1$, define

$$\widehat{\boldsymbol{\xi}}_{\mathcal{I}}^{\pi,a} = \arg\min_{\boldsymbol{\xi}} \sum_{i\in\mathcal{I}} \left(A_i - \sum_{j=1}^{K_0+4} N_j(X_i^{(2)})\xi_j \right)^2 \mathbb{I}(X_i^{(1)}=a),$$

Table 7.2: ACP and AL of the CIs with standard errors in parenthesis

Setting (A)	Proposed		Online		Oracle	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
500	93.6 (0.8)	11.1 (0.02)	94.1 (0.7)	12.8 (0.02)	94.0 (0.8)	13.1 (0.02)
1000	93.7 (0.8)	7.8 (0.01)	93.9 (0.8)	8.8 (0.01)	94.1 (0.7)	9.0 (0.01)
Setting (B)	Proposed		Online		Oracle	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
500	95.1 (0.7)	11.1 (0.01)	93.9 (0.8)	11.5 (0.02)	95.4 (0.7)	11.2 (0.01)
1000	95.3 (0.7)	7.8 (0.01)	95.5 (0.7)	7.9 (0.01)	95.3 (0.7)	7.8 (0.01)
Setting (C)	Proposed		Online		Oracle	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
500	94.1 (0.7)	36.8 (0.04)	92.7 (0.8)	41.1 (0.06)	94.3 (0.7)	38.0 (0.08)
1000	93.9 (0.7)	25.9 (0.02)	93.4 (0.8)	27.4 (0.03)	94.5 (0.7)	26.3 (0.02)
Setting (D)	Proposed		Online		Oracle	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
500	95.1 (0.7)	36.6 (0.04)	93.2 (0.8)	40.6 (0.05)	93.7 (0.8)	38.0 (0.20)
1000	94.9 (0.7)	25.7 (0.02)	93.1 (0.8)	27.1 (0.02)	94.2 (0.7)	25.9 (0.02)
Setting (E)	Proposed		Online		Oracle	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
500	94.7 (0.7)	22.6 (0.02)	88.2 (1.0)	25.8 (0.03)	94.0 (0.8)	24.6 (0.09)
1000	95.5 (0.7)	15.9 (0.01)	91.9 (0.9)	17.2 (0.01)	95.3 (0.7)	16.7 (0.02)
Setting (F)	Proposed		Online		Oracle	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
500	92.4 (0.8)	21.3 (0.04)	87.5 (1.0)	23.3 (0.03)	93.8 (0.8)	24.3 (0.36)
1000	94.3 (0.7)	14.8 (0.01)	90.8 (0.9)	15.6 (0.01)	94.1 (0.7)	15.3 (0.03)

and

$$\widehat{\boldsymbol{\xi}}_{\mathscr{I}}^{h_1,a} = \arg\min_{\boldsymbol{\xi}} \sum_{i\in\mathscr{I}} \left(Y_i - \sum_{j=1}^{K_0+4} N_j(X_i^{(2)})\xi_j \right)^2 \mathbb{I}(A_i = 1, X_i^{(1)} = a),$$
$$\widehat{\boldsymbol{\xi}}_{\mathscr{I}}^{h_0,a} = \arg\min_{\boldsymbol{\xi}} \sum_{i\in\mathscr{I}} \left(Y_i - \sum_{j=1}^{K_0+4} N_j(X_i^{(2)})\xi_j \right)^2 \mathbb{I}(A_i = 0, X_i^{(1)} = a),$$

where $N_1(\cdot),\ldots,N_{K_0+4}(\cdot)$ stand for the cubic B-spline basis, and K_0 denotes the number of interior knots. Given K_0, the interior knots are placed at equally spaced sample quantiles of $\{X_i^{(2)}\}_{i\in\mathscr{I}_0}$. The hyperparameter K_0 is selected via 5-fold cross-validation. After computing $\widehat{\boldsymbol{\xi}}_{\mathscr{I}}^{\pi,a}$, $\widehat{\boldsymbol{\xi}}_{\mathscr{I}}^{h_1,a}$ and $\widehat{\boldsymbol{\xi}}_{\mathscr{I}}^{h_0,a}$, we set

$$\widehat{\pi}_{\mathscr{I}}(1,\boldsymbol{x}_1) = \min\left(\sum_{\substack{a=\{0,1\}\\ 1\le j\le K+4}} \mathbb{I}(x_1^{(1)} = a) N_j(x_1^{(2)}) \widehat{\xi}_{\mathscr{I},j}^{\pi,a}, 0.05 \right), \tag{7.17}$$
$$\widehat{\pi}_{\mathscr{I}}(0,\boldsymbol{x}_1) = \min\{1-\widehat{\pi}_{\mathscr{I}}(1,\boldsymbol{x}_1), 0.05\}, \tag{7.18}$$
$$\widehat{h}_{\mathscr{I}}(1,\boldsymbol{x}_1) = \sum_{a=0,1} \mathbb{I}(x_1^{(1)} = a) \sum_{j=1}^{K+4} N_j(x_1^{(2)}) \widehat{\xi}_{\mathscr{I},j}^{h_1,a},$$
$$\widehat{h}_{\mathscr{I}}(0,\boldsymbol{x}_1) = \sum_{a=0,1} \mathbb{I}(x_1^{(1)} = a) \sum_{j=1}^{K+4} N_j(x_1^{(2)}) \widehat{\xi}_{\mathscr{I},j}^{h_0,a}.$$

Truncation is used in (7.17) and (7.18) to avoid extreme weights, resulting in a more "stabilized" value estimator. The estimated contrast function is defined as

$$\widehat{\tau}_{\mathscr{I}}(\boldsymbol{x}_1) = \widehat{h}_{\mathscr{I}}(1,\boldsymbol{x}_1) - \widehat{h}_{\mathscr{I}}(0,\boldsymbol{x}_1).$$

To calculate the CI in (7.3), we set $s_n = \lfloor n^{6/7} \rfloor$ where $\lfloor z \rfloor$ denotes the largest integer smaller than or equal to z. Such a choice of s_n satisfies the condition $s_n = o(n)$. To implement the online one-step method, we need to specify l_n. In general, the length of the CI in (7.11) increases as l_n increases. Nonetheless, l_n should be large enough to guarantee that the bias $V(\widehat{d}_{\mathscr{I}_{l_n}}) - V_0$ is negligible. In Scenario (A) and (B), we set $l_n = 50$. In Scenario (C)-(F), we find that when $l_n = 50$, the resulting CIs have very poor coverage probabilities. Therefore, we set

$l_n = 100$ in these scenarios. The variance estimator $\widetilde{\sigma}^2_{\mathscr{I}_{(j)}}$ is computed by

$$\widetilde{\sigma}^2_{\mathscr{I}_{(j)}} = \frac{1}{j(j-1)} \sum_{i=1}^{j} \left(\widehat{V}_{\{i\}}(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}}) - \widehat{V}_{\mathscr{I}_{(j)}}(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}}) \right)^2.$$

We implement the simulation program in R. Some subroutines are written in C with the GNU Scientific Library (Galassi et al. 2015) to facilitate the computation.

Reported in Table 7.2 are the average coverage probability (ACP) and average length (AL) of the CIs in (i)-(iii). Results are aggregated over 1000 replications. It can be seen that all three CIs achieve nominal coverage in Scenario (A)-(D). However, the CIs based on the online one-step method and the oracle method are wider than the proposed CIs in all cases. Take Scenario (A) as an example. ALs of our proposed method are at least 13% smaller than other competing methods. In Scenario (E) and (F), ACPs of the online one-step method are smaller than 90% when $n = 500$. In contrast, ACPs of the proposed CIs are close to the nominal level in all cases. In addition, the proposed CIs achieve smaller ALs in these scenarios.

Notice that in Scenario (B), (D) and (F), the contrast function is almost surely nonzero. In theory, when $l_n = o(n)$, the lengths of all three CIs should be asymptotically the same. However, it can been seen from Table 7.2 that in finite samples, ALs of the CIs based on our proposed method are always smaller than other competing methods.

7.5.2 Multiple time point study

Consider the following model:

$$\begin{aligned} Y_0 = \Phi(X_0^{(1)(1)}, X_0^{(2)(1)}, A_0^{(1)}, X_0^{(2)}) + A_0^{(2)} \tau(X_0^{(1)(1)}, X_0^{(2)(1)}, A_0^{(1)}, X_0^{(2)}) + e_0^{(2)}, \\ X_0^{(2)} = A_0^{(1)} X_0^{(1)(1)} + e_0^{(1)}, \end{aligned}$$

where $X_0^{(1)(1)}$ and $X_0^{(2)(1)}$ are the baseline covariates, $A_0^{(1)}$ and $A_0^{(2)}$ denote the first and second treatment a patient receives at t_1 and t_2, $X_0^{(2)}$ stands for the intermediate covariate collected between t_1 and t_2. Variables $A_0^{(1)}$, $A_0^{(2)}$, $X_0^{(1)(1)}$, $X_0^{(2)(1)}$, $e_0^{(1)}$ and $e_0^{(2)}$ are all independent. In addition, we assume $A_0^{(1)}, A_0^{(2)} \sim \text{Ber}(0.5)$, $e_0^{(1)}, e_0^{(2)} \sim N(0, 0.25)$ and $X_0^{(1)(1)}, X_0^{(2)(1)} \sim \text{Unif}[-2, 2]$ where $\text{Unif}[a, b]$ denotes the uniform distribution on the interval $[a, b]$.

We consider three scenarios. The functional forms of Φ and τ and the optimal value

Table 7.3: Simulation setting

	(G)	(H)	(I)
$\Phi(x_1^{(1)}, x_1^{(2)}, a_1, x_2)$	$x_1^{(1)2} - a_1(0.25 + x_1^{(1)2})$	x_2^2	0
$\tau(x_1^{(1)}, x_1^{(2)}, a_1, x_2)$	$a_1 x_2^2$	0	x_2^2
V_0	1.33	1.58	1.58

function V_0 under these scenarios are reported in Table 7.3. In Scenario (G), we have

$$h_1(a_1, \boldsymbol{x}_1) = \mathbb{E}\{A_0^{(1)} X_0^{(2)} + \Phi(X_0^{(1),(1)}, X_0^{(2),(1)}, A_0^{(1)}, X_0^{(2)}) | X_0^{(1),(1)} = x_1^{(1)}, X_0^{(2),(1)} = x_1^{(2)}, A_0^{(1)} = a_1\} = a_1\{0.25 + (x_1^{(1)})^2\} + (x_1^{(1)})^2 - a_1\{0.25 + (x_1^{(1)})^2\} = x_1^{(1)2},$$

where $\boldsymbol{x}_1 = (x_1^{(1)}, x_1^{(2)})$. Therefore, the first stage contrast function $\tau_1(\cdot)$ equals zero. In Scenario (H), the second stage contrast function $\tau_2(\cdot,\cdot)$ equals zero. Hence, the ODTR is not unique in these two scenarios. In the last scenario, we have

$$h_2(\bar{\boldsymbol{a}}_2, \bar{\boldsymbol{x}}_2) = x_2^2, h_1(a_1, \boldsymbol{x}_1) = \mathbb{E}\{(X_0^{(2),(2)})^2 | \bar{\boldsymbol{X}}_0^{(1)} = \boldsymbol{x}_1, A_0^{(1)} = a_1\} = a_1(x_1^{(1)})^2 + 0.25,$$

where $\bar{\boldsymbol{x}}_2 = (x_1^{(1)}, x_1^{(2)}, x_2)$. In this scenario, the ODTR is uniquely defined and we have $d_1^{opt}(\boldsymbol{x}_1) = 1$, $d_2^{opt}(a_1, \bar{\boldsymbol{x}}_2) = 1$.

We compare our proposed CI (see (7.16)) with the CI based on the online one-step method, defined as $[\widehat{V}^{on} - z_{\alpha/2}\widehat{\sigma}^{on}/\sqrt{n - l_n}, \widehat{V}^{on} + z_{\alpha/2}\widehat{\sigma}^{on}/\sqrt{n - l_n}]$ where

$$\begin{aligned}
\widehat{V}^{on} &= \left(\sum_{j=l_n}^{n-1} \widetilde{\sigma}_{\mathscr{I}_{(j)}}^{-1}\right)^{-1} \left(\sum_{j=l_n}^{n-1} \widetilde{\sigma}_{\mathscr{I}_{(j)}}^{-1} \widehat{V}_{\{j+1\}}(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}})\right), \\
\widetilde{\sigma}_{\mathscr{I}_{(j)}}^2 &= \frac{1}{j(j-1)} \sum_{i=1}^{j} \left(\widehat{V}_{\{i\}}(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}}) - \widehat{V}_{\mathscr{I}_{(j)}}(\widehat{d}_{\mathscr{I}_{(j)}}; \widehat{\pi}_{\mathscr{I}_{(j)}}, \widehat{h}_{\mathscr{I}_{(j)}})\right)^2, \\
\widehat{\sigma}^{on} &= \left(\sum_{j=l_n}^{n-1} \widetilde{\sigma}_{\mathscr{I}_{(j)}}^{-1} / (n - l_n)\right)^{-1},
\end{aligned}$$

for some divergent sequence l_n. Notice that both methods require to calculate $\widehat{h}_{\mathscr{I}} = \{\widehat{h}_{\mathscr{I},k}\}_{k=1}^2$, $\widehat{\pi}_{\mathscr{I}} = \{\widehat{\pi}_{\mathscr{I},k}\}_{k=1}^2$, $\widehat{d}_{\mathscr{I}} = \{\widehat{d}_{\mathscr{I},k}\}_{k=1}^2$. These estimators are computed based on cubic B-spline methods. To save space, we present the detailed estimating procedure in Section **??** of the supplementary article.

We consider two sample sizes, $n = 600$ and $n = 1200$. Similar to Section 7.5.1, we set

Table 7.4: ACP and AL of the CIs with standard errors in parenthesis

Setting (G)	Proposed		Online ($l_n = 200$)		Online ($l_n = 400$)	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
600	93.0 (0.8)	27.4 (0.13)	82.1 (1.2)	38.3 (0.12)	90.8 (0.9)	54.0 (0.18)
1200	92.9 (0.8)	18.2 (0.05)	87.9 (1.0)	24.2 (0.06)	90.7 (0.9)	27.0 (0.07)
Setting (H)	Proposed		Online ($l_n = 200$)		Online ($l_n = 400$)	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
600	92.2 (0.8)	37.2 (0.09)	84.1 (1.2)	45.3 (0.10)	91.4 (0.9)	64.0 (0.13)
1200	92.8 (0.8)	25.4 (0.04)	89.5 (1.0)	28.6 (0.05)	92.8 (0.8)	32.0 (0.05)
Setting (I)	Proposed		Online ($l_n = 200$)		Online ($l_n = 400$)	
n	ACP(%)	AL*100	ACP(%)	AL*100	ACP(%)	AL*100
600	91.8 (0.9)	39.4 (0.14)	84.5 (1.2)	44.5 (0.10)	90.1 (0.9)	63.1 (0.13)
1200	93.0 (0.8)	26.4 (0.04)	90.5 (0.9)	28.4 (0.05)	92.5 (0.8)	31.8 (0.05)

$s_n = \lfloor n^{6/7} \rfloor$ when implementing (7.16). In Table 7.4, we report the ACP and AL of the proposed CI and the CI based on online one-step method, with $l_n = 200$ and $l_n = 400$. It can be seen that ACPs of our proposed CIs are close to the nominal level in almost all cases. In contrast, ACPs of the CIs based on the online one-step method are well below the nominal level in Scenario (G) and (H). Moreover, CIs based on our proposed method are much shorter than those based on the online one-step method.

7.6 Real data analysis

In this section, we apply the proposed method to a data from AIDS Clinical Trials Group Protocol 175 (ACTG175). We focus on a subset of the data which consists of 1046 patients that were treated with either ZDV + zalcitabine (zal) ($A = 0$) or ZDV + didanosine (ddI) ($A = 1$). The outcome of interests were CD4 count (cells/mm^3) at 20 ± 5 weeks after receiving the treatment. Fan et al. (2017) found that patient's age is the only variable that has significant interaction with the treatment. Therefore, in the following, we use age to construct the OTR. Since ACTG175 is a randomized trial, the no unmeasured confounders assumption (A2) automatically holds.

In Table 7.5, we report the estimated optimal value function and its 95% CI based on our proposed method and the online one-step method with $l_n = 50, 100$ and 200. To construct these CIs, we set $\widehat{\pi}_{\mathscr{I}} = 0.5$ for any $\mathscr{I} \subseteq \mathscr{I}_0$. The conditional mean functions are estimated using cubic B-splines. The detailed estimating procedure is very similar to that in Section

Table 7.5: Estimated value functions and confidence intervals

Method	Estimated value function	95% CI	Length of CI
Proposed	399.5	[387.9, 411.2]	23.2
Online ($l_n = 50$)	399.2	[385.6, 412.7]	27.1
Online ($l_n = 100$)	398.3	[384.4, 412.2]	27.7
Online ($l_n = 200$)	403.9	[389.5, 418.4]	28.9

7.5.1 and is hence omitted for brevity. In addition, we set $s_n = \lfloor n^{6/7} \rfloor$, as in simulations.

It can be seen from Table 7.5 that all methods yield similar estimated optimal value functions. These estimated values are larger than those based on linear decision rules (see Section 4 in Fan et al. 2017). Besides, we notice that our proposed CI is at least 16% shorter compared to those based on the online one-step method. Such phenomenon is consistent with our theoretical findings and simulation results.

7.7 Proof of Theorem 7.2.1

For any $\mathscr{I} = \{i_1, i_2, \ldots, i_s\} \subseteq \mathscr{I}_0$, the estimated treatment regime $|\widehat{d}_{\mathscr{I}}(\cdot)|$ is upper bounded by 1. It follows from the ANOVA decomposition of Efron and Stein (1981) that

$$\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) = p_s(\boldsymbol{x}) + \sum_{i \in \mathscr{I}} d_{s,1}(O_i; \boldsymbol{x}) + \sum_{\substack{i,j \in \mathscr{I} \\ i \neq j}} d_{s,2}(O_i, O_j; \boldsymbol{x}) \tag{7.19}$$

$$+ \sum_{\substack{i,j,k \in \mathscr{I} \\ i \neq j, i \neq k, j \neq k}} d_{s,3}(O_i, O_j, O_k; \boldsymbol{x}) + \cdots + d_{s,s}(O_{i_1}, O_{i_2}, \ldots, O_{i_s}; \boldsymbol{x}), \;\; \forall \boldsymbol{x},$$

where $p_s(x) = \mathbb{E}\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) = \mathbb{P}(\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) = 1)$, is the grand mean; $d_{s,1}(o; \boldsymbol{x}) = \mathbb{E}\{\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) | O_{i_1} = o\} - p_s(\boldsymbol{x})$, is the main effect;

$$d_{s,2}(o_1, o_2; \boldsymbol{x}) = \mathbb{E}\{\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) | O_{i_1} = o_1, O_{i_2} = o_2\}$$

$$- \;\; \mathbb{E}\{\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) | O_{i_1} = o_1\} - \mathbb{E}\{\widehat{d}_{\mathscr{I}}(\boldsymbol{x}) | O_{i_2} = o_2\} + p_s(\boldsymbol{x}),$$

is the second-order interaction; etc.

All the 2^s random variables on the right-hand side (RHS) of (7.19) are uncorrelated.

Therefore,

$$\sum_{k=1}^{s}\binom{s}{k}\mathbb{E}d_{s,k}^2(O_{i_1},O_{i_2},\dots,O_{i_k};\boldsymbol{x})=\mathbb{VAR}\{\widehat{d}_{\mathscr{I}}(\boldsymbol{x})\}\le\mathbb{E}\widehat{d}_{\mathscr{I}}^2(\boldsymbol{x})\le 1. \tag{7.20}$$

In the following, we show $\eta_2=o_p(n^{-1/2})$. By definition, this implies $\widehat{V}_\infty=\eta_1+o_p(n^{-1/2})$. Notice that

$$\begin{aligned}\eta_2 &= \underbrace{\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\frac{(2A_i-1)R_{\mathscr{I}}(\boldsymbol{X}_i)}{\pi(A_i,\boldsymbol{X}_i)}\{Y_i-h(A_i,\boldsymbol{X}_i)\}}_{\eta_3}\\ &+ \underbrace{\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\tau(\boldsymbol{X}_i)R_{\mathscr{I}}(\boldsymbol{X}_i)}_{\eta_4}.\end{aligned}$$

Below, we break the proof into two steps. In the first step, we show $\eta_3=o_p(n^{-1/2})$. In the second step, we prove $\eta_4=o_p(n^{-1/2})$.

Step 1: For $i=0,1,\dots,n$, let $\omega_{0,i}=(1-A_i)\{Y_i-h(0,\boldsymbol{X}_i)\}/\pi(0,\boldsymbol{X}_i)$ and $\omega_{1,i}=A_i\{Y_i-h(1,\boldsymbol{X}_i)\}/\pi(1,\boldsymbol{X}_i)$. We have

$$\eta_3=-\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\omega_{0,i}R_{\mathscr{I}}(\boldsymbol{X}_i)+\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\omega_{1,i}R_{\mathscr{I}}(\boldsymbol{X}_i).$$

Below, we show

$$\eta_3^{(1)}\equiv\frac{1}{(n-s_n)\binom{n}{s_n}}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\omega_{0,i}R_{\mathscr{I}}(\boldsymbol{X}_i)=o_p(n^{-1/2}). \tag{7.21}$$

It follows from (7.19) that

$$\eta_3^{(1)}=\frac{1}{\binom{n}{s_n}(n-s_n)}\sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}}\sum_{i\in\mathscr{I}^c}\omega_{0,i}\left(\sum_{k=1}^{s_n}\sum_{\{j_1,\dots,j_k\}\subseteq\mathscr{I}}d_{s_n,k}(O_{j_1},\dots,O_{j_k};\boldsymbol{X}_i)\right).$$

Notice that $(n-s_n)\binom{n}{s_n}=(n-s_n)\binom{n}{n-s_n}=n\binom{n-1}{n-s_n-1}=n\binom{n-1}{s_n}$. With some calculations, we have

$$\eta_3^{(1)} = \frac{1}{n}\sum_{i=1}^{n}\omega_{0,i}\sum_{k=1}^{s_n}\frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}} d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i),$$

and thus

$$\begin{aligned}\eta_3^{(1)} &= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\omega_{0,i}\sum_{k=1}^{l_0-1}\frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}} d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i)}_{\eta_3^{(2)}}\\ &+ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\omega_{0,i}\sum_{k=l_0}^{s_n}\frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}} d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i),}_{\eta_3^{(3)}}\end{aligned}$$

for some fixed integer $l_0 \geq 2$ that satisfies $l_0 > 1/(1-\beta_0)$.

By C1 and C2, we have for any $i = 1,\ldots,n$,

$$\mathbb{E}(\omega_{0,i}|\boldsymbol{X}_i)=\mathbb{E}\left(\frac{1-A_i}{\pi(0,\boldsymbol{X}_i)}\{Y_i^*(0)-h(0,\boldsymbol{X}_i)\}|\boldsymbol{X}_i\right)=\mathbb{E}[\{Y_i^*(0)-h(0,\boldsymbol{X}_i)\}|\boldsymbol{X}_i]=0.$$

By Condition C3 and C5, we have

$$\begin{aligned}\max_{i\in\mathscr{I}_0}\mathbb{E}(\omega_{0,i}^2|\boldsymbol{X}_i) &\leq \max_{i\in\mathscr{I}_0}\mathbb{E}\left(\frac{\{Y_i^*(0)-h(0,\boldsymbol{X}_i)\}^2}{\pi^2(0,\boldsymbol{X}_i)}\middle|\boldsymbol{X}_i\right)\\ &\leq \max_{i\in\mathscr{I}_0}\frac{1}{\epsilon_0^2}\mathbb{E}[\{Y_i^*(0)-h(0,\boldsymbol{X}_i)\}^2|\boldsymbol{X}_i]\\ &\leq \max_{i\in\mathscr{I}_0}\frac{1}{\epsilon_0^2}\mathbb{E}[\{Y_i^*(0)\}^2|\boldsymbol{X}_i]\leq\frac{1}{c_0^2}\sup_{\boldsymbol{x}}\mathbb{E}[\{Y_0^*(0)\}^2|\boldsymbol{X}_0=\boldsymbol{x}]\leq\bar{c}_*, \quad (7.22)\end{aligned}$$

for some constant $\bar{c}_* > 0$. Here, the first inequality in (7.22) is due to that $\mathbb{E}\{Y_i^*(0)|\boldsymbol{X}_i\} =$

$h(0, \boldsymbol{X}_i)$. Therefore, we have

$$
\begin{aligned}
n\mathbb{E}(\eta_3^{(3)})^2 &\le \mathbb{E}\sum_{i=1}^{n}\omega_{0,i}^2\left(\sum_{k=l_0}^{s_n}\frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}} d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i)\right)^2\\
\le\ & \sum_{i=1}^{n}\mathbb{E}\left\{\mathbb{E}^{O_i}\omega_{0,i}^2\left(\sum_{k=l_0}^{s_n}\frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}} d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i)\right)^2\right\}\\
=\ & \sum_{i=1}^{n}\mathbb{E}\left(\omega_{0,i}^2\mathbb{E}^{O_i}\sum_{k=l_0}^{s_n}\frac{\binom{n-1-k}{s_n-k}^2}{\binom{n-1}{s_n}^2}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}} d_{s_n,k}^2(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i)\right)\\
=\ & n\mathbb{E}\left(\omega_{0,1}^2\mathbb{E}^{O_1}\sum_{k=l_0}^{s_n}\frac{\binom{n-1-k}{s_n-k}^2}{\binom{n-1}{s_n}^2}\binom{n-1}{k} d_{s_n,k}^2(O_2,\ldots,O_{k+1};\boldsymbol{X}_1)\right)\\
\le\ & n\max_{k\ge l_0}\frac{\binom{n-1-k}{s_n-k}^2\binom{n-1}{k}}{\binom{n-1}{s_n}^2\binom{s_n}{k}}\mathbb{E}\omega_{0,1}^2 = \max_{k\ge l_0}\frac{n\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\mathbb{E}(\mathbb{E}^{\boldsymbol{X}_1}\omega_{0,1}^2)\le \bar{c}_* n\frac{s_n^{l_0}}{(n-1)^{l_0}},
\end{aligned}
$$

where $\mathbb{E}^{O_i}$ and $\mathbb{E}^{\boldsymbol{X}_i}$ denote the conditional expectation given O_i and $\boldsymbol{X}_i$, respectively, the first inequality is due to Cauchy-Schwarz inequality, the first equality follows by the fact that $d_{s_n,k_1}(O_{j_1^{(1)}},\cdots,O_{j_{k_1}^{(1)}};\boldsymbol{x})$ and $d_{s_n,k_2}(O_{j_1^{(2)}},\cdots,O_{j_{k_2}^{(2)}};\boldsymbol{x})$ are uncorrelated for any $\{j_1^{(1)},\ldots,j_{k_1}^{(1)}\}\neq\{j_1^{(2)},\ldots,j_{k_2}^{(2)}\}$, the third inequality is due to (7.20) and the last inequality is due to (7.22).

By the definitions of β_0 and l_0, we have $ns_n^{l_0}/(n-1)^{l_0}\asymp n^{1+\beta_0 l_0}/n^{l_0}=o(1)$. This together with Chebyshev's inequality gives

$$
\eta_3^{(3)} = o_p(n^{-1/2}). \tag{7.23}
$$

In addition, it follows from Cauchy-Schwarz inequality that

$$
n\mathbb{E}(\eta_3^{(2)})^2 \le \frac{l_0-1}{n}\sum_{k=1}^{l_0-1}\mathbb{E}\left(\frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\sum_{i=1}^{n}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}}\omega_{0,i}d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i)\right)^2.
$$

For any $1\le i^{(1)}, i^{(2)}\le n$ and $\{j_1^{(1)},\ldots,j_k^{(1)}\}\subseteq\mathscr{I}_{(-i^{(1)})}$, $\{j_1^{(2)},\ldots,j_k^{(2)}\}\subseteq\mathscr{I}_{(-i^{(2)})}$, we have

$$
\mathbb{E}\omega_{0,i^{(1)}}d_{s_n,k}(O_{j_1^{(1)}},\ldots,O_{j_k^{(1)}};\boldsymbol{X}_{i^{(1)}})\omega_{0,i^{(2)}}d_{s_n,k}(O_{j_1^{(2)}},\ldots,O_{j_k^{(2)}};\boldsymbol{X}_{i^{(2)}})=0,
$$

for any $\{i^{(1)}, j_1^{(1)},\ldots,j_k^{(1)}\}\neq\{i^{(2)}, j_1^{(2)},\ldots,j_k^{(2)}\}$. Let $\sigma_k^2=\mathbb{E}\omega_{0,1}^2d_{s_n,k}^2(O_2,\ldots,O_{k+1};\boldsymbol{X}_1)$. It follows from Cauchy-Schwarz inequality that

$$\begin{aligned}
& |\mathbb{E}\omega_{0,i^{(1)}} d_{s_n,k}(O_{j_1^{(1)}},\ldots,O_{j_k^{(1)}};\boldsymbol{X}_{i^{(1)}})\omega_{0,i^{(2)}} d_{s_n,k}(O_{j_1^{(2)}},\ldots,O_{j_k^{(2)}};\boldsymbol{X}_{i^{(2)}})| \\
\leq\ & \frac{1}{2}\mathbb{E}|\omega_{0,i^{(1)}} d_{s_n,k}(O_{j_1^{(1)}},\ldots,O_{j_k^{(1)}};\boldsymbol{X}_{i^{(1)}})|^2+\frac{1}{2}\mathbb{E}|\omega_{0,i^{(2)}} d_{s_n,k}(O_{j_1^{(2)}},\ldots,O_{j_k^{(2)}};\boldsymbol{X}_{i^{(2)}})|^2 \\
=\ & \sigma_k^2.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
n\mathbb{E}(\eta_3^{(2)})^2 &\leq \frac{l_0-1}{n}\sum_{k=1}^{l_0-1}\frac{\binom{n-1-k}{s_n-k}^2}{\binom{n-1}{s_n}^2}\left(\sum_{i=1}^{n}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}}\mathbb{E}\omega_{0,i}^2 d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_i)^2\right. \\
&\left.+\sum_{\substack{1\leq i\leq n\\1\leq j\leq n\\ i\neq j}}\ \sum_{\substack{\{i_1,\ldots,i_k\}\subseteq\mathscr{I}_{(-i)}\\ \{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-j)}\\ \{i,i_1,\ldots,i_k\}=\{j,j_1,\ldots,j_k\}}}\mathbb{E}\omega_{0,i} d_{s_n,k}(O_{i_1},\ldots,O_{i_k};\boldsymbol{X}_i)\omega_{0,j} d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\boldsymbol{X}_j)\right) \\
&\leq \frac{l_0-1}{n}\sum_{k=1}^{l_0-1}\frac{\binom{n-1-k}{s_n-k}^2}{\binom{n-1}{s_n}^2}\left(\sum_{i=1}^{n}\sum_{\{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-i)}}\sigma_k^2+\sum_{\substack{1\leq i\leq n\\1\leq j\leq n\\ i\neq j}}\ \sum_{\substack{\{i_1,\ldots,i_k\}\subseteq\mathscr{I}_{(-i)}\\ \{j_1,\ldots,j_k\}\subseteq\mathscr{I}_{(-j)}\\ \{i,i_1,\ldots,i_k\}=\{j,j_1,\ldots,j_k\}}}\sigma_k^2\right) \\
&= \frac{l_0-1}{n}\sum_{k=1}^{l_0-1}\frac{\binom{n-1-k}{s_n-k}^2}{\binom{n-1}{s_n}^2}\left\{n\binom{n-1}{k}\sigma_k^2+kn\binom{n-1}{k}\sigma_k^2\right\}\asymp\sum_{k=1}^{l_0-1}\frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}}\binom{s_n}{k}\sigma_k^2,
\end{aligned}$$

where the second inequality is due to Cauchy-Schwarz inequality. By (7.20) and (7.22), we have

$$\begin{aligned}
& \binom{s_n}{k}\sigma_k^2 \leq \binom{s_n}{k}\mathbb{E}\omega_1^2 d_{s_n,k}^2(O_2,\ldots,O_{k+1};\boldsymbol{X}_1) \\
\leq\ & \binom{s_n}{k}\mathbb{E}\{\mathbb{E}(\omega_{0,1}^2|\boldsymbol{X}_1)d_{s_n,k}^2(O_2,\ldots,O_{k+1};\boldsymbol{X}_1)\} \\
\leq\ & \bar{c}_*\binom{s_n}{k}\mathbb{E}\{d_{s_n,k}^2(O_2,\ldots,O_{k+1};\boldsymbol{X}_1)\}\leq\bar{c}_*.
\end{aligned}$$

Therefore,

$$n\mathbb{E}(\eta_3^{(2)})^2\asymp\sum_{k=1}^{l_0-1}(k+1)\frac{s_n^k}{(n-1)^k}\leq l_0^2\frac{s_n}{n-1}\to 0.$$

By Cauchy-Schwarz inequality, we obtain $\eta_3^{(2)} = o_p(n^{-1/2})$. This together with (7.23) gives $\eta_3^{(1)} = o_p(n^{-1/2})$. Similarly, we can show

$$\frac{1}{(n-s_n)\binom{n}{s_n}} \sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}} \sum_{i\in\mathscr{I}^c} \omega_{1,i} R_{\mathscr{I}}(\boldsymbol{X}_i) = o_p(n^{-1/2}).$$

This together with (7.21) proves $\eta_3 = o_p(n^{-1/2})$.

Step 2: Notice that

$$\begin{aligned}
\eta_4 &= \frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\mathscr{I}\subseteq\mathscr{I}_0,|\mathscr{I}|=s_n} \sum_{i\in\mathscr{I}^c} \tau(\boldsymbol{X}_i)[\widehat{d}_{\mathscr{I}}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)] \\
&= \underbrace{\frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\mathscr{I}\subseteq\mathscr{I}_0,|\mathscr{I}|=s_n} \sum_{i\in\mathscr{I}^c} \tau(\boldsymbol{X}_i)[\widehat{d}_{\mathscr{I}}(\boldsymbol{X}_i) - \mathbb{I}\{\tau(\boldsymbol{X}_i) > 0\}]}_{\eta_4^{(1)}} \\
&- \underbrace{\frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\mathscr{I}\subseteq\mathscr{I}_0,|\mathscr{I}|=s_n} \sum_{i\in\mathscr{I}^c} \tau(\boldsymbol{X}_i)[p_{s_n}(\boldsymbol{X}_i) - \mathbb{I}\{\tau(\boldsymbol{X}_i) > 0\}]}_{\eta_4^{(2)}}.
\end{aligned}$$

To prove $\eta_4 = o_p(n^{-1/2})$, it suffices to show $\eta_4^{(1)}, \eta_4^{(2)} = o_p(n^{-1/2})$.

With some calculations, we have

$$\begin{aligned}
\eta_4^{(1)} &= \underbrace{\frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}} \sum_{i\in\mathscr{I}^c} \tau(\boldsymbol{X}_i)[\widehat{d}_{\mathscr{I}}(\boldsymbol{X}_i) - \mathbb{I}\{\tau(\boldsymbol{X}_i) > 0\}]\mathbb{I}\{|\tau(\boldsymbol{X}_i)| \le \tau_n\}}_{\eta_4^{(3)}} \\
&+ \underbrace{\frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}} \sum_{i\in\mathscr{I}^c} \tau(\boldsymbol{X}_i)[\widehat{d}_{\mathscr{I}}(\boldsymbol{X}_i) - \mathbb{I}\{\tau(\boldsymbol{X}_i) > 0\}]\mathbb{I}\{|\tau(\boldsymbol{X}_i)| > \tau_n\}}_{\eta_4^{(4)}},
\end{aligned}$$

for some sequence τ_n that will be specified later.

It follows from Condition C4 that

$$\begin{aligned}
&\mathbb{E}|\eta_4^{(3)}| \le \frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\mathscr{I}\subseteq\mathscr{I}_0,|\mathscr{I}|=s_n} \sum_{i\in\mathscr{I}^c} \mathbb{E}|\tau(\boldsymbol{X}_i)|\mathbb{I}\{|\tau(\boldsymbol{X}_i)| \le \tau_n\} \qquad (7.24)\\
&= \frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\mathscr{I}\subseteq\mathscr{I}_0,|\mathscr{I}|=s_n} \sum_{i\in\mathscr{I}^c} \mathbb{E}|\tau(\boldsymbol{X}_i)|\mathbb{I}\{0 < |\tau(\boldsymbol{X}_i)| \le \tau_n\} \le \bar{c}\tau_n^{1+\alpha},
\end{aligned}$$

for any τ_n such that $\tau_n \le \delta_0$. Moreover, since $\widehat{d}_{\mathscr{I}}(\boldsymbol{X}_i) \ne \mathbb{I}\{\tau(\boldsymbol{X}_i) > 0\}$ only when $|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_i) - \tau(\boldsymbol{X}_i)| > |\tau(\boldsymbol{X}_i)|$, $\mathbb{E}|\eta_4^{(4)}|$ can be upper bounded by

$$\begin{aligned}
&\frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}} \sum_{i\in\mathscr{I}^c} \mathbb{E}|\tau(\boldsymbol{X}_i)|\mathbb{I}(|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_i)-\tau(\boldsymbol{X}_i)| > |\tau(\boldsymbol{X}_i)|)\mathbb{I}\{|\tau(\boldsymbol{X}_i)| > \tau_n\} \\
&\le \frac{1}{\binom{n}{s_n}(n-s_n)} \sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}} \sum_{i\in\mathscr{I}^c} \mathbb{E}|\tau(\boldsymbol{X}_i)| \frac{|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_i)-\tau(\boldsymbol{X}_i)|^2}{|\tau(\boldsymbol{X}_i)|^2} \mathbb{I}\{|\tau(\boldsymbol{X}_i)| > \tau_n\} \\
&\le \frac{1}{\tau_n\binom{n}{s_n}(n-s_n)} \sum_{\substack{\mathscr{I}\subseteq\mathscr{I}_0\\|\mathscr{I}|=s_n}} \sum_{i\in\mathscr{I}^c} \mathbb{E}|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_i)-\tau(\boldsymbol{X}_i)|^2 \le \frac{s_n^{-\kappa_0}}{\tau_n} \asymp \frac{1}{\tau_n n^{\beta_0\kappa_0}}, \qquad (7.25)
\end{aligned}$$

where the last inequality is due to Condition C6. Combining (7.25) together with (7.24) gives $\mathbb{E}|\eta_4^{(1)}| = O(\tau_n^{1+\alpha} + \tau_n^{-1} n^{-\beta_0\kappa_0})$. Set $\tau_n = n^{-\beta_0\kappa_0/(2+\alpha)}$, we obtain that

$$\mathbb{E}|\eta_4^{(1)}| = O\left(n^{-\beta_0\kappa_0(1+\alpha)/(2+\alpha)}\right) = o(n^{-1/2}),$$

where the last equality is due to the condition $\beta_0 > (2+\alpha)/\{\kappa_0(1+\alpha)\}$. By Markov's inequality, we obtain $\eta_4^{(1)} = o_p(n^{-1/2})$. As for $\eta_4^{(2)}$, we have

$$\begin{aligned}
\eta_4^{(2)} &= \underbrace{\frac{1}{n}\sum_{i=1}^n \tau(\boldsymbol{X}_i)[p_{s_n}(\boldsymbol{X}_i) - \mathbb{I}\{\tau(\boldsymbol{X}_i)>0\}]\mathbb{I}\{|\tau(\boldsymbol{X}_i)| \le \tau_n\}}_{\eta_4^{(5)}} \\
&+ \underbrace{\frac{1}{n}\sum_{i=1}^n \tau(\boldsymbol{X}_i)[p_{s_n}(\boldsymbol{X}_i) - \mathbb{I}\{\tau(\boldsymbol{X}_i)>0\}]\mathbb{I}\{|\tau(\boldsymbol{X}_i)| > \tau_n\}}_{\eta_4^{(6)}}.
\end{aligned}$$

Similar to (7.24), we can show

$$\mathbb{E}|\eta_4^{(5)}| \le \bar{c}\,\tau_n^{1+\alpha}, \qquad (7.26)$$

for any τ_n such that $\tau_n \le \delta_0$.

Besides, it follows from Chebysev's inequality that

$$\begin{aligned}
&|p_{s_n}(\boldsymbol{X}_0) - \mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}| = |\mathbb{P}^{\boldsymbol{X}_0}\{\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_0)>0\} - \mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}| \\
\le\ & \mathbb{P}^{\boldsymbol{X}_0}\{|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_0)-\tau(\boldsymbol{X}_0)| \ge |\tau(\boldsymbol{X}_0)|\} \le \mathbb{E}^{\boldsymbol{X}_0}\frac{|\widehat{\tau}_{\mathscr{I}}(\boldsymbol{X}_0)-\tau(\boldsymbol{X}_0)|^2}{|\tau(\boldsymbol{X}_0)|^2}, \qquad (7.27)
\end{aligned}$$

where $\mathbb{P}^{\boldsymbol{X}_0}(\cdot)$ denotes the conditional probability given $\boldsymbol{X}_0$, and $\mathscr{I}$ is an arbitrary subset of $\{1,\ldots,n\}$ with $|\mathscr{I}|=s_n$. Therefore, using similar arguments in bounding $\mathbb{E}|\eta_4^{(4)}|$, we can show $\mathbb{E}|\eta_4^{(6)}|=O(\tau_n^{-1}n^{-\beta_0\kappa_0})$. Combining this together with (7.26), we've shown

$$\mathbb{E}|\eta_4^{(2)}|=O\left(\tau_n^{1+\alpha}+\frac{1}{\tau_n n^{\beta_0\kappa_0}}\right).$$

Set $\tau_n = n^{-\beta_0\kappa_0/(2+\alpha)}$, by Markov's inequality, we obtain $\eta_4^{(2)} = o_p(n^{-1/2})$. This proves $\eta_4 = o_p(n^{-1/2})$.

To summarize, we've shown $\eta_2 = o_p(n^{-1/2})$. Next, we show $V_0 = \mathbb{E}\eta_1 + o(n^{-1/2})$. Let $\mathbb{E}^{A_0,\boldsymbol{X}_0}$ denote the conditional expectation given A_0 and $\boldsymbol{X}_0$. It follows from the definitions of V_0 and η_1 that

$$\begin{aligned}
& V_0-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{\mathrm{g}\{A_i,p_{s_n}(\boldsymbol{X}_i)\}}{\pi(A_0,\boldsymbol{X}_i)}\{Y_i-h(A_i,\boldsymbol{X}_i)\}+h(p_{s_n}(\boldsymbol{X}_i),\boldsymbol{X}_i)\right) \qquad (7.28)\\
= \;& V_0-\mathbb{E}\left(\frac{\mathrm{g}\{A_0,p_{s_n}(\boldsymbol{X}_0)\}}{\pi(A_0,\boldsymbol{X}_0)}\{Y_0-h(A_0,\boldsymbol{X}_0)\}+h(p_{s_n}(\boldsymbol{X}_0),\boldsymbol{X}_0)\right)\\
= \;& V_0-\mathbb{E}\left(\frac{\mathrm{g}\{A_0,p_{s_n}(\boldsymbol{X}_0)\}}{\pi(A_0,\boldsymbol{X}_0)}\mathbb{E}^{A_0,\boldsymbol{X}_0}\{Y_0-h(A_0,\boldsymbol{X}_0)\}+h(p_{s_n}(\boldsymbol{X}_0),\boldsymbol{X}_0)\right)\\
= \;& V_0-\mathbb{E}[p_{s_n}(\boldsymbol{X}_0)h(1,\boldsymbol{X}_0)+\{1-p_{s_n}(\boldsymbol{X}_0)\}h(0,\boldsymbol{X}_0)]\\
= \;& \mathbb{E}[h(0,\boldsymbol{X}_0)+\tau(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}]-\mathbb{E}[p_{s_n}(\boldsymbol{X}_0)h(1,\boldsymbol{X}_0)\\
+ \;& \{1-p_{s_n}(\boldsymbol{X}_0)\}h(0,\boldsymbol{X}_0)]=\mathbb{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}-p_{s_n}(\boldsymbol{X}_0)].
\end{aligned}$$

Using similar arguments in bounding $\mathbb{E}|\eta_4^{(1)}|$, we can show

$$\mathbb{E}|\tau(\boldsymbol{X})||p_{s_n}(\boldsymbol{X}_0)-\mathbb{I}\{\tau(\boldsymbol{X}_0>0)\}|=o(n^{-1/2}).$$

In view of (7.28), this yields $V_0=\mathbb{E}\eta_1+o(n^{-1/2})$. The proof is hence completed.

7.8 Discussion

In this chapter, we propose to construct the confidence interval for the optimal value function based on subsample aggregating and refitted cross-validation. Such an inference method can be applied to some other non-regular problems as well, such as inference for low dimensional regression coefficients in high-dimensional models. Alternative to our method, one may generate multiple bootstrap samples, estimate the OTR based on each bootstrap sample, evaluate its value using the remaining out-of-bag samples, and average over all value estimators. However, it remains unknown what the asymptotic distribution of the aggregated value estimator would be.

REFERENCES

Athreya, K. B. and Lahiri, S. N. (2006). *Measure theory and probability theory*. Springer Texts in Statistics. Springer, New York.

Audibert, J.-Y. and Tsybakov, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.*, 35(2):608–633.

Avi-Itzhak, H., Van Mieghem, J. A., and Rub, L. (1995). Multiple subclass pattern recognition: a maximin correlation approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(4):418–431.

Baker, S. G., Cook, N. R., Vickers, A., and Kramer, B. S. (2009). Using relative utility curves to evaluate risk prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(4):729âĂŞ748.

Banerjee, M., Durot, C., and Sen, B. (2019). Divide and conquer in nonstandard problems and the super-efficiency phenomenon. *Ann. Statist.*, 47(2):720–757.

Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2):345–366.

Biau, G. (2012). J. mach. learn. res. *Journal of Machine Learning Research*, 13:1063–1095.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.*, 37(4):1705–1732.

Boers, M., Verhoeven, A. C., Markusse, H. M., van de Laar, M. A., Westhovens, R., van Denderen, J. C., van Zeben, D., Dijkmans, B. A., Peeters, A. J., Jacobs, P., et al. (1997). Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *The Lancet*, 350(9074):309–318.

Brown, L. D., Low, M. G., and Zhao, L. H. (1997). Superefficiency in nonparametric function estimation. *Ann. Statist.*, 25(6):2607–2625.

Bühlmann, P. and Meinshausen, N. (2016). Magging: maximin aggregation for inhomogeneous large-scale data. *Proceedings of the IEEE*, 104(1):126–135.

Bühlmann, P. and Yu, B. (2002). Analyzing bagging. *Ann. Statist.*, 30(4):927–961.

Candès, E. and Tao, T. (2007). Rejoinder: "The Dantzig selector: statistical estimation when p is much larger than n" [Ann. Statist. **35** (2007), no. 6, 2313–2351; mr2382644]. *Ann. Statist.*, 35(6):2392–2404.

Cannings, T. I. and Samworth, R. J. (2017). Random-projection ensemble classification. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(4):959–1035. With discussions and a reply by the authors.

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(4):408–417.

Chakraborty, B., Murphy, S., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat. Methods Med. Res.*, 19(3):317–343.

Chang, M., Lee, S., and Whang, Y.-J. (2015). Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements. *Econom. J.*, 18(3):307–346.

Chatterjee, S. (2005). An error bound in the sudakov-fernique inequality. *arXiv preprint math/0510424*.

Chen, H., Manning, A. K., and Dupuis, J. (2012). A method of moments estimator for random effect multivariate meta-analysis. *Biometrics*, 68(4):1278–1284.

Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771.

Chen, X. and Xie, M.-g. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statist. Sinica*, 24(4):1655–1684.

Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.*, 16:31–41.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.*, 42(4):1564–1597.

Csörgő, M., Csörgő, S., Horváth, L., and Révész, P. (1985). On weak and strong approximations of the quantile process. In *Proceedings of the seventh conference on probability theory (Braşov, 1982)*, pages 81–95. VNU Sci. Press, Utrecht.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188.

Ding, Y., Li, Y., and Song, R. (2019). *Statistical learning for personalized wealth management*. Under review.

Dunn, G. and Bentall, R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in medicine*, 26(26):4719–4745.

Durot, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields*, 122(2):222–240.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *Ann. Statist.*, 9(3):586–596.

Fan, A., Lu, W., and Song, R. (2016). Sequential advantage selection for optimal treatment regime. *Ann. Appl. Stat.*, 10(1):32–53.

Fan, C., Lu, W., Song, R., and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(5):1565–1582.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica*, 20(1):101–148.

Fan, J. and Lv, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory*, 57(8):5467–5484.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London.

Farragher, T. M., Lunt, M., Fu, B., Bunn, D., and Symmons, D. P. (2010). Early treatment with, and time receiving, first disease-modifying antirheumatic drug predicts long-term function in patients with inflammatory polyarthritis. *Annals of the rheumatic diseases*, 69(4):689–695.

Fava, M., Rush, A. J., Trivedi, M. H., Nierenberg, A. A., Thase, M. E., Sackeim, H. A., Quitkin, F. M., Wisniewski, S., Lavori, P. W., Rosenbaum, J. F., et al. (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (star* d) study. *Psychiatric Clinics of North America*, 26(2):457–494.

Gail, M. H. (2009). Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *Journal of the National Cancer Institute*, 101(13):959–963.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., Rossi, F., and Ulerich, R. (2015). *GNU Scientific Library Reference Manual (Version 2.1)*.

Giné, E., Mason, D. M., and Zaitsev, A. Y. (2003). The L_1-norm density estimator process. *Ann. Probab.*, 31(2):719–768.

Groeneboom, P., Hooghiemstra, G., and Lopuhaä, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.*, 27(4):1316–1347.

Gunter, L., Zhu, J., and Murphy, S. A. (2011). Variable selection for qualitative interactions. *Stat. Methodol.*, 8(1):42–55.

Härdle, W. (1990). *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Hsu, Y.-C. (2017). Consistent tests for conditional treatment effects. *The Econometrics Journal*, 20(1):1–22.

Huang, Y., Laber, E. B., and Janes, H. (2015). Characterizing expected benefits of biomarkers in treatment selection. *Biostatistics*, 16(2):383–399.

Jackson, D., White, I. R., and Thompson, S. G. (2010). Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat. Med.*, 29(12):1282–1297.

Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1.

Kay, S. R., Fiszbein, A., and Opfer, L. A. (1987). The positive and negative syndrome scale (panss) for schizophrenia. *Schizophrenia bulletin*, 13(2):261.

Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., Markowitz, J. C., Nemeroff, C. B., Russell, J. M., Thase, M. E., et al. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine*, 342(20):1462–1470.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.*, 18(1):191–219.

Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M. I. (2014). A scalable bootstrap for massive data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 76(4):795–816.

Koltchinskii, V. I. (1994). Komlos-Major-Tusnady approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab.*, 7(1):73–118.

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Laber, E. B. and Zhao, Y. Q. (2015). Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514.

Lee, T., Moon, T., Kim, S. J., and Yoon, S. (2016). Regularization and kernelization of the maximin correlation approach. *IEEE Access*, 4:1385–1392.

Li, K.-C. and Duan, N. (1989). Regression analysis under link violation. *Ann. Statist.*, 17(3):1009–1052.

Li, P., Hastie, T. J., and Church, K. W. (2006). Very sparse random projections. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 287–296. ACM.

Liang, S., Lu, W., Song, R., and Wang, L. (2018). Sparse concordance-assisted learning for optimal treatment decision. *J. Mach. Learn. Res.*, 18:1–26.

Lopes, M., Jacob, L., and Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems*, pages 1206–1214.

Lu, W., Zhang, H. H., and Zeng, D. (2013). Variable selection for optimal treatment decision. *Stat. Methods Med. Res.*, 22(5):493–504.

Luedtke, A. R. and van der Laan, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.*, 44(2):713–742.

Lv, J. and Fan, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.*, 37(6A):3498–3528.

Mason, D. M. and Polonik, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.*, 19(3):1108–1142.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London. Second edition [of MR0727836].

Meinshausen, N. and Bühlmann, P. (2015). Maximin effects in inhomogeneous large-scale data. *Ann. Statist.*, 43(4):1801–1830.

Mendelson, S., Pajor, A., and Tomczak-Jaegermann, N. (2007). Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.*, 17(4):1248–1282.

Mendelson, S., Pajor, A., and Tomczak-Jaegermann, N. (2008). Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.*, 28(3):277–289.

Milman, V. and Pajor, A. (2003). Regularization of star bodies by random hyperplane cut off. *Studia Math*, 159(2):247–261.

Milman, V. D. and Pajor, A. (1989). Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. In *Geometric aspects of functional analysis*, pages 64–104. Springer.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(2):331–366.

Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Stat. Med.*, 24(10):1455–1481.

Nelson, J. and Nguyên, H. L. (2013). Osnap: Faster numerical linear algebra algorithms via sparser subspace embeddings. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 117–126. IEEE.

Omidiran, D. and Wainwright, M. J. (2010). High-dimensional variable selection with sparse random projections: measurement sparsity and statistical efficiency. *Journal of Machine Learning Research*, 11(Aug):2361–2386.

Pimentel, L. P. R. (2014). On the location of the maximum of a continuous stochastic process. *J. Appl. Probab.*, 51(1):152–161.

Piquette-Miller, M. and Grant, D. (2007). The art and science of personalized medicine. *Clinical Pharmacology & Therapeutics*, 81:311–315.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.*, 39(2):1180–1210.

Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields*, 98(1):21–45.

Robins, J., Hernan, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiol.*, 11:550–560.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, volume 179 of *Lecture Notes in Statist.*, pages 189–326. Springer, New York.

Rubin, D. B. (2005). Causal inference using potential outcomes: design, modeling, decisions. *J. Amer. Statist. Assoc.*, 100(469):322–331.

Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., Thase, M. E., Nierenberg, A. A., Quitkin, F. M., Kashner, T. M., et al. (2004). Sequenced treatment alternatives to relieve depression (star* d): rationale and design. *Controlled clinical trials*, 25(1):119–142.

Sakhanenko, A. I. (2006). Estimates in the invariance principle in terms of truncated power moments. *Sibirsk. Mat. Zh.*, 47(6):1355–1371.

Shergin, V. V. (1990). The central limit theorem for finitely dependent random variables. In *Probability theory and mathematical statistics, Vol. II (Vilnius, 1989)*, pages 424–431. “Mokslas”, Vilnius.

Shi, C., Fan, A., Song, R., and Lu, W. (2018a). High-dimensional *A*-learning for optimal dynamic treatment regimes. *Ann. Statist.*, 46(3):925–957.

Shi, C., Lu, W., and Song, R. (2018b). A massive data framework for M-estimators with cubic-rate. *J. Amer. Statist. Assoc.*, 113(524):1698–1709.

Shi, C., Lu, W., and Song, R. (2019a). A sparse random projection-based test for overall qualitative treatment effects. *Journal of the American Statistical Association*, accepted.

Shi, C., Song, R., and Lu, W. (2016). Robust learning for optimal treatment decision with NP-dimensionality. *Electron. J. Stat.*, 10(2):2894–2921.

Shi, C., Song, R., and Lu, W. (2019b). *Concordance and value information criteria for optimal treatment decision.* Under review.

Shi, C., Song, R., and Lu, W. (2019c). On testing conditional qualitative treatment effects. *Annals of Statistics*, accepted.

Shi, C., Song, R., and Lu, W. (2019d). *Subagging for inference of the mean outcome under optimal treatment regimes.* under review.

Shi, C., Song, R., Lu, W., and Fu, B. (2018c). Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80(4):681–702.

Shi, C., Song, R., Lu, W., and Zhang, D. (2019e). *Inference of the mean outcome under an optimal treatment regime in high dimensions.*

Song, R., Wang, W., Zeng, D., and Kosorok, M. R. (2015). Penalized Q-learning for dynamic treatment regimens. *Statist. Sinica*, 25(3):901–920.

Steinwart, I. and Christmann, A. (2008). *Support vector machines.* Information Science and Statistics. Springer, New York.

Symmons, D., Tricker, K., Roberts, C., Davies, L., Dawes, P., and Scott, D. (2005). The british rheumatoid outcome study group (brosg) randomised controlled trial to compare the effectiveness and cost-effectiveness of aggressive versus symptomatic therapy in established rheumatoid arthritis. *Health technology assessment (Winchester, England)*, 9(34):iii–iv.

Tarrier, N., Lewis, S., Haddock, G., Bentall, R., Drake, R., Kinderman, P., Kingdon, D., Siddle, R., Everitt, J., Leadley, K., et al. (2004). Cognitive-behavioural therapy in first-episode and early schizophrenia. *The British Journal of Psychiatry*, 184(3):231–239.

Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.*, 32(1):135–166.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.*, 42(3):1166–1202.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *J. Amer. Statist. Assoc.*, 113(523):1228–1242.

Wand, M. P. and Jones, M. C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. Amer. Statist. Assoc.*, 88(422):520–528.

Wang, H. (2009). Forward regression for ultra-high dimensional variable screening. *J. Amer. Statist. Assoc.*, 104(488):1512–1524.

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

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Xu, Z., Li, Z., Guan, Q., Zhang, D., Li, Q., Nan, J., Liu, C., Bian, W., and Ye, J. (2018). Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 905–913. ACM.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 76(1):217–242.

Zhang, X. and Cheng, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.*, 112(518):757–768.

Zhang, Y., Duchi, J., and Wainwright, M. (2013). Divide and conquer kernel ridge regression. In *Conference on Learning Theory*, pages 592–617.

Zhang, Y., Laber, E. B., Davidian, M., and Tsiatis, A. A. (2018). Interpretable dynamic treatment regimes. *J. Amer. Statist. Assoc.*, 113(524):1541–1549.

Zhang, Y., Laber, E. B., Tsiatis, A., and Davidian, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904.

Zhao, T., Cheng, G., and Liu, H. (2016). A partially linear framework for massive heterogeneous data. *Ann. Statist.*, 44(4):1400–1437.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.*, 107(499):1106–1118.

Zhao, Y.-Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.*, 110(510):583–598.

Zhou, S. (2009). Restricted eigenvalue conditions on subgaussian random matrices. *arXiv: 0912.4045*.

Zhou, S., Shen, X., and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.*, 26(5):1760–1782.

Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. R. (2015). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, (just-accepted):00–00.

Zhu, R., Zhao, Y.-Q., Chen, G., Ma, S., and Zhao, H. (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics*, 73(2):391–400.