EFFICIENCY AND ROBUSTNESS IN INDIVIDUALIZED DECISION MAKING

Weibin Mo

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2021

Approved by:

Yufeng Liu

Shankar Bhamidi

Jan Hannig

Quoc Tran-Dinh

Donglin Zeng

ABSTRACT

Weibin Mo: Efficiency and Robustness in Individualized Decision Making
(Under the direction of Yufeng Liu)

Recent development in data-driven decision science has seen great advances in individualized decision making. Given data with covariates, treatment assignments and outcomes, one common goal is to find individualized decision rules that map the individual characteristics or contextual information to the treatment assignment, such that the overall expected outcome can be optimized. In this dissertation, we propose several new approaches to learn efficient and robust individualized decision rules. In the first project, we consider the robust learning problem when training and testing distributions can be different. A novel framework of the *Distributionally Robust Individualized Treatment Rule (DR-ITR)* is proposed to maximize the worst-case value function under distributional changes. The testing performance among a set of distributions close to training can be guaranteed reasonably well. For the second project, we consider the problem of treatment-free effect misspecification and heteroscedasticity. We propose an *Efficient Learning (E-Learning)* framework for finding an optimal ITR with improved efficiency in the multiple treatment setting. The proposed E-Learning is optimal among a regular class of semiparametric estimates that can allow treatment-free effect misspecification and heteroscedasticity. We demonstrate its effectiveness when one of or both misspecified treatment-free effect and heteroscedasticity exist. For the third project, we study the multi-stage multi-treatment decision problem. A new *Backward Change Point Structural Nested Mean Model (BCP-SNMM)* is developed to allow an unknown backward change point of the SNMM. We further propose the *Dynamic Efficient Learning (DE-Learning)* framework that is optimal under the BCP-SNMM and enjoys more robustness. Compared with the existing G-Estimation, DE-Learning is a tractable procedure for rigorous semiparametric efficient estimation, with much fewer nuisance functions to estimate and can be implemented in a backward stagewise manner.

To my parents,

Yanping Deng and Jianmin Mo,

and my beloved fiancée,

Xiyue Li,

who have supported me throughout my life.

ACKNOWLEDGEMENTS

The success of this dissertation cannot be achieved without my family, mentors and fellows who stood by me all the way along my doctorate journey. I have been enjoying the exploration of academic problems in my research area. Such an experience will be one of the most treasurable wealth in my life.

I would like to express the most sincere gratitude to my advisor, Dr. Yufeng Liu, who has paved the way of my current research and offered important instructions without any reservation. During our research discussion, he always shared his professional points of view on our work and provided the most helpful suggestions for substantial improvement. One of the most important lessons I have learned from is to illustrate a comprehensive idea in the simplest way, which has been an essential ingredient of my research projects. As a consequence, the significance of two research projects can eventually be recognized by the peer reviews of top journals. Besides the device of simplification, I have also learned a lot to address the most relevant research questions, to begin with the study of concrete examples in a highly complicated problem, and many of others. With his great encouragement and support to continue with my research, I am thrilled to make more academic contributions in the future. Therefore, a significant part of my current academic honors, and the future accomplishments if possible, should go to my advisor.

The completion of my dissertation should also acknowledge the devotion of the committee members: Shankar Bhamidi, Jan Hannig, Quoc Tran-Dinh and Donglin Zeng. In particular, as a pioneer researcher on the individualized decision making problem, Dr. Donglin Zeng has highlighted many important punchlines from this area during the discussion. The involvement of other committee members has seen the meeting of minds extensively with probabilistic tools, statistical decision theories and modern optimization techniques in this field. Besides the helpful comments of the committee on the research projects, the dissertation has also benefited from the graduate courses of Measure Theory and Probability by Prof. Shankar Bhamidi, Mathematical Statistics by

Prof. Jan Hannig, and Convex Optimization by Prof. Quoc Tran-Dinh. I am grateful to study and do research in such a supportive environment surrounded by their intelligence.

During my doctorate research, I am also thankful to my collaborators: Drs. Zhengling Qi at the George Washington University, Junlong Zhao at Beijing Normal University, Ji Zhu at University of Michigan, and the Ph.D. students José Ángel Sánchez Gómez at UNC-Chapel Hill, as well as Weijing Tang and Songkai Xue at University of Michigan. They have been supportive and complementary in our collaborative projects and allowed me to conduct more extensive research in many different areas.

Other than the memorable research life, I also rejoice over the companion of the friends I have met at Chapel Hill, North Carolina, including my roommates Chen Xing, Saurabh Nagda, Yiyun Luo and Ruituo Fan, and the colleagues Gang Li, Miheer Dewaskar, Brendan Brown, Zhengling Qi, Jianyu Liu, Hongsheng Liu, Yifan Cui, Haodong Wang, José Ángel Sánchez Gómez, Wei Liu, and others from Department of STOR at UNC-Chapel Hill.

Besides the study and life at the graduate school, I would also like to express the appreciation to my undergraduate mentors. Prof. Jun Yang at Zhejiang University has always been encouraging me to adhere to my dream, to be down-to-earth, and to work on criticality. I have been benefiting too much from his words during these years. Profs. Rong Li at Syracuse University, Changliang Zou and Li Tian at Nankai University had enlightened and supported me a lot to pursue an academic profession in Statistics when I was on a completely different track. They were the oracle on finding the best way for my self-fulfilment. Thanks to their recognition, I shall keep pursing my interdisciplinary advantages in the future of my career.

Finally, I would like to give the sincere thanks to my parents, Yanping Deng and Jianmin Mo, who have been supporting and dedicating to my life with no complaints. My fiancée, Xiyue Li, has been the most important person of my life that shares my positive and negative thoughts and emotions. I cannot express how much I am grateful of her companion when I was experiencing the most challenging growth during the doctorate study. The most sincere wish is given to her accomplishment of and growth from her doctorate dissertation as well, during which I shall stand by her side as always.

TABLE OF CONTENTS

CHAPTER 1
Introduction

Data-driven individualized decision making problems are commonly seen in practice and have been studied intensively in the literature. In disease management, the physician may decide whether to introduce or switch a therapy for a patient based on his/her characteristics in order to achieve better clinical outcome (Bertsimas et al., 2017). In public policy making, a policy that allocates the resource based on the characteristics of the targets can improve the overall resource allocation efficiency (Kube et al., 2019). In a context-based recommender system, the use of the contextual information such as time, location and social connection can improve the effectiveness of the recommendation process (Aggarwal, 2016). One common goal of the individualized decision making problem is to find decision rules that map the individual characteristics or contextual information to the treatment assignment, such that the overall expected outcome can be optimized. In this dissertation, we mainly focus on the efficiency and robustness of the individualized decision making problem and investigate several new approaches.

In this chapter, we provide the general background and review some existing techniques in the literature. In Section 1.1, we discuss the single-stage decision problem, on which many existing methods have been developed. In Section 1.2, we consider the multi-stage decision problem that can be more challenging due to the existence of time-varying treatment effects. In Section 1.3, we introduce three main problems that this dissertation focuses on, and outline the organization of the subsequent chapters.

1.1 Single-Stage Decision Problems

For a single-stage decision problem, each data point consists of a covariate vector $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ incorporating the individual characteristics or contextual information, a treatment assignment $A \in \mathcal{A}$ and an outcome $Y \in \mathcal{Y} \subseteq \mathbb{R}$. Assume without loss of generality that a larger outcome Y is

better. One important goal is to find the optimal *Individualized Treatment Rule (ITR)* $d : \mathcal{X} \to \mathcal{A}$ that maximizes the expected outcome (Manski, 2004):

$$d^\star \in \underset{d:\mathcal{X}\to\mathcal{A}}{\operatorname{argmax}} \Big\{ \overbrace{\mathcal{V}(d) := \mathbb{E}[Y|A = d(\boldsymbol{X})]}^{\text{single-stage value function}} \Big\}.$$

One approach for estimating an optimal ITR, known as the *regression-based approach*, is to estimate some conditional mean functions associated with the optimal ITR. Specifically, consider the *Q-function* $Q(\boldsymbol{x}, a) := \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}, A = a)$ as a function of the covariates $\boldsymbol{x} \in \mathcal{X}$ and the treatment assignment $a \in \mathcal{A}$. Then the optimal ITR is induced by $d^\star(\boldsymbol{x}) = \operatorname{argmin}_{a\in\mathcal{A}} Q(\boldsymbol{x}, a)$. The approach based on the Q-function estimate is known as the *Q-Learning* (Watkins, 1989; Qian and Murphy, 2011). In the binary treatment case $\mathcal{A} = \{0, 1\}$, such an approach at the population level is equivalent to estimating the treatment-covariate interaction effect $C(\boldsymbol{x}) := Q(\boldsymbol{x}, 1) - Q(\boldsymbol{x}, 0)$. This is also known as the *Conditional Average Treatment Effect (CATE)* in the causal inference literature, and $a \times C(\boldsymbol{x}) = Q(\boldsymbol{x}, a) - Q(\boldsymbol{x}, 0)$ is the *blip-to-zero function* in Robins (1994)'s *Structural Mean Model (SMM)*. In particular, the optimal ITR at the population level can be represented by the sign of the CATE $d^*(\boldsymbol{x}) = \mathbb{1}[C(\boldsymbol{x}) \geqslant 0]$.

In the binary treatment case, the regression-based optimal ITR can be further obtained via two main different strategies. The first strategy, known as the *A-Learning* (Murphy, 2003), estimates the CATE from the semiparameteric model:

$$Y = m(\boldsymbol{X}) + A \times C(\boldsymbol{X}; \boldsymbol{\beta}) + \epsilon; \quad \mathbb{E}(\epsilon|\boldsymbol{X}, A) = 0. \tag{1.1}$$

Here, the CATE $C(\boldsymbol{X}; \boldsymbol{\beta})$ is modeled by a p-dimensional parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$ of interest, and $m(\boldsymbol{X})$ is a nuisance function of covariates. When p is large, a sparse linear model $C(\boldsymbol{X}; \boldsymbol{\beta}) = \boldsymbol{X}_{\mathcal{S}}^\intercal \boldsymbol{\beta}_{\mathcal{S}}$ is assumed to select a relevant variable subset $\mathcal{S} \subseteq \{1, 2, \cdots, p\}$ (Imai and Ratkovic, 2013; Lu et al., 2013; Shi et al., 2016; Zhao et al., 2017; Jeng et al., 2018; Nie and Wager, 2020). When targeting the CATE parameter $\boldsymbol{\beta}$, the treatment-free effect function $m(\cdot)$ is nuisance and assumed nonparametric. To achieve the semiparametric efficiency when estimating $\boldsymbol{\beta}$, Lu et al. (2013); Zhao et al. (2017);

Nie and Wager (2020) considered Robinson (1988)'s transformation:

$$Y - \mathbb{E}(Y|\boldsymbol{X}) = [A - \mathbb{E}(A|\boldsymbol{X})]C(\boldsymbol{X};\boldsymbol{\beta}) + \epsilon.$$

Denote the nuisance functions $\mu(\boldsymbol{x}) := \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x})$ and $\pi(\boldsymbol{x}) := \mathbb{E}(A|\boldsymbol{X} = \boldsymbol{x})$. Let $(\widehat{\mu}, \widehat{\pi})$ be some estimates of (μ, π). Then the CATE parameter $\boldsymbol{\beta}$ can be estimated from the least-squares problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\mathbb{E}_n\Big\{Y - \widehat{\mu}(\boldsymbol{X}) - [A - \widehat{\pi}(\boldsymbol{X})]C(\boldsymbol{X};\boldsymbol{\beta})\Big\}^2.$$

The corresponding estimating function is equivalent to the *G-Estimation* for the single-stage doubly robust SMM (Bickel and Kwon, 2001):

$$\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \widehat{\mu}, \widehat{\pi}) = \Big\{Y - \widehat{\mu}(\boldsymbol{X}) - [A - \widehat{\pi}(\boldsymbol{X})]C(\boldsymbol{X};\boldsymbol{\beta})\Big\}[A - \widehat{\pi}(\boldsymbol{X})]\frac{\partial}{\partial\boldsymbol{\beta}}C(\boldsymbol{X};\boldsymbol{\beta}).$$

In particular, the G-estimator $\widehat{\boldsymbol{\beta}}_n$ as the solution to the empirical estimating equation $\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \widehat{\mu}_n, \widehat{\pi}_n)] = \mathbf{0}$ is consistent and asymptotic normal if at least one of $\widehat{\mu} = \mu$ and $\widehat{\pi} = \pi$. This is known as the *double robustness* property. If we assume that $(\widehat{\mu}, \widehat{\pi}) = (\mu, \pi)$ and $\mathsf{Var}(\epsilon|\boldsymbol{X}, A)$ is a constant, then $\widehat{\boldsymbol{\beta}}_n$ is *semiparametric efficient*. In order to ensure the estimation effects from $(\widehat{\mu}_n, \widehat{\pi}_n)$ are $\sqrt{n}$-negligible in $\widehat{\boldsymbol{\beta}}_n = \widehat{\boldsymbol{\beta}}_n(\widehat{\mu}_n, \widehat{\pi}_n)$, it requires that $(\widehat{\mu}_n, \widehat{\pi}_n) = (\mu, \pi) + o_{\mathbb{P}}(n^{-1/2})$. For a less restrictive rate requirement on $(\widehat{\mu}_n, \widehat{\pi}_n)$, Zheng and van der Laan (2010); Chernozhukov et al. (2018a,b) considered the nuisance function estimates $(\widehat{\mu}_n^{(-i)}(\boldsymbol{X}_i), \widehat{\pi}_n^{(-i)}(\boldsymbol{X}_i))$ at the i-th sample point, where $(\widehat{\mu}_n^{(-i)}, \widehat{\pi}_n^{(-i)})$ are obtained from a sub-sample excluding the i-th sample point. The corresponding *cross-fitting estimate* of $\boldsymbol{\beta}$ solving $(1/n)\sum_{i=1}^n \boldsymbol{\phi}_{\text{eff},i}\left(\boldsymbol{\beta}; \widehat{\mu}_n^{(-i)}(\boldsymbol{X}_i), \widehat{\pi}_n^{(-i)}(\boldsymbol{X}_i)\right) = \mathbf{0}$ is semiparametric efficient under the looser condition $\left\|\widehat{\mu}_n^{(-i)} - \mu\right\|_{L^2(\mathbb{P})}\left\|\widehat{\pi}_n^{(-i)} - \pi\right\|_{L^2(\mathbb{P})} = o_{\mathbb{P}}(n^{-1/2})$. Such a property is known as the *locally double robustness* (Chernozhukov et al., 2018c) or the *rate double robustness* (Rotnitzky et al., 2021).

Estimating a parametric CATE in A-Learning relies on the parametric model assumption, and hence may suffer from potential model misspecification. It can be desirable to approximate it using flexible nonparametric regression or machine learning approaches. This problem has been intensively studied in the causal inference literature (Dorie et al., 2019; Guo et al., 2020), and many flexible modeling methods have been proposed. Some main examples include nonparametric

regression of the doubly robust transformed outcome (Kennedy et al., 2017; Semenova and Chernozhukov, 2017; Kennedy, 2020; Curth et al., 2020), the index models (Song et al., 2017; Liang and Yu, 2020; Guo et al., 2021), the Generalized Additive Model (GAM) (Moodie et al., 2014), the local methods (Abrevaya et al., 2015; Bertsimas and Kallus, 2020), the ClassificAtion-Regression Tree (CART) (Su et al., 2009; Athey and Imbens, 2016; Bertsimas et al., 2019), the Multivariate Adaptive Regression Spline (MARS) and the boosting estimates (Powers et al., 2018), the random regression forest models (Foster et al., 2011; Wager and Athey, 2018; Friedberg et al., 2020), the Bayesian Additive Regression Tree (BART) (Hill, 2011; Hahn et al., 2020), the Gaussian process (Alaa and van der Schaar, 2017, 2018), the Reproducing Kernel Hilbert Space (RKHS) (Bertsimas and Koduri, 2021), the neural network (Johansson et al., 2016; Shalit et al., 2017; Louizos et al., 2017; Yoon et al., 2018; Yao et al., 2018; Johansson et al., 2020; Curth and van der Schaar, 2021), and the meta learners (Künzel et al., 2019).

Instead of estimating the CATE function $\boldsymbol{x} \mapsto C(\boldsymbol{x})$ from Model (1.1), another strategy is to directly estimate the optimal ITR $\boldsymbol{x} \mapsto \mathsf{sign}[C(\boldsymbol{x})]$ from a weighted loss minimization problem:

$$\min_{f:\mathcal{X}\to[-1,1]} \mathbb{E}\left\{w(\boldsymbol{X},A)\ell\Big(Y,(A-1/2)\times f(\boldsymbol{X})\Big)\right\}, \tag{1.2}$$

where $w(\boldsymbol{X},A)\geqslant 0$ is a weight function, and $\ell(y,\widehat{y})\geqslant 0$ is a general loss function. In particular, it requires the weight function $w(\boldsymbol{X},A)$ to satisfy the following balancing condition (Wallace and Moodie, 2015):

$$w(\boldsymbol{x},1)\pi(\boldsymbol{x}) = w(\boldsymbol{x},0)[1-\pi(\boldsymbol{x})]; \quad \forall \boldsymbol{x}\in\mathcal{X}. \tag{1.3}$$

As a concrete example, the *Inverse Probability Weights (IPWs)* $w(\boldsymbol{x},1)=\pi(\boldsymbol{x})^{-1}$ and $w(\boldsymbol{x},0)=[1-\pi(\boldsymbol{x})]^{-1}$ satisfies (1.3). Furthermore, the *overlap weights* (Crump et al., 2006, 2009) $w(\boldsymbol{x},1)=1-\pi(\boldsymbol{x})$ and $w(\boldsymbol{x},0)=\pi(\boldsymbol{x})$ can be another example. Here, if we define $\mathbb{P}_w \ll \mathbb{P}$ such that $\mathrm{d}\mathbb{P}_w/\mathrm{d}\mathbb{P} := w(\boldsymbol{X},A)$, then $\mathbb{P}_w(A=1|\boldsymbol{X})=\mathbb{P}_w(A=0|\boldsymbol{X})=1/2$, and for any $h:\mathcal{X}\to\mathbb{R}$, we have $\mathbb{E}_w[h(\boldsymbol{X})|A=1]=\mathbb{E}_w[h(\boldsymbol{X})|A=0]=\mathbb{E}_w[h(\boldsymbol{X})]$. That is, the weight $w(\boldsymbol{X},A)$ is a specific form of Rosenbaum and Rubin (1983)'s balancing score in the sense that $\boldsymbol{X}\perp\!\!\!\perp A$ under $\mathbb{P}_w$. Therefore, the balancing condition (1.2) can correspond to the more general *inverse Covariate Balancing*

Propensity Score (CBPS) weight (Imai and Ratkovic, 2014, 2015; Li et al., 2018; Wong and Chan, 2018; Fong et al., 2018; Zhao, 2019; Li and Li, 2019; Wang and Zubizarreta, 2020; Ning et al., 2020; Bennett et al., 2020; Josey et al., 2020; Fan et al., 2020).

It further requires the loss function $\ell(y, \widehat{y})$ in (1.2) to satisfy the following two conditions under the general Subgroup Identification framework (Chen et al., 2017).

- The score function $\mathcal{S}(y, \widehat{y}) := (\partial/\partial\widehat{y})\ell(y, \widehat{y})$ is strictly increasing in $\widehat{y}$ for every $y \in \mathcal{Y}$.
- The utility function $\mathcal{U}(y) := -(\partial/\partial\widehat{y})\ell(y, 0)$ is strictly monotone in y.

Without loss of generality, assume that $\mathcal{U}(y)$ is strictly increasing in y. Given the conditions on $w(\boldsymbol{X}, A)$ and $\ell(y, \widehat{y})$, the solution $f^\star$ to (1.2) satisfies that for any $\boldsymbol{x} \in \mathcal{X}$, $f^\star(\boldsymbol{x}) \geqslant 0$ if and only if $\mathbb{E}[\mathcal{U}(Y)|\boldsymbol{X} = \boldsymbol{x}, A = 1] \geqslant \mathbb{E}[\mathcal{U}(Y)|\boldsymbol{X} = \boldsymbol{x}, A = 0]$. As a special case, suppose the distribution of Y belongs to the exponential family with the canonically parametrized negative log-likelihood function $\ell(y, \eta) := -y\eta + \psi(\eta) - \log h(y)$, where η is the canonical parameter and $\psi(\eta) = \log \int h(y) e^{\eta y} \mathrm{d}y$ is the log-partition function. Then the corresponding utility function is $\mathcal{U}(Y) = Y - \psi'(0)$, so that $\mathsf{sign}[f^\star(\boldsymbol{x})] = \mathsf{sign}[C(\boldsymbol{x})]$ for all $\boldsymbol{x} \in \mathcal{X}$. Therefore, the solution $f^\star$ to (1.2) is Fisher consistent with the optimal ITR. In this case, (1.2) with the negative log-likelihood loss function $\ell(y, \widehat{y})$ corresponds to the Maximal Likelihood Estimate (MLE) under the working model $(\psi')^{-1}[\mathbb{E}(Y|\boldsymbol{X}, A)] = (A - 1/2) \times f(\boldsymbol{X})$ with zero treatment-free effect. In particular, we have $\psi'[(1/2)f^\star(\boldsymbol{x})] - \psi'[-(1/2)f^\star(\boldsymbol{x})] = C(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$, where the function $\eta \mapsto \psi'(\eta/2) - \psi'(-\eta/2)$ is strictly increasing. In this way, the estimation of the treatment-free effect can be avoided. Based on this framework, Tian et al. (2014); Xu et al. (2015); Chen et al. (2017); Qi and Liu (2018); Qi et al. (2020) can consider the ITR problem for continuous, binary and survival outcomes with a flexible decision function class $f \in \mathcal{F}$.

Problem (1.2) with the squared loss $\ell(y, \widehat{y}) = (1/2)(y - \widehat{y})^2$ corresponds to the solution $f^\star(\boldsymbol{x}) = C(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathcal{X}$. In this case, the general use of the weight function $w(\boldsymbol{X}, A)$ satisfying (1.3) was studied in Huang et al. (2014); Wallace and Moodie (2015); Simoneau et al. (2020); Schulz and Moodie (2021). Specifically, they considered the estimating function:

$$\boldsymbol{\phi}(\boldsymbol{\beta}, \boldsymbol{\eta}; \widehat{w}) := \widehat{w}(\boldsymbol{X}, A)\Big\{Y - \boldsymbol{X}^\intercal\boldsymbol{\eta} - (A - 1/2)\boldsymbol{X}^\intercal\boldsymbol{\beta}\Big\} \begin{pmatrix} \boldsymbol{X} \\ (A - 1/2)\boldsymbol{X} \end{pmatrix}. \tag{1.4}$$

Here, $\boldsymbol{X}^\intercal\boldsymbol{\eta}$ and $\boldsymbol{X}^\intercal\boldsymbol{\beta}$ are the parametric models for the treatment-free effect and the CATE respectively. The parameters $(\boldsymbol{\beta}, \boldsymbol{\eta})$ are simultaneously estimated by solving the empirical estimating equations based on (1.4). The estimating function (1.4) is doubly robust if either the estimated weight function $\widehat{w}(\boldsymbol{X}, A)$ satisfies (1.3), or the working linear model $\boldsymbol{X}^\intercal\boldsymbol{\eta}$ is correct for the treatment-free effect.

Other than the regression-based approach, another approach for estimating the optimal ITR, known as the *direct-search* approach, is to directly estimate the value function $\mathcal{V}(d)$ for every ITR d using the *IPW Estimate (IPWE)*

$$\widehat{\mathcal{V}}_{\mathrm{IPWE},n}(d) := \mathbb{E}_n\left\{\frac{\mathbb{1}[d(\boldsymbol{X}) = A]}{p_{\mathscr{A}}(A|\boldsymbol{X})}Y\right\},$$

where $p_{\mathscr{A}}(a|\boldsymbol{x}) := \mathbb{P}(A = a|\boldsymbol{X} = \boldsymbol{x})$. The corresponding optimal ITR is $\widehat{d}_{\mathrm{IPWE},n} \in \mathrm{argmax}_{d\in\mathcal{D}}\,\widehat{\mathcal{V}}_{\mathrm{IPWE}}(d)$, where $\mathcal{D} \subseteq \{d : \mathcal{X} \to \mathcal{A}\}$ is a pre-specified function class of ITRs. Beygelzimer and Langford (2009); Laber and Zhao (2015); Zhu et al. (2017); Kallus (2017) considered $\mathcal{D}$ as the class of decision trees, and introduced the splitting criteria that maximize the IPWE in the corresponding CART algorithm. Kitagawa and Tetenov (2018) considered $\mathcal{D}$ as a general VC class of ITRs, and established the $\sqrt{n}$-regret bound and the minimax rate optimality of $\widehat{d}_{\mathrm{IPWE},n}$. For implementation, they used the Mixed Integer Programming (MIP) to maximize the IPWE over the linear ITR class $\mathcal{D}$. In order to overcome the challenge of nonconvex optimization, Zhao et al. (2012) reformulated the IPWE maximization problem into the minimization of an outcome-weighted misclassification error. The corresponding *Outcome Weighted Learning (OWL)* problem is

$$\min_{f\in\mathcal{F}}\left\{\mathbb{E}_n\left(\frac{Y}{p_{\mathscr{A}}(A|\boldsymbol{X})}\phi[(2A-1)f(\boldsymbol{X})]\right) + \lambda_n\|f\|_{\mathcal{F}}^2\right\},$$

where ϕ is a margin-based convex surrogate loss function, and $\mathcal{F}$ is a pre-specified function class of $\{f : \mathcal{X} \to \mathbb{R}\}$, and $\lambda_n\|\cdot\|_{\mathcal{F}}$ is the functional penalty associated with $\mathcal{F}$. The OWL framework can allow general types of outcomes, such as the binary (Huang and Fong, 2014) and survival (Zhao et al., 2015b; Cui et al., 2017) outcomes, and the applications of any supervised learning methods, such as the bagging and neural network (Mi et al., 2019). To reduce the finite sample variance, Zhou et al. (2017); Liu et al. (2018) further proposed the *Residual Weighted Learning (RWL)* with

Y replaced by $Y - g(\boldsymbol{X})$ for some function $g : \mathcal{X} \to \mathbb{R}$. To handle the possibly negative weights and gain more robustness in presence of covariate outliers, Huang and Fong (2014); Zhou et al. (2017); Qiu et al. (2018) considered the nonconvex ramp loss function for ψ. In particular, the weighted loss functions $\frac{Y}{p_{\mathscr{A}}(A|\boldsymbol{X})}\phi[(2A-1)f(\boldsymbol{X})]$ and $\frac{|Y|}{p_{\mathscr{A}}(A|\boldsymbol{X})}\phi[(2A-1)\mathsf{sign}(Y)f(\boldsymbol{X})]$ are equivalent in this case. Moreover, the ramp loss with a well tuned bandwidth parameter can converge to the 0-1 loss. When the number of variables p is large, sparse penalties can be further incorporated in the OWL framework (Song et al., 2015a; Xu et al., 2015). For multiple and continuous treatment problems, the extensions were studied in Chen et al. (2016, 2018); Lou et al. (2018); Zhou et al. (2018a); Liang et al. (2018); Huang et al. (2019); Fu et al. (2019); Zhang et al. (2020); Meng et al. (2020).

In observational studies, the propensity score function $p_{\mathscr{A}}(A|\boldsymbol{X})$ needs to be estimated from data. In order to protect the risk of misspecifying the propensity score model, the *Augmented IPWE (AIPWE)* of $\mathcal{V}(d)$ was introduced in the literature:

$$\begin{aligned}\widehat{\mathcal{V}}_{\text{AIPWE},n}(d;\widehat{Q},\widehat{p}_{\mathscr{A}}) :=& \mathbb{E}_n\left\{\frac{\mathbb{1}[d(\boldsymbol{X})=A]}{\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})}Y - \frac{\mathbb{1}[d(\boldsymbol{X})=A]-\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})}{\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})}\widehat{Q}(\boldsymbol{X},d)\right\}\\ =& \mathbb{E}_n\left\{\widehat{Q}(\boldsymbol{X},d) + \frac{\mathbb{1}[d(\boldsymbol{X})=A]}{\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})}[Y-\widehat{Q}(\boldsymbol{X},A)]\right\},\end{aligned}$$

where for a general function $h(\boldsymbol{x},a)$ and an ITR $d : \mathcal{X} \to \mathcal{A}$, we denote $h(\boldsymbol{x},d) := \sum_{a\in\mathcal{A}} h(\boldsymbol{x},a)\mathbb{1}[d(\boldsymbol{x})=a]$. The first definition can be obtained from the efficient influence function under the missing data framework (Robins et al., 1994) or Targeted Minimum Loss-based Estimation (TMLE) (van der Laan and Rubin, 2006; van der Laan and Rose, 2018). The second equivalent definition is represented with the additive augmented term $\frac{\mathbb{1}[d(\boldsymbol{X})=A]}{\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})}[Y-\widehat{Q}(\boldsymbol{X},A)]$ to the predicted Q-function $\widehat{Q}(\boldsymbol{X},d)$. The AIPWE is doubly robust in the sense that either $\widehat{Q}=Q$ or $\widehat{p}_{\mathscr{A}}=p_{\mathscr{A}}$ implies that the $\widehat{\mathcal{V}}_{\text{AIPWE},n}(d;\widehat{Q},\widehat{\pi})$ is a consistent estimate of $\mathcal{V}(d)$ (Dudík et al., 2011; Zhang et al., 2012b). Notice that $\max_{d:\mathcal{X}\to\mathcal{A}}\widehat{\mathcal{V}}_{\text{AIPWE},n}(d;\widehat{Q},\widehat{p}_{\mathscr{A}})$ is equivalent to $\max_{d:\mathcal{X}\to\mathcal{A}}\left\{\widehat{\mathcal{V}}_{\text{AIPWE},n}(d;\widehat{Q},\widehat{p}_{\mathscr{A}}) - \widehat{\mathcal{V}}_{\text{AIPWE},n}(-d;\widehat{Q},\widehat{p}_{\mathscr{A}}) := \mathbb{E}_n[\widehat{C}_{\text{AIPWE}}(\boldsymbol{X})][2d(\boldsymbol{X})-1]\right\}$, where

$$\widehat{C}_{\text{AIPWE}}(\boldsymbol{X}) = \widehat{C}_{\text{AIPWE}}(\boldsymbol{X};\widehat{Q},\widehat{p}_{\mathscr{A}}) := \widehat{Q}(\boldsymbol{X},1) - \widehat{Q}(\boldsymbol{X},0) + \frac{2A-1}{\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})}[Y-\widehat{Q}(\boldsymbol{X},A)].$$

This is also known as the *doubly robust score* in Zhou et al. (2018b); Athey and Wager (2021) and the *Doubly Robust (DR) pseudo outcome* of Kennedy (2020); Curth et al. (2020).

Based on the AIPWE, Zhang et al. (2012b) considered the direct maximization of $\widehat{\mathcal{V}}_{\text{AIPWE},n}(d_\eta; \widehat{Q}, \widehat{p}_{\mathscr{A}})$ over a parametric ITR class $\mathcal{D}_\eta = \{d_\eta : \eta \in \Xi\}$ using the genetic algorithm, while Zhang et al. (2015) considered the class of decision lists for $\mathcal{D}$ and proposed an approximation algorithm. Zhang et al. (2012a) proposed *C-Learning* that minimizes the CATE-weighted misclassification error $\mathbb{E}\left\{\left|\widehat{C}_{\text{AIPWE}}(\boldsymbol{X})\right| \times \mathbb{1}\left[d(\boldsymbol{X}) \neq \mathbb{1}\left(\widehat{C}_{\text{AIPWE}}(\boldsymbol{X}) \geqslant 0\right)\right]\right\}$. In a slightly different way, Dudík et al. (2011) also considered cost-sensitive classification algorithms but based on $\widehat{\mathcal{V}}_{\text{AIPWE},n}(\boldsymbol{X}; \widehat{Q}, \widehat{p}_{\mathscr{A}})$ directly. Similar to the regression-based doubly robust estimates, the cross-fitting AIPWE can also be considered, where the estimates $\left(\widehat{Q}_n^{(-i)}(\boldsymbol{X}_i, a), \widehat{p}_{\mathscr{A}}^{(-i)}(a|\boldsymbol{X}_i) : a \in \mathcal{A}\right)$ are used for the AIPWE of the i-th sample point. In this way, Zhou et al. (2018b); Athey and Wager (2021) maximized the cross-fitting AIPWE over the ITR class of decision trees. The performance guarantee of AIPWE maximization over a Donsker ITR class $\mathcal{D}$ was justified in Luedtke and Chambaz (2020); Athey and Wager (2021). Alternatively, Zhao et al. (2019a); Liang et al. (2020) proposed the *Efficient Augmentation and Relaxation Learning (EARL)* with convex surrogate loss relaxation analogous to OWL. Similarly, based on the surrogate loss relaxation, Bennett and Kallus (2020a) proposed the *Efficient Surrogate Policy Risk Minimization (ESPRM)* that solves the variational method-of-moment problem (Bennett and Kallus, 2020b) and established a $\sqrt{n}$-regret bound with the optimal constant dependency.

The IPW in an (A)IPWE can have unbounded variance if there exists some covariate domain on which the propensity score $p_{\mathscr{A}}(a|\boldsymbol{x})$ is close to zero. Swaminathan and Joachims (2015a,b) proposed to trim the IPW from above, and introduced the variance penalty to trade off the trimming bias and the reduced variance:

$$\min_{d\in\mathcal{D}} \left\{ -\mathbb{E}_n \left[\left(\frac{\mathbb{1}[d(\boldsymbol{X}) = A]}{\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})} \wedge M \right) Y \right] + \lambda_n \sqrt{\frac{1}{n} \mathsf{Var}_n \left[\left(\frac{\mathbb{1}[d(\boldsymbol{X}) = A]}{\widehat{p}_{\mathscr{A}}(A|\boldsymbol{X})} \wedge M \right) Y \right]} \right\}.$$

The same strategy was also taken in Kallus and Zhou (2018) when considering the continuous treatment problem. Let $W^d := \frac{\mathbb{1}[d(\boldsymbol{X})=A]}{p_{\mathscr{A}}(A|\boldsymbol{X})}$ and consider the following decompositions:

$$\begin{aligned}
\widehat{\mathcal{V}}_{\text{IPWE},n}(d) &= \mathbb{E}_n[W^d Q(\boldsymbol{X}, A)] &+ \mathbb{E}_n(W^d \epsilon); \\
\widehat{\mathcal{V}}_{\text{AIPWE},n}(d) &= \mathbb{E}_n\left\{\widehat{Q}(\boldsymbol{X}, d) + W^d[Q(\boldsymbol{X}, A) - \widehat{Q}(\boldsymbol{X}, A)]\right\} &+ \mathbb{E}_n(W^d \epsilon).
\end{aligned}$$

Denote $\sigma_i^2 := \mathbb{E}(\epsilon^2|\boldsymbol{X}_i, A_i)$ for $1 \leqslant i \leqslant n$. Then for $\widehat{\mathcal{V}}_n(d) = \widehat{\mathcal{V}}_{\text{IPWE},n}(d)$ (*resp.* $\widehat{\mathcal{V}}_n(d) = \widehat{\mathcal{V}}_{\text{AIPWE},n}(d)$), the *Conditional Mean Square Error (CMSE)* is given by

$$\begin{aligned}\text{CMSE}\left\{\widehat{\mathcal{V}}_n(d)\middle|\{\boldsymbol{X}_i, A_i\}_{i=1}^n\right\} :=& \mathbb{E}\left\{\left[\widehat{\mathcal{V}}_n(d) - \mathbb{E}_n Q(\boldsymbol{X}, d)\right]^2\middle|\{\boldsymbol{X}_i, A_i\}_{i=1}^n\right\} \\ =& \mathbb{E}_n\left\{[W^d q(\boldsymbol{X}, A) - q(\boldsymbol{X}, d)]^2\right\} + \frac{1}{n}\mathbb{E}_n[\sigma^2 (W^d)^2],\end{aligned}$$

where $q := Q$ (*resp.* $q = Q - \widehat{Q}$). In particular, the decomposition of the CMSE entails the bias-variance trade-off due to the sample weights $\{W_i^d\}_{i=1}^n$. This also explains why the variance of $\widehat{\mathcal{V}}_n(d)$ can be large if $p_{\mathscr{A}}(A_i|\boldsymbol{X}_i) \approx 0 \Leftrightarrow W_i(d) \gg 0$. In order to minimize the CMSE, Hirshberg and Wager (2017); Kallus (2018, 2020); Kallus et al. (2021) considered the criteria

$$\mathfrak{E}^2(\boldsymbol{W}_n, d; \mathcal{Q}, \boldsymbol{\sigma}_n^2) := \sup_{q\in\mathcal{Q}}\left\{\frac{1}{n}\sum_{i=1}^n \left[W_i q(\boldsymbol{X}_i, A_i) - q\big(\boldsymbol{X}_i, d(\boldsymbol{X}_i)\big)\right]\right\}^2 + \frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 W_i^2,$$

where $\mathcal{Q}$ is a pre-specified function class. Then Kallus (2018) proposed the *balanced policy learning* to obtain the optimal ITR:

$$\min_{d\in\mathcal{D}}\left\{-\widehat{\mathcal{V}}_n(d; \widehat{Q}, \boldsymbol{W}_n^*) + \lambda\mathfrak{E}(\boldsymbol{W}_n^*, d; \mathcal{Q}, \boldsymbol{\sigma}_n^2) : \boldsymbol{W}_n^* \in \operatorname*{argmin}_{\boldsymbol{W}_n\in\Delta^{n-1}} \mathfrak{E}^2(\boldsymbol{W}_n, d; \mathcal{Q}, \boldsymbol{\sigma}_n^2)\right\}.$$

1.2 Multi-Stage Decision Problems

For a T-stage decision problem, each data point consists of a longitudinal trajectory $(\boldsymbol{X}_t, A_t, Y_t : 1 \leqslant t \leqslant T)$, with the time-varying covariates $\boldsymbol{X}_t \in \mathcal{X}_t \subseteq \mathbb{R}^{p_t}$, treatment $A_t \in \mathcal{A}_t$ and outcome $Y_t \in \mathcal{Y}_t \subseteq \mathbb{R}$ for $1 \leqslant t \leqslant T$. A *Dynamic Treatment Regime (DTR)* is defined as a sequence of stagewise decision rules $\boldsymbol{d}_{1:T} = (d_1, d_2, \cdots, d_T) \in \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_T = \boldsymbol{\mathcal{D}}_{1:T}$, where $\mathcal{D}_t = \{d_t : \mathcal{H}_t \to \mathcal{A}_t\}$ consists of all mappings from the stage-t pre-treatment history $\boldsymbol{H}_t := \big((\boldsymbol{X}_s, A_s, Y_s : 1 \leqslant s \leqslant t-1), \boldsymbol{X}_t\big) \in \mathcal{H}_t$ to the stage-t treatment assignment $A_t \in \mathcal{A}_t$. The goal is to find the optimal DTR that maximizes the expected cumulative outcome

$$\boldsymbol{d}_{1:T}^\star \in \operatorname*{argmax}_{\boldsymbol{d}_{1:T}\in\boldsymbol{\mathcal{D}}_{1:T}}\left\{\overbrace{\mathcal{V}(\boldsymbol{d}_{1:T}) := \sum_{t=1}^T \mathbb{E}[Y_t|A_t = d_t(\boldsymbol{H}_t)]}^{T\text{-stage value function}}\right\}.$$

One key challenge of the DTR problem is that the stage-t treatment A_t can have time-varying effects on the post-treatment variables $\big(Y_t, (\boldsymbol{X}_u, A_u, Y_u : t+1 \leqslant u \leqslant T)\big)$. However, a standard regression analysis conditioning on the observed trajectory may cut off all indirect effects such as $A_t \to \boldsymbol{X}_{t+1} \to Y_{t+1}$ (Almirall et al., 2010). As a consequence, $\operatorname{argmax}_{a_t \in \mathcal{A}_t} \mathbb{E}\left\{\sum_{u=t}^{T} Y_u \middle| \boldsymbol{H}_t, A_t = a_t\right\}$ is the not same as the stage-t optimal decision rule $d_t^\star(\boldsymbol{H}_t)$. Therefore, the observed data should be adjusted to unveil the time-varying treatment effects.

The first approach to adjust for the cross-stage treatment effects is to perform stagewise model-based outcome transformations. Specifically, consider the Bellman equations (Bellman, 1966) that recursively define the stagewise *state-value functions* $\{\mathcal{V}_t(\boldsymbol{H}_t)\}_{t=1}^T$ and *Q-functions* $\{\mathcal{Q}_t(\boldsymbol{H}_t, A_t)\}_{t=1}^T$:

$$\begin{aligned} \mathcal{V}_T(\boldsymbol{H}_T) &:= \max_{a_T \in \mathcal{A}_T} \underbrace{\mathbb{E}(Y_T | \boldsymbol{H}_T, A_T = a_T)}_{:=\mathcal{Q}_T(\boldsymbol{H}_T, a_T)}; \\ \mathcal{V}_t(\boldsymbol{H}_t) &:= \max_{a_t \in \mathcal{A}_t} \underbrace{\mathbb{E}\Big\{ Y_t + \mathcal{V}_{t+1}(\boldsymbol{H}_{t+1}) \Big| \boldsymbol{H}_t, A_t = a_t \Big\}}_{:=\mathcal{Q}_t(\boldsymbol{H}_t, a_t)}; \quad t = T-1, T-2, \cdots, 1. \end{aligned} \tag{1.5}$$

Then the stage-t optimal decision rule can be induced by $d_t^\star(\boldsymbol{H}_t) = \operatorname{argmax}_{a_t \in \mathcal{A}_t} \mathcal{Q}_t(\boldsymbol{H}_t, a_t)$. In particular, the stage-T problem with data $(\boldsymbol{H}_T, A_T, Y_T)$ can be handled by any single-stage methods in Section 1.1. For the stage-$t(< T)$ problem, we can consider the q-*outcome* as $Y_t^{(\mathrm{q})} := Y_t + \mathcal{Q}_{t+1}\big(\boldsymbol{H}_{t+1}, d_{t+1}^\star(\boldsymbol{H}_{t+1})\big) = Y_t + \max_{a_{t+1} \in \mathcal{A}_{t+1}} \mathcal{Q}_{t+1}(\boldsymbol{H}_{t+1}, a_{t+!})$. By (1.5), we have $\mathbb{E}(Y_t^{(\mathrm{q})} | \boldsymbol{H}_t, A_t) = \mathcal{Q}_t(\boldsymbol{H}_t, A_t)$. That is, the stage-$t$ problem can be solved as a single-stage problem with data $(\boldsymbol{H}_t, A_t, Y_t^{(\mathrm{q})})$. Notice that both the optimal DTR $(d_1^\star, d_2^\star, \cdots, d_T^\star)$ and the stagewise q-outcomes $\{Y_t^{(\mathrm{q})}\}_{t=1}^T$ require the complete knowledge of $\{\mathcal{Q}_t(\boldsymbol{H}_t, A_t)\}_{t=1}^T$. We can consider statistical models for the stagewise Q-functions and perform estimation in a backward stagewise manner. Specifically, at stage $t(< T)$, we first obtain the estimated q-outcome $\widehat{Y}_t^{(\mathrm{q})} := Y_t + \max_{a_{t+1} \in \mathcal{A}_{t+1}} \widehat{\mathcal{Q}}_{t+1}(\boldsymbol{H}_{t+1}, a_{t+1})$ based on the stage-$(t+1)$ Q-function estimate $\widehat{\mathcal{Q}}_{t+1}$. Then we can consider a single-stage regression problem using $\widehat{Y}_t^{(\mathrm{q})}$ as the response and $(\boldsymbol{H}_t, A_t)$ as the covariates to estimate $\widehat{\mathcal{Q}}_t$. Such an approach gives the T-stage *Q-Learning* (Watkins, 1989; Murphy, 2005; Zhao et al., 2009; Goldberg and Kosorok, 2012; Murray et al., 2018; Zhang et al., 2018; Zhu et al., 2019; Ertefaie et al., 2021).

If we further assume that the Q-functions are stationary $\mathbb{E}(Y_t^{(\mathrm{q})} | \boldsymbol{H}_t, A_t) = \mathcal{Q}(\boldsymbol{X}_t, A_t)$ with stationary covariate space $\boldsymbol{X}_t \in \mathcal{X} \subseteq \mathbb{R}^p$ and stationary treatment space $A_t \in \mathcal{A}$ across stages

$1 \leqslant t \leqslant T$, then Q-Learning can be extended to the infinite-horizon setting ($T = +\infty$) as a *reinforcement learning* problem (Sutton and Barto, 2018; Ertefaie and Strawderman, 2018; Shi et al., 2020b; Liao et al., 2020, 2021). As a method closely connected to the infinite-horizon Q-Learning, Luckett et al. (2020) proposed the *V-Learning* framework that estimates the stationary state-value function instead.

In the binary treatment case $\mathcal{A}_t = \{0, 1\}$, we define the stagewise CATE functions $\mathcal{C}_t(\boldsymbol{H}_t) := \mathcal{Q}_t(\boldsymbol{H}_t, 1) - \mathcal{Q}_t(\boldsymbol{H}_t, 0)$ for $1 \leqslant t \leqslant T$. Then the optimal DTR becomes $d_t^\star(\boldsymbol{H}_t) = \mathbb{1}[\mathcal{C}_t(\boldsymbol{H}_t) \geqslant 0]$ $(1 \leqslant t \leqslant T)$. We further introduce the stagewise g-*outcome* as $Y_t^{(\mathrm{g})} := \sum_{u=t}^T Y_u - \sum_{u=t+1}^T \{A_u - \mathbb{1}[\mathcal{C}_u(\boldsymbol{H}_u) \geqslant 0]\} \mathcal{C}_u(\boldsymbol{H}_u)$. It can be shown that $\mathbb{E}(Y_t^{(\mathrm{g})} | \boldsymbol{H}_t, A_t) = \mathcal{Q}_t(\boldsymbol{H}_t, A_t) = \mathcal{Q}_t(\boldsymbol{H}_t, 0) + A_t \mathcal{C}_t(\boldsymbol{H}_t)$. Then the stage-$t$ problem can be solved based on the single-stage semiparametric model:

$$Y_t^{(\mathrm{g})} = m_t(\boldsymbol{H}_t) + A_t \times C_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t) + e_t^{(\mathrm{g})}; \quad \mathbb{E}(e_t^{(\mathrm{g})} | \boldsymbol{H}_t, A_t) = 0. \tag{1.6}$$

This modeling approach is an instance of the optimal *Structural Nested Mean Model (SNMM)* (Robins, 2004).

There are three different estimation strategies for the optimal DTR based on the SNMM. The first strategy, known as the *stagewise A-Learning* (Blatt et al., 2004; Shi et al., 2018a), is implemented analogously to the Q-Learning. Specifically, at stage $t(< T)$, we obtain the estimated g-outcome $\widehat{Y}_t^{(\mathrm{g})} := \sum_{u=t}^T Y_u - \sum_{u=t+1}^T \left\{A - \mathbb{1}[C_u(\boldsymbol{H}_u; \widehat{\boldsymbol{\beta}}_{u,n}) \geqslant 0]\right\} C_u(\boldsymbol{H}_u; \widehat{\boldsymbol{\beta}}_{u,n})$ based on the estimated CATE parameters $\{\widehat{\boldsymbol{\beta}}_{u,n}\}_{u=t+1}^T$ from the subsequent stages. Then we consider the single-stage semiparametric regression problem (1.6) with $\widehat{Y}_t^{(\mathrm{g})}$ as the response. Following this strategy, Shi et al. (2018a) considered the linear working model $C_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t) := \boldsymbol{H}_t^\intercal \boldsymbol{\beta}_t$ and the single-stage efficient estimating function

$$\boldsymbol{\phi}_{\mathrm{eff},t}(\boldsymbol{\beta}; \widehat{\boldsymbol{\eta}}_t, \widehat{\boldsymbol{\alpha}}_t) = [\widehat{Y}_t^{(\mathrm{g})} - m_t(\boldsymbol{H}_t; \widehat{\boldsymbol{\eta}}_t) - A_t \boldsymbol{H}_t^\intercal \boldsymbol{\beta}_t][A_t - \pi_t(\boldsymbol{H}_t; \widehat{\boldsymbol{\alpha}}_t)]\boldsymbol{H}_t,$$

where $m_t(\boldsymbol{H}_t; \boldsymbol{\eta}_t)$ and $\pi_t(\boldsymbol{H}_t; \boldsymbol{\alpha}_t)$ are the parametric models for $\mathbb{E}[Y_t^{(\mathrm{g})} | \boldsymbol{H}_t, A_t = 0]$ and $\mathbb{P}(A_t = 1 | \boldsymbol{H}_t)$ respectively, and $(\widehat{\boldsymbol{\eta}}_t, \widehat{\boldsymbol{\alpha}}_t)$ are the corresponding estimates. On the other hand, Huang et al. (2014); Wallace and Moodie (2015); Simoneau et al. (2020) considered the balancing weight func-

tion (1.3) and proposed the *dynamic Weighted Ordinary Least Squares (dWOLS)* that solves the weighted least-squares problem using $\widehat{Y}_t^{(\mathrm{g})}$ as the response and $(\boldsymbol{H}_t, A_t\boldsymbol{H}_t)$ as the covariates. Both the stagewise A-Learning and dWOLS are doubly robust.

The second strategy to estimate the optimal SNMM, known as the *regret regression*, (Murphy, 2003; Almirall et al., 2010; Henderson et al., 2010; Almirall et al., 2014), exploits the following cross-stage representation that is equivalent to (1.6):

$$\begin{aligned} &\sum_{t=1}^{T} Y_t = \mathcal{V}_0 + \sum_{t=1}^{T} \{A_t - \mathbb{1}[C_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t) \geqslant 0]\}\, C_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t) + \sum_{t=1}^{T+1} \Delta\mathcal{M}_t(\boldsymbol{H}_t), \\ &\text{subject to} \quad \mathbb{E}[\Delta\mathcal{M}_{t+1}(\boldsymbol{H}_{t+1})|\boldsymbol{H}_t, A_t] = 0; \quad 0 \leqslant t \leqslant T. \end{aligned}$$

Here, we denote $(\boldsymbol{H}_0, A_0) := \varnothing$ and $\boldsymbol{H}_{T+1} := (\boldsymbol{H}_T, A_T, Y_T)$ for convenience. Then the following fully parametric least-squares problem is considered:

$$\min_{\left((\boldsymbol{\beta}_t,\boldsymbol{\eta}_t:1\leqslant t\leqslant T),v_0\right)} \mathbb{E}_n \left\{ \sum_{t=1}^{T} Y_t - v_0 - \sum_{t=1}^{T} \{A_t - \mathbb{1}[C_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t) \geqslant 0]\}\, C_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t) - \sum_{t=1}^{T} \Delta m_t(\boldsymbol{H}_t;\boldsymbol{\eta}_t) \right\}^2,$$

where $\{\Delta m_t(\boldsymbol{H}_t;\boldsymbol{\eta}_t)\}_{t=1}^T$ are the parametric models for $\{\Delta\mathcal{M}_t(\boldsymbol{H}_t)\}_{t=1}^T$ subject to $\mathbb{E}[\Delta m_t(\boldsymbol{H}_t;\boldsymbol{\eta}_t)|\boldsymbol{H}_{t-1}, A_{t-1}] = 0$ for $1 \leqslant t \leqslant T$. Although the regret regression can enjoy better efficiency than the stagewise A-Learning if the nuisance models $\{\Delta m_t(\boldsymbol{H}_t;\boldsymbol{\eta}_t)\}_{t=1}^T$ are correct, it can be vulnerable to nuisance model misspecifications.

The third strategy is the *G-Estimation* (Robins, 2004) for the optimal SNMM. Under Model (1.6), the stage-t working g-outcomes is $Y_t^{(\mathrm{g})}(\boldsymbol{\beta}_{(t+1):T}) = \sum_{u=t}^{T} Y_u - \sum_{u=t+1}^{T} \{A_u - \mathbb{1}[C_u(\boldsymbol{H}_u;\boldsymbol{\beta}_u) \geqslant 0]\}\, C_u(\boldsymbol{H}_u;\boldsymbol{\beta}_u)$, and the characterizing moment condition of (1.6) is

$$\mathbb{E}\Bigg\{ \overbrace{\left[Y_t^{(\mathrm{g})}(\boldsymbol{\beta}_{(t+1):T}) - A_t C_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)\right]}^{:=\mathfrak{H}_t(\boldsymbol{\beta}_{t:T})} - \mathbb{E}\left[Y_t^{(\mathrm{g})}(\boldsymbol{\beta}_{(t+1):T}) - A_t C_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)\Big|\boldsymbol{H}_t\right] \Bigg| \boldsymbol{H}_t, A_t \Bigg\} = 0.$$

Let $\boldsymbol{G}_t : \mathcal{H}_t \times \mathcal{A}_t \to \mathbb{R}^p$ be some user-defined instrument function, where $p := \sum_{t=1}^T p_t$ is the dimension of the parameter $\boldsymbol{\beta}_{1:T}$. Then the G-estimating function is

$$\boldsymbol{\phi}(\boldsymbol{\beta}_{1:T}) := \sum_{t=1}^{T} \Big\{ \mathfrak{H}_t(\boldsymbol{\beta}_{t:T}) - \mathbb{E}[\mathfrak{H}_t(\boldsymbol{\beta}_{t:T})|\boldsymbol{H}_t] \Big\} \Big\{ \boldsymbol{G}_t(\boldsymbol{H}_t, A_t) - \mathbb{E}[\boldsymbol{G}_t(\boldsymbol{H}_t, A_t)|\boldsymbol{H}_t] \Big\}.$$

At stage t, it requires the nuisance functions of the treatment-free effect $\mathbb{E}[\mathfrak{H}_t(\boldsymbol{\beta}_{t:T})|\boldsymbol{H}_t]$ and the propensity score $p_{\mathscr{A},t}(a_t|\boldsymbol{h}_t) := \mathbb{P}(A_t = a_t|\boldsymbol{H}_t = \boldsymbol{h}_t)$ in evaluating $\mathbb{E}[\boldsymbol{G}_t(\boldsymbol{H}_t, A_t)|\boldsymbol{H}_t]$. If either of the models of $\mathbb{E}[\mathfrak{H}_t(\boldsymbol{\beta}_{t:T})|\boldsymbol{H}_t]$ and $p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)$ is correct for $1 \leqslant t \leqslant T$, then we have $\mathbb{E}[\boldsymbol{\phi}(\boldsymbol{\beta}_{1:T})] = \mathbf{0}$ at the true parameter $\boldsymbol{\beta}_{1:T}$. This gives the *stagewise double robustness* of G-Estimation. If the models of $\{\mathbb{E}[\mathfrak{H}_t(\boldsymbol{\beta}_{t:T})|\boldsymbol{H}_t]\}_{t=1}^T$ and $\{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)\}_{t=1}^T$ are correct, then there exists optimal instrument functions $\{\boldsymbol{G}_{\text{eff},t}(\boldsymbol{H}_t, A_t)\}_{t=1}^T$ such that the corresponding estimating function $\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}_{1:T})$ is *semiparametric efficient* (Robins, 1994, 2004). However, the closed forms of $\{\boldsymbol{G}_{\text{eff},t}(\boldsymbol{H}_t, A_t)\}_{t=1}^T$ are intractable unless we assume the condition $\mathsf{Var}[\mathfrak{H}_t(\boldsymbol{\beta}_{t:T})|\boldsymbol{H}_t, A_t] = \mathsf{Var}[\mathfrak{H}_t(\boldsymbol{\beta}_{t:T})|\boldsymbol{H}_t]$ for $1 \leqslant t \leqslant T$. Even in this case, the efficient instrument functions $\{\boldsymbol{G}_{\text{eff},t}(\boldsymbol{H}_t, A_t)\}_{t=1}^T$ can involve too many vector-valued nuisance functions and be hard to estimate (Vansteelandt and Joffe, 2014).

The G-estimating equations are solved simultaneously across stages. If we assume that the t-th block of the estimating function $\boldsymbol{\phi}_t(\boldsymbol{\beta}_{1:T})$ corresponding to $\boldsymbol{\beta}_t$ satisfies

$$\boldsymbol{\phi}_t(\boldsymbol{\beta}_{1:T}) = \boldsymbol{\phi}_t(\boldsymbol{\beta}_{t:T}) = \sum_{u=t}^{T} \Big\{\mathfrak{H}_u(\boldsymbol{\beta}_{u:T}) - \mathbb{E}[\mathfrak{H}_u(\boldsymbol{\beta}_{u:T})|\boldsymbol{H}_u]\Big\}\Big\{\boldsymbol{G}_{tu}(\boldsymbol{H}_u, A_u) - \mathbb{E}[\boldsymbol{G}_{tu}(\boldsymbol{H}_u, A_u)|\boldsymbol{H}_u]\Big\},$$

then G-Estimation is equivalent to a backward stagewise procedure in Robins (2004, Seciton 7.2) and Moodie et al. (2007, Section 3.3.2). Specifically, at stage $t(< T)$, we obtain the estimated g-outcomes $\{Y_u^{(\text{g})}(\widehat{\boldsymbol{\beta}}_{(u+1):T})\}_{u=t}^T$ for the current and subsequent stages based on the estimated parameters $\{\widehat{\boldsymbol{\beta}}_u\}_{u=t+1}^T$. Then $\{\mathfrak{H}_u(\widehat{\boldsymbol{\beta}}_{u:T})\}_{u=t+1}^T$ can be computed, and $\mathfrak{H}_t(\boldsymbol{\beta}_t, \widehat{\boldsymbol{\beta}}_{(t+1):T}) = Y_t^{(\text{g})}(\widehat{\boldsymbol{\beta}}_{(t+1):T}) - A_t C_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t)$ is expressed as a function of $\boldsymbol{\beta}_t$ only. The stage-t estimate $\widehat{\boldsymbol{\beta}}_t$ is obtained by solving the estimating equations $\mathbb{E}_n[\boldsymbol{\phi}_t(\boldsymbol{\beta}_t, \widehat{\boldsymbol{\beta}}_{(t+1):T})] = \mathbf{0}$. If we further assume that $\boldsymbol{G}_{tu}(\boldsymbol{H}_u, A_u) = A_t \times (\partial/\partial\boldsymbol{\beta}_t) C_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t) \mathbb{1}(u = t)$ for $1 \leqslant t, u \leqslant T$, then G-Estimation is equivalent to the stagewise A-Learning in this case (Schulte et al., 2014).

The q-outcome and g-outcome are both nonsmooth functions of the parameters of interest. Specifically, in the binary treatment case where we consider the semiparametric model (1.6), we have for $1 \leqslant t \leqslant T-1$,

$$\begin{aligned}
Y_t^{(\text{q})} &= Y_t^{(\text{q})}(\boldsymbol{\beta}_{t+1}) = Y_t + m_{t+1}(\boldsymbol{H}_{t+1}) + \mathbb{1}[C_{t+1}(\boldsymbol{H}_{t+1}; \boldsymbol{\beta}_{t+1}) \geqslant 0] C_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t); \\
Y_t^{(\text{g})} &= Y_t^{(\text{g})}(\boldsymbol{\beta}_{(t+1):T}) = \sum_{u=t}^{T} Y_u - \sum_{u=t+1}^{T} \{A_u - \mathbb{1}[C_u(\boldsymbol{H}_u; \boldsymbol{\beta}_u) \geqslant 0]\} C_u(\boldsymbol{H}_u; \boldsymbol{\beta}_u).
\end{aligned}$$

In particular, the indicator function $\mathbb{1}[C_u(\boldsymbol{H}_u;\boldsymbol{\beta}_u) \geqslant 0]$ is nonsmooth in $\boldsymbol{\beta}_u$ no matter how $C_u(\boldsymbol{H}_u;\boldsymbol{\beta}_u)$ depends on $\boldsymbol{\beta}_u$. This can result in an exceptional law if $\mathbb{P}[C_u(\boldsymbol{H}_u;\boldsymbol{\beta}_u) = 0] > 0$ that leads to a biased parameter estimate of G-Estimation (Robins, 2004; Moodie and Richardson, 2010). The same nonregularity was also studied for Q-Learning (Chakraborty et al., 2010; Laber et al., 2014b) and dWOLS (Simoneau et al., 2018).

There are several strategies in the literature to overcome the challenge of nonregularity. If the main goal is to perform hypothesis testings on the treatment effect parameters $\boldsymbol{\beta}_{1:T}$ or to construct their confidence intervals, then an adaptive m-out-of-n Bootstrap procedure was proposed to obtain valid confidence intervals, where the Bootstrap sample size m is chosen adaptively to the data for proper coverage (Chakraborty et al., 2013; Simoneau et al., 2018). If the estimation properties are of the main concern, then several shrinkage estimates were proposed to modify the estimated q-and g-outcomes. Specifically, the estimated optimal CATE $C_t(\boldsymbol{H}_t;\widehat{\boldsymbol{\beta}}_{t,n})\mathbb{1}[C_t(\boldsymbol{H}_t;\widehat{\boldsymbol{\beta}}_{t,n}) \geqslant 0]$ can be replaced by the *Zeroing Instead of Plugging In (ZIPI)* estimate $C_t(\boldsymbol{H}_t;\widehat{\boldsymbol{\beta}}_{t,n})\mathbb{1}[C_t(\boldsymbol{H}_t;\widehat{\boldsymbol{\beta}}_{t,n}) \geqslant \lambda_n]$ (Moodie and Richardson, 2010; Chakraborty et al., 2010; Zhu et al., 2019). Alternatively, Song et al. (2015b); Goldberg et al. (2013) introduced a subject-specific shrinkage penalty $\lambda_n J[C_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)]$ to the stage-t estimation problem. Other than the shrinkage estimates, Laber et al. (2014a); Linn et al. (2017) proposed to estimate the treatment-free effects $\{m_t(\boldsymbol{H}_t)\}_{t=1}^T$ and the conditional distributions of $\{\mathcal{C}_t(\boldsymbol{H}_t)|(\boldsymbol{H}_{t-1},A_{t-1})\}_{t=1}^T$. Then the q-outcomes can be obtained by $\widehat{Y}_t^{(\mathrm{q})} = Y_t + \widehat{m}_{t+1}(\boldsymbol{H}_{t+1}) + \widehat{\mathbb{E}}_{\mathcal{C}_{t+1}|\boldsymbol{H}_t,A_t}(\mathcal{C}_{t+1}^+|\boldsymbol{H}_t,A_t)$ for $1 \leqslant t \leqslant T-1$, which are smooth functions of data.

Besides the model-based approach, another framework, known as the *Marginal Structural Mean Model (MSMM)* (Robins, 1998), allows the direct estimation of the stagewise state-value functions. Fix a DTR $\boldsymbol{d}_{1:T} \in \boldsymbol{\mathcal{D}}_{1:T}$. Define the DTR-specific state-value functions $\{\mathcal{V}_t^{\boldsymbol{d}}(\boldsymbol{H}_t)\}_{t=1}^T$, Q-functions $\{\mathcal{Q}_t^{\boldsymbol{d}}(\boldsymbol{H}_t,A_t)\}_{t=1}^T$, and *Bellman-error* functions $\{\Delta\mathcal{M}_t^{\boldsymbol{d}}(\boldsymbol{H}_t)\}_{t=2}^{T+1}$ from the Bellman equations

$$
\left\{\begin{array}{lcl}
\mathcal{Q}_T^{\boldsymbol{d}}(\boldsymbol{H}_T,A_T) & := & \mathbb{E}(Y_T|\boldsymbol{H}_T,A_T) \\
\Delta\mathcal{M}_{T+1}^{\boldsymbol{d}}(\boldsymbol{H}_{T+1}) & := & Y_T - \mathcal{Q}_T^{\boldsymbol{d}}(\boldsymbol{H}_T,A_T) \\
\mathcal{V}_T^{\boldsymbol{d}}(\boldsymbol{H}_T) & := & \mathcal{Q}_T^{\boldsymbol{d}}\big(\boldsymbol{H}_T,d_T(\boldsymbol{H}_T)\big)
\end{array}\right\};
$$

$$
\left\{\begin{array}{lcl}
\mathcal{Q}_t^{\boldsymbol{d}}(\boldsymbol{H}_t,A_t) & := & \mathbb{E}\Big\{Y_t + \mathcal{V}_{t+1}^{\boldsymbol{d}}(\boldsymbol{H}_{t+1})\Big|\boldsymbol{H}_t,A_t\Big\} \\
\Delta\mathcal{M}_{t+1}^{\boldsymbol{d}}(\boldsymbol{H}_{t+1}) & := & Y_t + \mathcal{V}_{t+1}^{\boldsymbol{d}}(\boldsymbol{H}_{t+1}) - \mathcal{Q}_t^{\boldsymbol{d}}(\boldsymbol{H}_t,A_t) \\
\mathcal{V}_t^{\boldsymbol{d}}(\boldsymbol{H}_t) & := & \mathcal{Q}_t^{\boldsymbol{d}}\big(\boldsymbol{H}_t,d_t(\boldsymbol{H}_t)\big)
\end{array}\right| \; t = T-1,\cdots,1 \Bigg\}. \tag{1.7}
$$

Then the stage-1 marginal value $\mathcal{V}_0^{\boldsymbol{d}} := \mathbb{E}[\mathcal{V}_1^{\boldsymbol{d}}(\boldsymbol{H}_1)]$ gives the T-stage value function $\mathcal{V}(\boldsymbol{d}_{1:T})$. Denote $V_{s:t} := \sum_{u=s}^{t} Y_u$ and $W_{s:t}^{\boldsymbol{d}} := \prod_{u=s}^{t} \frac{\mathbb{1}[d_u(\boldsymbol{H}_u)=A_u]}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}$ for $1 \leqslant s \leqslant t \leqslant T$. Murphy et al. (2001) studied the estimation of the MSMM $\mathcal{V}_0^{\boldsymbol{d}}(\boldsymbol{Z};\boldsymbol{\beta}) := \boldsymbol{Z}^\intercal\boldsymbol{\beta}$ using the IPW estimating function $\boldsymbol{\phi}^{\boldsymbol{d}}(\boldsymbol{\beta}) = W_{1:T}^{\boldsymbol{d}}(V_{1:T} - \boldsymbol{Z}^\intercal\boldsymbol{\beta})\boldsymbol{Z}$, where $\boldsymbol{Z}$ is the subject-specific covariate vector, and $\boldsymbol{\beta}$ is the parameter of interest. As a special case, if $Z = 1$, then the estimating function of the MSMM $\mathcal{V}_0^{\boldsymbol{d}}(\beta) := \beta$ simplifies as $\phi^{\boldsymbol{d}}(\beta) = W_{1:T}^{\boldsymbol{d}}(V_{1:T} - \beta)$. The corresponding estimate is $\widehat{\mathcal{V}}_{\text{IPWE},n}(\boldsymbol{d}_{1:T}) = [\mathbb{E}_n(W_{1:T}^{\boldsymbol{d}})]^{-1}[\mathbb{E}_n(W_{1:T}^{\boldsymbol{d}}V_{1:T})]$, which gives the T-stage Hájek IPWE of $\mathcal{V}(\boldsymbol{d}_{1:T})$. To obtain the optimal DTR, Orellana et al. (2010a,b) proposed to first estimate the MSMM $\mathcal{V}_0^{\boldsymbol{d}_\eta}(\boldsymbol{Z};\boldsymbol{\beta})$ with the parametric DTR $\boldsymbol{d}_\eta \in \boldsymbol{\mathcal{D}}_\eta = \{\boldsymbol{d}_\eta : \eta \in \Xi\}$, and then solve $\widehat{\eta}(\boldsymbol{Z}) \in \operatorname{argmax}_{\eta\in\Xi} \mathcal{V}_0^{\boldsymbol{d}_\eta}(\boldsymbol{Z};\widehat{\boldsymbol{\beta}})$ for the optimal DTR parameter. Despite the simplicity of the MSMM compared to the SNMM, the parametric DTR class $\boldsymbol{\mathcal{D}}_\eta$ is typically restrictive and cannot handle too complicated functional forms.

To learn the optimal DTR of flexible functional forms, Zhao et al. (2015a) proposed the T-stage extension of OWL that directly maximizes $\widehat{\mathcal{V}}_{\text{IPWE},n}(\boldsymbol{d}_{1:T}) = \mathbb{E}_n(W_{1:T}^{\boldsymbol{d}}V_{1:T})$. They considered two strategies to relax the T-stage nonsmooth nonconvex function $\boldsymbol{d}_{1:T} \mapsto W_{1:T}^{\boldsymbol{d}}$. The first approach, known as the *Backward OWL (BOWL)*, solves T single-stage OWL problems in a backward stage-wise manner:

$$\widehat{f}_{t,n} \in \operatorname*{argmin}_{f_t\in\mathcal{F}_t} \left\{ \mathbb{E}_n \left(\frac{W_{(t+1):T}^{\widehat{\boldsymbol{f}}_n} V_{t:T}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \phi_t[(2A_t - 1)f_t(\boldsymbol{H}_t)] \right) + \lambda_{t,n}\|f_t\|_{\mathcal{F}_t}^2 \right\}; \quad t = T, \cdots, 1,$$

where $W_{(t+1):T}^{\widehat{\boldsymbol{f}}_n} := \prod_{u=t+1}^{T} \frac{\mathbb{1}[(2A_u-1)\widehat{f}_{u,n}(\boldsymbol{H}_u)\geqslant 0]}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}$ based on $\{\widehat{f}_{u,n}\}_{u=t+1}^{T}$ from the subsequent stages. Jiang et al. (2019) specifically considered the entropy loss functions and established the asymptotic properties of the DTR parameter estimate. The second approach, known as the *Simultaneous OWL (SOWL)*, utilizes a multivariate surrogate loss $\phi : \mathbb{R}^T \to \mathbb{R}_+$ that approximates the multivariate 0-1 loss function $\boldsymbol{u}_{1:T} \mapsto 1 - \prod_{t=1}^{T} \mathbb{1}(u_t \geqslant 0)$. Then SOWL solves the multi-dimensional large-margin classification problem:

$$\widehat{\boldsymbol{f}}_{1:T,n} \in \operatorname*{argmin}_{\boldsymbol{f}_{1:T}\in\boldsymbol{\mathcal{F}}_{1:T}} \left\{ \mathbb{E}_n \left[\frac{V_{1:T}}{\prod_{t=1}^{T} p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \phi \begin{pmatrix} (2A_1 - 1)f_1(\boldsymbol{H}_1) \\ \vdots \\ (2A_T - 1)f_T(\boldsymbol{H}_T) \end{pmatrix} \right] + \lambda_n \|\boldsymbol{f}_{1:T}\|_{\boldsymbol{\mathcal{F}}_{1:T}}^2 \right\}.$$

In particular, when considering the multivariate hinge loss $\phi(\boldsymbol{u}_{1:T}) = 1 + \bigwedge_{t=1}^{T}(u_t - 1)^+$, the dual problem of SOWL is a Quadratic Programming (QP) problem.

Analogous to the single-stage problem, the T-stage IPWE can correspond to an efficient estimating function, which gives the T-stage AIPWE as in the following Theorem 1.1. The same result was also used for Optimal Policy Evaluation (OPE) in the reinforcement learning literature (Jiang and Li, 2016; Thomas and Brunskill, 2016; Kallus and Uehara, 2020).

Theorem 1.1 (T-Stage AIPWE)**.** *Consider a semiparametric model identified from the IPW estimating function $\phi^{\boldsymbol{d}}(\beta) = W^{\boldsymbol{d}}_{1:T}(V_{1:T} - \beta)$. Then the efficient estimating function is*

$$\phi^{\boldsymbol{d}}_{\text{eff}}(\beta) = \mathcal{V}^{\boldsymbol{d}}_1(\boldsymbol{H}_1) + \sum_{t=1}^{T} W^{\boldsymbol{d}}_{1:t}\Delta\mathcal{M}^{\boldsymbol{d}}_{t+1}(\boldsymbol{H}_{t+1}) - \beta.$$

The corresponding semiparametric efficient estimate is

$$\widehat{\mathcal{V}}_{\text{AIPWE},n}(\boldsymbol{d}) = \mathbb{E}_n\left\{\mathcal{V}^{\boldsymbol{d}}_1(\boldsymbol{H}_1) + \sum_{t=1}^{T} W^{\boldsymbol{d}}_{1:t}\Delta\mathcal{M}^{\boldsymbol{d}}_{t+1}(\boldsymbol{H}_{t+1})\right\}.$$

Proof of Theorem 1.1.

$$\phi^{\boldsymbol{d}}_{\text{eff}}(\beta) = \phi^{\boldsymbol{d}}(\beta) - \sum_{t=1}^{T}\Big\{\mathbb{E}[\phi^{\boldsymbol{d}}(\beta)|\boldsymbol{H}_t, A_t] - \mathbb{E}[\phi^{\boldsymbol{d}}(\beta)|\boldsymbol{H}_t]\Big\} \tag{1.8}$$

$$= \quad W^{\boldsymbol{d}}_{1:T}(V_{1:T} - \beta)$$

$$- \sum_{t=1}^{T}\Big\{W^{\boldsymbol{d}}_{1:t}[V_{1:(t-1)} + \mathcal{Q}^{\boldsymbol{d}}_t(\boldsymbol{H}_t, A_t) - \beta] - W^{\boldsymbol{d}}_{1:(t-1)}[V_{1:(t-1)} + \mathcal{V}^{\boldsymbol{d}}_t(\boldsymbol{H}_t) - \beta]\Big\} \tag{1.9}$$

$$= W^{\boldsymbol{d}}_{1:T}[Y_T - \mathcal{Q}^{\boldsymbol{d}}_T(\boldsymbol{H}_T, A_T)] + \sum_{t=1}^{T-1} W^{\boldsymbol{d}}_{1:t}[Y_t - \mathcal{Q}^{\boldsymbol{d}}_t(\boldsymbol{H}_t, A_t) + \mathcal{V}^{\boldsymbol{d}}_{t+1}(\boldsymbol{H}_{t+1})] + \mathcal{V}^{\boldsymbol{d}}_1(\boldsymbol{H}_1) - \beta$$

$$= \sum_{t=1}^{T} W^{\boldsymbol{d}}_{1:t}\Delta\mathcal{M}^{\boldsymbol{d}}_{t+1}(\boldsymbol{H}_{t+1}) + \mathcal{V}^{\boldsymbol{d}}_1(\boldsymbol{H}_1) - \beta,$$

where (1.8) follows from Robins (2000, Theorem 4.2), and (1.9) follows from

$$\begin{array}{rcll} \mathbb{E}(W_{1:T}^{\boldsymbol{d}}|\boldsymbol{H}_t, A_t) & = & W_{1:t}^{\boldsymbol{d}}; & \\ \mathbb{E}(W_{1:T}^{\boldsymbol{d}} V_{1:T}|\boldsymbol{H}_t, A_t) & = & W_{1:t}^{\boldsymbol{d}} & \times \quad [V_{1:(t-1)} + \mathcal{Q}_t^{\boldsymbol{d}}(\boldsymbol{H}_t, A_t)]; \\ \mathbb{E}(W_{1:T}^{\boldsymbol{d}}|\boldsymbol{H}_t) & = & W_{1:(t-1)}^{\boldsymbol{d}}; & \\ \mathbb{E}(W_{1:T}^{\boldsymbol{d}} V_{1:T}|\boldsymbol{H}_t) & = & W_{1:(t-1)}^{\boldsymbol{d}} & \times \quad [V_{1:(t-1)} + \mathcal{V}_t^{\boldsymbol{d}}(\boldsymbol{H}_t)]. \end{array}.$$

□

Based on the AIPWE, Zhang et al. (2013) proposed to directly maximize $\widehat{\mathcal{V}}_{\text{AIPWE},n}(\boldsymbol{d}_\eta)$ over a parametric DTR class $\boldsymbol{\mathcal{D}}_\eta = \{\boldsymbol{d}_\eta : \eta \in \Xi\}$. However, the nuisance functions $\{\mathcal{Q}_t^{\boldsymbol{d}}(\boldsymbol{H}_t, A_t)\}_{t=1}^T$ also depend on $\boldsymbol{d}$, which can result in a challenging computational problem. Nie et al. (2021) considered a special class of the when-to-treat DTRs $\boldsymbol{\mathcal{D}}_{\text{when-to-treat}}$. The AIPWE maximization problem can be further simplified by alternatively estimating the value differences $\{\mathcal{V}(\boldsymbol{d}) - \mathcal{V}(\boldsymbol{d}') : \boldsymbol{d}, \boldsymbol{d}' \in \boldsymbol{\mathcal{D}}_{\text{when-to-treat}}\}$, which have special structures.

Jiang and Li (2016); Zhang and Zhang (2018) pointed out that $\widehat{\mathcal{V}}_{\text{AIPWE},n}(\boldsymbol{d}) = \mathbb{E}_n(V_1^{\boldsymbol{d}})$, where $V_1^{\boldsymbol{d}}$ can be computed from $V_{T+1}^{\boldsymbol{d}} = 0$ and $V_T^{\boldsymbol{d}}, \cdots, V_1^{\boldsymbol{d}}$ in the following backward stagewise manner according to (1.7):

$$V_t^{\boldsymbol{d}} = \mathcal{V}_t^{\boldsymbol{d}}(\boldsymbol{H}_t) + W_t^{d_t}\left[Y_t + V_{t+1}^{\boldsymbol{d}} - \mathcal{Q}_t^{\boldsymbol{d}}(\boldsymbol{H}_t, A_t)\right]; \quad t = T, \cdots, 1.$$

At stage t, if $\boldsymbol{d} = (\boldsymbol{d}_{1:t}, \boldsymbol{d}_{(t+1):T})$ is replaced by $\boldsymbol{d} = (\boldsymbol{d}_{1:t}, \boldsymbol{d}^\star_{(t+1):T})$, then we further have

$$\begin{array}{rcll} V_t^{d_t, \boldsymbol{d}^\star_{(t+1):T}} & = & \mathcal{Q}_t(\boldsymbol{H}_t, d_t) + W_t^{d_t}\left[Y_t + V_{t+1}^{\boldsymbol{d}^\star} - \mathcal{Q}_t(\boldsymbol{H}_t, A_t)\right]; & \\ d_t^\star & \in & \underset{d_t \in \mathcal{D}_t}{\operatorname{argmax}}\, \mathbb{E}\left\{V_t^{d_t, \boldsymbol{d}^\star_{(t+1):T}}\right\}; & t = T, \cdots, 1. \end{array}$$

This can lead to a method of *stagewise AIPWE maximization*: $\widehat{V}_{T+1}^{(\text{a})} = 0$,

$$\begin{array}{rcll} \widehat{d}_{t,n} & \in & \underset{d_t \in \mathcal{D}_t}{\operatorname{argmax}}\, \mathbb{E}_n\left\{\mathcal{Q}_t(\boldsymbol{H}_t, d_t) + W_t^{d_t}\left[Y_t + \widehat{V}_{t+1}^{(\text{a})} - \mathcal{Q}_t(\boldsymbol{H}_t, A_t)\right]\right\}; & \\ \widehat{V}_t^{(\text{a})} & = & \mathcal{Q}_t(\boldsymbol{H}_t, \widehat{d}_{t,n}) + W_t^{\widehat{d}_{t,n}}\left(Y_t + \widehat{V}_{t+1}^{(\text{a})} - \mathcal{Q}_t(\boldsymbol{H}_t, A_t)\right); & t = T, \cdots, 1. \end{array} \quad (1.10)$$

In particular, the nuisance functions $\{\mathcal{Q}_t(\boldsymbol{H}_t, A_t)\}_{t=1}^T$ can be estimated from Q-Learning.

Based on the stagewise computation of the AIPWE, Liu et al. (2018) proposed the *Augmented OwL (AOL)* framework that can be equivalent to (1.10). Specifically, the stage-t AOL problem is

$$\begin{aligned}
\widehat{f}_{t,n} &\in \underset{f_t \in \mathcal{F}_t}{\operatorname{argmax}} \left\{ \mathbb{E}_n \left(\frac{\left| Y_t + \widehat{V}_{t+1}^{(\mathrm{a})} - g_t(\boldsymbol{H}_t) \right|}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \phi_t \left\{ (2A_t - 1)\mathsf{sign}\left[Y_t + \widehat{V}_{t+1}^{(\mathrm{a})} - g_t(\boldsymbol{H}_t) \right] f_t(\boldsymbol{H}_t) \right\} \right) + \lambda_t \|f_t\|_{\mathcal{F}_t}^2 \right\}; \\
\widehat{V}_t^{(\mathrm{a})} &= \mathcal{Q}_t \left(\boldsymbol{H}_t, \mathbb{1}[\widehat{f}_{t,n}(\boldsymbol{H}_t) \geqslant 0] \right) + \frac{\mathbb{1}[(2A_t - 1)\widehat{f}_{t,n}(\boldsymbol{H}_t) \geqslant 0]}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \left(Y_t + \widehat{V}_{t+1}^{(\mathrm{a})} - \mathcal{Q}_t(\boldsymbol{H}_t, A_t) \right).
\end{aligned}$$

Here, $g_t : \mathcal{H}_t \to \mathbb{R}$ can be an arbitrary function for efficiency augmentation, and $g_t(\boldsymbol{H}_t) = p_{\mathscr{A},t}(\boldsymbol{H}_t, 0)\mathcal{Q}_t(\boldsymbol{H}_t, 1) + p_{\mathscr{A},t}(\boldsymbol{H}_t, 1)\mathcal{Q}_t(\boldsymbol{H}_t, 0)$ corresponds to the case that $\widehat{f}_{t,n}$ maximizes the AIPWE (Zhou and Kosorok, 2017). Instead of maximizing the stagewise value function estimates, Zhang and Zhang (2018) proposed the T-stage *C-Learning*: $\widehat{V}_{t+1}^{(\mathrm{g})} = 0$,

$$\begin{aligned}
\widehat{C}_t^{(\mathrm{a})} &= \mathcal{Q}_t(\boldsymbol{H}_t, 1) - \mathcal{Q}_t(\boldsymbol{H}_t, 0) + \frac{2A_t - 1}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \left[Y_t + \widehat{V}_{t+1}^{(\mathrm{g})} - \mathcal{Q}_t(\boldsymbol{H}_t, A_t) \right]; \\
\widehat{d}_{t,n} &\in \underset{d_t \in \mathcal{D}_t}{\operatorname{argmin}} \mathbb{E}_n \left\{ \left| \widehat{C}_t^{(\mathrm{a})} \right| \mathbb{1}\left[d_t(\boldsymbol{H}_t) \neq \mathbb{1}\left(\widehat{C}_t^{(\mathrm{a})} \geqslant 0 \right) \right] \right\}; \qquad t = T, \cdots, 1. \\
\widehat{V}_t^{(\mathrm{g})} &= Y_t + \widehat{V}_{t+1}^{(\mathrm{g})} - \left\{ A_t - \mathbb{1}\left[\widehat{d}_{t,n}(\boldsymbol{H}_t) \geqslant 0 \right] \right\} \left[\mathcal{Q}_t(\boldsymbol{H}_t, 1) - \mathcal{Q}_t(\boldsymbol{H}_t, 0) \right];
\end{aligned}$$

Here, analogous to the single-stage C-Learning, $\widehat{d}_{t,n}$ is equivalent to maximizing a single-stage AIPWE. Instead of the a-outcomes $\{\widehat{V}_t^{(\mathrm{a})}\}_{t=1}^T$ in (1.10), the g-outcomes $\{\widehat{V}_t^{(\mathrm{g})}\}_{t=1}^T$ from the SNMM are used instead. For implementation, C-Learning minimizes the CATE-weighted misclassification rate over the class of decision trees at each stage. Extending from the binary treatment case, the general T-stage $K(\geqslant 2)$-treatment setting was further studied by Tao and Wang (2017); Tao et al. (2018).

When considering the statistical inference of the estimated value at the optimal DTR $\widehat{\mathcal{V}}_{\mathrm{AIPWE},n}(\widehat{\boldsymbol{d}}_{1:T,n})$, the nonregularity problem can occur. The construction of valid confidence intervals was studied in van der Laan and Luedtke (2014, 2015); Luedtke and van der Laan (2016); Shi et al. (2020a).

1.3 New Contributions and Outline

In Sections 1.1 and 1.2, we have introduced the main frameworks for the individualized decision making problem in the literature and discussed their advantages and disadvantages. However, there are still a few open problems to be addressed. First of all, existing methods rely on the assumption that the training and testing distributions are identical, while much less work has been

done on the problem when potential distributional changes exist. Secondly, while double robustness can guarantee the estimation consistency in presence of at most one model misspecification, the consequence towards efficiency remains unclear. In particular, when one model misspecification exists, we are able to show that a doubly robust estimate can suffer from downgraded efficiency. Other than potential misspecified nuisance models, most existing methods do not account for the heteroscedastic noise, which can greatly affect the estimation efficiency as well. Thirdly, for the $T(\geqslant 2)$-stage $K(\geqslant 3)$-treatment decision problem, there exists gaps between the theory and practice for semiparametric efficient methods. In particular, the rigorous semiparametric efficient estimation procedure is rarely used in practice. This dissertation mainly aims to address all these problems.

The remaining chapters are organized as follows.

- In Chapter 2, we consider the problem when training and testing distributions can be different. We make use of the development in the literature on Distributionally Robust Optimization (DRO) and propose a novel *Distributionally Robust ITR (DR-ITR)* framework that maximizes the worst-case value function across the values under a set of underlying distributions that are "close" to the training distribution. The DR-ITR can guarantee the performance among all such distributions reasonably well. We further propose the calibration procedures that tune the DR-ITR adaptively to a small amount of calibration data generated from a specific testing distribution. In this way, the calibrated DR-ITR enjoys better generalizability than the standard ITR in many different testing datasets. In our illustrating example, we show that the standard ITR can have very poor values on many testing distributions, while our calibrated DR-ITRs still enjoy relatively good performance. In particular, our proposed calibration procedures can pick reasonably good DR-constants based on the small calibrating sample. To solve the worst-case optimization problem, we make use of the Difference-of-Convex (DC) relaxation of the nonsmooth indicator, and propose two algorithms to solve the nonconvex problems of different scenarios. We also provide the finite sample approximation guarantees for the proposed DR-ITR. Finally, we apply our proposed DR-ITR to the AIDS clinical dataset ACTG 175 and evaluate its generalizability on the women patient subgroup. The manuscript of this chapter is accepted by *Journal of the American Statistical Association* with discussion and our rejoinder (Mo et al., 2021a,b).

- In Chapter 3, we consider the problem of potential treatment-free effect misspecification and heteroscedasticity. We demonstrate that the consequences of misspecified treatment-free effect and heteroscedasticity can be unified as a covariate-treatment dependent variance of residuals. To improve efficiency of the estimated ITR, we propose an *Efficient Learning (E-Learning)* framework for finding an optimal ITR in the multi-treatment setting. We show that the proposed E-Learning is optimal among a regular class of semiparametric estimates that can allow treatment-free misspecification and heteroscedasticity. In our simulation study, E-Learning demonstrates its effectiveness if one of or both misspecified treatment-free effect and heteroscedasticity exist. Our analysis of a *Type 2 Diabetes Mellitus (T2DM)* observational study also suggests the improved efficiency of E-Learning.

- In Chapter 4, we consider the multi-stage multi-treatment decision problem. We first introduce a novel *Backward Change Point SNMM (BCP-SNMM)*, where there exists an unknown backward change point, such that the data generating process is completely nonparametric before the change point, and then follows the SNMM starting from the change point to the end. The BCP-SNMM can allow more robustness against model misspecifications. Any violations of the SNMMs at previous stages do not affect the estimation properties at the current stage, including consistency and semiparametric efficiency. Based on the BCP-SNMM, we further propose the *Dynamic Efficient Learning (DE-Learning)* that solves the semiparametric efficient estimating equations under the multiple treatment setting. DE-Learning is optimal under the BCP-SNMM even in presence of heteroscedasticity and treatment-free effect misspecifications. It enjoys stagewise double robustness in addition to the robustness with respect to backward model misspecifications. Moreover, DE-Learning is a tractable procedure for rigorous semiparametric efficient estimation, with much fewer nuisance functions than G-Estimation and can be implemented in a backward stagewise manner. The superiority of DE-Learning is demonstrated in our simulation studies with stagewise misspecified treatment-free effects and heteroscedasticity.

CHAPTER 2

Learning Optimal Distributionally Robust Individualized Treatment Rules

2.1 Introduction

Consider the single-stage problem to estimate an optimal ITR. When the training and testing distributions are different, an estimated optimal ITR may not generalize well on the testing data (Zhao et al., 2019b). Similar phenomenon for causal inference in randomized controlled trials (RCTs) has also been pointed out by Muller (2014); Gatsonis and Morton (2017). Specifically, due to the inclusion and exclusion criteria of an RCT, the training sample can be unrepresentative of the testing population we are interested in. Therefore, the corresponding casual evidence may not be broadly applicable or relevant for the real-world practice. In causal inference literature, it is common to regard the training data as a selected sample from the pooled population of training and testing. The selection bias can be adjusted by reweighing or stratifying the training data according to the relationship between training and testing (O'Muircheartaigh and Hedges, 2014; Buchanan et al., 2018). However, it requires strong assumptions on completely measuring the selection confounders and correctly specifying the selection model, and thus can only work well on a prespecified testing population. There are many other practical scenarios where the difference between the training and testing distributions is unknown. One example is that the training data can be confounded by some unidentified effects such as batch effects, which may cause potential covariate shifts (Luo et al., 2010). Another possibility is that the testing distribution may evolve over time (Hand, 2006). There is also a widely studied scenario that multiple datasets are aggregated to perform combined analysis (Alyass et al., 2015; Shi et al., 2018b; Li et al., 2020). Aggregating data from various sources can benefit from sharing common information, transferring knowledge from different but related samples, and maintaining certain privacy. However, due to the heterogeneity among data sources, standard approaches of finding pooled optimal ITRs may not generalize well on all these sources. One way of handling the heterogeneity is to formulate it as a problem of distributional changes,

where we train on the mixture of subpopulations while testing on one of the subpopulations (Duchi et al., 2019). In all these applications, an optimal ITR that is robust to unattended distributional differences is of great interest.

Despite a vast literature in ITR, much less work has been done on the problem when the training and testing distributions are different. Imai and Ratkovic (2013) and Johansson et al. (2018) estimated the CTE function by reweighing the training loss to ensure the estimators generalizable on a prespecified testing distribution. Zhao et al. (2019b) aimed to find an ITR that optimizes the worst-case quality assessment among all testing covariate distributions satisfying some moment conditions. However, since their method only requires some moment conditions, the uncertainty set of the testing distributions can be very large. Recent developments in the distributionally robust optimization (DRO) literature provide the opportunities to quantify the difference between the training and testing distributions more precisely (Ben-Tal et al., 2013; Duchi and Namkoong, 2018; Rahimian and Mehrotra, 2019). Motivated by the DRO literature, we develop a new robust optimal ITR framework in this chapter.

In this chapter, we consider the problem of finding an optimal ITR from a restricted ITR class, where there is some unknown covariate changes between the training and testing distributions. We propose to use the *distributionally robust ITR (DR-ITR)* that maximizes the defined worst-case value function among value functions under a set of underlying distributions. More specifically, value functions are evaluated under all testing covariate distributions that are "close" to the training distribution, and the worst-case situation takes a minimal one. Our distributionally robust ITR framework is different from the existing doubly robust ITR framwork that uses an AIPWE. In particular, an AIPWE robustifies the model specification assumptions, while our DR-ITR robustifes the underlying distributions. The DR-ITR aims to guarantee reasonable performance across all testing distributions in an uncertainty set around the training distribution by optimizing the worst-case scenarios. In particular, we parameterize the amount of "closeness" by the *distributional robustness-constant (DR-constant)*, where the smallest possible DR-constant corresponds to the *standard ITR* that maximizes the value function under the training distribution. To ensure the performance of the DR-ITR on a specific testing distribution, we fit a class of DR-ITRs for a spectrum of DR-constants at the training stage, and calibrate the DR-constant based on a small amount of the calibrating data from the testing distribution. In this way, the correctly calibrated

DR-constant ensures that the DR-ITR performs at least as well as, often much better than, the standard ITR. Using our illustrative example, we show that the standard ITR can have very poor values on many testing distributions, while our calibrated DR-ITRs still maintain relatively good performance. In particular, our proposed calibrating procedures can tune DR-constants based on the small calibrating sample. To solve the worst-case optimization problem, we make use of the difference-of-convex (DC) relaxation of the nonsmooth indicator, and propose two algorithms to solve the related nonconvex optimization problems. We also provide the finite sample regret bound for the proposed DR-ITR.

The rest of this chapter is organized as follows. In Section 2.2, we discuss an illustrative example that the optimality of an ITR can be sensitive to the underlying distribution, and introduce the DR-ITR that can generalize well across all testing distributions considered in this example. Then we propose the DR-ITR framework and the corresponding learning problem. In Section 2.3, we justify the theoretical guarantees of the finite sample approximations for the learning problem. In Section 2.4, we evaluate the generalizability of our proposed DR-ITR on two simulation studies: the problem of covariate shifts and the problem of mixture of multiple subgroups. We apply our proposed DR-ITR on the AIDS clinical dataset ACTG 175 and evaluate its generalizability on the subgroup of female patients in Section 2.5. Some related discussions and extensions are given in Section 2.6. The implementation details, technical proofs and some additional numerical results are all given in Section 2.7.

2.2 Methodology

In this section, we introduce the value maximization framework in the current literature, and discuss its limitation when the training and testing distributions are different. Then we propose the DR-value function that optimizes the worse-case value function across all distributions within an uncertainty set around the training distribution.

2.2.1 Maximizing the Value Function

Consider the training data $(\boldsymbol{X}, A, Y) \sim \mathbb{P}$, where $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ denotes the covariates, $A \in \mathcal{A} = \{+1, -1\}$ is the binary treatment assignment, and $Y \in \mathcal{Y} \subseteq \mathbb{R}$ is the observed outcome. We

assume that the larger outcome is better. Let $Y(+1), Y(-1)$ be the potential outcomes. Consider a prespecified ITR class $\mathcal{D} \subseteq \{\pm 1\}^{\mathcal{X}}$. For $d \in \mathcal{D}$, denote $Y(d) := Y(1)\mathbb{1}[d(\boldsymbol{X}) = 1] + Y(-1)\mathbb{1}[d(\boldsymbol{X}) = -1]$ as the potential outcome following the treatment assignment prescribed by the ITR d. Then the value function under the training distribution $\mathbb{P}$ is defined as

$$\mathcal{V}(d) := \mathbb{E}[Y(d)].$$

Denote $\pi(a|\boldsymbol{x}) := \mathbb{P}(A = a|\boldsymbol{X} = \boldsymbol{x})$ as the training propensity score function for treatment assignment. If we assume 1) the *consistency* of the observed outcome $Y = Y(A)$; 2) the *strict overlap* $\pi(\pm 1|\boldsymbol{x}) \geqslant \tau > 0$ for any $\boldsymbol{x} \in \mathcal{X}$; and 3) the *strong ignorability* $(Y(+1), Y(-1)) \perp\!\!\!\perp A|\boldsymbol{X}$ (Rubin, 1974), then we can identify $\mathcal{V}(d)$ in terms of the observed data $(\boldsymbol{X}, A, Y)$ by the IPWE of $\mathbb{E}\left(\frac{\mathbb{1}[d(\boldsymbol{X})=A]}{\pi(A|\boldsymbol{X})}Y\right)$.

Instead of targeting the value function directly, we instead consider the CTE function as $C(\boldsymbol{x}) := \mathbb{E}[Y(+1) - Y(-1)|\boldsymbol{X} = \boldsymbol{x}]$ under the training distribution $\mathbb{P}$. Note that for an ITR d and all $\boldsymbol{x} \in \mathcal{X}$, the prescribed treatment assignment satisfies $d(\boldsymbol{x}) \in \{\pm 1\}$. Then we have $C(\boldsymbol{x})d(\boldsymbol{x}) = \mathbb{E}[Y(d) - Y(-d)|\boldsymbol{X} = \boldsymbol{x}]$. Based on this representation, we define another value function

$$\mathcal{V}_1(d) := \mathbb{E}[C(\boldsymbol{X})d(\boldsymbol{X})] = \mathbb{E}[Y(d) - Y(-d)]. \tag{2.1}$$

Since $Y(d) + Y(-d) \equiv Y(1) + Y(-1)$, it can be observed that $\mathcal{V}_1(d) = 2\left[\mathcal{V}(d) - \frac{\mathbb{E}[Y(+1)+Y(-1)]}{2}\right] = 2[\mathcal{V}(d) - \mathcal{V}(d_{\text{rand}})]$, where $d_{\text{rand}}(\boldsymbol{x}) = +1$ with probability $1/2$ and -1 with probability $1/2$. Therefore, $\mathcal{V}_1(d)$ can be interpreted as the value improvement of the ITR d upon the completely random treatment rule d_{rand}. In terms of the optimal ITR, the resulting rules by optimizing the value functions $\mathcal{V}_1(d)$ and $\mathcal{V}(d)$ over d are equivalent.

By the definition (2.1), we have $\mathcal{V}_1(d) \leqslant \mathbb{E}[|C(\boldsymbol{X})|]$ with equality if $d(\boldsymbol{X}) = \mathbf{sign}[C(\boldsymbol{X})]$ almost surely. Such an ITR is the global optimal ITR when $\mathcal{D}$ consists of all measurable functions from $\mathcal{X}$ to $\{\pm 1\}$. To obtain the global optimal ITR, we can estimate $C(\boldsymbol{X})$ from data using flexible nonparametric techniques, such as the Bayesian additive regression tree (BART) (Hill, 2011), or the casual forest (Wager and Athey, 2018). However, in general, the global optimal ITR

$\boldsymbol{x} \mapsto \mathbf{sign}[C(\boldsymbol{x})]$ can take a very complicated functional form, while decision makers may want to have a simpler ITR (Kitagawa and Tetenov, 2018). Then the ITR class $\mathcal{D}$ is often considered as a restricted subset of measurable functions from $\mathcal{X}$ to $\{\pm 1\}$. The following two-step procedure can be implemented to estimate the restricted optimal ITR on $\mathcal{D}$: first we estimate the CTE function $\boldsymbol{x} \mapsto \widehat{C}(\boldsymbol{x})$ using flexible nonparametric techniques; and then we estimate the ITR by solving $\max_{d\in\mathcal{D}} \mathbb{E}_n[\widehat{C}(\boldsymbol{X})d(\boldsymbol{X})]$ on the restricted ITR class $\mathcal{D}$ (Zhang et al., 2012a). Here, $\mathbb{E}_n$ is the empirical average based on the training data.

2.2.2 Covariate Changes

It can be observed that the value functions defined in Section 2.2.1 depend on the underlying distribution. Suppose we are interested in a testing distribution $\mathbb{P}_{\text{test}}$ that may be different from the training distribution $\mathbb{P}$ to some extent. Then ITRs estimated by most existing methods may not be able to perform well on our target population. In order to address this problem, we first make the following assumption on the potential difference between $\mathbb{P}_{\text{test}}$ and $\mathbb{P}$.

Assumption 2.1 (Covariate Changes)**.** For every training distribution $\mathbb{P}$ and testing distribution $\mathbb{P}_{\text{test}}$ considered in this chapter, we assume the followings:

(I) $\mathbb{P}_{\text{test}} \ll \mathbb{P}$;

(II) There exists $w : \mathcal{X} \to \mathbb{R}_+$ such that $\mathbb{E}_{\mathbb{P}} w(\boldsymbol{X}) = 1$, and $\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P} = w(\boldsymbol{X})$.

Assumption 2.1 (I) requires that the support of the testing distribution cannot go beyond the training distribution. Assumption 2.1 (II) is mathematically equivalent to assuming that the differences between $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$ only appear in the covariate distributions. The treatment-response relationship conditional on covariates remains unchanged across training and testing distributions. Specifically, let $p_{\boldsymbol{X}}(\boldsymbol{x})p_{Y|\boldsymbol{X}}(y(1), y(-1)|\boldsymbol{x})$ and $q_{\boldsymbol{X}}(\boldsymbol{x})q_{Y|\boldsymbol{X}}(y(1), y(-1)|\boldsymbol{x})$ be the training and testing densities of the data $(\boldsymbol{X}, Y(1), Y(-1))$. Then the density ratio $\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P}$ becomes

$$\frac{\mathrm{d}\mathbb{P}_{\text{test}}}{\mathrm{d}\mathbb{P}} = \frac{q_{\boldsymbol{X}}(\boldsymbol{X})}{p_{\boldsymbol{X}}(\boldsymbol{X})} \times \frac{q_{Y|\boldsymbol{X}}(Y(1), Y(-1)|\boldsymbol{X})}{p_{Y|\boldsymbol{X}}(Y(1), Y(-1)|\boldsymbol{X})}.$$

If $q_{Y|\boldsymbol{X}}(Y(1), Y(-1)|\boldsymbol{X}) = p_{Y|\boldsymbol{X}}(Y(1), Y(-1)|\boldsymbol{X})$, *i.e.*, the conditional distributions $(Y(1), Y(-1))|\boldsymbol{X}$ are identical under $\mathbb{P}_{\text{test}}$ and $\mathbb{P}$, then $\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P} = q_{\boldsymbol{X}}(\boldsymbol{X})/p_{\boldsymbol{X}}(\boldsymbol{X})$, which is the weighting function $w(\boldsymbol{X})$ in Assumption 2.1 (II).

The assumption of covariate changes is commonly seen in the setting of randomized trial. Consider the training and testing populations together as a pooled population with finite subjects. For each subject $i \in \{1, 2, \cdots, N\}$, let $S_i \in \{0, 1\}$ be a selection random variable such that $S_i = 1$ if i is a training sample point, and $S_i = 0$ if i is a testing sample point. Let the distributions of $(\boldsymbol{X}_i, Y_i(1), Y_i(-1))|(S_i = 1)$ and $(\boldsymbol{X}_i, Y_i(1), Y_i(-1))|(S_i = 0)$ be the training distribution $\mathbb{P}$ and the testing distribution $\mathbb{P}_{\text{test}}$ respectively. Denote $\bar{\mathbb{P}}$ as the joint distribution of $(\boldsymbol{X}_i, Y_i(1), Y_i(-1), S_i)$. Then conditions in Assumption 2.1 can correspond to the following (Hotz et al., 2005; Stuart et al., 2011):

- (Overlapping Support) $0 < \bar{\mathbb{P}}(S_i = 1|\boldsymbol{X}_i) < 1$;
- (Selection Unconfoundedness) $S_i \perp\!\!\!\perp (Y_i(1), Y_i(-1))|\boldsymbol{X}_i$.

In particular, under this finite population setting, the overlapping support condition is equivalent to that $\mathbb{P}_{\text{test}} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{P}_{\text{test}}$, and the selection unconfoundedness condition is equivalent to Assumption 2.1 (II). Such a correspondence can bring more intuitive implications of Assumption 2.1 under the randomized trial setting. Specifically, the overlapping support requires the chances of each subject being selected into the training and testing populations to be both positive. The selection unconfoundnedness requires that the selection mechanism is independent of the potential outcomes given the covariates. Both conditions can be satisfied by a successful trial design (Pearl and Bareinboim, 2014). The phenomenon of covariate changes between $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$ can exist if $\bar{\mathbb{P}}(S_i = 1|\boldsymbol{X}_i) \neq \bar{\mathbb{P}}(S_i = 0|\boldsymbol{X}_i)$ with a positive probability. This can be often the case if the subject needs to satisfy certain requirements before enrolling a trial.

As a consequence from Assumption 2.1, the CTE function $C(\boldsymbol{X}) = \mathbb{E}_{\mathbb{P}}[Y(1) - Y(-1)|\boldsymbol{X}] = \mathbb{E}_{\text{test}}[Y(1) - Y(-1)|\boldsymbol{X}]$ remains unchanged under $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$. Then it can be convenient to consider the value functions $\mathcal{V}_1(d) = \mathbb{E}_{\mathbb{P}}[C(\boldsymbol{X})d(\boldsymbol{X})]$ and $\mathcal{V}_{1,\text{test}}(d) = \mathbb{E}_{\text{test}}[C(\boldsymbol{X})d(\boldsymbol{X})]$ defined in (2.1). When the testing value function $\mathcal{V}_{1,\text{test}}(d)$ is of interest, maximizing the training value function $\mathcal{V}_1(d)$ may not be optimal. Alternatively, we can rewrite the testing value function $\mathcal{V}_{1,\text{test}}(d) = \mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})C(\boldsymbol{X})d(\boldsymbol{X})]$ where $w(\boldsymbol{X}) = \mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P}$. Then based on the training data from $\mathbb{P}$, we can

maximize $\mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})C(\boldsymbol{X})d(\boldsymbol{X})]$ that targets the correct objective. It amounts to determine the weighting function w that captures the differences between $\mathbb{P}_{\text{test}}$ and $\mathbb{P}$.

Remark 2.1. Notice that for any weighting function $w : \mathcal{X} \to \mathbb{R}_+$, we have $\mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})C(\boldsymbol{X})d(\boldsymbol{X})] \leqslant \mathbb{E}_{\mathbb{P}}[w(\boldsymbol{X})|C(\boldsymbol{X})|]$ with equality if $d(\boldsymbol{X}) = \mathbf{sign}[C(\boldsymbol{X})]$. That is, if $\mathcal{D}$ consists of all measurable functions from $\mathcal{X}$ to $\{\pm 1\}$, then the global optimal ITR is *not* sensitive to any covariate changes in the testing distribution. However, the problem of covariate changes induces a challenge if $\mathcal{D}$ is a restricted ITR class.

Remark 2.2. Our methodology only relies on the fact that $C(\boldsymbol{X})$ remains unchanged under $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$. Therefore, it can be possible to relax Assumption 2.1 to allowing distributional changes in $(Y(1), Y(-1))|\boldsymbol{X}$, while assuming that the CTE function $C(\cdot)$ remains identical across $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$. Furthermore, our methodology can also be meaningful if the testing CTE function can be different from training, but the optimal treatment assignment remains unchanged. We will discuss this extension in Remark 2.5.

2.2.3 An Illustrative Example

In this section, we begin with an example as in Figure 2.1 that the optimality of an ITR depends on the underlying distribution. There are two underlying bivariate normal distributions of means $(0, 0)^\intercal$ (training) and $(1.47, 1.69)^\intercal$ (testing) respectively. We obtain the standard ITR by maximizing the value function $\mathcal{V}_1(d)$ under the training distribution over the linear ITR class. We also obtain the DR-ITR by maximizing the DR-value function $\mathcal{V}_c^k(d)$ to be introduced in Section 2.2.4 over the linear ITR class. Then the DR-ITR is compared with the standard ITR through the value functions $\mathcal{V}_1$ under the training distribution and $\mathcal{V}_{1,\text{test}}$ under the testing distribution as in Table 2.1. Since the values can be comparable only through the same value function but not across different value functions, we further define the criteria *relative regret* of an ITR as $[\text{value(LB-ITR)} - \text{value(ITR)}]/|\text{value(LB-ITR)}|$, where "value" can be $\mathcal{V}_1$ or $\mathcal{V}_{1,\text{test}}$, and the LB-ITR maximizes the corresponding value function over the linear ITR class. In this sense, value(LB-ITR) is the best achievable value among the linear ITR class for the corresponding value function, and becomes the benchmark reference for the relative regret criteria.

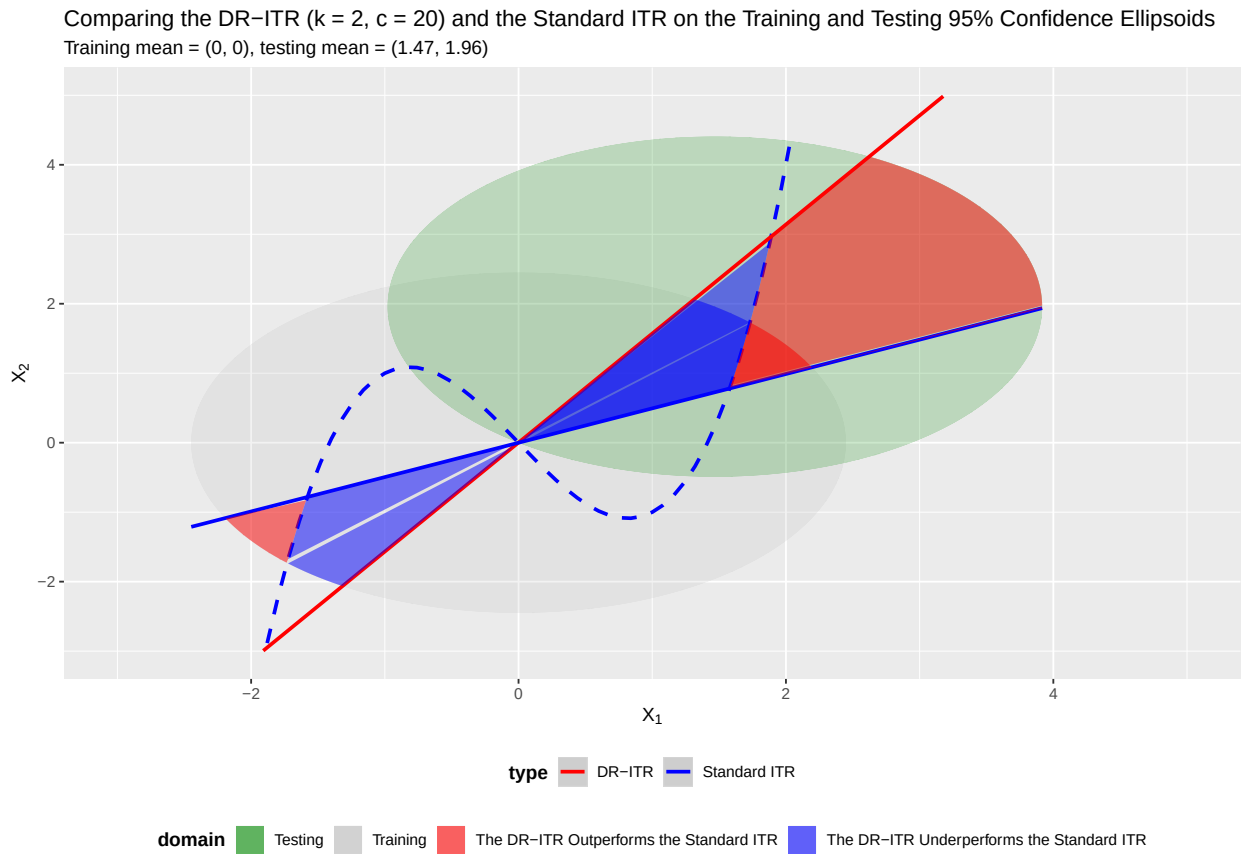


Figure 2.1: ITRs and the 95% confidence ellipsoids of the training distribution $(X_1, X_2) \sim \mathcal{N}_2\big((0,0)^\intercal, \mathbf{I}_2\big)$ and the testing distribution $(X_1, X_2) \sim \mathcal{N}_2\big((1.47, 1.96)^\intercal, \mathbf{I}_2\big)$. The blue dashed curve is the underlying CTE boundary $C(X_1, X_2) = X_2 - (X_1^3 - 2X_1) = 0$.

Table 2.1: Testing Values (Relative Regrets) Comparisons of ITRs

Value \ ITR	DR-ITR	Standard ITR	LB-ITR
Training $\mathcal{V}_1$	0.6253 (37.36%)	0.9982 (0%)	0.9982
Testing $\mathcal{V}_{1,\text{test}}$	4.8230 (9.16%)	0.2927 (94.49%)	5.3096

[1] DR-ITR maximizes $\mathcal{V}_c^k(d)$ defined in (2.4) with $k = 2$ and $c = 20$ over the linear ITR class.
[2] Standard ITR maximizes $\mathcal{V}_1(d)$ over the linear ITR class.
[3] LB-ITR maximizes $\mathcal{V}_1(d)$ or $\mathcal{V}_{1,\text{test}}(d)$ over the linear ITR class.
[4] Values (larger the better) can be comparable within rows but incomparable between rows.
[5] Relative regret(ITR) = [value(LB-ITR) − value(ITR)]/|value(LB-ITR)| (smaller the better).
[6] A size-10,000 sample is generated for fitting DR-ITR and LB-ITRs, and an independent size-100,000 sample is generated for evaluation under $\mathcal{V}_1$ and $\mathcal{V}_{1,\text{test}}$.

Two facts can be concluded from Table 2.1: 1) the optimality of an ITR can be different across different distributions; and 2) maximizing the training value function may have poor testing performance when covariate changes exist. In Table 2.1, even though the standard ITR is optimal

under the training distribution, it can be far from optimal (94.49% off in terms of relative regret) under the testing distribution. In contrast, the DR-ITR may not enjoy high training value, but can have much better testing performance (only 9.16% off in terms of relative regret).

Remark 2.3. Figure 2.1 also illustrates how the covariate changes affect the optimality of ITRs. Specifically, we can divide the covariate domain into two types of subdomains, annotated in blue and red, on which the DR-ITR and standard ITR have different treatment assignments. On the blue subdomain, the standard ITR assignment shares the same sign with the CTE function, while the DR-ITR does not. In this case, the standard ITR outperforms the DR-ITR with the difference of value $|C(\boldsymbol{X})|$ at the individual level. The case reverses on the red subdomain on which the DR-ITR outperforms the standard ITR. The overall difference of values integrates the individual difference with respect to the training or testing density.

The overall outperformance of the DR-ITR under the testing distribution can be explained from the following three perspectives: 1) the 95% confidence ellipsoid of the training domain only covers a small area of the red subdomain, while that of the testing domain covers a much larger area; 2) the distance of the red subdomain from the testing centroid is much closer than its distance from the training centroid. Then the red subdomain concentrates higher testing density than training; and 3) the individual value differences $|C(\boldsymbol{X})|$'s are generally larger on the red subdomain intersected with the testing domain than that intersected with the training domain. Therefore, the DR-ITR performs much better than the standard ITR on the testing distribution.

2.2.4 Maximizing the Distributionally Robust Value (DR-Value) Function

We begin to introduce our DR-ITR that can show strong generalizability as in Figure 2.1. As discussed in Section 2.1, our goal in this chapter is not to find an ITR that is generalizable on a specific testing distribution, but rather, to find an ITR that guarantees reasonable performance across an uncertain set of testing distributions. We first define the k-th *power uncertainty set* in two equivalent ways under Assumption 2.1:

$$\mathcal{P}_c^k(\mathbb{P}) := \left\{ \mathbb{Q} \ll \mathbb{P} \;\middle|\; \|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^k(\mathbb{P})} \leqslant c \right\} \tag{2.2}$$

$$= \left\{ \mathbb{Q} \ll \mathbb{P} \;\middle|\; w : \mathcal{X} \to \mathbb{R}_+,\ \mathbb{E}_{\mathbb{P}} w(\boldsymbol{X}) = 1,\ \mathbb{E}_{\mathbb{P}} w(\boldsymbol{X})^k \leqslant c^k,\ \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = w(\boldsymbol{X}) \right\}. \tag{2.3}$$

The set $\mathcal{P}_c^k(\mathbb{P})$ consists of the probability distributions $\mathbb{Q}$ such that the $L^k(\mathbb{P})$-norm of the density ratio $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ is bounded above by the DR-constant c. The definition (2.3) highlights that the density ratio is a weighting function w of $\boldsymbol{X}$, and the distribution $\mathbb{Q}$ in $\mathcal{P}_c^k(\mathbb{P})$ can be characterized by the weighting function w satisfying the conditions in (2.3). Here the DR-constant $c \geqslant 1$ controls the degree of the distributional robustness that measures how "close" $\mathbb{Q}$ is from $\mathbb{P}$. In particular, $c = 1$ reduces the power uncertainty set $\mathcal{P}_1^k(\mathbb{P})$ to the singleton $\{\mathbb{P}\}$. The power order $1 < k \leqslant +\infty$ parametrizes the measurement of the distance of $\mathbb{Q}$ from $\mathbb{P}$. In particular, the power uncertainty set $\mathcal{P}_c^k(\mathbb{P})$ increases in c as k is fixed, and decreases in k as c is fixed. The latter one is due to the Lyapunov's inequality: $\|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^k(\mathbb{P})} \leqslant \|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^{k'}(\mathbb{P})}$ whenever $1 < k \leqslant k' \leqslant +\infty$. In Section 2.7, we will discuss the explicit form of $\mathcal{P}_c^k(\mathbb{P})$ in the context of specific parametric families of distributions, and how it depends on the DR-constant c and the power k. One important conclusion from Example 2.2 in Section 2.7 for the mean-shifted p-dimensional normal distribution is that $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p) \in \mathcal{P}_c^k\big(\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)\big)$ if and only if $\|\boldsymbol{\mu}\|_2^2 \leqslant \frac{2\log c}{k-1}$.

With the power uncertainty set $\mathcal{P}_c^k(\mathbb{P})$, we propose to robustly maximize the following worst-case value function among the values under $\mathbb{Q} \in \mathcal{P}_c^k(\mathbb{P})$:

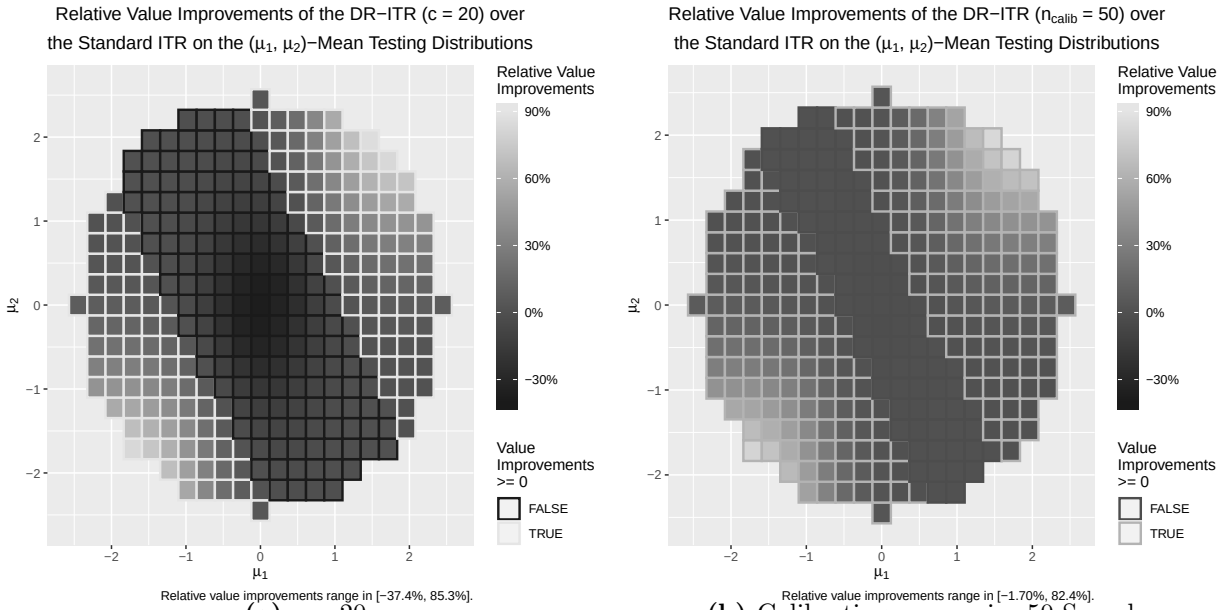
$$\mathcal{V}_c^k(d) := \inf_{\mathbb{Q} \in \mathcal{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}[C(\boldsymbol{X})d(\boldsymbol{X})], \tag{2.4}$$

which we term as the *DR-value function*. In particular, $c = 1$ reduces the DR-value function $\mathcal{V}_1^k(d)$ to the standard value function $\mathcal{V}_1(d) = \mathbb{E}_{\mathbb{P}}[C(\boldsymbol{X})d(\boldsymbol{X})]$ in the definition (2.1).

Remark 2.4 (Optimality)**.** The "optimality" of the DR-ITR is with respect to the DR-value function $\mathcal{V}_c^k$, which highlights its difference from the traditional "optimal" ITR with respect to the standard value function $\mathcal{V}_1$.

In the example in Section 2.2.3, the standard ITR maximizes the value function under the training distribution over the linear ITR class, while the DR-ITR maximizes the DR-value function $\mathcal{V}_c^k(d)$ of $k = 2$ and $c = 20$ over the linear ITR class. In particular, the randomness of $\mathbb{P}$ comes from the training covariate distribution $\mathcal{N}_2(\mathbf{0}_2, \mathbf{I}_2)$. Such a choice of $\mathcal{P}_c^k(\mathbb{P})$ contains the mean-shifted normal distributions $\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{I}_2)$ for all $\boldsymbol{\mu} \in \big\{(\mu_1, \mu_2)^\intercal : \mu_1^2 + \mu_2^2 \leqslant 4\log 5\big\}$. In Figure 2.2a, we enumerate such mean-shifted normal distributions as the testing distributions, and evaluate the *relative improvement* of the DR-ITR over the standard ITR as the difference of their relative regrets.

Among all testing distributions, the relative improvements of the DR-ITR span from -37.4% to 85.3%, suggesting that the potential of improvement can be large. Besides the DR-constant $c = 20$, we also consider the case $c = 2.71, 6.51, 10.31$ in Section 2.7. As c increases, the range of relative improvements becomes wider. The increase in the relative improvement upper bound is in general much larger than the decrease in the lower bound.



(a) $c = 20$ **(b)** Calibrating c on a size-50 Sample

Figure 2.2: Relative improvements of the DR-ITR over the standard ITR as the difference of relative regrets on testing distributions $\mathcal{N}_2(\boldsymbol{\mu}, \mathbf{I}_2)$ of $\boldsymbol{\mu} \in \{(\mu_1, \mu_2)^\intercal \in \mathbb{R}^2 : \mu_1^2 + \mu_2^2 \leqslant 4 \log 5\}$ (lighter the better).

Based on these observations, the DR-constant c should be carefully chosen. On one hand, as can be seen from Figure 2.2a, the DR-ITR for a fixed DR-constant c may or may not improve over the standard ITR on a specific testing distribution within $\mathcal{P}_c^k(\mathbb{P})$. When the DR-constant c can be tuned adaptive to the specific testing distribution, then the DR-ITR can perform at least as well as the standard ITR. On the other hand, we may not even have any prior information on c to ensure that the power uncertainty set $\mathcal{P}_c^k(\mathbb{P})$ contains the testing distribution of interest. Both cases ask for additional information to calibrate the choice of c so that the DR-ITR performs well on a specific testing distribution. Suppose we are able to obtain a small size of calibrating sample from the testing distribution. We propose the following training-calibrating procedure to choose c: 1) at the training stage, we estimate DR-ITRs $\{\widehat{d}_c\}_{c \in \mathcal{C}}$ where c is the DR-constant to compute $\widehat{d}_c$, and $\mathcal{C}$ is a set of candidate DR-constants; 2) we obtain a calibrating sample from the testing

distribution, on which we estimate the testing values of $\{\widehat{d}_c\}_{c\in\mathcal{C}}$; 3) we select the $\widehat{c}$ that maximizes the value of $\widehat{d}_c$ among $c \in \mathcal{C}$.

In order to estimate the value function under the testing distribution, we consider the following two possible calibration scenarios: 1) the calibrating sample is a randomized controlled trial (RCT) dataset $(\boldsymbol{X}, A, Y)$ from the testing distribution; and 2) the calibrating sample only consists of the covariates $\boldsymbol{X}$ from the testing distribution. Scenario 1 will be more ideal than Scenario 2 since we have the testing information of both the treatment and the outcome. We can evaluate an ITR d using the IPWE $\widehat{\mathcal{V}}^{\text{IPWE}}_{\text{calib}}(d) = \mathbb{E}_{n_{\text{calib}}}\{\mathbb{1}[d(\boldsymbol{X}) = A]Y/\pi_{\text{calib}}(A|\boldsymbol{X})\}$, where $\mathbb{E}_{n_{\text{calib}}}$ is the empirical average over the calibrating sample, π_{calib} is the corresponding propensity score function, and π_{calib} is known or estimable from the calibrating data. We call the corresponding calibrate DR-ITR as *RCT-DR-ITR*. In Scenario 2, we do not have the treatment-response information from the testing distribution. We can instead use the value function estimate $\widehat{\mathcal{V}}^{\text{CTE}}_{\text{calib}}(d) = \mathbb{E}_{n_{\text{calib}}}[\widehat{C}_n(\boldsymbol{X})d(\boldsymbol{X})]$ to evaluate d, where $\widehat{C}_n(\boldsymbol{X})$ is estimated at the training stage. However, the CTE estimate $\widehat{C}_n(\cdot)$ may also suffer from a potential generalizability problem on the testing distribution. Practitioners need to be careful of the generalizability of the CTE estimate when performing the calibration. We call the corresponding DR-ITR as *CTE-DR-ITR*.

RCT-DR-ITR and CTE-DR-ITR are different in their use of information for calibration. Specifically, the RCT-DR-ITR makes use of $(\boldsymbol{X}, A, Y)$ from the testing distribution, while the CTE-DR-ITR only makes use of $\boldsymbol{X}$ from the testing distribution, and the underlying CTE function $C(\boldsymbol{X})$. In practice, $C(\boldsymbol{X})$ is estimated from training data. It requires Assumption 2.1 to generalize the CTE estimate $\widehat{C}_n(\boldsymbol{X})$ from training to testing. If Assumption 2.1 holds, then CTE-DR-ITR can have better performance than RCT-DR-ITR, since CTE-DR-ITR captures less variance from calibrated data. If Assumption 2.1 is violated, which will be illustrated in Section 2.4.2, then CTE-DR-ITR can have poorer performance than RCT-DR-ITR, since the testing value function estimate of CTE-DR-ITR can be biased.

In Figure 2.2b, we generate a calibrating RCT sample from $\mathbb{P}_{\text{test}}$ of size 50. It shows that across the mean-shifted testing distributions, the relative improvements of the calibrated DR-ITRs range from -1.70% to 82.4%. It suggests that the small sample size 50 is sufficient for a reasonably good calibration, with the positive relative improvements being maintained.

Remark 2.5 (Extending Covariate Changes)**.** Consider the case that Assumption 2.1 is violated. Let C_{test} be the testing CTE function that can be different from the training CTE function C. We use the notations $\mathbb{P}$ and $\mathbb{P}_{\text{test}}$ to refer to the training and testing covariate distributions. Assume that $\mathbf{sign}[C_{\text{test}}(\boldsymbol{X})] = \mathbf{sign}[C(\boldsymbol{X})]$ almost surely. Then we can still represent the value function under the testing distribution as follows:

$$\mathbb{E}_{\text{test}}[C_{\text{test}}(\boldsymbol{X})d(\boldsymbol{X})] = \mathbb{E}_{\mathbb{P}}\left\{\frac{\mathrm{d}\mathbb{P}_{\text{test}}}{\mathrm{d}\mathbb{P}}\frac{C_{\text{test}}(\boldsymbol{X})}{C(\boldsymbol{X})}\mathbb{1}[C(\boldsymbol{X}) \neq 0] \times C(\boldsymbol{X})d(\boldsymbol{X})\right\}.$$

The definition of the DR-value function (2.4) can be robust with respect to the change of $(\mathbb{P}_{\text{test}}, C_{\text{test}})$ from $(\mathbb{P}, C)$, such that $w(\boldsymbol{X}) := (\mathrm{d}\mathbb{P}_{\text{test}}/\mathrm{d}\mathbb{P}) \times [C_{\text{test}}(\boldsymbol{X})/C(\boldsymbol{X})]\mathbb{1}[C(\boldsymbol{X}) \neq 0]$ satisfies $\mathbb{E}_{\mathbb{P}}w(\boldsymbol{X}) = 1$ and $\mathbb{E}_{\mathbb{P}}w(\boldsymbol{X})^k \leqslant c^k$.

Remark 2.6. The calibration procedure ensures that among the DR-ITRs of various DR-constants, the best one is chosen to maximize the testing value function. In this sense, the calibrated DR-ITR can have potential of improving the generalizability from training to testing. However, if the testing distribution is very far from the training distribution, one cannot expect that an ITR estimated by any method from the training data can perform well on the test data, even though our proposed method may be able to protect against such a distributional change to some extent. Therefore, in practice, we suggest to use our method when training and testing distributions are relatively close.

2.2.5 Distributionally Robust Expectation

In this section, we first discuss the rationale of considering the L^k-norm of the density ratio as the measurement of distributional distance. We show that the k-th power uncertainty set $\mathcal{P}_c^k(\mathbb{P})$ is equivalent to the distributional ball induced by the ϕ-divergence (Pardo, 2005) for some specific divergence ϕ. Then we derive the dual form of the worst-case expectation over $\mathcal{P}_c^k(\mathbb{P})$, which provides a more tractable optimization problem.

2.2.5.1 Equivalence to the Divergence-Based Distributional Ball

As a generalization of the conventional likelihood-based framework which corresponds to the Kullback-Leibler (KL) divergence, the framework of general ϕ-divergence between distributions has been well studied in the context of parameter estimation and hypothesis testing (Pardo, 2005).

The ϕ-divergence between two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ such that $\mathbb{Q} \ll \mathbb{P}$ is defined as follows:

$$D_\phi(\mathbb{Q}\|\mathbb{P}) := \int \phi\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right)\mathrm{d}\mathbb{P} = \mathbb{E}_{\mathbb{P}}\phi\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right); \quad \phi \in \Phi,$$

where Φ is a class of convex functions on $\mathbb{R}$ that satisfies the regularity conditions: $\phi(w) = +\infty$ for $w < 0$, $\phi(1) = \phi'(1) = 0$, and $\lim_{w\to 0_+} w\phi(p/w) = \lim_{w\to+\infty}\phi(w)/w$ for $p > 0$. The definition with various choices of ϕ's includes the empirical likelihood $\phi_{\mathrm{EL}}(w) = -\log w + w - 1$, the KL divergence $\phi_{\mathrm{KL}}(w) = w\log w - w + 1$, and the χ^2-divergence $\phi_{\chi^2}(w) = \frac{1}{2}(w-1)^2$. There is another important special case that relates to the power uncertainty set of $k = +\infty$. Consider the optimization indicator for $c \geqslant 1$: $\phi_{\infty,c} = 0$ if $u \in [0, c]$ and $+\infty$ otherwise, for which $D_{\phi_{\infty,c}}(\mathbb{Q}\|\mathbb{P}) = 0$ if $\|\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}\|_{L^\infty(\mathbb{P})} \leqslant c$, and $+\infty$ otherwise. Then $D_{\phi_{\infty,c}}(\mathbb{Q}\|\mathbb{P}) = 0$ if and only if $\mathbb{Q} \in \mathcal{P}_c^\infty(\mathbb{P})$.

Although D_ϕ is not a proper metric between probability distributions since it is asymmetric, we can still define a D_ϕ-distributional ball as $\mathscr{P}_\rho^\phi(\mathbb{P}) := \{\mathbb{Q} \ll \mathbb{P} : D_\phi(\mathbb{Q}\|\mathbb{P}) \leqslant \rho\}$, where $\mathbb{P}$ is the center and $\rho \geqslant 0$ is the radius. Then for any $\rho \geqslant 0$, the $D_{\phi_{\infty,c}}$-distributional ball $\mathscr{P}_\rho^{\phi_{\infty,c}}(\mathbb{P}) \equiv \{\mathbb{Q} \ll \mathbb{P} : D_{\phi_{\infty,c}}(\mathbb{Q}\|\mathbb{P}) = 0\}$, which coincides with the power uncertainty set $\mathcal{P}_c^\infty(\mathbb{P})$ defined in (2.2) for $k = \infty$. Such an equivalence can be extended to all finite $k \in (1, +\infty)$ when a Cressie-Read (CR) family (Cressie and Read, 1984) of divergence functions $\Phi_{\mathrm{CR}} \subseteq \Phi$ is taken into consideration. For $k > 1$, the corresponding $\phi_k \in \Phi_{\mathrm{CR}}$ is defined as

$$\phi_k(w) := \frac{w^k - kw + k - 1}{k(k-1)}; \quad w \geqslant 0.$$

Here, ϕ_k effectively measures the probability-distributional distance by the k-th moment of the density ratio, since $D_{\phi_k}(\mathbb{Q}\|\mathbb{P}) = \frac{1}{k(k-1)}[\mathbb{E}_{\mathbb{P}}(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P})^k - 1]$ as long as $\mathbb{Q}$ is a probability distribution. Then it can be inferred that the D_{ϕ_k}-distributional ball $\mathscr{P}_\rho^{\phi_k}(\mathbb{P})$ is actually equivalent to the power uncertainty set $\mathcal{P}_{c_k(\rho)}^k(\mathbb{P})$ in (2.2). Here, there is a one-to-one correspondence between the DR-constant c and the radius ρ of the D_{ϕ_k}-distributional ball with $c_k(\rho) := [k(k-1)\rho + 1]^{1/k}$. We conclude the case $k = +\infty$ and $1 < k < +\infty$ with the following:

$$\mathscr{P}_\rho^{\phi_{\infty,c}}(\mathbb{P}) = \mathcal{P}_c^\infty(\mathbb{P}); \quad \mathscr{P}_\rho^{\phi_k}(\mathbb{P}) = \mathcal{P}_{c_k(\rho)}^k(\mathbb{P}); \quad \rho \geqslant 0. \tag{2.5}$$

2.2.5.2 Dual Representation

We begin with a general result on the dual representation of the ϕ-divergence-based distributionally robust expectation. We state the following lemma and refer readers to Duchi and Namkoong (2018, Proposition 1).

Lemma 2.1. *Fix a random variable Z on $\mathbb{R}$ with distribution $\mathbb{P}$. Let $\phi \in \Phi$ be a legitimate divergence function. Define the convex conjugate of ϕ as*

$$\phi^\star(x^\star) := \sup_{x\in\mathbb{R}}\{\langle x^\star, x\rangle - \phi(x)\}; \quad x^\star \in \mathbb{R}.$$

Then for $\rho > 0$,

$$\sup_{\mathbb{Q}\in\mathscr{P}_\rho^\phi(\mathbb{P})} \mathbb{E}_\mathbb{Q} Z = \inf_{\substack{\lambda\geqslant 0\\ \eta\in\mathbb{R}}} \left\{ \mathbb{E}_\mathbb{P}\left[\lambda\phi^\star\left(\frac{Z-\eta}{\lambda}\right)\right] + \lambda\rho + \eta \right\}. \tag{2.6}$$

Let $c \geqslant 1$. Lemma 2.1 can be directly applied to the optimization indicator: $\phi_{\infty,c}(u) := 0$ if $u \in [0, c]$ and $+\infty$ otherwise, whose convex conjugate is given by $\phi^\star_{\infty,c}(u) = c\max\{u, 0\}$. Then λ in (2.6) attains the infimum at $\lambda = 0$, so that

$$\sup_{\mathbb{Q}\in\mathscr{P}_\rho^{\phi_{\infty,c}}(\mathbb{P})} \mathbb{E}_\mathbb{Q} Z = \inf_{\eta\in\mathbb{R}} \left\{ c\mathbb{E}_\mathbb{P}(Z-\eta)_+ + \eta \right\}. \tag{2.7}$$

In particular, the right hand side of (2.7) is solved by the $(1-1/c)$-*value-at-risk* $\mathsf{VaR}_{1-1/c}$ in finance, or equivalently, the $(1-1/c)$-quantile of Z under the center distribution $\mathbb{P}$. The right hand side of (2.7) itself is defined as the $(1-1/c)$-*conditional value-at-risk* $\mathsf{CVaR}_{1-1/c}$ (Rockafellar and Uryasev, 2000). Next, we apply Lemma 2.1 to the k-th power divergence ϕ_k to derive the dual problem of the worst-case expectation over $\mathcal{P}_c^k(\mathbb{P})$.

Lemma 2.2. *Let Φ_{CR} be the Cressie-Read family of divergence functions, $k, k^\star \in (1, +\infty)$ be conjugate numbers, i.e., $\frac{1}{k} + \frac{1}{k^\star} = 1$, and $\phi_k \in \Phi_{\mathrm{CR}}$. Then we have following conclusions:*

(I) The convex conjugate of ϕ_k is given by

$$\phi_k^\star(z) = \frac{1}{k}\left\{[(k-1)z+1]_+^{k^\star} - 1\right\}.$$

(II) Fix a probability measure $\mathbb{P}$ and a random variable Z on $\mathbb{R}$. Then for $\rho \geqslant 0$,

$$\sup_{\mathbb{Q}\in\mathscr{P}_\rho^{\phi_k}(\mathbb{P})} \mathbb{E}_{\mathbb{Q}} Z = \inf_{\eta\in\mathbb{R}} \left\{ c_k(\rho)[\mathbb{E}_{\mathbb{P}}(Z-\eta)_+^{k^\star}]^{1/k^\star} + \eta \right\}, \tag{2.8}$$

where $c_k(\rho) = [k(k-1)\rho + 1]^{1/k}$.

Note that the right hand side of (2.8) and its optimizer η are both coherent risk measures as the higher-order generalizations of the CVaR and VaR (Krokhmal, 2007).

Using the equivalence in (2.5), the worst-case expectation over the power uncertainty set $\mathcal{P}_c^k(\mathbb{P})$ for $k \in (1, \infty]$ and $k^\star = \frac{k}{k-1}$ (in particular, $k = \infty \Leftrightarrow k^\star = 1$) unifies (2.7) and (2.8) as follows:

$$\sup_{\mathbb{Q}\in\mathcal{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}} Z = \inf_{\eta\in\mathbb{R}} \left\{ c[\mathbb{E}_{\mathbb{P}}(Z-\eta)_+^{k^\star}]^{1/k^\star} + \eta \right\}; \quad c \geqslant 1. \tag{2.9}$$

By inspecting the dual problem (2.9), the right hand side is computationally more tractable than the left hand side, since instead of optimizing over an infinite-dimensional probability measure $\mathbb{Q}$, we only need to optimize over a univariate variable η.

In order to apply the duality result to the DR-ITR problem, we negate the DR-value maximization to a risk minimization problem. Denote the *risk function* under the training distribution $\mathbb{P}$ as $\mathcal{R}_1(d) := -\mathcal{V}_1(d) = \mathbb{E}_{\mathbb{P}}\{C(\boldsymbol{X})[-d(\boldsymbol{X})]\}$. Then for $k \in (1, +\infty]$ and $c \geqslant 1$, the *DR-risk function* is defined as

$$\mathcal{R}_c^k(d) := \sup_{\mathbb{Q}\in\mathcal{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}\{C(\boldsymbol{X})[-d(\boldsymbol{X})]\}.$$

Using the fact $Z = -C(\boldsymbol{X})d(\boldsymbol{X}) = C(\boldsymbol{X})\mathbb{1}[d(\boldsymbol{X}) = -1] + [-C(\boldsymbol{X})]\mathbb{1}[d(\boldsymbol{X}) = 1]$, the dual representation (2.9) can be expressed in the following particular form (2.10).

Corollary 2.3 (Dual Representation of the DR-Risk Function)**.** *Let $k \in (1, +\infty]$, $k^\star = \frac{k}{k-1}$ if $k < +\infty$ and $k^\star = 1$ if $k = +\infty$, $c \geqslant 1$. Then the DR-risk function $\mathcal{R}_c^k$ has the following dual representation:*

$$\mathcal{R}_c^k(d) = \inf_{\eta\in\mathbb{R}} \left\{ c\left[\mathbb{E}\left([C(\boldsymbol{X})-\eta]_+^{k^\star}\mathbb{1}[d(\boldsymbol{X}) = -1] + [-C(\boldsymbol{X})-\eta]_+^{k^\star}\mathbb{1}[d(\boldsymbol{X}) = 1]\right)\right]^{1/k^\star} + \eta \right\}. \tag{2.10}$$

2.2.6 Implementation

In this section, we introduce the implementation of DR-risk minimization based on the empirical data. We cast the learning problem as finding a decision function $f : \mathcal{X} \to \mathbb{R}$ that induces an ITR based on its sign: $d(\boldsymbol{x}) = \mathbf{sign}[f(\boldsymbol{x})]$. The ITR class $\mathcal{D}$ can correspond to a prespecified decision function class $\mathcal{F}$. The DR-risk function as a functional of the decision function becomes $\mathcal{R}_c^k(f) = \sup_{\mathbb{Q}\in\mathcal{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}\big\{C(\boldsymbol{X})\mathbf{sign}[-f(\boldsymbol{X})]\big\}$. However, directly optimizing the risk $\mathcal{R}_c^k(f)$ is challenging, since the $\mathbf{sign}(\cdot)$ operation is nonconvex and nonsmooth. We consider a specific difference-of-convex (DC) relaxation of the sign operator.

We propose to relax the indicators in the dual form (2.10) by the following robust smoothed ramp loss (Zhou et al., 2017): $\psi(u) := (1-u)^2\mathbb{1}(0 \leqslant u \leqslant 1) + [2-(1+u)^2]\mathbb{1}(-1 \leqslant u \leqslant 0) + 2\mathbb{1}(u \leqslant -1)$. The DC representation is given by $\psi(u) = \psi_+(u) - \psi_-(u)$, where $\psi_+(u) = (1-u)^2\mathbb{1}(0 \leqslant u \leqslant 1) + (1-2u)\mathbb{1}(u \leqslant 0)$, $\psi_-(u) = u^2\mathbb{1}(-1 \leqslant u \leqslant 0) + (-1-2u)\mathbb{1}(u \leqslant -1)$. The advantages of using the symmetric nonconvex loss can be: 1) to protect from outliers in $\boldsymbol{X}$ and improve generalizability (Shen et al., 2003; Wu and Liu, 2007), and 2) to equally indicate $f(\boldsymbol{X}) < 0$ and $f(\boldsymbol{X}) > 0$. We would like to point out that $\mathbb{1}[f(\boldsymbol{X}) < 0] + \mathbb{1}[f(\boldsymbol{X}) > 0] \equiv 1$ will be preserved to $\frac{\psi[f(\boldsymbol{X})]}{2} + \frac{\psi[-f(\boldsymbol{X})]}{2} \equiv 1$ in this surrogate loss. Then we define the DR-ψ-risk function as

$$\mathcal{R}_{c,\psi}^k(f) := \inf_{\eta\in\mathbb{R}} \left\{ c\left[\mathbb{E}\left([C(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[f(\boldsymbol{X})]}{2} + [-C(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[-f(\boldsymbol{X})]}{2}\right)\right]^{1/k^\star} + \eta \right\}. \quad (2.11)$$

Algebraically, we can invert (2.11) to its primal representation $\mathcal{R}_{c,\psi}^k(f) = \sup_{\mathbb{Q}\in\mathcal{P}_c^k(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}[C(\boldsymbol{X})\zeta_\psi(f)]$ by introducing a sign random variable $\zeta_\psi(f) \in \{\pm 1\}$ with $\mathbb{P}(\zeta_\psi(f) = \pm 1|\boldsymbol{X}) := \frac{\psi[\pm f(\boldsymbol{X})]}{2}$. That is, given the covariate $\boldsymbol{X}$, the original deterministic sign $\mathbf{sign}[-f(\boldsymbol{X})]$ is relaxed to the random sign $\zeta_\psi(f)$ with ± 1 probability $\frac{\psi[\pm f(\boldsymbol{X})]}{2}$. In particular, if $f(\boldsymbol{X}) > 0$, then $\mathbf{sign}[-f(\boldsymbol{X})] = -1$ is a hard sign while $\zeta_\psi(f)$ is a soft sign with $\mathbb{P}(\zeta_\psi(f) = -1|\boldsymbol{X}) = \frac{\psi[-f(\boldsymbol{X})]}{2} > \frac{\psi[f(\boldsymbol{X})]}{2} = \mathbb{P}(\zeta_\psi(f) = 1|\boldsymbol{X})$. When $c = 1$, the DR-risk function reduces to the risk function under the training distribution, and the DC relaxation here is equivalent to the relaxation in Zhou et al. (2017).

The DR-ψ-risk function provides the learning objective based on the empirical data. In particular, the population expectation $\mathbb{E}$ is replaced by the empirical average $\mathbb{E}_n$, and the CTE function

$C(\cdot)$ is replaced by a plug-in estimate $\widehat{C}_n(\cdot)$. The corresponding empirical objective is minimized over the decision function f and the auxiliary variables (η, λ) jointly:

$$\begin{aligned}
&\min_{f\in\mathcal{F},\eta\in\mathbb{R}} \left\{ c\left[\mathbb{E}_n\left([\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[f(\boldsymbol{X})]}{2}+[-\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[-f(\boldsymbol{X})]}{2}\right)\right]^{1/k^\star}+\eta\right\}\\
=&\min_{f\in\mathcal{F},\eta\in\mathbb{R},\lambda\geqslant 0}\left\{\frac{c}{k^\star\lambda^{k^\star-1}}\mathbb{E}_n\left([\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[f(\boldsymbol{X})]}{2}+[-\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[-f(\boldsymbol{X})]}{2}\right)+\frac{c\lambda}{k}+\eta\right\}.
\end{aligned}$$

The objective function is a summation of multiple products of DC functions. For $k<+\infty$, we consider a block successive upper-bound minimization algorithm (Razaviyayn et al., 2013) to alternatively minimize the convex upper bounds over the decision function f and the auxiliary variables (η, λ) respectively. For $k=+\infty$, it requires a further probabilistic enhancement to break ties at argmin and ensure the convergence to stationarity (Qi et al., 2019a,b). The implementation details are given in Section 2.7.

2.3 Theoretical Properties

In this section, we justify the validity of the DC relaxation and the empirical substitution. First of all, we introduce the following joint stochastic objectives:

$$\begin{aligned}
\ell_c^k(f,\eta,\lambda;\widetilde{C}) &:= \frac{c}{k^\star\lambda^{k^\star-1}}\left([\widetilde{C}(\boldsymbol{X})-\eta]_+^{k^\star}\mathbb{1}[f(\boldsymbol{X})<0]+[-\widetilde{C}(\boldsymbol{X})-\eta]_+^{k^\star}\mathbb{1}[f(\boldsymbol{X})>0]\right)+\frac{c\lambda}{k}+\eta;\\
\ell_{c,\psi}^k(f,\eta,\lambda;\widetilde{C}) &:= \frac{c}{k^\star\lambda^{k^\star-1}}\left([\widetilde{C}(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[f(\boldsymbol{X})]}{2}+[-\widetilde{C}(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[-f(\boldsymbol{X})]}{2}\right)+\frac{c\lambda}{k}+\eta.
\end{aligned}$$

Here, $\widetilde{C}$ can be the plug-in estimate $\widehat{C}_n$ or the underlying true CTE C. Denote $\mathcal{L}_c^k(f,\eta,\lambda) := \mathbb{E}\ell_c^k(f,\eta,\lambda;C)$, $\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) := \mathbb{E}\ell_{c,\psi}^k(f,\eta,\lambda;C)$. Then by Corollary 2.3, we have $\mathcal{R}_c^k(f) = \inf_{\eta\in\mathbb{R},\lambda\geqslant 0}\mathcal{L}_c^k(f,\eta,\lambda)$, $\mathcal{R}_{c,\psi}^k(f) = \inf_{\eta\in\mathbb{R},\lambda\geqslant 0}\mathcal{L}_{c,\psi}^k(f,\eta,\lambda)$. In the following proposition, we show the validity of the DC relaxation.

Proposition 2.4 (Fisher Consistency and Excess Risk)**.** *Suppose $\mathcal{R}_c^k$, $\mathcal{R}_{c,\psi}^k$, $\mathcal{L}_c^k$ and $\mathcal{L}_{c,\psi}^k$ are defined as above. Fix $k\in(1,+\infty]$, $k^\star=\frac{k}{k-1}$, $c\geqslant 1$, $\eta\in\mathbb{R}$, $\lambda>0$. Then the following results hold:*

(I) (Fisher Consistency)

$$\operatorname*{argmin}_{f:\mathcal{X}\to[-1,1]}\mathcal{L}_{c,\psi}^k(f,\eta,\lambda)=\operatorname*{argmin}_{f:\mathcal{X}\to\{\pm1\}}\mathcal{L}_c^k(f,\eta,\lambda),\qquad \min_{f:\mathcal{X}\to[-1,1]}\mathcal{L}_{c,\psi}^k(f,\eta,\lambda)=\min_{f:\mathcal{X}\to\{\pm1\}}\mathcal{L}_c^k(f,\eta,\lambda);$$

(II) (Excess Risk) Denote $\mathcal{L}_c^{k,*}(\eta,\lambda) := \min_{f\in\mathcal{X}\to\{\pm 1\}} \mathcal{L}_c^k(f,\eta,\lambda)$. *Then for* $f:\mathcal{X}\to\mathbb{R}$, *we have*

$$\mathcal{L}_c^k(f,\eta,\lambda) - \mathcal{L}_c^{k,*}(\eta,\lambda) \leqslant 2[\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) - \mathcal{L}_c^{k,*}(\eta,\lambda)].$$

Denote $\mathcal{R}_c^{k,*} := \inf_{\eta\in\mathbb{R},\lambda\geqslant 0} \mathcal{L}_c^{k,*}(\eta,\lambda)$. *Then for* $f:\mathcal{X}\to\mathbb{R}$, *we have*

$$\mathcal{L}_c^k(f,\eta,\lambda) - \mathcal{R}_c^{k,*} \leqslant 2[\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) - \mathcal{R}_c^{k,*}], \quad \mathcal{R}_c^k(f) - \mathcal{R}_c^{k,*} \leqslant 2[\mathcal{R}_{c,\psi}^k(f) - \mathcal{R}_c^{k,*}].$$

Suppose $\mathcal{F}$ is a functional class on $\mathcal{X}$ with norm $\|\cdot\|_{\mathcal{F}}$ that characterizes the complexity of function. Motivated by Steinwart and Scovel (2007, (6)), we define for $\gamma \geqslant 0$ the constrained version of the approximation error

$$\mathcal{A}_c^k(\gamma) := \inf_{f\in\mathcal{F}} \left\{\mathcal{R}_{c,\psi}^k(f) : \|f\|_{\mathcal{F}} \leqslant \gamma\right\} - \mathcal{R}_c^{k,*}.$$

Similarly to that in Steinwart and Scovel (2007), $\mathcal{A}_c^k(\gamma)$ with the appropriately chosen tuning parameter γ can trade off the learnability and the approximatability of $\mathcal{F}$ towards the population Bayes rule $\operatorname{argmin}_{f:\mathcal{X}\to\{\pm 1\}} \mathcal{R}_c^k(f)$. Specifically, as γ increases, the population approximation error ("bias") $\mathcal{A}_c^k(\gamma)$ decreases with γ, while the empirical complexity ("variance") increases with γ. The trade-off will be stated more explicitly in the following Assumption 2.5.

Next, we make the following assumptions to show the regret bound for the empirical minimization of the ψ-risk $\mathbb{E}_n \ell_{c,\psi}^k(f,\eta,\lambda;\widehat{C}_n)$. Without loss of generality, we restrict to consider the functional class $\mathcal{F}$ as the Reproducing Kernel Hilbert Space (RKHS) with the Gaussian radial basis function kernels, where $\|\cdot\|_{\mathcal{F}}$ is the RKHS-norm. General results can be established by adopting the covering number argument as in Zhao et al. (2019a, Theorem 3.1).

Assumption 2.2 (Boundedness)**.** There exists $M < +\infty$ such that $|C(\boldsymbol{X})| \leqslant M$ almost surely.

Assumption 2.3 (Diffuse Property)**.** The distribution of $C(\boldsymbol{X})$ has a uniformly bounded density with respect to the Lebesgue measure.

Assumption 2.4 (Convergence of the Plug-in CTE)**.** For the CTE estimate $\widehat{C}_n(\boldsymbol{X})$, we assume that $\|\widehat{C}_n - C\|_\infty := \sup_{\boldsymbol{x}\in\mathcal{X}} \left|\widehat{C}_n(\boldsymbol{x}) - C(\boldsymbol{x})\right| \xrightarrow{\mathbb{P}} 0$.

Assumption 2.5 (Approximation Error Rate)**.** There exists $\beta \in (0,1]$ and $K_{\mathcal{A}} < +\infty$ such that for all small enough $\gamma > 0$, we have $\mathcal{A}_c^k(\gamma) \leqslant K_{\mathcal{A}} \gamma^{-\beta}$.

As a remark, we note that Assumption 2.2 can hold if the difference of potential outcomes $Y(1) - Y(-1)$ is uniformly bounded, or $\mathcal{X}$ is compact and $\boldsymbol{x} \mapsto C(\boldsymbol{x})$ is continuous. Assumption 2.3 holds if $\boldsymbol{X}$ has a diffuse distribution, *i.e.*, $\boldsymbol{X}$ doesn't contain points with positive mass; and $\boldsymbol{x} \mapsto C(\boldsymbol{x})$ is injective. Assumption 2.3 is the key assumption to bound λ away from 0. This assumption will not be necessary if $k = +\infty$ and $k^\star = 1$. Assumption 2.4 can be met if $\mathcal{X}$ is compact and $\widehat{C}_n$ is a random forest estimate (Wager and Walther, 2015). Following Steinwart and Scovel (2007, Theorem 2.7), Assumption 2.5 can be shown valid if the Tsybakov's noise assumption on the population margin is met and the kernel bandwidth parameter is chosen appropriately. In the following proposition, we establish the regret bound.

Proposition 2.5 (Regret Bound)**.** *Suppose $\mathcal{R}_c^k$, $\mathcal{R}_{c,\psi}^k$, $\mathcal{L}_c^k$ and $\mathcal{L}_{c,\psi}^k$ are defined as above. Fix $k \in (1, +\infty]$, $k^\star = \frac{k}{k-1}$, $c > 1$. Assume that Assumptions 2.2-2.5 hold. Let*

$$(\widehat{f}_n, \widehat{\eta}_n, \widehat{\lambda}_n) \in \operatorname*{argmin}_{f \in \mathcal{F}, \eta \in \mathbb{R}, \lambda \geqslant 0} \left\{ \mathbb{E}_n \ell_{c,\psi}^k(f, \eta, \lambda; \widehat{C}_n) : \|f\|_{\mathcal{F}} \leqslant \gamma_n \right\},$$

with the tuning parameter γ_n satisfying $\gamma_n = \mathcal{O}(n^{-\frac{1}{2\beta+1}})$ as $n \to \infty$. Then there exists constants $K_0 = K_0(c, M) < +\infty$ and $K_1 = K_1(c, M) < +\infty$ such that for $0 < \delta < 1$, with probability at least $1 - \delta$, we have

$$\mathcal{R}_c^k(\widehat{f}_n) - \mathcal{R}_c^{k,*} \leqslant \mathcal{L}_c^k(\widehat{f}_n, \widehat{\eta}_n, \widehat{\lambda}_n) - \mathcal{R}_c^{k,*} \leqslant K_0 \sqrt{\log(2/\delta)} n^{-\frac{\beta}{2\beta+1}} + K_1 \|\widehat{C}_n - C\|_\infty.$$

In particular, there exists $K_{01}, K_{02}, K_{11}, K_{12} < +\infty$ not depending on c, M, such that

$$K_0(c,M) = \begin{cases} K_{01} \frac{c^{\frac{(k^\star+1)(2k^\star-1)}{k^\star-1}+\frac{1}{2}}}{(c-1)^{k^\star+1/2}} M^{k^\star+1/2}, & k < +\infty; \\ K_{02} c M^{3/2}, & k = +\infty; \end{cases} \quad K_1(c,M) = \begin{cases} K_{11} \frac{c^{2k^\star+1}}{(c-1)^{k^\star-1}} M^{k^\star-1}, & k < +\infty; \\ K_{12} c, & k = +\infty. \end{cases}$$

In Proposition 2.5, it can be of theoretical interest to understand how the regret bound depends on the DR-constant c and the power order k. Specifically, as $c \to +\infty$, η approaches to the essential supremum of $[C(\boldsymbol{X}) - \eta]_+^{k^\star} \frac{\psi[f(\boldsymbol{X})]}{2} + [-C(\boldsymbol{X}) - \eta]_+^{k^\star} \frac{\psi[-f(\boldsymbol{X})]}{2}$ (Krokhmal, 2007, Example 2.3). Then

λ vanishes to 0 so that $1/\lambda$ tends to $+\infty$. Since the Lipschitz constant of $\ell^k_{c,\psi}(f,\eta,\lambda)$ with respect to λ scales with $1/\lambda^{k^\star}$, the universal constants K_0 and K_1 grow to $+\infty$ as well.

Another important fact is that the conjugate number $k^\star$ of k appears in the polynomial orders of c and M respectively in the universal constants K_0 and K_1. In particular, for a large conjugate order $k^\star$, the universal constants K_0 and K_1 increase with the DR-constant c and the CTE bound M more rapidly. In order to achieve a tighter finite sample regret bound, a smaller $k^\star$ and hence a larger k is preferred. Such a phenomenon complements the fact that the power uncertainty set $\mathcal{P}^k_c(\mathbb{P})$ decreases in k. Specifically, as the power order k increases, its conjugate order $k^\star$ decreases, and the regret bound in Proposition 2.5 becomes tighter. On the contrary, the power uncertainty set $\mathcal{P}^k_c(\mathbb{P})$ gets smaller, and the worst-case objective is less distributionally robust. Therefore, the power order k trades off between the distributional robustness in terms of the size of $\mathcal{P}^k_c(\mathbb{P})$, and the finite sample regret bound.

2.4 Simulation Studies

In this section, we carry out two simulation studies to evaluate the generalizability of the DR-ITR on the testing distributions that are different from the training distribution. The first simulation considers the covariat shifts. The second simulation considers the mixture of subgroups.

2.4.1 Covariate Shifts

In this section, we extend the motivating example in Section 2.2.3 to a more practical simulation setting. Consider the training data generating process: $n = 1,000$, $p = 10$, $\boldsymbol{X} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$, $A|\boldsymbol{X} \sim \mathsf{Bernoulli}(1/2)$ and $Y|(\boldsymbol{X}, A) = m(\boldsymbol{X}) + (A - 1/2)C(\boldsymbol{X}) + \mathcal{N}(0,1)$, where $m(\boldsymbol{x}) = 1 + \frac{1}{p}\sum_{j=1}^p x_j$, $C(\boldsymbol{x}) = x_2 - (x_1^3 - 2x_1)$.

At the training stage, we first obtain a CTE function estimate $\widehat{C}_n$ by fitting a casual forest (Wager and Athey, 2018) on the training data. Then we obtain the out-of-bag prediction at the training covariates $\widehat{C}_n(\boldsymbol{X})$. Next we fit the standard ITR by empirically minimizing $\mathbb{E}_n\big\{\widehat{C}_n(\boldsymbol{X})\,(\psi[f(\boldsymbol{X})] - 1)\big\}$ as the ψ-relaxation of the empirical risk function $\mathbb{E}_n\big\{\widehat{C}_n(\boldsymbol{X})\mathbf{sign}[-f(\boldsymbol{X})]\big\}$, over the linear function class $\mathcal{F}_\gamma := \{f(\boldsymbol{x}) = b + \boldsymbol{\beta}^\intercal \boldsymbol{x} : b \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^p,\ \|\boldsymbol{\beta}\|_2 \leqslant \gamma\}$. The tuning parameter $\gamma \geqslant 0$ is determined by 10-fold cross-validation among

$\{0.1, 0.5, 1, 2, 4\}$. Finally, we fit the DR-ITRs for $k = 2$ and $c \in \mathcal{C} = \{1.19, 1.38, \cdots, 20\}$ from the function class $\mathcal{F}_\gamma$, where γ is the same as that of the standard ITR.

We consider the mean-shifted testing distribution $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ for various covariate centroids $\boldsymbol{\mu}$'s. In order to calibrate the DR-constant c for every fixed $\boldsymbol{\mu}$, we generate a calibrating dataset of size $n_{\text{calib}} = 50$ from the testing distribution. The following two scenarios for the calibrating data are considered here: 1) a randomized controlled trial (RCT) dataset $(\boldsymbol{X}, A, Y)$ is generated, with $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ and (A, Y) as before; and 2) only the covariate vector $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ is generated. In Scenario 1, we use the IPWE of the calibrating value function $\widehat{\mathcal{V}}^{\text{IPWE}}_{\text{calib}}(\widehat{f}_c) := \mathbb{E}_{n_{\text{calib}}}\{Y\mathbb{1}[(2A-1)\widehat{f}_c(\boldsymbol{X}) > 0]/(1/2)\}$ to evaulate the DR-constant c, while in Scenario 2, we use the CTE-based calibrating value function $\widehat{\mathcal{V}}^{\text{CTE}}_{\text{calib}}(\widehat{f}_c) := \mathbb{E}_{n_{\text{calib}}}\{\widehat{C}_n(\boldsymbol{X})\mathbf{sign}[\widehat{f}_c(\boldsymbol{X})]\}$ instead. Here, the estimated CTE function $\widehat{C}_n$ is obtained from the training stage.

For comparison, we consider the following: 1) the LB-ITR that maximizes the value function under the testing distribution; 2) the ℓ_1-penalized least-square (ℓ_1-PLS) (Qian and Murphy, 2011) of $Q(\boldsymbol{X}, A) = \mathbb{E}(Y|\boldsymbol{X}, A)$ on $(1, \boldsymbol{X}, A, A\boldsymbol{X})$ and the corresponding estimated ITR $\widehat{d}(\boldsymbol{x}) \in \text{argmin}_{a\in\{\pm 1\}} \widehat{Q}_n(\boldsymbol{x}, a)$; 3) the standard ITR; 4) the RCT-DR-ITR for the calibrating Scenario 1; and 5) the CTE-DR-ITR for the calibrating Scenario 2. We compare the testing values $\mathbb{E}_{n_{\text{test}}}[C(\boldsymbol{X})\widehat{d}(\boldsymbol{X})]$ based on an independent testing dataset of size $n_{\text{test}} = 100,000$ for every testing distribution. The testing values across different testing distributions are not comparable. For a specific testing distribution, the LB-ITR can be a benchmark to be compared to, since its testing value is the best achievable in theory among the linear ITR class. The training-calibrating-testing procedure is replicated for 500 times. The testing values (standard errors) for $n_{\text{calib}} = 50$ are reported in Table 2.2.

When the testing distribution is the same as training $(\mu_1, \mu_2) = (0, 0)$, the calibration procedures for the DR-ITRs are expected to choose $c = 1$, which corresponds to the standard ITR. With the finite calibrating sample, some DR-constant c greater than 1 can be possibly chosen, leading to smaller testing values for the DR-ITRs in Table 2.2. In particular, the testing value of the CTE-DR-ITR is higher than that of the RCT-DR-ITR, and is closer to the testing value of the standard ITR in this case. The reason is that, the RCT-based calibrating value function estimate $\widehat{\mathcal{V}}^{\text{IPWE}}_{\text{calib}}$ depends on $(\boldsymbol{X}, A, Y)$ in the calibrating data, while the CTE-based one $\widehat{\mathcal{V}}^{\text{CTE}}_{\text{calib}}$ depends on

$\boldsymbol{X}$ only. As a consequence, the CTE-based calibration can be more accurate than the RCT-based one.

Table 2.2: Testing Values (Standard Errors) on the Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	type	0	0.734	1.469	1.958
1.958	LB-ITR	*2.333 (0.00244)*	*2.907 (0.011)*	*5.334 (0.0362)*	*9.27 (0.0154)*
	ℓ_1-PLS	**2.124** (0.0022)	2.235 (0.011)	3.613 (0.0505)	6.32 (0.103)
	Standard ITR	2.089 (0.00158)	1.735 (0.013)	1.348 (0.0595)	1.567 (0.13)
	RCT-DR-ITR	2.085 (0.00444)	2.286 (0.0114)	4.545 (0.0255)	8.371 (0.0451)
	CTE-DR-ITR	2.098 (0.00348)	**2.304** (0.0106)	**4.551** (0.0238)	**8.459** (0.0424)
1.469	LB-ITR	*1.893 (0.00712)*	*2.627 (0.00656)*	*5.28 (0.0213)*	*9.379 (0.0128)*
	ℓ_1-PLS	1.667 (0.00307)	**2.021** (0.0076)	4.095 (0.0342)	7.573 (0.0706)
	Standard ITR	**1.674** (0.00152)	1.645 (0.0127)	2.377 (0.0553)	4.011 (0.119)
	RCT-DR-ITR	1.627 (0.00688)	1.987 (0.00997)	4.484 (0.0192)	8.611 (0.0285)
	CTE-DR-ITR	1.663 (0.00326)	1.997 (0.00992)	**4.55** (0.0163)	**8.686** (0.0269)
0.734	LB-ITR	*1.227 (0.00244)*	*2.144 (0.00609)*	*5.269 (0.00931)*	*9.608 (0.00898)*
	ℓ_1-PLS	1.094 (0.00418)	**1.676** (0.00442)	4.587 (0.0151)	8.8 (0.0314)
	Standard ITR	**1.174** (0.00149)	1.553 (0.00806)	3.739 (0.0379)	7.06 (0.0763)
	RCT-DR-ITR	1.094 (0.00753)	1.651 (0.00675)	4.622 (0.0109)	9.036 (0.015)
	CTE-DR-ITR	1.152 (0.00292)	1.667 (0.00588)	**4.648** (0.0113)	**9.06** (0.0161)
0.000	LB-ITR	*0.9942 (0.00202)*	*1.774 (0.0034)*	*5.232 (0.00559)*	*9.767 (0.0068)*
	ℓ_1-PLS	0.8296 (0.00454)	1.648 (0.0036)	**4.914** (0.00501)	**9.476** (0.0103)
	Standard ITR	**0.9437** (0.00153)	1.679 (0.00336)	4.654 (0.017)	8.895 (0.0342)
	RCT-DR-ITR	0.8374 (0.00821)	1.647 (0.00574)	4.868 (0.00797)	9.444 (0.00841)
	CTE-DR-ITR	0.9206 (0.00272)	**1.688** (0.00289)	4.888 (0.00698)	9.442 (0.00999)

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Values (larger the better) can be comparable for the same (μ_1, μ_2) but incomparable across different (μ_1, μ_2).
[3] LB-ITR maximizes the testing value function at (μ_1, μ_2) over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.

When $(\mu_1, \mu_2) \neq (0, 0)$, the testing distribution is different from training, and the performance of the standard ITR deteriorates while the DR-ITRs still maintain reasonably good performance. The phenomenon is more evident when $\mu_1, \mu_2 \in \{1.469, 1.958\}$. In particular at $(\mu_1, \mu_2) = (1.958, 1.958)$, the value of the standard ITR can be as low as 17% of the best achievable value among the linear ITR class, while the DR-ITRs can maintain more than 90%. In fact, such a phenomenon is general. In Figure 2.3a, we further enumerate the testing covariate centroid $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ for $\mu_1, \mu_2 \in [-2.448, 2.448]$ and compute the relative regrets of the standard ITR and the RCT-DR-ITR. Across all mean-shifted testing distributions, the relative regrets of the standard ITRs can be as high as 108%, in which case the standard ITR value is negative, and hence even worse than the completely random treatment rule d_{rand}. On the contrary, the relative regrets for the RCT-DR-ITR ($n_{\text{calib}} = 50$) shown in Figure 2.3b are at most 24% across all testing centroids. This suggests that

the RCT-DR-ITR maintains relatively good performance on all such testing distributions, while the standard ITR fails. Figure 2.4 further shows that the DR-ITR provides substantial testing value improvements over the standard ITR. This demonstrates that the small sample size $n_{\text{calib}} = 50$ is sufficient for calibrating the DR-ITR with significant testing improvement.

From Table 2.2, it can be also observed that ℓ_1-PLS can have better performance than the standard ITR when training and testing distributions are different. The reason is that, the objective of ℓ_1-PLS does not target the value function under the training distribution directly, but rather, the mean squared error of the linear approximation to $Q(\boldsymbol{X}, A)$ under the training distribution. Such a linear approximation can perform well when the testing distribution is not far from the training distribution. However, in the case $\mu_1, \mu_2 \in \{1.469, 1.958\}$ in the sense that the testing distribution deviates more from the training one, the DR-ITRs enjoy notably higher testing values than ℓ_1-PLS.

In Section 2.7, we provide more detailed results for other comparisons including the relative regrets/improvements on all mean-shifted covariate domains of all centroids, the misclassification rates on all mean-shifted covariate domains of all centroids, the comparison with some other methods in relative regrets and misclassification rates, and the case of $k \in \{1.25, 1.5, 2, 3, \infty\}$. In particular, the misclassification rates inform similar conclusions as the relative regrets/improvements. If we increase the calibrating sample size from 50 to 100, then the testing values of DR-ITRs can be further improved. We also find that among our simulation scenarios, the testing values of the DR-ITR are not very sensitive to difference choices of k.

2.4.2 Performance on the Mixture of Subgroups

In this section, we consider a population that consists of two subgroups, with each following a distinct CTE function. We aim to find an ITR that can generalize well on different mixtures of subgroups.

We modify the simulation setup in Section 2.4.1 as follows: $\boldsymbol{X}|\xi \sim \xi\mathcal{N}_p(\boldsymbol{\mu}_1, \mathbf{I}_p) + (1 - \xi)\mathcal{N}_p(\boldsymbol{\mu}_0, \mathbf{I}_p)$, where $\xi \sim \mathsf{Bernoulli}(p_{\text{mix}})$ is the unobservable mixture/subgroup indicator with subgroup 1 probability p_{mix} and subgroup 0 probability $1 - p_{\text{mix}}$, and the subgroup means $\boldsymbol{\mu}_1 = (-1/2, 1/2, 0, \cdots, 0)^\intercal$ and $\boldsymbol{\mu}_0 = -\boldsymbol{\mu}_1$. We consider the CTE function $C(\boldsymbol{x}; \xi) := (2\xi - 1)\beta_0 + \beta_1 x_1 + \beta_2 x_2$ that is linear in the covariate vector, but with a subgroup-dependent intercept

(a) Standard ITR **(b)** RCT-DR-ITR ($n_{\text{calib}} = 50$)

Figure 2.3: Relative Regrets on the Mean-Shifted Covariate Domains (lighter the better).

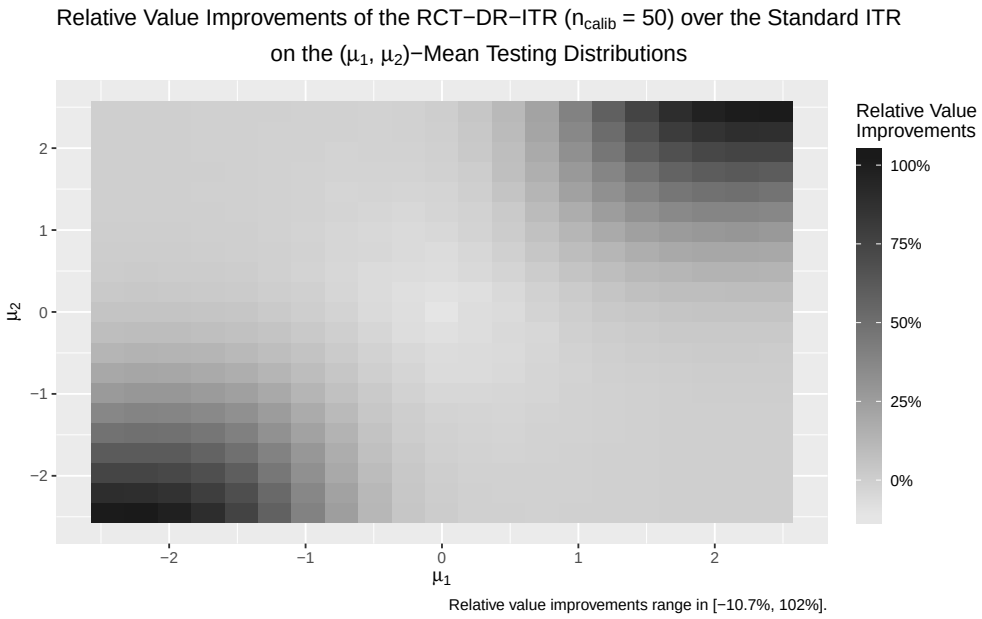


Figure 2.4: Relative improvements of the RCT-DR-ITR over the standard ITR as the difference of their relative regrets on the mean-shifted covariate domains ($n_{\text{calib}} = 50$, darker the better).

$(2\xi - 1)\beta_0$, and $(\beta_0, \beta_1, \beta_2) := (-3/2, -2, 1)$. The unconditional CTE function is nonlinear:

$$C(\boldsymbol{x}) := \mathbb{E}[C(\boldsymbol{x}; \xi) | \boldsymbol{X} = \boldsymbol{x}] = \frac{p_{\text{mix}} \exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_1\|_2^2/2) - (1 - p_{\text{mix}}) \exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_0\|_2^2/2)}{p_{\text{mix}} \exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_1\|_2^2/2) + (1 - p_{\text{mix}}) \exp(-\|\boldsymbol{x} - \boldsymbol{\mu}_0\|_2^2/2)} \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

In particular, the unconditional CTE function $C(\boldsymbol{x})$ depends on the subgroup 1 probability p_{mix}. The distributional changes are due to the subgroup 1 probability. Specifically, the training subgroup 1 probability is 0.75, while the testing subgroup 1 probability varies in $\{0.1, 0.25, 0.5, 0.75, 0.9\}$. Since the training and testing CTE functions can be different, Assumption 2.1 cannot be fully met. Therefore, our proposed DR-ITR can be robust to such distributional changes only to some extent.

We consider the same training-calibrating-testing procedure as that in Section 2.4.1, except that the DR-constant c ranges in $\{1.18, 1.27, \cdots, 10\}$. The testing values of the ITRs are reported in Table 2.3. When the training and testing distributions are the same at $p_{\text{mix}} = 0.75$, all ITRs have similar testing performance. The standard ITRs have higher testing values than the DR-ITRs in this case. When the testing p_{mix} becomes smaller, the DR-ITRs show better testing performance than the standard ITR. When the testing $p_{\text{mix}} = 0.25$ or 0.1, the RCT-DR-ITR has the highest testing values among all. Since the true testing CTE function changes along with the testing p_{mix}, the corresponding estimate $\widehat{C}_n$ based on the training data can suffer from the generalizability problem. Therefore, the CTE-based calibration performs slightly worse than the RCT-based calibration in this case. However, the CTE-based DR-ITR is superior to the standard ITR, and is comparable to the ℓ_1-PLS. More detailed comparisons and the case $n_{\text{calib}} = 100$ are provided in Section 2.7.

Table 2.3: Testing Values (Standard Errors) on the Mixture of Subgroups ($n_{\text{calib}} = 50$)

	Testing Subgroup 1 Probability				
type	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**
LB-ITR	*1.665 (0.0067)*	*1.537 (0.00618)*	*1.444 (0.00412)*	*1.545 (0.00537)*	*1.679 (0.00585)*
ℓ_1-PLS	1.182 (0.00191)	1.264 (0.0014)	**1.399** (0.000591)	**1.537** (0.000333)	1.624 (0.000781)
Standard ITR	1.143 (0.00434)	1.232 (0.00329)	1.383 (0.0015)	1.535 (0.000543)	**1.632** (0.00142)
RCT-DR-ITR	**1.267** (0.0066)	**1.305** (0.00423)	1.395 (0.00256)	1.52 (0.00212)	1.614 (0.00234)
CTE-DR-ITR	1.16 (0.00409)	1.247 (0.00323)	1.388 (0.00137)	1.534 (0.00055)	1.628 (0.00149)

[1] Testing subgroup 1 probability = 0.75 is the same as the training one.
[2] Values (larger the better) can be comparable for the same subgroup 1 probability but incomparable across different subgroup 1 probabilities
[3] LB-ITR maximizes the testing value function over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.

2.5 Application to the ACTG 175 Trial Data

In this section, we evaluate the generalizability of our proposed DR-ITR on a clinical trial dataset from the "AIDS clinical trial group study 175" (Hammer et al., 1996). The goal of this study was to compare four treatment arms among 2,139 randomly assigned subjects with human immunodeficiency virus type 1 (HIV-1), whose CD4 counts were 200-500 cells/mm^3. The four treatments are

the zidovudine (ZDV) monotherapy, the didanosine (ddI) monotherapy, the ZDV combined with ddI, and the ZDV combined with zalcitabine (ZAL).

The evidence found from the AIDS trial data can have some generalizability problems. When studying women living with HIV and women at risk for HIV infection in the USA cohort, the Women's Interagency HIV Study (WIHS) (Bacon et al., 2005) has been considered to be representative. However, it was reported in Gandhi et al. (2005) that 28-68% of the HIV positive women in WIHS were excluded from the eligibility criteria of many ACTG studies. In the ACTG 175 dataset, the number of female patients is only 368 out of 2139. Thus we suspect that the female patients may be underrepresented in this dataset, and the ITR based on the dataset may not generalize well on the women subgroup. In this section, we study the generalizability of DR-ITR when the testing dataset consists of female patients only. Specifically, the training dataset is a subsample from ACTG 175 with original male/female proportion, while the testing dataset is a subsample from the female patients of ACTG 175, and there is no overlap across training and testing. We try to resemble the ideal world that we can have independent testing data from the female population.

We consider the outcome Y as the difference between the early stage (at 20 ± 5 weeks from baseline) CD4 cell counts and the CD4 counts at baseline. We focus on the treatment comparison between the ZDV + ZAL ($A = 1$) and the ddI ($A = -1$), and the corresponding patients from the dataset. In particular, only 180 of them are women. The average treatment effects on the male and female subgroups are -8.97 and -1.39 respectively, which suggests that there is treatment effect discrepancy between these subgroups. We sample the training data from the ACTG 175 dataset in the ZDV + ZAL or ddI arm of sample size $1,085 \times 60\% = 651$ stratified to the gender. In particular, the training dataset includes $180 \times 60\% = 108$ female patients. The remaining female data ($180 - 108 = 72$) are used for testing. We only consider female patients in testing. We further sample 50 from the testing female data for calibration, and the remaining ($72 - 50 = 22$) are the testing dataset. We also consider 12 selected baseline covariates $\boldsymbol{X}$ as was studied in Lu et al. (2013). There are 5 continuous covariates: age (year), weight (kg, coded as `wtkg`), CD4 count (cells/mm^3) at baseline, Karnofsky score (scale of 0-100, coded as `karnof`), CD8 count (cells/mm^3) at baseline. They are centered and scaled before further analysis. In addition, there are 7 binary variables: gender (1 = male, 0 = female), homosexual activity (`homo`, 1 = yes, 0 = no), race (1 = nonwhite, 0 = white), history of intravenous drug use (`drug`, 1 = yes, 0 = no), symptomatic status

(`symptom`, 1 = symptomatic, 0 = asymptomatic), antiretroviral history (`str2`, 1 = experienced, 0 = naive) and hemophilia (`hemo`, 1 = yes, 0 = no).

Before fitting ITRs, we estimate the CTE function $C(\boldsymbol{X})$ by the following regress-and-subtract procedure: first we fit two separate random forests by regressing Y on $\boldsymbol{X}$ restricted on $A = 1$ and $A = -1$ respectively; then we subtract two regression models to obtain the CTE function estimate $\widehat{C}_n(\boldsymbol{X})$. We follow the same implementation as in Section 2.4.1 to fit the standard ITR and DR-ITRs over a constrained linear function class $\mathcal{F}_\gamma := \{f(\boldsymbol{x}) = b + \boldsymbol{\beta}^\intercal \boldsymbol{x} : b \in \mathbb{R},\ \boldsymbol{\beta} \in \mathbb{R}^p,\ \|\boldsymbol{\beta}\|_2 \leqslant \gamma\}$ on the training data. The testing performance is evaluated by the IPWE of the value function on the testing data. The training-calibrating-testing procedure is repeated for 1,500 times. The testing values are reported in Table 2.4, where the value can be interpreted as the expected CD4 count improvement from baseline at the early stage (20 ± 5 weeks). In addition to the calibrated DR-ITRs, we also include the value of the *best DR-ITR* that enjoys the highest testing performance among all DR-constants. For comparison, we include the results of residual weighted learning (RWL) (Zhou et al., 2017) with linear kernel. Both RWL and the standard ITR share similar implementation, except that RWL can be shown equivalently using $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$ as a plug-in CTE estimate.

The testing results show that our proposed DR-ITRs can have better values than the standard ITR and RWL. In particular, the improvement of the best DR-ITR is substantial, while the improvements of the calibrated ITRs are not as strong. We plot the testing values of the DR-ITRs against the corresponding DR-constants in Figure 2.5. It suggests that the testing values generally increase with the DR-constant. In this analysis, the calibrated DR-constants are not close to the optimal DR-constant. As a result, the testing performance of the calibrated DR-ITRs is not as good as the best DR-ITR. One reason for this phenomenon can be that the outcome Y has a heavy tail distribution, as was highlighted in Qi et al. (2019b), so that the value function estimate is highly variable based on the small calibrating sample. Another reason can be that the random forest regress-and-subtract estimate of the CTE function does not generalize well on the testing distribution.

On the overall dataset, we fit the DR-ITRs and report their fitted coefficients in Table 2.5 for selected DR-constants. To stabilize the randomness from the random forest estimate of the CTE function, we refit the random forest 20 times and average the corresponding DR-ITR coefficients.

Table 2.4: Expected CD4 Count Improvement (cells/mm^3) from Baseline at the Early Stage (20±5 weeks) and Standard Errors on the ACTG-175 Female Patients (higher the better).

RWL	Standard ITR	Best DR-ITR	RCT-DR-ITR	CTE-DR-ITR
10.7617 (0.8636)	10.593 (0.8627)	**13.9423** (0.8378)	11.8133 (0.8357)	11.1563 (0.8514)

Standard errors are computed based on 1,500 replications.

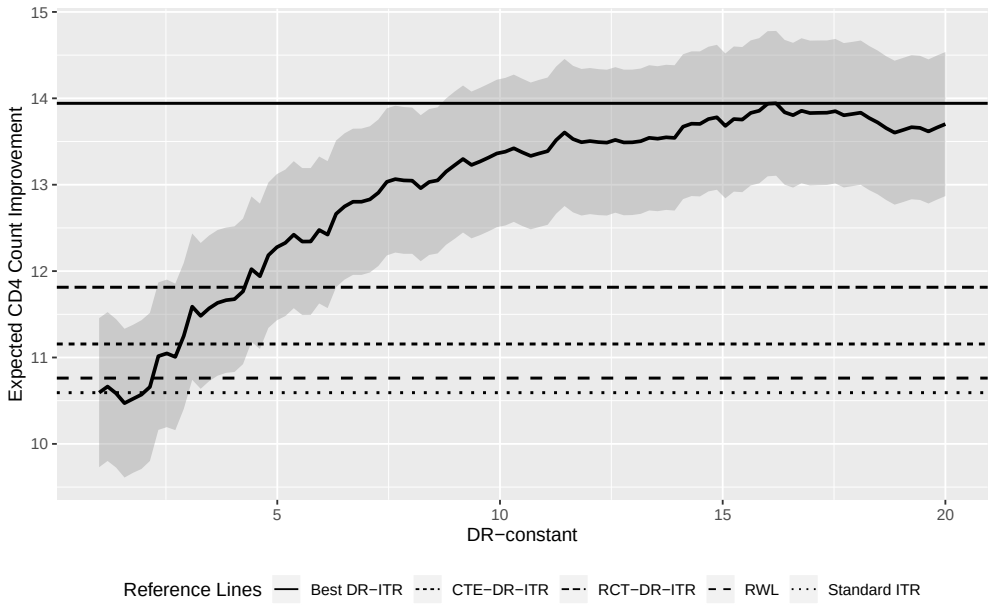


Figure 2.5: Expected CD4 Count Improvement (cells/mm^3) from Baseline at the Early Stage (20±5 weeks) of the DR-ITRs of Various DR-Constants on the ACTG 175 Female Patients (higher the better)

We find that there are noticeable changes in the coefficients of the intercept and the homosexual activity when the DR-constant gets large. Within the ACTG 175 dataset (ZDV + ZAL or ddI), we find that only 6 female patients have homosexual activity. Four of them are treated with ZDV + ZAL, and the change of their CD4 counts are 123, 34, −11 and 158 respectively. Two of them are treated with ddI, and the change of their CD4 counts are −41, −182. Therefore, the ZDV + ZAL ($A = +1$) may have more benefits compared to the ddI ($A = -1$) on these patients. This helps to explain why the larger coefficients in homosexual activity for the larger DR-constants can be beneficial for the female patients.

Table 2.5: Linear Coefficients of the DR-ITRs Fitted on the ACTG 175 Dataset

DR-constant	Intercept	age	wtkg	cd40	karnof	cd80	gender	homo	race	drugs	symptom	str2	hemo
1	−0.02	−0.25	0.06	−0.58	−0.06	0.53	−0.16	−0.4	0.16	0.16	0.16	0.16	0.09
4.8	−0.31	−0.23	0.12	−0.67	0.11	0.55	−0.12	−0.21	0.2	0.12	0.1	−0.06	0.09
8.6	−0.43	−0.23	0.11	−0.64	0.16	0.54	−0.11	−0.05	0.12	0.04	0.07	−0.24	0.01
12.4	−0.54	−0.22	0.1	−0.64	0.19	0.51	−0.04	0.01	0.08	0.05	0.04	−0.27	−0.02
16.2	−0.61	−0.23	0.1	−0.64	0.2	0.51	0	0.03	0.06	0.05	0.02	−0.27	−0.02
20	−0.64	−0.24	0.09	−0.63	0.22	0.5	0.01	0.03	0.05	0.07	0.01	−0.26	−0.01

[1] DR-constant = 1 corresponds to the standard ITR; DR-constant = 16.2 has the highest testing value in Figure 2.5.

2.6 Discussion

In this chapter, we propose a new framework for learning a distributionally robust ITR by maximizing the worst-case value function among values under distributions within the power uncertainty set. We introduce two possible calibration scenarios under which the DR-constant can be tuned adaptively to a small amount of the calibrating data from the target population. In this way, when the training and testing distributions are identical, the calibrated DR-ITRs can achieve similar performance as compared to the standard ITR. When the testing distribution deviates from the training distribution, we show that there are many possible scenarios that the standard ITR generalizes poorly, while the calibrated DR-ITRs maintain relatively good testing performance. Our simulation studies and an application to the ACTG 175 dataset demonstrate the competitive generalizability of our proposed DR-ITR.

The main assumption on the changes of covariates in our DR-ITR framework is equivalent to the selection unconfoundedness assumption in a randomized controlled trial. In practice, there may exist unmeasured selection confounding problems for the trial data, and the distributional changes affect both the covariates and the CTE function. One possible extension is to consider the simultaneous changes of the covariate distribution and the CTE function, and leverage more general robustness measure against these changes.

In our DR-ITR framework, we require an estimate of the CTE function based on the flexible nonparametric techniques. The performance of our DR-ITR can depend on the quality of the CTE function estimate. An alternative strategy is to avoid plugging in a CTE estimate. Instead, the dual representation (2.10) can be identified from $(\boldsymbol{X}, A, Y)$ directly using a variational representation of $[\pm C(\boldsymbol{X}) - \eta]_+^{k^\star}$ (Duchi et al., 2019). This can be a possible extension of our framework.

Another possible extension is to consider the problem of high-dimensional covariates. Our current formulation involves an ℓ_2-constraint to control the model complexity. It can be extended

to obtain sparse solutions when a ℓ_1-constraint is used instead. Besides the high-dimensional extension, our current theoretical results assume that $C(\boldsymbol{X})$ is uniformly bounded. It will be interesting to relax the assumption, such as sub-Gaussianity. Further investigations along these lines can be pursued.

2.7 Appendix

2.7.1 Explicit Forms of the Power Uncertainty Set

In this section, we study the explicit forms of the power uncertainty set $\mathcal{P}_c^k(\mathbb{P})$ on certain parameteric families of distributions, and how they depend on the DR-constant c and the power k. We first examine the family of Bernoulli distributions and the normal distributions, and show that their power uncertainty sets depend on c and k differently. Then the general exponential family will be discussed.

Example 2.1 (Bernoulli Distributional Ball)**.** Consider two Bernoulli distributions $\mathsf{Bernoulli}(p)$ and $\mathsf{Bernoulli}(q)$ for some $p, q \in [0,1]$. We have $\left\| \frac{\mathrm{d}\mathsf{Bernoulli}(q)}{\mathrm{d}\mathsf{Bernoulli}(p)} \right\|_{L^k(\mathsf{Bernoulli}(p))} = \left[p\left(\frac{q}{p}\right)^k + (1-p)\left(\frac{1-q}{1-p}\right)^k \right]^{1/k}$. If $p \leqslant q$, then the above becomes $\frac{q}{p}\left[p + (1-p)\left(\frac{p(1-q)}{q(1-p)}\right)^k \right]^{1/k} \in [(q/p) \times p^{1/k}, q/p]$. If $p \geqslant q$, then the above becomes $\frac{1-q}{1-p}\left[p\left(\frac{q(1-p)}{p(1-q)}\right)^k + 1 - p \right]^{1/k} \in \left[\frac{1-q}{1-p} \times (1-p)^{1/k}, \frac{1-q}{1-p} \right]$. As $k \to +\infty$, the above both approach to $\frac{q}{p} \vee \frac{1-q}{1-p}$. For fixed p and every $k \in [1, +\infty)$, we have

$$\mathcal{P}_c^k(\mathsf{Bernoulli}(p)) \supseteq \{\mathsf{Bernoulli}(q) : q \in [0,1],\ 1 - c(1-p) \leqslant q \leqslant cp\},$$

and

$$\mathcal{P}_c^k(\mathsf{Bernoulli}(p))^{\complement} \supseteq \left\{ \mathsf{Bernoulli}(q) : q \in [0,1],\ q > \frac{cp}{p^{1/k}} \text{ or } q < 1 - \frac{c(1-p)}{(1-p)^{1/k}} \right\},$$

with the meaningful $c \leqslant \frac{1}{p \wedge (1-p)}$. In particular as the large enough k increases while $1 < c \leqslant \frac{1}{p \wedge (1-p)}$ is fixed, $\mathcal{P}_c^k(\mathsf{Bernoulli}(p))$ contains less Bernoulli distributions, down to that of success probabilities in $[1 - c(1-p), cp]$ only.

Example 2.2 (Normal Distributional Ball of Mean Shifts)**.** Consider two p-dimensional normal distributions $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ and $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$ for some center parameter $\boldsymbol{\mu} \in \mathbb{R}^p$. The density ratio of $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p)$

w.r.t. $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ is given by $\frac{\exp(-\|\boldsymbol{x}-\boldsymbol{\mu}\|_2^2/2)}{\exp(-\|\boldsymbol{x}\|_2^2/2)} = e^{-\|\boldsymbol{\mu}\|_2^2/2} \times e^{\boldsymbol{\mu}^\intercal \boldsymbol{x}}$. Then the L^k-norm of the density ratio under $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ can be calculated analytically as $e^{-\|\boldsymbol{\mu}\|_2^2/2} \left(\int_{\mathbb{R}^p} e^{k\boldsymbol{\mu}^\intercal \boldsymbol{x}} \times (2\pi)^{-p/2} e^{-\|\boldsymbol{x}\|_2^2/2} \mathrm{d}\boldsymbol{x} \right)^{1/k} = e^{(k-1)\|\boldsymbol{\mu}\|_2^2/2}$. Then $\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{I}_p) \in \mathcal{P}_c^k(\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p))$ if and only if $e^{(k-1)\|\boldsymbol{\mu}\|_2^2/2} \leqslant c \Leftrightarrow \|\boldsymbol{\mu}\|_2^2 \leqslant \frac{2\log c}{k-1}$.

Note that the conclusion is presented in terms of the L^2-difference of the mean vectors $\|\boldsymbol{\mu}\|_2$ between two normal components. It can be extended to two p-dimensional normal distributions of the same covariance matrix: $\mathcal{N}_p(\boldsymbol{\mu}_1, \Sigma) \in \mathcal{P}_c^k(\mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma))$ if and only if $\exp\left\{\frac{k-1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\right\} \leqslant c \Leftrightarrow (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \leqslant \frac{2\log c}{k-1}$. Then we have

$$\mathcal{P}_c^k(\mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma)) \supseteq \left\{ \mathcal{N}_p(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^p,\ (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) \leqslant \frac{2\log c}{k-1} \right\},$$

and

$$\mathcal{P}_c^k(\mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma))^\complement \supseteq \left\{ \mathcal{N}_p(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^p,\ (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) > \frac{2\log c}{k-1} \right\}.$$

In particular as k increases with $c > 1$ fixed, $\mathcal{P}_c^k(\mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma))$ contains less normal distributions of covariance matrix Σ.

Example 2.3 (Normal Distributional Ball of Covariance Scales)**.** Consider two p-dimensional normal distributions $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ and $\mathcal{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{I}_p)$ for some scale parameter $\sigma^2 > 0$. The density ratio of $\mathcal{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{I}_p)$ *w.r.t.* $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ is given by $\frac{\sigma^{-p} \exp\{-\|\boldsymbol{x}\|_2^2/(2\sigma^2)\}}{\exp(-\|\boldsymbol{x}\|_2^2/2)} = \sigma^{-p} e^{-(\sigma^{-2}-1)\|\boldsymbol{x}\|_2^2/2}$. Then the L^k-norm of the density ratio under $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ can be calculated analytically as $\sigma^{-p} \left(\int_{\mathbb{R}^p} e^{-k(\sigma^{-2}-1)\|\boldsymbol{x}\|_2^2/2} \times (2\pi)^{-p/2} e^{-\|\boldsymbol{x}\|_2^2/2} \mathrm{d}\boldsymbol{x} \right)^{1/k} = \sigma^{-p}[k(\sigma^{-2}-1)+1]^{-p/(2k)}$, which is a nonlinear function in σ^2 ranging in $(0, k^\star)$ and attaining the minimum at $\sigma^2 = 1$. Then $\mathcal{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{I}_p) \in \mathcal{P}_c^k(\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p))$ if and only if $\sigma^{-p}[k(\sigma^{-2}-1)+1]^{-p/(2k)} \leqslant c \Leftrightarrow \underline{\sigma}_k^2(c) \leqslant \sigma^2 \leqslant \bar{\sigma}_k^2(c)$ where $\underline{\sigma}_k^2(c) \in (0,1)$ and $\bar{\sigma}_k^2(c) \in (1, k^\star)$ are the unique roots solving the nonlinear equation $\sigma^{-p}[k(\sigma^{-2}-1)+1]^{-p/(2k)} = c$ $\Leftrightarrow \sigma^{-2k} - c^{2k/p}[k(\sigma^{-2}-1)+1] \overset{t:=\sigma^{-2}-1}{=\!=} (t+1)^k - c^{2k/p}(kt+1) = 0$ on the interval $t \in (c^{2k^\star/p} - 1, +\infty)$ $\Leftrightarrow \sigma^2 \in (0, c^{-2k^\star/p})$ and $t \in (-1/k, 0) \Leftrightarrow \sigma^2 \in (1, k^\star)$ respectively. In particular as k increases while c is fixed, the lower root $\underline{\sigma}_k^2(c)$ increases to 1 while the upper root $\bar{\sigma}_k^2(c)$ decreases to 1, so that $\mathcal{P}_c^k(\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p))$ contains fewer and fewer distributions of the form $\mathcal{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{I}_p)$ with $\sigma^2 \in [\underline{\sigma}_k^2(c), \bar{\sigma}_k^2(c)]$.

The result is general if the mean vector $\mathbf{0}_p$ is replaced by any vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and the covariance matrix $\mathbf{I}_p$ is replaced by some positive semi-definite matrix Σ:

$$\mathcal{P}_c^k(\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)) \supseteq \left\{\mathcal{N}_p(\boldsymbol{\mu}, \sigma^2\Sigma) : \underline{\sigma}_k^2(c) \leqslant \sigma^2 \leqslant \overline{\sigma}_k^2(c)\right\},$$

and

$$\mathcal{P}_c^k(\mathcal{N}_p(\boldsymbol{\mu}, \Sigma))^{\complement} \supseteq \left\{\mathcal{N}_p(\boldsymbol{\mu}, \sigma^2\Sigma) : \sigma^2 < \underline{\sigma}_k^2(c) \text{ or } \sigma^2 > \overline{\sigma}_k^2(c)\right\}.$$

As an extension of the Bernoulli and the normal distribution, we can also consider the mixture of two fixed normal components.

Lemma 2.6 (Upper Bound of the Mixture ϕ-Divergence)**.** *Suppose $\mathbb{P}_0, \mathbb{P}_1$ are probability distributions, $p, q \in [0,1]$. Denote $\mathbb{P}_p := p\mathbb{P}_1 + (1-p)\mathbb{P}_0$, $\mathbb{P}_q := q\mathbb{P}_1 + (1-q)\mathbb{P}_0$. Let $\phi \in \Phi$ be a legitimate divergence function. Then*

$$D_\phi(\mathbb{P}_q \| \mathbb{P}_p) \leqslant D_\phi\big(q\mathbb{P}_1 \big\| (1-p)\mathrm{d}\mathbb{P}_0\big) + D_\phi\big((1-q)\mathbb{P}_0 \big\| p\mathbb{P}_1\big).$$

Proof.

$$\begin{aligned}
&D_\phi(\mathbb{P}_q \| \mathbb{P}_p) \\
=&\int \phi\left(\frac{q\mathrm{d}\mathbb{P}_1 + (1-q)\mathrm{d}\mathbb{P}_0}{p\mathrm{d}\mathbb{P}_1 + (1-p)\mathrm{d}\mathbb{P}_0}\right)[p\mathrm{d}\mathbb{P}_1 + (1-p)\mathrm{d}\mathbb{P}_0] \\
=&\int \phi\left(\frac{(1-p)\mathrm{d}\mathbb{P}_0}{p\mathrm{d}\mathbb{P}_1 + (1-p)\mathrm{d}\mathbb{P}_0} \times \frac{q\mathrm{d}\mathbb{P}_1}{(1-p)\mathrm{d}\mathbb{P}_0} + \frac{p\mathrm{d}\mathbb{P}_1}{p\mathrm{d}\mathbb{P}_1 + (1-p)\mathrm{d}\mathbb{P}} \times \frac{(1-q)\mathrm{d}\mathbb{P}_0}{p\mathrm{d}\mathbb{P}_1}\right)[p\mathrm{d}\mathbb{P}_1 + (1-p)\mathrm{d}\mathbb{P}_0] \\
\overset{\text{Jensen}}{\leqslant}&\int \phi\left(\frac{q\mathrm{d}\mathbb{P}_1}{(1-p)\mathrm{d}\mathbb{P}_0}\right)(1-p)\mathrm{d}\mathbb{P}_0 + \int \phi\left(\frac{(1-q)\mathrm{d}\mathbb{P}_0}{p\mathrm{d}\mathbb{P}_1}\right)p\mathrm{d}\mathbb{P}_1 \\
=&D_\phi\big(q\mathbb{P}_1 \big\| (1-p)\mathrm{d}\mathbb{P}_0\big) + D_\phi\big((1-q)\mathbb{P}_0 \big\| p\mathbb{P}_1\big).
\end{aligned}$$

□

Remark 2.7. The conclusion can be stated in terms of the k-th moment of the density ratio. Suppose $\mathbb{P}_0 \ll \mathbb{P}_1$ and $\mathbb{P}_1 \ll \mathbb{P}_0$. Then

$$\left\|\frac{\mathrm{d}\mathbb{P}_q}{\mathrm{d}\mathbb{P}_p}\right\|_{L^k(\mathbb{P}_p)}^k \leqslant (1-p)\left(\frac{q}{1-p}\right)^k \left\|\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}\right\|_{L^k(\mathbb{P}_0)}^k + p\left(\frac{1-q}{p}\right)^k \left\|\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}_1}\right\|_{L^k(\mathbb{P}_1)}^k.$$

Remark 2.8 (Mixture Normal Distributional Ball)**.** Consider two mixture normal distributions $\mathsf{GMM}_p(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0; \Sigma) := p\mathcal{N}_d(\boldsymbol{\mu}_1, \Sigma) + (1-p)\mathcal{N}_d(\boldsymbol{\mu}_0, \Sigma)$ and $\mathsf{GMM}_q(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0; \Sigma) := q\mathcal{N}_d(\boldsymbol{\mu}_1, \Sigma) + (1-q)\mathcal{N}_d(\boldsymbol{\mu}_0, \Sigma)$ with the same components and different mixture probabilities $p, q \in [0,1]$. Example 2.2, Lemma 2.6 and Example 2.1 together imply that

$$\begin{aligned}
&\left\| \frac{\mathrm{d}\mathsf{GMM}_q(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0; \Sigma)}{\mathrm{d}\mathsf{GMM}_p(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0; \Sigma)} \right\|_{L^k(\mathsf{GMM}_p(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0; \Sigma))} \\
&\leqslant \left[(1-p) \left(\frac{q}{1-p} \right)^k + p \left(\frac{1-q}{p} \right)^k \right]^{1/k} \exp\left\{ \frac{k-1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \right\} \qquad (2.12) \\
&\leqslant \left(\frac{q}{1-p} \vee \frac{1-q}{p} \right) \exp\left\{ \frac{k-1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \right\}.
\end{aligned}$$

Consequently, if $c \geqslant \exp\left\{ \frac{k-1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \right\}$, then $\mathcal{P}_c^k\big(\mathsf{GMM}_p(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0; \Sigma)\big)$ contains all those $\mathsf{GMM}_q(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0; \Sigma)$ with mixture probability q such that $1 - c\exp\left\{ \frac{k-1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \right\} p \leqslant q \leqslant c\exp\left\{ \frac{k-1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\intercal \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \right\} (1-p)$. However, since the inequality (2.12) applies the Jensen Inequality to $(\cdot)^k$, the right hand side can be loose when k is large.

Next we proceed to discuss the exponential family in its abstract canonical form. Depending on the growth of the log-partition function, the power divergence might or might not increase with the power k. And consequently when the distributional constant is held fixed, the power uncertainty set $\mathcal{P}_c^k(\mathbb{P})$ might or might not vanish.

Example 2.4 (Canonical Exponential Family Distributional Ball)**.** Consider a canonical parameterized exponential family with density as $f(\boldsymbol{x}; \boldsymbol{\eta}) = h(\boldsymbol{x}) \exp(\langle \boldsymbol{\eta}, \boldsymbol{x} \rangle - A(\boldsymbol{\eta}))$ where $\boldsymbol{\eta} \in \mathbb{R}^p$ is the canonical parameter, $A(\boldsymbol{\eta}) = \log \int h(\boldsymbol{x}) e^{\langle \boldsymbol{\eta}, \boldsymbol{x} \rangle} \mathrm{d}\boldsymbol{x}$ is the log-partition function. Note that $A(\cdot + \boldsymbol{\eta}_0) - A(\boldsymbol{\eta}_0)$ is the logarithm of the moment generating function of the sufficient statistic. Then for fixed $\boldsymbol{\eta}_1, \boldsymbol{\eta}_0 \in \mathbb{R}^p$,

$$\begin{aligned}
\left\| \frac{f(\cdot; \boldsymbol{\eta}_1)}{f(\cdot; \boldsymbol{\eta}_0)} \right\|_{L^k(\boldsymbol{\eta}_0)}^k &= e^{-k[A(\boldsymbol{\eta}_1) - A(\boldsymbol{\eta}_0)] - A(\boldsymbol{\eta}_0)} \int h(\boldsymbol{x}) e^{\langle k(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0) + \boldsymbol{\eta}_0, \boldsymbol{x} \rangle} \mathrm{d}\boldsymbol{x} \\
&= \exp\Big(A[k(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0) + \boldsymbol{\eta}_0] - k[A(\boldsymbol{\eta}_1) - A(\boldsymbol{\eta}_0)] - A(\boldsymbol{\eta}_0) \Big). \qquad (2.13)
\end{aligned}$$

Note that the relationship of (2.13) and k depends on the functional form of the log-partition function $A(\cdot)$. In Example 2.2, $A(\boldsymbol{\eta}) = \boldsymbol{\eta}^\intercal \Sigma \boldsymbol{\eta} + \log\det(\Sigma)$ is a quadratic function in the scaled mean vector $\boldsymbol{\eta} = \Sigma^{-1}\boldsymbol{\mu}$ as the canonical parameter (where the covariance matrix Σ is assumed known and fixed), and hence (2.13) is a quadratic function in k in the exponential, which coincides with the conclusion from Example 2.2 that the L^k-norm of the density ratio is exponentially linear in k. In Example 2.1, the partition function $A(\eta) = \log(1+e^\eta) = \eta + \log(1+e^{-\eta})$ is at most linear in the log-odd $\eta = \log\left(\frac{p}{1-p}\right)$ as the canonical parameter. Then the L^k-norm of the density ratio should be bounded when k varies.

In general, the L^k-norm of the density ratio of distributions from the exponential family increases with k if A is super-linear: $\frac{A(\boldsymbol{\eta})}{\|\boldsymbol{\eta}\|} \to +\infty$ as $\|\boldsymbol{\eta}\| \to +\infty$.

2.7.2 Implementation Details

To practically optimize the DR-ITR ψ-risk $\mathcal{R}^k_{c,\psi}(f)$ based on the empirical data, we first estimate the CTE function $\widehat{C}_n(\cdot)$ using flexible nonparametric techniques. Then we replace the CTE function $C(\cdot)$ by its estimate $\widehat{C}_n(\cdot)$, and the population expectation $\mathbb{E}$ by its empirical version $\mathbb{E}_n$. We solve the following joint minimization problem based on the training data:

$$\min_{f\in\mathcal{F},\eta}\left\{c\left[\mathbb{E}_n\left([\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[f(\boldsymbol{X})]}{2}+[-\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[-f(\boldsymbol{X})]}{2}\right)\right]^{1/k^\star}+\eta\right\}.$$

In this section, we discuss more implementation details of $k<+\infty$ and $k=+\infty$.

2.7.2.1 Optimization when $k < +\infty$

When $k<+\infty$ and $k^\star>1$, the $k^\star$-moment makes the direct optimization more challenging. To reduce the power $1/k^\star$, we introduce the auxiliary variable $\lambda \geqslant 0$ and consider $(\cdot)^{1/k^\star} = \inf_{\lambda\geqslant 0}\left(\frac{(\cdot)}{k^\star\lambda^{k^\star-1}}+\frac{\lambda}{k}\right)$, where due to the AM-GM Inequality, $\frac{1}{k^\star}\left(\frac{(\cdot)}{\lambda^{k^\star-1}}+\underbrace{\lambda+\cdots+\lambda}_{k^\star-1}\right)\geqslant(\cdot)^{1/k^\star}$ with equality if and only if $\lambda=(\cdot)^{1/k^\star}>0$. Then we consider the following joint objective to minimize:

$$L(f,\eta,\lambda) := \frac{c}{k^\star\lambda^{k^\star-1}}\mathbb{E}_n\left([\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[f(\boldsymbol{X})]}{2}+[-\widehat{C}_n(\boldsymbol{X})-\eta]_+^{k^\star}\frac{\psi[-f(\boldsymbol{X})]}{2}\right)+\frac{c\lambda}{k}+\eta. \tag{2.14}$$

Note that the joint objective (2.14) as multiple sum-products of DC functions is difference-of-convex in (f, η, λ), but the DC representation can be messy. Instead of using a direct DC algorithm, we apply the BSUM algorithm (Razaviyayn et al., 2013) to alternatively optimize over (η, λ) and f respectively, where the the upper-bound of the objective in f is a convex majorant. Specifically, we fix a small $\epsilon > 0$ and alternatively implement the following two steps:

Step I: For fixed $\widehat{f}_t$, we implement the t-th step optimization of $(\widehat{\eta}_t, \widehat{\lambda}_t)$ by solving

$$\begin{cases} \widehat{\eta}_t \in \operatorname*{argmin}_{\eta \in \mathbb{R}} \left\{ c \left[\mathbb{E}_n \left(\frac{\psi[\widehat{f}_t(\boldsymbol{X})]}{2} [\widehat{C}_n(\boldsymbol{X}) - \eta]_+^{k^\star} + \frac{\psi[-\widehat{f}_t(\boldsymbol{X})]}{2} [-\widehat{C}_n(\boldsymbol{X}) - \eta]_+^{k^\star} \right) \right]^{1/k^\star} + \eta \right\} \\ \widehat{\lambda}_t := \left[\mathbb{E}_n \left(\frac{\psi[\widehat{f}_t(\boldsymbol{X})]}{2} [\widehat{C}_n(\boldsymbol{X}) - \widehat{\eta}_t]_+^{k^\star} + \frac{\psi[-\widehat{f}_t(\boldsymbol{X})]}{2} [-\widehat{C}_n(\boldsymbol{X}) - \widehat{\eta}_t]_+^{k^\star} \right) \right]^{1/k^\star} \vee \underline{\lambda} \end{cases}. \quad (2.15)$$

The objective in η is univariate and continuously differentiable and can be minimized by any univariate solver. The $\underline{\lambda} > 0$ is a prespecified small constant such that the updated $\widehat{\lambda}_t$ is trimmed at $\underline{\lambda}$ from below for better numerical stability.

Step II: For fixed $(\widehat{f}_t, \widehat{\eta}_t, \widehat{\lambda}_t)$, we solve the $(t+1)$-th step $\widehat{f}_{t+1}$ by minimizing the following convex upper-bound over $\mathcal{F}$:

$$\begin{aligned} \widetilde{L}(f; \widehat{f}_t, \widehat{\eta}_t, \widehat{\lambda}_t) := \mathbb{E}_n \Bigg(& \frac{c}{2k^\star \widehat{\lambda}_t^{k^\star - 1}} [+\widehat{C}_n(\boldsymbol{X}) - \widehat{\eta}_t]_+^{k^\star} \widetilde{\psi}[+f(\boldsymbol{X}); +\widehat{f}_t(\boldsymbol{X})] + \\ & \frac{c}{2k^\star \widehat{\lambda}_t^{k^\star - 1}} [-\widehat{C}_n(\boldsymbol{X}) - \widehat{\eta}_t]_+^{k^\star} \widetilde{\psi}[-f(\boldsymbol{X}); -\widehat{f}_t(\boldsymbol{X})] \Bigg), \end{aligned}$$

where given $u_0 \in \mathbb{R}$, $\widetilde{\psi}(\cdot; u_0)$ is a first-order convex majorant of ψ expanded at u_0:

$$\widetilde{\psi}(u; u_0) := \psi_+(u) - \psi_-(u_0) - \psi'_-(u_0)(u - u_0); \quad u \in \mathbb{R}.$$

In particular for fixed u_0, $\widetilde{\psi}$ satisfies: 1) the majorization $\widetilde{\psi}(u; u_0) \geqslant \psi(u)$ with equality if $u = u_0$; 2) the convexity of $\widetilde{\psi}(\cdot; u_0)$; and 3) the first-order condition $\widetilde{\psi}'(u; u_0) = \psi'_+(u) - \psi'_-(u_0)$ and $\widetilde{\psi}'(u_0; u_0) = \psi'(u_0)$, where $\widetilde{\psi}'(u; u_0)$ is taken over u. To organize the computation, define

$$Z_t^{(\pm)} := \frac{c}{2k^\star \widehat{\lambda}_t^{k^\star - 1}} [\pm \widehat{C}_n(\boldsymbol{X}) - \widehat{\eta}_t]_+^{k^\star}; \quad S_t := Z_t^{(+)} \psi'_-[+\widehat{f}_t(\boldsymbol{X})] - Z_t^{(-)} \psi'_-[-\widehat{f}_t(\boldsymbol{X})]. \quad (2.16)$$

Then at the t-th step, we only need to keep track of $Z_t^{(\pm)}, S_t$ and minimize

$$\widetilde{L}(f; Z_t^{(\pm)}, S_t) := \mathbb{E}_n \left(Z_t^{(+)} \psi_+[+f(\boldsymbol{X})] + Z_t^{(-)} \psi_+[-f(\boldsymbol{X})] - S_t \times f(\boldsymbol{X}) \right), \tag{2.17}$$

over $\mathcal{F}$. We summarize the algorithm for learning the DR-ITR when $k < +\infty$ in Algorithm 2.1.

Algorithm 2.1: Learning the DR-ITR ($k < +\infty$)

Input: Data $\{\boldsymbol{X}_i, \widehat{C}_n(\boldsymbol{X}_i)\}_{i=1}^n$, initial $\widehat{f}_0 \in \mathcal{F}$, $c \geqslant 1$, $\underline{\lambda} > 0$, and tolerance $\epsilon_{\text{tol}} > 0$.
Repeat for $t = 0, 1, \cdots$, do until $|\widehat{f}_{t+1} - \widehat{f}_t| \leqslant (|\widehat{f}_t| \vee 1)\epsilon_{\text{tol}}$:
    Solve $(\widehat{\eta}_t, \widehat{\lambda}_t)$ by (2.15);
    Update $(Z_t^{(\pm)}, S_t)$ as in (2.16);
    Solve $\widehat{f}_{t+1}$ by optimizing the objective $\widetilde{L}(\cdot; Z_t^{(\pm)}, S_t)$ as in (2.17);
Output: $\widehat{f}_{t+1}$.

2.7.2.2 Optimization when $k = +\infty$

For $k = +\infty$ and $c > 1$, it is possible that the BSUM algorithm introduced in Algorithm 2.1 suffers potential convergence problems when the minimizer $\widehat{\eta}_t$ given $\widehat{f}_t$ in (2.15) is non-unique. Following Qi et al. (2019b, Proposition 3.1), we see that the joint objective is minimized with respect to η at one of the $2n$ knots $\{\eta_j^*\}_{j=1}^{2n} := \{\pm \widehat{C}_n(\boldsymbol{X}_i)\}_{i=1}^n$. Then the joint minimization problem boils down to

$$\min_{f \in \mathcal{F}} \min_{1 \leqslant j \leqslant 2n} \left\{ L_j(f) := \frac{c}{2} \mathbb{E}_n \left([\widehat{C}_n(\boldsymbol{X}) - \eta_j^*]_+ \psi[+f(\boldsymbol{X})] + [-\widehat{C}_n(\boldsymbol{X}) - \eta_j^*]_+ \psi[-f(\boldsymbol{X})] \right) + \eta_j^* \right\}.$$

That is, the minimization with respect to η can attain at only finitely many candidates $\{\eta_j^*\}_{j=1}^{2n}$. For $1 \leqslant j \leqslant 2n$, we define the convex upper bound of L_j at f_0 as

$$\widetilde{L}_j(f; f_0) := \mathbb{E}_n \left(\frac{c}{2} [+\widehat{C}_n(\boldsymbol{X}) - \eta_j^*]_+ \widetilde{\psi}[+f(\boldsymbol{X}); +f_0(\boldsymbol{X})] + \frac{c}{2} [-\widehat{C}_n(\boldsymbol{X}) - \eta_j^*]_+ \widetilde{\psi}[-f(\boldsymbol{X}); -f_0(\boldsymbol{X})] \right),$$

where $\widetilde{\psi}$ is the first-order convex majorant of ψ as before. Then the previously discussed BSUM algorithm iteratively updates the following two steps: (I) for fixed $\widehat{f}_t$, solve for the t-th step $\widehat{j}_t \in \operatorname{argmin}_{1 \leqslant j \leqslant 2n} L_j(\widehat{f}_t)$; (II) for fixed $(\widehat{f}_t, \widehat{j}_t)$, solve for the $(t+1)$-th step $\widehat{f}_{t+1}$ by minimizing $\widetilde{L}_{\widehat{j}_t}(\cdot; \widehat{f}_t)$. Notice that the non-uniqueness of the minimizer $\widehat{\eta}_t$ given $\widehat{f}_t$ now becomes the non-uniqueness of the index $\widehat{j}_t$.

To overcome the difficulty due to the non-uniqueness, Pang et al. (2016, Section 5) showed that the following two requirements should be met to ensure the convergence to stationarity: (1) minimizing the surrogate function $\widetilde{L}_{\widehat{j}_t}(\cdot; \widehat{f}_t)$ of the chosen index $\widehat{j}_t$ should let the true objective function L descend the most; (2) the most descent requirement (1) holds with respect to the indices chosen among the following ϵ-argmin index set for some fixed $\epsilon > 0$:

$$\mathscr{M}_\epsilon(\widehat{f}_t) := \left\{ 1 \leqslant j \leqslant 2n : L_j(\widehat{f}_t) \leqslant \min_{1 \leqslant l \leqslant 2n} L_l(\widehat{f}_t) + \epsilon \right\}, \tag{2.18}$$

rather than the traditional argmin index set $\mathscr{M}_0(\widehat{f}_t)$. To avoid too many surrogate functions to be minimized at each step, Pang et al. (2016, Section 5.2) proposed to randomly choose $\widehat{j}_t \in \mathscr{M}_\epsilon(\widehat{f}_t)$ with a positive probability, so that at least for some positive chance the most descent index can be picked. To ensure the true objective is strictly decreasing, we accept the minimizer $\widetilde{f}_{t+1} \in \operatorname{argmin}_f \widetilde{L}_{\widehat{j}_t}(f; \widehat{f}_t)$ only when $\widetilde{L}_{\widehat{j}_t}(\widetilde{f}_{t+1}; \widehat{f}_t) \leqslant L(\widehat{f}_t)$, or equivalently,

$$L_{\widehat{j}_t}(\widehat{f}_t) - \min_{1 \leqslant j \leqslant 2n} L_j(\widehat{f}_t) \leqslant \widetilde{L}_{\widehat{j}_t}(\widehat{f}_t; \widehat{f}_t) - \widetilde{L}_{\widehat{j}_t}(\widetilde{f}_{t+1}; \widehat{f}_t).$$

That is, the descent in terms of the surrogate objective $\widetilde{L}_{\widehat{j}_t}(\cdot; \widehat{f}_t)$ is no less than the excess value (up to ϵ) of the chosen $\widehat{j}_t$-th objective $L_{\widehat{j}_t}$ at $\widehat{f}_t$.

To organize the computation, we again define for $1 \leqslant j \leqslant 2n$ and f_0

$$Z_j^{(\pm)} := \frac{c}{2}[\pm \widehat{C}_n(\boldsymbol{X}) - \eta_j^*]_+; \quad S_j(f_0) := Z_j^{(+)} \psi_-'[+f_0(\boldsymbol{X})] - Z_j^{(-)} \psi_-'[-f_0(\boldsymbol{X})], \tag{2.19}$$

similarly as in (2.16), but with the index t replaced by j. Then at the t-th step, we first randomly pick $\widehat{j}_t \in \mathscr{M}_\epsilon(\widehat{f}_t)$ uniformly and keep the excess value $\epsilon_t := L_{\widehat{j}_t}(\widehat{f}_t) - \min_{1 \leqslant j \leqslant 2n} L_j(\widehat{f}_t)$. Then we keep $Z_t := Z_{\widehat{j}_t}^{(\pm)}$ and $S_t := S_{\widehat{j}_t}(\widehat{f}_t)$ and minimize $\widetilde{L}(\cdot; Z_t^{(\pm)}, S_t)$ as in (2.17). Finally, we accept the minimizer $\widetilde{f}_{t+1} \in \operatorname{argmin}_f \widetilde{L}(f; Z_t^{(\pm)}, S_t)$ if $\widetilde{L}(\widetilde{f}_t; Z_t^{(\pm)}, S_t) - \widetilde{L}(\widetilde{f}_{t+1}; Z_t^{(\pm)}, S_t) \geqslant \epsilon_t$. We summarize the algorithm for learning the DR-ITR when $k = +\infty$ in Algorithm 2.2.

Algorithm 2.2: Learning the DR-ITR ($k = +\infty$)

Input: Data $\{\boldsymbol{X}_i, \widehat{C}_n(\boldsymbol{X}_i)\}_{i=1}^n$, initial $\widehat{f}_0 \in \mathcal{F}$, $c > 1$, $\epsilon > 0$, and tolerance $\epsilon_{\text{tol}} > 0$.
For $t = 0, 1, \cdots$, do until $\|\widehat{f}_{t+1} - \widehat{f}_t\| \leqslant (\|\widehat{f}_t\| \vee 1)\epsilon_{\text{tol}}$:
  Choose $\widehat{j}_t \in \mathscr{M}_\epsilon(\widehat{f}_t)$ in (2.18) uniformly and randomly, and keep $\epsilon_t := L_{\widehat{j}_t}(\widehat{f}_t) - \min_{1 \leqslant j \leqslant 2n} L(\widehat{f}_t)$;
  Update $Z_t^{(\pm)} = Z_{\widehat{j}_t}^{(\pm)}$ and $S_t = S_{\widehat{j}_t}(\widehat{f}_t)$ as in (2.19);
  Solve $\widetilde{f}_{t+1}$ by optimizing the objective $\widetilde{L}(\cdot; Z_t^{(\pm)}, S_t)$ as in (2.17);
  If $\widetilde{L}(\widehat{f}_t; Z_t^{(\pm)}, S_t) - \widetilde{L}(\widetilde{f}_{t+1}; Z_t^{(\pm)}, S_t) \geqslant \epsilon_t$, then set $\widehat{f}_{t+1} = \widetilde{f}_{t+1}$; otherwise, set $\widehat{f}_{t+1} = \widehat{f}_t$.
Output: $\widehat{f}_{t+1}$.

2.7.3 Technical Proofs

2.7.3.1 Proof of Lemma 2.2

(I) follows from direct calculation. Now we admit (I) and prove (II). First notice that

$$\lambda\phi_k^\star(z/\lambda) = \frac{(k-1)^{k^\star}/k}{\lambda^{1/(k-1)}}\left(z - \eta + \frac{\lambda}{k-1}\right)_+^{k^\star} - \frac{\lambda}{k},$$
$$\nabla\phi_k^\star(z/\lambda) = \frac{(k-1)^{1/(k-1)}}{\lambda^{1/(k-1)}}\left(z - \eta + \frac{\lambda}{k-1}\right)_+^{1/(k-1)}.$$

Now using the (2.6)-R.H.S., the Cressie-Read family defining worst-case expectation is further solved by

$$\min_{\lambda \geqslant 0, \eta \in \mathbb{R}} \ [(k-1)^{k^\star}/k] \times \lambda^{-1/(k-1)}\mathbb{E}_{\mathbb{P}}\left(Z - \eta + \tfrac{\lambda}{k-1}\right)_+^{k^\star} + \lambda\left(\rho - \tfrac{1}{k}\right) + \eta, \quad W^* = \tfrac{(k-1)^{1/(k-1)}}{(\lambda^*)^{1/(k-1)}}\left(Z - \eta^* + \tfrac{\lambda^*}{k-1}\right)_+^{1/(k-1)},$$
$$\Leftrightarrow \min_{\lambda \geqslant 0, \eta \in \mathbb{R}} \ [(k-1)^{k^\star}/k] \times \lambda^{-1/(k-1)}\mathbb{E}_{\mathbb{P}}(Z - \eta)_+^{k^\star} + \lambda\left(\rho + \tfrac{1}{k(k-1)}\right) + \eta, \qquad W^* = \tfrac{(k-1)^{1/(k-1)}}{(\lambda^*)^{1/(k-1)}}(Z - \eta^*)_+^{1/(k-1)}.$$

where (λ, η) can be optimized stagewise.

Fix $\eta \in \mathbb{R}$.

$$\begin{aligned}
& [(k-1)^{k^\star}/k] \times \lambda^{-1/(k-1)}\mathbb{E}_{\mathbb{P}}(Z - \eta)_+^{k^\star} + \lambda\left(\rho + \frac{1}{k(k-1)}\right) \\
= \ & (k-1) \times \lambda^{-1/(k-1)}\left(\frac{(k-1)^{1/(k-1)}}{k}\right)\mathbb{E}_{\mathbb{P}}(Z - \eta)_+^{k^\star} + \lambda\left(\rho + \frac{1}{k(k-1)}\right) \\
\geqslant \ & k\left[\left(\frac{(k-1)^{1/(k-1)}}{k}\right)^{k-1}[\mathbb{E}_{\mathbb{P}}(Z - \eta)_+^{k^\star}]^{k-1}\left(\rho + \frac{1}{k(k-1)}\right)\right]^{1/k} \qquad \text{(by AM-GM Inequality)} \\
= \ & [k(k-1)\rho + 1]^{1/k}[\mathbb{E}_{\mathbb{P}}(Z - \eta)_+^{k^\star}]^{1/k^\star}.
\end{aligned}$$

Denote $c_k(\rho) := [k(k-1)\rho + 1]^{1/k}$. Then the objective in η becomes

$$\min_{\eta \in \mathbb{R}} \quad c_k(\rho)[\mathbb{E}_{\mathbb{P}}(Z-\eta)_+^{k^\star}]^{1/k^\star} + \eta.$$

2.7.3.2 Proof of Proposition 2.4

Define

$$\widetilde{C}_{\eta,\lambda}^{(\pm)}(\boldsymbol{X}) := \frac{c}{k^\star \lambda^{k^\star - 1}} \mathbb{E}\left([\pm C(\boldsymbol{X}) - \eta]_+^{k^\star} \Big| \boldsymbol{X} \right), \quad \widetilde{C}_{\eta,\lambda}(\boldsymbol{X}) := \widetilde{C}_{\eta,\lambda}^{(+)}(\boldsymbol{X}) - \widetilde{C}_{\eta,\lambda}^{(-)}(\boldsymbol{X}).$$

Then by conditioning on $\boldsymbol{X}$,

$$\begin{aligned}
\mathcal{L}_c^k(f,\eta,\lambda) &= \mathbb{E}\left(\widetilde{C}_{\eta,\lambda}^{(+)}(\boldsymbol{X}) \mathbb{1}[f(\boldsymbol{X}) < 0] + \widetilde{C}_{\eta,\lambda}^{(-)}(\boldsymbol{X}) \mathbb{1}[f(\boldsymbol{X}) > 0] \right) + \frac{c\lambda}{k} + \eta, \\
\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) &= \mathbb{E}\left(\widetilde{C}_{\eta,\lambda}^{(+)}(\boldsymbol{X}) \frac{\psi[f(\boldsymbol{X})]}{2} + \widetilde{C}_{\eta,\lambda}^{(-)}(\boldsymbol{X}) \frac{\psi[-f(\boldsymbol{X})]}{2} \right) + \frac{c\lambda}{k} + \eta.
\end{aligned}$$

(I) (Fisher Consistency) Notice that for our ramp surrogate loss ψ, $f \geqslant 1$ implies that $\frac{\psi(f)}{2} = 0$, and $f \leqslant -1$ implies that $\frac{\psi(f)}{2} = 1$. Then without loss of generality, we might restrict to consider $f \in [-1,1]$ for which $f = 1$ if and only if $\frac{\psi(f)}{2} = 0$ and $f = -1$ if and only if $\frac{\psi(f)}{2} = 1$. Then for fixed $\boldsymbol{x} \in \mathcal{X}$,

$$\begin{aligned}
&\min_{f \in \{\pm 1\}} \left\{ \widetilde{C}_{\eta,\lambda}^{(+)}(\boldsymbol{x}) \mathbb{1}(f < 0) + \widetilde{C}_{\eta,\lambda}^{(-)}(\boldsymbol{x}) \mathbb{1}(f > 0) \right\} \\
=& \min_{f \in [-1,1]} \left\{ \widetilde{C}_{\eta,\lambda}^{(+)}(\boldsymbol{x}) \frac{\psi(f)}{2} + \widetilde{C}_{\eta,\lambda}^{(-)}(\boldsymbol{x}) \frac{\psi(-f)}{2} \right\} \\
=& \widetilde{C}_{\eta,\lambda}^{(+)}(\boldsymbol{x}) \wedge \widetilde{C}_{\eta,\lambda}^{(-)}(\boldsymbol{x})
\end{aligned}$$

attained at the common function value $f_{\eta,\lambda}^*(\boldsymbol{x}) := \mathbf{sign}[\widetilde{C}_{\eta,\lambda}(\boldsymbol{x})]$. Define $\mathcal{L}_c^{k,*}(\eta,\lambda) := \mathcal{L}_c^k(f_{\eta,\lambda}^*,\eta,\lambda) = \mathbb{E}[\widetilde{C}_{\eta,\lambda}^{(+)}(\boldsymbol{X}) \wedge \widetilde{C}_{\eta,\lambda}^{(-)}(\boldsymbol{X})] + \frac{c\lambda}{k} + \eta$. Then

$$\begin{aligned}
\min_{f:\mathcal{X}\to\{\pm 1\}} \mathcal{L}_c^k(f,\eta,\lambda) &= \min_{f:\mathcal{X}\to[-1,1]} \mathcal{L}_{c,\psi}^k(f,\eta,\lambda) = \mathcal{L}_c^{k,*}(\eta,\lambda), \\
\operatorname*{argmin}_{f:\mathcal{X}\to\{\pm 1\}} \mathcal{L}_c^k(f,\eta,\lambda) &= \operatorname*{argmin}_{f:\mathcal{X}\to[-1,1]} \mathcal{L}_{c,\psi}^k(f,\eta,\lambda) = f_{\eta,\lambda}^*(\boldsymbol{X}) \quad a.s.
\end{aligned}$$

(II) (Excess Risk) For fixed $f : \mathcal{X} \to \mathbb{R}$,

$$\mathcal{L}_c^k(f,\eta,\lambda) - \mathcal{L}_c^{k,*}(\eta,\lambda) = \mathbb{E}\left[\widetilde{C}_{\eta,\lambda}(\boldsymbol{X}) \times \left(\mathbb{1}[f(\boldsymbol{X}) < 0] - \mathbb{1}[f^*_{\eta,\lambda}(\boldsymbol{X}) < 0]\right) + \widetilde{C}^{(-)}_{\eta,\lambda}(\boldsymbol{X})\right],$$
$$\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) - \mathcal{L}_c^{k,*}(\eta,\lambda) = \mathbb{E}\left[\widetilde{C}_{\eta,\lambda}(\boldsymbol{X}) \times \frac{\psi[f(\boldsymbol{X})] - \psi[f^*_{\eta,\lambda}(\boldsymbol{X})]}{2} + \widetilde{C}^{(-)}_{\eta,\lambda}(\boldsymbol{X})\right],$$

where the second equation follows from the fact that $\psi(u) + \psi(-u) \equiv 2$. For fixed $\boldsymbol{x} \in \mathcal{X}$, if $\widetilde{C}_{\eta,\lambda}(\boldsymbol{x}) > 0$, then $f^*_{\eta,\lambda}(\boldsymbol{x}) = 1$, and

$$\mathbb{1}[f(\boldsymbol{x}) < 0] - \mathbb{1}[f^*_{\eta,\lambda}(\boldsymbol{x}) < 0] = \mathbb{1}[f(\boldsymbol{x}) < 0] \leqslant 2 \times \frac{\psi[f(\boldsymbol{x})]}{2} = 2 \times \frac{\psi[f(\boldsymbol{x})] - \psi[f^*_{\eta,\lambda}(\boldsymbol{x})]}{2};$$

otherwise if $\widetilde{C}_{\eta,\lambda}(\boldsymbol{x}) < 0$, then $f^*_{\eta,\lambda}(\boldsymbol{x}) = -1$, and

$$2 \times \frac{\psi[f(\boldsymbol{x})] - \psi[f^*_{\eta,\lambda}(\boldsymbol{x})]}{2} = -\psi[-f(\boldsymbol{x})] \leqslant -\mathbb{1}[-f(\boldsymbol{x}) \leqslant 0] = \mathbb{1}[f(\boldsymbol{x}) < 0] - \mathbb{1}[f^*_{\eta,\lambda}(\boldsymbol{x}) < 0].$$

Therefore,

$$\mathcal{L}_c^k(f,\eta,\lambda) - \mathcal{L}_c^{k,*}(\eta,\lambda) \leqslant 2[\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) - \mathcal{L}_c^{k,*}(\eta,\lambda)].$$

Finally, by rearranging $\mathcal{L}_c^{k^\star}(\eta,\lambda)$ to the same side and infimizing its $(\eta,\lambda) \in \mathbb{R} \times \mathbb{R}_+$, we have

$$\begin{aligned} & \mathcal{L}_c^k(f,\eta,\lambda) \leqslant 2\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) - \mathcal{L}_c^{k,*}(\eta,\lambda) \leqslant 2\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) - \mathcal{R}_c^{k,*} \\ \Leftrightarrow \quad & \mathcal{L}_c^k(f,\eta,\lambda) - \mathcal{R}_c^{k,*} \leqslant 2[\mathcal{L}_{c,\psi}^k(f,\eta,\lambda) - \mathcal{R}_c^{k,*}]. \end{aligned}$$

And by partially infimizing $(\eta,\lambda) \in \mathbb{R} \times \mathbb{R}_+$ on both sides, we have

$$\mathcal{R}_c^k(f) - \mathcal{R}_c^{k,*} \leqslant 2[\mathcal{R}_{c,\psi}^k(f) - \mathcal{R}_c^{k,*}].$$

2.7.3.3 Proof of Proposition 2.5

By Assumption 2.4, without loss of generality, we also assume that Assumptions 2.2 and 2.3 also hold for $\{\widehat{C}_n(\boldsymbol{X})\}_{n\in\mathbb{N}}$ uniformly.

First assume $k < +\infty$ and $k^\star > 1$. We first provide a few boundedenss results implied by Assumption 2.2. For $f : \mathcal{X} \to \mathbb{R}$, define

$$\eta_f^* := \operatorname*{argmin}_{\eta\in\mathbb{R}} \left\{ c\left[\mathbb{E}\left(\frac{\psi[f(\boldsymbol{X})]}{2}[C(\boldsymbol{X})-\eta]_+^{k^\star} + \frac{\psi[-f(\boldsymbol{X})]}{2}[-C(\boldsymbol{X})-\eta]_+^{k^\star}\right)\right]^{1/k^\star} + \eta \right\}, \quad (2.20)$$

$$\lambda_f^* := \left[\mathbb{E}\left(\frac{\psi[f(\boldsymbol{X})]}{2}[C(\boldsymbol{X})-\eta_f^*]_+^{k^\star} + \frac{\psi[-f(\boldsymbol{X})]}{2}[-C(\boldsymbol{X})-\eta_f^*]_+^{k^\star}\right)\right]^{1/k^\star}. \quad (2.21)$$

By Assumption 2.2 that $|C(\boldsymbol{X})| \leqslant M$, the optimal objective of (2.20) is bounded from above by $\min_{\eta\in\mathbb{R}}\{c(M-\eta)_+ + \eta\} = M$. And for any fixed $\eta \in \mathbb{R}$, the objective of (2.20) is bounded from below by $c(-M-\eta)_+ + \eta$. Then by the optimality of η_f^*, we have $c(-M-\eta_f^*)_+ + \eta_f^* \leqslant M \Leftrightarrow -\frac{c+1}{c-1}M \leqslant \eta_f^* \leqslant M$.

As for λ_f^*, since the optimal value of (2.20) is $c\lambda_f^* + \eta_f^*$, we have $c\lambda_f^* + \eta_f^* \leqslant M \Rightarrow \lambda_f^* \leqslant \frac{M-\eta_f^*}{c} \leqslant \frac{2M}{c-1}$. On the other hand, we need to elaborate more to give the lower bound (away from 0) on λ_f^*. The following lemma is a useful tool to motivate our analysis.

Lemma 2.7. *Suppose Z is a bounded random variable, $k \geqslant 1$, $c \geqslant 1$. Define*

$$\eta^* := \operatorname*{argmin}_{\eta\in\mathbb{R}} \left\{ c[\mathbb{E}(Z-\eta)_+^k]^{1/k} + \eta \right\}.$$

Then $\mathbb{P}(Z \geqslant \eta^*) \geqslant c^{-k}$.

Proof. For $k = 1$, η^* as the VaR (Krokhmal, 2007) can be obtained by $\eta^* = \inf\{\eta \in \mathbb{R} : \mathbb{P}(Z \leqslant \eta) \geqslant 1 - c^{-1}\}$. Then for any $\epsilon > 0$, $\mathbb{P}(Z \leqslant \eta - \epsilon) < 1 - c^{-1} \Leftrightarrow \mathbb{P}(Z > \eta - \epsilon) \geqslant c^{-1}$. Let $\epsilon \to 0^+$ and by upper semi-continuity, we have $\mathbb{P}(Z \geqslant \eta^*) \geqslant c^{-1}$.

Suppose $k > 1$. If $\mathbb{P}(Z = \operatorname{ess.sup} Z) \geqslant c^{-k}$, then by $\eta^* \leqslant \operatorname{ess.sup} Z$, $\mathbb{P}(Z \geqslant \eta^*) \geqslant \mathbb{P}(Z = \operatorname{ess.sup} Z) \geqslant c^{-k}$ holds trivially. Now assume $\mathbb{P}(Z = \operatorname{ess.sup} Z) < c^{-k}$. By lower semi-continuity, there exists $\epsilon_0 > 0$, such that for any $0 \leqslant \epsilon \leqslant \epsilon_0$, $\mathbb{P}(Z \geqslant \operatorname{ess.sup} Z - \epsilon) < c^{-k}$. Then

$$c[\mathbb{E}(Z - \operatorname{ess.sup} Z + \epsilon)_+^k]^{1/k} + \operatorname{ess.sup} Z - \epsilon \leqslant c\epsilon\mathbb{P}(Z \geqslant \operatorname{ess.sup} - \epsilon)^{1/k} + \operatorname{ess.sup} Z - \epsilon < \operatorname{ess.sup} Z.$$

As a result, $\eta^* < \operatorname{ess.sup} Z - \epsilon_0$, hence

$$\mathbb{E}(Z-\eta^*)_+^k \geqslant (\epsilon_0/2)^k\mathbb{P}(Z \geqslant \eta^* + \epsilon_0/2) \geqslant (\epsilon_0/2)^k\mathbb{P}(Z \geqslant \operatorname{ess.sup} Z - \epsilon_0/2) > 0.$$

Finally, the first order condition for η^* is given by

$$-\frac{c\mathbb{E}(Z-\eta^*)_+^{k-1}}{\mathbb{E}[(Z-\eta^*)_+^k]^{1-1/k}}+1=0 \quad \Leftrightarrow \quad \frac{\|(Z-\eta^*)_+\|_{L^{k-1}}}{\|(Z-\eta^*)_+\|_{L^k}}=c^{-\frac{1}{k-1}}.$$

On the other hand, by Hölder Inequality,

$$\mathbb{E}(Z-\eta^*)_+^{k-1}=\mathbb{E}[(Z-\eta^*)_+^{k-1}\mathbb{1}(Z\geqslant\eta^*)]\leqslant[\mathbb{E}(Z-\eta^*)_+^k]^{\frac{k-1}{k}}\mathbb{P}(Z\geqslant\eta^*)^{1/k}.$$

We have

$$c^{-\frac{1}{k-1}}=\frac{\|(Z-\eta^*)_+\|_{L^{k-1}}}{\|(Z-\eta^*)_+\|_{L^k}}\leqslant\mathbb{P}(Z\geqslant\eta^*)^{1/[k(k-1)]} \quad \Leftrightarrow \quad \mathbb{P}(Z\geqslant\eta^*)\geqslant c^{-k}.$$

□

Next, we introduce the sign variable $\zeta_\psi(f)\in\{\pm1\}$ such that $\mathbb{P}[\zeta_\psi(f)=\pm1|\boldsymbol{X}]=\frac{\psi[\pm f(\boldsymbol{X})]}{2}$. Then $\eta_f^*\in\operatorname{argmin}_{\eta\in\mathbb{R}}\left\{c\left(\mathbb{E}[C(\boldsymbol{X})\zeta_\psi(f)-\eta]_+^{k^\star}\right)^{1/k^\star}+\eta\right\}$. By Lemma 2.7, we immediately have $\mathbb{P}[C(\boldsymbol{X})\zeta_\psi(f)\geqslant\eta_f^*]\geqslant c^{-k}$. Next by Assumption 2.3, $C(\boldsymbol{X})$ has uniformly bounded density h with respect to the Lebesgue measure. Then $C(\boldsymbol{X})\zeta_\psi(f)$ also has density $h_{\psi,f}(c)\leqslant h(c)\vee h(-c)$ with respect to the Lebesgue measure, and $h_{\psi,f}$ is uniformly bounded as well: $\|h_{\psi,f}\|_\infty\leqslant\|h\|_\infty<+\infty$. Then for any fixed $\underline{c}\leqslant\bar{c}$, we have $\mathbb{P}\{C(\boldsymbol{X})\zeta_\psi(f)\in[\underline{c},\bar{c}]\}\leqslant(\bar{c}-\underline{c})\|h\|_\infty$. In particular, for any $t>0$,

$$\mathbb{P}[C(\boldsymbol{X})\zeta_\psi(f)\geqslant\eta_f^*+t]\geqslant c^{-k}-t\|h\|_\infty.$$

In particular, by taking $t:=1/(2\|h\|_\infty c^k)$, we have

$$\begin{aligned}
\lambda_f^* &= \left(\mathbb{E}[C(\boldsymbol{X})\zeta_\psi(f)-\eta_f^*]_+^{k^\star}\right)^{1/k^\star} \\
&\geqslant 1/(2\|h\|_\infty c^k)\mathbb{P}[C(\boldsymbol{X})\zeta_\psi(f)\geqslant\eta_f^*+1/(2\|h\|_\infty c^k)]^{1/k^\star} \\
&\geqslant 1/(2\|h\|_\infty c^k)[c^{-k}-1/(2\|h\|_\infty c^k)\times\|h\|_\infty]^{1/k^\star} \\
&= 1/(2^{(2k-1)/k}\|h\|_\infty c^{2k-1})>0,
\end{aligned}$$

which decreases in c of order $c^{-(2k-1)}$. Note that the lower bound on λ_f^* depends on the order k, while its upper bound doesn't. In particular, as k increases, the vanishing rate of λ_f^* as $c \to +\infty$ gets faster.

We conclude the preceding boundedness results by denoting $\eta_f^* \in [\underline{\eta}, \bar{\eta}] := \big[-\frac{c+1}{c-1}M, M \big]$ and $\lambda_f^* \in [\underline{\lambda}, \bar{\lambda}] := \big[\frac{1}{2^{(2k-1)/k}\|h\|_\infty c^{2k-1}}, \frac{2M}{c-1} \big]$. Note that all those bounds above also hold when $\mathbb{E}$ is replaced by $\mathbb{E}_n$, $\frac{\psi(\cdot)}{2}$ is replaced by $\mathbb{1}(\cdot < 0)$, and $C(\cdot)$ is replaced by $\widehat{C}_n(\cdot)$. As an immediate result, we further have boundedness $\ell_c^k, \ell_{c,\psi}^k \in [\underline{\ell}_c^k, \bar{\ell}_c^k]$ where $\underline{\ell}_c^k := \frac{c}{k}\underline{\lambda} + \underline{\eta}$, and

$$
\begin{aligned}
\bar{\ell}_c^k &:= \max_{(\eta,\lambda) \in \{\underline{\eta}, \bar{\eta}\} \times \{\underline{\lambda}, \bar{\lambda}\}} \left\{ \frac{c}{k^\star \lambda^{k^\star - 1}} (M - \eta)^{k^\star} + \frac{c\lambda}{k} + \eta \right\} \\
&= \max \begin{cases} \frac{2}{k}\frac{c}{c-1}M + M, & \eta = \bar{\eta},\ \lambda = \bar{\lambda}; \\ \frac{2^{2k^\star - 1/k^\star}\|h\|_\infty^{k^\star - 1}}{k^\star} \frac{c^{2k^\star+2}}{(c-1)^{k^\star}} M^{k^\star} + \frac{1}{k 2^{1+1/k^\star}\|h\|_\infty} \frac{1}{c^{2/(k^\star-1)}} - \frac{c+1}{c-1}M, & \eta = \underline{\eta},\ \lambda = \underline{\lambda}; \\ \frac{2}{k^\star}\frac{c^{k^\star+1}}{c-1}M + \frac{2}{k}\frac{c}{c-1}M - \frac{c+1}{c-1}M, & \eta = \underline{\eta},\ \lambda = \bar{\lambda}. \end{cases}
\end{aligned}
$$

Notice that as $c, (c-1)^{-1}, M \to +\infty$, the leading order term is $\mathcal{O}\left(\frac{c^{2k^\star+2}}{(c-1)^{k^\star}} M^{k^\star} \right)$. To conclude all boundedness results, we introduce the joint parameter space

$$
\theta := (f, \eta, \lambda) \in \Theta_n := \mathcal{F}_n \times \Pi_n \times \Lambda_n,
$$

where $\mathcal{F}_n := \{ f \in \mathcal{F} : \|f\|_{\mathcal{F}} \leqslant \gamma_n \}$, $\Pi_n := [\underline{\eta}, \bar{\eta}]$ and $\Lambda_n := [\underline{\lambda}, \bar{\lambda}]$. Moreover, we have

$$
\left| \ell_c^k(\theta; \widehat{C}_n) - \ell_c^k(\theta; C) \right| \leqslant \underbrace{\frac{2c}{\underline{\lambda}^{k^\star - 1}} (M - \underline{\eta})^{k^\star - 1}}_{L_C} \left| \widehat{C}_n(\boldsymbol{X}) - C(\boldsymbol{X}) \right|,
$$

and

$$
\left| \ell_{c,\psi}^k(\theta; \widehat{C}_n) - \ell_c^k(\theta; C) \right| \leqslant L_C \left| \widehat{C}_n(\boldsymbol{X}) - C(\boldsymbol{X}) \right|.
$$

In particular, $L_C = \frac{2^{2k^\star - 1/k^\star}\|h\|_\infty^{k^\star - 1}}{k^\star} \frac{c^{2k^\star+1}}{(c-1)^{k^\star - 1}} M^{k^\star - 1}$.

Next, we begin to prove the regret bound. Recall that the empirical minimizer is $\widehat{\theta}_n := (\widehat{f}_n, \widehat{\eta}_n, \widehat{\lambda}_n) \in \operatorname*{argmin}_{(f,\eta,\lambda) \in \Theta_n} \mathbb{E}_n \ell_{c,\psi}^k(f, \eta, \lambda; \widehat{C}_n)$ where the distributions of (η, λ) can be constrained to $\Pi_n \times \Lambda_n = [\underline{\eta}, \bar{\eta}] \times [\underline{\lambda}, \bar{\lambda}]$ due to the previous boundedness. We also define the within-Θ_n oracle

$\theta^*_\gamma := (f^*_\gamma, \eta^*_\gamma, \lambda^*_\gamma) \in \underset{(f,\eta,\lambda)\in\Theta_n}{\operatorname{argmin}} \mathcal{L}^k_{c,\psi}(f,\eta,\lambda)$. Then, by definition, we have $\mathcal{L}^k_{c,\psi}(\theta^*_\gamma) - \mathcal{R}^{k,*}_c = \mathcal{A}^k_c(\gamma_n)$. By Proposition 2.4, we have

$$
\begin{aligned}
&\mathcal{L}^k_c(\widehat{\theta}_n) - \mathcal{R}^{k,*}_c \\
\leqslant& 2[\mathcal{L}^k_{c,\psi}(\widehat{\theta}_n) - \mathcal{R}^{k,*}_c] \\
=& 2\left(\mathcal{L}^k_{c,\psi}(\widehat{\theta}_n) - \mathbb{E}_n \ell^k_{c,\psi}(\widehat{\theta}_n; C)\right) + 2\mathbb{E}_n[\ell^k_{c,\psi}(\widehat{\theta}_n; C) - \ell^k_{c,\psi}(\widehat{\theta}_n; \widehat{C}_n)] + 2\Big(\mathbb{E}_n \ell^k_{c,\psi}(\widehat{\theta}_n; \widehat{C}_n) - \mathcal{R}^{k,*}_c\Big) \\
\leqslant& 2 \sup_{\theta\in\Theta_n} (\mathbb{P} - \mathbb{P}_n)\ell^k_{c,\psi}(\theta; C) + 2L_C \|\widehat{C}_n - C\|_\infty + 2\Big(\mathbb{E}_n \ell^k_{c,\psi}(\theta^*_\gamma; \widehat{C}_n) - \mathcal{R}^{k,*}_c\Big) \\
\leqslant& 2 \sup_{\theta\in\Theta_n} (\mathbb{P} - \mathbb{P}_n)\ell^k_{c,\psi}(\theta; C) + 4L_C \|\widehat{C}_n - C\|_\infty + 2\left(\mathbb{E}_n \ell^k_{c,\psi}(\theta^*_\gamma; C) - \mathcal{L}^k_{c,\psi}(\theta^*_\gamma)\right) + 2\mathcal{A}(\gamma_n) \\
\leqslant& 2 \sup_{\theta\in\Theta_n} (\mathbb{P} - \mathbb{P}_n)\ell^k_{c,\psi}(\theta; C) + 4L_C \|\widehat{C}_n - C\|_\infty + 2 \sup_{\theta\in\Theta_n} (\mathbb{P}_n - \mathbb{P})\ell^k_{c,\psi}(\theta; C) + 2\mathcal{A}(\gamma_n).
\end{aligned}
$$

It follows standard routine to propose a Rademacher complexity bound. Fix $\delta > 0$. First by McDiarmid Inequality (Bartlett and Mendelson, 2002, Theorem 9), with probability $\geqslant 1 - \delta$,

$$
\begin{aligned}
\sup_{\theta\in\Theta_n} (\mathbb{P} - \mathbb{P}_n)\ell^k_{c.\psi}(\theta; C) &\leqslant \mathbb{E} \sup_{\theta\in\Theta_n} (\mathbb{P} - \mathbb{P}_n)\ell^k_{c,\psi}(\theta; C) + (\overline{\ell}^k_c - \underline{\ell}^k_c)\sqrt{\frac{\log(2/\delta)}{2n}}, \\
\sup_{\theta\in\Theta_n} (\mathbb{P}_n - \mathbb{P})\ell^k_{c.\psi}(\theta; C) &\leqslant \mathbb{E} \sup_{\theta\in\Theta_n} (\mathbb{P}_n - \mathbb{P})\ell^k_{c,\psi}(\theta; C) + (\overline{\ell}^k_c - \underline{\ell}^k_c)\sqrt{\frac{\log(2/\delta)}{2n}}.
\end{aligned}
$$

Next we define the Rademacher complexity on Θ_n as follows:

$$
\mathsf{R}_n(\Theta_n) := \mathbb{E}_{(\boldsymbol{X},\sigma)\sim\mathbb{P}} \sup_{\theta\in\Theta_n} \mathbb{E}_n[\sigma \ell^k_{c,\psi}(\theta; C)],
$$

where σ is the Rademacher variable independent of $(\boldsymbol{X}, A, Y)$ under $\mathbb{P}$. Then by standard symmetrization arguments, we have

$$
\mathbb{E} \sup_{\theta\in\Theta_n} (\mathbb{P} - \mathbb{P}_n)\ell^k_{c,\psi}(\theta; C) \leqslant 2\mathsf{R}_n(\Theta_n), \quad \mathbb{E} \sup_{\theta\in\Theta_n} (\mathbb{P}_n - \mathbb{P})\ell^k_{c,\psi}(\theta; C) \leqslant 2\mathsf{R}_n(\Theta_n).
$$

To obtain an error bound on $\mathsf{R}_n(\Theta_n)$, we decouple Θ_n by exploiting the ℓ^1-Lipschitzness of $\ell^k_{c,\psi}$. For ease of notation, we suppress the dependency on C in $\ell^k_{c,\psi}$. Note that for $\theta_i = (f_i, \eta_i, \lambda_i)$

$(i = 1, 2)$,

$$
\begin{aligned}
&|\ell^k_{c,\psi}(\theta_1) - \ell^k_{c,\psi}(\theta_2)| \\
\leqslant &|\ell^k_{c,\psi}(f_1, \eta_1, \lambda_1) - \ell^k_{c,\psi}(f_1, \eta_1, \lambda_2)| + |\ell^k_{c,\psi}(f_1, \eta_1, \lambda_2) - \ell^k_{c,\psi}(f_1, \eta_2, \lambda_2)| + |\ell^k_{c,\psi}(f_1, \eta_2, \lambda_2) - \ell^k_{c,\psi}(f_2, \eta_2, \lambda_2)| \\
\leqslant &\frac{c}{k^\star}\left(\frac{2cM}{c-1}\right)^{k^\star}\left|\frac{1}{\lambda_1^{k^\star-1}} - \frac{1}{\lambda_1^{k^\star-1}}\right| + \frac{c}{k}|\lambda_1 - \lambda_2| + \\
&\frac{c}{k^\star \underline{\lambda}^{k^\star-1}}\left[\frac{\psi[+f(\boldsymbol{X})]}{2}\left|\left(+\widehat{C}_n(\boldsymbol{X}) - \eta_1\right)_+^{k^\star} - \left(+\widehat{C}_n(\boldsymbol{X}) - \eta_2\right)_+^{k^\star}\right| + \right. \\
&\qquad \left.\frac{\psi[-f(\boldsymbol{X})]}{2}\left|\left(-\widehat{C}_n(\boldsymbol{X}) - \eta_1\right)_+^{k^\star} - \left(-\widehat{C}_n(\boldsymbol{X}) - \eta_2\right)_+^{k^\star}\right|\right] + |\eta_1 - \eta_2| + \\
&\frac{c}{k^\star \underline{\lambda}^{k^\star-1}}\left(\frac{2cM}{c-1}\right)^{k^\star}|\psi[f_1(\boldsymbol{X})] - \psi[f_2(\boldsymbol{X})]| \\
\leqslant &L_\lambda|\lambda_1 - \lambda_2| + L_\eta|\eta_1 - \eta_2| + L_f|f_1(\boldsymbol{X}) - f_2(\boldsymbol{X})|,
\end{aligned}
$$

where

$$
\begin{cases}
L_\lambda := \frac{c}{k^\star}\left(\frac{2cM}{c-1}\right)^{k^\star} \times \frac{k^\star-1}{\underline{\lambda}^{k^\star}} + \frac{c}{k} & = \frac{2^{2k^\star+1}\|h\|_\infty^{k^\star}}{k}\frac{c^{\frac{(k^\star+1)(2k^\star-1)}{k^\star-1}}}{(c-1)^{k^\star}}M^{k^\star} + \frac{c}{k}; \\
L_\eta := \frac{c}{k^\star \underline{\lambda}^{k^\star-1}} \times k^\star\left(\frac{2cM}{c-1}\right)^{k^\star-1} + 1 & = \frac{2^{2k^\star-1/k^\star-1}\|h\|_\infty^{k^\star-1}}{k^\star}\frac{c^{2k^\star+1}}{(c-1)^{k^\star-1}}M^{k^\star-1} + 1; \\
L_f := \frac{c}{k^\star \underline{\lambda}^{k^\star-1}}\left(\frac{2cM}{c-1}\right)^{k^\star} \times 2 & = \frac{2^{2k^\star-1/k^\star+1}\|h\|_\infty^{k^\star-1}}{k^\star}\frac{c^{2k^\star+2}}{(c-1)^{k^\star}}M^{k^\star}.
\end{cases}
$$

We Denote $L_\ell := L_f \vee L_\eta \vee L_\lambda$. Notice that the leading order term as $c, (c-1)^{-1}, M \to +\infty$ is $L_\lambda = \mathcal{O}\left(\frac{c^{\frac{(k^\star+1)(2k^\star-1)}{k^\star-1}}}{(c-1)^{k^\star}}M^{k^\star}\right)$. And we also define the marginal Rademacher complexities

$$
\mathsf{R}_n(\mathcal{F}_n) := \mathbb{E}_{(\boldsymbol{X},\sigma)\sim\mathbb{P}}\sup_{f\in\mathcal{F}_n}\mathbb{E}_n[\sigma f(\boldsymbol{X})]; \quad \mathsf{R}_n(\Pi_n) := \mathbb{E}_\sigma\sup_{\eta\in\Pi_n}(\eta\mathbb{E}_n\sigma); \quad \mathsf{R}_n(\Lambda_n) := \mathbb{E}_\sigma\sup_{\lambda\in\Lambda_n}(\lambda\mathbb{E}_n\sigma).
$$

Then by the multidimensional version (Qi et al., 2019b, Lemma 3.1) of the Rademarcher complexity of the Lipschitz composition (Boucheron et al., 2005, Theorem 3.3), we have

$$
\mathsf{R}_n(\Theta_n) \leqslant L_\ell[\mathsf{R}_n(\mathcal{F}_n) + \mathsf{R}_n(\Pi_n) + \mathsf{R}_n(\Lambda_n)],
$$

where by Vapnik-Chervonenkis Inequality (Boucheron et al., 2005, Theorem 3.4), there exists a universal constant C_{VC} such that $\mathsf{R}_n(\Pi_n) \leqslant C_{\mathrm{VC}}\sqrt{2(|\overline{\eta}| \vee |\underline{\eta}|)/n}$ and $\mathsf{R}_n(\Lambda_n) \leqslant C_{\mathrm{VC}}\sqrt{2\overline{\lambda}/n}$, and by Bartlett and Mendelson (2002, Lemma 22), $\mathsf{R}_n(\mathcal{F}_n) \leqslant 2\sqrt{\gamma_n/n}$. Combining the above results,

our regret bound becomes

$$\mathcal{L}_c^k(\widehat{f}_n, \widehat{\eta}_n, \widehat{\lambda}_n) - \mathcal{R}_c^{k,*} \leqslant 8L_\ell \left(2\sqrt{\gamma_n/n} + C_{\mathrm{VC}}\sqrt{2(|\overline{\eta}| \vee |\underline{\eta}|)/n} + C_{\mathrm{VC}}\sqrt{2\overline{\lambda}/n}\right) + 4(\overline{\ell}_c^k - \underline{\ell}_c^k)\sqrt{\frac{\log(2/\delta)}{2n}} + 4L_C\|\widehat{C}_n - C\|_\infty + 2\mathcal{A}_c^k(\gamma_n).$$

Finally by Assumption 2.5 that $\mathcal{A}_c^k(\gamma) = K_{\mathcal{A}}\gamma^\beta$, we choose $\gamma_n := n^{\frac{1}{2\beta+1}}$ to obtain the desired regret bound of rate $\mathcal{O}(n^{-\frac{\beta}{2\beta+1}})$ as $n \to \infty$, with the universal constant K_0 as

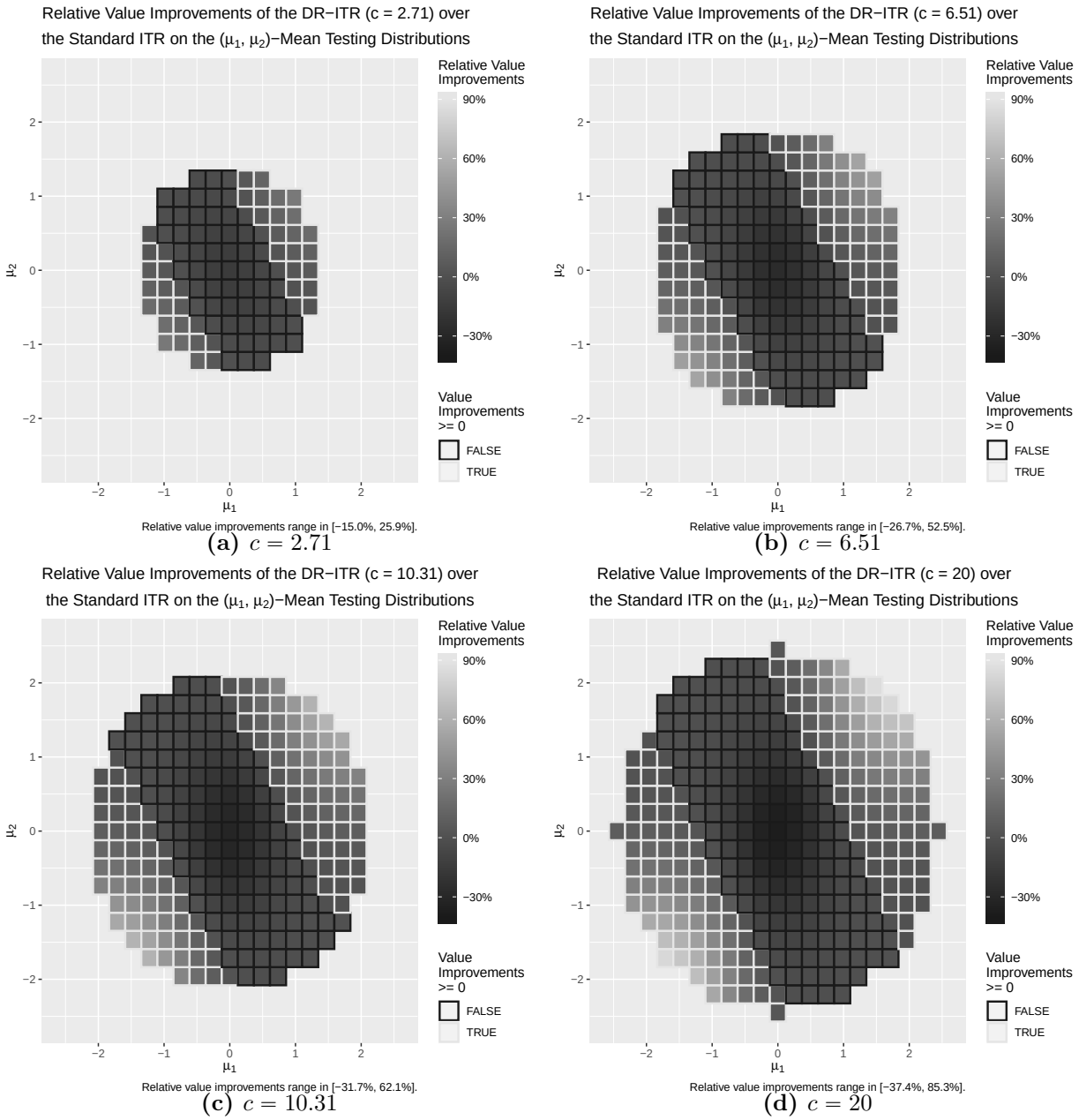
$$\begin{aligned} K_0 &= 8L_\ell \left(2 + C_{\mathrm{VC}}\sqrt{2}(|\overline{\eta}| \vee |\underline{\eta}|)^{1/2} + C_{\mathrm{VC}}\sqrt{2}\overline{\lambda}^{1/2}\right) + 2\sqrt{2}(\overline{\ell}_c^k - \underline{\ell}_c^k) + 2K_{\mathcal{A}} \\ &= \mathcal{O}\left(L_\ell[(|\overline{\eta}| \vee |\underline{\eta}|)^{1/2} + \overline{\lambda}^{1/2}] + \overline{\ell}_c^k - \underline{\ell}_c^k\right) \\ &= \mathcal{O}\left(\frac{c^{\frac{(k^\star+1)(2k^\star-1)}{k^\star-1}+\frac{1}{2}}}{(c-1)^{k^\star+1/2}} M^{k^\star+1/2}\right), \end{aligned}$$

and $K_1 = 4L_C = \mathcal{O}\left(\frac{c^{2k^\star+1}}{(c-1)^{k^\star-1}} M^{k^\star-1}\right)$.

Consider the spectial case $k = +\infty$ and $k^\star = 1$. Consider η_f^* as in (2.20). Since for any $\eta \leqslant -M$, the objective (2.20) remains constant. Then we have $-M \leqslant \eta \leqslant M$. The regret bound analysis follows the same as above except that λ is redundant in $\ell_{c,\psi}^1$. For the bounds on $\ell_{c,\psi}^1$, have $\overline{\ell}_c^1 = (2c+1)M$ and $\underline{\ell}_c^1 = -M$. The Lipschitz constants are refined to be $L_C = 2c$, $L_\eta = c+1$, $L_f = 4cM$. And the final universal constants become

$$K_0 = \mathcal{O}\left(L_\ell(|\overline{\eta}| \vee |\underline{\eta}|)^{1/2} + \overline{\ell}_c^k - \underline{\ell}_c^k\right) = \mathcal{O}(cM^{3/2}); \quad K_1 = 8c = \mathcal{O}(c).$$

2.7.4 Additional Tables and Figures



(a) $c = 2.71$ **(b)** $c = 6.51$

(c) $c = 10.31$ **(d)** $c = 20$

Figure 2.6: Comparing the testing values of the DR-ITR for various c's with the standard ITR on testing distributions $\mathcal{N}_2\big(\boldsymbol{\mu}, \mathbf{I}_2\big)$ of means $\boldsymbol{\mu} \in \big\{(\mu_1, \mu_2)^\intercal \in \mathbb{R}^2 : \mu_1^2 + \mu_2^2 \leqslant 4 \log 5\big\}$.

Table 2.6: Relative Regrets (%) of Standard ITRs on Mean-Shifted Covariate Domains

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	0	0	0	0	2	8	27	58	91	107	108
1.96	0	0	0	0	2	10	28	54	75	83	80
1.47	0	0	0	0	2	12	28	46	55	57	52
0.979	1	1	1	0	1	11	25	35	38	35	31
0.49	3	3	3	2	2	2	16	23	22	19	16
0	7	9	11	10	3	5	3	10	11	9	7
−0.49	16	19	22	23	17	3	1	2	3	3	3
−0.979	30	35	38	34	26	10	1	0	1	1	1
−1.47	52	57	55	45	27	12	2	0	0	0	0
−1.96	79	82	75	53	29	11	2	0	0	0	0
−2.45	108	107	91	58	27	9	2	0	0	0	0

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Relative regret(ITR) = [value(LB-ITR) − value(ITR)]/|value(LB-ITR)|

Table 2.7: Misclassification Rates (%) of Standard ITRs on Mean-Shifted Covariate Domains

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	1	1	2	3	4	6	10	18	30	43	53
1.96	2	3	5	7	8	10	13	20	29	38	44
1.47	3	6	10	13	15	16	19	23	28	33	35
0.979	6	10	16	20	23	25	26	27	28	27	26
0.49	9	15	22	27	30	32	32	30	27	23	19
0	13	19	26	31	34	35	34	30	26	19	13
−0.49	18	23	27	30	32	32	30	27	21	15	9
−0.979	26	27	28	27	26	25	23	20	16	11	6
−1.47	34	33	28	23	19	16	15	13	10	6	3
−1.96	44	38	29	20	14	10	8	7	5	3	2
−2.45	53	43	30	18	10	6	4	3	2	1	1

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.

Table 2.8: Relative Regrets (%) of RCT-DR-ITRs on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	0	0	0	1	3	8	16	19	16	10	6
1.96	0	0	1	1	4	11	19	21	15	10	5
1.47	0	0	1	2	4	14	23	23	15	8	4
0.979	0	0	1	2	6	15	24	22	14	7	3
0.49	1	2	2	3	7	9	18	18	11	5	2
0	1	3	7	9	8	16	9	10	7	3	1
−0.49	2	5	11	17	19	10	7	3	2	1	1
−0.979	3	7	14	21	23	14	5	2	1	0	0
−1.47	3	7	14	22	21	13	4	1	0	0	0
−1.96	5	9	15	21	19	10	3	1	0	0	0
−2.45	6	9	15	18	15	8	2	1	0	0	0

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Relative regret(ITR) = [value(LB-ITR) − value(ITR)]/|value(LB-ITR)|

Table 2.9: Relative Regrets (%) of RCT-DR-ITRs on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 100$)

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	0	0	0	1	3	7	14	16	14	9	6
1.96	0	0	0	1	3	10	18	19	13	8	4
1.47	0	0	0	1	3	12	21	20	14	7	3
0.979	0	0	1	2	4	13	22	20	13	6	2
0.49	1	1	2	2	4	7	17	17	10	4	2
0	1	3	6	8	5	11	6	8	6	3	1
−0.49	2	4	10	16	17	7	4	2	2	1	1
−0.979	2	6	13	20	22	12	3	1	1	0	0
−1.47	3	7	13	20	20	12	3	1	0	0	0
−1.96	4	8	13	18	17	10	3	1	0	0	0
−2.45	5	8	14	16	13	7	2	0	0	0	0

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Relative regret(ITR) = [value(LB-ITR) − value(ITR)]/|value(LB-ITR)|

Table 2.10: Relative Value Improvements (%) of RCT-DR-ITRs over Standard ITRs on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	0	0	0	−1	−1	1	11	40	75	98	102
1.96	0	0	−1	−1	−2	0	9	32	60	73	75
1.47	0	0	0	−2	−3	−3	6	23	40	49	48
0.979	0	0	0	−2	−4	−5	2	13	24	28	28
0.49	2	2	1	−2	−6	−7	−2	5	11	14	14
0	6	6	4	1	−5	−11	−6	0	4	6	5
−0.49	13	14	11	6	−2	−6	−5	−2	1	2	2
−0.979	27	29	24	13	2	−4	−4	−2	0	0	0
−1.47	48	49	41	23	6	−1	−3	−1	0	0	0
−1.96	74	73	60	33	10	0	−1	−1	0	0	0
−2.45	102	98	76	40	12	1	−1	−1	0	0	0

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Relative value improvement = difference of relative regrets.

Table 2.11: Misclassification Rates (%) of RCT-DR-ITRs on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	1	2	3	4	5	7	12	16	20	19	15
1.96	2	3	6	7	10	12	15	20	21	20	15
1.47	3	7	11	14	17	19	22	24	24	21	15
0.979	6	11	17	22	26	28	29	30	27	21	14
0.49	9	15	23	30	34	35	35	33	28	21	13
0	11	19	27	34	37	39	37	33	27	19	11
−0.49	13	21	28	33	35	35	34	30	23	15	9
−0.979	14	21	27	29	29	28	25	22	17	11	6
−1.47	14	20	24	24	21	19	17	14	11	7	3
−1.96	15	19	21	20	15	12	9	8	6	3	2
−2.45	15	18	19	16	11	7	5	4	2	1	1

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.

Table 2.12: Misclassification Rates (%) of RCT-DR-ITRs on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 100$)

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	1	2	3	4	5	7	11	16	19	19	15
1.96	2	3	6	7	9	12	15	20	21	20	14
1.47	3	7	10	14	17	18	21	24	24	21	14
0.979	6	11	17	22	25	27	28	29	27	21	14
0.49	9	15	23	29	32	34	34	33	28	21	13
0	11	19	27	33	36	37	36	33	27	19	11
−0.49	12	21	28	32	34	34	33	29	23	15	9
−0.979	13	21	27	29	28	27	25	22	17	11	6
−1.47	14	20	24	24	21	18	16	14	11	7	3
−1.96	14	19	21	19	15	11	9	7	6	3	2
−2.45	15	18	19	16	11	7	5	3	2	1	1

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.

Table 2.13: Misclassification Improvements (%) of RCT-DR-ITRs over Standard ITRs on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	−2.45	−1.96	−1.47	−0.979	−0.49	0	0.49	0.979	1.47	1.96	2.45
2.45	0	0	−1	−1	−1	−1	−1	2	10	24	38
1.96	0	0	−1	−1	−1	−2	−2	0	8	18	29
1.47	0	0	−1	−1	−2	−3	−3	−1	3	12	20
0.979	0	0	−1	−2	−3	−3	−3	−3	1	6	12
0.49	1	0	−1	−3	−3	−3	−3	−3	−1	2	6
0	2	0	−2	−3	−3	−4	−4	−3	−2	1	2
−0.49	6	3	−1	−3	−3	−3	−3	−3	−1	0	1
−0.979	12	7	1	−2	−3	−3	−2	−2	−1	0	0
−1.47	20	12	4	−1	−2	−2	−2	−1	−1	0	0
−1.96	29	18	8	0	−2	−2	−1	−1	0	0	0
−2.45	38	24	11	3	−1	−1	−1	−1	0	0	0

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.

Table 2.14: Testing Values (Standard Errors) on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	type	0	0.734	1.469	1.958
	LB-ITR	*2.333 (0.00244)*	*2.907 (0.011)*	*5.334 (0.0362)*	*9.27 (0.0154)*
	ℓ^1-PLS	**2.124** (0.0022)	2.235 (0.011)	3.613 (0.0505)	6.32 (0.103)
	RWL	2.067 (0.00125)	1.59 (0.0104)	0.7237 (0.0488)	0.2045 (0.108)
1.958	Standard ITR	2.089 (0.00158)	1.735 (0.013)	1.348 (0.0595)	1.567 (0.13)
	RCT-ITR	1.913 (0.0082)	1.969 (0.026)	4.168 (0.034)	7.838 (0.0388)
	RCT-DR-ITR	2.085 (0.00444)	2.286 (0.0114)	4.545 (0.0255)	8.371 (0.0451)
	CTE-DR-ITR	2.098 (0.00348)	**2.304** (0.0106)	**4.551** (0.0238)	**8.459** (0.0424)
	LB-ITR	*1.893 (0.00712)*	*2.627 (0.00656)*	*5.28 (0.0213)*	*9.379 (0.0128)*
	ℓ^1-PLS	1.667 (0.00307)	**2.021** (0.0076)	4.095 (0.0342)	7.573 (0.0706)
	RWL	1.655 (0.00131)	1.501 (0.0106)	1.798 (0.0472)	2.791 (0.102)
1.469	Standard ITR	**1.674** (0.00152)	1.645 (0.0127)	2.377 (0.0553)	4.011 (0.119)
	RCT-ITR	1.414 (0.0094)	1.597 (0.025)	4.075 (0.0299)	8.022 (0.0334)
	RCT-DR-ITR	1.627 (0.00688)	1.987 (0.00997)	4.484 (0.0192)	8.611 (0.0285)
	CTE-DR-ITR	1.663 (0.00326)	1.997 (0.00992)	**4.55** (0.0163)	**8.686** (0.0269)
	LB-ITR	*1.227 (0.00244)*	*2.144 (0.00609)*	*5.269 (0.00931)*	*9.608 (0.00898)*
	ℓ^1-PLS	1.094 (0.00418)	**1.676** (0.00442)	4.587 (0.0151)	8.8 (0.0314)
	RWL	1.168 (0.00134)	1.462 (0.00729)	3.357 (0.0344)	6.323 (0.0696)
0.734	Standard ITR	**1.174** (0.00149)	1.553 (0.00806)	3.739 (0.0379)	7.06 (0.0763)
	RCT-ITR	0.7323 (0.011)	1.152 (0.021)	4.157 (0.0238)	8.534 (0.0299)
	RCT-DR-ITR	1.094 (0.00753)	1.651 (0.00675)	4.622 (0.0109)	9.036 (0.015)
	CTE-DR-ITR	1.152 (0.00292)	1.667 (0.00588)	**4.648** (0.0113)	**9.06** (0.0161)
	LB-ITR	*0.9942 (0.00202)*	*1.774 (0.0034)*	*5.232 (0.00559)*	*9.767 (0.0068)*
	ℓ^1-PLS	0.8296 (0.00454)	1.648 (0.0036)	**4.914** (0.00501)	**9.476** (0.0103)
	RWL	**0.9457** (0.00126)	1.645 (0.00339)	4.494 (0.0165)	8.589 (0.0329)
0.000	Standard ITR	0.9437 (0.00153)	1.679 (0.00336)	4.654 (0.017)	8.895 (0.0342)
	RCT-ITR	0.4303 (0.0109)	1.161 (0.0145)	4.518 (0.0172)	8.983 (0.034)
	RCT-DR-ITR	0.8374 (0.00821)	1.647 (0.00574)	4.868 (0.00797)	9.444 (0.00841)
	CTE-DR-ITR	0.9206 (0.00272)	**1.688** (0.00289)	4.888 (0.00698)	9.442 (0.00999)

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Values (larger the better) can be comparable for the same (μ_1, μ_2) but incomparable across different (μ_1, μ_2).
[3] LB-ITR maximizes the testing value function at (μ_1, μ_2) over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.
[4] RWL (Zhou et al., 2017) implements the same routine as Standard ITR except that $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$.
[5] RCT-ITR fits RWL on the calibrating RCT dataset directly.

Table 2.15: Testing Values (Standard Errors) on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 100$)

μ_2 \ μ_1	type	0	0.734	1.469	1.958
	LB-ITR	*2.333 (0.00244)*	*2.907 (0.011)*	*5.334 (0.0362)*	*9.27 (0.0154)*
	ℓ^1-PLS	**2.124** (0.0022)	2.235 (0.011)	3.613 (0.0505)	6.32 (0.103)
	RWL	2.067 (0.00125)	1.59 (0.0104)	0.7237 (0.0488)	0.2045 (0.108)
1.958	Standard ITR	2.089 (0.00158)	1.735 (0.013)	1.348 (0.0595)	1.567 (0.13)
	RCT-ITR	2.015 (0.00565)	**2.593** (0.0132)	**4.996** (0.0158)	**8.588** (0.0208)
	RCT-DR-ITR	2.109 (0.00342)	2.349 (0.00905)	4.62 (0.0219)	8.5 (0.0394)
	CTE-DR-ITR	2.099 (0.00392)	2.34 (0.00954)	4.602 (0.0215)	8.488 (0.0393)
	LB-ITR	*1.893 (0.00712)*	*2.627 (0.00656)*	*5.28 (0.0213)*	*9.379 (0.0128)*
	ℓ^1-PLS	1.667 (0.00307)	2.021 (0.0076)	4.095 (0.0342)	7.573 (0.0706)
	RWL	1.655 (0.00131)	1.501 (0.0106)	1.798 (0.0472)	2.791 (0.102)
1.469	Standard ITR	**1.674** (0.00152)	1.645 (0.0127)	2.377 (0.0553)	4.011 (0.119)
	RCT-ITR	1.54 (0.00529)	**2.286** (0.0129)	**4.846** (0.017)	8.713 (0.0183)
	RCT-DR-ITR	1.662 (0.00367)	2.044 (0.00721)	4.566 (0.0153)	8.711 (0.0254)
	CTE-DR-ITR	1.67 (0.00286)	2.044 (0.00818)	4.577 (0.0144)	**8.734** (0.0251)
	LB-ITR	*1.227 (0.00244)*	*2.144 (0.00609)*	*5.269 (0.00931)*	*9.608 (0.00898)*
	ℓ^1-PLS	1.094 (0.00418)	1.676 (0.00442)	4.587 (0.0151)	8.8 (0.0314)
	RWL	1.168 (0.00134)	1.462 (0.00729)	3.357 (0.0344)	6.323 (0.0696)
0.734	Standard ITR	**1.174** (0.00149)	1.553 (0.00806)	3.739 (0.0379)	7.06 (0.0763)
	RCT-ITR	0.8905 (0.00647)	1.651 (0.0138)	**4.701** (0.0168)	9.011 (0.013)
	RCT-DR-ITR	1.134 (0.00408)	1.662 (0.0065)	4.671 (0.00885)	9.094 (0.0122)
	CTE-DR-ITR	1.156 (0.00251)	**1.68** (0.00573)	4.699 (0.00824)	**9.132** (0.0112)
	LB-ITR	*0.9942 (0.00202)*	*1.774 (0.0034)*	*5.232 (0.00559)*	*9.767 (0.0068)*
	ℓ^1-PLS	0.8296 (0.00454)	1.648 (0.0036)	4.914 (0.00501)	9.476 (0.0103)
	RWL	**0.9457** (0.00126)	1.645 (0.00339)	4.494 (0.0165)	8.589 (0.0329)
0.000	Standard ITR	0.9437 (0.00153)	1.679 (0.00336)	4.654 (0.017)	8.895 (0.0342)
	RCT-ITR	0.6198 (0.00875)	1.388 (0.00857)	4.745 (0.00861)	9.376 (0.00737)
	RCT-DR-ITR	0.8879 (0.00506)	1.671 (0.00389)	4.901 (0.00451)	9.489 (0.0068)
	CTE-DR-ITR	0.925 (0.00233)	**1.689** (0.00262)	**4.916** (0.00496)	**9.508** (0.00626)

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Values (larger the better) can be comparable for the same (μ_1, μ_2) but incomparable across different (μ_1, μ_2).
[3] LB-ITR maximizes the testing value function at (μ_1, μ_2) over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.
[4] RWL (Zhou et al., 2017) implements the same routine as Standard ITR except that $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$.
[5] RCT-ITR fits RWL on the calibrating RCT dataset directly.

Table 2.16: Testing Misclassification Rates (Standard Errors) on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	type	0	0.734	1.469	1.958
	LB-ITR	*0.05348 (0.000259)*	*0.0301 (0.000804)*	*0.02702 (0.0038)*	*0.02554 (0.00337)*
	ℓ^1-PLS	0.113 (0.000781)	**0.1625** (0.000913)	0.2239 (0.0015)	0.247 (0.00305)
	RWL	**0.09857** (0.000358)	0.1675 (0.000346)	0.3093 (0.00126)	0.4145 (0.00255)
1.958	Standard ITR	0.0988 (0.000392)	0.1628 (0.000402)	0.29 (0.00163)	0.3802 (0.00322)
	RCT-ITR	0.1148 (0.00191)	0.1783 (0.00334)	0.2567 (0.00477)	0.2687 (0.00374)
	RCT-DR-ITR	0.118 (0.00135)	0.1785 (0.00196)	**0.2148** (0.00178)	**0.1997** (0.00192)
	CTE-DR-ITR	0.1142 (0.00114)	0.1879 (0.0021)	0.236 (0.00237)	0.209 (0.00201)
	LB-ITR	*0.11 (0.00149)*	*0.05955 (0.000487)*	*0.0374 (0.00328)*	*0.03026 (0.00324)*
	ℓ^1-PLS	0.1904 (0.00113)	0.2229 (0.00132)	**0.2353** (0.0011)	0.2251 (0.00203)
	RWL	**0.1616** (0.000581)	0.2099 (0.000599)	0.2972 (0.00124)	0.3601 (0.00255)
1.469	Standard ITR	0.1637 (0.00067)	**0.2066** (0.000681)	0.2781 (0.00153)	0.326 (0.00307)
	RCT-ITR	0.1875 (0.00248)	0.2381 (0.00365)	0.2895 (0.00471)	0.2744 (0.00324)
	RCT-DR-ITR	0.1927 (0.00205)	0.2306 (0.00196)	0.2437 (0.00199)	**0.2109** (0.00173)
	CTE-DR-ITR	0.181 (0.00132)	0.2373 (0.00221)	0.2514 (0.00208)	0.2155 (0.00168)
	LB-ITR	*0.2575 (0.000703)*	*0.144 (0.00177)*	*0.07107 (0.00288)*	*0.04661 (0.00282)*
	ℓ^1-PLS	0.3275 (0.00147)	0.3291 (0.00165)	**0.273** (0.00104)	0.2085 (0.00091)
	RWL	**0.2764** (0.000746)	**0.2877** (0.000915)	0.2858 (0.000886)	0.2747 (0.00184)
0.734	Standard ITR	0.283 (0.000914)	0.2898 (0.00109)	0.2747 (0.00101)	0.2519 (0.00205)
	RCT-ITR	0.333 (0.00275)	0.3537 (0.0036)	0.3333 (0.00393)	0.2615 (0.00234)
	RCT-DR-ITR	0.3178 (0.00237)	0.3203 (0.00214)	0.2778 (0.00192)	0.2102 (0.00128)
	CTE-DR-ITR	0.2974 (0.00129)	0.3147 (0.00189)	0.2771 (0.00173)	**0.2076** (0.00118)
	LB-ITR	*0.3246 (0.000396)*	*0.2802 (0.0015)*	*0.1293 (0.00214)*	*0.08388 (0.00267)*
	ℓ^1-PLS	0.3988 (0.0016)	0.3649 (0.00139)	0.2742 (0.000873)	0.1875 (0.000467)
	RWL	**0.3358** (0.000755)	**0.3147** (0.000808)	0.2582 (0.000556)	0.2033 (0.000881)
0.000	Standard ITR	0.3452 (0.000963)	0.3211 (0.001)	**0.2564** (0.000666)	0.1942 (0.000918)
	RCT-ITR	0.4085 (0.0025)	0.4158 (0.00234)	0.3261 (0.00214)	0.2349 (0.00169)
	RCT-DR-ITR	0.3864 (0.00274)	0.3529 (0.0021)	0.2726 (0.0015)	0.1889 (0.000857)
	CTE-DR-ITR	0.3575 (0.00126)	0.3345 (0.00123)	0.264 (0.00106)	**0.1848** (0.000668)

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] LB-ITR maximizes the testing value function at (μ_1, μ_2) over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.
[3] RWL (Zhou et al., 2017) implements the same routine as Standard ITR except that $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$.
[4] RCT-ITR fits RWL on the calibrating RCT dataset directly.

Table 2.17: Testing Values of RCT-DR-ITRs of Various k's on Mean-Shifted Covariate Domains ($n_{\text{calib}} = 50$)

μ_2 \ μ_1	k	0	0.734	1.47	1.96
	1.25	2.08(0.004443)	2.25(0.01238)	4.4(0.03824)	8.17(0.07266)
	1.5	2.09(0.004052)	2.28(0.01154)	4.47(0.0317)	8.27(0.05863)
1.96	2	2.09(0.004445)	2.29(0.01139)	4.54(0.02549)	8.37(0.04507)
	3	2.08(0.005431)	2.25(0.01187)	4.52(0.02422)	8.37(0.0428)
	∞	2.1(0.004169)	2.27(0.01313)	4.54(0.02419)	8.43(0.03522)
	1.25	1.64(0.005444)	1.99(0.009954)	4.42(0.02606)	8.45(0.04875)
	1.5	1.64(0.005729)	2(0.009707)	4.42(0.02437)	8.52(0.04136)
1.47	2	1.63(0.006885)	1.99(0.009965)	4.48(0.01924)	8.61(0.02852)
	3	1.64(0.006302)	1.98(0.01028)	4.47(0.01846)	8.63(0.02501)
	∞	1.64(0.006803)	1.98(0.01093)	4.51(0.01848)	8.63(0.02595)
	1.25	1.11(0.006071)	1.64(0.006628)	4.58(0.01659)	8.95(0.02455)
	1.5	1.12(0.005547)	1.64(0.007019)	4.58(0.01508)	8.97(0.02298)
0.734	2	1.09(0.007527)	1.65(0.006753)	4.62(0.01089)	9.04(0.01496)
	3	1.1(0.007473)	1.62(0.008308)	4.59(0.01228)	9.02(0.01563)
	∞	1.12(0.00672)	1.62(0.008311)	4.61(0.01417)	9.04(0.01468)
	1.25	0.859(0.007158)	1.65(0.005616)	4.87(0.007131)	9.43(0.01052)
	1.5	0.859(0.007117)	1.64(0.006172)	4.88(0.006802)	9.43(0.0116)
0	2	0.837(0.008205)	1.65(0.005744)	4.87(0.007969)	9.44(0.008415)
	3	0.854(0.007488)	1.64(0.006564)	4.86(0.006542)	9.46(0.007206)
	∞	0.888(0.005782)	1.64(0.005722)	4.85(0.008767)	9.45(0.008676)

[1] $\boldsymbol{\mu} = (\mu_1, \mu_2, 0, \cdots, 0)^\intercal$ with μ_1 in column and μ_2 in row is the testing covariate centroid.
[2] Values (larger the better) can be comparable for the same (μ_1, μ_2) but incomparable across different (μ_1, μ_2).

Table 2.18: Testing Values (Standard Errors) on Mixture of Subgroups ($n_{\text{calib}} = 50$)

	Testing Subgroup 1 Probability				
type	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**
LB-ITR	*1.665 (0.0067)*	*1.537 (0.00618)*	*1.444 (0.00412)*	*1.545 (0.00537)*	*1.679 (0.00585)*
ℓ^1-PLS	1.182 (0.00191)	1.264 (0.0014)	**1.399** (0.000591)	**1.537** (0.000333)	1.624 (0.000781)
RWL	1.092 (0.00349)	1.194 (0.00265)	1.363 (0.00123)	1.535 (0.00046)	**1.64** (0.00114)
Standard ITR	1.143 (0.00434)	1.232 (0.00329)	1.383 (0.0015)	1.535 (0.000543)	1.632 (0.00142)
RCT-ITR	1.251 (0.0108)	1.116 (0.011)	1.046 (0.0108)	1.144 (0.0101)	1.275 (0.0102)
RCT-DR-ITR	**1.267** (0.0066)	**1.305** (0.00423)	1.395 (0.00256)	1.52 (0.00212)	1.614 (0.00234)
CTE-DR-ITR	1.16 (0.00409)	1.247 (0.00323)	1.388 (0.00137)	1.534 (0.00055)	1.628 (0.00149)

[1] Testing subgroup 1 probability = 0.75 is the same as the training one.
[2] Values (larger the better) can be comparable for the same subgroup 1 probability but incomparable across different subgroup 1 probabilities
[3] LB-ITR maximizes the testing value function over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.
[4] RWL (Zhou et al., 2017) implements the same routine as Standard ITR except that $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$.
[5] RCT-ITR fits RWL on the calibrating RCT dataset directly.

Table 2.19: Testing Values (Standard Errors) on Mixture of Subgroups ($n_{\text{calib}} = 100$)

	Testing Subgroup 1 Probability				
type	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**
LB-ITR	*1.665 (0.0067)*	*1.537 (0.00618)*	*1.444 (0.00412)*	*1.545 (0.00537)*	*1.679 (0.00585)*
ℓ^1-PLS	1.182 (0.00191)	1.264 (0.0014)	1.399 (0.000591)	**1.537** (0.000333)	1.624 (0.000781)
RWL	1.092 (0.00349)	1.194 (0.00265)	1.363 (0.00123)	1.535 (0.00046)	**1.64** (0.00114)
Standard ITR	1.143 (0.00434)	1.232 (0.00329)	1.383 (0.0015)	1.535 (0.000543)	1.632 (0.00142)
RCT-ITR	**1.493** (0.00431)	**1.354** (0.00499)	1.25 (0.00489)	1.359 (0.0049)	1.5 (0.0046)
RCT-DR-ITR	1.284 (0.00654)	1.324 (0.00421)	**1.402** (0.00195)	1.524 (0.00191)	1.613 (0.00233)
CTE-DR-ITR	1.165 (0.00403)	1.247 (0.00305)	1.389 (0.00134)	1.535 (0.000584)	1.628 (0.00147)

[1] Testing subgroup 1 probability = 0.75 is the same as the training one.
[2] Values (larger the better) can be comparable for the same subgroup 1 probability but incomparable across different subgroup 1 probabilities
[3] LB-ITR maximizes the testing value function over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.
[4] RWL (Zhou et al., 2017) implements the same routine as Standard ITR except that $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$.
[5] RCT-ITR fits RWL on the calibrating RCT dataset directly.

Table 2.20: Testing Misclassification Rates (Standard Errors) on Mixture of Subgroups ($n_{\text{calib}} = 50$)

	Testing Subgroup 1 Probability				
type	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**
LB-ITR	*0.06691 (0.0017)*	*0.1556 (0.0014)*	*0.2296 (0.00078)*	*0.153 (0.0012)*	*0.06668 (0.0015)*
ℓ^1-PLS	0.3059 (0.00044)	0.2775 (0.00027)	0.2291 (0.00016)	0.1789 (0.00041)	0.149 (0.00058)
RWL	0.3242 (0.00071)	0.2885 (4e-04)	0.2283 (0.00021)	**0.1664** (0.00069)	**0.1302** (0.00099)
Standard ITR	0.3103 (0.00097)	0.2785 (0.00058)	**0.2238** (0.00017)	0.1676 (0.00074)	0.1342 (0.0011)
RCT-ITR	**0.2472** (0.0027)	0.2822 (0.0025)	0.3001 (0.0022)	0.2763 (0.0023)	0.2436 (0.0026)
RCT-DR-ITR	0.2751 (0.0023)	**0.2614** (0.0013)	0.2266 (0.00052)	0.1809 (0.0012)	0.147 (0.0014)
CTE-DR-ITR	0.3068 (0.00093)	0.2759 (0.00059)	0.2242 (0.00019)	0.1701 (0.00074)	0.1379 (0.0011)

[1] Testing subgroup 1 probability = 0.75 is the same as the training one.
[2] LB-ITR maximizes the testing value function over the linear ITR class. The corresponding testing value is the best achievable among the linear ITR class.
[3] RWL (Zhou et al., 2017) implements the same routine as Standard ITR except that $\widehat{C}_n(\boldsymbol{X}) = \widehat{Q}_n(\boldsymbol{X}, 1) - \widehat{Q}_n(\boldsymbol{X}, -1) + 2A[Y - \widehat{Q}_n(\boldsymbol{X}, A)]$.
[4] RCT-ITR fits RWL on the calibrating RCT dataset directly.

CHAPTER 3

Efficient Learning of Optimal Individualized Treatment Rules for Heteroscedastic or Misspecified Treatment-Free Effect Models

3.1 Introduction

Among the methods of finding an optimal ITR, the *double robustness* property has been studied and advocated to protect from potential model misspecifications. In the model-based approaches, the optimal ITR only depends on the interaction effect between covariates and treatment within the outcome mean model. Then the treatment-free effect that only depends on covariates can be a nuisance component. Robins (2004) investigated the incorrectly specified parametric model for the treatment-free effect, and introduced the G-estimating equation that can incorporate additional information from the propensity score. The G-estimator can be doubly robust in the sense that the estimate remains consistent even if one of the treatment-free effect model and the propensity score model is misspecified. As special cases, Lu et al. (2013); Ertefaie et al. (2021) developed least-squares approaches that can equivalently solve the G-estimating equation and enjoy double robustness. Wallace and Moodie (2015); Meng and Qiao (2020) took a different approach to hedge the risk of treatment-free effect misspecification. Specifically, they proposed the weighted least-squares problem that utilizes the propensity score information to construct balancing weights, and the resulting estimates can also be doubly robust. In the direct-search approaches, the AIPWE of the value function is doubly robust in a slightly different way. Specifically, the AIPWE incorporates the outcome mean function and the propensity score function. When estimating the outcome mean and propensity score functions, even if one of their model specifications is incorrect, the corresponding AIPWE can still remain consistent.

The double robustness property has also been widely studied in the causal inference literature (Robins et al., 1994, 1995; Ding and Li, 2018). One problem of particular interest is to study the case when one of or both model misspecifications happen. Kang and Schafer (2007) provided a

comprehensive empirical study on how model misspecification can affect the resulting estimates. They concluded that the misspecified outcome mean model can be generally more harmful than the misspecified propensity score model. When both models are misspecified, the doubly robust estimate can perform even worse than the IPWE. Later studies further developed improved estimates and inference procedures to overcome such challenges (Tan, 2010; Rotnitzky et al., 2012; Vermeulen and Vansteelandt, 2015; Benkeser et al., 2017). These studies have also motivated some improvement of the AIPWE for the ITR problem. Specifically, when the outcome mean model is incorrectly specified, Cao et al. (2009) proposed an optimal estimation strategy for the misspecified outcome mean model in the sense that the resulting AIPWE can have the smallest variance. Pan and Zhao (2021) further extended this work to the ITR problem, and utilized augmented inverse-probability weighted estimating equations for the outcome mean model estimation.

Motivated from Kang and Schafer (2007) that the misspecified treatment-free effect can have more severe consequence, we focus on addressing this challenge. In our study, we find that the misspecified treatment-free effect in the model-based approach can have a consequence similar to heteroscedasticity (Carroll, 1982). More specifically, both misspecified treatment-free effect and heteroscedasticity can cause the variance of residuals being dependent on covariates and treatment. Therefore, we take the approach of semiparametric efficient estimation under heteroscedasticity (Ma et al., 2006) and propose an *Efficient Learning (E-Learning)* framework for the optimal ITR in the multi-armed treatment setting. Our proposed E-Learning can enjoy the following properties:

1. When nuisance models are correctly specified, E-Learning performs semiparametric efficient estimation. Our framework can allow the variance of outcome depends on covariates and treatment, and hence is more general than existing semiparametric efficient procedures such as G-Estimation and its equivalents;

2. E-Learning is doubly robust with respect to the treatment-free effect model and the propensity score model;

3. In presence of misspecified treatment-free effect, E-Learning is optimal with the minimal $\sqrt{n}$-asymptotic variance among a regular class of semiparametric estimates based on the given working treatment-free effect function. Our optimality incorporates the standard semiparametric efficiency (Tsiatis, 2007) as a special case for the ITR problem.

This chapter contributes to existing literature in terms of the followings:

1. Parallel to the improved doubly robust procedure in Pan and Zhao (2021) for direct-search approaches, E-Learning is an improved doubly robust method for model-based approaches. Specifically, E-Learning performs optimal efficiency improvement when one of or both misspecified treatment-free effect and heteroscedasticity exist;

2. E-Learning incorporates many existing approaches as special cases, including Q-Learning, G-Estimation, A-Learning, dWOLS, Subgroup Identification, D-Learning and RD-Learning. It provides a more general framework to study the double robustness and estimation efficiency for these methods;

3. We develop E-Learning for the setting with multiple treatments. In particular, E-Learning utilizes a generalized equiangular coding of multiple treatment arms to develop the efficient estimating function. This can be the first work to incorporate equiangularity in the semiparametric framework among those utilizing the equiangular coding (Zhang and Liu, 2014; Zhang et al., 2020; Qi et al., 2020; Meng et al., 2020; Xue et al., 2021);

4. In our simulation study, our proposed E-Learning demonstrates superior performance over existing methods when one of or both misspecified treatment-free effect and heteroscedasticity exist, which confirms the superior performance of the proposed E-Learning. In the analysis of a *Type 2 Diabetes Mellitus (T2DM)* observational study, E-Learning also demonstrates its improved efficiency compared to other methods.

The rest of this chapter is organized as follows. In Section 3.2, we introduce the methodology of E-Learning. In particular, mathematical setups and notations are introduced in Section 3.2.1. A motivating example is discussed in Section 3.2.2 to demonstrate the consequence of misspecified treatment-free effect and heteroscedasticity. Semiparametric efficient estimating equation is developed in Section 3.2.3. E-Learning and its implementation details are proposed in Sections 3.2.4 and 3.2.5. In Section 3.3, we discuss the connection of E-Learning with the existing literature. In Section 3.4, we establish theoretical results for E-Learning. Simulation studies and the application to the T2DM dataset are provided in Sections 3.5 and 3.6 respectively. Some discussions are given in Section 3.7. Additional discussions, including an analysis of the ACTG 175 dataset, technical

proofs, additional tables and figures can be found in Section 3.8. The implementation based on `R` of this chapter is available at `https://github.com/harrymok/E-Learning.git`.

3.2 Methodology

In this section, we first introduce the ITR problem as a semiparametric estimation problem. Then we study the semiparametric efficient estimation procedure and propose E-Learning.

3.2.1 Setup

Consider the data $(\boldsymbol{X}, A, Y)$, where $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ denotes the covariates, $A \in \mathcal{A} = \{1, 2, \cdots, K\}$ is the treatment assignment with K treatment options, and $Y \in \mathbb{R}$ is the observed outcome. For $1 \leqslant k \leqslant K$, let $Y(k)$ be the potential outcome under the assigned treatment k. An ITR is a mapping from covariates to treatment assignment $d : \mathcal{X} \to \mathcal{A}$. The *value function* of an ITR is defined as $\mathcal{V}(d) := \mathbb{E}[Y(d(\boldsymbol{X}))]$. Assuming that a larger outcome is better, the goal is to find the optimal ITR that maximizes the value function $d^\star \in \operatorname{argmax}_{d:\mathcal{X}\to\mathcal{A}} \mathcal{V}(d)$.

Assume the identifiability conditions (Rubin, 1974): (*consistency*) $Y = Y(A)$; (*unconfoundedness*) $A \perp\!\!\!\perp \{Y(k)\}_{k=1}^K | \boldsymbol{X}$; (*strict overlap*) for $1 \leqslant k \leqslant K$, $\mathbb{P}(A = k|\boldsymbol{X}) \geqslant \underline{p}_{\mathscr{A}}$ for some $\underline{p}_{\mathscr{A}} > 0$. Then the value function can be written as $\mathcal{V}(d) = \mathbb{E}[Y|A = d(\boldsymbol{X})] = \mathbb{E}\left\{\sum_{k=1}^K \mathbb{E}(Y|\boldsymbol{X}, A = k)\mathbb{1}[d(\boldsymbol{X}) = k]\right\}$. Consequently, the optimal ITR satisfies $d^\star(\boldsymbol{x}) \in \operatorname{argmax}_{1\leqslant k\leqslant K} \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}, A = k)$ for any $\boldsymbol{x} \in \mathcal{X}$. This motivates us to study the following semiparametric model:

$$\begin{aligned} & Y = \mu_0(\boldsymbol{X}) + \gamma(\boldsymbol{X}, A; \boldsymbol{\beta}) + \epsilon, \\ \text{subject to } & \sum_{k=1}^K \gamma(\boldsymbol{X}, k; \boldsymbol{\beta}) = 0; \quad \mathbb{E}(\epsilon|\boldsymbol{X}, A) = 0; \quad \sigma^2(\boldsymbol{X}, A) := \mathbb{E}(\epsilon^2|\boldsymbol{X}, A) < +\infty; \\ & (\boldsymbol{X}, A, \epsilon) \sim p_{\mathscr{X}}(\boldsymbol{x}) p_{\mathscr{A}}(a|\boldsymbol{x}) p_\epsilon(\epsilon|\boldsymbol{x}, a). \end{aligned} \tag{3.1}$$

Here, $\mu_0(\boldsymbol{X})$ is the *treatment-free effect*, and $\gamma(\boldsymbol{X}, A; \boldsymbol{\beta})$ is the *interaction effect* between $\boldsymbol{X}$ and A that is parametrized by the p-dimensional parameter vector $\boldsymbol{\beta} \in \mathscr{B} \subseteq \mathbb{R}^p$. In particular, it requires that the parametrized interaction effect satisfies a sum-to-zero constraint for identifiability. The dependency on $\boldsymbol{\beta}$ may be suppressed for ease of notation in our later presentation. Moreover, $\sigma^2(\boldsymbol{X}, A)$ is the *variance function* of ϵ that can depend on $(\boldsymbol{X}, A)$. Finally, $p_{\mathscr{X}}(\boldsymbol{x})$, $p_{\mathscr{A}}(a|\boldsymbol{x})$

and $p_\epsilon(\epsilon|\boldsymbol{x}, a)$ are density functions. Then the nuisance component $\eta := (p_{\mathscr{X}}, p_{\mathscr{A}}, p_\epsilon, \mu_0)$ is left unspecified only with the moment restriction $\int \epsilon p_\epsilon(\epsilon|\boldsymbol{x}, a)\mathrm{d}\epsilon = 0$.

Given the true parameter $\boldsymbol{\beta}$ in Model (3.1), the optimal ITR is $d^\star(\boldsymbol{x}) \in \operatorname{argmax}_{1\leqslant k\leqslant K} \gamma(\boldsymbol{x}, k; \boldsymbol{\beta})$. In Theorem 3.1 below, we show that maximizing the value function can be directly related to finding a good estimate of the interaction effect $\gamma(\boldsymbol{X}, A)$ in Model (3.1).

Theorem 3.1 (Estimation and Regret Bound)**.** *Consider Model (3.1). Let $\widehat{\gamma}_n(\boldsymbol{X}, A)$ be an estimate of $\gamma(\boldsymbol{X}, A)$, $\widehat{d}_n(\boldsymbol{x}) \in \operatorname{argmax}_{1\leqslant k\leqslant K} \widehat{\gamma}_n(\boldsymbol{x}, k)$, and $d^\star(\boldsymbol{x}) \in \operatorname{argmax}_{1\leqslant k\leqslant K} \gamma(\boldsymbol{x}, k)$. Then*

$$\mathcal{V}(d^\star) - \mathcal{V}(\widehat{d}_n) \leqslant 2 \max_{1\leqslant k\leqslant K} \mathbb{E}\left|\widehat{\gamma}_n(\boldsymbol{X}, k) - \gamma(\boldsymbol{X}, k)\right|.$$

Here, $\widehat{\gamma}_n$ is fixed and $\mathbb{E}$ takes expectation over $\boldsymbol{X}$.

The proof of Theorem 3.1 is similar to Murphy (2005, Lemma 2) and is included in Section 3.8. It implies that minimizing the estimation error of the interaction effects $\{\gamma(\boldsymbol{X}, k)\}_{k=1}^K$ can also minimize the regret. In this chapter, we focus on finding an efficient estimate of the parametric interaction effect $\gamma(\boldsymbol{X}, A; \boldsymbol{\beta})$.

3.2.2 A Motivating Example

We introduce a motivating example to demonstrate that several existing approaches, including Q-Learning, G-Estimation, A-Learning, dWOLS, Subgroup Identification, D-Learning and RD-Learning, may not be optimal if either the treatment-free effect $\mu_0(\boldsymbol{X})$ is misspecified, or the variance function $\mathbb{E}(\epsilon^2|\boldsymbol{X}, A)$ depends on $(\boldsymbol{X}, A)$. In contrast, the E-Learning estimate can be much more efficient. All these methods are compared in Section 3.3.

Consider the covariate X with a symmetric distribution on $\mathbb{R}$, the treatment $A \sim \mathsf{Bernoulli}(1/2)$, and the error term $\epsilon \sim \mathcal{N}(0, 1)$, where X, A, ϵ are mutually independent. Suppose the outcome Y is generated by

$$Y = \underbrace{c_1|X|}_{\text{treatment-free effect}} + \underbrace{(A - 1/2)\beta_0}_{\text{interaction effect}} + \sqrt{\underbrace{1 + 2c_2^2 A X^2}_{\text{variance function}}}\,\epsilon,$$

for some $\beta_0 \geqslant 0$. When estimating from the training data, suppose that we specify $X\eta$ for the treatment-free effect with η to be estimated, and $(A - 1/2)\beta$ for the interaction effect with β to be estimated. If $c_1 = 0$, then the treatment-free effect is correctly specified, with the true parameter

$\eta = 0$; otherwise, the treatment-free effect is misspecified. If $c_2 = 0$, then the variance function is 1, and homogeneous with respect to (X, A); otherwise, we have a heteroscedastic model with the variance of error depending on (X, A).

Denote $\mathbb{E}_n$ as the empirical average over the training dataset of size n. Then for this particular example, Q-Learning (Watkins, 1989), G-Estimation (Robins, 2004), A-Learning (Murphy, 2003), dWOLS (Wallace and Moodie, 2015), Subgroup Identification (Tian et al., 2014), D-Learning (Qi and Liu, 2018) and RD-Learning (Meng and Qiao, 2020) are equivalent to the following *Ordinary Least-Squares (OLS)* problem:

$$(\widehat{\eta}_n, \widehat{\beta}_n) \in \underset{\eta,\beta\in\mathbb{R}}{\operatorname{argmin}} \mathbb{E}_n[Y - X\eta - (A - 1/2)\beta]^2. \tag{3.2}$$

Note that if $c_1 = c_2 = 0$ with correctly specified treatment-free effect and homoscedasticity, then $\widehat{\beta}_n$ is semiparametric efficient. For the general c_1 and c_2, the OLS estimates $\widehat{\beta}_n$ and $\widehat{\eta}_n$ are asymptotically independent, with $\sqrt{n}\widehat{\eta}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \nu^2)$ for some $\nu^2 > 0$ and

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = \sqrt{n}[\mathbb{E}(A - 1/2)^2 + o_{\mathbb{P}}(1)]^{-1}\mathbb{E}_n\left[(A - 1/2)\left(c_1|X| + \sqrt{1 + 2c_2^2AX^2}\epsilon\right)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, v^2),$$

where the $\sqrt{n}$-asymptotic variance of $\widehat{\beta}_n$ is given by $v^2 = 4\mathbb{E}[1 + (c_1^2 + c_2^2)X^2] = 4\mathbb{E}(1 + c^2X^2)$ with $c^2 := c_1^2 + c_2^2$. Notice that the residual is $\widehat{e} = Y - X\widehat{\eta}_n - (2A - 1)\widehat{\beta}_n = c_1|X| + \sqrt{1 + 2c_2^2AX^2}\epsilon + \mathcal{O}_{\mathbb{P}}(n^{-1/2})$. Then we have $\mathbb{E}(\widehat{e}^2|X) = 1 + c^2X^2 + \mathcal{O}_{\mathbb{P}}(n^{-1})$, which clearly depends on X.

Motivated from the heteroscedastic residual, we define $\breve{v}_\epsilon(x) := 4(1 + c^2x^2)$. Consider the solutions to the generalized least-squares problem

$$(\widehat{\eta}_{\text{eff},n}, \widehat{\beta}_{\text{eff},n}) \in \underset{\eta,\beta\in\mathbb{R}}{\operatorname{argmin}} \mathbb{E}_n\Big\{\breve{v}_\epsilon^{-1}(X)[Y - X\eta - (A - 1/2)\beta]^2\Big\}. \tag{3.3}$$

Then $\widehat{\beta}_{\text{eff},n}$ and $\widehat{\eta}_{\text{eff},n}$ are asymptotically independent, with $\sqrt{n}\widehat{\eta}_{\text{eff},n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \widetilde{\nu}^2)$ for some $\widetilde{\nu}^2 > 0$,

$$\sqrt{n}(\widehat{\beta}_{\text{eff},n} - \beta_0) = \sqrt{n}\left\{\mathbb{E}\left[\frac{(A - 1/2)^2}{4(1 + c^2X^2)}\right] + o_{\mathbb{P}}(1)\right\}^{-1}\mathbb{E}_n\left[\frac{(A - 1/2)\left(c_1|X| + \sqrt{1 + 2c_2^2AX^2}\epsilon\right)}{4(1 + c^2X^2)}\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, v_{\text{eff}}^2),$$

where the $\sqrt{n}$-asymptotic variance of $\widehat{\beta}_{\text{eff},n}$ is given by $v_{\text{eff}}^2 = 4\left[\mathbb{E}(1 + c^2X^2)^{-1}\right]^{-1}$. The asymptotic relative efficiency of $\widehat{\beta}_{\text{eff},n}$ with respect to $\widehat{\beta}_n$ is $v^2/v_{\text{eff}}^2 = \mathbb{E}(1 + c^2X^2)\mathbb{E}\left(\frac{1}{1+c^2X^2}\right) \geqslant 1$. That is,

$\widehat{\beta}_{\text{eff},n}$ has a smaller $\sqrt{n}$-asymptotic variance than $\widehat{\beta}_n$. The strict inequality generally holds if $c \neq 0$ and X is non-degenerate.

Next we consider an extreme case to illustrate that $\widehat{\beta}_{\text{eff},n}$ can be much more efficient than $\widehat{\beta}_n$. Suppose $X \sim qf^{(M)}(x) + (1-q)f^{(\infty)}(x)$, where $f^{(M)}(x)$ is a symmetric *probability density function (PDF)* with compact support on $[-M, M]$, $f^{(\infty)}(x)$ is a symmetric PDF on $\mathbb{R}$ with $\int_{\mathbb{R}} x^2 f^{(\infty)}(x)\mathrm{d}x = +\infty$, and $q \in (0, 1]$ is the mixture probability. Then for $c \neq 0$, $v^2 \geqslant 4[1 + c^2\mathbb{E}_{X\sim f^{(\infty)}}(X^2)] = +\infty$, while $v_{\text{eff}}^2 \leqslant \left[q\mathbb{E}_{X\sim f^{(M)}}(1 + c^2X^2)^{-1}\right]^{-1} \leqslant 4(1 + c^2M^2)/q$. Here, $v^2 = +\infty$ implies that $\widehat{\beta}_n$ cannot even be $\mathcal{O}_{\mathbb{P}}(n^{-1/2})$, while in contrast, $\widehat{\beta}_{\text{eff},n}$ has a bounded $\sqrt{n}$-asymptotic variance v_{eff}^2. Therefore, if either the treatment-free effect is misspecified ($c_1 \neq 0$), or the variance function is not homogeneous ($c_2 \neq 0$), then $\widehat{\beta}_n$ can have much worse perforamance than the more efficient estimate $\widehat{\beta}_{\text{eff},n}$.

From the motivating example above, we can conclude that the efficiency of many existing approaches can be improved when either misspecified treatment-free effect or heteroscedasticity happens. In fact, our example shows that misspecified treatment-free effect or heteroscedasticity can cause the dependency of $\mathbb{E}(\hat{e}^2|X) = 1 + c^2X^2$ on X. Motivated from efficient estimation under heteroscedasticity (Ma et al., 2006) and our motivating example, we introduce the working variance function $v_{\text{opt}}(X) = 1 + c^2X^2$, and consider the generalized least-squares estimate as in (3.3). The estimation efficiency can be greatly improved in this case.

Xiao et al. (2019, Theorem 6) pointed out a phenomenon similar to our finding in Section 3.2.2, while their methodology and theoretical properties differ from ours. To be specific, Xiao et al. (2019) replaced the squared loss by general robust loss functions. Under the assumption $\epsilon \perp\!\!\!\perp A|\boldsymbol{X}$, their estimate based on the quantile loss function can be shown consistent and $\sqrt{n}$-asymptotic normal. However, it remains unclear whether the $\sqrt{n}$-asymptotic normality still holds, and if so, how large the corresponding $\sqrt{n}$-asymptotic variance is, when treatment-free effect misspecification and heteroscedasticity exist. In contrast, we show in Theorem 3.10 that, under a more general setting, our proposed estimation strategy using the working variance function $\breve{v}_\epsilon(\boldsymbol{x})$ is optimal, with the smallest $\sqrt{n}$-asymptotic variance, for heteroscedastic and misspecified treatment-free effect models. This implies that E-Learning is more general with better optimality guarantee than Xiao et al. (2019).

The methodology introduced in this section is special in the sense that the treatment assignment is binary, *i.e.* $A \in \{0, 1\}$. For multiple treatment options $A \in \{1, 2, \cdots, K\}$ with $K > 2$,

the estimation problem is no longer an inverse-variance weighted least-squares problem. We will motivate our general methodology from the semiparametric efficient estimate of Model (3.1).

3.2.3 Semiparametric Efficient Estimate

In this section, we derive the semiparametric efficient estimate of $\boldsymbol{\beta}$ for Model (3.1). The efficient estimating function can be related to some existing methods in the literature. The connections are discussed in Sections 3.3.1 and 3.3.2.

3.2.3.1 Efficient Score

In order to obtain the corresponding estimating equation, we first show the procedures to calculate the semiparametric efficient score following Tsiatis (2007). To that end, we take the following steps to derive: 1) the nuisance tangent space; 2) the efficient score; 3) the efficient estimating function.

We first derive the nuisance tangent space with respect to η following Tsiatis (2007, Chapter 7). The same result was also used in Ma et al. (2006); Liang and Yu (2020).

Lemma 3.2 (Nuisance Tangent Space)**.** *Consider Model (3.1). Define* $\mathcal{H} := \{\boldsymbol{h}(\boldsymbol{X}, A, \epsilon) \mid \boldsymbol{h} : \mathcal{X} \times \mathcal{A} \times \mathbb{R} \to \mathbb{R}^p,\ \mathbb{E}\boldsymbol{h}(\boldsymbol{X}, A, \epsilon) = \mathbf{0},\ \mathbb{E}\|\boldsymbol{h}(\boldsymbol{X}, A, \epsilon)\|_2^2 < +\infty\}$*, which is equipped with the norm* $\|\cdot\| := (\mathbb{E}\|\cdot\|_2^2)^{1/2}$*. Then the nuisance tangent space is*

$$\Lambda = \Big\{\boldsymbol{H} \in \mathcal{H} : \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}, A) = \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X})\Big\}.$$

The proof of Lemma 3.2 is included in Section 3.8.

Next we discuss how to obtain the efficient score of Model (3.1). The efficient score is defined as the projection of the score vector onto the orthogonal complement $\Lambda^{\perp}$ of the nuisance tangent space. Notice that the moment restriction in Lemma 3.2 is equivalent to

$$\mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}, A = 1) = \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}, A = 2) = \cdots = \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}, A = K).$$

Then we can introduce a set of coding vectors $\{\boldsymbol{\omega}_k\}_{k=1}^K \subseteq \mathbb{R}^{K-1}$, such that $\sum_{k=1}^K c_k\boldsymbol{\omega}_k = \mathbf{0}$ if and only if $c_1 = c_2 = \cdots = c_K$. Equivalently, we can let $\Omega := \sqrt{1 - 1/K}[\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_K]^\intercal \in \mathbb{R}^{K\times(K-1)}$, and require that $(1/\sqrt{K})\mathbf{1}_{K\times 1}$ is the only left singular vector corresponding to the singular value 0

of Ω. In the following Lemma 3.3, we show that any coding vectors satisfying such a requirement are equiangular up to normalization.

Lemma 3.3 (Equiangularity)**.** *Let* $\Omega := \sqrt{1-1/K}[\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_K]^\intercal \in \mathbb{R}^{K\times(K-1)}$ *such that* $(1/\sqrt{K})\mathbf{1}_{K\times 1}$ *is the only left singular vector corresponding to the singular value* 0*. Then* $\{(\Omega^\intercal\Omega)^{-1/2}\boldsymbol{\omega}_k\}_{k=1}^K$ *are equiangular.*

The equiangular coding representation in Zhang and Liu (2014); Zhang et al. (2020); Qi et al. (2020) is an example that satisfies Lemma 3.3. The equiangular coding vectors $\{\boldsymbol{\omega}_k\}_{K=1}^K$ can be useful to define the following $\mathbb{R}^{K-1}$-valued decision function associated with the interaction effect.

Lemma 3.4 (Angle-Based Decision Function)**.** *Consider Model (3.1). For the coding vectors* $\{\boldsymbol{\omega}_k\}_{k=1}^K \subseteq \mathbb{R}^{K-1}$ *as in Lemma 3.3, define an* $\mathbb{R}^{K-1}$*-valued decision function* $\vec{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\beta}) := (\Omega^\intercal\Omega)^{-1}\sum_{k=1}^K \gamma(\boldsymbol{x},k;\boldsymbol{\beta})\boldsymbol{\omega}_k$*. Then we have*

$$\gamma(\boldsymbol{x},k;\boldsymbol{\beta}) = \left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_k, \vec{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\beta})\rangle; \quad 1\leqslant k\leqslant K.$$

Moreover, the optimal ITR is given by

$$d^\star(\boldsymbol{x}) \in \underset{1\leqslant k\leqslant K}{\operatorname{argmax}}\langle\boldsymbol{\omega}_k, \vec{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\beta})\rangle. \tag{3.4}$$

Without loss of generality, assume that $\|\boldsymbol{\omega}_k\|_2 = 1$ for $1\leqslant k\leqslant K$. For ease of notation, we denote $\vec{\boldsymbol{f}} = \vec{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\beta})$. Then the angle between $\boldsymbol{\omega}_k$ and $\vec{\boldsymbol{f}}$ satisfies $\cos\angle(\boldsymbol{\omega}_k, \vec{\boldsymbol{f}}) = \langle\boldsymbol{\omega}_k, \vec{\boldsymbol{f}}\rangle/\|\vec{\boldsymbol{f}}\|_2$. The decision rule (3.4) is equivalent to $\operatorname{argmin}_{1\leqslant k\leqslant K}\angle(\boldsymbol{\omega}_k, \vec{\boldsymbol{f}})$. That is, among K coding vectors $\{\boldsymbol{\omega}_k\}_{k=1}^K$, the decision function $\vec{\boldsymbol{f}}$ seeks for the arm that the corresponding coding vector has the least angle with respect to $\vec{\boldsymbol{f}}$.

Based on the coding vectors, the tangent space in Lemma 3.2 can be rewritten as

$$\Lambda = \left\{\boldsymbol{H}\in\mathcal{H} : \mathsf{O}_{p\times(K-1)} = \sum_{k=1}^K \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}, A=k)\boldsymbol{\omega}_k^\intercal = \mathbb{E}\left(\frac{\boldsymbol{H}\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})}\middle|\boldsymbol{X}\right)\right\}.$$

Then we can obtain $\Lambda^\perp$ and the projection operator onto it as in the following Lemma 3.5. For a vector $\boldsymbol{a}$, we denote $\boldsymbol{a}^{\otimes 2} := \boldsymbol{a}\boldsymbol{a}^\intercal$.

Lemma 3.5 (Projection onto $\Lambda^\perp$)**.** *Let Λ be the tangent space in Lemma 3.2, $\{\boldsymbol{\omega}_k\}_{k=1}^K \subseteq \mathbb{R}^{K-1}$ be the coding vectors satisfying $\sum_{k=1}^K c_k\boldsymbol{\omega}_k = \mathbf{0}$ if and only if $c_1 = c_2 = \cdots = c_K$. Then*

$$\Lambda^\perp = \left\{ \frac{\mathsf{H}(\boldsymbol{X})\boldsymbol{\omega}_A\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \mathsf{H} : \mathcal{X} \to \mathbb{R}^{p\times(K-1)} \right\}.$$

Furthermore, the projection operator onto $\Lambda^\perp$ is

$$\mathbb{E}(\boldsymbol{H}|\Lambda^\perp) = \mathbb{E}\left\{ \frac{\boldsymbol{H}\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\} \mathsf{V}_\epsilon(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})},$$

where $\mathsf{V}_\epsilon(\boldsymbol{X}) := \sum_{k=1}^K \frac{\sigma^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(k|\boldsymbol{X})} \in \mathbb{R}^{(K-1)\times(K-1)}$. Here, if $\mathsf{V}_\epsilon(\boldsymbol{X})$ is degenerate, then $\mathsf{V}_\epsilon(\boldsymbol{X})^{-1}$ represents its measurable generalized inverse.

The efficient score of the semiparametric model (3.1) is defined as the projection of the score vector, the gradient of the log-likelihood with respect to $\boldsymbol{\beta}$, onto $\Lambda^\perp$ (Tsiatis, 2007). Proposition 3.6 provides the explicit form of the efficient score.

Proposition 3.6 (Efficient Score)**.** *Consider Model (3.1), the coding vectors $\{\boldsymbol{\omega}_k\}_{k=1}^K \subseteq \mathbb{R}^{K-1}$ as in Lemma 3.3, and the angle-based representation in Lemma 3.4. The semiparametric efficient score is*

$$\boldsymbol{S}_{\text{eff}}(\boldsymbol{\beta}) = \dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal\Omega^\intercal\Omega\mathsf{V}_\epsilon(\boldsymbol{X})^{-1} \times \frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})} \times \epsilon,$$

where $\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta}) := (\partial/\partial\boldsymbol{\beta}^\intercal)\vec{\boldsymbol{f}}(\boldsymbol{X};\boldsymbol{\beta}) \in \mathbb{R}^{(K-1)\times p}$, and $\mathsf{V}_\epsilon(\boldsymbol{X})^{-1}$ is the same as in Lemma 3.5.

As a consequence of Proposition 3.6, we can finally define the efficient estimating function:

$$\begin{aligned}
&\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\breve{\mu}_0,\breve{p}_{\mathscr{A}},\breve{\sigma}^2) \\
&:= \underbrace{\left[Y - \breve{\mu}_0(\boldsymbol{X}) - \left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_A, \vec{\boldsymbol{f}}(\boldsymbol{X};\boldsymbol{\beta})\rangle \right]}_{\text{residual}} \times \underbrace{\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal\Omega^\intercal\Omega \left[\sum_{k=1}^K \frac{\breve{\sigma}^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A}}(k|\boldsymbol{X})} \right]^{-1} \frac{\boldsymbol{\omega}_A}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}}_{\text{efficient instrument}},
\end{aligned} \tag{3.5}$$

which depends on the nuisance functions $\breve{\mu}_0(\boldsymbol{X})$, $\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})$ and $\breve{\sigma}^2(\boldsymbol{X},A)$. In particular, $\boldsymbol{S}_{\text{eff}}(\boldsymbol{\beta}) = \boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\mu_0,p_{\mathscr{A}},\sigma^2)$. That is, if the parameters $\boldsymbol{\beta}$ of interest and all nuisance functions $(\mu_0,p_{\mathscr{A}},\sigma^2)$ match with the truth in Model (3.1), then the estimating function becomes the efficient score.

3.2.4 E-Learning

In Section 3.2.3, we have obtained the efficient estimating function $\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2)$ from (3.5). An E-Learning estimate of $\boldsymbol{\beta}$ solves

$$\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \widehat{\mu}_{0,n}, \widehat{p}_{\mathscr{A},n}, \widehat{\sigma}_n^2)] = \mathbf{0}, \tag{3.6}$$

where $(\widehat{\mu}_{0,n}, \widehat{p}_{\mathscr{A},n})$ are the finite-sample estimates of treatment-free effect and treatment assignment probability in Model (3.1). Furthermore, $\widehat{\sigma}_n^2(\boldsymbol{X}, A)$ is an estimate of the optimal variance function

$$\sigma^2_{\text{opt}}(\boldsymbol{X}, A; \widehat{\mu}_{0,n}) := [\widehat{\mu}_{0,n}(\boldsymbol{X}) - \mu_0(\boldsymbol{X})]^2 + \sigma^2(\boldsymbol{X}, A). \tag{3.7}$$

The optimality of $\sigma^2_{\text{opt}}(\boldsymbol{X}, A; \widehat{\mu}_{0,n})$ is justified in Theorem 3.10 in Section 3.4.1.2. However, (3.7) can depend on the true treatment-free effect function $\mu_0(\boldsymbol{X})$ and variance function $\sigma^2(\boldsymbol{X}, A)$, which are unknown. Motivated from the example in Section 3.2.2, we can consider the working residual $\widehat{e} := Y - \widehat{\mu}_{0,n}(\boldsymbol{X}) - \gamma(\boldsymbol{X}, A; \boldsymbol{\beta})$, such that $\mathbb{E}(\widehat{e}^2 | \boldsymbol{X}, A, \widehat{\mu}_{0,n}) = [\widehat{\mu}_{0,n}(\boldsymbol{X}) - \mu_0(\boldsymbol{X})]^2 + \sigma^2(\boldsymbol{X}, A) = \sigma^2_{\text{opt}}(\boldsymbol{X}, A; \widehat{\mu}_{0,n})$. Therefore, $\widehat{\sigma}_n^2(\boldsymbol{X}, A)$ can be obtained by regressing $\widehat{e}^2$ on $(\boldsymbol{X}, A)$.

Similar to the general methodology in Davidian and Carroll (1987), the E-Learning estimate of $\boldsymbol{\beta}$ can be solved by the following three steps:

Step 1. Obtain a consistent estimate $\widehat{\boldsymbol{\beta}}_n^{(0)}$ of $\boldsymbol{\beta}$. This can be done by solving (3.6) with $\widehat{\sigma}_n^{(0)2} = 1$ that results in a consistent estimate of $\boldsymbol{\beta}$. The consistency is guaranteed by Proposition 3.7;

Step 2. Obtain $\widehat{\sigma}_n^2(\boldsymbol{X}, A)$. Specifically, we first compute the working residual $\widehat{e} = Y - \widehat{\mu}_{0,n}(\boldsymbol{X}) - \gamma(\boldsymbol{X}, A; \widehat{\boldsymbol{\beta}}_n^{(0)})$, and then perform a nonparametric regression using $\widehat{e}^2$ as the response and $(\boldsymbol{X}, A)$ as the covariates to estimate the optimal working variance function;

Step 3. Solve (3.6) again using $\widehat{\sigma}_n^2(\boldsymbol{X}, A)$ from **Step 2** to obtain the E-Learning estimate $\widehat{\boldsymbol{\beta}}_n$.

More implementation details are discussed in Section 3.2.5 below.

3.2.5 Implementation

For the implementation of E-Learning, we first need to estimate the treatment assignment probabilities $\{p_{\mathscr{A}}(k|\boldsymbol{X})\}_{k=1}^K$ and the treatment-free effect $\mu_0(\boldsymbol{X})$. Then we follow the three-step procedures in Section 3.2.4 for E-Learning estimation.

3.2.5.1 Estimating the Propensity Score Function

Suppose the treatment assignment probability $p_{\mathscr{A}}$ is unknown. The first approach of estimating $p_{\mathscr{A}}$ is to consider the penalized multinomial logistic regression (Friedman et al., 2010). Specifically, consider the multinomial logistic working model $\breve{p}_{\mathscr{A}}(k|\boldsymbol{X};\tau_1,\tau_2,\cdots,\tau_K) := \frac{\exp(\tau_k^\intercal \boldsymbol{X})}{\sum_{k'=1}^K \exp(\tau_{k'}^\intercal \boldsymbol{X})}$. The propensity score parameters $\tau_1,\tau_2,\cdots,\tau_K \in \mathbb{R}^p$ can be estimated by the following penalized log-likelihood maximization:

$$\max_{\tau_1,\cdots,\tau_K\in\mathbb{R}^p} \left\{ \mathbb{E}_n \left[\sum_{k=1}^K \tau_k^\intercal \boldsymbol{X} \mathbb{1}(A=k) - \log\left(\sum_{k'=1}^K e^{\tau_{k'}^\intercal \boldsymbol{X}} \right) \right] - \lambda_{\mathscr{A}} \sum_{j=1}^p \left(\sum_{k=1}^K \tau_{jk}^2 \right)^{1/2} \right\},$$

where the group-LASSO penalty $\sum_{j=1}^p \left(\sum_{k=1}^K \tau_{jk}^2\right)^{1/2}$ takes $\{\tau_{jk}\}_{k=1}^K$ for the j-th variable across all treatments as a group, and $\lambda_{\mathscr{A}}$ is a tuning parameter and can be chosen using cross validation.

In observational studies, the propensity scores can be vulnerable to model misspecification. Another approach for estimating $p_{\mathscr{A}}$ is to consider flexible nonparametric regression using the regression forest (Athey et al., 2019). Specifically, for each $1 \leqslant k \leqslant K$, we run a regression forest using $\mathbb{1}(A=k)$ as the response and $\boldsymbol{X}$ as the covariates. Then each fitted regression forest provides a prediction for $\mathbb{E}[\mathbb{1}(A=k)|\boldsymbol{X}]$. The final estimate of $p_{\mathscr{A}}(k|\boldsymbol{X})$ is the prediction after normalization such that the summation over $k=1,\cdots,K$ is one.

3.2.5.2 Estimating the Treatment-Free Effect Function

Similar to Section 3.2.5.1, the treatment-free effect function μ_0 can be estimated from a parametric model or nonparametric regression. For parametric estimation, we consider the linear working model $\breve{\mu}_0(\boldsymbol{X};\boldsymbol{\eta}) = \boldsymbol{\eta}^\intercal \boldsymbol{X}$. In this case, the outcome mean model in (3.1) is fully parametrized. For example, if $\gamma(\boldsymbol{X},A;\mathsf{B}) = (1-1/K)\langle \boldsymbol{\omega}_A, \mathsf{B}^\intercal \boldsymbol{X}\rangle$, then we can consider the following joint penalized

inverse-probability weighted least-squares problem with the ℓ_1-penalty:

$$\min_{\boldsymbol{\eta}\in\mathbb{R}^p,\mathsf{B}\in\mathbb{R}^{p\times(K-1)}} \left\{ \mathbb{E}_n \left[\frac{1}{\widehat{p}_{\mathscr{A},n}(A|\boldsymbol{X})} \left(Y - \boldsymbol{\eta}^\intercal \boldsymbol{X} - \left(1-\frac{1}{K}\right) \langle \boldsymbol{\omega}_A, \mathsf{B}^\intercal \boldsymbol{X} \rangle \right)^2 \right] + \lambda_{\mu_0} \left(\|\boldsymbol{\eta}\|_1 + \|\mathsf{B}\|_1 \right) \right\},$$

where $\widehat{p}_{\mathscr{A},n}$ is the estimated treatment assignment probability, λ_{μ_0} is a tuning parameter and can be chosen using cross validation. Here, if $\widehat{p}_{\mathscr{A},n}(A|\boldsymbol{X})$ is the correct treatment assignment probability, then the above estimate for $\boldsymbol{\eta}$ can be consistent even if the model for the interaction effect $\gamma(\boldsymbol{X}, A; \boldsymbol{\beta})$ is incorrect. If the model for the interaction effect $\gamma(\boldsymbol{X}, A; \boldsymbol{\beta})$ is correct, then the above estimate for $\boldsymbol{\eta}$ can also be consistent for any arbitrary $\widehat{p}_{\mathscr{A},n}$ besides the correct one.

For nonparametric regression, we first divide the data into K subsets according to the received treatments. For each $1 \leqslant k \leqslant K$, we use Y as the response and $\boldsymbol{X}$ as the covariates to fit a regression forest on the data subset $\{(\boldsymbol{X}_i, Y_i) : A_i = k\}$. Then each fitted regression forest corresponds to the prediction of $\mathbb{E}(Y|\boldsymbol{X}, A = k)$. We average the predictions over $k = 1, \cdots, K$ to obtain the treatment-free effect estimate.

3.2.5.3 Estimating the Variance Function

Suppose $\widehat{e}$ is the working residual in **Step 2**. In order to estimate the variance function, we specifically consider the regression forest using $\widehat{e}^2$ as the response and $(\boldsymbol{X}, A)$ as the covariates. Then $\widehat{\sigma}_n^2(\boldsymbol{X}, k)$ is the regression forest prediction at $(\boldsymbol{X}, k)$ for $1 \leqslant k \leqslant K$.

In the simulation study in Section 3.5.3, we also study another two nonparametric regression methods, the *Multivariate Adaptive Regression Splines (MARS)* (Friedman, 1991) and the *COmponent Selection and Smoothing Operator (COSSO)* (Lin and Zhang, 2006). Here, the COSSO estimate of the working variance function is based on the following *Smoothing Spline ANalysis Of VAriance (SS-ANOVA)* model: $\mathbb{E}(\widehat{e}^2|\boldsymbol{X}, A) = \nu_0 + \sum_{j=1}^p f_j(X_j) + \sum_{k=1}^K \alpha_k + \sum_{j=1}^p \sum_{k=1}^K f_{jk}(X_j) + u$, where ν_0 is the global main effect, $\{f_j(X_j)\}_{j=1}^p$ are the covariate main effects, $\{\alpha_k\}_{k=1}^K$ are the treatment main effects, $\{f_{jk}(X_j)\}_{1\leqslant j\leqslant p, 1\leqslant k\leqslant K}$ are the covariate-treatment interaction effects, and u is the remainder term that is not modeled.

3.2.5.4 Solving the Regularized E-Learning Estimating Equation

In this section, we consider further regularization $J(\boldsymbol{\beta})$ on the parameters of interest. One example from Qi et al. (2020) is to consider the linear angle-based decision function $\vec{\boldsymbol{f}}(\boldsymbol{X};\mathsf{B}) = \mathsf{B}^\intercal \boldsymbol{X}$ in Lemma 3.4, where the covariate vector $\boldsymbol{X} \in \mathbb{R}^p$ can be high-dimensional. They introduced the row-wise group-LASSO penalty on the matrix coefficient $\mathsf{B} = [\beta_{jk}]_{p\times(K-1)} \in \mathbb{R}^{p\times(K-1)}$ as $J(\mathsf{B}) := \|\mathsf{B}\|_{2,1} = \sum_{j=1}^{p}(\sum_{k=1}^{K-1}\beta_{jk}^2)^{1/2}$, which encourages sparsity among input covariates. Another example can be the extension to nonlinear modeling of the decision function $\vec{\boldsymbol{f}}(\boldsymbol{X})$, where a functional penalty $J(\vec{\boldsymbol{f}})$ is applied.

To incorporate regularization in E-Learning from (3.6), we solve the penalized estimating equations (Johnson et al., 2008):

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\left\{\frac{1}{2}\left\|\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\widehat{\mu}_{0,n},\widehat{p}_{\mathscr{A},n},\widehat{\sigma}_n^2)]\right\|_{\mathsf{W}}^2 + \lambda J(\boldsymbol{\beta})\right\}, \tag{3.8}$$

where $\|\boldsymbol{x}\|_{\mathsf{W}}^2 := \boldsymbol{x}^\intercal\mathsf{W}\boldsymbol{x}$ with some weighting matrix $\mathsf{W} \in \mathbb{R}^{p\times p}$. A typical choice of W can be $\mathsf{I}_{p\times p}$ or the inverse of the empirical information matrix $\left\{\mathbb{E}_n[(\partial/\partial\boldsymbol{\beta}^\intercal)\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\widehat{\mu}_{0,n},\widehat{p}_{\mathscr{A},n},\widehat{\sigma}_n^2)]\right\}^{-1}$. Problem (3.8) can be solved by the accelerated proximal gradient method (Nesterov, 2013) with the gradient $\boldsymbol{\beta}\mapsto\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\widehat{\mu}_{0,n},\widehat{p}_{\mathscr{A},n},\widehat{\sigma}_n^2)]$. A comprehensive lists of the proximal operators on various penalties $J(\boldsymbol{\beta})$ can be found in Mo and Liu (2021). For a fixed λ, the estimation procedure follows the three steps in Section 3.2.4. The parameter λ can be further tuned by cross validation. The IPWE of the value function is used as the tuning criteria. Denote $\widehat{\boldsymbol{\beta}}_n(\lambda)$ as the solution to (3.8). The corresponding ITR becomes $\widehat{d}_n(\boldsymbol{X};\lambda) := \text{argmax}_{1\leqslant k\leqslant K}\gamma(\boldsymbol{X},k;\widehat{\boldsymbol{\beta}}_n(\lambda))$. Let $\{(\boldsymbol{X}_i,A_i,Y_i)\}_{i=1}^{n_{\text{valid}}}$ be the validation dataset. Then the criteria for λ is $\frac{1}{n_{\text{valid}}}\sum_{i=1}^{n_{\text{valid}}}\frac{\mathbb{1}[\widehat{d}_n(\boldsymbol{X}_i;\lambda)=A_i]}{\widehat{p}_{\mathscr{A},n}(A_i|\boldsymbol{X}_i)}Y_i$, which is larger the better.

More implementation details for E-Learning are discussed in Sections 3.8.3 and 3.8.4 in Section 3.8.

3.3 Connections to Existing Literature

In this section, we discuss the connection of the E-Learning estimating function (3.5) to several methods in the existing literature. It can be shown that with more assumptions in addition to Model

(3.1), several existing methods can be equivalent to (3.5). That is, E-Learning can incorporate these methods as special cases. The motivating example in Section 3.2.2 is such a special case. In Sections 3.3.1 and 3.3.2, we discuss the equivalence and the specific additional assumptions. In Section 3.3.3, we further provide the general comparisons for these methods and some other nonparametric methods in the literature.

3.3.1 Binary Treatment

We first consider the binary treatment case $K = 2$ and relate the efficient estimating function (3.5) to some existing methods. We follow the convention to denote $\mathcal{A} = \{0, 1\}$. Then we have one-dimensional coding for two treatment arms as ω_0, ω_1, which satisfies $c_0\omega_0 + c_1\omega_1 = 0$ if and only if $c_0 = c_1$. Then we have $\omega_1 = -\omega_0$. Without loss of generality, we can assume that $\omega_1 = 1$ and $\omega_0 = -1$, which become the sign coding of treatments. Then $\Omega^\intercal\Omega = 1$.

The variance matrix from Proposition 3.6 becomes a scalar: $v_\epsilon(\boldsymbol{X}) := \frac{\sigma^2(\boldsymbol{X},1)}{p_{\mathscr{A}}(1|\boldsymbol{X})} + \frac{\sigma^2(\boldsymbol{X},0)}{p_{\mathscr{A}}(0|\boldsymbol{X})}$. The decision function $f(\boldsymbol{X};\boldsymbol{\beta})$ is $\mathbb{R}$-valued, such that $\gamma(\boldsymbol{X}, A;\boldsymbol{\beta}) = (1/2)\omega_A f(\boldsymbol{X};\boldsymbol{\beta})$. Then the E-Learning efficient estimating function (3.5) becomes

$$\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\breve{\mu}_0,\breve{p}_{\mathscr{A}},\breve{\sigma}^2) = [Y - \breve{\mu}_0(\boldsymbol{X}) - (1/2)\omega_A f(\boldsymbol{X};\boldsymbol{\beta})]\frac{\breve{v}_\epsilon^{-1}(\boldsymbol{X})\omega_A}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}\dot{\mathbf{f}}(\boldsymbol{X};\boldsymbol{\beta}), \tag{3.9}$$

where $\dot{\mathbf{f}}(\boldsymbol{X};\boldsymbol{\beta}) := (\partial/\partial\boldsymbol{\beta})f(\boldsymbol{X};\boldsymbol{\beta}) \in \mathbb{R}^p$. Moreover, (3.9) is also equivalent to the following weighed least-squares problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}_n\left\{\frac{\breve{v}_\epsilon^{-1}(\boldsymbol{X})}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}[Y - \breve{\mu}_0(\boldsymbol{X}) - (1/2)\omega_A f(\boldsymbol{X};\boldsymbol{\beta})]^2\right\}. \tag{3.10}$$

There are some connections for this formulation to several methods in the existing literature.

Q-Learning Consider the additional assumptions: (a) homoscedasticity $\sigma^2(\boldsymbol{X}, 1) = \sigma^2(\boldsymbol{X}, 0) = \sigma^2$; and (b) complete-at-random treatment assignment $p_{\mathscr{A}}(1|\boldsymbol{X}) = p_{\mathscr{A}}(0|\boldsymbol{X}) = 1/2$. Then E-Learning (3.10) reduces to an OLS problem. If we also assume that: (c) the treatment-free effect satisfies $\mu_0(\boldsymbol{X}) = \boldsymbol{X}^\intercal(\boldsymbol{\eta} + \boldsymbol{\beta}/2)$, where $(\boldsymbol{\beta}, \boldsymbol{\eta})$ are jointly estimated, then E-Learning (3.10) can be

equivalent to the standard *Q-Learning* (Watkins, 1989) in this case:

$$\min_{\boldsymbol{\eta},\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}_n(Y - \boldsymbol{X}^\intercal\boldsymbol{\eta} - A\boldsymbol{X}^\intercal\boldsymbol{\beta})^2.$$

G-Estimation, A-Learning and dWOLS Consider the additional assumption: (a) homoscedasticity $\sigma^2(\boldsymbol{X},1) = \sigma^2(\boldsymbol{X},0) = \sigma^2$. Then $v_\epsilon^{-1}(\boldsymbol{X}) = \sigma^{-2}p_{\mathscr{A}}(1|\boldsymbol{X})p_{\mathscr{A}}(0|\boldsymbol{X})$. Without loss of generality, we can further assume that $\sigma^2 = 1$. Denote $\pi_{\mathscr{A}}(\boldsymbol{X}) := p_{\mathscr{A}}(1|\boldsymbol{X}) = \mathbb{E}(A|\boldsymbol{X})$. Then we have $\frac{v_\epsilon^{-1}(\boldsymbol{X})\omega_A}{p_{\mathscr{A}}(A|\boldsymbol{X})} = A - \pi_{\mathscr{A}}(\boldsymbol{X}) = |A - \pi_{\mathscr{A}}(\boldsymbol{X})|\omega_A$ and $\frac{v_\epsilon^{-1}(\boldsymbol{X})}{p_{\mathscr{A}}(A|\boldsymbol{X})} = |A - \pi_{\mathscr{A}}(\boldsymbol{X})|$.

Robins (2004) proposed the *G-Estimation* strategy for dynamic treatment regimes, which is equivalent to the standard *A-Learning* (Murphy, 2003) in the single-stage setting. In particular, G-Estimation solves the estimating equation

$$\mathbb{E}_n\Big\{ \big[Y - \widehat{\mu}_{0,n}(\boldsymbol{X}) - A\boldsymbol{X}^\intercal\boldsymbol{\beta}\big]\big[A - \widehat{\pi}_{\mathscr{A},n}(\boldsymbol{X})\big]\,\boldsymbol{X}\Big\} = \mathbf{0},$$

while A-Learning is equivalent to the estimating equation

$$\mathbb{E}_n\Big\{ \big[Y - \widehat{m}_{0,n}(\boldsymbol{X}) - \big(A - \widehat{\pi}_{\mathscr{A},n}(\boldsymbol{X})\big)\boldsymbol{X}^\intercal\boldsymbol{\beta}\big]\big[A - \widehat{\pi}_{\mathscr{A},n}(\boldsymbol{X})\big]\,\boldsymbol{X}\Big\} = \mathbf{0}.$$

Then G-Estimation and A-Learning are equivalent to E-Learning (3.9) in this case up to reparametrization, where $\widehat{\mu}_{0,n}(\boldsymbol{X})$ is replaced by $\widehat{m}_{0,n}(\boldsymbol{X}) - \widehat{\pi}_{\mathscr{A},n}(\boldsymbol{X})\boldsymbol{X}^\intercal\boldsymbol{\beta}$.

Wallace and Moodie (2015) proposed the dWOLS method. In the single-stage setting, they considered the following weighted least-squares problem:

$$\min_{\boldsymbol{\eta},\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}_n\Big\{w(\boldsymbol{X},A)(Y - \boldsymbol{X}^\intercal\boldsymbol{\eta} - A\boldsymbol{X}^\intercal\boldsymbol{\beta})^2\Big\},$$

where $w(\boldsymbol{X},A)$ satisfies the balancing condition $\pi_{\mathscr{A}}(\boldsymbol{X})w(\boldsymbol{X},1) = [1 - \pi_{\mathscr{A}}(\boldsymbol{X})]w(\boldsymbol{X},0)$. Note that $w(\boldsymbol{X},A) = |A - \pi_{\mathscr{A}}(\boldsymbol{X})|$ meets this balancing condition. Assume that: (b) the treatment assignment probability $\pi_{\mathscr{A}}(\boldsymbol{X}) = p_{\mathscr{A}}(1|\boldsymbol{X})$ is known; and (c) the treatment-free effect satisfies $\mu_0(\boldsymbol{X}) = \boldsymbol{X}^\intercal(\boldsymbol{\eta} + \boldsymbol{\beta}/2)$, where $(\boldsymbol{\beta},\boldsymbol{\eta})$ are jointly estimated. Then dWOLS with $w(\boldsymbol{X},A) = |A -$

$\pi_{\mathscr{A}}(\boldsymbol{X})|$ is equivalent to E-Learning (3.10):

$$\min_{\boldsymbol{\eta},\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}_n\Big\{|A-\pi_{\mathscr{A}}(\boldsymbol{X})|(Y-\boldsymbol{X}^\intercal\boldsymbol{\eta}-A\boldsymbol{X}^\intercal\boldsymbol{\beta})^2\Big\}.$$

Subgroup Identification, D-Learning and RD-Learning Consider the additional assumptions: (a) the variance function satisfies $v_\epsilon(\boldsymbol{X}) = \frac{\sigma^2(\boldsymbol{X},1)}{p_{\mathscr{A}}(1|\boldsymbol{X})} + \frac{\sigma^2(\boldsymbol{X},0)}{p_{\mathscr{A}}(0|\boldsymbol{X})} = v_\epsilon$, which is a constant; (b) the treatment assignment probability $p_{\mathscr{A}}(A|\boldsymbol{X})$ is known; and (c) the treatment-free effect satisfies $\mu_0(\boldsymbol{X}) = 0$. Then E-Learning (3.10) is equivalent to the standard *Subgroup Identification* (Tian et al., 2014; Chen et al., 2017) and the binary *D-Learning* (Qi and Liu, 2018):

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}_n\left\{\frac{1}{p_{\mathscr{A}}(A|\boldsymbol{X})}[Y-(1/2)\omega_A\boldsymbol{X}^\intercal\boldsymbol{\beta}]^2\right\}.$$

If both (b) and (c) are relaxed, then E-Learning (3.10) is equivalent to the augmented Subgroup Identification (Chen et al., 2017, Web Appendix B) and the binary *RD-Learning* (Meng and Qiao, 2020):

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}_n\left\{\frac{1}{\widehat{p}_{\mathscr{A},n}(A|\boldsymbol{X})}[Y-\widehat{\mu}_{0,n}(\boldsymbol{X})-(1/2)\omega_A\boldsymbol{X}^\intercal\boldsymbol{\beta}]^2\right\}.$$

3.3.2 Multiple Treatments and Partially Linear Model

For general $K \geqslant 3$, we consider the linear decision function $\vec{\boldsymbol{f}}(\boldsymbol{X};\mathsf{B}) = \mathsf{B}^\intercal\boldsymbol{X}$, where $\mathsf{B}\in\mathbb{R}^{p\times(K-1)}$ is a parameter matrix. By Lemma 3.4, Model (3.1) becomes

$$Y = \mu_0(\boldsymbol{X}) + \left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_A, \mathsf{B}^\intercal\boldsymbol{X}\rangle + \epsilon; \quad \mathbb{E}(\epsilon|\boldsymbol{X},A)=0; \quad \sigma^2(\boldsymbol{X},A)=\mathbb{E}(\epsilon^2|\boldsymbol{X},A)<+\infty, \tag{3.11}$$

which is a *Heteroscedasticitic Partially Linear Model (HPLM)* (Ma et al., 2006).

Denote $\mathsf{vec}(\mathsf{B}) \in \mathbb{R}^{p(K-1)}$ as the vectorization of B. The we further have $\vec{\boldsymbol{f}}(\boldsymbol{X};\mathsf{B}) = (\mathbf{I}_{(K-1)\times(K-1)}\otimes\boldsymbol{X})^\intercal\mathsf{vec}(\mathsf{B})$ and $\dot{\mathsf{F}}(\boldsymbol{X};\mathsf{B}) = [\partial/\partial\mathsf{vec}(\mathsf{B})^\intercal]\vec{\boldsymbol{f}}(\boldsymbol{X};\boldsymbol{\beta}) = (\mathbf{I}_{(K-1)\times(K-1)}\otimes\boldsymbol{X})^\intercal$, where

$\otimes$ denotes the Kronecker product. The E-Learning efficient estimating function (3.5) becomes

$$\boldsymbol{\phi}_{\text{eff}}(\mathsf{B};\breve{\mu}_0,\breve{p}_{\mathscr{A}},\breve{\sigma}^2)=\underbrace{[(\Omega^\intercal\Omega)\otimes\mathbf{I}_{d\times d}]}_{\text{constant matrix}}\times\left[Y-\breve{\mu}_0(\boldsymbol{X})-\left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_A,\mathsf{B}^\intercal\boldsymbol{X}\rangle\right]\times\frac{\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\boldsymbol{\omega}_A}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}\otimes\boldsymbol{X},\tag{3.12}$$

where $\breve{\mathsf{V}}_\epsilon(\boldsymbol{X}):=\sum_{k=1}^K\frac{\breve{\sigma}^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes2}}{\breve{p}_{\mathscr{A}}(k|\boldsymbol{X})}$, and $\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}$ denotes the generalized inverse if not invertible.

Consider the additional assumption: (a) the variance function satisfies $\mathsf{V}_\epsilon(\boldsymbol{X})=\sum_{k=1}^K\frac{\sigma^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes2}}{p_{\mathscr{A}}(k|\boldsymbol{X})}=\mathsf{V}_\epsilon$, which is a constant matrix. Then E-Learning (3.12) is equivalent to the multi-arm RD-Learning:

$$\min_{\mathsf{B}\in\mathbb{R}^{p\times(K-1)}}\mathbb{E}_n\left\{\frac{1}{\widehat{p}_{\mathscr{A},n}(A|\boldsymbol{X})}\left[Y-\widehat{\mu}_{0,n}(\boldsymbol{X})-\left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_A,\mathsf{B}^\intercal\boldsymbol{X}\rangle\right]^2\right\}.$$

Notice that the multi-arm D-Learning (Qi et al., 2020) cannot be equivalent to E-Learning. In fact, D-Learning solves the following vectorized least-squares problem:

$$\min_{\mathsf{B}\in\mathbb{R}^{p\times(K-1)}}\mathbb{E}_n\left\{\frac{1}{2Kp_{\mathscr{A}}(A|\boldsymbol{X})}\left\|KY\boldsymbol{\omega}_A-\mathsf{B}^\intercal\boldsymbol{X}\right\|_2^2\right\}.\tag{3.13}$$

The estimating function of (3.13) is

$$\underbrace{\left[Y-\left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_A,\mathsf{B}^\intercal\boldsymbol{X}\rangle\right]\times\frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}\otimes\boldsymbol{X}}_{\text{efficient estimating function if (a) and }\mu_0(\boldsymbol{X})=0}+\underbrace{\frac{1}{p_{\mathscr{A}}(A|\boldsymbol{X})}\left[\left(1-\frac{1}{K}\right)\boldsymbol{\omega}_A^{\otimes2}-\frac{1}{K}\mathsf{I}_{(K-1)\times(K-1)}\right]\mathsf{vec}(\boldsymbol{X}^{\otimes2}\mathsf{B})}_{:=\boldsymbol{\phi}_{\text{D}}(\boldsymbol{X},A)}.$$

Note that $\mathbb{E}[\boldsymbol{\phi}_{\text{D}}(\boldsymbol{X},A)|\boldsymbol{X}]=\mathbf{0}$ and $\mathbb{E}[\boldsymbol{\phi}_{\text{D}}(\boldsymbol{X},A)^{\otimes2}]$ is strictly positive definite, which contributes an extra term to the $\sqrt{n}$-asymptotic variance of the D-Learning estimate. This suggests that when $K\geqslant3$, the D-Learning estimate can generally have a larger asymptotic variance than E-Learning.

3.3.3 General Comparisons

In Table 3.1, we provide the comparisons of the methods discussed in Sections 3.3.1 and 3.3.2. We also compare several popular nonparametric approaches including *Outcome Weighted Learning (OWL)* (Zhao et al., 2012), *Residual Weighted Learning (RWL)* (Zhou et al., 2017; Liu et al., 2018), *Efficient Augmentation and Relaxation Learning (EARL)* (Zhao et al., 2019a), and Policy Learning

(Athey and Wager, 2021; Zhou et al., 2018b). In particular, EARL and Policy Learning utilize the AIPWE of the value function, which incorporates the outcome and propensity score models and is doubly robust. The listed methods are also compared in the simulation studies in Section 3.5.2.

Table 3.1: Comparisons of E-Learning with Several Existing Methods in the Literature

Method		**Nuisance Models**		**Doubly Robust**	**Assumptions for Being Optimal**			**Allow** $K \geqslant 3$
		Outcome	*Propensity*		*Treatment-Free Effect*	*Propensity*	*Variance*	
E-Learning		Yes	Yes	Yes	Arbitrary	Correct	Hetero.	Yes
Q-Learning		Yes	No	No	Correct	$1/K$	Homo.	Yes
G-Estimation		Yes	Yes	Yes	Correct	Correct	Homo.	No
A-Learning		Yes	Yes	Yes				No
dWOLS		Yes	Yes	Yes				No
Subgroup Identification	*Std.*	No	Yes	No	0	Known	Const.	Yes
	Aug.	Yes	Yes	No	Correct	Known	Const.	Yes
RD-Learning		Yes	Yes	Yes	Correct	Correct	Const.	Yes
D-Learning	$K = 2$	No	Yes	No	0	Known	Const.	Yes
	$K \geqslant 3$				N/A			
OWL		No	Yes	No				No
RWL		Yes	Yes	No				No
EARL		Yes	Yes	Yes				No
Policy Learning		Yes	Yes	Yes				Yes

[1] "Being optimal" is defined as the estimate of $\boldsymbol{\beta}$ in Model (3.1) achieves the smallest $\sqrt{n}$-asymptotic variance among the class of estimates in Definition 3.1.
[2] Methods of Subgroup Identification include the standard (`std.`) and augmented (`aug.`) versions.
[3] Variance assumptions are: `homo.` $\Leftrightarrow$ constant $\sigma^2(\boldsymbol{X}, A)$; `hetero.` $\Leftrightarrow$ general $\sigma^2(\boldsymbol{X}, A)$; `const.` $\Leftrightarrow$ $\mathsf{V}_\epsilon(\boldsymbol{X}) = \sum_{k=1}^{K} \frac{\sigma^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(k|\boldsymbol{X})}$ is a constant matrix.

We also discuss the estimation optimality for $\boldsymbol{\beta}$ in Table 3.1. Note that the nonparametric methods do not assume Model (3.1). Therefore, the estimation optimality for $\boldsymbol{\beta}$ is not available. In Theorem 3.10 in Section 3.4.1.2, we establish that the E-Learning estimate of $\boldsymbol{\beta}$ achieves the smallest $\sqrt{n}$-asymptotic variance among the class of estimates in Definition 3.1. This is also referred as "being optimal" in Table 3.1. Since the methods discussed in Sections 3.3.1 and 3.3.2, except for D-Learning with $K \geqslant 3$, are equivalent to E-Learning under specific additional assumptions, this also implies that the equivalent methods are optimal under those specific additional assumptions. However, this is not true for the general case. In contrast, our proposed E-Learning remains optimal under the most general scenario among all these methods.

3.4 Theoretical Properties

We investigate some theoretical properties of E-Learning. In particular, in Section 3.4.1, we establish estimation properties based on the efficient estimating function (3.5). In Section 3.4.2, we further relate the asymptotic properties to the regret bound of the estimated ITR.

3.4.1 Asymptotic Properties

We first focus on estimation properties of the proposed E-Learning. In Proposition 3.7, we show the double robustness property of the estimating function (3.5).

Proposition 3.7 (Double Robustness)**.** *Consider Model (3.1) and the estimating function (3.5). Suppose* $\breve{\mu}_0(\boldsymbol{X})$*,* $\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})$ *and* $\breve{\sigma}^2(\boldsymbol{X}, A)$ *are arbitrary nuisance functions. Then we have*

$$\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)] = \mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \mu_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2)] = \mathbf{0}.$$

If either $\breve{\mu}_0 = \mu_0$ or $\breve{p}_{\mathscr{A}} = p_{\mathscr{A}}$, then $\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2)] = \mathbf{0}$ at the true parameter $\boldsymbol{\beta}$ in Model (3.1). By assuming the positivity of the information matrix at $\boldsymbol{\beta}$ (Assumption 3.4.3), the consistency of $\widehat{\boldsymbol{\beta}}_n \in \operatorname{argmin}_{\boldsymbol{\beta}\in\mathscr{B}} \frac{1}{2} \left\| \mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\widehat{\boldsymbol{\beta}}_n; \breve{\mu}_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2)] \right\|_2^2$ can be established by the consistency of an M-estimator (van der Vaart and Wellner, 1996, Corollary 3.2.3). This implies the doubly robust property of $\widehat{\boldsymbol{\beta}}_n$. If $(\breve{\mu}_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2)$ are replaced by their finite-sample estimate $(\widehat{\mu}_{0,n}, \widehat{p}_{\mathscr{A},n}, \widehat{\sigma}_n^2)$, then Lemma 3.8 can be further applied to obtain consistency. Based on the connections from Section 3.3, Proposition 3.7 provides a more general framework to explain the double robustness property discussed in Robins (2004); Lu et al. (2013); Wallace and Moodie (2015); Meng and Qiao (2020).

Our next goal is to study how model specifications can affect estimation efficiency. In Section 3.4.1.1, we study the asymptotic properties of the parameter estimate under correctly specified models. In Section 3.4.1.2, we further consider the case of misspecified treatment-free effect, and show that there exists an optimal choice of the working variance function for efficiency improvement.

3.4.1.1 Correctly Specified Models

For simplicity, we assume that the treatment assignment probability $p_{\mathscr{A}}$ is known, so that the estimating function is consistent due to Proposition 3.7. This assumption can be relaxed to as-

suming a consistent estimate $\widehat{p}_{\mathscr{A},n}$ of $p_{\mathscr{A}}$, and the theoretical results can be extended following the cross-fitting argument in Ertefaie et al. (2021). For example, we can assume a correctly specified parametric model for $p_{\mathscr{A}}$.

We make additional assumptions on the squared integrability of Model (3.1) and the convergence of the plug-in treatment-free effect and variance function estimates. The estimated variance function $\widehat{\sigma}_n^2(\boldsymbol{X}, A)$ is furthered assumed uniformly bounded away from 0 to ensure that the smallest eigenvalue of $\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X}) = \sum_{k=1}^{K}[\widehat{\sigma}_n^2(\boldsymbol{X}, k)/p_{\mathscr{A}}(k|\boldsymbol{X})]\boldsymbol{\omega}_k^{\otimes 2}$ is uniformly bounded away from 0, so that the largest eigenvalue of $\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1}$ can be bounded from above. This can also be relaxed by considering a specific generalized inverse of $\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})$ to extend the theoretical results.

Assumption 3.1 (Treatment Assignment Probability)**.** The treatment assignment probability $p_{\mathscr{A}}$ is known, such that for some $\underline{p}_{\mathscr{A}} > 0$, we have $p_{\mathscr{A}}(a|\boldsymbol{x}) \geqslant \underline{p}_{\mathscr{A}}$ for all $\boldsymbol{x} \in \mathcal{X}$ and $a \in \mathcal{A}$.

Assumption 3.2 (Squared Integrability)**.** Consider Model (3.1) and the angle-based decision function $\vec{\boldsymbol{f}}(\boldsymbol{X}; \boldsymbol{\beta})$ in Lemma 3.4. We assume the following:

- $\mathbb{E}[\mu(\boldsymbol{X})^2] < +\infty$;
- $\mathbb{E}\sup_{\breve{\boldsymbol{\beta}} \in \mathscr{B}} \gamma(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})^2 < +\infty$;
- $\mathbb{E}(\epsilon^2) = \mathbb{E}\sigma^2(\boldsymbol{X}, A) < +\infty$;
- $\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}) = (\partial/\partial\boldsymbol{\beta}^{\intercal})\vec{\boldsymbol{f}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}) \in \mathbb{R}^{(K-1)\times p}$ exists for $\breve{\boldsymbol{\beta}} \in \mathscr{B}$, and $\mathbb{E}\sup_{\breve{\boldsymbol{\beta}} \in \mathscr{B}} \|\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})\|_2^2 < +\infty$, where $\|\cdot\|_2$ is the spectral norm on $\mathbb{R}^{(K-1)\times p}$.

Assumption 3.3 (Convergence of Plug-in Estimates)**.**

- There exists some $\breve{\mu}_0 : \mathcal{X} \to \mathbb{R}$, such that $\mathbb{E}[\breve{\mu}_0(\boldsymbol{X})^2] < +\infty$ and $(1/n)\sum_{i=1}^{n}[\widehat{\mu}_{0,n}(\boldsymbol{X}_i) - \breve{\mu}_0(\boldsymbol{X}_i)]^2 = o_{\mathbb{P}}(n^{-1})$.
- There exists some $0 < \underline{\sigma}^2 \leqslant \bar{\sigma}^2 < +\infty$ and $\breve{\sigma}^2 : \mathcal{X} \times \mathcal{A} \to \mathbb{R}_+$, such that $\underline{\sigma}^2 \leqslant \widehat{\sigma}_n^2(\boldsymbol{x}, a), \breve{\sigma}^2(\boldsymbol{x}, a) \leqslant \bar{\sigma}^2$, and $\|\widehat{\sigma}_n^2 - \breve{\sigma}^2\|_{\infty} = \sup_{\boldsymbol{x}\in\mathcal{X}, a\in\mathscr{A}} |\widehat{\sigma}_n^2(\boldsymbol{x}, a) - \breve{\sigma}^2(\boldsymbol{x}, a)| = o_{\mathbb{P}}(n^{-1/2})$.

Given Assumptions 3.1-3.3, we show in Lemma 3.8 that the plug-in estimating equation associated with (3.5) is $\sqrt{n}$-asymptotically equivalent to the limiting estimating equation.

Lemma 3.8 (Plug-in Estimating Equation)**.** *Consider Model (3.1) and the estimating function (3.5). Under Assumptions 3.1-3.3, we have*

$$\sup_{\breve{\boldsymbol{\beta}}\in\mathscr{B}} \mathbb{E}_n \left\| \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2) - \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2) \right\|_2 = o_{\mathbb{P}}(n^{-1/2}).$$

Lemma 3.8 implies that the plug-in estimates $(\widehat{\mu}_{0,n}, \widehat{\sigma}_n^2)$ do not affect the $\sqrt{n}$-asymptotic properties of the estimating function (3.5). Then we can show the asymptotic normality of $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ as the solution to $\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\widehat{\boldsymbol{\beta}}_{\text{eff},n}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)] = \mathbf{0}$ following the argument in Newey (1994). Moreover, if the treatment-free effect and the variance function are correctly specified, *i.e.*, $(\breve{\mu}_0, \breve{\sigma}^2) = (\mu_0, \sigma^2)$ in Model (3.1), then $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ is semiparametric efficient, in the sense that its $\sqrt{n}$-asymptotic variance achieves the semiparametric variance lower bound. We summarize the regularity conditions in Assumption 3.4.

Assumption 3.4 (Regularity Conditions)**.** Consider Model (3.1) and the angle-based representation in Lemma 3.4. We assume the following:

3.4.1 $\mathscr{B}$ is a compact subset in $\mathbb{R}^p$ and the true parameter $\boldsymbol{\beta} \in \mathring{\mathscr{B}}$, where $\mathring{\mathscr{B}}$ is the interior of $\mathscr{B}$;

3.4.2 $\ddot{\mathsf{F}}(\boldsymbol{x}; \breve{\boldsymbol{\beta}}) = (\partial/\partial\boldsymbol{\beta})\dot{\mathsf{F}}(\boldsymbol{x}; \breve{\boldsymbol{\beta}}) \in \mathbb{R}^{(K-1)\times p\times p}$ exists and satisfies $\mathbb{E}\sup_{\breve{\boldsymbol{\beta}}\in\mathscr{B}} \|\ddot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})\|_2^2 < +\infty$, where $\|\cdot\|_2$ is the operator norm of $(\mathbb{R}^p, \|\cdot\|_2) \to (\mathbb{R}^{(K-1)\times p}, \|\cdot\|_2)$;

3.4.3 Define $\breve{\mathsf{V}}_\epsilon(\boldsymbol{X}) := \sum_{k=1}^K \frac{\breve{\sigma}^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A}}(k|\boldsymbol{X})}$ and $\breve{\mathcal{I}}(\boldsymbol{\beta}) := \mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \Omega^\intercal \Omega \dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right]$, where $\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}$ denotes the generalized inverse if not invertible. Assume $\breve{\mathcal{I}}(\boldsymbol{\beta})$ is positive definite;

3.4.4 The true parameter $\boldsymbol{\beta}$ is a unique solution to $\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)] = \mathbf{0}$.

Note that the definition of $\breve{\mathcal{I}}(\boldsymbol{\beta})$ only depends on the working variance function $\breve{\sigma}^2$ through $\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})$. We denote $\breve{\mathcal{I}}$ to reflect that it depends on $\breve{\sigma}^2$. It can be shown that for any $\breve{\boldsymbol{\beta}} \in \mathscr{B}$, we have $\breve{\mathcal{I}}(\breve{\boldsymbol{\beta}}) = \mathbb{E}[-(\partial/\partial\boldsymbol{\beta}^\intercal)\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)]$. In Theorem 3.9, we establish the semiparametric efficiency of E-Learning. For symmetric matrices A and B, the matrix inequality $\mathsf{A} \preccurlyeq \mathsf{B}$ means that $\mathsf{B} - \mathsf{A}$ is positive semi-definite.

Theorem 3.9 (Semiparametric Efficiency under Correct Specification)**.** *Consider Model (3.1) and the angle-based representation in Lemma 3.4. Suppose $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ is the solution to the estimating equa-*

tion $\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\widehat{\boldsymbol{\beta}}_{\text{eff},n}; \widehat{\mu}_n, p_{\mathscr{A}}, \widehat{\sigma}_n^2)] = \mathbf{0}$ *from (3.6). Then under Assumptions 3.1-3.4, we have*

$$\widehat{\boldsymbol{\beta}}_{\text{eff},n} - \boldsymbol{\beta} = \breve{\mathcal{I}}(\boldsymbol{\beta})^{-1}\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)] + o_{\mathbb{P}}(n^{-1/2}).$$

Moreover, if $(\breve{\mu}_0, \breve{\sigma}^2) = (\mu_0, \sigma^2)$*, then* $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ *is semiparametric efficient, in the sense that for any other Regular and Asymptotic Linear (RAL) estimate* $\widehat{\boldsymbol{\beta}}_n$*, we have*

$$\mathcal{I}(\boldsymbol{\beta})^{-1} = \lim_{n\to\infty} n\mathsf{Var}(\widehat{\boldsymbol{\beta}}_{\text{eff},n}) \leqslant \lim_{n\to\infty} n\mathsf{Var}(\widehat{\boldsymbol{\beta}}_n),$$

where $\mathcal{I}(\boldsymbol{\beta}) := \mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \mathsf{V}_\epsilon(\boldsymbol{X})^{-1} \Omega^\intercal \Omega \dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right]$ *is the semiparametric information matrix.*

For specific parametric models of $\vec{f}(\boldsymbol{X};\boldsymbol{\beta})$, the information matrix can be simplified. In the binary treatment case discussed in Section 3.3.1, we have $v_\epsilon(\boldsymbol{X}) = \frac{\sigma^2(\boldsymbol{X},1)}{p_{\mathscr{A}}(1|\boldsymbol{X})} + \frac{\sigma^2(\boldsymbol{X},0)}{p_{\mathscr{A}}(0|\boldsymbol{X})}$, which becomes a scalar weight. It is shown that E-Learning is equivalent to the generalized least-squares problem (3.10) with the weight $v_\epsilon^{-1}(\boldsymbol{X})p_{\mathscr{A}}(A|\boldsymbol{X})^{-1}$, which is also the overlap weight under heteroscedasticity (Crump et al., 2006; Li and Li, 2019). Then the information is $\mathcal{I}(\boldsymbol{\beta}) = \mathbb{E}[v_\epsilon(\boldsymbol{X})^{-1}\dot{\mathbf{f}}(\boldsymbol{X};\boldsymbol{\beta})^{\otimes 2}]$, where $\dot{\mathbf{f}}(\boldsymbol{X};\boldsymbol{\beta}) := (\partial/\partial\boldsymbol{\beta})f(\boldsymbol{X};\boldsymbol{\beta})$. For HPLM (3.11) in the multiple treatment case (Section 3.3.2), the information matrix becomes $\mathcal{I}(\mathsf{B}) = \mathbb{E}\left[\mathsf{V}_\epsilon(\boldsymbol{X})^{-1} \otimes \boldsymbol{X}^{\otimes 2}\right]$.

In Theorem 3.9, if $\breve{\mu}_0 \neq \mu_0$, then $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ is not semiparametric efficient. A natural question is to ask whether there exists some $\breve{\sigma}^2$ such that $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ is still "optimal" in certain sense. This motivates our discussion in Section 3.4.1.2.

3.4.1.2 Misspecified Treatment-Free Effect Model

Going beyond the double robustness and semiparametric efficiency of the estimating function (3.5), we are further interested in certain optimality when misspecified treatment-free effect happens. Specifically, we first define the *regular* class of semiparametric estimates of $\boldsymbol{\beta}$.

Definition 3.1 (Regular Class of Semiparametric Estimates)**.** Denote $\widehat{\boldsymbol{\beta}}_n = \widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0)$ as an estimate based on n observations independent and identically distributed from Model (3.1), and take the working treatment-free effect function $\breve{\mu}_0$ as its input. We define a regular class of semiparametric estimates $\mathcal{B}_n(\breve{\mu}_0)$ as follows. For any $\widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0) \in \mathcal{B}_n(\breve{\mu}_0)$, there exists some $\boldsymbol{h} : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^p$, which can depend on $(\boldsymbol{\beta}, \breve{\eta})$, such that:

- The estimate $\widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0)$ corresponds to the estimating function

$$\boldsymbol{\phi}(\boldsymbol{\beta}; \breve{\mu}_0) = [Y - \breve{\mu}_0(\boldsymbol{X}) - \gamma(\boldsymbol{X}, A; \boldsymbol{\beta})]\boldsymbol{h}(\boldsymbol{X}, A; \boldsymbol{\beta}, \breve{\eta}).$$

  That is, $\mathbb{E}_n[\boldsymbol{\phi}(\widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0); \breve{\eta})] = \mathbf{0}$;

- (Consistency) $\mathbb{E}[\boldsymbol{h}(\boldsymbol{X}, A; \boldsymbol{\beta}, \breve{\eta})|\boldsymbol{X}] = \mathbf{0}$.

Note that the consistency condition is equivalent to $\mathbb{E}[\boldsymbol{\phi}(\boldsymbol{\beta}; \breve{\mu}_0)] = \mathbf{0}$ for any $\breve{\mu}_0 : \mathcal{X} \to \mathbb{R}$. This can be concluded from that $\mathbb{E}[\boldsymbol{h}(\boldsymbol{X}, A; \boldsymbol{\beta}, \breve{\eta})|\boldsymbol{X}] = \mathbf{0}$ if and only if for any $\breve{\mu}_0$, we have $\mathbf{0} = \mathbb{E}[\boldsymbol{\phi}(\boldsymbol{\beta}; \breve{\mu}_0)] = \mathbb{E}\Big\{[\mu_0(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})]\mathbb{E}[\boldsymbol{h}(\boldsymbol{X}, A; \boldsymbol{\beta}, \breve{\eta})|\boldsymbol{X}]\Big\}$. The consistency can be met by any doubly robust estimates with a correct propensity score, such as G-Estimation, A-Learning, dWOLS, and RD-Learning.

If $\breve{\mu}_0$ is the true treatment-free effect μ_0 in Model (3.1), then by Tsiatis (2007, Theorem 4.2), any semiparametric RAL estimate of $\boldsymbol{\beta}$ must have an influence function in the form of $\boldsymbol{\phi}(\boldsymbol{\beta}; \mu_0)$ in Definition 3.1. That is, for any RAL estimate $\widetilde{\boldsymbol{\beta}}_n$, there exists some $\widehat{\boldsymbol{\beta}}_n(\mu_0) \in \mathcal{B}_n(\mu_0)$, such that $\widetilde{\boldsymbol{\beta}}_n = \widehat{\boldsymbol{\beta}}_n(\mu_0) + o_{\mathbb{P}}(n^{-1/2})$ under Model (3.1). Therefore, $\mathcal{B}_n(\mu_0)$ can represent the equivalent classes of RAL estimates, where two RAL estimates are "equivalent" if and only if their $\sqrt{n}$-asymptotic variances are the same. In particular, $\mathcal{B}_n(\mu_0)$ consists of the "regular versions" such that their estimating functions coincide with their IFs.

Definition 3.1 provides a useful class of estimates with a specific form of dependency on the working treatment-free effect function $\breve{\mu}_0$. In fact, the following Theorem 3.10 shows that, given a working treatment-free effect function $\breve{\mu}_0$, there exists some optimal RAL estimate among the regular class $\mathcal{B}_n(\breve{\mu}_0)$, in the sense that its $\sqrt{n}$-asymptotic variance is the smallest.

Theorem 3.10 (Optimal Efficiency Improvement under Misspecification)**.** *Given a working treatment-free effect function $\breve{\mu}_0 : \mathcal{X} \to \mathbb{R}$, consider Model (3.1) and the regular class of semiparametric estimates $\mathcal{B}_n(\breve{\mu}_0)$ in Definition 3.1. Define*

$$\sigma^2_{\text{opt}}(\boldsymbol{X}, A; \breve{\mu}_0) := [\breve{\mu}_0(\boldsymbol{X}) - \mu_0(\boldsymbol{X})]^2 + \sigma^2(\boldsymbol{X}, A),$$

and $\widehat{\boldsymbol{\beta}}_{\text{eff},n}(\breve{\mu}_0) \in \mathcal{B}_n(\breve{\mu}_0)$ *as the solution to* $\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\widehat{\boldsymbol{\beta}}_{\text{eff},n}(\breve{\mu}_0); \breve{\mu}_0, p_{\mathscr{A}}, \sigma^2_{\text{opt}})] = \mathbf{0}$ *from (3.6). Then under Assumptions 3.1-3.4, we have*

$$\mathcal{I}(\boldsymbol{\beta}; \breve{\mu}_0)^{-1} = \lim_{n\to\infty} n\mathsf{Var}[\widehat{\boldsymbol{\beta}}_{\text{eff},n}(\breve{\mu}_0)] \leqslant \lim_{n\to\infty} n\mathsf{Var}[\widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0)]; \quad \forall \widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0) \in \mathcal{B}_n(\breve{\mu}_0),$$

where $\mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0) := \sum_{k=1}^K \frac{\sigma^2_{\text{opt}}(\boldsymbol{X}, A; \breve{\mu}_0)\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(k|\boldsymbol{X})}$ *and* $\mathcal{I}(\boldsymbol{\beta}; \breve{\mu}_0) := \mathbb{E}\Big[\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0)^{-1} \Omega^\intercal \Omega \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})\Big]$.

Note that Theorem 3.10 can be more general than the semiparametric efficiency in Theorem 3.9, in the sense that the optimality in Theorem 3.10 is for a general working treatment-free effect function. Specifically, if $\breve{\mu}_0 = \mu_0$, then $\mathcal{B}_n(\mu_0)$ in Definition 3.1 represents the equivalent classes of RAL estimates with the $\sqrt{n}$-asymptotic variance as the equivalence relationship. In that case, Theorem 3.10 recovers Theorem 3.9 that $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ has the smallest $\sqrt{n}$-asymptotic variance. As a remark, we would like to point out that Theorem 3.10 can be extended to the estimating equation $\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\widehat{\boldsymbol{\beta}}_{\text{eff},n}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}^2_n)] = \mathbf{0}$ with plug-in nuisance function estimates $(\widehat{\mu}_{0,n}, \widehat{\sigma}^2_n)$. The argument is similar to Theorem 3.9, and we omit the details here.

If the working treatment-free effect function $\breve{\mu}_0$ is not identical to the true treatment-free effect function μ_0 in Model (3.1), then Theorem 3.10 suggests an optimal variance function $\sigma^2_{\text{opt}}(\boldsymbol{X}, A; \breve{\mu}_0)$. For the binary treatment case, the optimal working variance function can correspond to

$$v_\epsilon(\boldsymbol{x}; \breve{\mu}_0) = \frac{\sigma^2_{\text{opt}}(\boldsymbol{x}, 1; \breve{\mu}_0)}{p_{\mathscr{A}}(1|\boldsymbol{x})} + \frac{\sigma^2_{\text{opt}}(\boldsymbol{x}, 0; \breve{\mu}_0)}{p_{\mathscr{A}}(0|\boldsymbol{x})} = \frac{\sigma^2(\boldsymbol{x}, 1)}{p_{\mathscr{A}}(1|\boldsymbol{x})} + \frac{\sigma^2(\boldsymbol{x}, 0)}{p_{\mathscr{A}}(0|\boldsymbol{x})} + \frac{[\breve{\mu}_0(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]^2}{p_{\mathscr{A}}(1|\boldsymbol{x})p_{\mathscr{A}}(0|\boldsymbol{x})}.$$

The corresponding generalized least-squares estimate from (3.10) can achieve the smallest $\sqrt{n}$-asymptotic variance among the regular class of estimates $\mathcal{B}_n(\breve{\mu}_0)$. The motivating example in Section 3.2.2 is a special case when we further assume $p_{\mathscr{A}}(1|\boldsymbol{X}) = p_{\mathscr{A}}(0|\boldsymbol{X}) = 1/2$.

Remark 3.1 (General Asymptotic Variance)**.** It can be useful to compute the $\sqrt{n}$-asymptotic variance for arbitrary working treatment-free effect and variance function $(\breve{\mu}_0, \breve{\sigma}^2)$. Suppose $\widehat{\boldsymbol{\beta}}_n$ is the solution to $\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\widehat{\boldsymbol{\beta}}_n; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)] = \mathbf{0}$. Then we have

$$\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)^{\otimes 2}] = \mathbb{E}\Big[\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0) \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \Omega^\intercal \Omega \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})\Big], \tag{3.14}$$

and $\mathbb{E}[-(\partial/\partial\boldsymbol{\beta}^\intercal)\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2)] = \breve{\mathcal{I}}(\boldsymbol{\beta})$. The $\sqrt{n}$-asymptotic variance is given by the sandwich form $\lim_{n\to\infty} n\mathsf{Var}(\widehat{\boldsymbol{\beta}}_n) = \breve{\mathcal{I}}(\boldsymbol{\beta})^{-1}[(3.14)]\breve{\mathcal{I}}(\boldsymbol{\beta})^{-1}$.

Remark 3.2 (Incorrect Propensity Score)**.** In our theoretical analysis, we assume that the propensity score is known or can be consistently estimated. In Section 3.8.2 in Section 3.8, we further discuss the case when the propensity score is incorrect. Although the optimality of Theorem 3.10 cannot be recovered in this case, the covariate-dependent variance adjustment for the optimal working variance function $\breve{\sigma}^2_{\text{opt}}(\boldsymbol{X},A)$ can still be helpful. We demonstrate in our simulation studies (Section 3.5) that E-Learning still outperforms other methods even with incorrect propensity.

In Theorem 3.10, we establish the optimality of using the working variance function $\sigma^2_{\text{opt}}(\boldsymbol{X},A;\breve{\mu}_0)$ in the proposed E-Learning. As discussed in Section 3.2.4, the optimal working variance function can be identified by the expectation of the squared working residual. This can confirm the optimality of the E-Learning estimate.

3.4.2 Regret Bound

In this section, we relate the theoretical results for estimation in Section 3.4.1 to the regret bound for the estimated ITR. Recall from Theorem 3.1 that the estimation error of the interaction effect can dominate the regret. We further make compactness assumption on covariates to establish the regret bound.

Assumption 3.5 (Compact Covariate Domain)**.** The support of the distribution $p_{\mathscr{X}}(\boldsymbol{x})$ is compact.

Theorem 3.11 (Regret Bound for RAL Estimate)**.** *Consider Model (3.1) and the angle-based representation in Lemma 3.4. Suppose* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0},\Sigma)$ *for some* $\Sigma > 0$. *Define* $\widehat{d}_n(\boldsymbol{x}) := \text{argmax}_{1\leqslant k\leqslant K}\langle\boldsymbol{\omega}_k, \vec{\boldsymbol{f}}(\boldsymbol{x};\widehat{\boldsymbol{\beta}}_n)\rangle$ *and* $d^\star(\boldsymbol{x}) := \text{argmax}_{1\leqslant k\leqslant K}\langle\boldsymbol{\omega}_k, \vec{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\beta})\rangle$. *Then under Assumptions 3.2, 3.4.2 and 3.5, we have*

$$\begin{aligned}\limsup_{n\to\infty}\sqrt{n}[\mathcal{V}(d^\star) - \mathbb{E}\mathcal{V}(\widehat{d}_n)] &\leqslant 2\lim_{n\to\infty}\left\{n\sum_{k=1}^{K}\mathbb{E}[\gamma(\boldsymbol{X},k;\widehat{\boldsymbol{\beta}}_n) - \gamma(\boldsymbol{X},k;\boldsymbol{\beta})]^2\right\}^{1/2}\\ &= 2\left(1-\frac{1}{K}\right)^{1/2}\mathsf{Tr}\left\{\mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right]\Sigma\right\}^{1/2}.\end{aligned}$$

The regret bound in Theorem 3.11 can be tight compared to Theorem 3.1, since Theorem 3.11 only relaxes the absolute estimation error to the squared estimation error, and the maximization to the summation. Theorem 3.11 further implies that the regret bound and the estimation error are both in $\sqrt{n}$-order, where the leading constant depends on the $\sqrt{n}$-asymptotic variance Σ of the estimated parameters $\widehat{\boldsymbol{\beta}}_n$. In particular, denote $\|\cdot\|_{\mathrm{F}}$ as the Frobenius norm. Then we have

$$\mathsf{Tr}\left\{\mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^{\intercal}\Omega^{\intercal}\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right]\Sigma\right\} \leqslant \left\|\mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^{\intercal}\Omega^{\intercal}\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right]\right\|_{\mathrm{F}} \times \|\Sigma\|_{\mathrm{F}},$$

with equality if $\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})$ contains $\boldsymbol{X}$ and we take supremum over all possible covariate distribution $p_{\mathscr{X}}$ on both sides. This suggests that an RAL estimate with the smallest $\sqrt{n}$-asymptotic variance Σ can achieve the minimal regret bound. This complements the theoretical results in Sections 3.4.1.1 and 3.4.1.2 that establish the optimality of E-Learning estimate of $\boldsymbol{\beta}$ in terms of the $\sqrt{n}$-asymptotic variance. In particular, if we use the efficient estimate $\widehat{\boldsymbol{\beta}}_{\mathrm{eff},n}(\breve{\mu}_0)$ with the optimal choice of working variance function $\sigma^2_{\mathrm{opt}}(\boldsymbol{X}, A; \breve{\mu}_0)$, then the $\sqrt{n}$-asymptotic variance Σ becomes $\mathcal{I}(\boldsymbol{\beta}; \breve{\mu}_0)^{-1}$, and the regret bound above is the smallest among all RAL estimates in $\mathcal{B}(\breve{\mu}_0)$.

To conclude this section, we have established that E-Learning is doubly robust and optimal with the smallest $\sqrt{n}$-asymptotic variance among the class of regular semiparametric estimates in Definition 3.1, which can allow multiple treatments, heteroscedasticity and misspecified treatment-free effect. The corresponding regret bound can also have an optimal leading constant in the $n^{-1/2}$-order.

3.5 Simulation Study

We consider several simulation studies to compare the proposed E-Learning with existing methods from the literature and demonstrate the superiority of E-Learning.

3.5.1 Data Generating Process and Model Specifications

The synthetic data generation process is as follows. Let $n \in \{100, 200, 400, 800, 1600\}$ be the training sample size, $p \in \{10, 50, 100\}$ be the number of variables, and $K \in \{2, 3, 5, 7\}$ be the number of treatments. First, we generate the coefficients of the treatment-covariate-interaction effect by $(\widetilde{\beta}_{0k}, \widetilde{\beta}_{1k}, \widetilde{\beta}_{2k}, \widetilde{\beta}_{3k}, \widetilde{\beta}_{4k}, \widetilde{\beta}_{5k}) \sim \mathsf{Uniform}\{\boldsymbol{u} \in \mathbb{R}^6 : \|\boldsymbol{u}\|_2 = 1\}$ independently for $1 \leqslant k \leqslant K$, $\beta_{jk} :=$

$\widetilde{\beta}_{jk} - \frac{1}{K}\sum_{k'=1}^{K}\widetilde{\beta}_{jk'}$ for $0 \leqslant j \leqslant 5$, and $\boldsymbol{\beta}_k := (\beta_{0k}, \beta_{1k}, \beta_{2k}, \beta_{3k}, \beta_{4k}, \beta_{5k}, \overbrace{0, \cdots, 0}^{p-5})^\intercal$ for $1 \leqslant k \leqslant K$. Then we generate the data from:

$$\boldsymbol{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_{p\times p}); \quad \mathbb{P}(A = k|\boldsymbol{X}) = \underbrace{p_{\mathscr{A}}(k|\boldsymbol{X})}_{\text{propensity score}} = \frac{e(\boldsymbol{X}, k)}{\sum_{k'=1}^{K} e(\boldsymbol{X}, k')}; \quad 1 \leqslant k \leqslant K;$$

$$Y|(\boldsymbol{X}, A) = \underbrace{\mu_0(\boldsymbol{X}) - \mathbb{E}[\mu_0(\boldsymbol{X})]}_{\text{treatment-free effect}} + \underbrace{\boldsymbol{\beta}_A^\intercal(1, \boldsymbol{X}^\intercal)^\intercal}_{\text{interaction effect}} + \underbrace{\sigma(\boldsymbol{X}, A)}_{\text{variance function}} \times \mathcal{N}(0, 1).$$

For the coefficient vectors $\{\boldsymbol{\beta}_k\}_{k=1}^K$, the optimal ITR is $d^\star(\boldsymbol{x}) = \text{argmax}_{1\leqslant k\leqslant K} \boldsymbol{\beta}_k^\intercal(1, \boldsymbol{x}^\intercal)^\intercal$. Here, the true treatment-free effect function $\mu_0(\boldsymbol{x})$, variance functions $\{\sigma^2(\boldsymbol{x}, k)\}_{k=1}^K$ and the propensity score functions $\{e(\boldsymbol{x}, k)\}_{k=1}^K$ are defined according to Table 3.2.

Table 3.2: True Models and the Implying Model Specifications in the Simulation Studies

		Correctly Specified		*Misspecified*		***Treatment-Free Effect***		
		Homo-scedastic	*Hetero-scedastic*	*Homo-scedastic*	*Hetero-scedastic*	***Variance***		
Truth	$\mu_0(\boldsymbol{x}) =$	$\frac{1}{\sqrt{K}}\sum_{k'=1}^{K} x_{k'}$		$\frac{1}{K}\sum_{k'=1}^{K} e^{\sqrt{2}x_{k'}}$		$e(\boldsymbol{x}, k) = e^{x_k/2}$	*Correctly Specified*	***Propensity Score***
	$\sigma^2(\boldsymbol{x}, k) =$	1	$e^{2\sqrt{2}x_k}$	1	$e^{2\sqrt{2}x_k}$			
	$\mu_0(\boldsymbol{x}) =$	$\frac{1}{\sqrt{K}}\sum_{k'=1}^{K} x_{k'}$		$\frac{1}{K}\sum_{k'=1}^{K} e^{\sqrt{2}x_{k'}}$		$e(\boldsymbol{x}, k) = \lvert x_k\rvert^{1/2}$	*Mis-specified*	
	$\sigma^2(\boldsymbol{x}, k) =$	1	$e^{2\sqrt{2}x_k}$	1	$e^{2\sqrt{2}x_k}$			

[1] The treatment-free effect is estimated by a linear working model.
[2] The propensity score is estimated by a multinomial logistic working model.

When estimating the treatment-free effect $\mu_0(\boldsymbol{X}) - \mathbb{E}[\mu_0(\boldsymbol{X})]$, we consider a linear working model $\breve{\mu}_0(\boldsymbol{X}; \eta) = \boldsymbol{\eta}^\intercal(1, \boldsymbol{X}^\intercal)^\intercal$. Then the treatment-free effect model is correctly specified if the truth is $\mu_0(\boldsymbol{x}) = \frac{1}{\sqrt{K}}\sum_{k'=1}^{K} x_{k'}$, while misspecified if the truth is $\mu_0(\boldsymbol{x}) = \frac{1}{K}\sum_{k'=1}^{K} e^{\sqrt{2}x_{k'}}$. In Figure 3.5 in Section 3.8, we provide the fitted treatment-free effect plots when the model is correctly and incorrectly specified. It shows that the estimated treatment-free effect is consistent if correctly specified, and deviates from the truth if misspecified. When estimating the propensity score functions $\{p_{\mathscr{A}}(k|\boldsymbol{X})\}_{k=1}^K$, we consider a multinomial logistic working model $\breve{p}_{\mathscr{A}}(k|\boldsymbol{X}; \tau_1, \tau_2, \cdots, \tau_K) = \frac{\exp[\tau_k^\intercal(1, \boldsymbol{X}^\intercal)^\intercal]}{\sum_{k'=1}^{K}\exp[\tau_{k'}^\intercal(1, \boldsymbol{X}^\intercal)^\intercal]}$. Then the propensity score model is correctly specified if the truth is generated from $e(\boldsymbol{x}, k) = e^{x_k/2}$, while misspecified if the truth is generated from $e(\boldsymbol{x}, k) = |x_k|^{1/2}$. In Figure 3.6 in Section 3.8, we provide the fitted propensity score plots when the

model is correctly and incorrectly speceified, and demonstrate how the misspecified model affects the fitted propensity scores. As discussed in Section 3.2.4, if one of or both misspecified treatment-free effect model and heteroscedasticity exist, the squared residuals can depend on $(\boldsymbol{X}, A)$. In Figures 3.7-3.10, we provide the residual plots in all these cases to demonstrate such dependencies.

3.5.2 Binary Treatments

In this section, we consider the binary treatment case ($K = 2$) and compare E-Learning with existing methods from literature discussed in Table 3.1 in Section 3.3.3. The implementation details of these methods are provided in Section 3.8.3 in Section 3.8.

For the implementation of E-Learning, we consider HPLM (3.11) and solve the regularized estimating equation in Section 3.2.5.4 with the row-wise group-LASSO penalty. We follow the implementation in Section 3.2.5 for the estimation of the treatment-free effect with the linear working model, the propensity score with the multinomial logistic working model, and the variance function with regression forest. The tuning parameter λ is chosen based on 10-fold cross validation. We consider the oracle working variance function $\sigma^2_{\text{opt}}(\boldsymbol{X}, A) = [\breve{\mu}_0(\boldsymbol{X}) - \mu_0(\boldsymbol{X})]^2 + \sigma^2(\boldsymbol{X}, A)$ and the estimated one from the regression forest using the squared residual as the response and $(\boldsymbol{X}, A)$ as the covariates. At the testing stage, a testing covariate sample $\{\boldsymbol{X}_i\}_{i=1}^{n_{\text{test}}=10000} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{I}_{p\times p})$ is generated, and the testing value of an estimated ITR $\widehat{d}$ is computed as $\widehat{\mathcal{V}}_{\text{test}}(\widehat{\boldsymbol{d}}) = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \sum_{k=1}^{K} \boldsymbol{\beta}_k^\intercal (1, \boldsymbol{X}_i^\intercal)^\intercal \mathbb{1}[\widehat{d}(\boldsymbol{X}_i) = k]$. Recall that the optimal ITR is $d^\star(\boldsymbol{x}) = \text{argmax}_{1\leqslant k\leqslant K} \boldsymbol{\beta}_k^\intercal (1, \boldsymbol{x}^\intercal)^\intercal$. Then we report the testing regret, $\widehat{\mathcal{V}}_{\text{test}}(d^\star) - \widehat{\mathcal{V}}_{\text{test}}(\widehat{d})$, and the testing misclassification rate, $\frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbb{1}[\widehat{d}(\boldsymbol{X}_i) \neq d^\star(\boldsymbol{X}_i)]$. The training-testing process is replicated for 100 times for each of the model specification scenarios in Table 3.2.

We first consider the low-dimensional setting ($p = 10$). Figure 3.1 reports the testing misclassification rates for the training sample sizes $n \in \{100, 200, 400, 800, 1600\}$ and each of the specification scenarios listed in Table 3.2, while Figure 3.2 provides more details for $n = 400$. In the case of correctly specified treatment-free effect, correctly specified propensity score, and homoscedasticity (upper-left panel of the plots), E-Learning, Q-Learning, G-Estimation, A-Learning, RD-Learning, dWOLS and Subgroup Identification have similar testing performance, since all of them leverage the correct parametric model assumption. Here, although Subgroup Identification does not rely on a specific parametric model assumption, it is equivalent to RD-Learning in this case as discussed

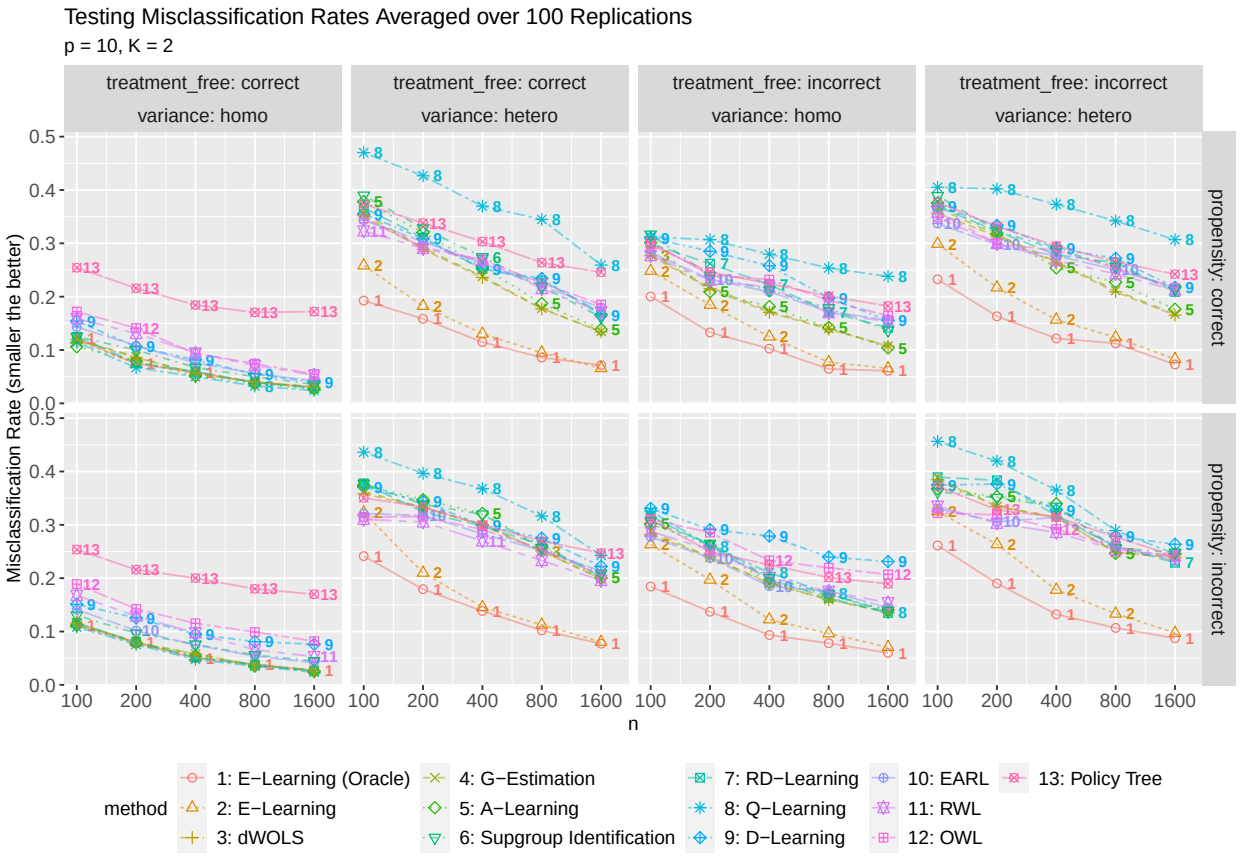


Figure 3.1: Testing misclassification rates (smaller the better) for $n \in \{100, 200, 400, 800, 1600\}$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees.

in Section 3.3.1. Therefore, it can enjoy similar performance as other model-based methods. In contrast, D-Learning, OWL, RWL, EARL and Policy Tree are based on nonparametric models, and can have inferior performance in this case. When one of or both misspecified treatment-free effect and heteroscedasticity happen (columns 2-4 of the plots), the E-Learning procedures with the oracle and estimated working variance function both demonstrate the best performance among all methods. In particular, the advantages of E-Learning are more evident as n increases. Such a superiority can still maintain even if the propensity score model is misspecified (second rows of the plots). This suggests that incorrect propensity score can have relatively small impacts.

In Section 3.8, we further provide more plots of misclassification rates for $n \in \{100, 200, 800, 1600\}$ (Figures 3.11-3.14) and testing regrets (Figures 3.15-3.20). All of them show the same patterns as in Figures 3.1 and 3.2. In order to further demonstrate the superiority of

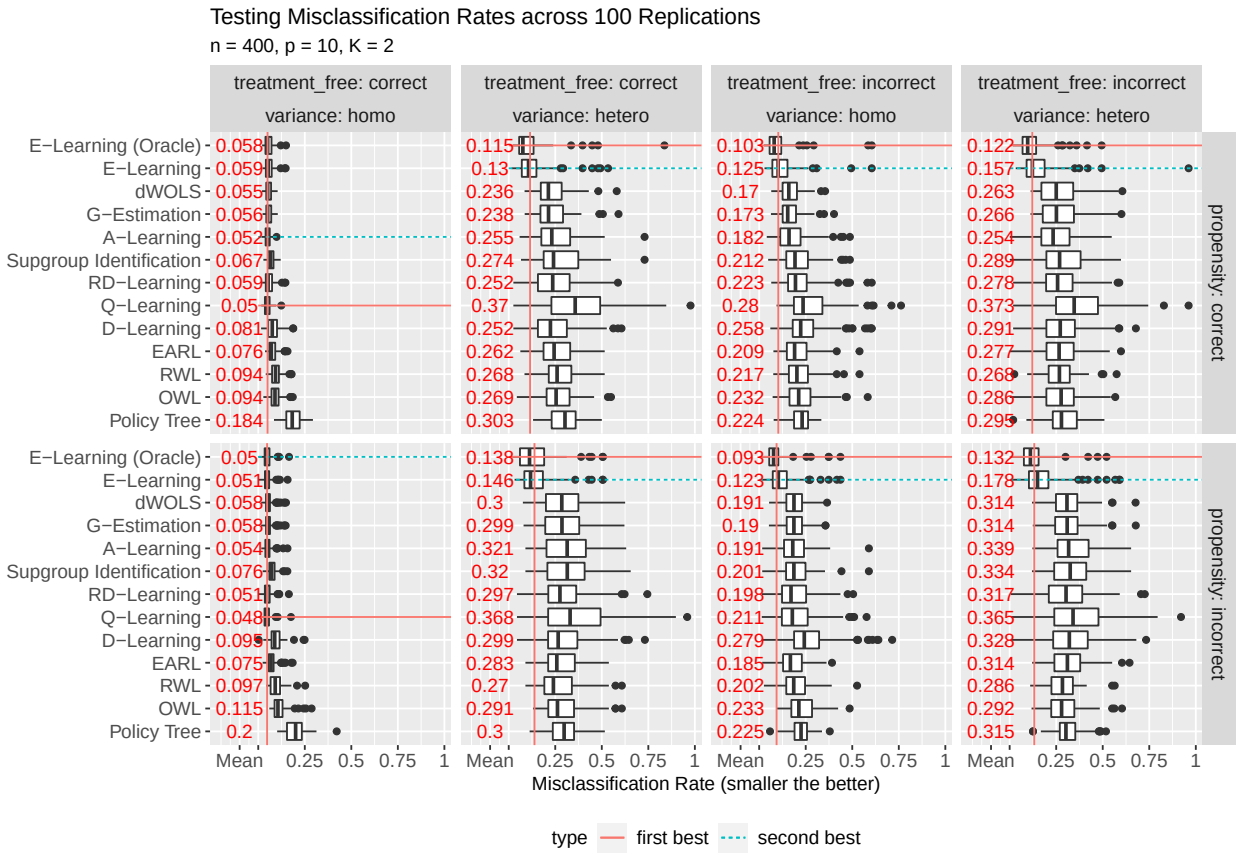


Figure 3.2: Testing misclassification rates (smaller the better) for $n = 400$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged misclassification rates are annotated in horizontal lines. The minimal averaged misclassification rate is shown by the vertical line.

E-Learning in presence of moderately large number of variables, we also study the case of $p = 50$ and report the testing performance in Figures 3.21 and 3.22. They show that even though the increase in p can result in worse performance of all methods, the efficiency gain in sufficiently large samples ($n = 200, 400, 800, 1600$) of E-Learning remains.

3.5.3 Multiple Treatments

We consider the multiple treatment case ($K = 3$) and compare E-Learning with model-based methods that can allow multiple treatments (Q-Learning, D-Learning, RD-Learning). In particular, we are interested in the following questions:

(I) Efficiency of different methods as n increases across all model specifications in Table 3.2;

(II) The impacts of increase in the number of variables p;

(III) The impacts of increase in the number of treatments K;

(IV) Effects of different nonparametric estimation methods for variance function on the performance of E-Learning.

For Question I, we consider the same setup as in Section 3.5.2 but with $K = 3$. The testing results are provided in Figure 3.23 in Section 3.8. In particular, E-Learning shows the same superiority over Q-Learning, D-Learning, RD-Learning as in the binary case. For Question II, we consider $K = 3$ and varying $p \in \{10, 50, 100\}$ (Figures 3.24 and 3.25). As the number of variables p increases, the performance of all methods become worse. For $p = 50, 100$, Q-Learning, D-Learning and RD-Learning have much worse performance when one of or both treatment-free effect misspecification and heteroscedasticity happen, even with the sample size $n = 1600$. The misclassification rates of these methods are 0.562, 0.429 and 0.433 respectively for incorrectly specified treatment-free effect and heteroscedasticity with $n = 1600$ and $p = 100$. In contrast, for sufficiently large sample sizes ($n = 400, 800, 1600$), the number of variables p has less impacts on E-Learning with the oracle working variance function, while it requires sizes ($n = 800, 1600$) for E-Learning with the estimated working variance function to have comparable performance across p's. The reason for requiring larger sample sizes is due to the challenge of the high-dimensional nonparametric estimation of the working variance function. The misclassification rates of E-Learning for incorrectly specified treatment-free effect and heteroscedasticity with $n = 1600$ and $p = 100$ are 0.167 for the oracle working variance function and 0.248 for the estimated working variance function respectively. These results can confirm the superiority of E-Learning even when the number of variables increases to 100.

In order to study Question III, we consider $p = 10$ and varying $K \in \{2, 3, 5, 7\}$. Notice that increasing the number of treatments can have two folds of effects. On one hand, the effective dimensionality generally increases in K. For HPLM (3.11), the interaction effect $\gamma(\boldsymbol{X}, A; \mathsf{B}) = (1 - 1/K)\langle \boldsymbol{\omega}_A, \mathsf{B}^\intercal \boldsymbol{X} \rangle$ is indexed by the matrix-valued parameter $\mathsf{B} \in \mathbb{R}^{p \times (K-1)}$. The effective dimension is $p(K-1)$ and increases with K. Moreover, the number of variance functions $\{\sigma^2_{\text{opt}}(\boldsymbol{X}, k)\}_{k=1}^K$ also increases in K, which means more nuisance functions to be nonparametrically estimated. On the other hand, more treatments can lead to a harder classification problem. In particular, the

misclassification rate of a random treatment rule d_{rand} with $\mathbb{P}[d_{\text{rand}}(\boldsymbol{X}) = k] = 1/K$ for $1 \leqslant k \leqslant K$ is $1 - 1/K$. Then the misclassification rate of the random treatment rule increases in K, which suggests that the difficulty of the learning problem is also increasing. In Figures 3.26 and 3.27 in Section 3.8, Q-Learning, D-Learning and RD-Learning have poor performance in presence of one of or both treatment-free effect misspecification and heteroscedasticity. When both treatment-free effect misspecification and heteroscedasticity exist, the misclassification rates of these methods with $n = 1600$ and $K = 7$ are 0.811, 0.648 and 0.645 respectively. Notice that the misclassification rate of the random treatment rule in this case is $1 - 1/7 = 0.857$, which suggests that the performance of Q-Learning is close to the random treatment rule. In contrast, the E-Learning procedures with oracle working variance and estimated working variance have misclassification rates 0.299 and 0.424 in this case, which significantly outperform other methods.

Finally, for Question IV, we consider $p \in \{10, 50, 100\}$, $K = 3$ and the comparisons among E-Learning procedures with the oracle optimal working variance function, the working variance function estimated by regression forest, MARS and COSSO. The numerical results in Figures 3.28 and 3.29 suggest that E-Learning with regression forest can have better performance than E-Learning with MARS or COSSO, and the superiority remains even for $p = 50, 100$. Therefore, we recommend using regression forest for the working variance function estimation in E-Learning.

3.6 Application to a Type 2 Diabetes Mellitus (T2DM) Study

We consider a T2DM dataset from an observational study based on the *Clinical Practice Research Datalink (CPRD)* (Herrett et al., 2015; Chen et al., 2018). The study population comprises T2DM patients of age $\geqslant$ 21 years (registered at a CPRD practice) who received at least one of the long-acting insulins (Glargine or Detemir), the intermediate-acting insulins, the short-acting insulins, and the *Glucagon-Like Peptide 1 Receptor Agonists (GLP-1 RAs)* of Exenatide and Liraglutide during 01/01/2012 - 12/31/2013. The treatment exposure A is defined as: 1) the long-acting insulins alone (with no addition of any short or intermediate-acting insulin within 60 days); 2) the intermediate-acting insulins alone (with no addition of any short or long-acting insulins within 60 days); 3) any insulin regimens including a short-acting insulin (the short-acting insulins either alone or in combinations with any long or intermediate-acting insulin); 4) the GLP-1 RAs alone.

Here, for patients who received one of the insulins as well as the GLP-1 RAs, the corresponding treatment is defined as the earliest received one.

The primary outcome Y of this study is the change of the *Hemoglobin A1c (HbA1c)* lab value (%, smaller the better) between Day 182 and Day 1 (defined as the first treatment date). The following individual covariates $\boldsymbol{X}$ are measured: age, gender, ethnicity, weight, height, *Body Mass Index (BMI)*, *High Density Lipoprotein (HDL)*, *Low Density Lipoprotein (LDL)*, baseline HbA1c, smoking status, and comorbidities (any of angina, congestive heart failure, myocardial infarction, stroke, retinopathy, macular edema, renal status, neuropathy, and lower extremity amputation). The total number of records from this study is 1139, with the primary outcome available for 591 records and missing for the rest. Among the 591 observations, there is a large proportion of missingness in HDL and LDL. Therefore, for HDL and LDL, we first discretize the available observations into two levels: if the observation is above the median, then set as `high`; otherwise, set as `low`. Then we code the missing observations as `n/a`. Consequently, all possible levels of LDL and HDL become: `high`, `low`, and `n/a`. For categorical variables (gender, ethnicity, smoking status and comorbidities), we also code the missing observations as `n/a` and combine it with the original levels of these variables. Finally, the remaining numerical variables (age, weight, height, BMI, baseline HbA1c) have mild missingness, and we remove the records that contain any missing entries among these variables. After pre-processing the dataset as above, there remains 430 records for further analysis.

Next, we estimate the propensity scores from the dataset using the regression forest estimator in Section 3.2.5.1. Then we are ready to apply E-Learning, RD-Learning, D-Learning, Q-Learning and Policy Tree to the analysis of this dataset. In order to estimate the expected change of HbA1c under the fitted ITRs, we randomly sample two disjoint subsets from the dataset for training and testing. We choose various training sample sizes as $n \in \{100, 200, 300\}$, and a testing sample size $n_{\text{test}} = 100$. On the training set, we consider estimation of the propensity scores based on regression forest in the same way as that on the full dataset. We also apply different estimation methods of the treatment-free effect for RD-Learning, including: 1) the linear model on $\boldsymbol{X}$ with the ℓ_1-penalty (fitted by `glmnet`) as in Section 3.5, 2) the regression forest on $\boldsymbol{X}$, 3) fitted treatment-free effect as the mean of the primary outcome on the training set, and 4) fitted treatment-free effect as 0. We find that the fitted treatment-free effect as 0 can result in better testing performance for RD-

Learning. Therefore, we also use 0 as the estimated treatment-free effect in E-Learning. Since a smaller outcome is better for this problem, we negate the outcome before fitting all models. Other implementation details of all these methods remain the same as in Section 3.5. On the testing dataset, we use the IPWE $\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\frac{\mathbb{1}[\widehat{d}(\boldsymbol{X}_i)=A_i]}{\widehat{p}_{\mathscr{A}}(A_i|\boldsymbol{X}_i)}Y_i$ to estimate the expected change in HbA1c under the estimated ITR $\widehat{d}$. The training-testing process is repeated for 500 times on this dataset.

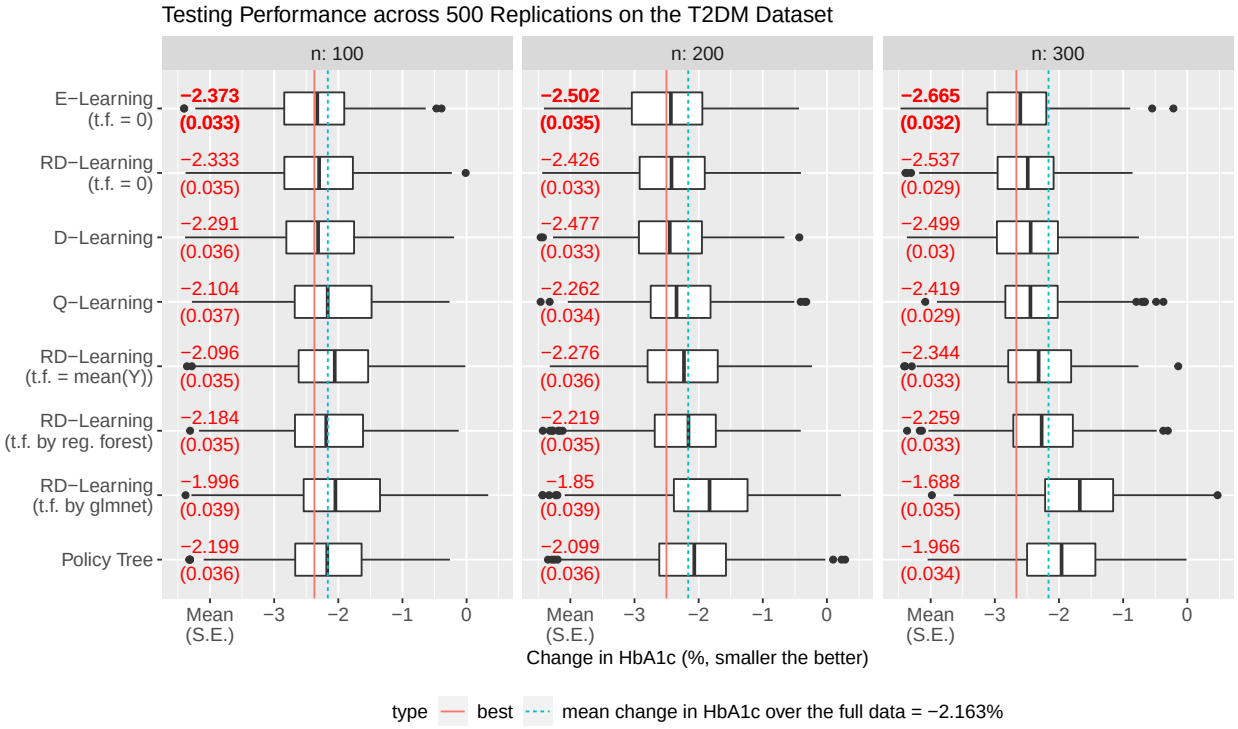


Figure 3.3: Testing changes in HbA1c (%, smaller the better) for training sample sizes $n \in \{100, 200, 300\}$ on the T2DM dataset. Here, *t.f.* represents the fitted treatment-free effect, and *reg. forest* corresponds to the regression forest.

The testing results are reported in Figure 3.3. E-Learning enjoys the best testing performance among all training sample sizes. As the training sample size n increases, the advantage of E-Learning is more evident compared with other methods. This can confirm the efficiency improvement of E-Learning by using an optimal working variance function on this dataset. Among patients in the T2DM dataset, E-Learning recommends 19.77% for long-acting insulins, 18.14% for intermediate-acting insulins, 30.23% for short-acting insulins, and 31.86% for GLP-1 RAs. The fitted E-Learning coefficients are reported in Table 3.3. In particular, short-acting insulins (`A = 3`) is recommended for the patients with average covariates. Patients as former smokers are more recommended for the short-acting insulins than other patients. The general benefits of short-acting insulins are consistent

Table 3.3: E-Learning Coefficients on the T2DM Dataset

	A = 1	A = 2	A = 3	A = 4
Intercept	**-0.164**	-0.053	**0.168**	0.048
gender (male)	-0.008	-0.014	0.029	-0.007
ethnic (others)				
ethnic (white)	-0.023	0.081	-0.029	-0.03
smoke (former)	-0.001	**-0.174**	**0.361**	**-0.186**
smoke (no)				
smoke (yes)				
comorbidity (yes)				
HDL (low)	0.033	-0.009	-0.053	0.029
HDL (high)				
LDL (low)	0.004	**0.115**	**-0.122**	0.003
LDL (high)	0.007	-0.002	-0.023	0.018
baseline HbA1c	**-0.492**	**0.139**	**0.168**	**0.185**
age	-0.058	**0.169**	**0.106**	**-0.217**
weight				
height				
BMI				

Note:
Larger coefficients encourage better outcome.
Coefficients are fitted at standardized scales of covariates.
Blank coefficients are 0's. Absolute value > 0.1 are bolded.

with the results in Chen et al. (2018); Meng et al. (2020). Moreover, it can be observed that the coefficients for `baseline HbA1c` in Table 3.3 increase in the treatment arm. In fact, the averaged baseline HbA1c values among recommended treatments $A = 1, 2, 3, 4$ are 7.35%, 10.67%, 10.91% and 11.18% respectively. This suggests that patients with worse baseline HbA1c are recommended for faster-acting therapies, where the GLP-1 RAs (`A = 4`) can be regarded as an alternative for the rapid-acting insulin (Ostroff, 2016). Such a phenomenon is also consistent with the recommended treatment ordinality pointed out by Chen et al. (2018).

3.7 Discussion

In this chapter, we propose E-Learning for learning an optimal ITR under heteroscedasticity or misspecified treatment-free effect. In particular, E-Learning is developed from semiparametric efficient estimation in the multi-armed treatment setting. When nuisance models are correctly specified, even if heteroscedasticity exists, the $\sqrt{n}$-asymptotic variance of the estimated parameters achieve the semiparametric variance lower bound. When the treatment-free effect model is misspecified, E-Learning targets the optimal working variance function, so that the $\sqrt{n}$-asymptotic variance of the estimated parameters is still the smallest among the class of regular semiparametric estimates. In summary, E-Learning extends the optimality of existing model-based methods to allow multi-

ple treatments, heteroscedasticity and treatment-free effect misspecification. The efficiency gain of E-Learning is demonstrated by our simulation studies when either of or both heteroscedasticity and misspecified treatment-free effect happen, where existing methods can have much worse performance. This also can be consistent with Kang and Schafer (2007)'s finding that the misspecified treatment-free effect can have severe consequences.

E-Learning is developed based on the parametric assumption on the covariate-treatment interaction effect $\gamma(\boldsymbol{X}, A; \boldsymbol{\beta})$. This can be further extended to flexible semiparametric or nonparametric models such as the single-index model (Liang and Yu, 2020) and nonlinear functions in the reproducing kernel Hilbert space (Zhao et al., 2012). Our proposed regularized estimation problem in Section 3.2.5.4 can be ready for nonlinear learning when a functional penalty is used. It requires further extensions of the efficient score from our Proposition 3.6 to semiparametric/nonparametric settings.

Another direction of future work can be the high-dimensional problem. In Section 3.2.5.4, we propose to solve the regularized estimating equation, which can handle high-dimensional parameter estimation. However, the nonparametric estimation of the working variance function is also a potential challenge when the dimension is growing. In our simulation study, our proposed E-Learning with estimated working variance function requires larger sample sizes in presence of increasing numbers of variables and treatments. In the literature, there exists three possible strategies to accommodate this challenge: 1) considering index models for the variance function that can allow dimension reduction (Zhu et al., 2013; Lian et al., 2015); 2) estimating the central variance subspace for sufficient dimension reduction (Zhu and Zhu, 2009; Luo et al., 2014; Ma and Zhu, 2019); 3) performing simultaneous nonlinear variable selection during nonparametric regression (Lin and Zhang, 2006; Lafferty and Wasserman, 2008; Zhang et al., 2011; Allen, 2013). These can have potential for further improvement of E-Learning.

3.8 Appendix

3.8.1 Analysis of the ACTG 175 Trial Data

We evaluate the effectiveness of our proposed E-Learning on a clinical trial dataset from the "AIDS clinical trial group study 175" (Hammer et al., 1996). The goal of this study was to compare four

treatment arms among 2,139 randomly assigned subjects with human immunodeficiency virus type 1 (HIV-1), whose CD4 counts were 200-500 cells/mm^3. The four treatment options of A are the zidovudine (ZDV) monotherapy, the didanosine (ddI) monotherapy, the ZDV combined with ddI, and the ZDV combined with zalcitabine (ZAL).

The primary outcome Y of our interest is the difference between the CD4 cell counts at early stage (20$\pm$5 weeks from baseline) and the CD4 counts at baseline, which is larger the better. We follow the analyses in Lu et al. (2013); Qi et al. (2020); Meng and Qiao (2020) and consider 12 selected baseline covariates $\boldsymbol{X}$. There are 5 continuous covariates: age (year), weight (kg, coded as `wtkg`), CD4 count (cells/mm^3) at baseline, Karnofsky score (scale of 0-100, coded as `karnof`), CD8 count (cells/mm^3) at baseline. They are centered and scaled before further analysis. In addition, there are 7 binary variables: gender (1 = male, 0 = female), homosexual activity (`homo`, 1 = yes, 0 = no), race (1 = nonwhite, 0 = white), history of intravenous drug use (`drug`, 1 = yes, 0 = no), symptomatic status (`symptom`, 1 = symptomatic, 0 = asymptomatic), antiretroviral history (`str2`, 1 = experienced, 0 = naive) and hemophilia (`hemo`, 1 = yes, 0 = no).

We consider the training sample size $n \in \{100, 200, 400, 800, 1600\}$ and the testing sample size $n_{\text{test}} = 400$. The full dataset is randomly split into training and testing according to the given sample sizes. Since the dataset is obtained from a randomized controlled trial, the propensity score function is known to be $p_{\mathscr{A}}(k|\boldsymbol{X}) = 1/4$ for $k = 1, 2, 3, 4$. For the treatment-free effect estimation, we consider a linear working model with the ℓ_1-penalty throughout the analysis, which will be different from the implementation in Meng and Qiao (2020). For this real-world data application, the underlying truth is unknown to us. We cannot verify whether any of misspecified treatment-free effect and heteroscedasticity on the original dataset exist. Nevertheless, after modifying the dataset according to the following Table 3.4, the treatment-free effect misspecification and heteroscedasticity can be anticipated. Note that the unmodified cases can also have treatment-free effect misspecification and heteroscedasticity as well. Our modification can enlarge such effects. The goal of our analysis is to demonstrate the efficiency improvement of E-Learning in presence of heavy treatment-free effect misspecification and heteroscedasticity. We further provide the residual plots in Figure 3.30 in Section 3.8.6. Residuals are computed from the fitted E-Learning on each modified dataset according to Table 3.4, and averaged over 10 replications. It confirms that the modifications can result in the squared residuals heavily depending on the variables `age` and `wtkg`.

Table 3.4: Modifications on the ACTG 175 Dataset and the Implying Model Specifications

Training Outcome Modification	Treatment-Free Effect	Variance Function
Original Y	*unmodified*	*unmodified*
$Y \leftarrow Y + e^{2.5\times\texttt{age}}$	***misspecified***	*unmodified*
$Y \leftarrow Y + 5e^{1.5\times\texttt{wtkg}} \times \xi$	*unmodified*	***heteroscedastic***
$Y \leftarrow Y + e^{2.5\times\texttt{age}} + 5e^{1.5\times\texttt{wtkg}} \times \xi$	***misspecified***	***heteroscedastic***

[1] The variables `age` and `wtkg` are centered and scaled at the data preparation stage.
[2] The additional noise ξ is randomly generated from $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = -1) = 1/2$ independent of $\boldsymbol{X}, A, Y$.
[3] We further round the modified outcomes to their nearest integers to respect the integer nature of Y.
[4] The treatment-free effect is estimated by a linear working model with the ℓ_1-penalty.

On the training sample, we implement the same procedures as in Section 3.5 to fit Q-Learning, D-Learning, RD-Learning, and our proposed E-Learning. On the testing dataset, we evaluate the an estimated ITR $\widehat{d}$ by the IPWE $(1/n_{\text{test}}) \sum_{i=1}^{n_{\text{test}}} Y_i \mathbb{1}[A_i = \widehat{d}(\boldsymbol{X}_i)]/(1/4)$, which is larger the better. Here, the testing outcome Y_i is unmodified in contrast to the training outcome to ensure comparable testing evaluation. Testing results based on 500 repeated training and testing for each of the four cases in Table 3.4 are reported in Figure 3.4.

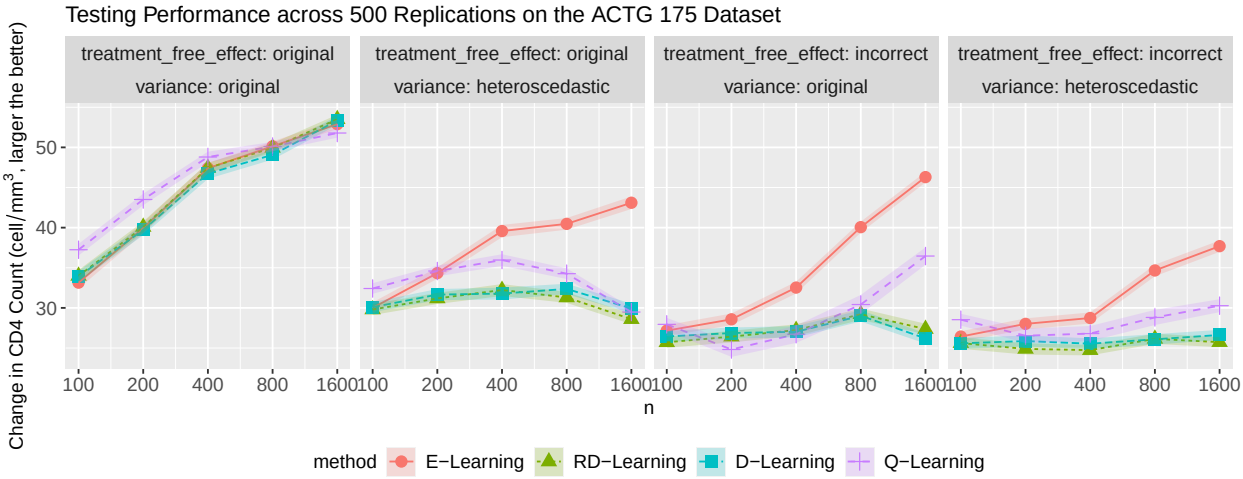


Figure 3.4: Testing changes in CD4 count (cell/mm^3, larger the better) on the ACTG 175 dataset.

When we use the original training outcome Y, the testing CD4 count improvements of all methods close to each other. In particular, Q-Learning demonstrates slightly better performance for $n = 100, 200, 400$, but all methods have similar performance when $n = 800$ and 1600. All these methods have improving testing performance as n increases. When we modify Y to incorporate heavy heteroscedasticity or/and treatment-free effect misspecification, E-Learning can maintain the improvements as n increases, while other methods can have much poorer performance. In particular,

other methods can even get worse as n increases in presence of heavy heteroscedasticity. We shall anticipate that the scientific findings for the analysis with original outcome Y will not be disturbed when we introduce additional treatment-free effect misspecification or/and heteroscedasticity. The results in Figure 3.4 show that E-Learning maintains the testing performance well during these modifications, while RD-Learning, D-Learning and Q-Learning are heavily affected. In this way, E-Learning demonstrates its superiority of efficiency gain in presence of misspecified treatment-free effect or/and heteroscedasticity.

We further report the estimated coefficients for D-Learning, RD-Learning and E-Learning in Table 3.5 and Figure 3.31 in Section 3.8.6. The fitted coefficients on the original data are consistent with existing literature. Specifically, `Intercept`, `age` and `cd40` are common important covariates that were frequently reported in the literature (Lu et al., 2013; Qi et al., 2020; Meng and Qiao, 2020). When we incorporate heavy treatment-free effect misspecification or/and heteroscedasticity in cases II, III and IV, the fitted coefficients of D-Learning and RD-Learning become highly unstable with many extreme coefficients. In contrast, the fitted E-Learning coefficients are relatively stable across these cases. This suggests that the E-Learning estimate can be more resilient to the training outcome modifications in Table 3.4 compared with the other methods.

3.8.2 Optimal Estimating Function under Misspecified Propensity Score Model

Let $\breve{p}_{\mathscr{A}}(a|\boldsymbol{x})$ be an arbitrary propensity score function. For any $\mathsf{H} : \mathcal{X} \to \mathbb{R}^{p\times(K-1)}$, which can depend on $\boldsymbol{\beta}$, consider the following estimating function:

$$\boldsymbol{\phi}(\boldsymbol{\beta};\breve{p}_{\mathscr{A}}) = [Y - \mu_0(\boldsymbol{X}) - \gamma(\boldsymbol{X}, A; \boldsymbol{\beta})]\frac{\mathsf{H}(\boldsymbol{X})\boldsymbol{\omega}_A}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}.$$

By Proposition 3.7, since the working treatment-free effect function μ_0 is true, for any working propensity score function $\breve{p}_{\mathscr{A}}$, we have $\mathbb{E}[\boldsymbol{\phi}(\boldsymbol{\beta};\breve{p}_{\mathscr{A}})] = \mathbf{0}$ at the true $\boldsymbol{\beta}$. Our goal is to find the optimal $\mathsf{H}(\boldsymbol{X})$ for a given working propensity score function $\breve{p}_{\mathscr{A}}$.

The following derivations are analogous to the proof of Theorem 3.10:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\phi}(\boldsymbol{\beta};\breve{p}_{\mathscr{A}})^{\otimes 2}] &= \mathbb{E}\left[\mathsf{H}(\boldsymbol{X})\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\epsilon^2}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})^2}\right)\mathsf{H}(\boldsymbol{X})^\intercal\right] \\
&= \mathbb{E}\left[\mathsf{H}(\boldsymbol{X})\mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}})\mathsf{H}(\boldsymbol{X})^\intercal\right]; \\
\mathbb{E}\left[-\frac{\partial\breve{\boldsymbol{\phi}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\intercal}\right] &= \mathbb{E}\left\{\mathsf{H}(\boldsymbol{X})\left[\left(1-\frac{1}{K}\right)\frac{\boldsymbol{\omega}_A^{\otimes 2}}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}\right]\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right\} - \underbrace{\mathbb{E}\left[\dot{\mathsf{H}}(\boldsymbol{X};\boldsymbol{\beta})\frac{\boldsymbol{\omega}_A\epsilon}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}\right]}_{=0} \\
&= \mathbb{E}\left[\mathsf{H}(\boldsymbol{X})\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right].
\end{aligned}$$

In the second equality, we define

$$\mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}}) := \sum_{k=1}^K \frac{p_{\mathscr{A}}(k|\boldsymbol{X})\sigma^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A}}(k|\boldsymbol{X})^2}.$$

In the forth equality, it follows from that $\mathbb{E}(\epsilon|\boldsymbol{X},A)=0$ and we define

$$\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}}) := \left(1-\frac{1}{K}\right)\sum_{k=1}^K \frac{p_{\mathscr{A}}(k|\boldsymbol{X})\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A}}(k|\boldsymbol{X})}.$$

Let $\widehat{\boldsymbol{\beta}}_n(\breve{p}_{\mathscr{A}})$ be the solution to $\mathbb{E}_n[\boldsymbol{\phi}(\boldsymbol{\beta};\breve{p}_{\mathscr{A}})]=\mathbf{0}$. Then under the same regularity conditions as in Theorems 3.9 and 3.10, we have

$$\begin{aligned}
&\lim_{n\to\infty} n\mathsf{Var}[\widehat{\boldsymbol{\beta}}_n(\breve{p}_{\mathscr{A}})] \\
=&\left\{\mathbb{E}\left[-\frac{\partial\boldsymbol{\phi}(\boldsymbol{\beta};\breve{p}_{\mathscr{A}})}{\partial\boldsymbol{\beta}^\intercal}\right]\right\}^{-1}\mathbb{E}[\breve{\boldsymbol{\phi}}(\boldsymbol{\beta})^{\otimes 2}]\left\{\mathbb{E}\left[-\frac{\partial\boldsymbol{\phi}(\boldsymbol{\beta};\breve{p}_{\mathscr{A}})^\intercal}{\partial\boldsymbol{\beta}}\right]\right\}^{-1} \\
=&\widetilde{\mathsf{B}}^{-1}\widetilde{\mathsf{A}}\widetilde{\mathsf{B}}^{-\intercal} = (\widetilde{\mathsf{B}}^\intercal\widetilde{\mathsf{A}}^{-1}\widetilde{\mathsf{B}})^{-1} \geqslant \widetilde{\mathsf{C}}^{-1},
\end{aligned}$$

where, analogous to Lemma 3.15, we define

$$\begin{aligned}
\widetilde{\mathsf{A}} &:= \mathbb{E}[\mathsf{H}(\boldsymbol{X})\mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}})\mathsf{H}(\boldsymbol{X})^\intercal]; \\
\widetilde{\mathsf{B}} &:= \mathbb{E}[\mathsf{H}(\boldsymbol{X})\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})]; \\
\widetilde{\mathsf{C}} &:= \mathbb{E}[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})\mathsf{V}_\epsilon(\boldsymbol{X};\breve{\mu}_0)^{-1}\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})],
\end{aligned}$$

with equality if and only if exists some non-singular constant matrix $\mathsf{H}_0 \in \mathbb{R}^{p\times p}$ such that

$$\mathsf{H}(\boldsymbol{X}) = \mathsf{H}_0\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal \mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})\mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}})^{-1}.$$

Therefore, the optimal estimating function under the working propensity score function $\breve{p}_{\mathscr{A}}$ is

$$\begin{aligned}\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta};\breve{p}_{\mathscr{A}}) :=& [Y-\mu_0(\boldsymbol{X})-\gamma(\boldsymbol{X},A;\boldsymbol{\beta})]\times \\ & \underbrace{\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal \overbrace{\left[\left(1-\frac{1}{K}\right)\sum_{k=1}^K \frac{p_{\mathscr{A}}(k|\boldsymbol{X})\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A}}(k|\boldsymbol{X})}\right]}^{:=\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})} \left[\overbrace{\sum_{k=1}^K \frac{p_{\mathscr{A}}(k|\boldsymbol{X})\sigma^2(\boldsymbol{X},k)\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A}}(k|\boldsymbol{X})^2}}^{:=\mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}})}\right]^{-1} \frac{\boldsymbol{\omega}_A}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}}_{\text{optimal instrument}}.\end{aligned}$$

Here, different from the case of misspecified treatment-free effect model discussed in Section 3.4.1.2, the optimal estimating function cannot be defined from an optimal working variance as in Theorem 3.10. It require different strategies to estimate the optimal variance components $\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})$ and $\mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}})$. One potential strategy is to identified them from

$$\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}}) = \mathbb{E}\left[\left(1-\frac{1}{K}\right)\frac{\boldsymbol{\omega}_A^{\otimes 2}}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}\middle|\boldsymbol{X}\right]; \quad \mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}}) = \mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\epsilon^2}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})^2}\middle|\boldsymbol{X}\right),$$

where ϵ is replaced by the working residual $e = Y-\mu_0(\boldsymbol{X})-\gamma(\boldsymbol{X},A;\boldsymbol{\beta})$. Therefore, we can perform nonparametric regression on the $\mathbb{R}^{(K-1)\times(K-1)}$-valued matrices $\left(1-\frac{1}{K}\right)\frac{\boldsymbol{\omega}_A^{\otimes 2}}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}$ and $\frac{\boldsymbol{\omega}_A^{\otimes 2}e^2}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})}$ on $\boldsymbol{X}$. However, such a strategy will be much different from the methodology proposed in this chapter. In particular, in this chapter, we only need to estimate an $\mathbb{R}^K$-valued function $\left(\breve{\sigma}^2_{\text{opt}}(\boldsymbol{X},k) : 1 \leqslant k \leqslant K\right)$, while the optimal estimating function under misspecified propensity score model require the estimation of two $\mathbb{R}^{(K-1)\times(K-1)}$-valued functions $\mathsf{V}_{\mathscr{A}}(\boldsymbol{X};\breve{p}_{\mathscr{A}})$ and $\mathsf{V}_\epsilon(\boldsymbol{X};\breve{p}_{\mathscr{A}})$.

3.8.3 More Implementation Details

- **E-Learning** (general K): When fitting the treatment-free effect and the propensity score functions, we consider the 10-fold cross-fitting strategy as in Chernozhukov et al. (2018a); Zhao et al. (2019a); Athey and Wager (2021). Specifically, the training sample is randomly divided into 10 folds. For the k-th fold fitting, we utilize the data other the k-th fold to fit a treatment-free effect

or propensity score model, and then predict the treatment-free effects or the propensity scores for the k-th fold data.

We use regression forest to estimate variance function from the squared working residual. The `regression_forest` function from the `grf` package in `R` is called. We also fit the MARS and COSSO estimates of variance function. The `earth` function from the `earth` package and a modified program based on the `cosso.Gaussian` function from the `cosso` package are applied. More details on the COSSO model and program are discussed in Section 3.8.4.

Before fitting E-Learning, we first center and scale each variables to ensure $(1/n)\sum_{i=1}^{n} X_{ij} = 0$ and $(1/n)\sum_{i=1}^{n} X_{ij}^2 = 1$. When solving the penalized minimization problem (3.8) by the accelerated proximal gradient descent, we call the `apg` function in `R` to perform optimization. When determining the tuning sequence of λ's, we take the strategy analog to `glmnet` (Friedman et al., 2010).

- **Q-Learning** (general K): We consider the linear model using Y as the response and $\big(1, \boldsymbol{X}^\intercal, \vec{\boldsymbol{A}}^\intercal, \vec{\boldsymbol{A}}^\intercal \otimes \boldsymbol{X}^\intercal\big)^\intercal$ as the covariates with the ℓ_1-penalty. Here, $\vec{\boldsymbol{A}} = \big(\mathbb{1}(A = 2), \mathbb{1}(A = 3), \cdots, \mathbb{1}(A = K)\big)^\intercal$, and $\otimes$ denotes the Kronecker product. The method is also known as the *ℓ_1-Penalized Least Square (ℓ_1-PLS)* (Qian and Murphy, 2011), and implemented by the `glmnet` function in `R`.

- **G-Estimation, dWOLS** ($K = 2$, $A \in \{0, 1\}$): The `DTRreg` function from the `DTRreg` package (Wallace et al., 2017) in `R` is called to fit G-Estimation (`method = "gest"`) and dWOLS (`method = "dwols"`). The treatment-free effect model (`tf.mod`) is specified as linear in $(1, \boldsymbol{X}^\intercal)^\intercal$. The propensity score model (`treat.mod`) is specified as the logistic model of A with respect to $(1, \boldsymbol{X}^\intercal)^\intercal$. The interaction effect model (`blip.mod`) is specified as linear in $(1, \boldsymbol{X}^\intercal)^\intercal$. For dWOLS, the weight function $w(\boldsymbol{X}, A) = |A - \widehat{\pi}_{\mathscr{A},n}(\boldsymbol{X})|$ in Section 3.3.1 is used.

- **A-Learning, Subgroup Identification** ($K = 2$, $A \in \{-1, 1\}$): The `fit.subgroup` function from the `personalized` package (Huling and Yu, 2018) in `R` is called to fit A-Learning and Subgroup Identification with the ℓ_1-penalty (`method = "a_learning"` and `method = "weighting"` respectively, `loss = "sq_loss_lasso"`). The propensity score model (`propensity.func`) is specified as the logistic model of A with respect to $(1, \boldsymbol{X}^\intercal)^\intercal$ with the ℓ_1-penalty, which is fitted by

`glmnet`. The treatment-free effect, also known as the augmentation function (`augment.func`), is specified as: for A-Learning, the linear model of Y with respect to $(1, \boldsymbol{X}^\intercal)^\intercal$ with the ℓ_1-penalty, fitted by `glmnet`; and for Subgroup Identification, the linear model of Y with respect to $(1, \boldsymbol{X}^\intercal, A, A\boldsymbol{X}^\intercal)^\intercal$ with the ℓ_1-penalty, fitted by `glmnet`, and outputting the arithmetic average of predictions at $A = 1$ and $A = -1$.

- **D-Learning, RD-Learning:** (general K) We consider the class of linear functions with the row-wise grouped LASSO penalty. The training process is performed by the accelerated proximal gradient descent using the `apg` function. The estimation of the treatment-free effect and propensity score functions, the fitting details and the tuning strategy are the same as in E-Learning.

- **OWL, RWL, EARL** ($K = 2$, $A \in \{0, 1\}$): The `owl`, `rwl` and `earl` functions are called from the `R` package `DynTxRegime` to fit OWL, RWL and EARL respectively. The propensity score model (`moPropen`) is specified as the logistic model of A with respect to $(1, \boldsymbol{X}^\intercal)^\intercal$ with the ℓ_1-penalty, which is fitted by `glmnet`. The outcome models, including the main effect model (`moMain`, used in `rwl`) and the contrast model (`moCont`, used in `rwl` and `earl`), are both specified as linear in $(1, \boldsymbol{X}^\intercal)^\intercal$ with the ℓ_1-penalty, which are fitted by `glmnet`. The corresponding outcome mean model is $\mathbb{E}(Y|\boldsymbol{X}, A) = \texttt{moMain}(\boldsymbol{X}) + A \times \texttt{moCont}(\boldsymbol{X})$. These methods are fitted with linear decision functions (`kernel = "linear"`). For `owl` and `earl`, the hinge surrogate loss is used (`surrogate = "hinge"`). For `rwl`, the surrogate loss is the smoothed ramp loss (Zhou et al., 2017). The tuning parameter λ for all methods is determined by 5-fold cross validation (`cvFolds = 5`). The sequence of λ's for tuning is determined analog to `glmnet` (Friedman et al., 2010).

- **Policy Learning:** (general K) We use the `policy_tree` function from the `policytree` package (Sverdrup et al., 2020) in `R` to fit policy learning with decision trees. The outcome mean function and the propensity score function are both fitted by `regression_forest` from the `grf` package.

3.8.4 COSSO Estimate of the Working Variance Function

In this section, we consider the implementation details of estimating $\sigma^2_{\text{opt}}(\boldsymbol{x}, A; \widehat{\mu}_{0,n})$ from COSSO (Lin and Zhang, 2006). Specifically, we perform nonparametric regression using the squared working residual $\widehat{e}^2$ as the response and $(\boldsymbol{X}, A)$ as the covariates.

First of all, we discuss an SS-ANOVA model in terms of the covariate vector $\boldsymbol{X} = (X_1, \cdots, X_p)^\intercal \in \mathbb{R}^p$ and the treatment variable $A \in \{1, 2, \cdots, K\}$. If the j-th variable X_j is continuously ranged, then the we first scale the domain of X_j to $[0, 1]$, and consider the j-th covariate function space as $\mathcal{H}_j = \mathcal{S}_2$, where $(\mathcal{S}_2, \|\cdot\|_{\mathcal{S}_2})$ is the second order Sobolev Hilbert space:

$$\begin{aligned}
\mathcal{S}_2 &= \left\{ f : [0,1] \to \mathbb{R} \;\middle|\; f \text{ and } f' \text{ are absolutely continuous, } \int f''(x)^2 \mathrm{d}x < +\infty \right\}; \\
\|f\|_{\mathcal{S}_2}^2 &= \left(\int_0^1 f(x) \mathrm{d}x \right)^2 + \left(\int_0^1 f'(x) \mathrm{d}x \right)^2 + \int_0^1 f''(x)^2 \mathrm{d}x.
\end{aligned}$$

In particular, $\mathcal{H}_j = \mathcal{S}_2$ can be decomposed as $\{1\} \oplus \bar{\mathcal{H}}_j$ (Gu, 2013, Equation (2.26)), where $\bar{\mathcal{H}}_j$ is the *reproducing kernel Hilbert space (RKHS)* corresponding to the kernel function

$$\bar{\kappa}_j(x, x') = k_1(x)k_1(x') + k_2(x)k_2(x') - k_4(|x - x'|); \quad x, x' \in [0, 1].$$

Here, $k_1(x) = x - 0.5$, $k_2(x) = (1/2)[k_1(x)^2 - 1/12]$, and $k_4(x) = (1/24)[k_1(x)^4 - k_1(x)^2/2 + 7/240]$. If X_j takes finitely many values in $\{1, 2, \cdots, L_j\}$, then we consider $\mathcal{H}_j = \mathbb{R}^{L_j}$, which can be further decomposed as $\{1\} \oplus \bar{\mathcal{H}}_j$. Here, $\bar{\mathcal{H}}_j = \{(\alpha_1, \alpha_2, \cdots, \alpha_{L_j})^\intercal \in \mathbb{R}^{L_j} : \sum_{l=1}^{L_j} \alpha_l = 0\}$, and can be regarded as an L_j-dimensional RKHS corresponding to the kernel matrix $[\bar{\kappa}_j(x, x')]_{x,x'=1}^{L_j} = \mathsf{I}_{L_j \times L_j} - (1/L_j)\vec{\mathbf{1}}_{L_j}\vec{\mathbf{1}}_{L_j}^\intercal$. Similarly, since the treatment variable A is valued in $\{1, 2, \cdots, K\}$, we also consider the treatment function space $\mathcal{H}_{\mathscr{A}} = \mathbb{R}^K$ with the decomposition $\mathcal{H}_{\mathscr{A}} = \{1\} \oplus \bar{\mathcal{H}}_{\mathscr{A}}$, where $\bar{\mathcal{H}}_{\mathscr{A}}$ is the subspace of $\mathbb{R}^K$ with the sum-to-zero constraint and corresponds to the kernel matrix $[\bar{\kappa}_{\mathscr{A}}(a, a')]_{a,a'=1}^K = \mathsf{I}_{K \times K} - (1/K)\vec{\mathbf{1}}_K\vec{\mathbf{1}}_K^\intercal$.

The SS-ANOVA model is based on the following tensor-product RKHS (Gu, 2013, Section 2.4.1):

$$\begin{aligned}\mathcal{H} := \left[\bigotimes_{j=1}^{d}\mathcal{H}_j\right]\otimes\mathcal{H}_{\mathscr{A}} &= \left[\bigotimes_{j=1}^{d}\left(\{1\}\oplus\bar{\mathcal{H}}_j\right)\right]\otimes\left(\{1\}\oplus\bar{\mathcal{H}}_{\mathscr{A}}\right)\\ &= \underbrace{\{1\}}_{\text{global main effect}} \oplus \underbrace{\left[\bigoplus_{j=1}^{d}\bar{\mathcal{H}}_j\right]}_{\text{covariate main effects}} \oplus \underbrace{\bar{\mathcal{H}}_{\mathscr{A}}}_{\text{treatment main effect}}\\ &\quad\oplus \underbrace{\left[\bigoplus_{j=1}^{d}\left(\bar{\mathcal{H}}_j\otimes\bar{\mathcal{H}}_{\mathscr{A}}\right)\right]}_{\text{covariate-treatment interaction effect}}\\ &\quad\oplus \underbrace{\left\{\bigoplus_{\substack{\mathscr{J}\subseteq\{1,2,\cdots,d\}\\ |\mathscr{J}|\geqslant 2}}\left[\left(\bigotimes_{j\in\mathscr{J}}\bar{\mathcal{H}}_j\right)\oplus\left(\bigotimes_{j\in\mathscr{J}}\bar{\mathcal{H}}_j\otimes\bar{\mathcal{H}}_{\mathscr{A}}\right)\right]\right\}}_{\text{higher-order interaction effects}}.\end{aligned}$$

Here, we only consider first four effects from the above tensor-sum decomposition and ignore the higher-order interaction effects. Then the SS-ANOVA model is

$$\begin{aligned}\mathbb{E}(\hat{e}^2|\boldsymbol{X},A) = &\underbrace{\nu_0}_{\text{global main effect}} + \underbrace{\sum_{j=1}^{p}f_j(X_j)}_{\text{covariate main effect}} + \underbrace{\sum_{k=1}^{K}\alpha_k}_{\text{treatment main effect}}\\ &+ \underbrace{\sum_{j=1}^{K}\sum_{k=1}^{K}f_{jk}(X_j)}_{\text{covariate-treatment interaction effect}} + \underbrace{u}_{\text{remainder}}.\end{aligned}$$

In particular, the tensor-product RKHS $\bar{\mathcal{H}}_j\otimes\bar{\mathcal{H}}_{\mathscr{A}}$, which models the covariate-treatment interaction effect, corresponds to the kernel function

$$(\bar{\kappa}_j\otimes\bar{\kappa}_{\mathscr{A}})\Big((x_j,a)^{\intercal},(x_j',a')^{\intercal}\Big) = \bar{\kappa}_j(x_j,x_j')\bar{\kappa}_{\mathscr{A}}(a,a').$$

Then the COSSO estimate $\widehat{\sigma}_n^2(\boldsymbol{X},A)$ of the working variance function is obtained by solving:

$$\min_{f\in\mathcal{H}}\left\{\frac{1}{n}\sum_{i=1}^{n}[e_i^2-f(\boldsymbol{X}_i,A_i)]^2+\lambda_{\sigma^2}\left(\sum_{j=1}^{d}\|f\|_{\bar{\mathcal{H}}_j}+\|f\|_{\bar{\mathcal{H}}_{\mathscr{A}}}+\sum_{j=1}^{d}\|f\|_{\bar{\mathcal{H}}_j\otimes\bar{\mathcal{H}}_{\mathscr{A}}}\right)\right\}.$$

Here, $\|\cdot\|_{\bar{\mathcal{H}}_j}$, $\|\cdot\|_{\bar{\mathcal{H}}_{\mathscr{A}}}$ and $\|\cdot\|_{\bar{\mathcal{H}}_j \otimes \bar{\mathcal{H}}_{\mathscr{A}}}$ are the RKHS-norms corresponding to the associated component spaces, and λ_{σ^2} is a tuning parameter.

For implementation, we define the empirical kernel matrices $\bar{\mathsf{K}}_j := [\bar{\kappa}_j(X_{ij}, X_{i'j})]_{i,i'=1}^n$ and $\bar{\mathsf{K}}_{\mathscr{A}} := [\bar{\kappa}_{\mathscr{A}}(A_i, A_{i'})]_{i,i'=1}^n$. Then $\bar{\mathsf{K}}_j$, $\bar{\mathsf{K}}_{\mathscr{A}}$ and $\bar{\mathsf{K}}_{j,\mathscr{A}} := \bar{\mathsf{K}}_j \odot \bar{\mathsf{K}}_{\mathscr{A}}$ are the empirical kernel matrices on $\bar{\mathcal{H}}_j$, $\bar{\mathcal{H}}_{\mathscr{A}}$ and $\bar{\mathcal{H}}_j \otimes \bar{\mathcal{H}}_{\mathscr{A}}$ respectively, where $\bar{\mathsf{K}}_j \odot \bar{\mathsf{K}}_{\mathscr{A}}$ is the elementwise product of $\bar{\mathsf{K}}_j$ and $\bar{\mathsf{K}}_{\mathscr{A}}$. For a vector $\boldsymbol{\theta} := (\theta_1, \cdots, \theta_d; \theta_{\mathscr{A}}; \theta_{1,\mathscr{A}}, \cdots, \theta_{d,\mathscr{A}})^\intercal \in \mathbb{R}_+^{2d+1}$ of kernel weights, we write $\bar{\mathsf{K}}_{\boldsymbol{\theta}} := \sum_{j=1}^d \theta_j \bar{\mathsf{K}}_j + \theta_{\mathscr{A}} \bar{\mathsf{K}}_{\mathscr{A}} + \sum_{j=1}^d \theta_{j,\mathscr{A}} \bar{\mathsf{K}}_{j,\mathscr{A}} \in \mathbb{R}^{n \times n}$ as the weighted sum of the empirical kernel matrices. For a vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ of representer coefficients, we write $\mathsf{G}_{\boldsymbol{\alpha}} := [\bar{\mathsf{K}}_1 \boldsymbol{\alpha}, \cdots, \bar{\mathsf{K}}_d \boldsymbol{\alpha}; \bar{\mathsf{K}}_{\mathscr{A}} \boldsymbol{\alpha}; \bar{\mathsf{K}}_{1,\mathscr{A}} \boldsymbol{\alpha}, \cdots, \bar{\mathsf{K}}_{d,\mathscr{A}} \boldsymbol{\alpha}] \in \mathbb{R}^{n \times (2d+1)}$ as the gram matrix of the componentwise prediction values. We also denote $\vec{\boldsymbol{e}}^2 := (e_1^2, \cdots, e_n^2)^\intercal$ as the empirical squared residual vector. Then we fit a COSSO model by calling the `R` function `cosso::cosso.Gaussian` with the aforementioned kernel matrices and the squared residual vector as inputs. In particular, a random subset of sample points with size $\max\{40, \lceil 12n^{2/9} \rceil\}$ is used for representers. The following two steps are alternatively implemented:

- For a given kernel weight vector $\boldsymbol{\theta}$, we solve

$$\min_{b, \boldsymbol{\alpha}} \left\{ \frac{1}{n} \left\| \vec{\boldsymbol{e}}^2 - b\vec{\mathbf{1}}_n - \bar{\mathsf{K}}_{\boldsymbol{\theta}} \boldsymbol{\alpha} \right\|_2^2 + \lambda_0 \boldsymbol{\alpha}^\intercal \bar{\mathsf{K}}_{\boldsymbol{\theta}} \boldsymbol{\alpha} \right\}$$

  for the representer coefficient vector $(b, \boldsymbol{\alpha}^\intercal)^\intercal$;

- For a given representer coefficient vector $(b, \boldsymbol{\alpha}^\intercal)^\intercal$, we solve

$$\min_{\boldsymbol{\theta}} \left\{ \frac{1}{n} \left\| \vec{\boldsymbol{e}}^2 - b\vec{\mathbf{1}}_n - \mathsf{G}_{\boldsymbol{\alpha}} \boldsymbol{\theta} \right\|_2^2 + \lambda_0 \boldsymbol{\alpha}^\intercal \mathsf{G}_{\boldsymbol{\alpha}} \boldsymbol{\theta} \quad \text{subject to} \quad \boldsymbol{\theta} \in \mathbb{R}_+^{2d+1},\ \vec{\mathbf{1}}_{2d+1}^\intercal \boldsymbol{\theta} \leqslant M \right\},$$

  for the kernel weight vector $\boldsymbol{\theta}$.

Here, the tuning parameters (λ_0, M) are chosen according to Lin and Zhang (2006, Section 6).

3.8.5 Technical Proofs

3.8.5.1 Proof of Theorem 3.1

Proof of Theorem 3.1.

$$
\begin{aligned}
\mathcal{V}(d^\star) - \mathcal{V}(\widehat{d}_n) =& \mathbb{E}\left\{\sum_{k=1}^{K} \gamma(\boldsymbol{X}, k)\{\mathbb{1}[d^\star(\boldsymbol{X}) = k] - \mathbb{1}[\widehat{d}_n(\boldsymbol{X}) = k]\}\right\} \\
\leqslant& \mathbb{E}\left\{\sum_{k \neq k'} |\gamma(\boldsymbol{X}, k) - \gamma(\boldsymbol{X}, k')|; d^\star(\boldsymbol{X}) = k,\ \widehat{d}_n(\boldsymbol{X}) = k'\right\} \\
=& \frac{1}{2}\mathbb{E}\Bigg\{\sum_{k \neq k'} |\gamma(\boldsymbol{X}, k) - \gamma(\boldsymbol{X}, k')|; \\
& \qquad d^\star(\boldsymbol{X}) = k \text{ and } \widehat{d}_n(\boldsymbol{X}) = k' \text{ or } d^\star(\boldsymbol{X}) = k' \text{ and } \widehat{d}_n(\boldsymbol{X}) = k\Bigg\} \\
\leqslant& \frac{1}{2}\mathbb{E}\Bigg\{\sum_{k \neq k'} |\gamma(\boldsymbol{X}, k) - \gamma(\boldsymbol{X}, k')| + |\widehat{\gamma}_n(\boldsymbol{X}, k) - \widehat{\gamma}_n(\boldsymbol{X}, k')|; \\
& \qquad d^\star(\boldsymbol{X}) = k \text{ and } \widehat{d}_n(\boldsymbol{X}) = k' \text{ or } d^\star(\boldsymbol{X}) = k' \text{ and } \widehat{d}_n(\boldsymbol{X}) = k\Bigg\} \\
=& \frac{1}{2}\mathbb{E}\Bigg\{\sum_{k \neq k'} \left|[\gamma(\boldsymbol{X}, k) - \gamma(\boldsymbol{X}, k')] - [\widehat{\gamma}_n(\boldsymbol{X}, k) - \widehat{\gamma}_n(\boldsymbol{X}, k')]\right|; \\
& \qquad d^\star(\boldsymbol{X}) = k \text{ and } \widehat{d}_n(\boldsymbol{X}) = k' \text{ or } d^\star(\boldsymbol{X}) = k' \text{ and } \widehat{d}_n(\boldsymbol{X}) = k\Bigg\} \qquad (\star) \\
\leqslant& \frac{1}{2}\mathbb{E}\Bigg\{\sum_{k \neq k'} |\gamma(\boldsymbol{X}, k) - \widehat{\gamma}_n(\boldsymbol{X}, k)| + |\gamma(\boldsymbol{X}, k') - \widehat{\gamma}_n(\boldsymbol{X}, k')|; \\
& \qquad d^\star(\boldsymbol{X}) = k \text{ and } \widehat{d}_n(\boldsymbol{X}) = k' \text{ or } d^\star(\boldsymbol{X}) = k' \text{ and } \widehat{d}_n(\boldsymbol{X}) = k\Bigg\} \\
\leqslant& \mathbb{E}\left|\gamma(\boldsymbol{X}, d^\star) - \widehat{\gamma}_n(\boldsymbol{X}, d^\star)\right| + \mathbb{E}\left|\gamma(\boldsymbol{X}, \widehat{d}_n) - \widehat{\gamma}_n(\boldsymbol{X}, \widehat{d}_n)\right| \\
\leqslant& 2 \max_{1 \leqslant k \leqslant K} \mathbb{E}\left|\gamma(\boldsymbol{X}, k) - \widehat{\gamma}_n(\boldsymbol{X}, k)\right|.
\end{aligned}
$$

The equality $(\star)$ holds since the event $d^\star(\boldsymbol{X}) = k$ and $\widehat{d}_n(\boldsymbol{X}) = k'$ or $d^\star(\boldsymbol{X}) = k'$ and $\widehat{d}_n(\boldsymbol{X}) = k$ implies that

$$
[\gamma(\boldsymbol{X}, k) - \gamma(\boldsymbol{X}, k')][\widehat{\gamma}_n(\boldsymbol{X}, k) - \widehat{\gamma}_n(\boldsymbol{X}, k')] < 0.
$$

□

3.8.5.2 Proof of Lemma 3.2

Proof of Lemma 3.2. Denote

$$\widetilde{\Lambda} := \Big\{ \boldsymbol{H} \in \mathcal{H} : \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}, A) = \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}) \Big\}.$$

Consider the following family of parametric submodels of (3.1):

$$\mathscr{P}_{\boldsymbol{\beta},\alpha} := \Big\{ (\boldsymbol{X}, A, Y) \sim p_{\mathscr{X}}(\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{X}}) \times p_{\mathscr{A}}(A|\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{A}}) \times p_{\epsilon}\big(\overbrace{Y - \mu_0(\boldsymbol{X}; \boldsymbol{\alpha}_{\mu_0}) - \gamma(\boldsymbol{X}, A; \boldsymbol{\beta})}^{\epsilon} \big| \boldsymbol{X}, A; \boldsymbol{\alpha}_{\epsilon} \big) \Big\},$$

where $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\mathscr{X}} \oplus \boldsymbol{\alpha}_{\mathscr{A}} \oplus \boldsymbol{\alpha}_{\epsilon} \oplus \boldsymbol{\alpha}_{\mu_0}$ are finite-dimensional parameters for the nuisance components $\eta = (p_{\mathscr{X}}, p_{\mathscr{A}}, p_{\epsilon}, \mu_0)$. Then the nuisance score vector is

$$\boldsymbol{S}_{\boldsymbol{\alpha}} = \begin{matrix} \boldsymbol{S}_{\mathscr{X}} \\ \boldsymbol{S}_{\mathscr{A}} \\ \boldsymbol{S}_{\epsilon} \\ \boldsymbol{S}_{\mu_0} \end{matrix} \begin{pmatrix} \frac{\partial \log p_{\mathscr{X}}(\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{X}})}{\partial \boldsymbol{\alpha}_{\mathscr{X}}} \\ \frac{\partial \log p_{\mathscr{A}}(A|\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{A}})}{\partial \boldsymbol{\alpha}_{\mathscr{A}}} \\ \frac{\partial \log p_{\epsilon}(\epsilon|\boldsymbol{X}, A; \boldsymbol{\alpha}_{\epsilon})}{\partial \boldsymbol{\alpha}_{\epsilon}} \\ -\frac{\partial \log p_{\epsilon}(\epsilon|\boldsymbol{X}, A; \boldsymbol{\alpha}_{\epsilon})}{\partial \epsilon} \frac{\partial \mu_0(\boldsymbol{X}; \boldsymbol{\alpha}_{\mu_0})}{\partial \boldsymbol{\alpha}_{\mu_0}} \end{pmatrix}.$$

Then the nuisance tangent space of the family of submodels $\mathscr{P}_{\boldsymbol{\beta},\boldsymbol{\alpha}}$ is defined as (Tsiatis, 2007):

$$\begin{aligned} \Lambda_{\boldsymbol{\alpha}} =& \Big\{ \mathsf{B}\boldsymbol{S}_{\boldsymbol{\alpha}} : \mathsf{B} \in \mathbb{R}^{p \times \dim(\boldsymbol{\alpha})} \Big\} \\ =& \Big\{ \mathsf{B}_{\mathscr{X}}\boldsymbol{S}_{\mathscr{X}} + \mathsf{B}_{\mathscr{A}}\boldsymbol{S}_{\mathscr{A}} + \mathsf{B}_{\epsilon}\boldsymbol{S}_{\epsilon} + \mathsf{B}_{\mu_0}\boldsymbol{S}_{\mu_0} : \\ & \mathsf{B}_{\mathscr{X}} \in \mathbb{R}^{p \times \dim(\boldsymbol{\alpha}_{\mathscr{X}})},\ \mathsf{B}_{\mathscr{A}} \in \mathbb{R}^{p \times \dim(\boldsymbol{\alpha}_{\mathscr{A}})}, \mathsf{B}_{\epsilon} \in \mathbb{R}^{p \times \dim(\boldsymbol{\alpha}_{\epsilon})}, \mathsf{B}_{\mu_0} \in \mathbb{R}^{p \times \dim(\boldsymbol{\alpha}_{\mu_0})} \Big\}. \end{aligned}$$

We aim to show that $\Lambda_{\boldsymbol{\alpha}} \subseteq \widetilde{\Lambda}$. By $\widetilde{\Lambda}$ is a linear space, it is equivalent to show that $\boldsymbol{S}_{\mathscr{X}}, \boldsymbol{S}_{\mathscr{A}}, \boldsymbol{S}_{\epsilon}, \boldsymbol{S}_{\mu_0} \in \widetilde{\Lambda}$. Denote $\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}$ as the expectation under the submodel parametrized by $(\boldsymbol{\beta}, \boldsymbol{\alpha})$.

- Since $\boldsymbol{S}_{\mathscr{X}}$ is a function of $\boldsymbol{X}$, we have that

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mathscr{X}}\epsilon|\boldsymbol{X}, A) = \boldsymbol{S}_{\mathscr{X}}\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\epsilon|\boldsymbol{X}, A) = \mathbf{0} = \boldsymbol{S}_{\mathscr{X}}\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\epsilon|\boldsymbol{X}) = \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mathscr{X}}\epsilon|\boldsymbol{X}).$$

That is, $\boldsymbol{S}_{\mathscr{X}} \in \widetilde{\Lambda}$.

- Since $\boldsymbol{S}_{\mathscr{A}}$ is a function of $(\boldsymbol{X}, A)$, we have that

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mathscr{A}}\epsilon|\boldsymbol{X}, A) = \boldsymbol{S}_{\mathscr{A}}\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\epsilon|\boldsymbol{X}, A) = \mathbf{0}.$$

On the other hand, by $\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\epsilon|\boldsymbol{X}) = 0$ for any $\boldsymbol{\alpha}$, we have

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\boldsymbol{\alpha}_{\mathscr{A}}}\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\epsilon|\boldsymbol{X}) \\
&= \frac{\partial}{\boldsymbol{\alpha}_{\mathscr{A}}}\int \epsilon p_{\mathscr{A}}(a|\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{A}})p_{\epsilon}(\epsilon|\boldsymbol{X}, a;\boldsymbol{\alpha}_{\epsilon})\mathrm{d}a\mathrm{d}\epsilon \\
&= \int \epsilon \times \frac{(\partial/\partial\boldsymbol{\alpha}_{\mathscr{A}})p_{\mathscr{A}}(a|\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{A}})}{p_{\mathscr{A}}(a|\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{A}})} \times p_{\mathscr{A}}(a|\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{A}})p_{\epsilon}(\epsilon|\boldsymbol{X}, a;\boldsymbol{\alpha}_{\epsilon})\mathrm{d}a\mathrm{d}\epsilon \\
&= \int \epsilon \times \frac{\partial \log p_{\mathscr{A}}(a|\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{A}})}{\partial\boldsymbol{\alpha}_{\mathscr{A}}} \times p_{\mathscr{A}}(a|\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{A}})p_{\epsilon}(\epsilon|\boldsymbol{X}, a;\boldsymbol{\alpha}_{\epsilon})\mathrm{d}a\mathrm{d}\epsilon \\
&= \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mathscr{A}}\epsilon|\boldsymbol{X}).
\end{aligned}$$

That is, $\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mathscr{A}}\epsilon|\boldsymbol{X}, A) = \mathbf{0} = \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mathscr{A}}\epsilon|\boldsymbol{X})$, and $\boldsymbol{S}_{\mathscr{A}} \in \widetilde{\Lambda}$.

- By $\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\epsilon|\boldsymbol{X}, A) = 0$ for any $\boldsymbol{\alpha}$, we have

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\boldsymbol{\alpha}_{\epsilon}}\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\epsilon|\boldsymbol{X}, A) \\
&= \frac{\partial}{\boldsymbol{\alpha}_{\epsilon}}\int \epsilon p_{\epsilon}(\epsilon|\boldsymbol{X}, A;\boldsymbol{\alpha}_{\epsilon})\mathrm{d}\epsilon \\
&= \int \epsilon \times \frac{(\partial/\partial\boldsymbol{\alpha}_{\epsilon})p_{\epsilon}(\epsilon|\boldsymbol{X}, A;\boldsymbol{\alpha}_{\epsilon})}{p_{\epsilon}(\epsilon|\boldsymbol{X}, A;\boldsymbol{\alpha}_{\epsilon})} \times p_{\epsilon}(\epsilon|\boldsymbol{X}, A;\boldsymbol{\alpha}_{\epsilon})\mathrm{d}\epsilon \\
&= \int \epsilon \times \frac{\partial \log p_{\epsilon}(\epsilon|\boldsymbol{X}, A;\boldsymbol{\alpha}_{\epsilon})}{\partial\boldsymbol{\alpha}_{\epsilon}} \times p_{\epsilon}(\epsilon|\boldsymbol{X}, A;\boldsymbol{\alpha}_{\epsilon})\mathrm{d}\epsilon \\
&= \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\epsilon}\epsilon|\boldsymbol{X}, A).
\end{aligned}$$

That is, $\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\epsilon}\epsilon|\boldsymbol{X}, A) = \mathbf{0} = \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\epsilon}\epsilon|\boldsymbol{X})$, and $\boldsymbol{S}_{\epsilon} \in \widetilde{\Lambda}$.

- Note that

$$\boldsymbol{S}_{\mu_0}\epsilon = -\frac{\epsilon\partial \log p_{\epsilon}(\epsilon|\boldsymbol{X}, A;\boldsymbol{\alpha}_{\epsilon})}{\partial\epsilon} \times \underbrace{\frac{\partial\mu_0(\boldsymbol{X};\boldsymbol{\alpha}_{\mu_0})}{\partial\boldsymbol{\alpha}_{\mu_0}}}_{\text{function of } \boldsymbol{X}}.$$

Lemma 3.12. *Suppose $X \sim p(x)$ where $p(x)$ is a probability density function with respect to the Lebesgue measure on $\mathbb{R}$, such that $|x|p(x) \to 0$ as $|x| \to +\infty$. Then*

$$\mathbb{E}\left(\frac{X \mathrm{d} \log p(X)}{\mathrm{d}x}\right) = -1.$$

By Lemma 3.12, we have

$$\mathbb{E}\left(\frac{\epsilon \partial \log p_\epsilon(\epsilon|\boldsymbol{X}, A; \boldsymbol{\alpha}_\epsilon)}{\partial \epsilon}\middle| \boldsymbol{X}, A\right) = -1. \tag{3.15}$$

Then

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mu_0}\epsilon|\boldsymbol{X}, A) = -\mathbb{E}\left(\frac{\epsilon \log p_\epsilon(\epsilon|\boldsymbol{X}, A; \boldsymbol{\alpha}_\epsilon)}{\partial \epsilon}\middle| \boldsymbol{X}, A\right) \frac{\partial \mu_0(\boldsymbol{X}; \boldsymbol{\alpha}_{\mu_0})}{\partial \boldsymbol{\alpha}_{\mu_0}} = \frac{\partial \mu_0(\boldsymbol{X}; \boldsymbol{\alpha}_{\mu_0})}{\partial \boldsymbol{\alpha}_{\mu_0}},$$

which is a function of $\boldsymbol{X}$. Consequently, $\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mu_0}\epsilon|\boldsymbol{X}, A) = \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\boldsymbol{S}_{\mu_0}\epsilon|\boldsymbol{X})$, and $\boldsymbol{S}_{\mu_0} \in \widetilde{\Lambda}$.

Therefore, we can conclude that $\Lambda_{\boldsymbol{\alpha}} \subseteq \widetilde{\Lambda}$. By the definition of nuisance tangent space of the semiparametric model, we have that $\Lambda \subseteq \widetilde{\Lambda}$.

Next, we aim to justify $\Lambda \supseteq \widetilde{\Lambda}$. Fix an $\boldsymbol{H} = \boldsymbol{h}(\boldsymbol{X}, A, \epsilon) \in \widetilde{\Lambda}$, *i.e.*,

$$\mathbb{E}(\boldsymbol{H}) = \boldsymbol{0}; \quad \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}, A) = \mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X}).$$

We need to construct a parametric submodel whose nuisance score vector is $\boldsymbol{H}$. Consider the following orthogonal decompositions:

$$\boldsymbol{H} = \underbrace{\mathbb{E}(\boldsymbol{H}|\boldsymbol{X})}_{:=\boldsymbol{H}_{\mathscr{X}}} + \underbrace{\mathbb{E}(\boldsymbol{H}|\boldsymbol{X}, A) - \mathbb{E}(\boldsymbol{H}|\boldsymbol{X})}_{:=\boldsymbol{H}_{\mathscr{A}}} + \underbrace{\boldsymbol{H} - \mathbb{E}(\boldsymbol{H}|\boldsymbol{X}, A)}_{:=\boldsymbol{H}_{\epsilon}}.$$

Without loss of generality, we assume that $\|\boldsymbol{H}\|_2 \leqslant M < +\infty$. Define

$$\begin{aligned}
p_{\mathscr{X}}(\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{X}}) &:= p_{\mathscr{X}}(\boldsymbol{X})(1 + \boldsymbol{\alpha}_{\mathscr{X}}^\intercal \boldsymbol{H}_{\mathscr{X}}); && \|\boldsymbol{\alpha}_{\mathscr{X}}\|_2 \leqslant 1/M; \\
p_{\mathscr{A}}(A|\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{A}}) &:= p_{\mathscr{A}}(A|\boldsymbol{X})(1 + \boldsymbol{\alpha}_{\mathscr{A}}^\intercal \boldsymbol{H}_{\mathscr{A}}); && \|\boldsymbol{\alpha}_{\mathscr{A}}\|_2 \leqslant 1/M; \\
p_{\widetilde{\epsilon}}(\widetilde{\epsilon}|\boldsymbol{X}, A; \boldsymbol{\alpha}_\epsilon) &:= p_\epsilon(\widetilde{\epsilon}|\boldsymbol{X}, A)(1 + \boldsymbol{\alpha}_\epsilon^\intercal \boldsymbol{H}_\epsilon); && \|\boldsymbol{\alpha}_\epsilon\|_2 \leqslant 1/M.
\end{aligned}$$

Here, $p_{\mathscr{X}}(\boldsymbol{X})$, $p_{\mathscr{A}}(A|\boldsymbol{X})$ and $p_\epsilon(\epsilon|\boldsymbol{X}, A)$ are the densities of Model (3.1). Then we consider the following data generating process (DGP) parametrized by $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ where $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\mathscr{X}} \oplus \boldsymbol{\alpha}_{\mathscr{A}} \oplus \boldsymbol{\alpha}_\epsilon$:

$$\begin{aligned} & Y = \mu_0(\boldsymbol{X}) + \gamma(\boldsymbol{X}, A; \boldsymbol{\beta}) + \widetilde{\epsilon}; \\ \text{subject to} \quad & \sum_{k=1}^{K} \gamma(\boldsymbol{X}, k; \boldsymbol{\beta}) = 0; \quad \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\widetilde{\epsilon}|\boldsymbol{X}, A) = \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\widetilde{\epsilon}|\boldsymbol{X}); \\ & (\boldsymbol{X}, A, \widetilde{\epsilon}) \sim p_{\mathscr{X}}(\boldsymbol{x}; \boldsymbol{\alpha}_{\mathscr{X}}) p_{\mathscr{A}}(a|\boldsymbol{x}; \boldsymbol{\alpha}_{\mathscr{A}}) p_{\widetilde{\epsilon}}(\widetilde{\epsilon}|\boldsymbol{x}, a; \boldsymbol{\alpha}_\epsilon). \end{aligned}$$

Notice that the above DGP can be transformed to the form of (3.1) by setting $\mu_0(\boldsymbol{X}; \boldsymbol{\alpha}_\epsilon) := \mu_0(\boldsymbol{X}) + \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\widetilde{\epsilon}|\boldsymbol{X})$ and $\epsilon := \widetilde{\epsilon} - \mu_0(\boldsymbol{X}; \boldsymbol{\alpha}_\epsilon)$. Next, we verify that the above DGP is well defined and corresponds to the nuisance score vectors $\boldsymbol{H}_{\mathscr{X}}$, $\boldsymbol{H}_{\mathscr{A}}$ and $\boldsymbol{H}_\epsilon$.

- By

  $$|\boldsymbol{\alpha}_{\mathscr{X}}^\intercal \boldsymbol{H}_{\mathscr{X}}| \leqslant \|\boldsymbol{\alpha}_{\mathscr{X}}\|_2 \|\boldsymbol{H}\|_2 \leqslant 1; \quad |\boldsymbol{\alpha}_{\mathscr{A}}^\intercal \boldsymbol{H}_{\mathscr{A}}| \leqslant \|\boldsymbol{\alpha}_{\mathscr{A}}\|_2 \|\boldsymbol{H}\|_2 \leqslant 1; \quad |\boldsymbol{\alpha}_\epsilon^\intercal \boldsymbol{H}_\epsilon| \leqslant \|\boldsymbol{\alpha}_\epsilon\|_2 \|\boldsymbol{H}\|_2 \leqslant 1,$$

  we have $p_{\mathscr{X}}(\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{X}}), p_{\mathscr{A}}(A|\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{A}}), p_{\widetilde{\epsilon}}(\widetilde{\epsilon}|\boldsymbol{X}, A; \boldsymbol{\alpha}_\epsilon) \geqslant 0$.

- By $\mathbb{E}(\boldsymbol{H}_{\mathscr{X}}) = \boldsymbol{0}$, $\mathbb{E}(\boldsymbol{H}_{\mathscr{A}}|\boldsymbol{X}) = \boldsymbol{0}$ and $\mathbb{E}(\boldsymbol{H}_\epsilon|\boldsymbol{X}, A) = \boldsymbol{0}$, we have

  $$\begin{aligned} \int p_{\mathscr{X}}(\boldsymbol{x}; \boldsymbol{\alpha}_{\mathscr{X}}) \mathrm{d}\boldsymbol{x} &= \int p_{\mathscr{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + \boldsymbol{\alpha}_{\mathscr{X}}^\intercal \mathbb{E}(\boldsymbol{H}_{\mathscr{X}}) = 1; \\ \int p_{\mathscr{A}}(a|\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{A}}) \mathrm{d}a &= \int p_{\mathscr{A}}(a|\boldsymbol{X}) \mathrm{d}a + \boldsymbol{\alpha}_{\mathscr{A}}^\intercal \mathbb{E}(\boldsymbol{H}_{\mathscr{A}}|\boldsymbol{X}) = 1; \\ \int p_{\widetilde{\epsilon}}(\widetilde{\epsilon}|\boldsymbol{X}, A; \boldsymbol{\alpha}_\epsilon) \mathrm{d}\widetilde{\epsilon} &= \int p_\epsilon(\widetilde{\epsilon}|\boldsymbol{X}, A) \mathrm{d}\widetilde{\epsilon} + \boldsymbol{\alpha}_\epsilon^\intercal \mathbb{E}(\boldsymbol{H}_\epsilon|\boldsymbol{X}, A) = 1. \end{aligned}$$

  Therefore, $p_{\mathscr{X}}(\boldsymbol{x}; \boldsymbol{\alpha}_{\mathscr{X}})$, $p_{\mathscr{X}}(a|\boldsymbol{X}; \boldsymbol{\alpha}_{\mathscr{A}})$ and $p_{\widetilde{\epsilon}}(\widetilde{\epsilon}|\boldsymbol{X}, A; \boldsymbol{\alpha}_\epsilon)$ are probability density functions.

- The conditional mean restriction becomes

  $$\begin{aligned} \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\widetilde{\epsilon}|\boldsymbol{X}, A) &= \int \widetilde{\epsilon}(1 + \boldsymbol{\alpha}_\epsilon^\intercal \boldsymbol{H}_\epsilon) p_\epsilon(\widetilde{\epsilon}|\boldsymbol{X}, A) \mathrm{d}\widetilde{\epsilon} \\ &= \mathbb{E}(\epsilon|\boldsymbol{X}, A) + \boldsymbol{\alpha}_\epsilon^\intercal \mathbb{E}(\boldsymbol{H}_\epsilon \epsilon|\boldsymbol{X}, A) \\ &= \boldsymbol{\alpha}_\epsilon^\intercal \mathbb{E}(\boldsymbol{H}_\epsilon \epsilon|\boldsymbol{X}, A), \end{aligned}$$

which is a function of $\boldsymbol{X}$, by the fact that

$$\mathbb{E}(\boldsymbol{H}_\epsilon\epsilon|\boldsymbol{X},A)=\mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X},A)-\underbrace{\mathbb{E}(\boldsymbol{H}|\boldsymbol{X},A)\mathbb{E}(\epsilon|\boldsymbol{X},A)}_{=\mathbf{0}}=\mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X})-\underbrace{\mathbb{E}(\boldsymbol{H}|\boldsymbol{X},A)\mathbb{E}(\epsilon|\boldsymbol{X})}_{=\mathbf{0}}=\mathbb{E}(\boldsymbol{H}_\epsilon\epsilon|\boldsymbol{X}).$$

Then it can be clear that

$$\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\widetilde{\epsilon}|\boldsymbol{X},A)=\mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\alpha}}(\widetilde{\epsilon}|\boldsymbol{X}).$$

- The nuisance score vectors are

$$\begin{array}{rclcrcl}\boldsymbol{S}_{\mathscr{X}} &=& \left[\frac{\partial\log p_{\mathscr{X}}(\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{X}})}{\partial\boldsymbol{\alpha}_{\mathscr{X}}}\right]_{\boldsymbol{\alpha}_{\mathscr{X}}=\mathbf{0}} &=& \left[\frac{\boldsymbol{H}_{\mathscr{X}}}{1+\boldsymbol{\alpha}_{\mathscr{X}}^\intercal\boldsymbol{H}_{\mathscr{X}}}\right]_{\boldsymbol{\alpha}_{\mathscr{X}}=\mathbf{0}} &=& \boldsymbol{H}_{\mathscr{X}};\\ \boldsymbol{S}_{\mathscr{A}} &=& \left[\frac{\partial\log p_{\mathscr{A}}(A|\boldsymbol{X};\boldsymbol{\alpha}_{\mathscr{A}})}{\partial\boldsymbol{\alpha}_{\mathscr{A}}}\right]_{\boldsymbol{\alpha}_{\mathscr{A}}=\mathbf{0}} &=& \left[\frac{\boldsymbol{H}_{\mathscr{A}}}{1+\boldsymbol{\alpha}_{\mathscr{X}}^\intercal\boldsymbol{H}_{\mathscr{A}}}\right]_{\boldsymbol{\alpha}_{\mathscr{A}}=\mathbf{0}} &=& \boldsymbol{H}_{\mathscr{A}};\\ \boldsymbol{S}_{\epsilon} &=& \left[\frac{\partial\log p_{\widetilde{\epsilon}}(\widetilde{\epsilon}|\boldsymbol{X},A;\boldsymbol{\alpha}_\epsilon)}{\partial\boldsymbol{\alpha}_\epsilon}\right]_{\boldsymbol{\alpha}_\epsilon=\mathbf{0}} &=& \left[\frac{\boldsymbol{H}_\epsilon}{1+\boldsymbol{\alpha}_\epsilon^\intercal\boldsymbol{H}_\epsilon}\right]_{\boldsymbol{\alpha}_\epsilon=\mathbf{0}} &=& \boldsymbol{H}_\epsilon.\end{array}$$

Therefore, the DGP above corresponds to the nuisance score vectors $\boldsymbol{H}$, so that $\Lambda\supseteq\widetilde{\Lambda}$. That is,

$$\Lambda=\Big\{\boldsymbol{H}\in\mathcal{H}:\mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X},A)=\mathbb{E}(\boldsymbol{H}\epsilon|\boldsymbol{X})\Big\}.$$

□

Proof of Lemma 3.12.

$$\begin{aligned}\mathbb{E}\left(\frac{X\partial\log p(X)}{\partial x}\right)&=\int\frac{x\partial\mathrm{d}p(x)}{\mathrm{d}x}p(x)\mathrm{d}x\\&=\int x\frac{\mathrm{d}p(x)}{\mathrm{d}x}\times\frac{1}{p(x)}\times p(x)\mathrm{d}x\\&=\int x\mathrm{d}p(x)\\&=\Big[xp(x)\Big]_{x\to-\infty}^{x\to+\infty}-\int p(x)\mathrm{d}x\\&=-1.\end{aligned}$$

□

3.8.5.3 Proof of Lemmas 3.3 and 3.4

Proof of Lemma 3.3. Since the column space of Ω is the orthogonal complement of the space spanned by $\mathbf{1}_K$, we have

$$\Omega(\Omega^\intercal\Omega)^{-1}\Omega^\intercal = \mathbf{I}_{K\times K} - \frac{1}{K}\mathbf{1}_{K\times 1}^{\otimes 2}.$$

In particular, for $1 \leqslant k, k' \leqslant K$, we have5

$$\langle(\Omega^\intercal\Omega)^{-1/2}\boldsymbol{\omega}_k, (\Omega^\intercal\Omega)^{-1/2}\boldsymbol{\omega}_{k'}\rangle = \left(1-\frac{1}{K}\right)^{-1}[\Omega(\Omega^\intercal\Omega)^{-1}\Omega^\intercal]_{kk'} = \mathbb{1}(k=k') - \frac{1}{K-1}\mathbb{1}(k\neq k').$$

Therefore, $\{\boldsymbol{\omega}_k\}_{k=1}^K$ are unit vectors and equiangular. □

Proof of Lemma 3.4. By Lemma 3.3, we have

$$\Omega(\Omega^\intercal\Omega)^{-1}\Omega^\intercal = \mathbf{I}_{K\times K} - (1/K)\vec{\mathbf{1}}_{K\times 1}^{\otimes 2}.$$

Denote $\vec{\boldsymbol{\gamma}}(\boldsymbol{x};\boldsymbol{\beta}) := \big(\gamma(\boldsymbol{x},1;\boldsymbol{\beta}), \gamma(\boldsymbol{x},2;\boldsymbol{\beta}), \cdots, \gamma(\boldsymbol{x},K;\boldsymbol{\beta})\big)^\intercal$. Then

$$\begin{aligned}
\sqrt{1-1/K}\,\Omega\vec{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\beta}) &= \Omega(\Omega^\intercal\Omega)^{-1}\Omega^\intercal\vec{\boldsymbol{\gamma}}(\boldsymbol{x};\boldsymbol{\beta}) \\
&= \vec{\boldsymbol{\gamma}}(\boldsymbol{x};\boldsymbol{\beta}) - \underbrace{\frac{1}{K}\sum_{k=1}^K \gamma(\boldsymbol{x},k;\boldsymbol{\beta})}_{=0}\vec{\mathbf{1}}_K \\
&= \vec{\boldsymbol{\gamma}}(\boldsymbol{x};\boldsymbol{\beta}).
\end{aligned}$$

That is,

$$\gamma(\boldsymbol{x},k;\boldsymbol{\beta}) = \left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_k, \vec{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\beta})\rangle; \quad 1 \leqslant k \leqslant K.$$

□

3.8.5.4 Proof of Lemma 3.5

Proof of Lemma 3.5. Suppose $\boldsymbol{H} \in \Lambda$ and $\mathsf{H}: \mathcal{X} \to \mathbb{R}^{p\times(K-1)}$. Then

$$\mathbb{E}\left[\boldsymbol{H}\left(\frac{\mathsf{H}(\boldsymbol{X})\boldsymbol{\omega}_A\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})}\right)^\intercal\right] = \mathbb{E}\left[\mathbb{E}\left(\frac{\boldsymbol{H}\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})}\middle|\boldsymbol{X}\right)\mathsf{H}(\boldsymbol{X})^\intercal\right] = \mathsf{O}_{p\times p}.$$

That is,

$$\Lambda \perp \left\{ \frac{\mathsf{H}(\boldsymbol{X})\boldsymbol{\omega}_A\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \mathsf{H} : \mathcal{X} \to \mathbb{R}^{p\times(K-1)} \right\}.$$

Now suppose $\boldsymbol{H} \in \mathcal{H}$. Define

$$\boldsymbol{H}_1 := \underbrace{\mathbb{E}\left\{ \frac{\boldsymbol{H}\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\} \mathsf{V}_\epsilon(\boldsymbol{X})^{-1}}_{\mathbb{R}^{p\times(K-1)}\text{-valued function of } \boldsymbol{X}} \frac{\boldsymbol{\omega}_A\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})}; \quad \boldsymbol{H}_2 := \boldsymbol{H} - \boldsymbol{H}_1.$$

Then the following shows that $\boldsymbol{H}_2 \in \Lambda$:

$$\mathbb{E}\left(\frac{\boldsymbol{H}_2\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right) = \mathbb{E}\left(\frac{\boldsymbol{H}\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right) - \mathbb{E}\left\{ \frac{\boldsymbol{H}\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\} \mathsf{V}_\epsilon(\boldsymbol{X})^{-1}\mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\epsilon^2}{p_{\mathscr{A}}(A|\boldsymbol{X})^2} \middle| \boldsymbol{X} \right) = \mathsf{O}_{p\times(K-1)},$$

where

$$\mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\epsilon^2}{p_{\mathscr{A}}(A|\boldsymbol{X})^2} \middle| \boldsymbol{X} \right) = \mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\sigma^2(\boldsymbol{X},A)}{p_{\mathscr{A}}(A|\boldsymbol{X})^2} \middle| \boldsymbol{X} \right) = \sum_{k=1}^{K} \frac{\sigma^2(\boldsymbol{X},A)\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(k|\boldsymbol{X})} = \mathsf{V}_\epsilon(\boldsymbol{X}).$$

Therefore,

$$\Lambda^\perp = \left\{ \frac{\mathsf{H}(\boldsymbol{X})\boldsymbol{\omega}_A\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \mathsf{H} : \mathcal{X} \to \mathbb{R}^{p\times(K-1)} \right\}.$$

□

3.8.5.5 Proof of Proposition 3.6

Proof of Proposition 3.6. The score vector of Model (3.1) is defined as (Tsiatis, 2007)

$$\boldsymbol{S}_{\boldsymbol{\beta}} = \frac{\partial \log p_\epsilon\Big(Y - \mu_0(\boldsymbol{X}) - \gamma(\boldsymbol{X},A;\boldsymbol{\beta}) \Big| \boldsymbol{X}, A\Big)}{\partial\boldsymbol{\beta}} = -\dot{\gamma}(\boldsymbol{X},A;\boldsymbol{\beta})\frac{\partial \log p_\epsilon(\epsilon|\boldsymbol{X},A)}{\partial\epsilon}.$$

In particular,

$$\mathbb{E}\left\{ \frac{\boldsymbol{S}_{\boldsymbol{\beta}}\boldsymbol{\omega}_A^\intercal\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\} = \mathbb{E}\left\{ \frac{\dot{\gamma}(\boldsymbol{X},A;\boldsymbol{\beta})\boldsymbol{\omega}_A^\intercal}{p_{\mathscr{A}}(A|\boldsymbol{X})} \left[-\frac{\epsilon\partial \log p_\epsilon(\epsilon|\boldsymbol{X},A)}{\partial\epsilon} \right] \middle| \boldsymbol{X} \right\} \overset{(3.15)}{=} \mathbb{E}\left\{ \frac{\dot{\gamma}(\boldsymbol{X},A;\boldsymbol{\beta})\boldsymbol{\omega}_A^\intercal}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\}.$$

Then by Lemma 3.5, the efficient score (Tsiatis, 2007) is

$$\boldsymbol{S}_{\text{eff}} = \mathbb{E}(\boldsymbol{S}_{\boldsymbol{\beta}}|\Lambda^\perp) = \mathbb{E}\left\{ \frac{\dot{\gamma}(\boldsymbol{X},A;\boldsymbol{\beta})\boldsymbol{\omega}_A^\intercal}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\} \mathsf{V}_\epsilon(\boldsymbol{X})^{-1}\frac{\boldsymbol{\omega}_A\epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})}.$$

Consider the angle-based decision function (3.4). We have

$$\dot{\gamma}(\boldsymbol{X}, A; \boldsymbol{\beta}) = \left(1 - \frac{1}{K}\right) \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^{\intercal} \boldsymbol{\omega}_A.$$

Then

$$\begin{aligned}
&\mathbb{E}\left\{ \frac{\dot{\gamma}(\boldsymbol{X}, A; \boldsymbol{\beta}) \boldsymbol{\omega}_A^{\intercal}}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\} = \left(1 - \frac{1}{K}\right) \mathbb{E}\left\{ \frac{\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^{\intercal} \boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right\} \\
=&\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^{\intercal} \left(1 - \frac{1}{K}\right) \sum_{k=1}^{K} \boldsymbol{\omega}_k^{\otimes 2} = \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^{\intercal} \Omega^{\intercal} \Omega.
\end{aligned}$$

Therefore,

$$\boldsymbol{S}_{\text{eff}} = \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^{\intercal} \Omega^{\intercal} \Omega \mathsf{V}_{\epsilon}(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A \epsilon}{p_{\mathscr{A}}(A|\boldsymbol{X})}.$$

□

3.8.5.6 Proof of Proposition 3.7

Proof of Proposition 3.7. It follows from direct calculation that

$$\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2) | \boldsymbol{X}, A] = [\mu_0(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})] \mathsf{H}(\boldsymbol{X}) \frac{p_{\mathscr{A}}(A|\boldsymbol{X})}{\breve{p}_{\mathscr{A}}(A|\boldsymbol{X})} \frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}.$$

If $\breve{\mu}_0 = \mu_0$, then the above is $\mathbf{0}$. If $\breve{p}_{\mathscr{A}} = p_{\mathscr{A}}$, the above becomes

$$\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2) | \boldsymbol{X}, A] = [\mu_0(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})] \mathsf{H}(\boldsymbol{X}) \frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}.$$

Then we have

$$\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, \breve{p}_{\mathscr{A}}, \breve{\sigma}^2) | \boldsymbol{X}] = [\mu_0(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})] \mathsf{H}(\boldsymbol{X}) \mathbb{E}\left(\frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})} \middle| \boldsymbol{X} \right) = \mathbf{0}.$$

□

3.8.5.7 Proof of Lemma 3.8

Proof of Lemma 3.8. Denote $\dot{F}(\boldsymbol{X}) := \sup_{\breve{\boldsymbol{\beta}} \in \mathscr{B}} \|\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})\|_2$. Define

$$\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X}) := \sum_{k=1}^{K} \frac{\widehat{\sigma}_n^2(\boldsymbol{X}, A)\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})},$$

and

$$\widehat{\boldsymbol{h}}_{\text{eff},n}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}}) := \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega \widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})};$$
$$\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}}) := \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}.$$

Then

$$\begin{aligned}
&\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2) - \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2) \\
= \quad & [\underbrace{Y - \breve{\mu}_0(\boldsymbol{X}) - \gamma(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})}_{:=\breve{e}(\breve{\boldsymbol{\beta}})}][\widehat{\boldsymbol{h}}_{\text{eff},n}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}}) - \breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})] && (3.16) \\
& - [\widehat{\mu}_{0,n}(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})]\breve{\boldsymbol{h}}_{\text{eff},n}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}}) && (3.17) \\
& - [\widehat{\mu}_{0,n}(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})][\widehat{\boldsymbol{h}}_{\text{eff},n}(\boldsymbol{X}, A, \breve{\boldsymbol{\beta}}) - \breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})]. && (3.18)
\end{aligned}$$

- We first relate $\widehat{\boldsymbol{h}}_{\text{eff},n} - \breve{\boldsymbol{h}}_{\text{eff}}$ to $\widehat{\sigma}_n^2 - \breve{\sigma}^2$. Note that

$$\begin{aligned}
\left\|\widehat{\boldsymbol{h}}_{\text{eff},n}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}}) - \breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})\right\|_2 &\leqslant \left\|\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})\right\|_2 \times \|\Omega\|_2^2 \times \left\|\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1} - \breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X})^{-1}\right\|_2 \times \frac{\|\boldsymbol{\omega}_A\|_2}{p_{\mathscr{A}}(A|\boldsymbol{X})} \\
&\leqslant \left(1 - \frac{1}{K}\right) \frac{\|\Omega\|_2^2 \|\Omega\|_{\text{F}}}{\underline{p}_{\mathscr{A}}} \times \dot{F}(\boldsymbol{X}) \times \left\|\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1} - \breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X})^{-1}\right\|_2.
\end{aligned}$$

Here, $\|\cdot\|_{\text{F}}$ is the Frobenius norm. By $\underline{\sigma}^2 \leqslant \widehat{\sigma}_n^2(\boldsymbol{X}, k), \breve{\sigma}^2(\boldsymbol{X}, k) \leqslant \bar{\sigma}^2$ for $1 \leqslant k \leqslant K$, we further have

$$\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X}),\ \breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X}) \geqslant \underline{\sigma}^2 \sum_{k=1}^{K} \boldsymbol{\omega}_k^{\otimes 2} = \underbrace{\underline{\sigma}^2 \left(1 - \frac{1}{K}\right)^{-1} \Omega^\intercal \Omega}_{:=\underline{\mathsf{V}}_{\epsilon}} > 0.$$

Here, "$\mathsf{A} \succcurlyeq \mathsf{B}$" means that $\mathsf{A}-\mathsf{B}$ is positive semi-definite, with strict inequality if $\mathsf{A}-\mathsf{B}$ is positive definite. Then

$$\begin{aligned}\left\|\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1} - \breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X})^{-1}\right\|_2 &\leqslant \|\underline{\mathsf{V}}_{\epsilon}^{-2}\|_2 \times \left\|\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X}) - \breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X})\right\|_2 \\ &= \frac{1}{\lambda_{\min}(\underline{\mathsf{V}}_{\epsilon})^2} \times \left\| \sum_{k=1}^{K} \frac{[\widehat{\sigma}_n^2(\boldsymbol{X},k) - \breve{\sigma}^2(\boldsymbol{X},k)]\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(k|\boldsymbol{X})} \right\|_2 \\ &\leqslant \frac{1}{\underline{p}_{\mathscr{A}} \lambda_{\min}(\underline{\mathsf{V}}_{\epsilon})^2} \Bigg\| \underbrace{\sum_{k=1}^{K} \boldsymbol{\omega}_k^{\otimes 2}}_{=(1-1/K)^{-1}\Omega^{\intercal}\Omega} \Bigg\|_2 \times \|\widehat{\sigma}_n^2 - \breve{\sigma}^2\|_{\infty}. \end{aligned} \tag{3.19}$$

Therefore,

$$\left\|\widehat{\boldsymbol{h}}_{\text{eff},n}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}}) - \breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})\right\|_2 \leqslant \text{constant} \times \dot{F}(\boldsymbol{X}) \times \|\widehat{\sigma}_n^2 - \breve{\sigma}^2\|_{\infty}.$$

- (Bound for Residual) First note that

$$\begin{aligned}\mathbb{E}_n[\breve{e}(\breve{\boldsymbol{\beta}})^2] &= \mathbb{E}[\breve{e}(\breve{\boldsymbol{\beta}})^2] + o_{\mathbb{P}}(1) && \text{(by SLLN)} \\ &\lesssim 5\mathbb{E}\Big\{\breve{\mu}_0(\boldsymbol{X})^2 + \mu_0(\boldsymbol{X})^2 + \gamma(\boldsymbol{X}, A; \boldsymbol{\beta})^2 + \gamma(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})^2 + \epsilon^2\Big\} + o_{\mathbb{P}}(1),\end{aligned}$$

  which is bounded by Assumption 3.2.

- (Convergence of (3.16))

$$\begin{aligned}[\mathbb{E}_n\|(3.16)\|_2]^2 &\leqslant \text{constant} \times \mathbb{E}_n\left\{\left|\breve{e}(\breve{\boldsymbol{\beta}})\right| \dot{F}(\boldsymbol{X})\right\}^2 \|\widehat{\sigma}_n^2 - \breve{\sigma}^2\|_{\infty}^2 \\ &\lesssim \text{constant} \times \Big\{\mathbb{E}[\breve{e}(\breve{\boldsymbol{\beta}})^2]\mathbb{E}[\dot{F}(\boldsymbol{X})^2] + o_{\mathbb{P}}(1)\Big\} \times \|\widehat{\sigma}_n^2 - \breve{\sigma}^2\|_{\infty}^2 \\ &= o_{\mathbb{P}}(n^{-1}).\end{aligned}$$

- (Convergence of (3.17)) First note that

$$\begin{aligned}\left\|\widehat{\boldsymbol{h}}_{\text{eff},n}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})\right\|_2 &\leqslant \left\|\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})\right\|_2 \times \|\Omega\|_2^2 \times \left\|\breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X})^{-1}\right\|_2 \times \frac{\|\boldsymbol{\omega}_A\|_2}{p_{\mathscr{A}}(A|\boldsymbol{X})} \\ &\leqslant \left(1 - \frac{1}{K}\right)^{-1/2} \frac{\|\Omega\|_2^2 \|\Omega\|_{\text{F}}}{\underline{p}_{\mathscr{A}} \lambda_{\min}(\underline{\mathsf{V}}_{\epsilon})} \times \dot{F}(\boldsymbol{X}).\end{aligned} \tag{3.20}$$

Here, $\lambda_{\min}(\cdot)$ is the smallest eigenvalue of a matrix. Then

$$\begin{aligned}[\mathbb{E}_n\|(3.17)\|_2]^2 &\leqslant \text{constant} \times \mathbb{E}_n[\dot{F}(\boldsymbol{X})^2] \times \mathbb{E}_n[\widehat{\mu}_{0,n}(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})]^2 \\ &\lesssim \text{constant} \times \left\{\mathbb{E}[\dot{F}(\boldsymbol{X})^2] + o_{\mathbb{P}}(1)\right\} \times \mathbb{E}_n[\widehat{\mu}_{0,n}(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})]^2 \\ &= o_{\mathbb{P}}(n^{-1}).\end{aligned}$$

- (Convergence of (3.18))

$$\begin{aligned}[\mathbb{E}_n\|(3.18)\|_2]^2 &\leqslant \text{constant} \times \mathbb{E}_n\left\{|\widehat{\mu}_{0,n}(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})| \times \dot{F}(\boldsymbol{X})\right\}^2 \times \|\widehat{\sigma}_n^2 - \breve{\sigma}^2\|_\infty^2 \\ &\leqslant \text{constant} \times \left\{\mathbb{E}[\dot{F}(\boldsymbol{X})^2] + o_{\mathbb{P}}(1)\right\} \times \mathbb{E}_n[\widehat{\mu}_{0,n}(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X})]^2 \times \|\widehat{\sigma}_n^2 - \breve{\sigma}^2\|_\infty^2 \\ &= o_{\mathbb{P}}(n^{-2}).\end{aligned}$$

Therefore,

$$\begin{aligned}&\sup_{\breve{\boldsymbol{\beta}} \in \mathscr{B}} \mathbb{E}_n\|\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2) - \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)\|_2 \\ \leqslant &\mathbb{E}_n\|(3.16)\|_2 + \mathbb{E}_n\|(3.17)\|_2 + \mathbb{E}_n\|(3.18)\|_2 \\ \lesssim &o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(n^{-1}) \\ = &o_{\mathbb{P}}(n^{-1/2}).\end{aligned}$$

□

3.8.5.8 Proof of Theorem 3.9

Proof of Theorem 3.9. We follow Newey (1994, Lemmas 5.1-5.3) to establish the asymptotic linear representation.

Step I: (Asymptotic Linear Representation) Our Lemma 3.8 can imply the asymptotic linear representation of the plug-in estimating function as in Newey (1994, Lemma 5.1):

$$\sqrt{n}\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)] = \sqrt{n}\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)] + o_{\mathbb{P}}(n^{-1/2}).$$

Step II: (Uniform Convergence and Consistency) We aim to establish the convergence of $\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)]$ and $\mathbb{E}_n[(\partial/\partial\boldsymbol{\beta}^\intercal)\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)]$ uniform for $\check{\boldsymbol{\beta}} \in \mathscr{B}$, and the consistency of $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$.

Recall from Lemma 3.8 that:

$$\sup_{\check{\boldsymbol{\beta}}\in\mathscr{B}} \mathbb{E}_n \left\| \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2) - \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2) \right\|_2 = o_{\mathbb{P}}(n^{-1/2}).$$

The same conclusion can be drawn for $(\partial/\partial\boldsymbol{\beta}^\intercal)\boldsymbol{\phi}_{\text{eff}}$ following the same argument.

Lemma 3.13. *Consider Model (3.1) and the estimating function (3.5). Under Assumptions 3.1-3.3 and 3.4.2, we have*

$$\sup_{\check{\boldsymbol{\beta}}\in\mathscr{B}} \mathbb{E}_n \left\| \frac{\partial \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)}{\partial \boldsymbol{\beta}^\intercal} - \frac{\partial \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2)}{\partial \boldsymbol{\beta}^\intercal} \right\|_2 = o_{\mathbb{P}}(n^{-1/2}).$$

Next, we apply Glivenko-Cantelli Theorem to replace $\mathbb{E}_n$ by $\mathbb{E}$. We establish the conditions in Lemma 3.14.

Lemma 3.14. *Consider Model (3.1) and the estimating function (3.5). Under Assumptions 3.1-3.3 and 3.4.2, we have:*

(I) There exists $L : \mathcal{X} \times \mathcal{A} \times \mathbb{R} \to \mathbb{R}_+$ such that $\mathbb{E}L(\boldsymbol{X}, A, \epsilon) < \infty$, and for any $\check{\boldsymbol{\beta}}_1, \check{\boldsymbol{\beta}}_2 \in \mathscr{B}$, we have

$$\left\| \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}_1; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2) - \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}_2; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2) \right\|_2 \leqslant L(\boldsymbol{X}, A, \epsilon) \left\| \check{\boldsymbol{\beta}}_1 - \check{\boldsymbol{\beta}}_2 \right\|_2.$$

(II) There exists $\widetilde{L} : \mathcal{X} \times \mathcal{A} \times \mathbb{R} \to \mathbb{R}_+$ such that $\mathbb{E}\widetilde{L}(\boldsymbol{X}, A, \epsilon) < \infty$, and for any $\check{\boldsymbol{\beta}}_1, \check{\boldsymbol{\beta}}_2 \in \mathscr{B}$, we have

$$\left\| \frac{\partial \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}_1; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2)}{\partial \boldsymbol{\beta}^\intercal} - \frac{\partial \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}_2; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2)}{\partial \boldsymbol{\beta}^\intercal} \right\|_2 \leqslant \widetilde{L}(\boldsymbol{X}, A, \epsilon) \left\| \check{\boldsymbol{\beta}}_1 - \check{\boldsymbol{\beta}}_2 \right\|_2.$$

(III) $\mathbb{E}\sup_{\check{\boldsymbol{\beta}}\in\mathscr{B}} \|\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2)\|_2 < +\infty$, $\mathbb{E}\sup_{\check{\boldsymbol{\beta}}\in\mathscr{B}} \|(\partial/\partial\boldsymbol{\beta}^\intercal)\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2)\|_2 < +\infty$.

By $\mathscr{B}$ is compact and Lemma 3.14, we can conclude that

$$\left\{ \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2) : \check{\boldsymbol{\beta}} \in \mathscr{B} \right\} \quad \text{and} \quad \left\{ (\partial/\partial\boldsymbol{\beta}^\intercal)\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2) : \check{\boldsymbol{\beta}} \in \mathscr{B} \right\}$$

are both $\mathbb{P}$-Glivenko-Cantelli. Then, by Glivenko-Cantelli Theorem, we have

$$\sup_{\widehat{\boldsymbol{\beta}}\in\mathscr{B}}(\mathbb{P}_n-\mathbb{P})\left\|\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\check{\mu}_0,p_{\mathscr{A}},\check{\sigma}^2)\right\|_2,\quad \sup_{\widehat{\boldsymbol{\beta}}\in\mathscr{B}}(\mathbb{P}_n-\mathbb{P})\left\|\frac{\partial\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\check{\mu}_0,p_{\mathscr{A}},\check{\sigma}^2)}{\partial\boldsymbol{\beta}^\intercal}\right\|_2\xrightarrow{\mathbb{P}}0. \tag{3.21}$$

Remark 3.3. From the proof of Lemma 3.14, we have

$$\begin{aligned}
&\mathbb{E}\left[-\frac{\partial\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\check{\mu}_0,p_{\mathscr{A}},\check{\sigma}^2)}{\partial\boldsymbol{\beta}^\intercal}\right]\\
=&\mathbb{E}(3.27)+\mathbb{E}(3.28)\\
=&\mathbb{E}\Big\{\dot{\mathsf{F}}(\boldsymbol{X};\check{\boldsymbol{\beta}})^\intercal\Omega^\intercal\Omega\check{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\underbrace{\mathbb{E}\left[\left(1-\frac{1}{K}\right)\frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})}\middle|\boldsymbol{X}\right]}_{=\Omega^\intercal\Omega}\dot{\mathsf{F}}(\boldsymbol{X};\check{\boldsymbol{\beta}})\Big\}\\
&-\mathbb{E}\Big\{\ddot{\mathsf{F}}(\boldsymbol{X};\check{\boldsymbol{\beta}})^\intercal\Omega^\intercal\Omega\check{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\underbrace{\mathbb{E}\left(\frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}\middle|\boldsymbol{X}\right)}_{=\mathbf{0}}\check{e}(\check{\boldsymbol{\beta}})\Big\}\\
=&\check{\mathcal{I}}(\check{\boldsymbol{\beta}}).
\end{aligned}$$

By Lemma 3.14 (II) and (3.21), we further have that $\check{\boldsymbol{\beta}}\to\check{\mathcal{I}}(\check{\boldsymbol{\beta}})$ is continuous.

Combining Lemmas 3.8, 3.13 and (3.21), we have

$$\sup_{\widehat{\boldsymbol{\beta}}\in\mathscr{B}}\left\|\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\widehat{\mu}_{0,n},p_{\mathscr{A}},\widehat{\sigma}_n^2)]-\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\check{\mu}_0,p_{\mathscr{A}},\check{\sigma}^2)]\right\|_2\xrightarrow{\mathbb{P}}0; \tag{3.22}$$

$$\sup_{\check{\boldsymbol{\beta}}\in\mathscr{B}}\left\|\mathbb{E}_n\left[\frac{\partial\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\widehat{\mu}_{0,n},p_{\mathscr{A}},\widehat{\sigma}_n^2)}{\partial\boldsymbol{\beta}^\intercal}\right]-\mathbb{E}\left[\frac{\partial\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\check{\mu}_0,p_{\mathscr{A}},\check{\sigma}^2)}{\partial\boldsymbol{\beta}^\intercal}\right]\right\|_2\xrightarrow{\mathbb{P}}0. \tag{3.23}$$

The consistency of $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ follows from that $\mathscr{B}$ is compact,

$$\widehat{\boldsymbol{\beta}}_{\text{eff},n}\in\operatorname*{argmax}_{\check{\boldsymbol{\beta}}\in\mathscr{B}}\left\|\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\widehat{\mu}_{0,n},p_{\mathscr{A}},\widehat{\sigma}_n^2)]\right\|_2^2;\quad \boldsymbol{\beta}\in\operatorname*{argmax}_{\check{\boldsymbol{\beta}}\in\mathscr{B}}\left\|\mathbb{E}[\boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}};\check{\mu}_0,p_{\mathscr{A}},\check{\sigma}^2)]\right\|_2^2,$$

and the uniform convergence in probability in (3.22).

Step III: By Mean Value Theorem, there exists some $\alpha_n \in [0,1]$ and $\widetilde{\boldsymbol{\beta}}_n = (1-\alpha_n)\widehat{\boldsymbol{\beta}}_{\text{eff},n} + \alpha_n \boldsymbol{\beta}$, such that

$$\begin{aligned}\mathbf{0} &= \mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\widehat{\boldsymbol{\beta}}_{\text{eff},n}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)] \\ &= \mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)] + \mathbb{E}_n\left[\frac{\partial \boldsymbol{\phi}_{\text{eff}}(\widetilde{\boldsymbol{\beta}}_n; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)}{\partial \boldsymbol{\beta}^\intercal}\right](\widehat{\boldsymbol{\beta}}_{\text{eff},n} - \boldsymbol{\beta}).\end{aligned}$$

That is,

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{eff},n} - \boldsymbol{\beta} &= \left\{\mathbb{E}_n\left[-\frac{\partial \boldsymbol{\phi}_{\text{eff}}(\widetilde{\boldsymbol{\beta}}_n; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)}{\partial \boldsymbol{\beta}^\intercal}\right]\right\}^{-1} \mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)] \\ &= \left\{\breve{\mathcal{I}}(\boldsymbol{\beta}) + o_{\mathbb{P}}(1)\right\}^{-1} \times \left\{\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)] + o_{\mathbb{P}}(n^{-1/2})\right\} \\ &= \breve{\mathcal{I}}(\boldsymbol{\beta})^{-1}\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)] + o_{\mathbb{P}}(n^{-1/2}).\end{aligned}$$

Here, the second equality follows from that

$$\begin{aligned}\mathbb{E}_n\left[-\frac{\partial \boldsymbol{\phi}_{\text{eff}}(\widetilde{\boldsymbol{\beta}}_n; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)}{\partial \boldsymbol{\beta}^\intercal}\right] &= \breve{\mathcal{I}}(\widetilde{\boldsymbol{\beta}}_n) + o_{\mathbb{P}}(1) \qquad \text{(by (3.23))} \\ &= \breve{\mathcal{I}}(\boldsymbol{\beta}) + o_{\mathbb{P}}(1). \qquad \text{(by } \widehat{\boldsymbol{\beta}}_{\text{eff},n} \xrightarrow{\mathbb{P}} \boldsymbol{\beta} \text{ and the continuity of } \breve{\mathcal{I}})\end{aligned}$$

Step IV: (Semiparametric Efficiency) If $(\breve{\mu}_0, \breve{\sigma}^2) = (\mu_0, \sigma^2)$, then $\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2) = \boldsymbol{S}_{\text{eff}}(\boldsymbol{\beta})$. Moreover,

$$\begin{aligned}\mathbb{E}[\boldsymbol{S}_{\text{eff}}(\boldsymbol{\beta})^{\otimes 2}] &= \mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \mathsf{V}_\epsilon(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A^{\otimes 2}\epsilon^2}{p_{\mathscr{A}}(A|\boldsymbol{X})^2} \mathsf{V}_\epsilon(\boldsymbol{X})^{-1}\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right] \\ &= \mathbb{E}\Bigg[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \mathsf{V}_\epsilon(\boldsymbol{X})^{-1} \underbrace{\mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\mathbb{E}(\epsilon^2|\boldsymbol{X}, A)}{p_{\mathscr{A}}(A|\boldsymbol{X})^2}\middle|\boldsymbol{X}\right)}_{=\mathsf{V}_\epsilon(\boldsymbol{X})} \mathsf{V}_\epsilon(\boldsymbol{X})^{-1}\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\Bigg] \\ &= \mathcal{I}(\boldsymbol{\beta}).\end{aligned}$$

That is, $\mathcal{I}(\boldsymbol{\beta})$ defined in Theorem 3.9 is the semiparametric Fisher information matrix. We further have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\text{eff},n} - \boldsymbol{\beta}) = \sqrt{n}\mathcal{I}(\boldsymbol{\beta})^{-1}\mathbb{E}_n[\boldsymbol{S}_{\text{eff}}(\boldsymbol{\beta})] + o_{\mathbb{P}}(1) \xrightarrow{\mathcal{D}} \mathcal{N}_p\big(\mathbf{0}, \mathcal{I}(\boldsymbol{\beta})^{-1}\big).$$

Therefore, $\widehat{\boldsymbol{\beta}}_{\text{eff},n}$ is semiparametric efficient.

□

Proof of Lemma 3.14. We follow the notations in Section 3.8.5.7. Denote $\ddot{F}(\boldsymbol{X}) := \sup_{\breve{\boldsymbol{\beta}}\in\mathscr{B}} \|\ddot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}})\|_2$.

(I) Fix $\breve{\boldsymbol{\beta}}_1, \breve{\boldsymbol{\beta}}_2 \in \mathscr{B}$. Then

$$\begin{aligned}
&\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}_1;\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2) - \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}_2;\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2)\\
=&-\left(1-\frac{1}{K}\right)\boldsymbol{\omega}_A^{\intercal}[\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)]\boldsymbol{h}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_1) \qquad (3.24)\\
&+\breve{e}(\breve{\boldsymbol{\beta}}_2)[\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_1)-\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_2)]. \qquad (3.25)
\end{aligned}$$

- (Lipschitz Bound on (3.24))
  - (Lipschitz Bound on $\vec{\boldsymbol{f}}$)

$$\begin{aligned}
\left\|\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)\right\|_2 &= \left\|\dot{\mathsf{F}}(\boldsymbol{X};\widetilde{\boldsymbol{\beta}})[\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2]\right\|_2 \qquad \left(\text{for some } \widetilde{\boldsymbol{\beta}}=(1-\alpha)\breve{\boldsymbol{\beta}}_1+\alpha\breve{\boldsymbol{\beta}}_2\right)\\
&\leqslant \left\|\dot{\mathsf{F}}(\boldsymbol{X};\widetilde{\boldsymbol{\beta}})\right\|_2 \times \left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2\\
&\leqslant \dot{F}(\boldsymbol{X}) \times \left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2.
\end{aligned}$$

  - Then we have

$$\begin{aligned}
\|(3.24)\|_2 &= \underbrace{\left(1-\frac{1}{K}\right)\|\boldsymbol{\omega}_A\|_2}_{\leqslant(1-1/K)^{1/2}\|\Omega\|_{\text{F}}} \left\|\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)\right\|_2 \underbrace{\left\|\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_1)\right\|_2}_{\leqslant(3.20)}\\
&\leqslant \text{constant} \times \dot{F}(\boldsymbol{X})^2 \times \left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2.
\end{aligned}$$

- (Lipschitz Bound on (3.25))
  - Note that

$$\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_1)-\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_2) = [\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)]^{\intercal}\Omega^{\intercal}\Omega\breve{\mathsf{V}}_{\epsilon}(\boldsymbol{X})^{-1}\frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}.$$

– (Lipschitz Bound on $\dot{\mathsf{F}}$)

$$\left\|\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)\right\|_2 \leqslant \ddot{F}(\boldsymbol{X})\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2.$$

– Then we have

$$\begin{aligned}
&\left\|\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_1)-\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X},A;\breve{\boldsymbol{\beta}}_2)\right\|_2\\
=&\left\|\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)\right\|_2 \underbrace{\|\Omega\|_2^2\left\|\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\right\|_2\frac{\|\boldsymbol{\omega}_A\|_2}{p_{\mathscr{A}}(A|\boldsymbol{X})}}_{\leqslant(1-1/K)^{-1/2}\|\Omega\|_2^2\|\Omega\|_{\text{F}}/[\underline{p}_{\mathscr{A}}\lambda_{\min}(\underline{\mathsf{V}}_\epsilon)]}\\
\leqslant&\text{constant}\times\ddot{F}(\boldsymbol{X})\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2.
\end{aligned}$$

– The residual $\breve{e}(\breve{\boldsymbol{\beta}}_2)$ in (3.25) can be bounded by

$$\begin{aligned}
\sup_{\breve{\boldsymbol{\beta}}\in\mathscr{B}}\left|\breve{e}(\breve{\boldsymbol{\beta}})\right| &\leqslant |\breve{\mu}_0(\boldsymbol{X})-\mu_0(\boldsymbol{X})|+\left(1-\frac{1}{K}\right)\sup_{\breve{\boldsymbol{\beta}}\in\mathscr{B}}\left|\boldsymbol{\omega}_A^{\intercal}[\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}})-\vec{\boldsymbol{f}}(\boldsymbol{X};\boldsymbol{\beta})]\right|+|\epsilon|\\
&\leqslant |\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\left(1-\frac{1}{K}\right)^{1/2}\|\Omega\|_{\text{F}}\sup_{\breve{\boldsymbol{\beta}}\in\mathscr{B}}\left\|\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}})-\vec{\boldsymbol{f}}(\boldsymbol{X};\boldsymbol{\beta})\right\|_2+|\epsilon|\\
&\leqslant |\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\text{constant}\times\dot{F}(\boldsymbol{X})\times\underbrace{\sup_{\breve{\boldsymbol{\beta}}\in\mathscr{B}}\left\|\breve{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_2}_{\leqslant\mathsf{diam}(\mathscr{B})}+|\epsilon|\\
&\leqslant |\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\text{constant}\times\dot{F}(\boldsymbol{X})+|\epsilon|. \qquad (3.26)
\end{aligned}$$

– Then we have

$$\|(3.25)\|_2\leqslant\text{constant}\times\left[|\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\dot{F}(\boldsymbol{X})+|\epsilon|\right]\times\ddot{F}(\boldsymbol{X})\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2$$

Combining the Lipschitz bounds on (3.24) and (3.25), we have

$$\begin{aligned}
&\left\|\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}_1;\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2)-\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}_2;\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2)\right\|_2\\
\leqslant&\underbrace{\text{constant}\times\left\{\dot{F}(\boldsymbol{X})^2+\left[|\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\dot{F}(\boldsymbol{X})+|\epsilon|\right]\times\ddot{F}(\boldsymbol{X})\right\}}_{:=L(\boldsymbol{X},A,\epsilon)}\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2.
\end{aligned}$$

In particular,

$$\begin{aligned}
&\mathbb{E}L(\boldsymbol{X}, A, \epsilon) \\
\leqslant&\text{constant} \times \Big[\mathbb{E}[\dot{F}(\boldsymbol{X})^2] + \Big(\left\{\mathbb{E}[\breve{\mu}_0(\boldsymbol{X})^2]\right\}^{1/2} + \left\{\mathbb{E}[\mu_0(\boldsymbol{X})^2]\right\}^{1/2} \\
&+ \left\{\mathbb{E}[\dot{F}(\boldsymbol{X})^2]\right\}^{1/2} + \left\{\mathbb{E}\epsilon^2\right\}^{1/2} \Big) \left\{\mathbb{E}[\ddot{F}(\boldsymbol{X})^2]\right\}^{1/2} \Big],
\end{aligned}$$

which is finite by Assumptions 3.2 and 3.4.2.

(II) Note that

$$-\frac{\partial \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)}{\partial \boldsymbol{\beta}^\intercal} = \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \left[\left(1 - \frac{1}{K}\right) \frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})} \right] \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}) \tag{3.27}$$

$$- \ddot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A \breve{e}(\breve{\boldsymbol{\beta}})}{p_{\mathscr{A}}(A|\boldsymbol{X})}. \tag{3.28}$$

Fix $\breve{\boldsymbol{\beta}}_1, \breve{\boldsymbol{\beta}}_2 \in \mathscr{B}$. Denote (3.27)($\breve{\boldsymbol{\beta}}_1$) and (3.27)($\breve{\boldsymbol{\beta}}_2$) as (3.27) with $\breve{\boldsymbol{\beta}}$ replaced by $\breve{\boldsymbol{\beta}}_1$ and $\breve{\boldsymbol{\beta}}_2$ respectively.

$$\begin{aligned}
(3.27)(\breve{\boldsymbol{\beta}}_1) - (3.27)(\breve{\boldsymbol{\beta}}_2) =& \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_1)^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \left[\left(1 - \frac{1}{K}\right) \frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})} \right] [\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_1) - \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_2)] + \\
& [\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_1) - \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_2)]^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \left[\left(1 - \frac{1}{K}\right) \frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})} \right] \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_2).
\end{aligned}$$

Then

$$\begin{aligned}
\left\| (3.27)(\breve{\boldsymbol{\beta}}_1) - (3.27)(\breve{\boldsymbol{\beta}}_2) \right\|_2 &\leqslant \frac{2\|\Omega\|_2^2 \|\Omega\|_{\text{F}}^2}{\underline{p}_{\mathscr{A}} \lambda_{\min}(\underline{\mathsf{V}}_\epsilon)} \times \dot{F}(\boldsymbol{X}) \times \left\| \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_1) - \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_2) \right\|_2 \\
&\leqslant \text{constant} \times \dot{F}(\boldsymbol{X}) \ddot{F}(\boldsymbol{X}) \times \left\| \breve{\boldsymbol{\beta}}_1 - \breve{\boldsymbol{\beta}}_2 \right\|_2.
\end{aligned}$$

Denote (3.28)($\breve{\boldsymbol{\beta}}_1$) and (3.28)($\breve{\boldsymbol{\beta}}_2$) as (3.28) with $\breve{\boldsymbol{\beta}}$ replaced by $\breve{\boldsymbol{\beta}}_1$ and $\breve{\boldsymbol{\beta}}_2$ respectively.

$$(3.28)(\breve{\boldsymbol{\beta}}_1) - (3.28)(\breve{\boldsymbol{\beta}}_2) = - [\dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_1) - \dot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_2)]^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A \breve{e}(\breve{\boldsymbol{\beta}}_1)}{p_{\mathscr{A}}(A|\boldsymbol{X})} \tag{3.29}$$

$$- \ddot{\mathsf{F}}(\boldsymbol{X}; \breve{\boldsymbol{\beta}}_2)^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1} \frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})} [\breve{e}(\breve{\boldsymbol{\beta}}_1) - \breve{e}(\breve{\boldsymbol{\beta}}_2)]. \tag{3.30}$$

Here,

$$\begin{aligned}\left|\breve{e}(\breve{\boldsymbol{\beta}}_1)-\breve{e}(\breve{\boldsymbol{\beta}}_2)\right| &= \left(1-\frac{1}{K}\right)\left|\boldsymbol{\omega}_A^{\intercal}[\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\vec{\boldsymbol{f}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)]\right| \\ &\leqslant \left(1-\frac{1}{K}\right)^{1/2}\|\Omega\|_{\mathrm{F}}\times\dot{F}(\boldsymbol{X})\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2.\end{aligned}$$

Then

$$\begin{aligned}\|(3.29)\|_2 \leqslant& \left(1-\frac{1}{K}\right)^{-1/2}\frac{\|\Omega\|_2^2\|\Omega\|_{\mathrm{F}}}{p_{\mathscr{A}}\lambda_{\min}(\underline{\mathsf{V}}_\epsilon)}\times\left[|\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\text{constant}\times\dot{F}(\boldsymbol{X})+|\epsilon|\right] \\ &\times\left\|\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_1)-\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}_2)\right\|_2 \\ \leqslant&\text{constant}\times\left[|\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\dot{F}(\boldsymbol{X})+|\epsilon|\right]\times\ddot{F}(\boldsymbol{X})\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2; \\ \|(3.30)\|_2 \leqslant& \left(1-\frac{1}{K}\right)^{-1/2}\frac{\|\Omega\|_2^2\|\Omega\|_{\mathrm{F}}}{p_{\mathscr{A}}\lambda_{\min}(\underline{\mathsf{V}}_\epsilon)}\times\ddot{F}(\boldsymbol{X})\times\left|\breve{e}(\breve{\boldsymbol{\beta}}_1)-\breve{e}(\breve{\boldsymbol{\beta}}_2)\right| \\ \leqslant&\text{constant}\times\dot{F}(\boldsymbol{X})\ddot{F}(\boldsymbol{X})\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2.\end{aligned}$$

And the Lipschitz bound for (3.28) is

$$\begin{aligned}&\left\|(3.28)(\breve{\boldsymbol{\beta}}_1)-(3.28)(\breve{\boldsymbol{\beta}}_2)\right\|_2 \\ \leqslant&\|(3.29)\|_2+\|(3.30)\|_2 \\ \leqslant&\text{constant}\times\left[|\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\dot{F}(\boldsymbol{X})+|\epsilon|\right]\times\ddot{F}(\boldsymbol{X})\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2\end{aligned}$$

Combining the Lipschitz bounds for (3.27) and (3.28), we have

$$\begin{aligned}&\left\|\frac{\partial\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}_1;\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2)}{\partial\boldsymbol{\beta}^{\intercal}}-\frac{\partial\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}_2;\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2)}{\partial\boldsymbol{\beta}^{\intercal}}\right\|_2 \\ \leqslant&\underbrace{\text{constant}\times\left[|\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\dot{F}(\boldsymbol{X})+|\epsilon|\right]\times\ddot{F}(\boldsymbol{X})}_{:=\widetilde{L}(\boldsymbol{X},A,\epsilon)}\times\left\|\breve{\boldsymbol{\beta}}_1-\breve{\boldsymbol{\beta}}_2\right\|_2\end{aligned}$$

In particular,

$$\mathbb{E}\widetilde{L}(\boldsymbol{X}, A, \epsilon) \leqslant \text{constant} \times \Big(\big\{\mathbb{E}[\breve{\mu}_0(\boldsymbol{X})^2]\big\}^{1/2} + \big\{\mathbb{E}[\mu_0(\boldsymbol{X})^2]\big\}^{1/2} + \Big\{\mathbb{E}[\dot{F}(\boldsymbol{X})^2]\Big\}^{1/2} + \big\{\mathbb{E}\epsilon^2\big\}^{1/2} \Big) \Big\{\mathbb{E}[\ddot{F}(\boldsymbol{X})^2]\Big\}^{1/2},$$

which is finite by Assumptions 3.2 and 3.4.2.

(III) Note that

$$\begin{aligned}\Big\|\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)\Big\|_2 &= \underbrace{\Big|\breve{e}(\breve{\boldsymbol{\beta}})\Big|}_{\leqslant (3.26)} \times \underbrace{\Big\|\breve{\boldsymbol{h}}_{\text{eff}}(\boldsymbol{X}, A; \breve{\boldsymbol{\beta}})\Big\|_2}_{\leqslant (3.20)} \\ &\leqslant \text{constant} \times \Big[|\breve{\mu}_0(\boldsymbol{X})| + |\mu_0(\boldsymbol{X})| + \dot{F}(\boldsymbol{X}) + |\epsilon|\Big] \times \dot{F}(\boldsymbol{X}).\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbb{E} \sup_{\breve{\boldsymbol{\beta}} \in \mathscr{B}} \Big\|\boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)\Big\|_2 \\ \leqslant &\text{constant} \times \Big(\big\{\mathbb{E}[\breve{\mu}_0(\boldsymbol{X})^2]\big\}^{1/2} + \big\{\mathbb{E}[\mu_0(\boldsymbol{X})^2]\big\}^{1/2} + \Big\{\mathbb{E}[\dot{F}(\boldsymbol{X})^2]\Big\}^{1/2} + \big\{\mathbb{E}\epsilon^2\big\}^{1/2} \Big) \Big\{\mathbb{E}[\dot{F}(\boldsymbol{X})^2]\Big\}^{1/2},\end{aligned}$$

which is finite by Assumptions 3.2 and 3.4.2. Next, we consider bounds for (3.27) and (3.28).

$$\begin{aligned}\|(3.27)\|_2 &\leqslant \frac{\|\Omega\|_2^2 \|\Omega\|_{\text{F}}^2}{p_{\mathscr{A}} \lambda_{\min}(\underline{\mathsf{V}}_\epsilon)} \times \dot{F}(\boldsymbol{X})^2; \\ \|(3.28)\|_2 &\leqslant \left(1 - \frac{1}{K}\right)^{-1/2} \frac{\|\Omega\|_2^2 \|\Omega\|_{\text{F}}}{p_{\mathscr{A}} \lambda_{\min}(\underline{\mathsf{V}}_\epsilon)} \times \ddot{F}(\boldsymbol{X}) \times \underbrace{\Big|\breve{e}(\breve{\boldsymbol{\beta}})\Big|}_{\leqslant (3.26)} \\ &\leqslant \text{constant} \times \Big[|\breve{\mu}_0(\boldsymbol{X})| + |\mu_0(\boldsymbol{X})| + \dot{F}(\boldsymbol{X}) + |\epsilon|\Big] \times \ddot{F}(\boldsymbol{X}).\end{aligned}$$

Then we have

$$\begin{aligned}&\mathbb{E}\left\|\frac{\partial \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}}; \breve{\mu}_0, p_{\mathscr{A}}, \breve{\sigma}^2)}{\partial \boldsymbol{\beta}^\intercal}\right\|_2 \leqslant \mathbb{E}\|(3.27)\|_2 + \mathbb{E}\|(3.28)\|_2 \\ \leqslant &\text{constant} \times \Big[\mathbb{E}[\dot{F}(\boldsymbol{X})^2] + \Big(\big\{\mathbb{E}[\breve{\mu}_0(\boldsymbol{X})^2]\big\}^{1/2} + \big\{\mathbb{E}[\mu_0(\boldsymbol{X})^2]\big\}^{1/2} + \Big\{\mathbb{E}[\dot{F}(\boldsymbol{X})^2]\Big\}^{1/2} \\ &+ \big\{\mathbb{E}\epsilon^2\big\}^{1/2} \Big) \Big\{\mathbb{E}[\ddot{F}(\boldsymbol{X})^2]\Big\}^{1/2} \Big],\end{aligned}$$

which is finite by Assumptions by 3.2 and 3.4.2.

□

Proof of Lemma 3.13. We follow the notations in Section 3.8.5.7 and the proof of Lemma 3.14. Note that

$$
\begin{aligned}
& -\frac{\partial \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}};\widehat{\mu}_{0,n},p_{\mathscr{A}},\widehat{\sigma}_n^2)}{\partial \boldsymbol{\beta}^\intercal}+\frac{\partial \boldsymbol{\phi}_{\text{eff}}(\breve{\boldsymbol{\beta}};\breve{\mu}_0,p_{\mathscr{A}},\breve{\sigma}^2)}{\partial \boldsymbol{\beta}^\intercal} \\
= \quad & \dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega[\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1}-\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}]\left[\left(1-\frac{1}{K}\right)\frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})}\right]\dot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}}) && (3.31)\\
& +\ddot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega \breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}[\widehat{\mu}_{0,n}(\boldsymbol{X})-\breve{\mu}_0(\boldsymbol{X})] && (3.32)\\
& -\ddot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega[\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1}-\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}]\frac{\boldsymbol{\omega}_A \breve{e}(\breve{\boldsymbol{\beta}})}{p_{\mathscr{A}}(A|\boldsymbol{X})} && (3.33)\\
& +\ddot{\mathsf{F}}(\boldsymbol{X};\breve{\boldsymbol{\beta}})^\intercal \Omega^\intercal \Omega[\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1}-\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}]\frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}[\widehat{\mu}_{0,n}(\boldsymbol{X})-\breve{\mu}_0(\boldsymbol{X})]. && (3.34)
\end{aligned}
$$

Then

$$
\begin{aligned}
\|(3.31)\|_2 &\leqslant \frac{\|\Omega\|_2^2\|\Omega\|_{\text{F}}^2}{p_{\mathscr{A}}}\times \dot{F}(\boldsymbol{X})^2 \times \underbrace{\left\|\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1}-\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\right\|_2}_{\leqslant (3.19)} \\
&\leqslant \text{constant}\times \dot{F}(\boldsymbol{X})^2 \times \|\widehat{\sigma}_n^2-\breve{\sigma}^2\|_\infty; \\
\|(3.32)\|_2 &\leqslant \left(1-\frac{1}{K}\right)^{-1/2}\frac{\|\Omega\|_2^2\|\Omega\|_{\text{F}}}{p_{\mathscr{A}}\lambda_{\min}(\underline{\mathsf{V}}_\epsilon)}\times \ddot{F}(\boldsymbol{X})\times |\widehat{\mu}_{0,n}(\boldsymbol{X})-\breve{\mu}_0(\boldsymbol{X})|\,; \\
\|(3.33)\|_2 &\leqslant \left(1-\frac{1}{K}\right)^{-1/2}\frac{\|\Omega\|_2^2\|\Omega\|_{\text{F}}}{p_{\mathscr{A}}}\times \ddot{F}(\boldsymbol{X})\times \underbrace{\left|\breve{e}(\breve{\boldsymbol{\beta}})\right|}_{\leqslant (3.26)}\times \underbrace{\left\|\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1}-\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\right\|_2}_{\leqslant (3.19)} \\
&\leqslant \text{constant}\times \left[|\breve{\mu}_0(\boldsymbol{X})|+|\mu_0(\boldsymbol{X})|+\dot{F}(\boldsymbol{X})+|\epsilon|\right]\times \ddot{F}(\boldsymbol{X})\times \|\widehat{\sigma}_n^2-\breve{\sigma}^2\|_\infty; \\
\|(3.34)\|_2 &\leqslant \left(1-\frac{1}{K}\right)^{-1/2}\frac{\|\Omega\|_2^2\|\Omega\|_{\text{F}}}{p_{\mathscr{A}}}\times \ddot{F}(\boldsymbol{X})\times \underbrace{\left\|\widehat{\mathsf{V}}_{\epsilon,n}(\boldsymbol{X})^{-1}-\breve{\mathsf{V}}_\epsilon(\boldsymbol{X})^{-1}\right\|_2}_{\leqslant (3.19)}\times |\widehat{\mu}_{0,n}(\boldsymbol{X})-\breve{\mu}_0(\boldsymbol{X})| \\
&\leqslant \text{constant}\times \ddot{F}(\boldsymbol{X})\times \|\widehat{\sigma}_n^2-\breve{\sigma}^2\|_\infty \times |\widehat{\mu}_{0,n}(\boldsymbol{X})-\breve{\mu}_0(\boldsymbol{X})|\,.
\end{aligned}
$$

Then we have

$$\begin{aligned}
&\mathbb{E}_n \left\| \frac{\partial \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \widehat{\mu}_{0,n}, p_{\mathscr{A}}, \widehat{\sigma}_n^2)}{\partial \boldsymbol{\beta}^\intercal} - \frac{\partial \boldsymbol{\phi}_{\text{eff}}(\check{\boldsymbol{\beta}}; \check{\mu}_0, p_{\mathscr{A}}, \check{\sigma}^2)}{\partial \boldsymbol{\beta}^\intercal} \right\|_2 \\
\leqslant & \mathbb{E}_n \|(3.31)\|_2 + \mathbb{E}_n \|(3.32)\|_2 + \mathbb{E}_n \|(3.33)\|_2 + \mathbb{E}_n \|(3.34)\|_2 \\
\lesssim & o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(n^{-1/2}) o_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(n^{-1}) \\
= & o_{\mathbb{P}}(n^{-1/2}).
\end{aligned}$$

□

3.8.5.9 Proof of Theorem 3.10

Proof of Theorem 3.10. By Theorem 3.9, we have

$$\widehat{\boldsymbol{\beta}}_{\text{eff},n}(\check{\mu}_0) - \boldsymbol{\beta} = \mathcal{I}(\boldsymbol{\beta}; \check{\mu}_0)^{-1} \mathbb{E}_n[\boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \check{\mu}_0, p_{\mathscr{A}}, \sigma^2_{\text{opt}})] + o_{\mathbb{P}}(n^{-1/2}).$$

Therefore, it suffices to study the asymptotic variance First of all, we derive the $\sqrt{n}$-asymptotic variance of $\widehat{\boldsymbol{\beta}}_{\text{eff},n}(\check{\mu}_0)$. Denote

$$\begin{aligned}
\check{\boldsymbol{\phi}}_{\text{eff}}(\boldsymbol{\beta}) &:= \boldsymbol{\phi}_{\text{eff}}(\boldsymbol{\beta}; \check{\mu}_0, p_{\mathscr{A}}, \sigma^2_{\text{opt}}) \\
&= \left[Y - \check{\mu}_0(\boldsymbol{X}) - \left(1 - \frac{1}{K}\right) \langle \boldsymbol{\omega}_A, \vec{\boldsymbol{f}}(\boldsymbol{X}; \boldsymbol{\beta}) \rangle \right] \boldsymbol{h}_{\text{eff}}(\boldsymbol{X}, A; \boldsymbol{\beta}, \check{\mu}_0) \\
&= [\underbrace{\mu_0(\boldsymbol{X}) - \check{\mu}_0(\boldsymbol{X}) + \epsilon}_{=\check{e}(\boldsymbol{\beta})}] \boldsymbol{h}_{\text{eff}}(\boldsymbol{X}, A; \boldsymbol{\beta}, \check{\mu}_0); \\
\boldsymbol{h}_{\text{eff}}(\boldsymbol{X}, A; \boldsymbol{\beta}, \check{\mu}_0) &:= \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \mathsf{V}_\epsilon(\boldsymbol{X}; \check{\mu}_0)^{-1} \frac{\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}.
\end{aligned}$$

First notice that

$$\begin{aligned}
\mathbb{E}[\breve{\boldsymbol{\phi}}_{\text{eff}}(\boldsymbol{\beta})^{\otimes 2}] &= \mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^{\intercal}\Omega^{\intercal}\Omega\mathsf{V}_{\epsilon}(\boldsymbol{X};\breve{\mu}_0)^{-1}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\breve{e}(\boldsymbol{\beta})^2}{p_{\mathscr{A}}(A|\boldsymbol{X})^2}\right)\mathsf{V}_{\epsilon}(\boldsymbol{X};\breve{\mu}_0)^{-1}\Omega^{\intercal}\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right] \\
&= \underbrace{\mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^{\intercal}\Omega^{\intercal}\Omega\mathsf{V}_{\epsilon}(\boldsymbol{X};\breve{\mu}_0)^{-1}\Omega^{\intercal}\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right]}_{=\mathcal{I}(\boldsymbol{\beta};\breve{\mu}_0)}; \\
\mathbb{E}\left[-\frac{\partial\breve{\boldsymbol{\phi}}_{\text{eff}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\intercal}}\right] &= \mathbb{E}\left\{\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^{\intercal}\Omega^{\intercal}\Omega\mathsf{V}_{\epsilon}(\boldsymbol{X};\breve{\mu}_0)^{-1}\left[\left(1-\frac{1}{K}\right)\frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})}\right]\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right\} \\
&\quad \underbrace{-\mathbb{E}\left\{\ddot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^{\intercal}\Omega^{\intercal}\Omega\mathsf{V}_{\epsilon}(\boldsymbol{X};\breve{\mu}_0)^{-1}\frac{[\mu_0(\boldsymbol{X})-\breve{\mu}_0(\boldsymbol{X})+\epsilon]\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}\right\}}_{=0} \\
&= \underbrace{\mathbb{E}\left[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^{\intercal}\Omega^{\intercal}\Omega\mathsf{V}_{\epsilon}(\boldsymbol{X};\breve{\mu}_0)^{-1}\Omega^{\intercal}\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\right]}_{=\mathcal{I}(\boldsymbol{\beta};\breve{\mu}_0)}.
\end{aligned}$$

Here, the second equality follows from that

$$\begin{aligned}
&\mathbb{E}[\breve{e}(\boldsymbol{\beta})^2|\boldsymbol{X},A] = [\breve{\mu}_0(\boldsymbol{X})-\mu_0(\boldsymbol{X})]^2+\sigma^2(\boldsymbol{X},A) = \sigma^2_{\text{opt}}(\boldsymbol{X},A;\breve{\mu}_0); \\
&\mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\breve{e}(\boldsymbol{\beta})^2}{p_{\mathscr{A}}(A|\boldsymbol{X})^2}\middle|\boldsymbol{X}\right) = \mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\sigma^2_{\text{opt}}(\boldsymbol{X},A;\breve{\mu}_0)}{p_{\mathscr{A}}(A|\boldsymbol{X})^2}\middle|\boldsymbol{X}\right) = \sum_{k=1}^{K}\frac{\sigma^2_{\text{opt}}(\boldsymbol{X},k;\breve{\mu}_0)\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(k|\boldsymbol{X})} = \mathsf{V}_{\epsilon}(\boldsymbol{X};\breve{\mu}_0).
\end{aligned}$$

The forth equality follows from that $\mathbb{E}[\boldsymbol{\omega}_A/p_{\mathscr{A}}(A|\boldsymbol{X})|\boldsymbol{X}] = \mathbf{0}$, $\mathbb{E}(\epsilon|\boldsymbol{X},A) = 0$ and

$$\left(1-\frac{1}{K}\right)\mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})}\middle|\boldsymbol{X}\right) = \left(1-\frac{1}{K}\right)\sum_{k=1}^{K}\boldsymbol{\omega}_k^{\otimes 2} = \Omega^{\intercal}\Omega.$$

Then

$$\lim_{n\to\infty} n\mathsf{Var}[\widehat{\boldsymbol{\beta}}_{\text{eff},n}(\breve{\mu}_0)] = \left\{\mathbb{E}\left[-\frac{\partial\breve{\boldsymbol{\phi}}_{\text{eff}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\intercal}}\right]\right\}^{-1}\mathbb{E}[\breve{\boldsymbol{\phi}}_{\text{eff}}(\boldsymbol{\beta})^{\otimes 2}]\left\{\mathbb{E}\left[-\frac{\partial\breve{\boldsymbol{\phi}}_{\text{eff}}(\boldsymbol{\beta})^{\intercal}}{\partial\boldsymbol{\beta}}\right]\right\}^{-1} = \mathcal{I}(\boldsymbol{\beta};\breve{\mu}_0)^{-1}.$$

Next, we study the $\sqrt{n}$-asymptotic variance of estimates from the regular class $\mathcal{B}_n(\breve{\mu}_0)$. Fix $\widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0) \in \mathcal{B}_n(\breve{\mu}_0)$ that corresponds to the estimating function $\boldsymbol{\phi}(\boldsymbol{\beta};\breve{\mu}_0)$, and denote $\breve{\boldsymbol{\phi}}(\boldsymbol{\beta}) := \boldsymbol{\phi}(\boldsymbol{\beta};\breve{\mu}_0)$. There exists $\boldsymbol{h}: \mathcal{X}\times A \to \mathbb{R}^p$, which can depend on $(\boldsymbol{\beta},\breve{\mu}_0)$, such that $\mathbb{E}[\boldsymbol{h}(\boldsymbol{X},A)|\boldsymbol{X}] = \mathbf{0}$ and

$$\breve{\boldsymbol{\phi}}(\boldsymbol{\beta}) = \Bigg[\underbrace{Y-\breve{\mu}_0(\boldsymbol{X})-\left(1-\frac{1}{K}\right)\langle\boldsymbol{\omega}_A,\vec{\boldsymbol{f}}(\boldsymbol{X};\boldsymbol{\beta})\rangle}_{=\breve{e}(\boldsymbol{\beta})}\Bigg]\boldsymbol{h}(\boldsymbol{X},A).$$

Here, we suppress the potential dependency of $\boldsymbol{h}$ on $(\boldsymbol{\beta}, \breve{\mu}_0)$ and only mention it if necessary. Note that $\mathbb{E}[\boldsymbol{h}(\boldsymbol{X}, A)|\boldsymbol{X}] = \mathbf{0}$ informs the representation

$$\boldsymbol{h}(\boldsymbol{X}, A) = \frac{\mathsf{H}(\boldsymbol{X})\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}; \quad \mathsf{H}(\boldsymbol{X}) := \left(1 - \frac{1}{K}\right) \sum_{k=1}^{K} p_{\mathscr{A}}(k|\boldsymbol{X}) \boldsymbol{h}(\boldsymbol{X}, k) \boldsymbol{\omega}_k^{\intercal} (\Omega^{\intercal}\Omega)^{-1} \in \mathbb{R}^{p \times (K-1)},$$

since by $\Omega^{\intercal}(\Omega^{\intercal}\Omega)^{-1}\Omega^{\intercal} = \mathbf{I}_{K\times K} - (1/K)\vec{\mathbf{1}}_K^{\otimes 2}$, we have

$$\mathsf{H}(\boldsymbol{X})\boldsymbol{\omega}_A = p_{\mathscr{A}}(A|\boldsymbol{X})\boldsymbol{h}(\boldsymbol{X}, A) - \frac{1}{K} \underbrace{\sum_{k=1}^{K} p_{\mathscr{A}}(k|\boldsymbol{X})\boldsymbol{h}(\boldsymbol{X}, k)}_{=\mathbb{E}[\boldsymbol{h}(\boldsymbol{X}, A)|\boldsymbol{X}] = \mathbf{0}} = p_{\mathscr{A}}(A|\boldsymbol{X})\boldsymbol{h}(\boldsymbol{X}, A).$$

Then

$$\begin{aligned}
\mathbb{E}[\breve{\boldsymbol{\phi}}(\boldsymbol{\beta})^{\otimes 2}] =& \mathbb{E}\left[\mathsf{H}(\boldsymbol{X}) \left(\frac{\boldsymbol{\omega}_A^{\otimes 2} \breve{e}(\boldsymbol{\beta})^2}{p_{\mathscr{A}}(A|\boldsymbol{X})^2}\right) \mathsf{H}(\boldsymbol{X})^{\intercal}\right] \\
=& \mathbb{E}\left[\mathsf{H}(\boldsymbol{X}) \mathsf{V}_{\epsilon}(\boldsymbol{X}; \breve{\mu}_0) \mathsf{H}(\boldsymbol{X})^{\intercal}\right]; \\
\mathbb{E}\left[-\frac{\partial \breve{\boldsymbol{\phi}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\intercal}}\right] =& \mathbb{E}\left\{\mathsf{H}(\boldsymbol{X}) \left[\left(1 - \frac{1}{K}\right) \frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})}\right] \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})\right\} \\
& \underbrace{- \mathbb{E}\left\{\dot{\mathsf{H}}(\boldsymbol{X}; \boldsymbol{\beta}, \breve{\mu}_0) \frac{[\mu_0(\boldsymbol{X}) - \breve{\mu}_0(\boldsymbol{X}) + \epsilon]\boldsymbol{\omega}_A}{p_{\mathscr{A}}(A|\boldsymbol{X})}\right\}}_{=0} \\
=& \mathbb{E}\left[\mathsf{H}(\boldsymbol{X})\Omega^{\intercal}\Omega \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})\right].
\end{aligned}$$

Here, the second equality follows from that $\mathbb{E}[\breve{e}(\boldsymbol{\beta})^2|\boldsymbol{X}, A] = \sigma_{\text{opt}}^2(\boldsymbol{X}, A)$ and

$$\mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\breve{e}(\boldsymbol{\beta})^2}{p_{\mathscr{A}}(A|\boldsymbol{X})^2}\middle|\boldsymbol{X}\right) = \mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}\sigma_{\text{opt}}^2(\boldsymbol{X}, A)}{p_{\mathscr{A}}(A|\boldsymbol{X})^2}\middle|\boldsymbol{X}\right) = \sum_{k=1}^{K} \frac{\sigma_{\text{opt}}^2(\boldsymbol{X}, A)\boldsymbol{\omega}_k^{\otimes 2}}{p_{\mathscr{A}}(k|\boldsymbol{X})} = \mathsf{V}_{\epsilon}(\boldsymbol{X}; \breve{\mu}_0).$$

The forth equality follows from that $\mathbb{E}[\boldsymbol{\omega}_A/p_{\mathscr{A}}(A|\boldsymbol{X})|\boldsymbol{X}] = \mathbf{0}$, $\mathbb{E}(\epsilon|\boldsymbol{X}, A) = 0$ and

$$\left(1 - \frac{1}{K}\right) \mathbb{E}\left(\frac{\boldsymbol{\omega}_A^{\otimes 2}}{p_{\mathscr{A}}(A|\boldsymbol{X})}\middle|\boldsymbol{X}\right) = \left(1 - \frac{1}{K}\right) \sum_{k=1}^{K} \boldsymbol{\omega}_k^{\otimes 2} = \Omega^{\intercal}\Omega.$$

Lemma 3.15 (Sandwich Variance Inequality)**.** *Define*

$$\begin{aligned}
&\mathsf{A} := \mathbb{E}[\mathsf{H}(\boldsymbol{X})\mathsf{V}_\epsilon(\boldsymbol{X};\breve{\mu}_0)\mathsf{H}(\boldsymbol{X})^\intercal];\\
&\mathsf{B} := \mathbb{E}[\mathsf{H}(\boldsymbol{X})\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})];\\
&\mathsf{C} := \mathbb{E}[\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal\Omega^\intercal\Omega\mathsf{V}_\epsilon(\boldsymbol{X};\breve{\mu}_0)^{-1}\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})];\\
&\mathsf{X} := \begin{pmatrix}\mathsf{A} & \mathsf{B}\\ \mathsf{B}^\intercal & \mathsf{C}\end{pmatrix}.
\end{aligned}$$

Then

$$\mathsf{X} \succcurlyeq 0 \quad \Leftrightarrow \quad \overbrace{\mathsf{X}/\mathsf{A} = \mathsf{C} - \mathsf{B}^\intercal\mathsf{A}^{-1}\mathsf{B}}^{\textit{Schur complement}} \succcurlyeq 0,$$

with equality if and only if there exists some non-singular constant matrix $\mathsf{H}_0 \in \mathbb{R}^{p\times p}$ *such that*

$$\mathsf{H}(\boldsymbol{X}) = \mathsf{H}_0\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal\Omega^\intercal\Omega\mathsf{V}_\epsilon(\boldsymbol{X};\breve{\mu}_0)^{-1}.$$

Following the notations in Lemma 3.15, we have

$$\begin{aligned}
&\lim_{n\to\infty} n\mathsf{Var}[\widehat{\boldsymbol{\beta}}_n(\breve{\mu}_0)]\\
=&\left\{\mathbb{E}\left[-\frac{\partial\breve{\boldsymbol{\phi}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\intercal}\right]\right\}^{-1}\mathbb{E}[\breve{\boldsymbol{\phi}}(\boldsymbol{\beta})^{\otimes 2}]\left\{\mathbb{E}\left[-\frac{\partial\breve{\boldsymbol{\phi}}(\boldsymbol{\beta})^\intercal}{\partial\boldsymbol{\beta}}\right]\right\}^{-1}\\
=&\mathsf{B}^{-1}\mathsf{A}\mathsf{B}^{-\intercal} = (\mathsf{B}^\intercal\mathsf{A}^{-1}\mathsf{B})^{-1} \succcurlyeq \mathsf{C}^{-1} \qquad \text{(by Lemma 3.15)}\\
=&\mathcal{I}(\boldsymbol{\beta};\breve{\mu}_0)^{-1},
\end{aligned}$$

with equality attained at

$$\mathsf{H}(\boldsymbol{X}) = \mathsf{H}_0\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})^\intercal\Omega^\intercal\Omega\mathsf{V}_\epsilon(\boldsymbol{X};\breve{\mu}_0)^{-1},$$

for some non-singular constant matrix $\mathsf{H}_0 \in \mathbb{R}^{p\times p}$. □

Proof of Lemma 3.15. For any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^p$, define

$$\begin{aligned}
&\boldsymbol{U} := \mathsf{V}_\epsilon(\boldsymbol{X};\breve{\mu}_0)^{1/2}\mathsf{H}(\boldsymbol{X})^\intercal\boldsymbol{u};\\
&\boldsymbol{V} := \mathsf{V}_\epsilon(\boldsymbol{X};\breve{\mu}_0)^{-1/2}\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X};\boldsymbol{\beta})\boldsymbol{v}.
\end{aligned}$$

Then

$$
\begin{aligned}
(\boldsymbol{u}^\intercal, \boldsymbol{v}^\intercal)\mathsf{X}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{pmatrix} &= \boldsymbol{u}^\intercal\mathsf{A}\boldsymbol{u} + 2\boldsymbol{v}^\intercal\mathsf{B}^\intercal\boldsymbol{u} + \boldsymbol{v}^\intercal\mathsf{C}\boldsymbol{v} \\
&= \mathbb{E}(\boldsymbol{U}^\intercal\boldsymbol{U} + 2\boldsymbol{V}^\intercal\boldsymbol{U} + \boldsymbol{V}^\intercal\boldsymbol{V}) \\
&= \mathbb{E}\|\boldsymbol{U} + \boldsymbol{V}\|_2^2 \\
&\geqslant 0,
\end{aligned}
$$

with equality if and only if

$$
\boldsymbol{U} = -\boldsymbol{V} \quad \Leftrightarrow \quad \mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0)^{1/2}\mathsf{H}(\boldsymbol{X})^\intercal\boldsymbol{u} = -\mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0)^{-1/2}\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})\boldsymbol{v} \quad a.s.
$$

That is, for some constant matrix $\mathsf{K} \in \mathbb{R}^{p\times p}$, we have $\boldsymbol{u} = -\mathsf{K}^\intercal\boldsymbol{v}$, and

$$
\begin{aligned}
&\mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0)^{1/2}\mathsf{H}(\boldsymbol{X})^\intercal\mathsf{K}^\intercal = \mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0)^{-1/2}\Omega^\intercal\Omega\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta}) \\
\Leftrightarrow \quad &\mathsf{K}\mathsf{H}(\boldsymbol{X}) = \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal\Omega^\intercal\Omega\mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0)^{-1} \quad a.s.
\end{aligned}
$$

This proves $\mathsf{X} \succcurlyeq 0$ and the equality condition. Finally, by $\mathsf{A} > 0$, we have

$$
\boldsymbol{v}^\intercal(\mathsf{X}/\mathsf{A})\boldsymbol{v} = \boldsymbol{v}^\intercal\mathsf{C}\boldsymbol{v} - \boldsymbol{v}^\intercal\mathsf{B}^\intercal\mathsf{A}^{-1}\mathsf{B}\boldsymbol{v} = \left[(\boldsymbol{u}^\intercal, \boldsymbol{v}^\intercal)\mathsf{X}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{pmatrix}\right]_{\boldsymbol{u}=-\mathsf{A}^{-1}\mathsf{B}\boldsymbol{v}} \geqslant 0.
$$

That is, $\mathsf{X}/\mathsf{A} \succcurlyeq 0$ with equality if and only if for $\mathsf{K} = \mathsf{B}^\intercal\mathsf{A}$ and $\mathsf{H}_0 := \mathsf{K}^{-1} = \mathsf{A}^{-1}\mathsf{B}^{-\intercal}$, we have

$$
\mathsf{H}(\boldsymbol{X}) = \mathsf{H}_0\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal\Omega^\intercal\Omega\mathsf{V}_\epsilon(\boldsymbol{X}; \breve{\mu}_0)^{-1} \quad a.s.
$$

□

3.8.5.10 Proof of Theorem 3.11

Proof of Theorem 3.11. By Assumption 3.5, the distribution of $\boldsymbol{X}$ has compact support. Without loss of generality, assume that $\mathcal{X}$ is compact. Recall from the proof of Lemma 3.13 that $\ddot{F}$ is

the envelop function of $\ddot{\mathsf{F}}$. Then by Assumption 3.4.2 that $\mathbb{E}\ddot{F}(\boldsymbol{X}) + \infty$, we further have that $\|\ddot{F}\|_\infty = \sup_{\boldsymbol{x}\in\mathcal{X}} \ddot{F}(\boldsymbol{x}) < +\infty$.

By $\widehat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \mathcal{O}_{\mathbb{P}}(n^{-1/2})$, we have

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \left\| \dot{\mathsf{F}}(\boldsymbol{x}, \widehat{\boldsymbol{\beta}}_n) - \dot{\mathsf{F}}(\boldsymbol{x}, \boldsymbol{\beta}) \right\|_2 \leqslant \|\ddot{F}\|_\infty \times \|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|_2 = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

Fix $\boldsymbol{x} \in \mathcal{X}$. By Mean Value Theorem, there exists some $\alpha_n \in [0,1]$ and $\widetilde{\boldsymbol{\beta}}_n = (1-\alpha_n)\widehat{\boldsymbol{\beta}}_n + \alpha_n\boldsymbol{\beta}$, such that

$$\begin{aligned}
\vec{\boldsymbol{f}}(\boldsymbol{x}; \widehat{\boldsymbol{\beta}}_n) - \vec{\boldsymbol{f}}(\boldsymbol{x}; \boldsymbol{\beta}) &= \dot{\mathsf{F}}(\boldsymbol{x}, \widetilde{\boldsymbol{\beta}}_n)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \\
&= [\dot{\mathsf{F}}(\boldsymbol{x}, \boldsymbol{\beta}) + \mathcal{O}_{\mathbb{P}}(n^{-1/2})](\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \\
&= \dot{\mathsf{F}}(\boldsymbol{x}, \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) + \mathcal{O}_{\mathbb{P}}(n^{-1}).
\end{aligned}$$

By $\lim_{n\to\infty} n\mathsf{Var}(\widehat{\boldsymbol{\beta}}_n) = \Sigma$, we have

$$\lim_{n\to\infty} n\mathsf{Var}[\vec{\boldsymbol{f}}(\boldsymbol{x}; \widehat{\boldsymbol{\beta}}_n)] = \dot{\mathsf{F}}(\boldsymbol{x}; \boldsymbol{\beta})\Sigma\dot{\mathsf{F}}(\boldsymbol{x}; \boldsymbol{\beta})^\intercal.$$

Suppose $\boldsymbol{X} \sim p_{\mathscr{X}}(\boldsymbol{x})$ and $\boldsymbol{X} \perp\!\!\!\perp \widehat{\boldsymbol{\beta}}_n$. Then

$$\begin{aligned}
&\limsup_{n\to\infty} n \sum_{k=1}^{K} \mathbb{E}[\gamma(\boldsymbol{X}, k; \widehat{\boldsymbol{\beta}}_n) - \gamma(\boldsymbol{X}, k; \boldsymbol{\beta})]^2 \\
=& \limsup_{n\to\infty} \left(1 - \frac{1}{K}\right)^2 \sum_{k=1}^{K} \boldsymbol{\omega}_k^\intercal \mathbb{E}[\vec{\boldsymbol{f}}(\boldsymbol{X}; \widehat{\boldsymbol{\beta}}_n) - \vec{\boldsymbol{f}}(\boldsymbol{X}; \boldsymbol{\beta})]^{\otimes 2} \boldsymbol{\omega}_k && \text{(by Lemma 3.4)} \\
=& \left(1 - \frac{1}{K}\right)^2 \sum_{k=1}^{K} \boldsymbol{\omega}_k^\intercal \mathbb{E}\Bigg[\lim_{n\to\infty} n \underbrace{\mathbb{E}\left\{ [\vec{\boldsymbol{f}}(\boldsymbol{X}; \widehat{\boldsymbol{\beta}}_n) - \vec{\boldsymbol{f}}(\boldsymbol{X}; \boldsymbol{\beta})]^{\otimes 2} \middle| \boldsymbol{X} \right\}}_{=\mathsf{Var}[\vec{\boldsymbol{f}}(\boldsymbol{x}; \widehat{\boldsymbol{\beta}}_n)]|_{\boldsymbol{x}=\boldsymbol{X}}} \Bigg] \boldsymbol{\omega}_k && \text{(by DCT)} \\
=& \left(1 - \frac{1}{K}\right)^2 \sum_{k=1}^{K} \boldsymbol{\omega}_k^\intercal \mathbb{E}[\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})\Sigma\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal] \boldsymbol{\omega}_k \\
=& \left(1 - \frac{1}{K}\right)^2 \mathsf{Tr}\left\{ \sum_{k=1}^{K} \boldsymbol{\omega}_k^\intercal \mathbb{E}[\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})\Sigma\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal] \boldsymbol{\omega}_k \right\} && \text{(trace of a scalar)} \\
=& \left(1 - \frac{1}{K}\right) \mathsf{Tr}\left\{ \mathbb{E}[\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})]\Sigma \right\}. && \text{(commutativity under trace)}
\end{aligned}$$

Finally, by Theorem 3.1, we have

$$\begin{aligned}
\limsup_{n\to\infty} \sqrt{n}[\mathcal{V}(d^\star) - \mathbb{E}\mathcal{V}(\widehat{d}_n)] &\leqslant 2 \limsup_{n\to\infty} \left\{ \sqrt{n} \max_{1\leqslant k\leqslant K} \mathbb{E} \left| \gamma(\boldsymbol{X}, k; \widehat{\boldsymbol{\beta}}_n) - \gamma(\boldsymbol{X}, k; \boldsymbol{\beta}) \right| \right\} \\
&\leqslant 2 \limsup_{n\to\infty} \left\{ n \max_{1\leqslant k\leqslant K} \mathbb{E}[\gamma(\boldsymbol{X}, k; \widehat{\boldsymbol{\beta}}_n) - \gamma(\boldsymbol{X}, k; \boldsymbol{\beta})]^2 \right\}^{1/2} \\
&\leqslant 2 \lim_{n\to\infty} \left\{ n \sum_{k=1}^{K} \mathbb{E}[\gamma(\boldsymbol{X}, k; \widehat{\boldsymbol{\beta}}_n) - \gamma(\boldsymbol{X}, k; \boldsymbol{\beta})]^2 \right\}^{1/2} \\
&= 2 \left(1 - \frac{1}{K} \right)^{1/2} \mathsf{Tr} \left\{ \mathbb{E}[\dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})^\intercal \Omega^\intercal \Omega \dot{\mathsf{F}}(\boldsymbol{X}; \boldsymbol{\beta})] \Sigma \right\}^{1/2}.
\end{aligned}$$

Here, $\mathbb{E}$ is taken over $(\boldsymbol{X}, \widehat{\boldsymbol{\beta}}_n)$. □

3.8.6 Additional Tables and Figures

Table 3.5: Estimated Coefficients on the ACTG175 Dataset (averaged over 10 replications)

variable	D-Learning				RD-Learning				E-Learning			
	ddI	ZDV	ZDV+ddI	ZDV+ZAL	ddI	ZDV	ZDV+ddI	ZDV+ZAL	ddI	ZDV	ZDV+ddI	ZDV+ZAL
I: original data												
Intercept	4.72	**-29.9**	**26.45**	-1.27	4.33	**-31.54**	**27.87**	-0.66	4.67	**-30.67**	**26.52**	-0.52
gender									-0.25	0.36	0.05	-0.16
homo	2.2	-0.12	-2.74	0.65	3.63	-0.53	-4.55	1.45	3.26	-0.4	-4.36	1.49
race	-0.12	0.4	-0.41	0.13	-0.41	1.04	-1.08	0.45	-0.57	2.56	-2.67	0.69
drugs	-3.29	-0.96	2.77	1.49	-3.3	-1.12	2.48	1.94	-4.26	-2.68	3.13	3.81
symptom									-0.08	0.08	-0.06	0.06
str2									0.11	0.07	-0.26	0.09
hemo									0.25	-0.5	-0.01	0.26
age	-0.54	-0.48	4.79	-3.77	-0.64	-0.55	**7.67**	**-6.48**	-0.2	0.1	**6.65**	**-6.55**
wtkg					0.24	-0.26	-0.19	0.21	0.5	-1.03	-0.36	0.9
cd40	**9.02**	3.22	**-13.54**	1.29	**7.94**	1.02	**-7.94**	-1.03	**5.82**	1.15	**-5.56**	-1.41
karnof	-0.02	-0.08	0.1	0					0.15	-0.25	-0.01	0.1
cd80									-0.22	-0.1	-0.03	0.35
II: modified treatment-free effect in age												
Intercept	1.69	-3.98	**16.8**	**-14.5**					3.44	**-15.74**	**11.17**	1.12
gender	-0.5	**5.59**	**-6.06**	0.97					0.95	-0.28	-1	0.34
homo	0.16	-0.19	0.81	-0.78					0.84	-0.31	-0.77	0.24
race	-3.52	2.19	-4.69	**6.03**					-0.76	0.46	-1.87	2.18
drugs	**-12.55**	**-16.86**	**24.37**	**5.04**					-0.64	-0.4	0.49	0.55
symptom	**13.4**	**-8.08**	**-9.37**	4.05								
str2	**-14.12**	-2.69	**9.25**	**7.56**					-0.01	-0.09	-0.06	0.16
hemo	-0.19	1.34	1.08	-2.23					0.05	-0.56	-0.91	1.43
age	3.77	**50.53**	**20.2**	**-74.49**					-0.07	0.02	0.01	0.04
wtkg	**-20.77**	0.78	**11.18**	**8.81**					1.08	-0.71	-1.79	1.42
cd40	**16.05**	**6.65**	**-22.85**	0.15					1.58	2.24	-4.34	0.53
karnof	**-9.9**	-1.56	**9.31**	2.14					-0.01	0	0.01	-0.01
cd80	**-18.92**	**7.55**	**17.75**	**-6.37**					-0.18	0.13	-0.24	0.28
III: modified variance function in wtkg												
Intercept									0.02	**-22.65**	**27.54**	-4.91
gender									0.74	3.24	**-8.76**	4.78
homo									4.36	-1.31	1.44	-4.49
race									-2.17	-0.28	3.72	-1.28
drugs									-1.77	-0.9	-0.51	3.18
symptom									-1.06	-0.78	-1.01	2.85
str2									0.72	-1.68	-1.18	2.14
hemo									1.6	-1.41	2.57	-2.76
age									2.2	-0.41	0.02	-1.8
wtkg	**-41.32**	**-13.14**	**87.41**	**-32.96**	**-15.54**	**-5.29**	**32.88**	**-12.04**	**-11.18**	-1.19	**16.93**	-4.56
cd40									1.94	-0.72	1.74	-2.96
karnof									2.84	-1.02	-3.35	1.53
cd80	**-20.1**	**-20.18**	**61**	**-20.73**	**-7.14**	**-8.31**	**24.04**	**-8.59**	-3.51	-4.5	**11.98**	-3.97
IV = II + III: modified treatment-free effect in age and modified variance function in wtkg												
Intercept	3.97	-0.11	-4.08	0.22					1.43	**-24.28**	**25.71**	-2.87
gender	**-7.29**	1.49	**11.15**	**-5.35**					**5.8**	2.22	**-5.33**	-2.68
homo	**5.02**	1.7	**-10.68**	3.96					**9.9**	**-5.34**	**-10.71**	**6.16**
race	**6.09**	**10.4**	**-27.53**	**11.04**					4.47	-0.29	**-13.69**	**9.51**
drugs	-3.77	-3.03	**9.26**	-2.45					-1.61	-2.18	-0.12	3.9
symptom	1.18	-0.57	0.56	-1.17					1.79	-0.47	-1.29	-0.03
str2	**-5.19**	0.95	**5.4**	-1.17					1.98	**-10.15**	**-5.35**	**13.52**
hemo	2.88	1.93	**-6.75**	1.94					4.35	**-8.64**	-4.36	**8.64**
age	**-9.42**	**11.74**	**25.67**	**-27.99**					-4.6	**5.59**	4.94	**-5.93**
wtkg	**65.69**	**17.05**	**-153.12**	**70.38**	**18.9**	**5.13**	**-43.57**	**19.53**	**6.31**	**-9.97**	-1.03	4.7
cd40	**16.05**	**14.91**	**-44.87**	**13.91**					-1.12	0.11	-0.22	1.23
karnof	**-9.03**	-1.3	**16.11**	**-5.78**					**-5.12**	0.38	0.97	3.77
cd80	**26.31**	**40.74**	**-95.75**	**28.7**	**5.37**	**6.4**	**-16.73**	4.96	0.98	-1.14	**-7.01**	**7.18**

Note:

Larger coefficients encourage better outcome.

Coefficients are fitted at standardized scales of covariates.

Coefficients at blank are 0's. Absolute values > 5 are bolded.

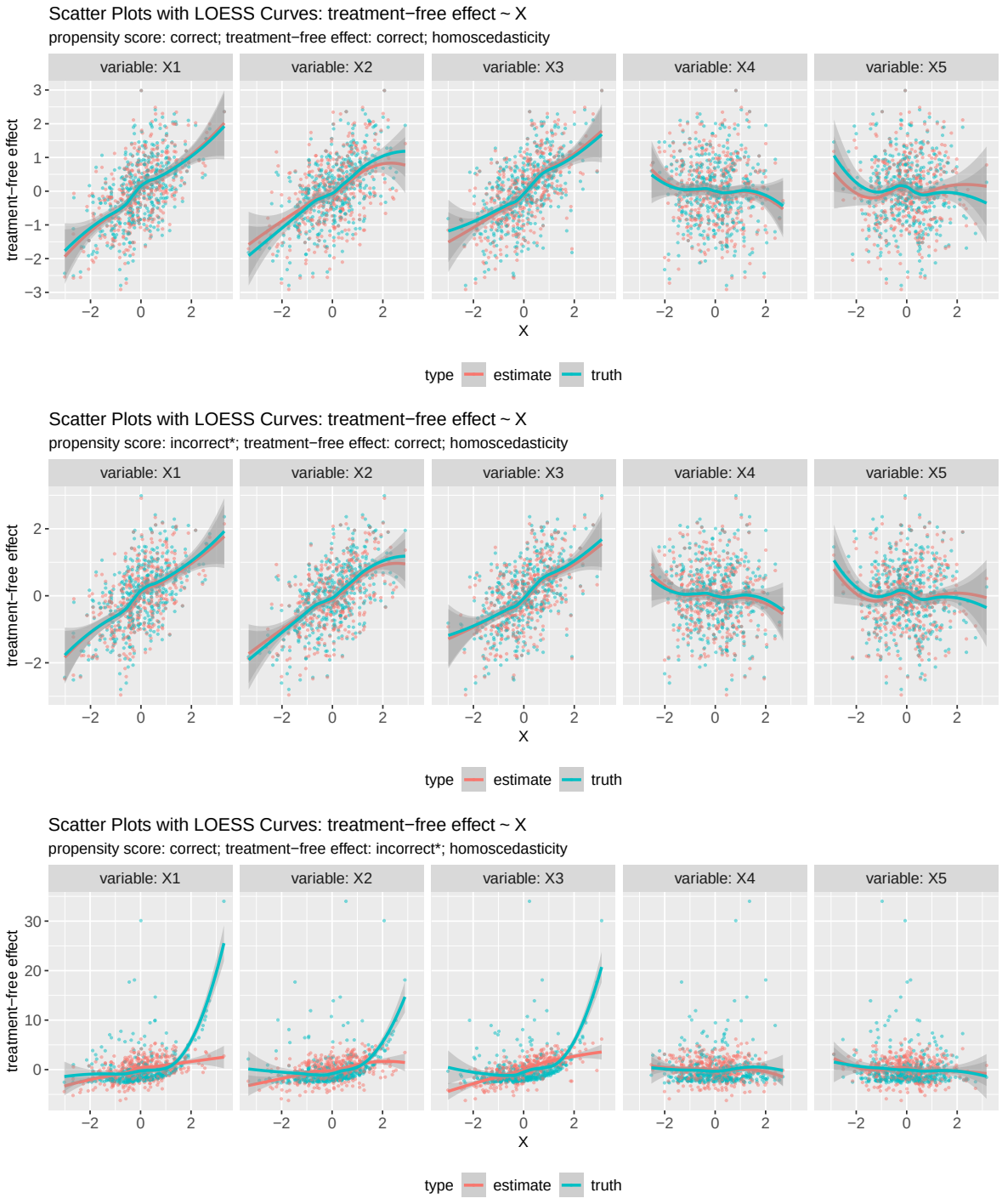


Figure 3.5: Fitted treatment-free effect plots with respect to X_k for $1 \leqslant k \leqslant 5$ for the simulation studies (Section 3.5) with $n = 400$, $p = 10$, $K = 3$. Curves are fitted by the *LOcally wEighted Scatterplot Smoothing (LOESS)* of cubic spline. When the treatment-free effect model is correctly specified (Rows 1 and 2), it can be consistently estimated. Note that the treatment-free effect estimation utilizes the estimated propensity scores according to Section 3.2.5.2. The correctness of the treatment-free effect is not affected by the correctness of the propensity score model. When the treatment-free effect model is misspecified (Row 3), the estimated treatment-free effect deviates from the truth.

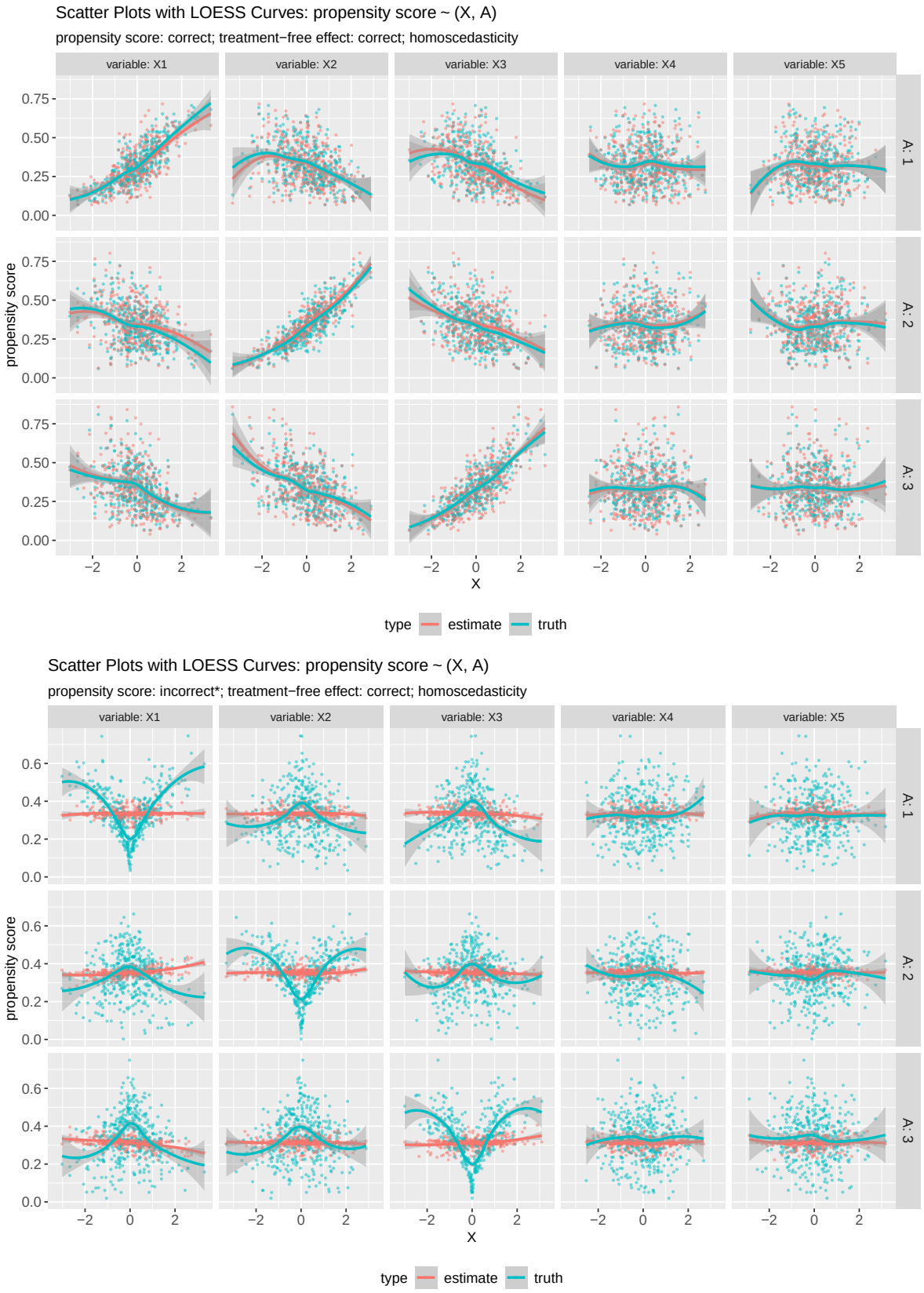


Figure 3.6: Fitted propensity score plots with respect to (X_k, A) for $1 \leqslant k \leqslant 5$ for the simulation studies (Section 3.5) with $n = 400$, $p = 10$, $K = 3$. Curves are fitted by the LOESS of cubic spline. When the propensity score model is correctly specified (Panel 1), it can be consistently estimated. When the propensity score model is misspecified (Panel 2), the estimated propensity score deviates from the truth.

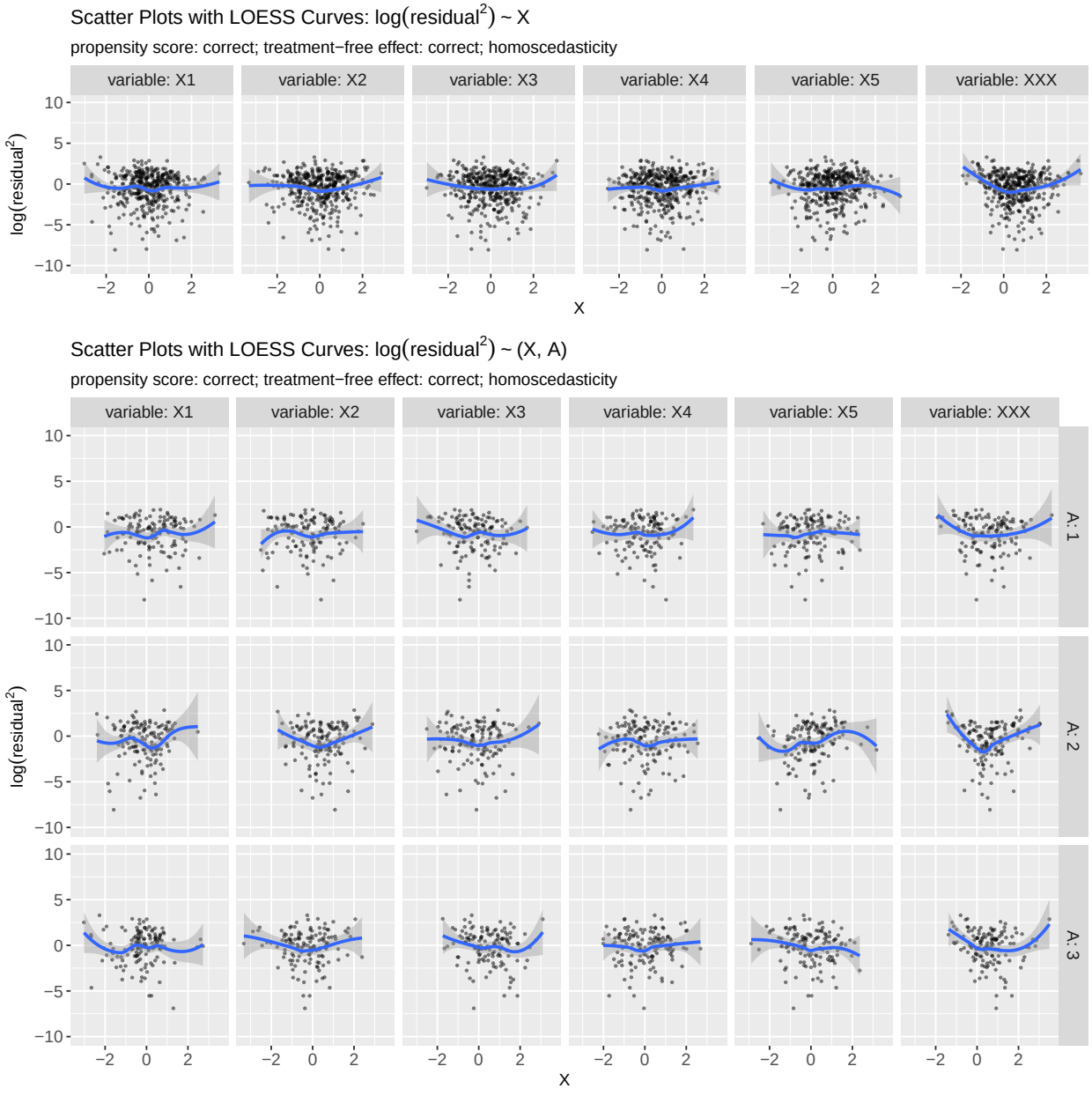


Figure 3.7: Residual plots with respect to X_k and (X_k, A) for $1 \leqslant k \leqslant 5$ for the simulation studies (Section 3.5) with $n = 400$, $p = 10$, $K = 3$, correctly specified treatment-free effect and homoscedasticity. Define $XXX := \log\left[\frac{1}{3}\left(e^{\sqrt{2}X_1} + e^{\sqrt{2}X_2} + e^{\sqrt{2}X_3}\right)\right]$. Residuals are computed from the fitted E-Learning. Curves are fitted by the LOESS of cubic spline. It shows no patterns of $\log(\texttt{residual}^2)$ with respect to $\boldsymbol{X}$ or A.

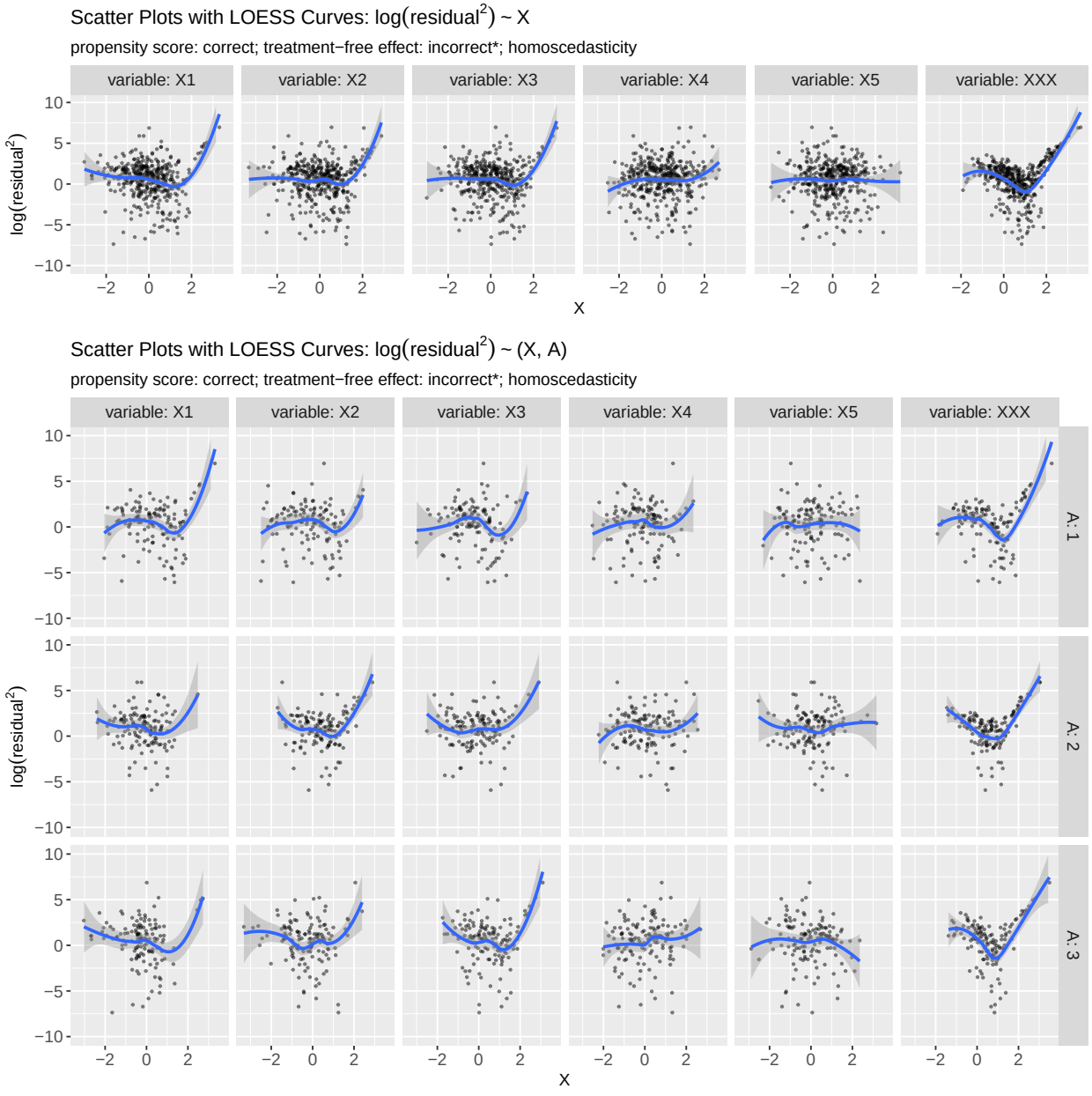


Figure 3.8: Residual plots with respect to X_k and (X_k, A) for $1 \leqslant k \leqslant 5$ for the simulation studies (Section 3.5) with $n = 400$, $p = 10$, $K = 3$, misspecified treatment-free effect and homoscedasticity. Define $XXX := \log\left[\frac{1}{3}\left(e^{\sqrt{2}X_1} + e^{\sqrt{2}X_2} + e^{\sqrt{2}X_3}\right)\right]$. Residuals are computed from the fitted E-Learning. Curves are fitted by the LOESS of cubic spline. It shows patterns of $\log(\texttt{residual}^2) \sim X_k$ for $k = 1, 2, 3$ and $\log(\texttt{residual}^2) \sim XXX$.

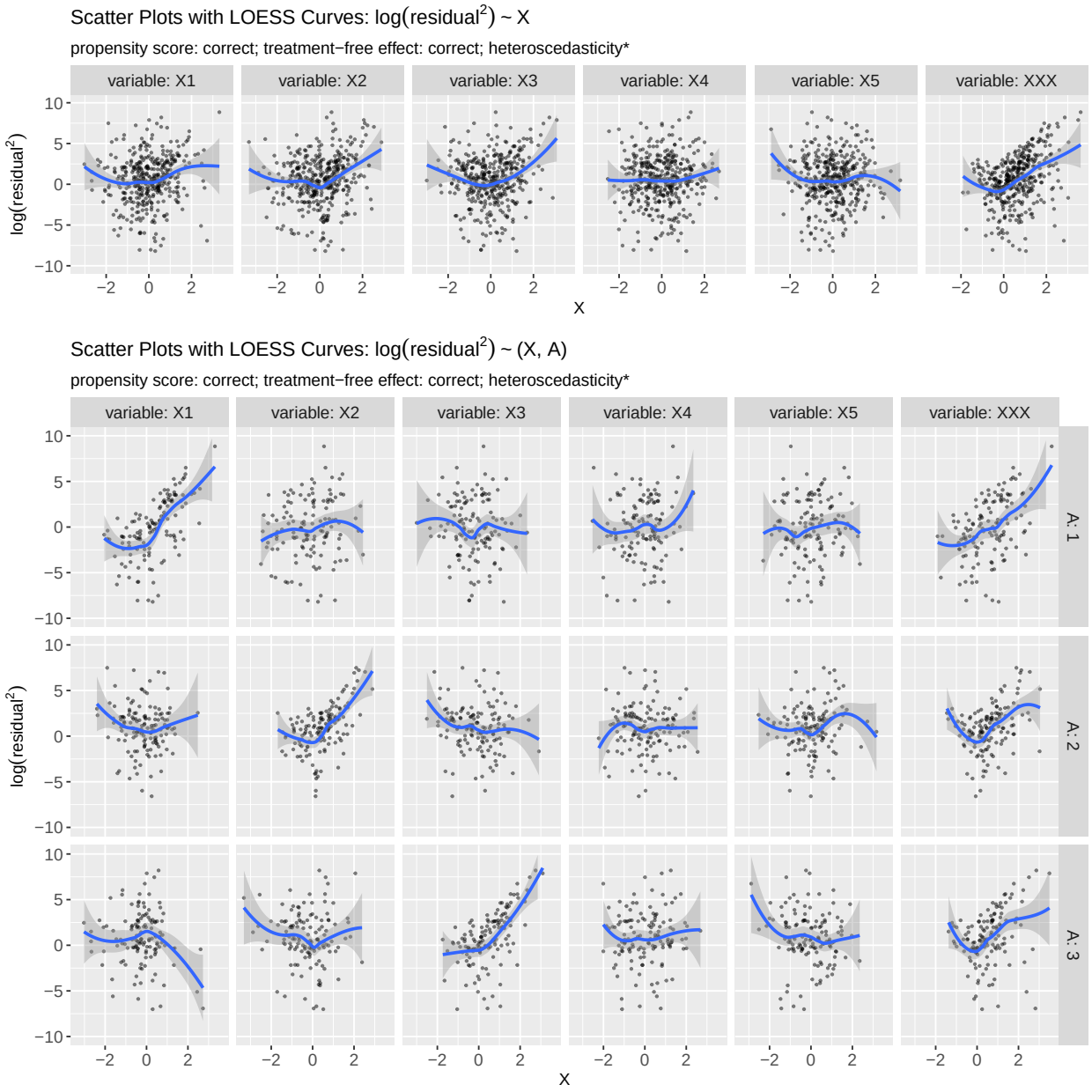


Figure 3.9: Residual plots with respect to X_k and (X_k, A) for $1 \leqslant k \leqslant 5$ for the simulation studies (Section 3.5) with $n = 400$, $p = 10$, $K = 3$, correctly specified treatment-free effect and heteroscedasticity. Define $XXX := \log\left[\frac{1}{3}\left(e^{\sqrt{2}X_1} + e^{\sqrt{2}X_2} + e^{\sqrt{2}X_3}\right)\right]$. Residuals are computed from the fitted E-Learning. Curves are fitted by the LOESS of cubic spline. It shows patterns of $\log(\texttt{residual}^2) \sim X_k$ on $A = k$ for $k = 1, 2, 3$ and $\log(\texttt{residual}^2) \sim XXX$.

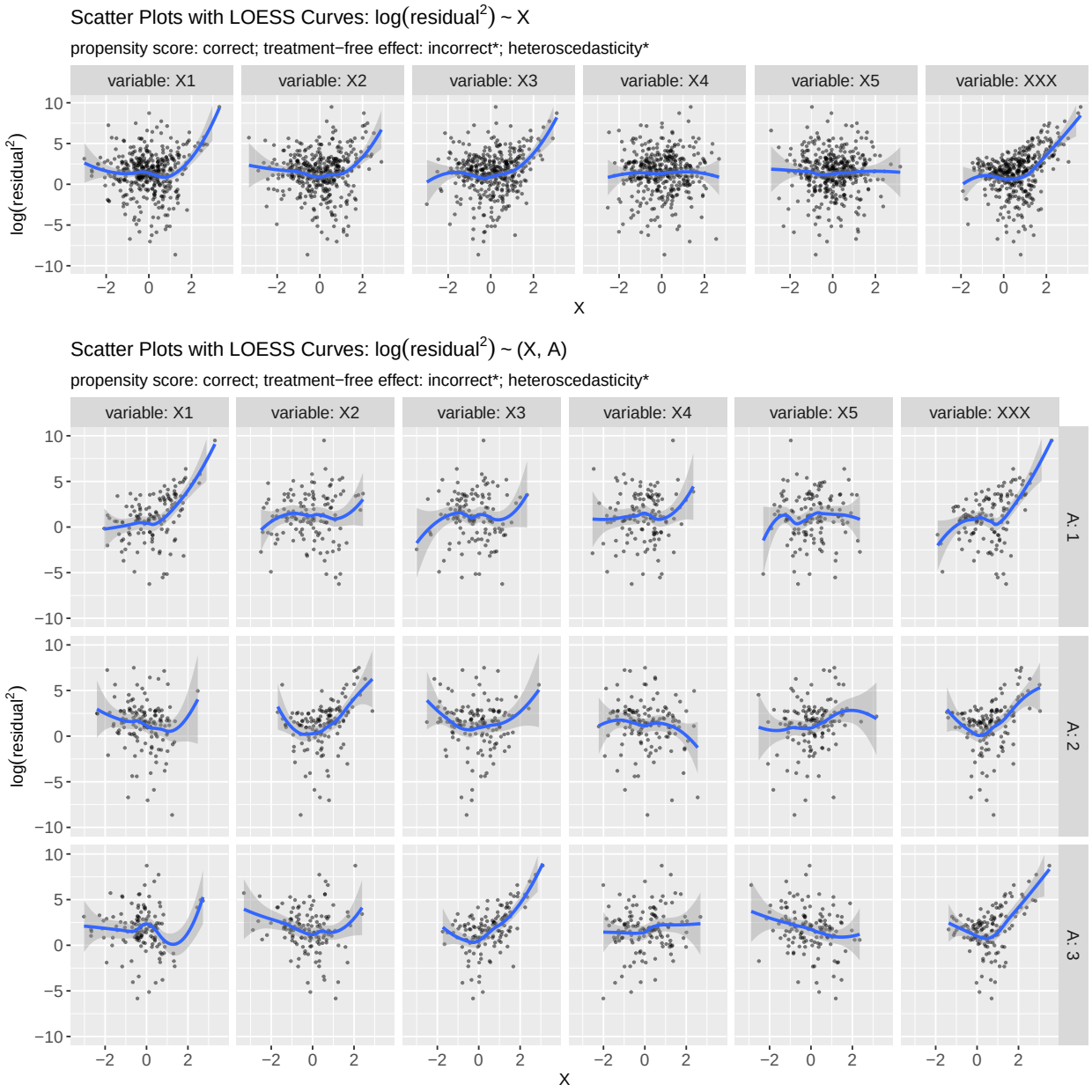


Figure 3.10: Residual plots with respect to X_k and (X_k, A) for $1 \leqslant k \leqslant 5$ for the simulation studies (Section 3.5) with $n = 400$, $p = 10$, $K = 3$, misspecified treatment-free effect and heteroscedasticity. Define $XXX := \log\left[\frac{1}{3}\left(e^{\sqrt{2}X_1} + e^{\sqrt{2}X_2} + e^{\sqrt{2}X_3}\right)\right]$. Residuals are computed from the fitted E-Learning. Curves are fitted by the LOESS of cubic spline. It shows patterns of $\log(\texttt{residual}^2) \sim X_k$ on $A = k$ for $k = 1, 2, 3$ and $\log(\texttt{residual}^2) \sim XXX$.

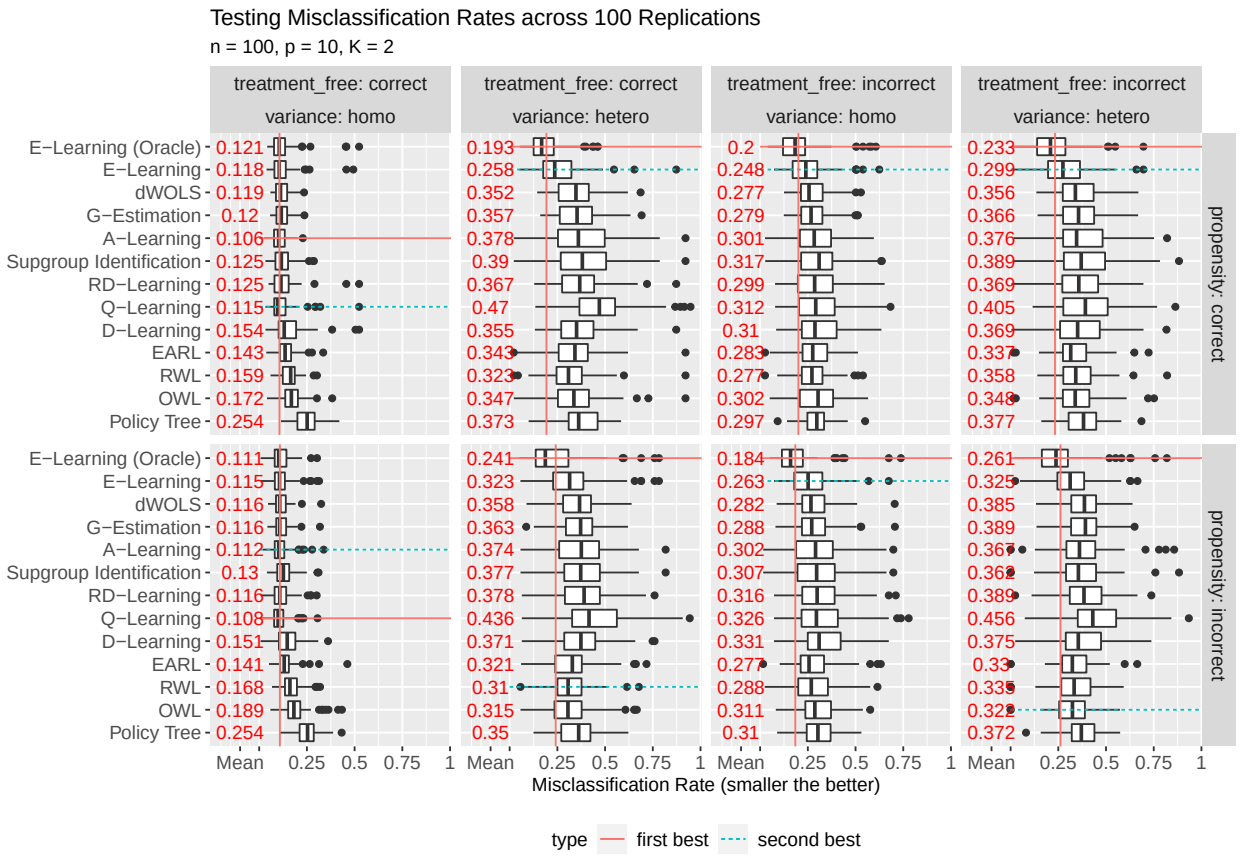


Figure 3.11: Testing misclassification rates (smaller the better) for $n = 100$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged misclassification rates are annotated in horizontal lines, while the minimal averaged misclassification rate is annotated in the vertical line.

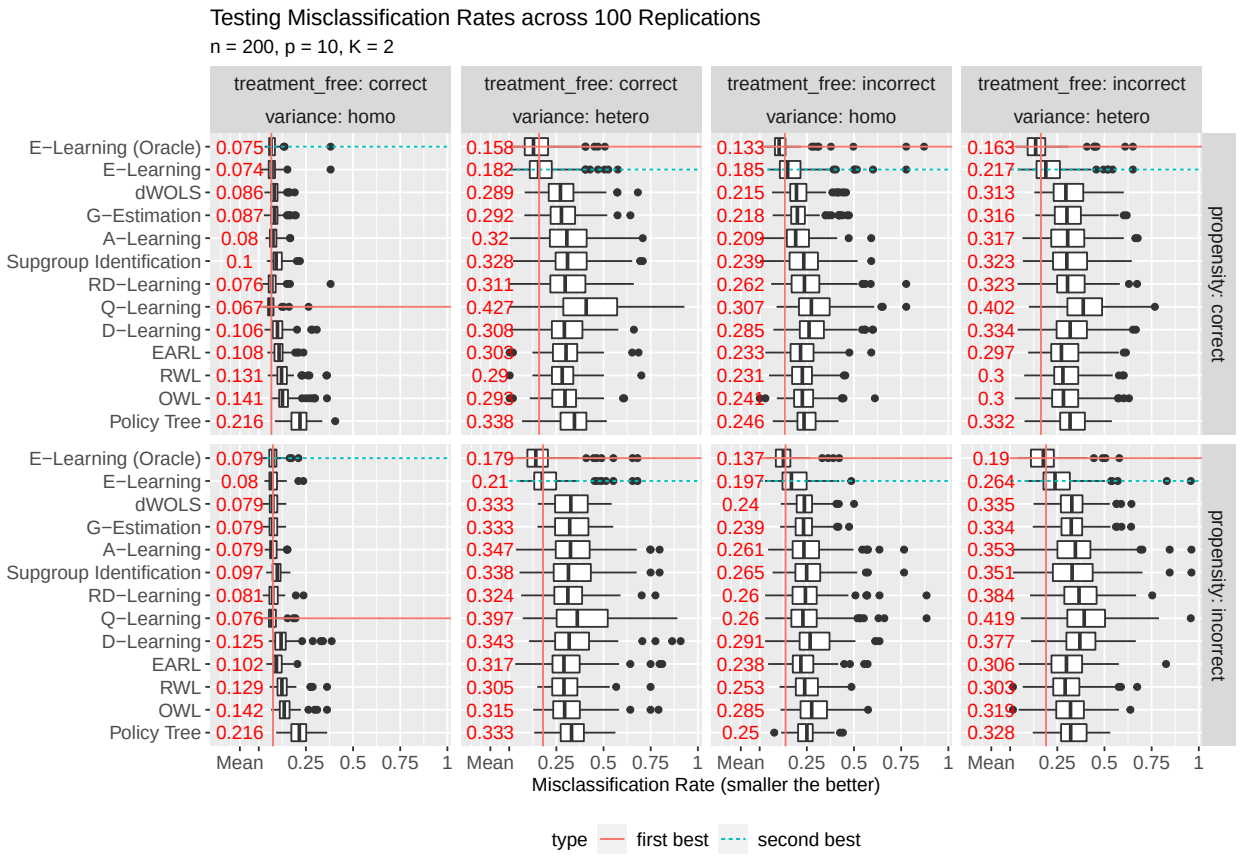


Figure 3.12: Testing misclassification rates (smaller the better) for $n = 200$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged misclassification rates are annotated in horizontal lines, while the minimal averaged misclassification rate is annotated in the vertical line.

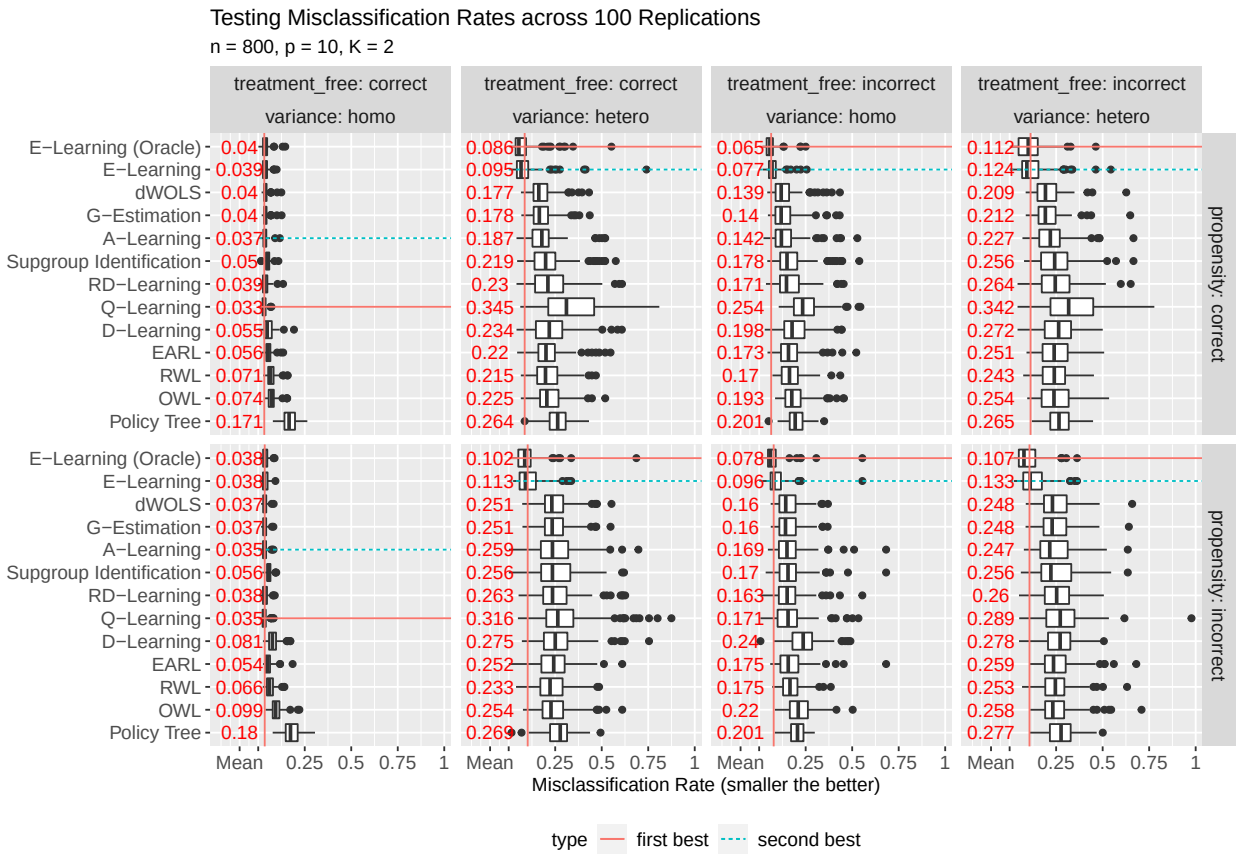


Figure 3.13: Testing misclassification rates (smaller the better) for $n = 800$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged misclassification rates are annotated in horizontal lines, while the minimal averaged misclassification rate is annotated in the vertical line.

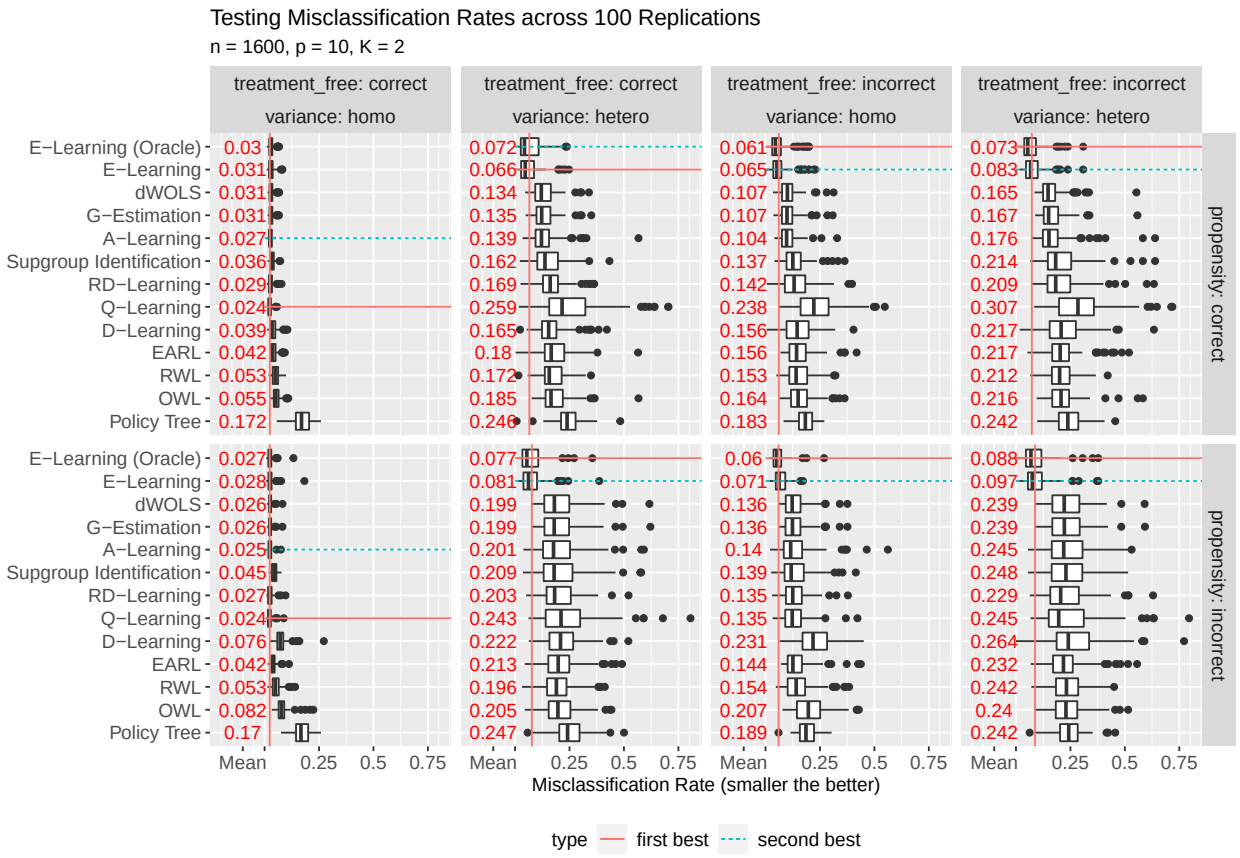


Figure 3.14: Testing misclassification rates (smaller the better) for $n = 1600$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged misclassification rates are annotated in horizontal lines, while the minimal averaged misclassification rate is annotated in the vertical line.

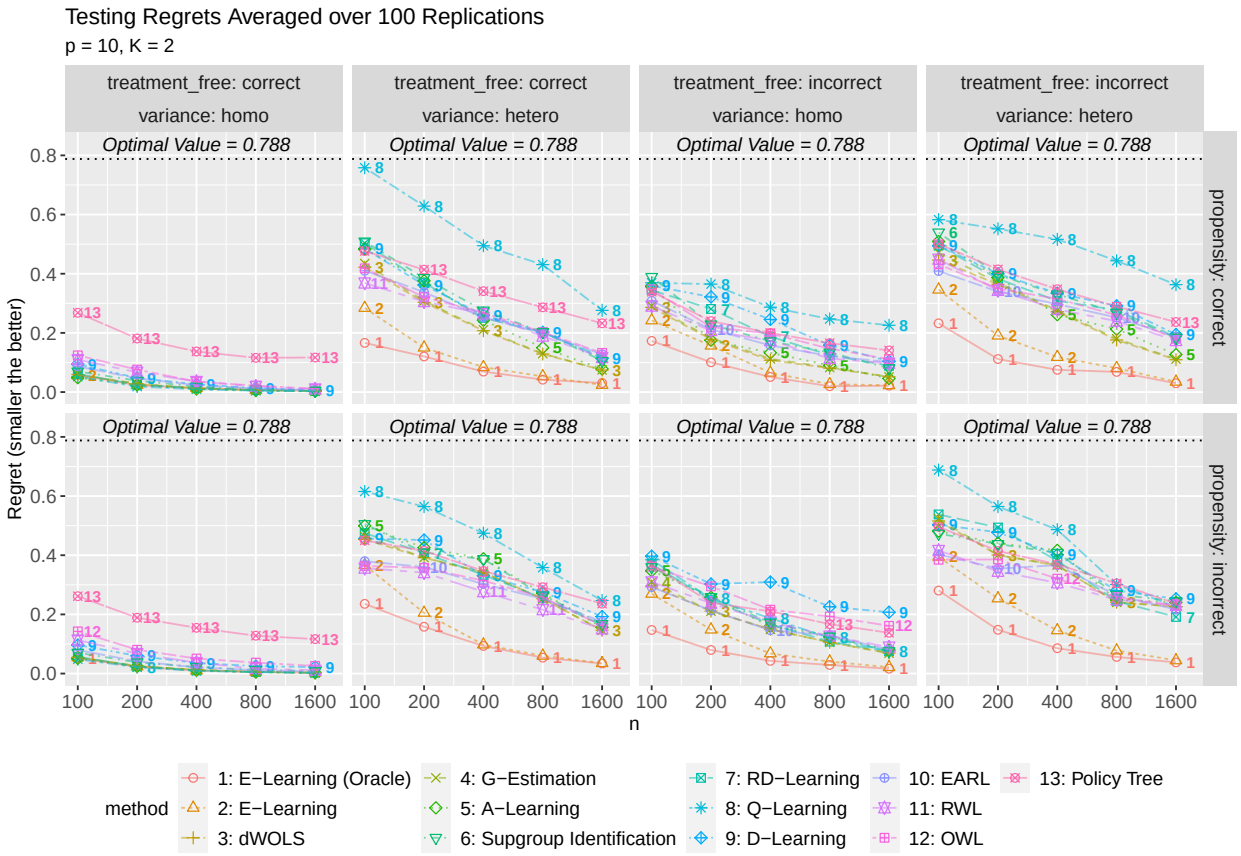


Figure 3.15: Testing regrets (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees.

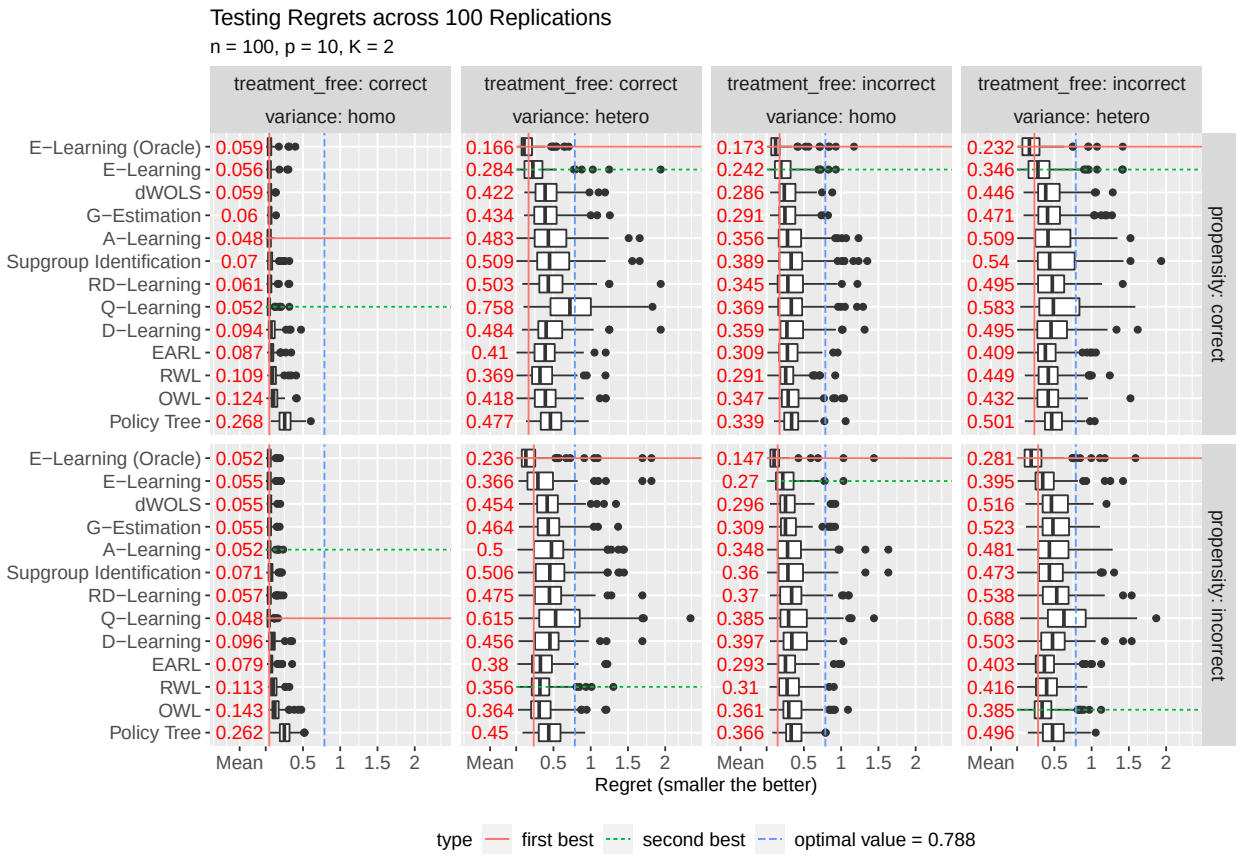


Figure 3.16: Testing regrets (smaller the better) for $n = 100$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged regrets are annotated in horizontal lines, while the minimal averaged regret is annotated in the vertical line. The optimal value is 0.788 and is annotated in the vertical long dashed line.

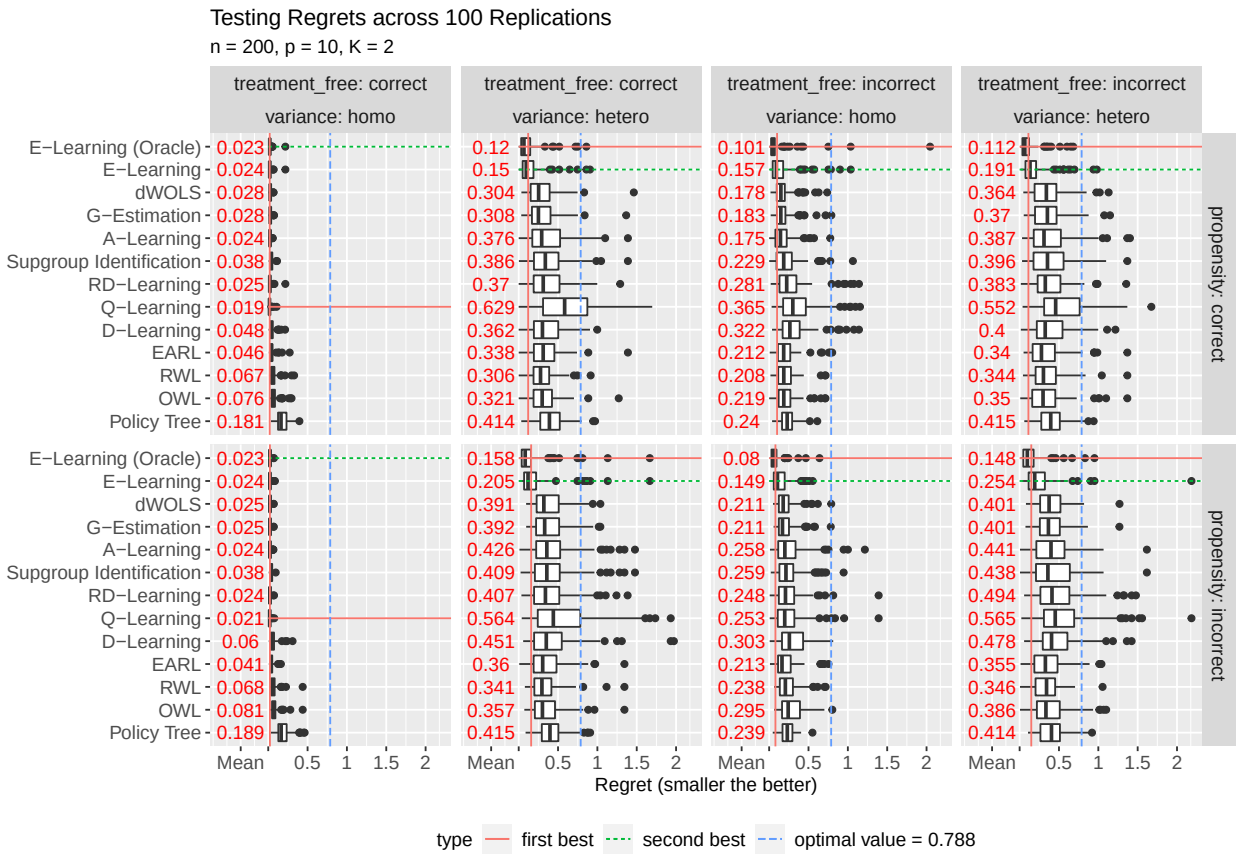


Figure 3.17: Testing regrets (smaller the better) for $n = 200$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged regrets are annotated in horizontal lines, while the minimal averaged regret is annotated in the vertical line. The optimal value is 0.788 and is annotated in the vertical long dashed line.

Figure 3.18: Testing regrets (smaller the better) for $n = 400$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged regrets are annotated in horizontal lines, while the minimal averaged regret is annotated in the vertical line. The optimal value is 0.788 and is annotated in the vertical long dashed line.

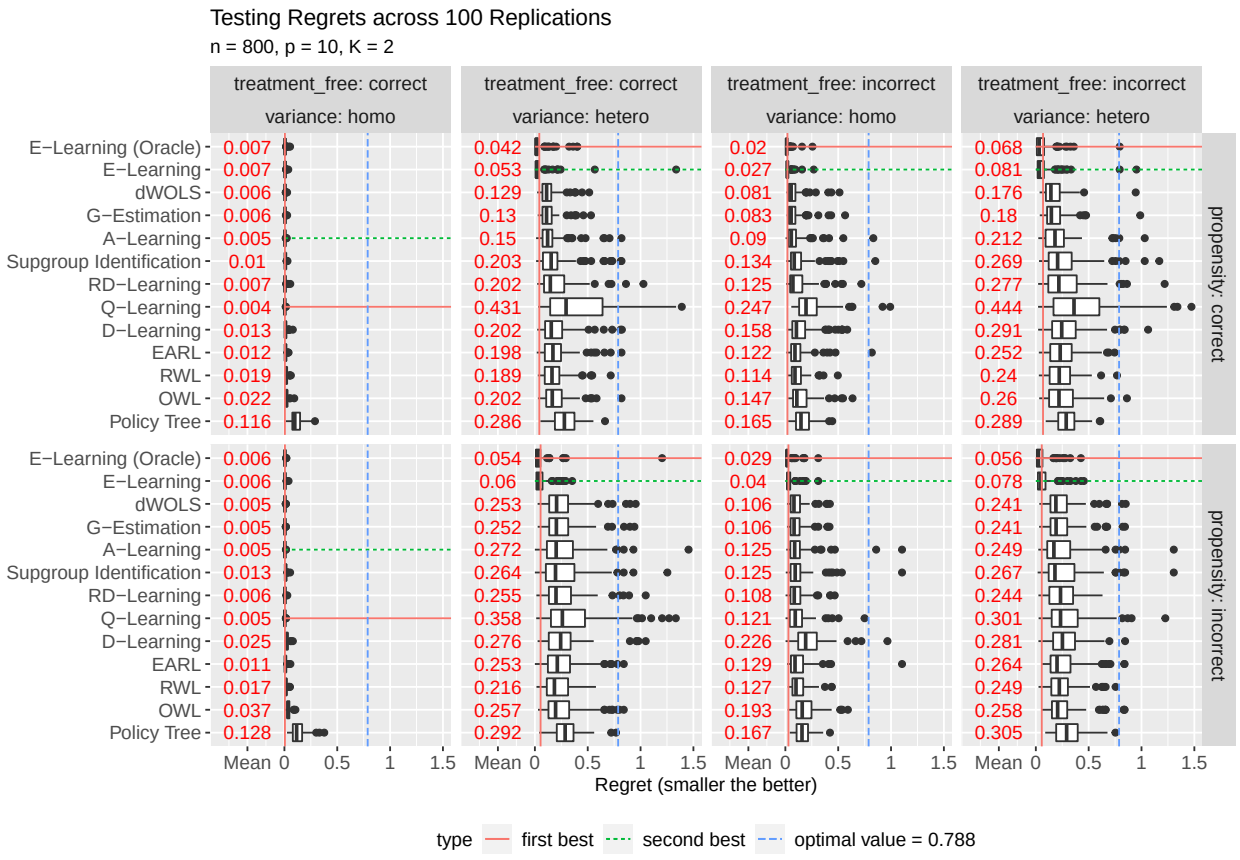


Figure 3.19: Testing regrets (smaller the better) for $n = 800$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged regrets are annotated in horizontal lines, while the minimal averaged regret is annotated in the vertical line. The optimal value is 0.788 and is annotated in the vertical long dashed line.

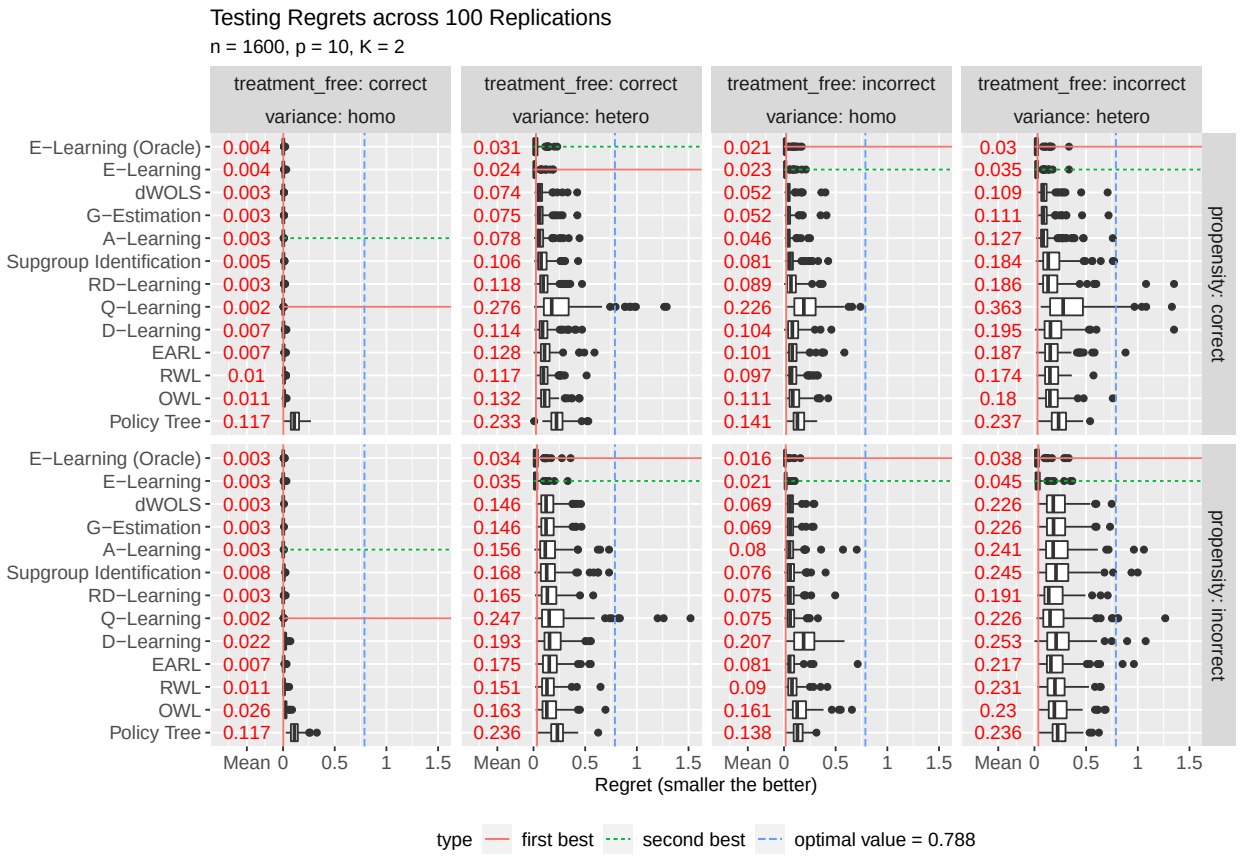


Figure 3.20: Testing regrets (smaller the better) for $n = 1600$, $p = 10$, $K = 2$ and each of the model specification scenarios in Table 3.2. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. First and second best methods in terms of the averaged regrets are annotated in horizontal lines, while the minimal averaged regret is annotated in the vertical line. The optimal value is 0.788 and is annotated in the vertical long dashed line.

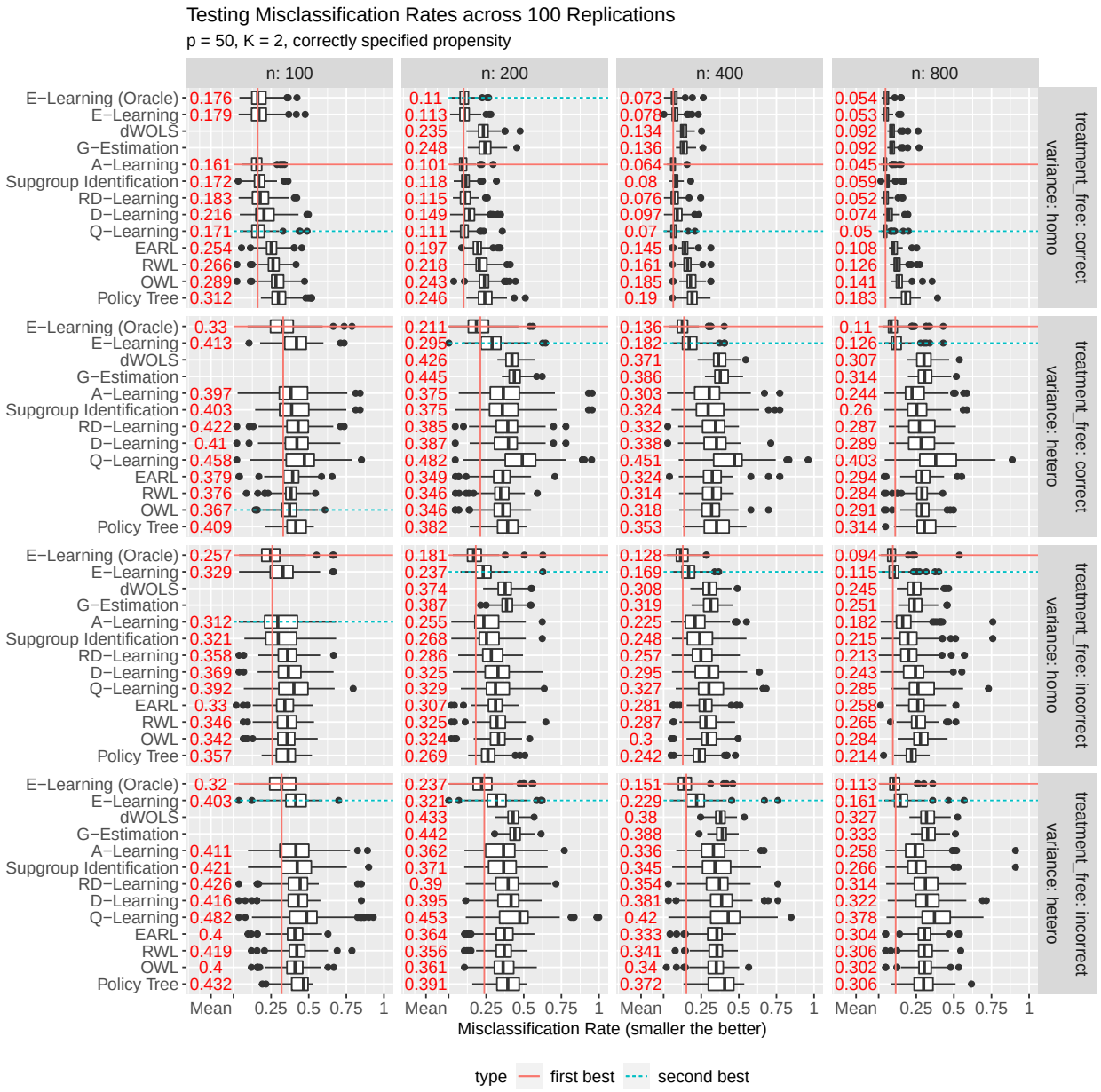


Figure 3.21: Testing misclassification rates (smaller the better) for $n \in \{100, 200, 400, 800\}$, $p = 50$, $K = 2$ and each of the model specification scenarios (correct propensity score) in Table 3.2. The optimal value is 0.788. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. *dWOLS* and *G-Estimation* for $n = 100$ cannot be implemented due to more number of parameters $2(p + 1)$ than the training sample size n. First and second best methods in terms of the averaged misclassification rates are annotated in horizontal lines, while the minimal averaged misclassification rates is annotated in the vertical line.

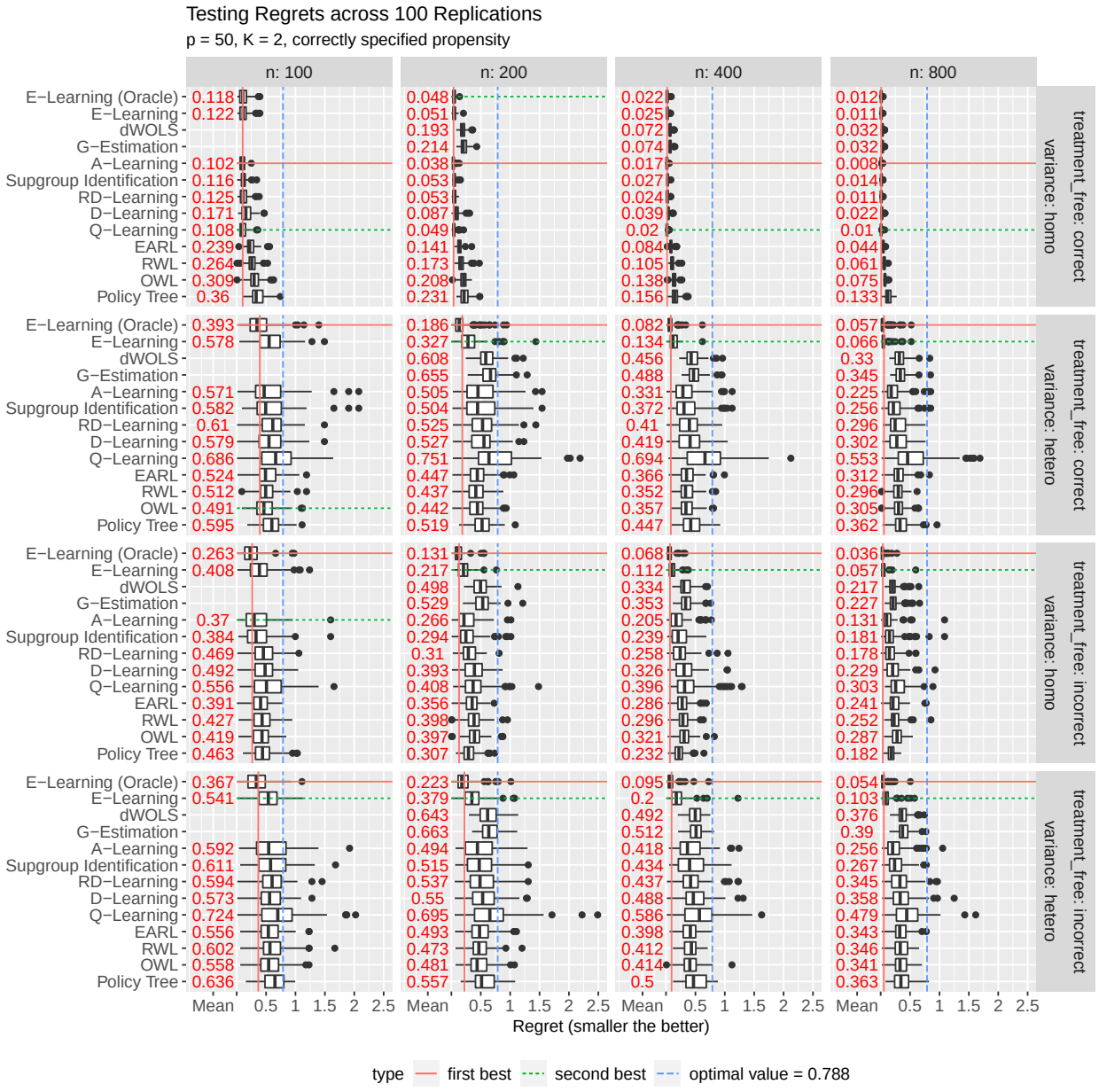


Figure 3.22: Testing regrets (smaller the better) for $n \in \{100, 200, 400, 800\}$, $p = 50$, $K = 2$ and each of the model specification scenarios (correct propensity score) in Table 3.2. The optimal value is 0.788. Methods in Table 3.1 are compared, where *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *Policy Tree* corresponds to *Policy Learning* with decision trees. *dWOLS* and *G-Estimation* for $n = 100$ cannot be implemented due to more number of parameters $2(p + 1)$ than the training sample size n. First and second best methods in terms of the averaged regrets are annotated in horizontal lines, while the minimal averaged regret is annotated in the vertical line.

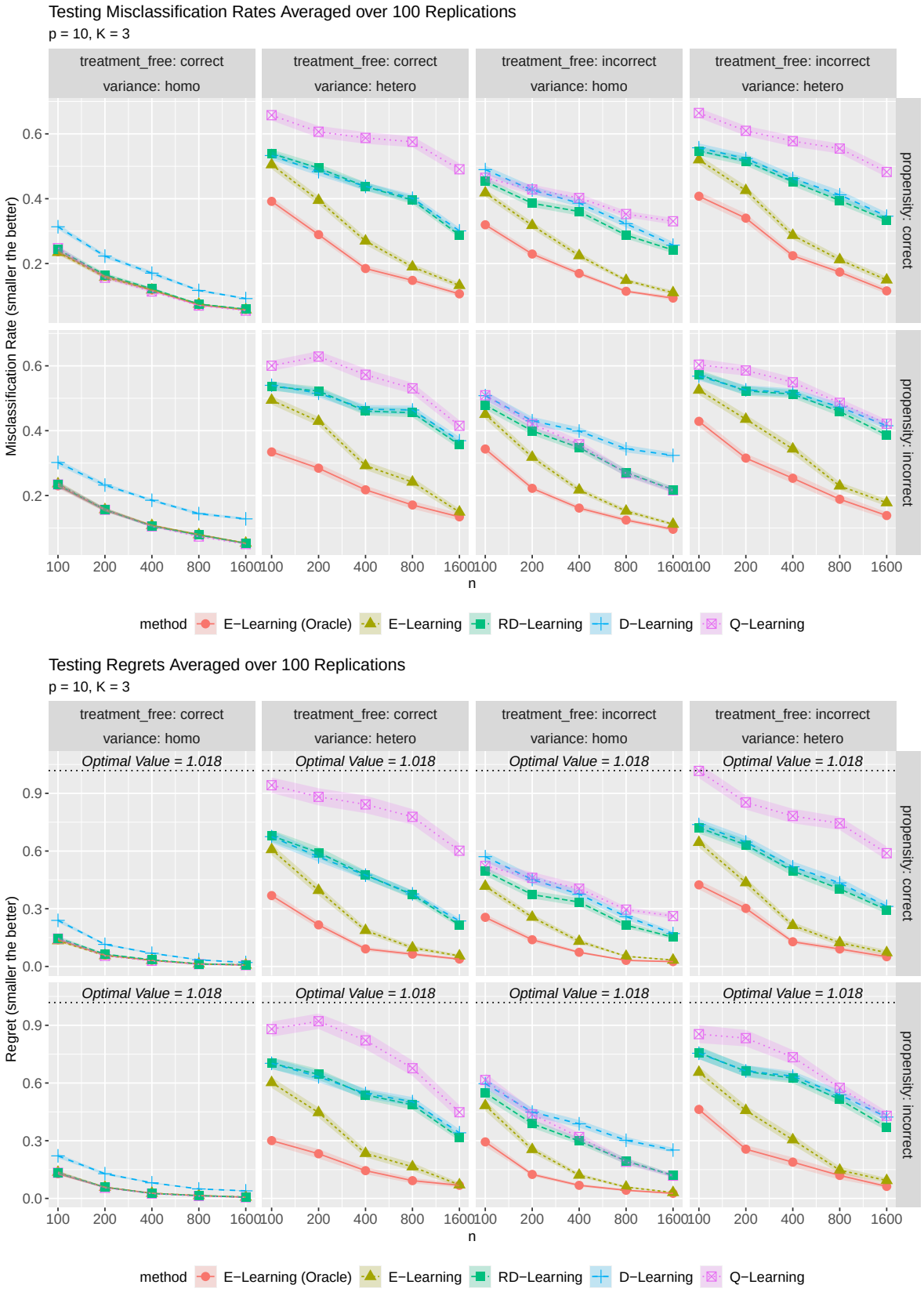


Figure 3.23: Testing misclassification rates and regrets (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p = 10$, $K = 3$ and each of the model specification scenarios in Table 3.2. *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *E-Learning* corresponds to E-Learning with the working variance function estimated by regression forest.

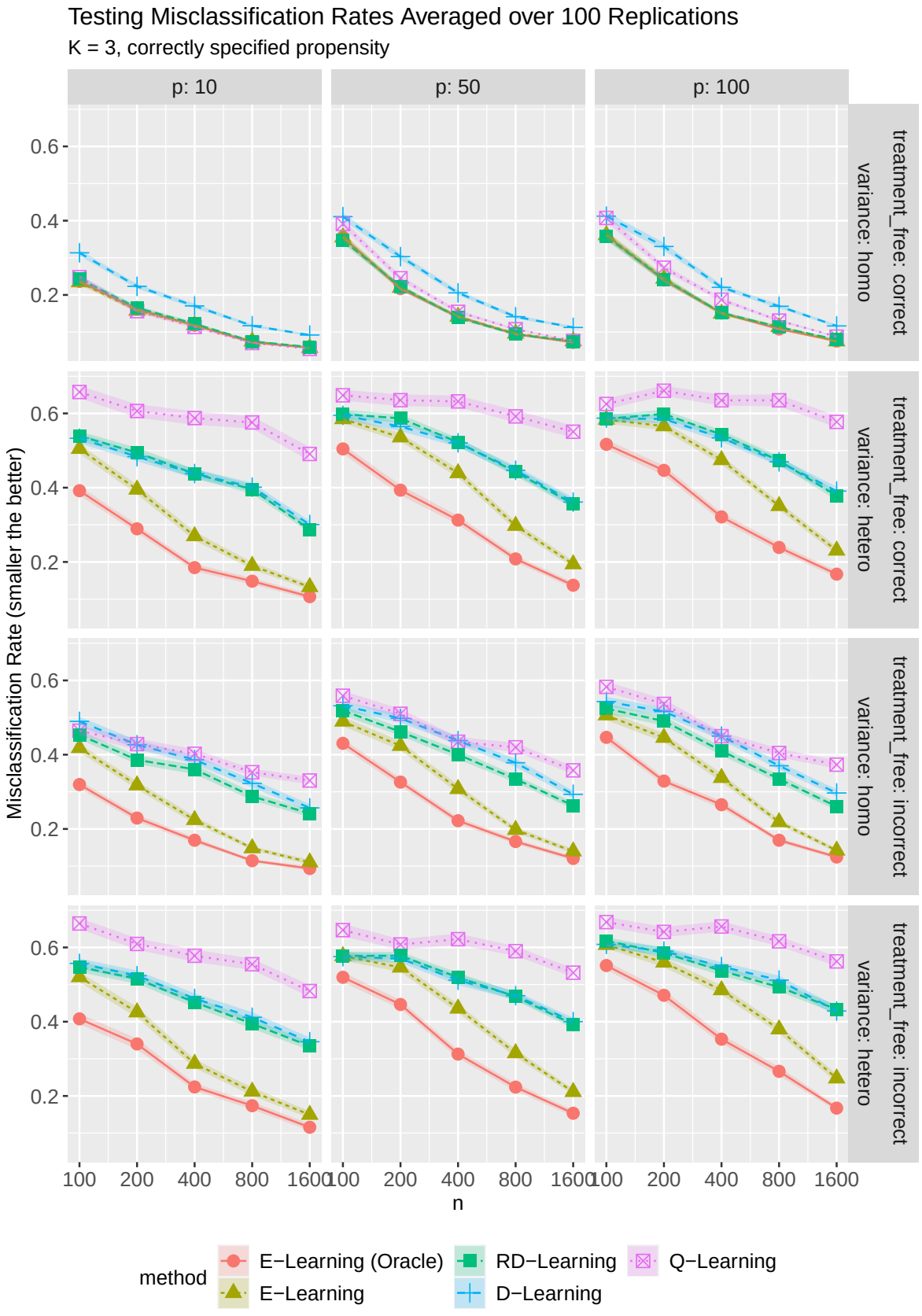


Figure 3.24: Testing misclassification rates (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p \in \{10, 50, 100\}$, $K = 3$ and each of the model specification scenarios with correctly specified propensity score in Table 3.2. *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *E-Learning* corresponds to E-Learning with the working variance function estimated by regression forest.

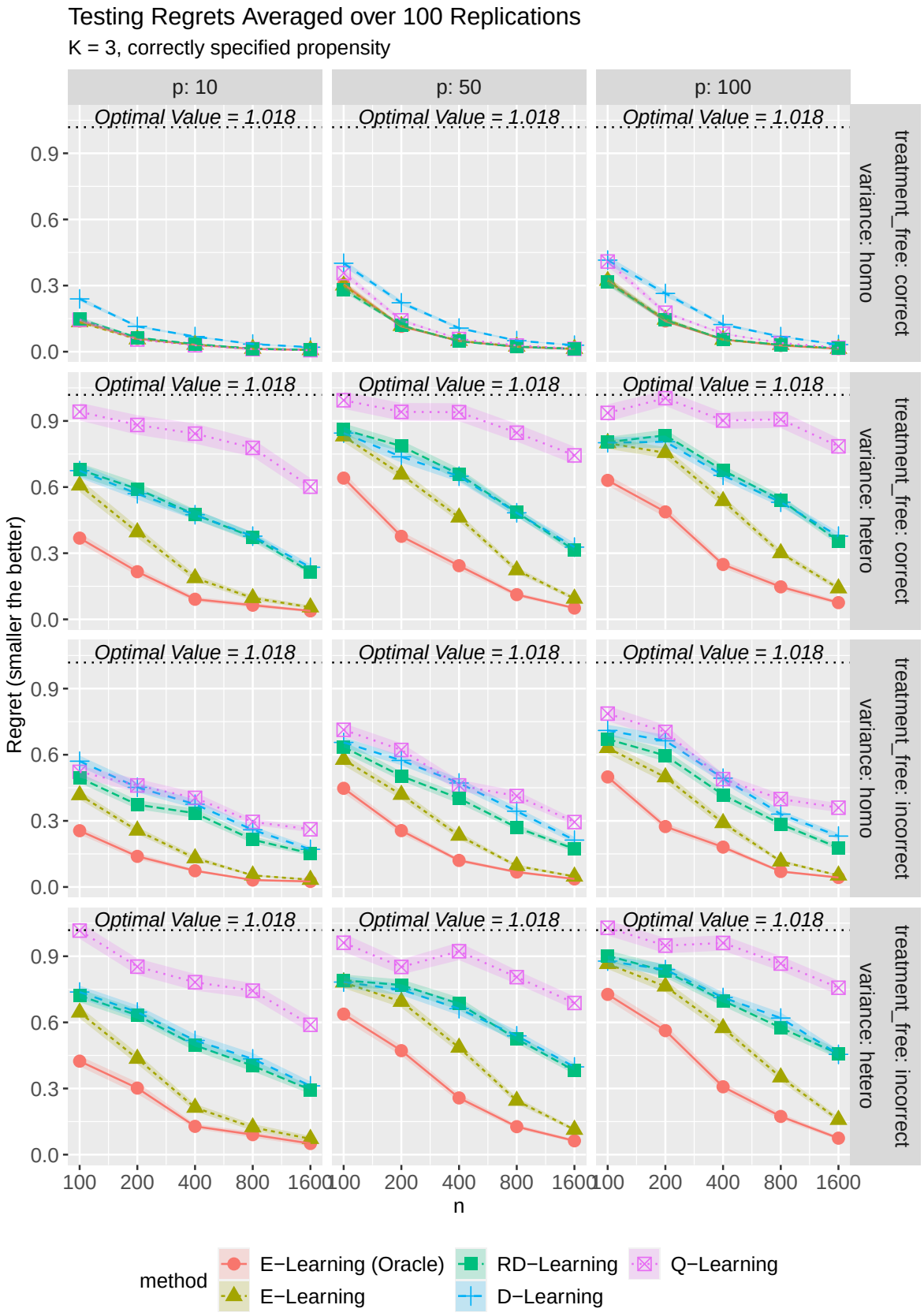


Figure 3.25: Testing regrets (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p \in \{10, 50, 100\}$, $K = 3$ and each of the model specification scenarios with correctly specified propensity score in Table 3.2. *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *E-Learning* corresponds to E-Learning with the working variance function estimated by regression forest.

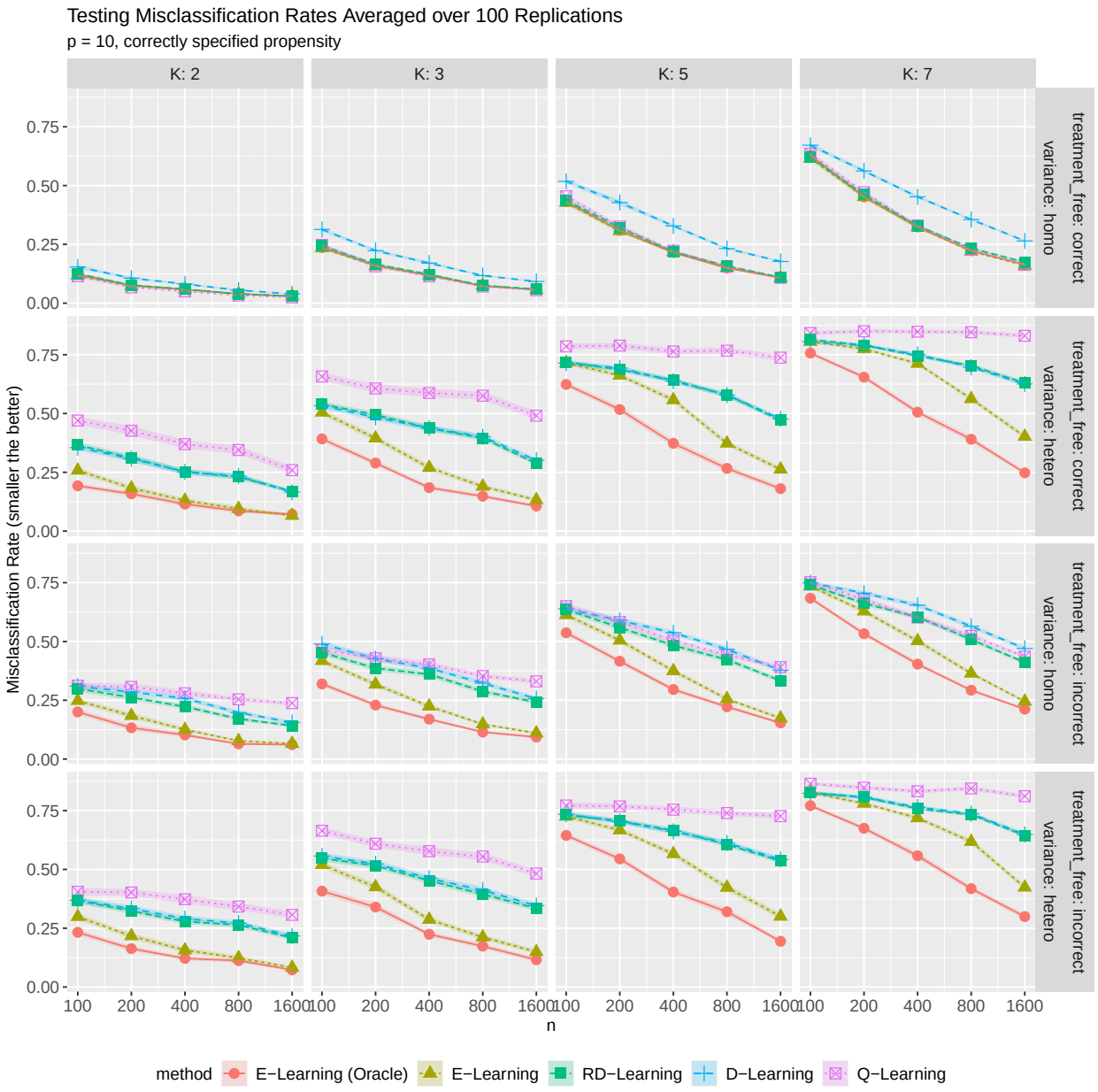


Figure 3.26: Testing misclassification rates (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p = 10$, $K \in \{2, 3, 5, 7\}$ and each of the model specification scenarios with correctly specified propensity score in Table 3.2. *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *E-Learning* corresponds to E-Learning with the working variance function estimated by regression forest.

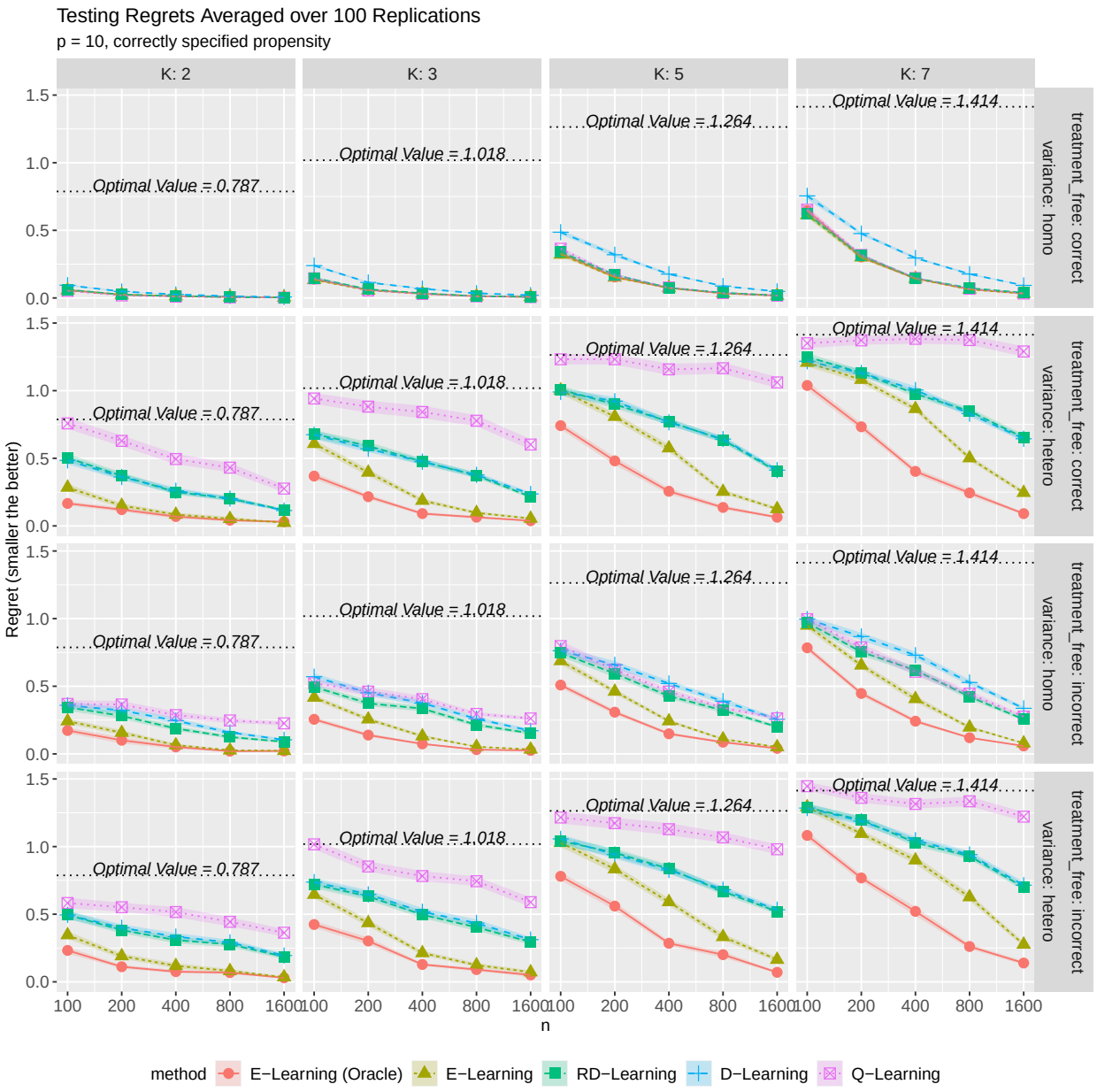


Figure 3.27: Testing regrets (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p = 10$, $K \in \{2, 3, 5, 7\}$ and each of the model specification scenarios with correctly specified propensity score in Table 3.2. *E-Learning (Oracle)* corresponds to E-Learning with the oracle working variance function, and *E-Learning* corresponds to E-Learning with the working variance function estimated by regression forest.

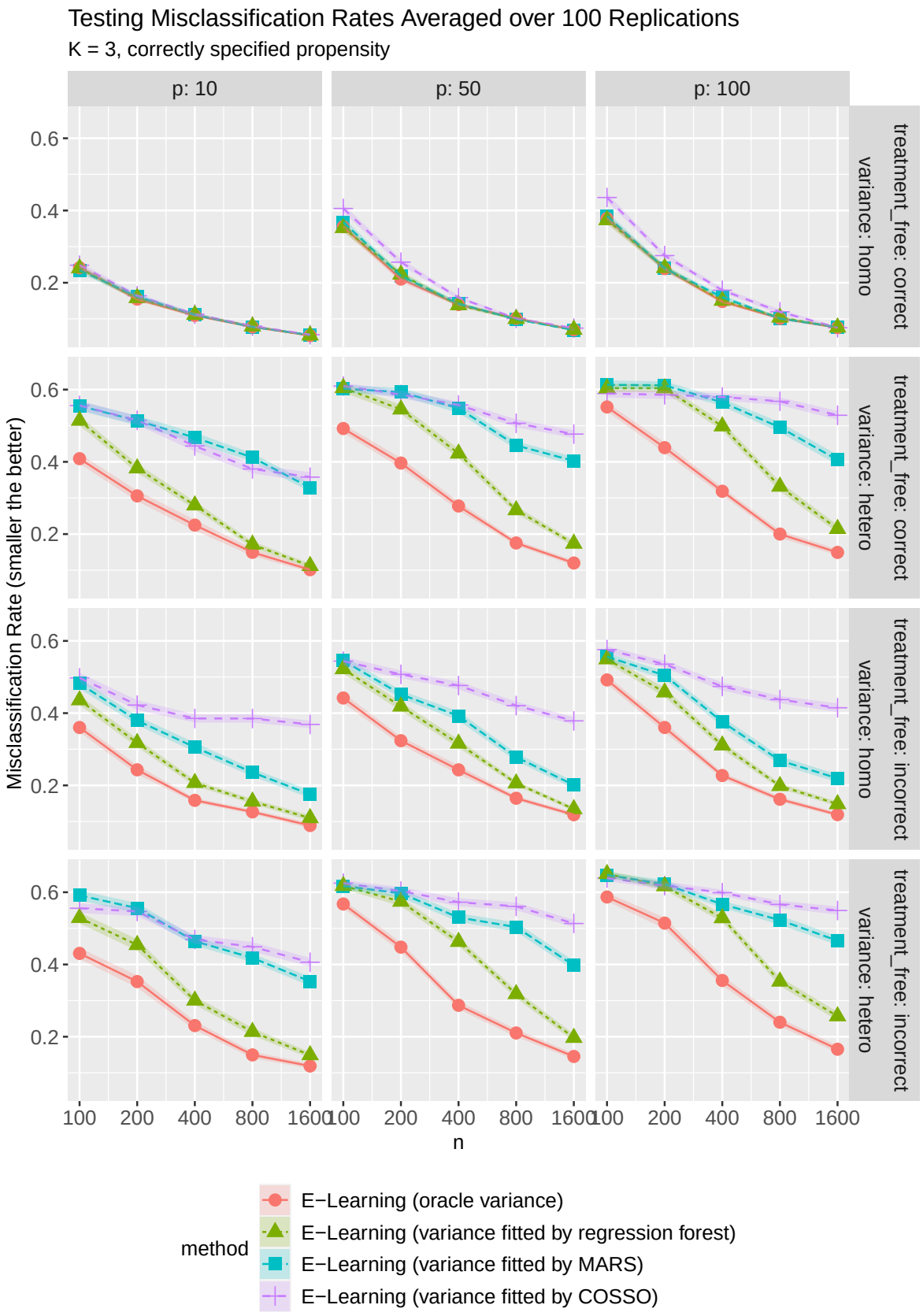


Figure 3.28: Testing misclassification rates (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p \in \{10, 50, 100\}$, $K = 3$ and each of the model specification scenarios with correctly specified propensity score in Table 3.2. The E-Learning procedures with different nonparametric estimation methods for variance function are compared.

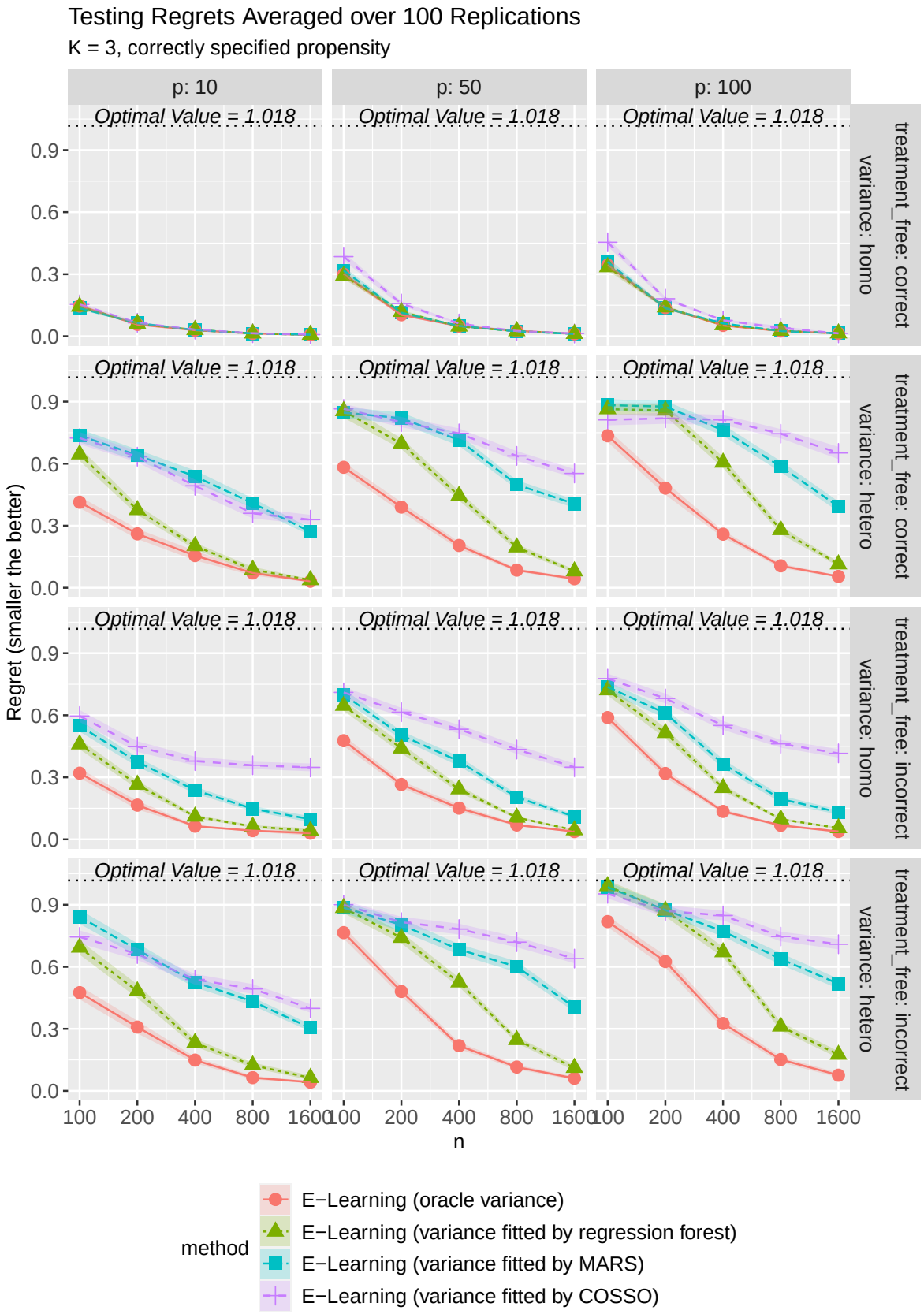


Figure 3.29: Testing regrets (smaller the better) for $n = \{100, 200, 400, 800, 1600\}$, $p \in \{10, 50, 100\}$, $K = 3$ and each of the model specification scenarios with correctly specified propensity score in Table 3.2. The E-Learning procedures with different nonparametric estimation methods for variance function are compared.

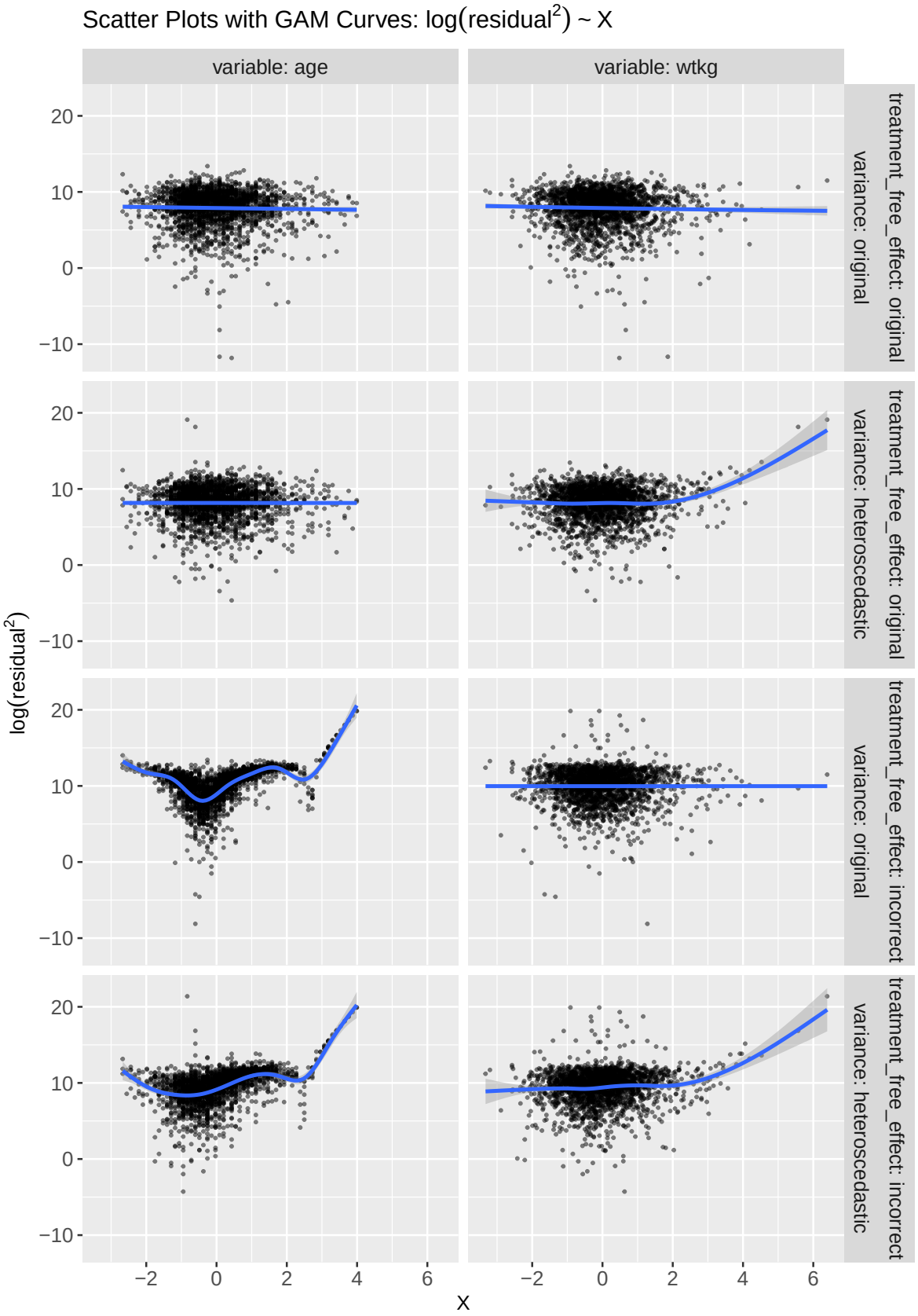


Figure 3.30: Residual plots with respect to `age` and `wtkg` on the ACTG175 dataset (Section 3.8.1). Curves are fitted by the *Generalized Additive Model (GAM)* of cubic spline. Residuals are computed from the fitted E-Learning on each modified dataset according to Table 3.4, and averaged over 10 replications.

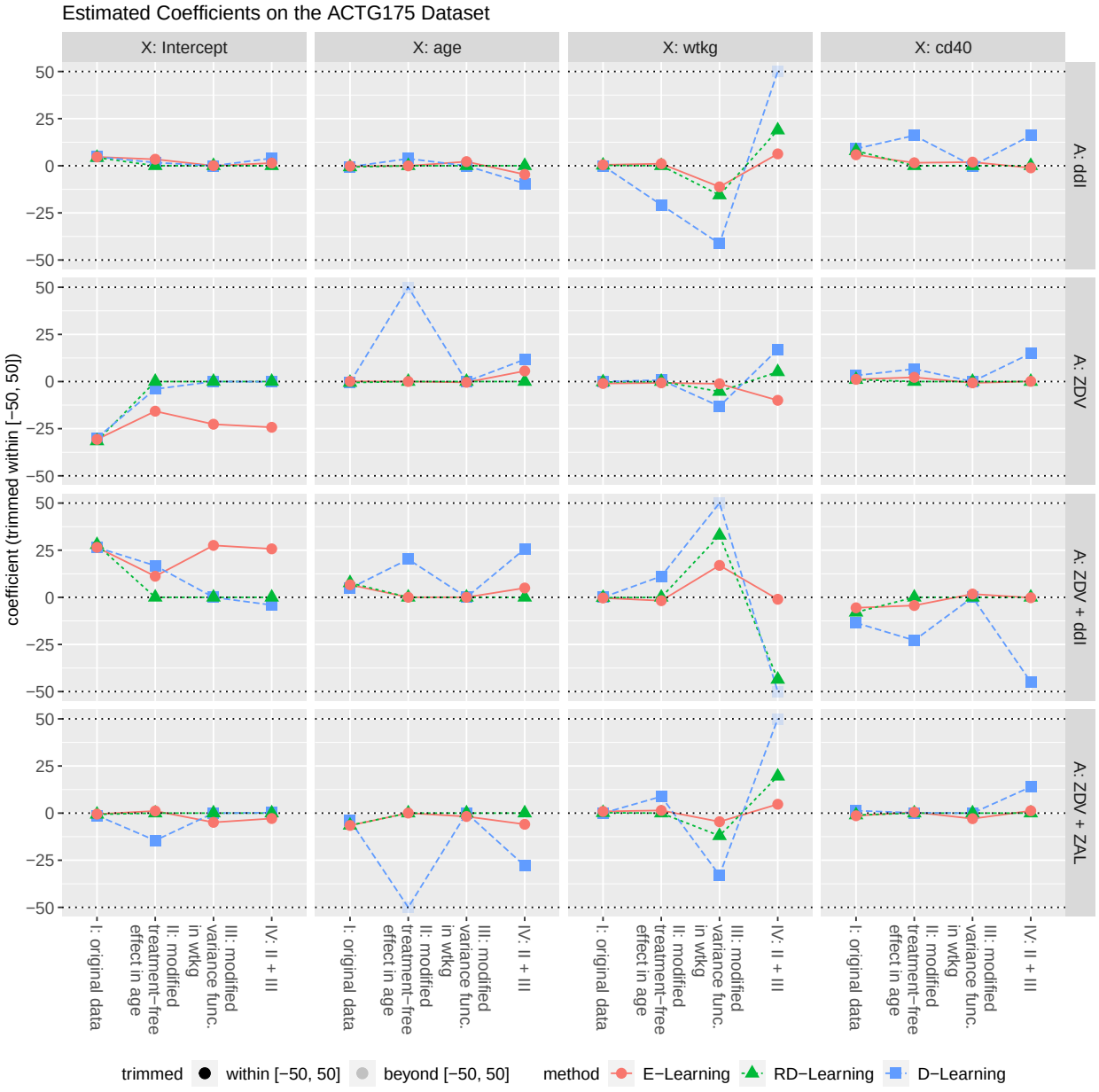


Figure 3.31: Fitted coefficients on each modified ACTG 175 dataset according to Table 3.4, and averaged over 10 replications.

CHAPTER 4
Efficient Learning for Optimal Dynamic Treatment Regimes

4.1 Introduction

In this chapter, we focus on the multi-stage decision problem. Among the existing approaches for estimating DTR, a model-based method can be more preferable if causal interpretations of the DTR are desired. In particular, the interaction effects in an SNMM can be interpreted as the stagewise individualized causal effects, which can be of primary interest when analyzing adaptive treatment strategies in randomized trials (Bembom and van der Laan, 2008). Moreover, under the correct model assumptions, the optimality of model-based methods can be established as the optimality of the parameter estimates, where the semiparametric efficiency theory can be applied (Robins, 1994, 2004). In practice, if the data generating process is close to the working semiparametric models, then model-based methods can generally enjoy superior performance (Shi et al., 2018a; Zhu et al., 2019; Ertefaie et al., 2021).

There remains gaps between the theory and practice for semiparametric efficient model-based methods. Specifically, Robins (2004) developed the G-Estimation procedures for an optimal DTR under the SNMM framework. The theoretical properties of G-Estimation can be applied to the analysis of other model-based methods due to the connections with Q-Learning (Chakraborty et al., 2010), A-Learning (Almirall et al., 2010) and dWOLS (Wallace and Moodie, 2015). The semiparametric efficiency can be established with the optimal estimating equations and correct nuisance models. However, the efficient G-estimating equations generally take a complicated form. The simplified versions under specific assumptions still require high-dimensional vector-valued nuisance functions, which can be hard to estimate in practice (Vansteelandt and Joffe, 2014). Moreover, there are conflicts in model specifications and the commonly used linear model can always misspecify the truth (Schulte et al., 2014). The residuals in the SNMM are generally heteroscedastic and positively correlated across stages. General practice can ignore these facts and implement a

suboptimal version of G-Estimation (Wallace et al., 2019). Last but not least, the algorithm based on backward recursive estimation is commonly used in practice, including Q-Learning (Watkins, 1989), recursive G-Estimation (Robins, 2004, Section 7.2), stagewise A-Learning (Shi et al., 2018a), and dWOLS (Wallace and Moodie, 2015). These methods do not solve the efficient G-estimating equations. Therefore, despite the well studied theoretical properties of G-Estimation, the rigorous semiparametric efficient procedure is rarely used in practice.

In this chapter, we first review the semiparametric theory of SNMM. The complicated semiparametric efficient score can be simplified if we consider a larger class of semiparametric estimates. Specifically, we propose a novel *Backward Change Point SNMM (BCP-SNMM)*, where there exists an unknown nuisance change point t_0, such that the data generating process is completely nonparametric for stages 1 to $t_0 - 1$, and then follows the SNMM from stage t_0 to the end. The BCP-SNMM can allow more robustness against model misspecifications. For any backward change point t_0 such that the SNMMs are violated before stage t_0, the properties of a *Regular and Asymptotically Linear (RAL)* estimate after stage t_0 remains, including consistency and semiparametric efficiency. The key observation is that an RAL estimate must be pivotal with respect to the nuisance change point t_0, and hence can only depend on the future model assumptions. In this way, many existing backward recursive estimates for the SNMM can be studied under the BCP-SNMM. We further propose *Dynamic Efficient Learning (DE-Learning)* that solves the semiparametric efficient estimating equations under the multiple treatment setting. In particular, DE-Learning enjoys the following properties:

1. (Optimality) Under correct model assumptions, DE-Learning is semiparametric efficient under the BCP-SNMM. In particular, it can handle the heteroscedasticity and cross-stage correlation with the efficient estimating equations. For general working treatment-free effect functions (possibly misspecified), the DE-Learning estimate achieves the smallest $\sqrt{n}$-asymptotic variance among a regular class of semiparametric estimates that allows misspecified treatment-free effects.

2. (Robustness) DE-Learning is stagewise doubly robust. For each stage, the corresponding estimate remains consistent when at most one of the treatment-free effect and propensity score is incorrect. Furthermore, DE-Learning is robust with respect to any backward model misspecifi-

cations. In particular, any violations of the SNMMs at stages $1, 2, \cdots, t-1$ do not affect the consistency and optimality of the stage-t estimate.

3. (Tractability) DE-Learning can be implemented in a backward stagewise manner. The nuisance functions required by DE-Learning are much fewer than that of the semiparametric efficient G-Estimation. More details on efficient G-Estimation are provided in Section 4.6.2.

This chapter makes the following contributions to the existing literature.

1. To our limited knowledge, this is the first work to establish the semiparametric efficiency of a backward stagewise estimate. The BCP-based model provides the framework for studying the optimality, robustness and cross-stage orthogonality of such an estimates.

2. DE-Learning is a tractable procedure for rigorous semiparametric efficient estimation. It can allow high-dimensional extensions with much fewer nuisance functions than G-Estimation.

3. In many practical scenarios, we show that the treatment-free effects in the SNMM can always be misspecified, and the stagewise heteroscedasticity generally exists. In presence of these challenges, DE-Learning remains optimal and enjoys significantly improved performance.

4. Under the BCP-SNMM, DE-Learning enjoys the cross-stage orthogonality, and hence can be less affected by the error propagation during backward stagewise estimation.

5. DE-Learning is developed for multiple treatments. We incorporate the equiangular coding in the semiparametric theory, which provides a tractable way of extending Robins (1994, 2004).

The rest of this chapter is organized as follows. In Section 4.2, we introduce the semiparametric models for the DTR problem. In particular, mathematical setups and notations are introduced in Section 4.2.1. The general SNMM and its semiparametric theories are discussed in Sections 4.2.2-4.2.4. The BCP-SNMM is proposed in Section 4.2.5. In Section 4.3, we propose DE-Learning and provide the implementation details. Simulation studies are provided in Section 4.4. General discussions and future work are given in Section 4.5. Additional discussions and technical proofs are provided in Section 4.6.

4.2 Semiparametric Models

In this section, we first consider the SNMM for the DTR problem. Then we propose the BCP-SNMM that can simplify the semiparametric theory of the standard SNMM.

4.2.1 Setup

Consider the observed data $\left\{\boldsymbol{O}_i := \left((\boldsymbol{X}_{it}, A_{it} : 1 \leqslant t \leqslant T), Y_i\right)\right\}_{i=1}^n$, where $\boldsymbol{X}_{it} \in \mathcal{X}_t \subseteq \mathbb{R}^d$ denotes the covariates at the t-th stage for the i-th subject, $A_{it} \in \mathcal{A} = \{1, 2, \cdots, K\}$ is the corresponding treatment assignment with K treatment options, and $Y_i \in \mathbb{R}$ is the corresponding observed outcome at the end. For $t = 1, \cdots, T$, we recursively define $\boldsymbol{H}_1 := \boldsymbol{X}_1 \in \mathcal{H}_1 = \mathcal{X}_1$, $\boldsymbol{H}_t := (\boldsymbol{H}_{t-1}^\intercal, A_{t-1}, \boldsymbol{X}_t^\intercal)^\intercal \in \mathcal{H}_t$ as the vector of pre-treatment historical information. We further introduce the stage-t *potential pre-treatment history* as $\boldsymbol{H}_t(\vec{\boldsymbol{a}}_1^{t-1})$, where $\vec{\boldsymbol{a}}_1^{t-1} := (a_1, a_2, \cdots, a_{t-1})^\intercal \in \mathcal{A}^{t-1}$ is a treatment assignment trajectory from stage 1 to stage $t-1$. At stage $t = 1$, we define $\vec{\boldsymbol{a}}_1^0 = \varnothing$, and the potential pre-treatment history $\boldsymbol{H}_1(\varnothing) = \boldsymbol{H}_1$. Analogously, $Y(\vec{\boldsymbol{a}}_1^T)$ is deifined as the *potential outcome* under the treatment assignment trajectory $\vec{\boldsymbol{a}}_1^T$. A *Dynamic Treatment Regime (DTR)* is defined as a sequence of mappings $\boldsymbol{d}_{1:T} = (d_1, d_2, \cdots, d_T) \in \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_T = \boldsymbol{\mathcal{D}}_{1:T}$, where $d_t \in \mathcal{D}_t := \{d_t : \mathcal{H}_t \to \mathcal{A}\}$ for $1 \leqslant t \leqslant T$. The *value function* of DTR is defined as

$$\mathcal{V}(\boldsymbol{d}_{1:T}) := \mathbb{E}\left\{Y(\vec{\boldsymbol{A}}_1^T)\middle| d_t\left[\boldsymbol{H}_t(\vec{\boldsymbol{A}}_1^{t-1})\right] = A_t \ (1 \leqslant t \leqslant T)\right\}.$$

Assuming that a larger outcome is better, the goal is to find the optimal DTR that maximizes the value function $\boldsymbol{d}_{1:T}^\star \in \operatorname{argmax}_{\boldsymbol{d}_{1:T} \in \boldsymbol{\mathcal{D}}_{1:T}} \mathcal{V}(\boldsymbol{d}_{1:T})$.

4.2.2 Structural Nested Mean Model (SNMM)

In order to identify $\mathcal{V}(\boldsymbol{d}_{1:T})$ from the observed data, we make the following identifiability conditions as in Robins (2004).

Assumption 4.1 (Consistency)**.** For $2 \leqslant t \leqslant T$, $\boldsymbol{H}_t = \boldsymbol{H}_t(\vec{\boldsymbol{A}}_1^{t-1})$; $Y = Y(\vec{\boldsymbol{A}}_1^T)$.

Assumption 4.2 (Sequential Ignorability)**.** For $1 \leqslant t \leqslant T$,

$$\left\{\left(\boldsymbol{H}_{t'}(\vec{\boldsymbol{a}}_1^{t'-1}) : 1 \leqslant t' \leqslant T\right), Y(\vec{\boldsymbol{a}}_1^T) : \vec{\boldsymbol{a}}_1^T \in \mathcal{A}^T\right\} \perp\!\!\!\perp A_t \middle| \boldsymbol{H}_t; \quad 1 \leqslant t \leqslant T.$$

Assumption 4.3 (Strict Overlap)**.** There exists some $\underline{p}_{\mathscr{A}} > 0$ such that $\mathbb{P}(A_t = k|\boldsymbol{H}_t) \geqslant \underline{p}_{\mathscr{A}}$ for $1 \leqslant k \leqslant K$ and $1 \leqslant t \leqslant T$.

Given Assumptions 4.1-4.3, the value function can be identified from the observed data by $\mathcal{V}(\boldsymbol{d}_{1:T}) = \mathbb{E}[Y|d_t(\boldsymbol{H}_t) = A_t \ (1 \leqslant t \leqslant T)]$. In order to obtain the optimal DTR in a stagewise manner, we introduce the *state value functions*, also known as the *V-functions*, as

$$\mathcal{V}_t(\boldsymbol{H}_t) := \max_{\boldsymbol{d}_{t:T} \in \boldsymbol{\mathcal{D}}_{t:T}} \mathbb{E}[Y|\boldsymbol{H}_t, d_u(\boldsymbol{H}_u) = A_u \ (t \leqslant u \leqslant T)]; \quad 1 \leqslant t \leqslant T. \tag{4.1}$$

Then $\{\mathcal{V}_t(\boldsymbol{H}_t)\}_{t=1}^T$ satisfy the following Bellman equations (Bellman, 1966):

$$\begin{aligned} \mathcal{V}_T(\boldsymbol{H}_T) &= \max_{1 \leqslant k \leqslant K} \underbrace{\mathbb{E}(Y|\boldsymbol{H}_T, A_T = k)}_{:=\mathcal{Q}_T(\boldsymbol{H}_T, k)}; \\ \mathcal{V}_t(\boldsymbol{H}_t) &= \max_{1 \leqslant k \leqslant K} \underbrace{\mathbb{E}\Big\{\mathcal{V}_{t+1}\left[\left(\boldsymbol{H}_t^\intercal, A_t, \boldsymbol{X}_{t+1}^\intercal\right)^\intercal\right] \Big| \boldsymbol{H}_t, A_t = k\Big\}}_{:=\mathcal{Q}_t(\boldsymbol{H}_t, k)}; \quad t = T-1, T-2, \cdots, 1. \end{aligned} \tag{4.2}$$

Here, $\{\mathcal{Q}_t(\boldsymbol{H}_t, A_t)\}_{t=1}^T$ are also known as the *state-action value functions* or the *Q-functions*. The optimal DTR $\boldsymbol{d}_{1:T}^\star$ satisfies that $d_t^\star(\boldsymbol{H}_t) \in \text{argmax}_{1 \leqslant k \leqslant K} \mathcal{Q}_t(\boldsymbol{H}_t, k)$ for $1 \leqslant t \leqslant T$.

The Q-functions can be interpreted as the conditional means for the *pseudo outcomes*, which are informally defined as $Y_t^* = Y_t^*(\vec{\boldsymbol{A}}_1^t) := Y(A_1, \cdots, A_t, d_{t+1}^\star, \cdots, d_T^\star)$ for $1 \leqslant t \leqslant T$, that is, the potential outcomes following the observed treatments up to stage t, while following the optimal treatments from stage $t+1$ to stage T. The precise definition is given in Section 4.6.1. The following Lemma 4.1 establishes the equivalence between the conditional mean of Y_t^* given $(\boldsymbol{H}_t, A_t)$ and the Q-function $\mathcal{Q}_t(\boldsymbol{H}_t, A_t)$.

Lemma 4.1 (Pseudo Outcome and Q-Functions)**.** *Consider the pseudo outcomes* $\{Y_t^*\}_{t=1}^T$ *in (4.16) in Section 4.6 and the Q-functions in* (4.2)*. Under Assumptions 4.1 and 4.2, we have*

$$\mathbb{E}(Y_t^*|\boldsymbol{H}_t, A_t) = \mathcal{Q}_t(\boldsymbol{H}_t, A_t); \quad 1 \leqslant t \leqslant T.$$

The proof of Lemma 4.1 is provided in Section 4.6. Lemma 4.1 implies that

$$\mathcal{V}_t(\boldsymbol{H}_t) = \max_{1 \leqslant k \leqslant K} \mathbb{E}(Y_t^*|\boldsymbol{H}_t, A_t = k); \quad d_t^\star(\boldsymbol{H}_t) \in \operatorname*{argmax}_{1 \leqslant k \leqslant K} \mathbb{E}(Y_t^*|\boldsymbol{H}_t, A_t = k).$$

The motivates us to study the following *Structural Nested Mean Model (SNMM)* (Robins, 1994, 2004). For $1 \leqslant t \leqslant T$, the stage-$t$ SNMM is defined as:

$$\begin{aligned} & Y_t^* = \mu_t(\boldsymbol{H}_t) + \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) + e_t^*; \\ \text{subject to } & \begin{cases} \sum_{k=1}^{K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) = 0; \\ \mathbb{E}(e_t^* | \boldsymbol{H}_t, A_t) = 0; \qquad \mathbb{E}[(e_t^*)^2] < +\infty. \end{cases} \end{aligned} \tag{SNMM}$$

For the rest of this chapter, we use $(\text{SNMM})_t$ to represent the stage-t SNMM, and $(\text{SNMM})_t(\boldsymbol{\beta}_t)$ to emphasize the true parameter $\boldsymbol{\beta}_t$. In $(\text{SNMM})_t$, $\mu_t(\boldsymbol{H}_t)$ is the stage-t *treatment-free effect*, and $\gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t)$ is the stage-t *history-treatment interaction effect*, also known as the "blip function" (Robins, 1994), which is parametrized by the p_t-dimensional parameter vector $\boldsymbol{\beta}_t \in \mathscr{B}_t \subseteq \mathbb{R}^{p_t}$. The sum-to-zero constraint $\sum_{k=1}^{K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) = 0$ is incorporated for identifiability. Since the stage-t Q-function is modeled in $(\text{SNMM})_t$ as $\mathcal{Q}_t(\boldsymbol{H}_t, A_t) = \mu_t(\boldsymbol{H}_t) + \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t)$, the induced stage-$t$ optimal decision rule becomes $d_t^\star(\boldsymbol{H}_t) \in \operatorname{argmax}_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t)$.

In the following Theorem 4.2, we further show that maximizing the value function can be directly related to finding good estimates of the interaction effects $\{\gamma_t(\boldsymbol{H}_t, A_t)\}_{t=1}^T$ in $(\text{SNMM})_1^T$.

Theorem 4.2 (Estimation and Regret Bound)**.** *Consider Model* $(\text{SNMM})_{t=1}^T$*. For* $1 \leqslant t \leqslant T$*, let* $\widehat{\gamma}_{t,n}(\boldsymbol{X}_t, A_t)$ *be an estimates of* $\gamma_t(\boldsymbol{X}_t, A_t)$*,* $\widehat{d}_{t,n}(\boldsymbol{H}_t) \in \operatorname{argmax}_{1 \leqslant k \leqslant K} \widehat{\gamma}_{t,n}(\boldsymbol{H}_t, k)$*, and* $d_t^\star(\boldsymbol{H}_t) \in \operatorname{argmax}_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k)$ *. Then*

$$\mathcal{V}(\boldsymbol{d}^\star) - \mathcal{V}(\widehat{\boldsymbol{d}}_n) \leqslant 2 \sum_{t=1}^{T} \max_{1 \leqslant k \leqslant K} \mathbb{E} \left| \widehat{\gamma}_{t,n}(\boldsymbol{H}_t, k) - \gamma_t(\boldsymbol{H}_t, k) \right|.$$

Here, $\{\widehat{\gamma}_{t,n}\}_{t=1}^T$ *are fixed and* $\mathbb{E}$ *takes expectation over* $\{\boldsymbol{H}_t\}_{t=1}^T$*.*

The proof is similar to Murphy (2005, Lemma 2), and is included in Section 4.6. Theorem 4.2 implies that minimizing the estimation error of γ can also minimize the regret. In this chapter, we focus on finding efficient estimates of the parametric interaction effects $\{\gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t)\}_{t=1}^T$.

4.2.3 Identification

Model $(\text{SNMM})_t$ is "structural" since the stage-t pseudo outcome Y_t^* is not directly observed from the data except for $t = T$. It is "nested" because $(\text{SNMM})_t$ depends on $(\text{SNMM})_{t+1}^T := \bigcap_{u=t+1}^T (\text{SNMM})_u$. Specifically, assume that Model $(\text{SNMM})_{t+1}^T(\boldsymbol{\beta}_{(t+1):T}) = \bigcap_{u=t+1}^T (\text{SNMM})_u(\boldsymbol{\beta}_u)$ is known. Define the stagewise g-*outcomes* from the observed data $\boldsymbol{O}$ as

$$Y_T^{(\text{g})} := Y; \quad Y_t^{(\text{g})} := Y - \sum_{u=t+1}^{T} \left\{ \gamma_u(\boldsymbol{H}_u, A_u; \boldsymbol{\beta}_u) - \max_{1 \leqslant k \leqslant K} \gamma_u(\boldsymbol{H}_u, k; \boldsymbol{\beta}_u) \right\}; \quad t = T-1, \cdots, 1. \tag{4.3}$$

The following Lemma 4.3 connects the pseudo outcome with the g-outcome.

Lemma 4.3 (Pseudo Outcome Identification)**.** *Fix* $1 \leqslant t \leqslant T$. *Consider the stage-t pseudo outcome* Y_t^* *in* (4.16) *and the* g-*outcome* $Y_t^{(\text{g})}$ *in* (4.3). *Then under Model* $(\text{SNMM})_{t+1}^T(\boldsymbol{\beta}_{(t+1):T})$, *Assumptions 4.1 and 4.2, we have*

$$\mathbb{E}_{\boldsymbol{\beta}_{(t+1):T}}(Y_t^{(\text{g})} | \boldsymbol{H}_t, A_t) = \mathbb{E}_{\boldsymbol{\beta}_{(t+1):T}}(Y_t^* | \boldsymbol{H}_t, A_t).$$

Here, $\mathbb{E}_{\boldsymbol{\beta}_{(t+1):T}}$ *denotes the expectation under the data generating process of Model* $(\text{SNMM})_{t+1}^T(\boldsymbol{\beta}_{(t+1):T})$.

In Lemma 4.3, the stage-t g-outcome $Y_t^{(\text{g})}$ can be obtained from the observed data $\{\boldsymbol{O}_i\}_{i=1}^n$ and the true parameters $\{\boldsymbol{\beta}_u\}_{u=t+1}^T$ in the subsequent-stage models. In this way, $(\text{SNMM})_t$ is identified from the g-outcome $Y_t^{(\text{g})}$, and the identification depends on $(\text{SNMM})_{t+1}^T$. We point out that $(\text{SNMM})_t$ does not depend on the previous-stage models $(\text{SNMM})_1^{t-1} := \bigcap_{s=1}^{t-1} (\text{SNMM})_s$. Therefore, the estimation of the stage-t parameter $\boldsymbol{\beta}_t$ can also be free from the model assumptions of $(\text{SNMM})_1^{t-1}$. This can provide the potential for robustness with respect to backward model misspecifications.

As a corollary of Lemma 4.3, Model $(\text{SNMM})_1^T$ can be characterized by some moment conditions on the observed data $\boldsymbol{O}$. Define the stage-t g-*residual* from the observed data $\boldsymbol{O}$ as:

$$e_t^{(\text{g})} = e_t^{(\text{g})}(\boldsymbol{\beta}_{t:T}; \mu_t) := Y - \sum_{u=t+1}^{T} \left\{ \gamma_u(\boldsymbol{H}_u, A_u; \boldsymbol{\beta}_u) - \max_{1 \leqslant k \leqslant K} \gamma_u(\boldsymbol{H}_u, k; \boldsymbol{\beta}_u) \right\} - \mu_t(\boldsymbol{H}_t) - \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t). \tag{4.4}$$

In the following Theorem 4.4, we establish the moment conditions for $(\text{SNMM})_1^T$ in terms of the observed data.

Theorem 4.4 (Characterizing Moment Conditions)**.** *Consider Model* $(\text{SNMM})_1^T(\boldsymbol{\beta}_{1:T})$ *with the true parameter* $\boldsymbol{\beta}_{1:T} = (\boldsymbol{\beta}_1^\intercal, \cdots, \boldsymbol{\beta}_T^\intercal)^\intercal \in \mathscr{B}_1 \times \mathscr{B}_2 \times \cdots \times \mathscr{B}_T = \mathscr{B}_{1:T}$*. Let* $\{e_t^{(\text{g})}(\breve{\boldsymbol{\beta}}_{t:T}; \breve{\mu}_t)\}_{t=1}^T$ *be the* g*-residuals in* (4.4) *based on the working parameter* $\breve{\boldsymbol{\beta}}_{1:T} \in \mathscr{B}_{1:T}$ *and the working treatment-free effect function* $\breve{\mu}_t(\boldsymbol{H}_t)$*. Then* $\breve{\boldsymbol{\beta}}_{1:T} = \boldsymbol{\beta}_{1:T}$ *if and only if*

$$\mathbb{E}_{\boldsymbol{\beta}_{t:T}}[e_t^{(\text{g})}(\breve{\boldsymbol{\beta}}_{t:T}; \breve{\mu}_t)|\boldsymbol{H}_t, A_t] = \mathbb{E}_{\boldsymbol{\beta}_{t:T}}[e_t^{(\text{g})}(\breve{\boldsymbol{\beta}}_{t:T}; \breve{\mu}_t)|\boldsymbol{H}_t]; \quad 1 \leqslant t \leqslant T. \tag{4.5}$$

Here, $\mathbb{E}_{\boldsymbol{\beta}_{t:T}}$ *denotes the expectation under the data generating process of Model* $(\text{SNMM})_t^T(\boldsymbol{\beta}_{t:T})$*.*

Robins (2004, Theorem 3.2 (ii)) also used similar moment conditions as our (4.5). These moment conditions define the nuisance tangent spaces in Lemma 4.6 in Section 4.2.4. In the following Corollary 4.5, we further obtain the equivalent data generating process implied by (4.5). Different from the stagewise models in $(\text{SNMM})_1^T$, Corollary 4.5 provides cross-stage the data generating process.

Corollary 4.5 (Equivalent Data Generating Process)**.** *Under Assumptions 4.1 and 4.2, Model* $(\text{SNMM})_1^T(\boldsymbol{\beta}_{1:T})$ *is equivalent to the following data generating process:*

$$\begin{aligned} & Y = \mathcal{V}_0 + \sum_{t=1}^T \Delta\mathcal{M}_t(\boldsymbol{H}_t) - \sum_{t=1}^T \left\{ \max_{1\leqslant k\leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) - \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) \right\} + \Delta\mathcal{M}_{T+1}(\boldsymbol{H}_{T+1}), \\ \textit{subject to} \quad & \begin{cases} \displaystyle\sum_{k=1}^K \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) = 0; & 1 \leqslant t \leqslant T; \\ \mathbb{E}[\Delta\mathcal{M}_{t+1}(\boldsymbol{H}_{t+1})|\boldsymbol{H}_t, A_t] = 0; \quad \mathbb{E}[\Delta\mathcal{M}_{t+1}(\boldsymbol{H}_{t+1})^2] < +\infty; & 0 \leqslant t \leqslant T, \end{cases} \end{aligned} \tag{4.6}$$

where $\boldsymbol{H}_0 = A_0 = \varnothing$*,* $\boldsymbol{H}_t = \left(\boldsymbol{H}_{t-1}^\intercal, A_{t-1}, \boldsymbol{X}_t^\intercal\right)^\intercal$ $(1 \leqslant t \leqslant T)$*,* $\boldsymbol{H}_{T+1} = (\boldsymbol{H}_T^\intercal, A_T, Y)^\intercal$*. In particular, the equivalent stage-t mean model on the* g*-outcome* (4.3) *is* $Y_t^{(\text{g})} = \mu_t(\boldsymbol{H}_t) + \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) + e_t^{(\text{g})}$*, where* $\mu_t(\boldsymbol{H}_t) = \mathcal{V}_t(\boldsymbol{H}_t) - \max_{1\leqslant k\leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t)$*,* $e_t^{(\text{g})} = \sum_{u=t+1}^{T+1} \Delta\mathcal{M}_u(\boldsymbol{H}_u)$*, and* $\mathcal{V}_t(\boldsymbol{H}_t) = \mathcal{V}_0 + \sum_{s=1}^t \Delta\mathcal{M}_s(\boldsymbol{H}_s) - \sum_{s=1}^{t-1} \{\max_{1\leqslant k\leqslant K} \gamma_s(\boldsymbol{H}_s, k; \boldsymbol{\beta}_s) - \gamma_s(\boldsymbol{H}_s, A_s; \boldsymbol{\beta}_s)\}$*.*

The stagewise g-residuals $\{e_t^{(\text{g})}\}_{t=1}^T$ are generally heteroscedastic and positively correlated:

$$\mathbb{E}\left(e_s^{(\text{g})} e_t^{(\text{g})} \middle| \boldsymbol{H}_t, A_t\right) = \mathbb{E}\left(e_t^{(\text{g})2} \middle| \boldsymbol{H}_t, A_t\right) = \sum_{u=t+1}^{T+1} \mathbb{E}[\Delta\mathcal{M}_u(\boldsymbol{H}_u)^2|\boldsymbol{H}_t, A_t]; \quad 1 \leqslant s \leqslant t \leqslant T. \tag{4.7}$$

Moreover, the stage-t-treatment-free effect $\mu_t(\boldsymbol{H}_t)$ consists of $\{\max_{1\leqslant k\leqslant K}\gamma_s(\boldsymbol{H}_s,k;\boldsymbol{\beta}_s)\}_{s=1}^t$. If $\gamma(\boldsymbol{H}_s,k;\boldsymbol{\beta}_s)$ is modeled as the linear model $\boldsymbol{H}_s^\intercal\boldsymbol{\beta}_{s,k}$ for $1\leqslant k\leqslant K$ and $1\leqslant s\leqslant t$, then $\mu_t(\boldsymbol{H}_t)$ is nonlinear in $\boldsymbol{H}_t$. This implies that existing strategies based on linear working models (Almirall et al., 2010; Henderson et al., 2010; Wallace and Moodie, 2015; Shi et al., 2018a; Zhu et al., 2019; Wallace et al., 2019) may always misspecify the treatment-free effect models (Laber et al., 2014a; Schulte et al., 2014).

Murphy (2003, Equation (12)) also obtained the same representation as our (4.6). In particular, $\{\Delta\mathcal{M}_t(\boldsymbol{H}_t)\}_{t=1}^{T+1}$ is a $(\boldsymbol{H}_t,A_t)_{t=1}^T$-martingale-difference sequence, where $\Delta\mathcal{M}_t(\boldsymbol{H}_t)=\mathcal{V}_t(\boldsymbol{H}_t)-\mathbb{E}[\mathcal{V}_t(\boldsymbol{H}_t)|\boldsymbol{H}_{t-1},A_{t-1}]$. The nonparametric function $\Delta\mathcal{M}_t:\mathcal{H}_t\to\mathbb{R}$ is part of the treatment-free effects $\{\mu_u(\boldsymbol{H}_u)\}_{u=t}^T$. The predicted quadratic variation $\mathbb{E}[\Delta\mathcal{M}_t(\boldsymbol{H}_t)^2|\boldsymbol{H}_s,A_s]$ is part of the variance function $\mathbb{E}(e_s^{(\mathrm{g})2}|\boldsymbol{H}_s,A_s)$ for $1\leqslant s\leqslant t-1$.

Almirall et al. (2010); Henderson et al. (2010) utilized this cross-stage data generating process representation, and estimated $\mathcal{V}_0$, $\{\Delta\mathcal{M}_t(\boldsymbol{H}_t)\}_{t=1}^T$ and $\{\gamma_t(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)\}_{t=1}^T$ simultaneously. However, $\{\Delta\mathcal{M}_t(\boldsymbol{H}_t)\}_{t=1}^T$ are nuisance components that can be vulnerable to model misspecifications.

Q-Learning (Watkins, 1989) utilized the nuisance components in a different way. Specifically, the following stagewise q-*outcomes* are considered:

$$Y_T^{(\mathrm{q})}:=Y;\quad Y_t^{(\mathrm{q})}:=\mu_{t+1}(\boldsymbol{H}_{t+1})+\max_{1\leqslant k\leqslant K}\gamma_{t+1}(\boldsymbol{H}_{t+1},k;\boldsymbol{\beta}_{t+1});\quad t=T-1,\cdots,1. \tag{4.8}$$

If the treatment-free effects $\{\mu_t(\boldsymbol{H}_t)\}_{t=2}^T$ are correctly specified, then the stage-t mean model on the q-outcome is $Y_t^{(\mathrm{q})}=\mu_t(\boldsymbol{H}_t)+\gamma_t(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)+e_t^{(\mathrm{q})}$, where the stage-$t$ q-*residual* is $e_t^{(\mathrm{q})}=\Delta\mathcal{M}_{t+1}(\boldsymbol{H}_{t+1})$. Here, different from the representation of g-residual in Corollary 4.5, the q-residual $e_t^{(\mathrm{q})}$ consists of fewer martingale-difference terms than the g-residual $e_t^{(\mathrm{g})}$. Therefore, Q-Learning under correct model assumptions can enjoy higher efficiency than methods based on the g-outcomes (Schulte et al., 2014). However, Q-Learning can also be vulnerable to model misspecifications. In particular, the stage-t q-outcome $Y_t^{(\mathrm{q})}$ can heavily depend on the stage-$(t+1)$ mean model, especially the treatment-free effect function $\mu_{t+1}(\boldsymbol{H}_{t+1})$.

To conclude this section, we use the following example to demonstrate (4.7) and the nonlinear treatment-free effects.

Example 4.1. Consider the data $(X_1, A_1, X_2, A_2, X_3, A_3, Y)$ as follows:

$$A_1, A_2, A_3, Z_1, Z_2, Z_3 \overset{\text{i.i.d.}}{\sim} 2 \times \mathsf{Bernoulli}(1/2) - 1;$$

$$X_1 = Z_1; \quad X_2 = Z_2 \mathbb{1}(X_1 = A_1 = 1); \quad X_3 = Z_3 \mathbb{1}(X_2 = A_2 = 1); \quad Y = \sum_{t=1}^{3} X_t - \sum_{t=1}^{3} (|X_t| - A_t X_t).$$

The stagewise pre-treatment histories are $\boldsymbol{H}_3 = (X_1, A_1, X_2, A_2, X_3)^\intercal$, $\boldsymbol{H}_2 = (X_1, A_1, X_2)^\intercal$, and $\boldsymbol{H}_1 = X_1$. Compared with (4.6), the martingale-difference sequence is $\{X_t\}_{t=1}^3$.

- The stage-3 mean model is $Y_3^{(\mathrm{g})} = Y = \big[\sum_{t=1}^{2}(X_t - |X_t| + A_t X_t) + X_3 - |X_3|\big] + (A_3 X_3)$, where the treatment-free effect is $\mu_3(\boldsymbol{H}_3) = \sum_{t=1}^{2}(X_t - |X_t| + A_t X_t) + X_3 - |X_3|$, the interaction effect is $\gamma_3(\boldsymbol{H}_3, A_3) = A_3 X_3$, and the g-residual is 0.

- The stage-2 mean model is $Y_2^{(\mathrm{g})} = Y - \big\{\gamma_3(\boldsymbol{H}_3, A_3) - \max_{a\in\{-1,1\}} \gamma_3(\boldsymbol{H}_3, a)\big\} = Y - (A_3 X_3 - |X_3|) = [(X_1 - |X_1| + A_1 X_1) + X_2 - |X_2|] + (A_2 X_2) + (X_3);$, where the treatment-free effect is $\mu_2(\boldsymbol{H}_2) = (X_1 - |X_1| + A_1 X_1) + X_2 - |X_2|$, the interaction effect is $\gamma_2(\boldsymbol{H}_2, A_2) = A_2 X_2$, and the g-residual is $e_2^{(\mathrm{g})} = X_3$.

- The stage-1 mean model is $Y_1^{(\mathrm{g})} = Y - \sum_{u=2}^{3} \big\{\gamma_u(\boldsymbol{H}_u, A_u) - \max_{a\in\{-1,1\}} \gamma_u(\boldsymbol{H}_u, a)\big\} = Y - (A_2 X_2 - |X_2|) - (A_3 X_3 - |X_3|) = (X_1 - |X_1|) + (A_1 X_1) + (X_2 + X_3)$, where the treatment-free effect is $\mu_1(\boldsymbol{H}_1) = X_1 - |X_1|$, the interaction effect is $\gamma_1(\boldsymbol{H}_1, A_1) = A_1 X_1$, and the g-residual is $e_1^{(\mathrm{g})} = X_2 + X_3$.

It can be clear that $\mu_3(\boldsymbol{H}_3)$, $\mu_2(\boldsymbol{H}_2)$ must be nonlinear functions. For the g-residuals, we have $\mathbb{E}(e_2^{(\mathrm{g})2}|\boldsymbol{H}_2, A_2) = \mathbb{E}(X_3^2|X_1, A_1, X_2, A_2) = \mathbb{1}(X_2 = A_2 = 1)$ and $\mathbb{E}(e_1^{(\mathrm{g})2}|\boldsymbol{H}_1, A_1) = \mathbb{E}[(X_2 + X_3)^2|X_1, A_1] = [\mathbb{E}(Z_2^2)\mathbb{1}(X_1 = 1) + \mathbb{E}(Z_3^2)\mathbb{P}(X_2 = A_2 = 1|X_1, A_1)]\mathbb{1}(A_1 = 1) = (5/4)\mathbb{1}(X_1 = A_1 = 1)$. That is, both $e_1^{(\mathrm{g})}$ and $e_2^{(\mathrm{g})}$ are heteroscedastic. Moreover, $\mathbb{E}(e_1^{(\mathrm{g})} e_2^{(\mathrm{g})}|\boldsymbol{H}_2, A_2) = \mathbb{E}[(X_2 + X_3)X_3|X_1, A_1, X_2, A_2] = \mathbb{E}(Z_3^2)\mathbb{1}(X_2 = A_2 = 1) = \mathbb{1}(X_2 = A_2 = 1)$. Then $\mathbb{E}(e_1^{(\mathrm{g})} e_2^{(\mathrm{g})}) = \mathbb{P}(X_2 = A_2 = 1) = 1/16$, which suggests that $e_1^{(\mathrm{g})}$ and $e_2^{(\mathrm{g})}$ are positively correlated.

4.2.4 Semiparametric Theory

Our next goal is to further study the semiparametric efficient estimation of $(\mathrm{SNMM})_1^T$. We first review several concepts in semiparametric inference. Consider an RAL estimate $\widehat{\boldsymbol{\beta}}_{1:T,n}$ of $\boldsymbol{\beta}_{1:T}$ with the $\sqrt{n}$-asymptotic linear representation: $\widehat{\boldsymbol{\beta}}_{1:T,n} - \boldsymbol{\beta}_{1:T} = \mathbb{E}_n(\mathsf{IF}) + o_{\mathbb{P}}(n^{-1/2})$. Here, IF is the *Influence Function (IF)* of $\widehat{\boldsymbol{\beta}}_{1:T,n}$, and $\mathbb{E}_n$ is the empirical average. Then under regularity conditions,

$\lim_{n\to\infty} n\mathsf{Var}(\widehat{\boldsymbol{\beta}}_{1:T,n}) = \mathbb{E}(\mathsf{IF}^{\otimes 2})$ where $\boldsymbol{a}^{\otimes 2} := \boldsymbol{a}\boldsymbol{a}^\intercal$. The goal is to find the semiparametric *Efficient IF (EIF)* with the smallest $\mathbb{E}(\mathsf{EIF}^{\otimes 2})$ among that of all RAL estimates. By Tsiatis (2007, Theorem 4.2 (ii)), $\mathsf{IF} \in \Lambda_{1:T}^{\perp}$, where $\Lambda_{1:T}$ is the *nuisance tangent space* of $(\text{SNMM})_1^T$, which can be characterized by the moment conditions (4.5) in Theorem 4.4. Then it remains to characterize the IFs in $\Lambda_{1:T}^{\perp}$ and choose the EIF with the minimal $\mathbb{E}(\mathsf{EIF}^{\otimes 2})$.

We first derive the nuisance tangent space following Robins (1994, Theorem 8).

Lemma 4.6 (Nuisance Tangent Spaces)**.** *Consider* $(\text{SNMM})_{t=1}^T$ *and the g-residuals* $\{e_t^{(\text{g})}\}_{t=1}^T$ *in (4.4). Define* $p := \sum_{t=1}^T p_t$ *and* $\mathcal{G} := \{\boldsymbol{g}(\boldsymbol{O}) \mid \boldsymbol{g} : \mathcal{O} \to \mathbb{R}^p,\ \mathbb{E}[\boldsymbol{g}(\boldsymbol{O})] = \boldsymbol{0},\ \mathbb{E}\|\boldsymbol{g}(\boldsymbol{O})\|_2^2 < +\infty\}$*, which is equipped with the norm* $\|\cdot\| := (\mathbb{E}\|\cdot\|_2^2)^{1/2}$*. Then the nuisance tangent space is* $\Lambda_{1:T} := \bigcap_{t=1}^T \Lambda_t$*, where*

$$\Lambda_t = \left\{\boldsymbol{G} \in \mathcal{G} : \mathbb{E}\left(\boldsymbol{G}e_t^{(\text{g})}\middle|\boldsymbol{H}_t, A_t\right) = \mathbb{E}\left(\boldsymbol{G}e_t^{(\text{g})}\middle|\boldsymbol{H}_t\right)\right\};\quad 1 \leqslant t \leqslant T.$$

By Tsiatis (2007, Theorem 4.3), the IF of an RAL estimate belongs to $\Lambda_{1:T}^{\perp} = \overline{\mathsf{span}}\left\{\Lambda_u^{\perp} : t \leqslant u \leqslant T\right\}$, where $\overline{\mathsf{span}}$ represents the closed linear span. Therefore, it suffices to study $\Lambda_t^{\perp}$ for each $1 \leqslant t \leqslant T$. Notice that the moment restriction in Lemma 4.6 is equivalent to

$$\mathbb{E}\left(\boldsymbol{G}e_t^{(\text{g})}\middle|\boldsymbol{H}_t, A_t = 1\right) = \mathbb{E}\left(\boldsymbol{G}e_t^{(\text{g})}\middle|\boldsymbol{H}_t, A_t = 2\right) = \cdots = \mathbb{E}\left(\boldsymbol{G}e_t^{(\text{g})}\middle|\boldsymbol{H}_t, A_t = K\right).$$

Then we can introduce a set of coding vectors $\{\boldsymbol{\omega}_k\}_{k=1}^K \subseteq \mathbb{R}^{K-1}$, such that $\sum_{k=1}^K c_k\boldsymbol{\omega}_k = \boldsymbol{0}$ if and only if $c_1 = c_2 = \cdots = c_K$. Equivalently, we can let $\Omega := \sqrt{1-1/K}[\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_K]^\intercal \in \mathbb{R}^{K\times(K-1)}$, and require that $(1/\sqrt{K})\mathbf{1}_{K\times 1}$ is the only left singular vector corresponding to the singular value 0 of Ω. In the following Lemma 4.7, we show that any coding vectors satisfying such a requirement are equiangular up to normalization.

Lemma 4.7 (Equiangularity)**.** *Let* $\Omega := \sqrt{1-1/K}[\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \cdots, \boldsymbol{\omega}_K]^\intercal \in \mathbb{R}^{K\times(K-1)}$ *such that* $(1/\sqrt{K})\mathbf{1}_{K\times 1}$ *is the only left singular vector corresponding to the singular value* 0*. Then* $\{(\Omega^\intercal\Omega)^{-1/2}\boldsymbol{\omega}_k\}_{k=1}^K$ *are equiangular.*

The equiangular coding representation in Zhang and Liu (2014); Qi et al. (2020); Zhang et al. (2020) is an example that satisfies Lemma 4.7. The equiangular coding vectors $\{\boldsymbol{\omega}_k\}_{K=1}^K$ can be useful to define the following $\mathbb{R}^{K-1}$-valued decision function associated with the interaction effect.

Lemma 4.8 (Angle-Based Decision Function)**.** *Consider* $(\text{SNMM})_t$*. For the coding vectors* $\{\boldsymbol{\omega}_k\}_{k=1}^K \subseteq \mathbb{R}^{K-1}$ *as in Lemma 4.7, define an* $\mathbb{R}^{K-1}$*-valued decision function as* $\vec{\boldsymbol{f}}_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t) :=$

$(\Omega^\intercal\Omega)^{-1}\sum_{k=1}^K \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t)\boldsymbol{\omega}_k$. *Then*

$$\gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) = \left(1 - \frac{1}{K}\right)\langle \boldsymbol{\omega}_k, \vec{\boldsymbol{f}}_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t)\rangle; \quad 1 \leqslant k \leqslant K.$$

Moreover, the stage-t optimal decision rule is given by

$$d_t^\star(\boldsymbol{H}_t; \boldsymbol{\beta}_t) \in \underset{1\leqslant k\leqslant K}{\operatorname{argmax}}\langle \boldsymbol{\omega}_k, \vec{\boldsymbol{f}}_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t)\rangle. \tag{4.9}$$

Denote $\vec{\boldsymbol{A}}_t := \boldsymbol{\omega}_{A_t}$. Based on the coding vectors, the stage-t nuisance tangent space in Lemma 4.6 can be rewritten as

$$\Lambda_t = \left\{\boldsymbol{G} \in \mathcal{G} : \mathbb{E}\left(\frac{\boldsymbol{G}\vec{\boldsymbol{A}}_t^\intercal e_t^{(\mathrm{g})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}\middle|\boldsymbol{H}_t\right) = \mathsf{O}_{p\times(K-1)}\right\}.$$

Then we can characterize $\Lambda_t^\perp$ as in the following Lemma 4.9.

Lemma 4.9 (Characterization of $\Lambda_t^\perp$)**.** *Let Λ_t be the stage-t nuisance tangent space in Lemma 4.6, $\{\boldsymbol{\omega}_k\}_{k=1}^K \subseteq \mathbb{R}^{K-1}$ be the coding vectors satisfying $\sum_{k=1}^K c_k\boldsymbol{\omega}_k = \mathbf{0}$ if and only if $c_1 = c_2 = \cdots = c_K$. Denote $\vec{\boldsymbol{A}}_t := \boldsymbol{\omega}_{A_t}$. Then*

$$\Lambda_t^\perp = \left\{\frac{\mathsf{G}_t(\boldsymbol{H}_t)\vec{\boldsymbol{A}}_t e_t^{(\mathrm{g})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}\middle|\mathsf{G}_t : \mathcal{H}_t \to \mathbb{R}^{p\times(K-1)}\right\}.$$

Here, $p_{\mathscr{A},t}(a_t|\boldsymbol{h}_t) := \mathbb{P}(A_t = a_t|\boldsymbol{H}_t = \boldsymbol{h}_t)$.

Based on Lemma 4.9, we have $\Lambda_{1:T}^\perp = \left\{\sum_{t=1}^T \frac{\mathsf{G}_t(\boldsymbol{H}_t)\vec{\boldsymbol{A}}_t e_t^{(\mathrm{g})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}\middle|\mathsf{G}_t : \mathcal{H}_t \to \mathbb{R}^{p\times(K-1)}\ (1 \leqslant t \leqslant T)\right\}$. Compared with Robins (2004, Equation (3.10)), our characterization of the nuisance tangent space $\Lambda_t^\perp$ utilize the equivangular coding $\{\boldsymbol{\omega}_k\}_{k=1}^K$ to re-express the working instruments $\boldsymbol{G}_t(\boldsymbol{H}_t, A_t) - \mathbb{E}[\boldsymbol{G}_t(\boldsymbol{H}_t, A_t)|\boldsymbol{H}_t]$ in Robins (2004) by $\frac{\mathsf{G}_t(\boldsymbol{H}_t)\vec{\boldsymbol{A}}_t}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}$, and can be more tractable to analyze.

Notice that $\{\Lambda_t^\perp\}_{t=1}^T$ are not mutually orthogonal, since $\mathbb{E}(e_s^{(\mathrm{g})}e_t^{(\mathrm{g})}|\boldsymbol{H}_t, A_t)$ is generally nonzero for $1 \leqslant s \leqslant t \leqslant T$ as in (4.7). In the next Lemma 4.10, we perform orthogonalization on $\{\Lambda_t^\perp\}_{t=1}^T$ to obtain a direct sum representation of $\Lambda_{1:T}^\perp$. For a vector $\boldsymbol{x}$ and a positive semi-definite matrix W with compatible dimensions, we define $\|\boldsymbol{x}\|_{\mathsf{W}}^2 := \boldsymbol{x}^\intercal\mathsf{W}\boldsymbol{x}$.

Lemma 4.10 (Orthogonalization)**.** *Consider* $\{\Lambda_t^\perp\}_{t=1}^T$ *in Lemma 4.9. Define*

$$\mathring{\Lambda}_t^\perp := \left\{ \frac{\mathsf{G}_t(\boldsymbol{H}_t)\vec{\boldsymbol{A}}_t e_t^{(\mathrm{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \middle| \mathsf{G}_t : \mathcal{H}_t \to \mathbb{R}^{p\times(K-1)} \right\}; \quad 1 \leqslant t \leqslant T.$$

Here, the ort*-residuals* $\{e_t^{(\mathrm{ort})}\}_{t=1}^T$ *are recursively defined from*

$$e_T^{(\mathrm{ort})} := e_T^{(\mathrm{g})}; \quad e_t^{(\mathrm{ort})} := e_t^{(\mathrm{g})} - \sum_{u=t+1}^{T} \tau_u e_u^{(\mathrm{ort})}; \quad t = T-1, \cdots, 1. \tag{4.10}$$

The stage-t *orthogonalization coefficient is* $\tau_t = \tau_t(\boldsymbol{H}_t, A_t) := \left\langle \sum_{k=1}^K \rho_t(\boldsymbol{H}_t, k)\boldsymbol{\omega}_k, \frac{\mathsf{V}_t(\boldsymbol{H}_t)^{-1}\vec{\boldsymbol{A}}_t}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \right\rangle$, *where* $\rho_t(\boldsymbol{H}_t, A_t) = \mathbb{E}(e_t^{(\mathrm{g})} e_t^{(\mathrm{ort})}|\boldsymbol{H}_t, A_t)$, $\sigma_t^2(\boldsymbol{H}_t, A_t) := \mathbb{E}(e_t^{(\mathrm{ort})2}|\boldsymbol{H}_t, A_t)$, *and* $\mathsf{V}_t(\boldsymbol{H}_t) := \sum_{k=1}^K \frac{\sigma_t^2(\boldsymbol{H}_t,k)\boldsymbol{\omega}_t^{\otimes 2}}{p_{\mathscr{A},t}(k|\boldsymbol{H}_t)}$. *Then* $\{\mathring{\Lambda}_t^\perp\}_{t=1}^T$ *are mutually orthogonal and* $\Lambda_{1:T}^\perp = \bigoplus_{t=1}^T \mathring{\Lambda}_t^\perp$.

If we assume the additional condition (Robins, 2004, Equation (3.11)): $\mathbb{E}(e_t^{(\mathrm{g})2}|\boldsymbol{H}_t, A_t) = \mathbb{E}(e_t^{(\mathrm{g})2}|\boldsymbol{H}_t)$ *for* $1 \leqslant t \leqslant T$, *then* $e_t^{(\mathrm{ort})} = e_t^{(\mathrm{g})}$ *and* $\mathring{\Lambda}_t^\perp = \Lambda_t^\perp$ *for* $1 \leqslant t \leqslant T$. *In this case,* $\{\Lambda_t^\perp\}_{t=1}^T$ *are mutually orthogonal, and* $\Lambda_{1:T}^\perp = \bigoplus_{t=1}^T \Lambda_t^\perp$.

Robins (1994, Theorem 9) also performed the same orthogonalization as in our Lemma 4.10 to derive the semiparametric efficient score. However, the form of efficient score in Robins (1994) can be too complicated to use without assuming the stronger condition: $\mathbb{E}(e_t^{(\mathrm{g})2}|\boldsymbol{H}_t, A_t) = \mathbb{E}(e_t^{(\mathrm{g})2}|\boldsymbol{H}_t)$ for $1 \leqslant t \leqslant T$, as in his Corollary A3.2. Note that this condition is not satisfied in our Example 4.1.

Given the direct-sum representation $\Lambda_{1:T}^\perp = \bigoplus_{t=1}^T \Lambda_t^\perp$, we are finally able to characterize the IF of an RAL estimate as in the following Theorem 4.11. For symmetric matrices A, B of compatible dimensions, $\mathsf{A} \leqslant \mathsf{B}$ means $\mathsf{B} - \mathsf{A}$ is positive semi-definite.

Theorem 4.11 (IF under the SNMM)**.** *Consider* $(\text{SNMM})_{t=1}^{T}$ *and the ort-residuals* $\{e_t^{(\text{ort})}\}_{t=1}^{T}$ *in Lemma 4.10. The IF of an RAL estimate* $\widehat{\boldsymbol{\beta}}_{1:T,n} = \big(\widehat{\boldsymbol{\beta}}_{1,n}^{\intercal}, \widehat{\boldsymbol{\beta}}_{2,n}^{\intercal}, \cdots, \widehat{\boldsymbol{\beta}}_{T,n}^{\intercal}\big)^{\intercal}$ *for* $\boldsymbol{\beta}_{1:T}$ *takes the form*

$$\mathsf{IF}(\mathsf{G}) = \begin{bmatrix} \mathsf{G}_{11}(\boldsymbol{H}_1) & \mathsf{G}_{12}(\boldsymbol{H}_2) & \cdots & \mathsf{G}_{1T}(\boldsymbol{H}_1) \\ \mathsf{G}_{21}(\boldsymbol{H}_1) & \mathsf{G}_{22}(\boldsymbol{H}_2) & \cdots & \mathsf{G}_{2T}(\boldsymbol{H}_T) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{G}_{T1}(\boldsymbol{H}_1) & \mathsf{G}_{T2}(\boldsymbol{H}_2) & \cdots & \mathsf{G}_{TT}(\boldsymbol{H}_T) \end{bmatrix} \begin{bmatrix} \frac{\vec{\boldsymbol{A}}_1 e_1^{(\text{ort})}}{p_{\mathscr{A},1}(A_1|\boldsymbol{H}_1)} \\ \frac{\vec{\boldsymbol{A}}_2 e_2^{(\text{ort})}}{p_{\mathscr{A},2}(A_2|\boldsymbol{H}_2)} \\ \vdots \\ \frac{\vec{\boldsymbol{A}}_T e_T^{(\text{ort})}}{p_{\mathscr{A},T}(A_T|\boldsymbol{H}_T)} \end{bmatrix};$$

$$\textit{subject to} \quad \mathbb{E}[\mathsf{IF}(\mathsf{G})\boldsymbol{S}^{\intercal}] = \mathbf{I}_{p\times p},$$

where $\mathsf{G} := [\mathsf{G}_{st} : 1 \leqslant s,t \leqslant T]$ *with the working instrument functions* $\mathsf{G}_{st} : \mathcal{H}_t \to \mathbb{R}^{p_s\times(K-1)}$ $(1 \leqslant s,t \leqslant T)$*, and* $\boldsymbol{S} = (\partial/\partial\boldsymbol{\beta}_1^T)\log[\texttt{likelihood}(\boldsymbol{\beta}_1^T)]$ *is the semiparametric score vector.*

Consider the lower-triangular instrument matrix $\mathsf{L} := [\mathsf{L}_{st} : 1 \leqslant s,t \leqslant T]$ *with* $\mathsf{L}_{st} := \mathsf{G}_{st}\mathbb{1}(s \geqslant t)$*. Then* $\mathsf{IF}(\mathsf{L})$ *is also an IF, and* $\mathbb{E}[\mathsf{IF}(\mathsf{L})^{\otimes 2}] \leqslant \mathbb{E}[\mathsf{IF}(\mathsf{G})^{\otimes 2}]$.

Our Theorem 4.11 extends the characterization in Robins (2004, Equation (3.10)). In particular, if the upper-triangular entries of the instrument matrix G is nonzero, then the corresponding IF is inadmissible. In this case, there exists another IF with a lower-triangular instrument matrix and a smaller $\sqrt{n}$-asymptotic variancee.

In Section 4.6.2, we discuss the semiparametric efficient estimate based on the IF characterization in Theorem 4.11. We also point out that the semiparametric efficient estimate requires many vector-valued nuisance functions, which can be challenging to estimate in practice. Assuming $\mathbb{E}(e_t^{(\text{g})2}|\boldsymbol{H}_t, A_t) = \mathbb{E}(e_t^{(\text{g})2}|\boldsymbol{H}_t)$ and $\vec{\boldsymbol{f}}_t(\boldsymbol{H}_t;\mathsf{B}_t) = \mathsf{B}_t^{\intercal}\boldsymbol{H}_t$ for $1 \leqslant t \leqslant T$, the required nuisance functions are $\boldsymbol{H}_s \mapsto (1-1/K)\mathbb{E}\{[\vec{\boldsymbol{A}}_t - \vec{\boldsymbol{d}}_t^{\star}(\boldsymbol{H}_t)]\otimes\boldsymbol{H}_t|\boldsymbol{H}_s, A_s = k\}$ for $1 \leqslant k \leqslant K$ and $1 \leqslant s \leqslant t-1 \leqslant T$, where $\otimes$ denotes the Kronecker product. There are $KT(T-1)/2$ vector-valued nuisance functions to estimate, and each of them can be generally nonlinear. This shows the challenge of implementing the efficient estimation procedure rigorously.

4.2.5 Backward Change Point SNMM (BCP-SNMM)

Motivated from the challenge of semiparametric efficient estimation, we introduce an unknown backward change point $t_0 \in \{1, 2, \cdots, T\}$, such that Model $(\text{SNMM})_{t_0}^{T}$ holds for $t_0 \leqslant t \leqslant T$, while

a nonparametric data distribution is allowed during $1 \leqslant t \leqslant t_0 - 1$. Such a model is defined as the BCP-SNMM:

$$\begin{aligned} &(\boldsymbol{X}_s, A_s : 1 \leqslant s \leqslant t_0 - 1) \sim \mathbb{P}_0; \\ &\big((\boldsymbol{X}_t, A_t : t_0 \leqslant t \leqslant T), Y\big) \sim (\text{SNMM})_{t_0}^T; \\ &t_0 \in \{1, 2, \cdots, T\},\ \mathbb{P}_0 \text{ is arbitrary.} \end{aligned} \tag{4.11}$$

It can be clear that (4.11) incorporates $\bigcup_{t_0=1}^T (\text{SNMM})_{t_0}^T$, which can allow any initial time $t_0 \in \{1, 2, \cdots, T\}$ for the SNMM. Therefore, Model (4.11) consists of a larger class of semiparametric models compared with Model $(\text{SNMM})_1^T$. The change point t_0 is an unknown nuisance parameter rather than a parameter of interest in most change point detection literature (Jirak, 2015; Wang and Samworth, 2018; Liu et al., 2020). Since the change point t_0 and the data distribution $\mathbb{P}_0$ before t_0 are unknown, a stage-t semiparametric estimate of $\boldsymbol{\beta}_t$ must be pivotal with respect to the moment characterizations (4.5) for stages $1, 2, \cdots, t-1$. In this way, we can further eliminate the lower-triangular entries of the instrument matrix in Theorem 4.11, and obtain the IF characterization with an upper-triangular instrument matrix as in the following Theorem 4.12. As a consequence, the admissible instrument matrix becomes diagonal.

Theorem 4.12 (IF under the BCP-SNMM)**.** *Consider Model* (4.11) *and the ort-residuals* $\{e_t^{(\text{ort})}\}_{t=1}^T$ *in Lemma 4.10. The IF of an RAL estimate* $\widehat{\boldsymbol{\beta}}_{t_0:T,n} = \big(\widehat{\boldsymbol{\beta}}_{t_0,n}^\intercal, \widehat{\boldsymbol{\beta}}_{t_0,n}^\intercal, \cdots, \widehat{\boldsymbol{\beta}}_{T,n}^\intercal\big)^\intercal$ *for* $\boldsymbol{\beta}_{t_0:T}$ *takes the form*

$$\mathsf{IF}(\mathsf{G}) = \begin{bmatrix} \mathsf{G}_{t_0t_0}(\boldsymbol{H}_{t_0}) & \mathsf{G}_{t_0,t_0+1}(\boldsymbol{H}_{t_0+1}) & \cdots & \mathsf{G}_{t_0T}(\boldsymbol{H}_T) \\ & \mathsf{G}_{t_0+1,t_0+1}(\boldsymbol{H}_{t_0+1}) & \cdots & \mathsf{G}_{t_0+1,T}(\boldsymbol{H}_T) \\ & & \ddots & \vdots \\ & & & \mathsf{G}_T(\boldsymbol{H}_T) \end{bmatrix} \begin{bmatrix} \frac{\vec{\boldsymbol{A}}_{t_0} e_{t_0}^{(\text{ort})}}{p_{\mathscr{A},t_0}(A_{t_0}|\boldsymbol{H}_{t_0})} \\ \frac{\vec{\boldsymbol{A}}_{t_0+1} e_{t_0+1}^{(\text{ort})}}{p_{\mathscr{A},t_0+1}(A_{t_0+1}|\boldsymbol{H}_{t_0+1})} \\ \vdots \\ \frac{\vec{\boldsymbol{A}}_T e_T^{(\text{ort})}}{p_{\mathscr{A},T}(A_T|\boldsymbol{H}_T)} \end{bmatrix};$$

subject to $\mathbb{E}[\mathsf{IF}(\mathsf{G})\boldsymbol{S}^\intercal] = \mathbf{I}$,

where $\mathsf{G} := [\mathsf{G}_{st} : t_0 \leqslant s \leqslant t \leqslant T]$ *with the working instrument functions* $\mathsf{G}_{st} : \mathcal{H}_t \to \mathbb{R}^{p_s \times (K-1)}$ $(t_0 \leqslant s \leqslant t \leqslant T)$*, and* $\boldsymbol{S} = (\partial/\partial \boldsymbol{\beta}_{t_0}^T) \log[\texttt{likelihood}(\boldsymbol{\beta}_1^T)]$ *is the semiparametric score vector.*

Consider the diagonal instrument matrix $\mathsf{D} := \mathsf{diag}\{\mathsf{G}_{tt}\}_{t=t_0}^T$*. Then* $\mathsf{IF}(\mathsf{D})$ *is also an IF, and* $\mathbb{E}[\mathsf{IF}(\mathsf{D})^{\otimes 2}] \leqslant \mathbb{E}[\mathsf{IF}(\mathsf{G})^{\otimes 2}]$.

Comparing Theorems 4.11 and 4.12, the lower-triangular entries of the instrument matrix are forced to zero under Model (4.11). The reason is that, as a stage-s RAL estimate of Model (4.11), $\widehat{\boldsymbol{\beta}}_{t,n}$ is pivotal to any $t_0 < s$, and hence can only depend on the future model assumptions $(\text{SNMM})_s^T$. For $1 \leqslant t \leqslant s-1$, the instrument function $\mathsf{G}_{st}(\boldsymbol{H}_t)$ corresponds to $\frac{\bar{\boldsymbol{A}}_t e_t^{(\text{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}$, where the leveraged moment condition $\mathbb{E}(e_t^{(\text{ort})}|\boldsymbol{H}_t, A_t) = 0$ is based on $(\text{SNMM})_t$. Therefore, $\mathsf{G}_{st}(\boldsymbol{H}_t)$ must be zero so that the s-th IF is pivotal to $(\text{SNMM})_t$.

Notice that an admissible IF with diagonal instrument matrix in Theorem 4.12 is cross-stage uncorrelated. Specifically, $\frac{\mathsf{G}_{tt}(\boldsymbol{H}_t)\vec{\boldsymbol{A}}_t e_t^{(\text{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \in \mathring{\Lambda}_t^{\perp}$ where, by Lemma 4.10, $\{\mathring{\Lambda}_t^{\perp}\}_{t=1}^T$ are orthogonal. This implies that an admissible RAL estimate of Model (4.11) is $\sqrt{n}$-asymptotically across-stages independent.

The advantage of the semiparametric estimates for (4.11) is that these estimates are robust to backward model misspecifications. Specifically, consider the stage-t estimate $\widehat{\boldsymbol{\beta}}_{t,n}$ that does not depend on model assumptions of $(\text{SNMM})_1^{t-1}$. Then if any of $(\text{SNMM})_1^{t-1}$ are incorrectly specified, the estimate $\widehat{\boldsymbol{\beta}}_{t,n}$ still remains consistent. In this way, studying the bigger class of semiparametric models in Model (4.11) can allow the gain of more robustness.

The fact that an RAL estimate of Model (4.11) can only depend on future model assumptions also suggests that Model (4.11) can incorporate several backward stagewise estimates, including Q-Learning (Watkins, 1989; Chakraborty et al., 2010), recursive G-Estimation (Robins, 2004, Section 7.2), stagewise A-Learning (Shi et al., 2018a), and dWOLS (Wallace and Moodie, 2015).

The cross-stage orthogonality of any admissible RAL estimates for Model (4.11) can simplify the semiparametric efficient estimate. Specifically, we can find the efficient estimates $\widehat{\boldsymbol{\beta}}_{\text{eff},t,n}$ for each stage separately, where the efficient working instrument function $\mathsf{G}_{tt}(\boldsymbol{H}_t)$ is chosen such that the $\sqrt{n}$-asymptotic variance of the stage-t estimate $\widehat{\boldsymbol{\beta}}_{\text{eff},t,n}$ is minimized regardless of its influence to other stages. The efficient estimate combines the stagewise estimates together. The following Theorem 4.13 describes this procedure in terms of the semiparametric efficient score.

Theorem 4.13 (Semiparametric Efficient Score under BCP-SNMM)**.** *Consider Model* (4.11)*, the angle-based representation in Lemma 4.8, and the ort-residuals* $\{e_t^{(\mathrm{ort})}\}_{t=1}^T$ *in Lemma 4.10. Assume* t_0 *is known. The semiparametric efficient score for* $\boldsymbol{\beta}_{t_0:T}$ *is* $\boldsymbol{S}_{\mathrm{eff}}^{(t_0)} := \big(\boldsymbol{S}_{\mathrm{eff},t_0}^{\intercal}, \boldsymbol{S}_{\mathrm{eff},t_0+1}^{\intercal}, \cdots, \boldsymbol{S}_{\mathrm{eff},T}^{\intercal}\big)^{\intercal}$ *where*

$$\boldsymbol{S}_{\mathrm{eff},t} := \dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^{\intercal}\Omega^{\intercal}\Omega\mathsf{V}_t(\boldsymbol{H}_t)^{-1}\frac{\vec{\boldsymbol{A}}_t e_t^{(\mathrm{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}; \quad 1 \leqslant t \leqslant T,$$

and $\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t) := (\partial/\partial\boldsymbol{\beta}_t^{\intercal})\vec{\boldsymbol{f}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)$*. The semiparametric Fisher information matrix is* $\mathcal{I}^{(t_0)}(\boldsymbol{\beta}_{1:T}) = \mathsf{diag}\{\mathcal{I}_t(\boldsymbol{\beta}_t)\}_{t=t_0}^T$*, where* $\mathcal{I}_t(\boldsymbol{\beta}_t) := \mathbb{E}\Big[\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^{\intercal}\Omega^{\intercal}\Omega\mathsf{V}_t(\boldsymbol{H}_t)^{-1}\Omega^{\intercal}\Omega\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)\Big]$ *for* $t_0 \leqslant t \leqslant T$*.*

Notice that Theorems 4.12 and 4.13 hold for stages after a given change point t_0, since for stage $t < t_0$, the data distribution in Model (4.11) is fully nonparametric, and the RAL estimate and semiparametric efficient score are not defined for these stages. In this chapter, our main focus is not to determine the change point t_0. Instead, we want to ensure that for any given t_0, the parameter estimates $\widehat{\boldsymbol{\beta}}_{t_0:T,n}$ for $(\mathrm{SNMM})_{t_0}^T$ are optimal, which can be guaranteed from the semiparametric efficient scores $\{\boldsymbol{S}_{\mathrm{eff},t}\}_{t=1}^T$ in Theorem 4.13 that are pivotal with respect to t_0.

Based on Theorem 4.13, we can finally define the efficient estimating function for Model (4.11). Denote $\breve{\eta}_{t:T} := (\breve{\mu}_{t:T}, \breve{p}_{\mathscr{A},t:T}, \breve{\sigma}^2_{t:T}, \breve{\rho}_{(t+1):T})$ as the nuisance components with the treatment-free effect functions $\{\breve{\mu}_u(\boldsymbol{H}_u)\}_{u=t}^T$, propensity score functions $\{\breve{p}_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)\}_{u=t}^T$, variance functions $\{\breve{\sigma}^2_u(\boldsymbol{H}_u, A_u)\}_{u=t}^T$, and covariance functions $\{\breve{\rho}_u(\boldsymbol{H}_u, A_u)\}_{u=t+1}^T$. Let $\breve{\boldsymbol{\beta}}_{(t+1):T}$ be working parameters for stages $t+1$ to T. We first introduce the stagewise ort-*outcomes*: $Y_T^{(\mathrm{ort})} := Y$, and for $t = T-1, \cdots, 1$,

$$\begin{aligned}
& Y_t^{(\mathrm{ort})} = Y_t^{(\mathrm{ort})}\Big(\boldsymbol{\beta}_t; \breve{\boldsymbol{\beta}}_{(t+1):T}, \breve{\eta}_{(t+1):T}\Big) \\
:= \; & Y - \sum_{u=t+1}^{T}\Big[\gamma_u(\boldsymbol{H}_u, A_u; \breve{\boldsymbol{\beta}}_u) - \max_{1\leqslant k\leqslant K}\gamma_u(\boldsymbol{H}_u, k; \breve{\boldsymbol{\beta}}_u)\Big] && \text{(g-outcome)} \\
& - \sum_{u=t+1}^{T}\Bigg\langle \sum_{k=1}^{K}\breve{\rho}_u(\boldsymbol{H}_u, k)\boldsymbol{\omega}_k, \Bigg[\sum_{k=1}^{K}\frac{\breve{\sigma}^2_u(\boldsymbol{H}_u, A_u)\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}\Bigg]^{-1}\frac{\vec{\boldsymbol{A}}_u}{\breve{p}_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}\Bigg\rangle \\
& \times\Big[Y_u^{(\mathrm{ort})} - \breve{\mu}_u(\boldsymbol{H}_u) - \gamma_u(\boldsymbol{H}_u, A_u; \breve{\boldsymbol{\beta}}_u)\Big]. && \text{(orthogonalization)}
\end{aligned} \tag{4.12}$$

Then the stage-t efficient estimating function is defined as

$$\begin{aligned}
&\boldsymbol{\phi}_{\text{eff},t}\left(\boldsymbol{\beta}_t; \breve{\boldsymbol{\beta}}_{(t+1):T}, \breve{\eta}_{t:T}\right) \\
&:= \underbrace{\left[Y_t^{(\text{ort})} - \breve{\mu}_t(\boldsymbol{H}_t) - \gamma_t(\boldsymbol{H}_t, A_t; \breve{\boldsymbol{\beta}}_t)\right]}_{\text{ort-residual}} \times \underbrace{\dot{\mathsf{F}}_t(\boldsymbol{H}_t; \boldsymbol{\beta}_t)^\intercal \Omega^\intercal \Omega \left[\sum_{k=1}^{K} \frac{\breve{\sigma}_t^2(\boldsymbol{H}_t, k)\boldsymbol{\omega}_k^{\otimes 2}}{\breve{p}_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}\right]^{-1} \frac{\vec{\boldsymbol{A}}_t}{\breve{p}_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}}_{\text{efficient instrument}}.
\end{aligned} \tag{4.13}$$

We obtain the stage-t estimate $\widehat{\boldsymbol{\beta}}_{t,n}$ by solving the estimating equation $\mathbb{E}_n[\boldsymbol{\phi}_{\text{eff},t}(\boldsymbol{\beta}_t; \widehat{\boldsymbol{\beta}}_{(t+1):T,n}, \breve{\mu}_{t:T})] = \mathbf{0}$, and the estimation proceeds with $t = T, T-1, \cdots, 1$.

4.3 Dynamic Efficient Learning (DE-Learning)

Based on Model (4.11), we are able to propose the DE-Learning the solves the corresponding semiparametric efficient estimation equations. We first consider the high-level procedures for DE-Learning in Section 4.3.1. Then we provide more implementation details in Section 4.3.2.

4.3.1 General Procedure

In Section 4.2.5, we have obtained the efficient estimating functions $\{\boldsymbol{\phi}_{\text{eff},t}(\boldsymbol{\beta}_t; \breve{\boldsymbol{\beta}}_{(t+1):T}, \breve{\eta}_{t:T})\}_{t=1}^T$ from (4.13). A DE-Learning estimate of $\boldsymbol{\beta}_{1:T}$ recursively solves:

$$\widehat{\boldsymbol{\beta}}_{t,n} \in \underset{\boldsymbol{\beta}_t \in \mathscr{B}_t}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbb{E}_n \left[\boldsymbol{\phi}_{\text{eff},t} \left(\boldsymbol{\beta}_t; \widehat{\boldsymbol{\beta}}_{(t+1):T,n}, \widehat{\eta}_{t:T,n} \right) \right] \right\|^2_{\mathcal{I}_t(\boldsymbol{\beta}_t)^{-1}} \right\}; \quad t = T, T-1, \cdots, 1, \tag{4.14}$$

where $\|\boldsymbol{x}\|^2_{\mathsf{W}} := \boldsymbol{x}^\intercal \mathsf{W} \boldsymbol{x}$, and $\widehat{\eta}_{t:T,n}$ are finite-sample estimates of nuisance functions. Given the working nuisance functions $\breve{\eta}_{t:T} = (\breve{\mu}_{t:T}, \breve{p}_{\mathscr{A},t:T}, \breve{\sigma}^2_{(t+1):T}, \breve{\rho}_{(t+1):T})$, it can be shown that the following stage-t working variance function is for optimal stage-t estimation:

$$\sigma^2_{\text{opt},t}(\boldsymbol{H}_t, A_t; \breve{\eta}_{t:T}) := \mathbb{E}\left[e_t^{(\text{ort})}(\boldsymbol{\beta}_{t:T}; \breve{\eta}_{t:T})^2 \middle| \boldsymbol{H}_t, A_t \right],$$

while the following stage-t working covariance function is for stage-t orthogonalization:

$$\rho_{\text{ort},t}(\boldsymbol{H}_t, A_t; \breve{\eta}_{t:T}) := \mathbb{E}\left[e_t^{(\text{g})}(\boldsymbol{\beta}_{t:T}; \breve{\eta}_{t:T}) e_t^{(\text{ort})}(\boldsymbol{\beta}_{t:T}; \breve{\eta}_{t:T}) \middle| \boldsymbol{H}_t, A_t \right].$$

Both the working variance and covariance functions can be identified from the computable ort-residual $e_t^{(\mathrm{ort})}$ from (4.10) and the g-residual $e_t^{(\mathrm{g})}$ from (4.10) with 0 orthogonalization. Therefore, we can obtain the estimated variance function $\widehat{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$ by regressing $e_t^{(\mathrm{ort})2}$ on $(\boldsymbol{H}_t, A_t)$, and obtain the estimated covariance function $\widehat{\rho}_{t,n}(\boldsymbol{H}_t, A_t)$ by regressing $e_t^{(\mathrm{g})} e_t^{(\mathrm{ort})}$ on $(\boldsymbol{H}_t, A_t)$.

The general procedures of DE-Learning are given as follows.

- Input data $\{(\boldsymbol{X}_{it}, A_{it} : 1 \leqslant t \leqslant T), Y_i\}_{i=1}^n$. Define $\boldsymbol{H}_1 = \boldsymbol{X}_1$, $\boldsymbol{H}_t := (\boldsymbol{H}_{t-1}^\intercal, A_{t-1}^\intercal, \boldsymbol{X}_t^\intercal)^\intercal$ for $2 \leqslant t \leqslant T$. Input or estimate the propensity score functions $\{\widehat{p}_{\not\mathscr{A},t,n}(\boldsymbol{H}_t, k)\}_{\substack{1 \leqslant k \leqslant K \\ 1 \leqslant t \leqslant T}}$.
- Set the initial g-outcome and ort-outcome as $Y_T^{(\mathrm{g})} \leftarrow Y_T^{(\mathrm{ort})} \leftarrow Y$. For stage $t = T, \cdots, 1$, do the following.

Step 1. Estimate the treatment-free effects $\widehat{\mu}_{t,n}(\boldsymbol{H}_t)$ using $Y_t^{(\mathrm{ort})}$ as the response.

Step 2. Obtain a consistent estimate $\widehat{\boldsymbol{\beta}}_{t,n}^{(0)}$ of $\boldsymbol{\beta}_t$ in $(\mathrm{SNMM})_t$. This can be done by solving (4.14) with the stage-t working variance function as 1.

Step 3. Compute the ort-residual $\widetilde{e}_t^{(\mathrm{ort})} \leftarrow Y_t^{(\mathrm{ort})} - \widehat{\mu}_{t,n}(\boldsymbol{H}_t) - \gamma_t(\boldsymbol{H}_t, A_t; \widehat{\boldsymbol{\beta}}_{t,n}^{(0)})$. Then perform a nonparametric regression using $\widetilde{e}_t^{(\mathrm{ort})2}$ as the response and $(\boldsymbol{H}_t, A_t)$ as the covariates to estimate the variance function $\widetilde{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$.

Step 4. Solve (4.14) again but with the stage-t working variance function $\widetilde{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$ from **Step 3** for the stage-t DE-Learning estimate $\widehat{\boldsymbol{\beta}}_{t,n}$.

Step 5. Compute the g-residual $e_t^{(\mathrm{g})} \leftarrow Y_t^{(\mathrm{g})} - \widehat{\mu}_{t,n}(\boldsymbol{H}_t) - \gamma_t(\boldsymbol{H}_t, A_t; \widehat{\boldsymbol{\beta}}_{t,n})$ and the ort-residual $e_t^{(\mathrm{ort})} \leftarrow Y_t^{(\mathrm{ort})} - \widehat{\mu}_{t,n}(\boldsymbol{H}_t) - \gamma_t(\boldsymbol{H}_t, A_t; \widehat{\boldsymbol{\beta}}_{t,n})$. Then:

- Perform a nonparametric regression using $e_t^{(\mathrm{ort})2}$ as the response and $(\boldsymbol{H}_t, A_t)$ as the covariates to estimate the variance function $\widehat{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$;
- Perform a nonparametric regression using $e_t^{(\mathrm{g})} e_t^{(\mathrm{ort})}$ as the response and $(\boldsymbol{H}_t, A_t)$ as the covariates to estimate the covariance function $\widehat{\rho}_{t,n}(\boldsymbol{H}_t, A_t)$.

Step 6. If $t \geqslant 2$, then update the stage-$(t-1)$ g-outcome and the ort-outcome as

$$
\begin{aligned}
Y_{t-1}^{(\mathrm{g})} \leftarrow & Y_t^{(\mathrm{g})} - \left(\gamma_t(\boldsymbol{H}_t, A_t; \widehat{\boldsymbol{\beta}}_{t,n}) - \max_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \widehat{\boldsymbol{\beta}}_{t,n})\right); \\
Y_{t-1}^{(\mathrm{ort})} \leftarrow & Y_t^{(\mathrm{ort})} - \left(\gamma_t(\boldsymbol{H}_t, A_t; \widehat{\boldsymbol{\beta}}_{t,n}) - \max_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \widehat{\boldsymbol{\beta}}_{t,n})\right) \\
& - \left\langle \sum_{k=1}^{K} \widehat{\rho}_{t,n}(\boldsymbol{H}_t, k)\boldsymbol{\omega}_k, \left[\sum_{k=1}^{K} \frac{\widehat{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)\boldsymbol{\omega}_k^{\otimes 2}}{\widehat{p}_{\mathscr{A},t,n}(A_t|\boldsymbol{H}_t)}\right]^{-1} \frac{\vec{\boldsymbol{A}}_t}{\widehat{p}_{\mathscr{A},t,n}(A_t|\boldsymbol{H}_t)} \right\rangle e_t^{(\mathrm{ort})},
\end{aligned}
$$

with $e_t^{(\mathrm{ort})}$, $\widehat{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$ and $\widehat{\rho}_{t,n}(\boldsymbol{H}_t, A_t)$ from **Step 5**.

Continue with the next t or stop if $t = 1$.

Notice that the stage-t variance function has been estimated twice. The first variance function estimate $\widetilde{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$ based on the initial consistent estimate $\widehat{\boldsymbol{\beta}}_{t,n}^{(0)}$ is used to obtain the DE-Learning estimate $\widehat{\boldsymbol{\beta}}_{t,n}$. The second variance function estimate $\widehat{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$ based on the DE-Learning estimate $\widehat{\boldsymbol{\beta}}_{t,n}$ is used for updating the ort-outcome $Y_t^{(\mathrm{ort})}$. The reason for estimating for the second time is to ensure $\widehat{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t)$ and $e_t^{(\mathrm{ort})}$ used in updating $Y_t^{(\mathrm{ort})}$ satisfy $\widehat{\sigma}_{t,n}^2(\boldsymbol{H}_t, A_t) \approx \mathbb{E}(e_t^{(\mathrm{ort})2}|\boldsymbol{H}_t, A_t)$. The second estimation of $\widehat{\sigma}_{t,n}^2$ can improve the performance.

4.3.2 Implementation

We provide more details for DE-Learning implementation in this section. In the input step and **Steps 1**, **3** and **5** of Section 4.3.1, we estimate the propensity score, treatment-free effect, variance and covariance functions. We provide more details in Sections 4.3.2.1-4.3.2.3. In Section 4.3.2.4, we further consider to solve the regularized version of DE-Learning (4.14).

4.3.2.1 Estimating the Propensity Score Function

Suppose the stage-t treatment assignment probability $p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)$ is unknown. The first approach of estimating $p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)$ is to consider the penalized multinomial logistic regression (Friedman et al., 2010). Specifically, consider the multinomial logistic working model $\breve{p}_{\mathscr{A},t}(k|\boldsymbol{H}_t; \tau_1, \tau_2, \cdots, \tau_K) := \frac{\exp(\tau_k^\intercal \boldsymbol{H}_t)}{\sum_{k'=1}^{K} \exp(\tau_{k'}^\intercal \boldsymbol{H}_t)}$. The propensity score parameters $\tau_1, \tau_2, \cdots, \tau_K \in \mathbb{R}^p$

can be estimated by the following penalized log-likelihood maximization:

$$\max_{\tau_1,\cdots,\tau_K\in\mathbb{R}^p}\left\{\mathbb{E}_n\left[\sum_{k=1}^{K}\tau_k^\intercal \boldsymbol{H}_t\mathbb{1}(A_t=k)-\log\left(\sum_{k'=1}^{K}e^{\tau_{k'}^\intercal \boldsymbol{H}_t}\right)\right]-\lambda_{\mathscr{A}}\sum_{j=1}^{p}\left(\sum_{k=1}^{K}\tau_{jk}^2\right)^{1/2}\right\},$$

where the group-LASSO penalty $\sum_{j=1}^{p}\left(\sum_{k=1}^{K}\tau_{jk}^2\right)^{1/2}$ takes $\{\tau_{jk}\}_{k=1}^{K}$ for the j-th variable across all treatments as a group, and $\lambda_{\mathscr{A}}$ is a tuning parameter and can be chosen using cross validation.

In observational studies, the propensity scores can be vulnerable to model misspecification. Another approach for estimating $p_{\mathscr{A}}(A_t|\boldsymbol{H}_t)$ is to consider flexible nonparametric regression using the regression forest (Athey et al., 2019). Specifically, for each $1\leqslant k\leqslant K$, we run a regression forest using $\mathbb{1}(A_t=k)$ as the response and $\boldsymbol{H}_t$ as the covariates. Then each fitted regression forest provides a prediction for $\mathbb{E}[\mathbb{1}(A_t=k)|\boldsymbol{H}_t]$. The final estimate of $p_{\mathscr{A},t}(k|\boldsymbol{H}_t)$ is the prediction after normalization such that the summation over $k=1,\cdots,K$ is one.

4.3.2.2 Estimating the Treatment-Free Effect Function

Similar to Section 4.3.2.1, the stage-t treatment-free effect function $\mu_t(\boldsymbol{H}_t)$ can be estimated from a parametric model or nonparametric regression. For parametric estimation, we consider the linear working model $\breve{\mu}_t(\boldsymbol{H}_t;\boldsymbol{\eta}_t)=\boldsymbol{H}_t^\intercal\boldsymbol{\eta}_t$ as in Wallace and Moodie (2015); Shi et al. (2018a); Zhu et al. (2019). As pointed out in the remark on Corollary 4.5, if we specify linear models for the interaction effects $\{\gamma_t(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)\}_{t=1}^{T}$, then the true treatment-free effects $\{\mu_t(\boldsymbol{H}_t)\}_{t=1}^{T}$ are generally nonlinear, and linear working models always misspecify the true model. Nevertheless, the linear working model has been widely used for implementation convenience and interpretability (Chakraborty et al., 2010; Wallace and Moodie, 2015; Zhu et al., 2019). In this case, we consider a joint estimation of μ_t and $\boldsymbol{\beta}_t$ by the following penalized inverse-probability weighted least-squares problem with the ℓ_1-penalty:

$$\min_{\boldsymbol{\eta}_t,\beta_t}\left\{\mathbb{E}_n\left[\frac{1}{\widehat{p}_{\mathscr{A},t,n}(A_t|\boldsymbol{H}_t)}\left(\widehat{Y}_t-\boldsymbol{H}_t^\intercal\boldsymbol{\eta}_t-\gamma_t(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)\right)^2\right]+\lambda_{\mu_t}\left(\|\boldsymbol{\eta}_t\|_1+\|\boldsymbol{\beta}_t\|_1\right)\right\},$$

where $\widehat{Y}_t$ is the stage-t working outcome, λ_{μ_t} is a tuning parameter and can be chosen using cross validation. Here, the weighted least-squares problem can be equivalent to solving an inefficient but consistent estimating equation for $(\text{SNMM})_t$. If $\widehat{p}_{\mathscr{A},t,n}(A_t|\boldsymbol{H}_t)$ is the correct propensity score, then the above estimate for $\boldsymbol{\eta}_t$ can be consistent even if the model for the interaction effect $\gamma_t(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)$

is incorrect. If the model for the interaction effect $\gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t)$ is correct, then the above estimate for $\boldsymbol{\eta}_t$ can also be consistent for any arbitrary $\widehat{p}_{\mathscr{A},t,n}$ besides the correct one.

For nonparametric estimation of $\mu_t(\boldsymbol{H}_t)$, which was also considered in Ertefaie et al. (2021), we first divide the data into K subsets according to the received treatments. For each $1 \leqslant k \leqslant K$, we use $\widehat{Y}_t$ as the response and $\boldsymbol{H}_t$ as the covariates to fit a regression forest on the data subset $\{(\boldsymbol{H}_{it}, Y_{\text{ort},it}) : A_{it} = k\}$. Then each fitted regression forest corresponds to the prediction of $\mathbb{E}(\widehat{Y}_t | \boldsymbol{H}_t, A_t = k)$. We average the predictions over $k = 1, \cdots, K$ to obtain the treatment-free effect estimate.

4.3.2.3 Estimating the Variance and Covariance Functions

Suppose $e_t^{(\text{g})}$ and $e_t^{(\text{ort})}$ are the working residuals in the general DE-Learning procedure. In order to estimate the variance function, we specifically consider the regression forest using $e_t^{(\text{ort})2}$ as the response and $(\boldsymbol{H}_t, A_t)$ as the covariates. Then $\widehat{\sigma}^2_{t,n}(\boldsymbol{H}_t, k)$ is the regression forest prediction at $(\boldsymbol{H}_t, k)$ for $1 \leqslant k \leqslant K$. Similarly, for the covariance function estimation, we consider $e_t^{(\text{g})} e_t^{(\text{ort})}$ as the response and $(\boldsymbol{H}_t, A_t)$ as the covariates to obtain the regression forest estimates $\{\widehat{\rho}_{t,n}(\boldsymbol{H}_t, k)\}_{k=1}^K$.

4.3.2.4 Solving the Regularized DE-Learning Estimating Equation

In this section, we consider a general penalty $J_t(\boldsymbol{\beta}_t)$ for the stage-t parameter estimation. To incorporate regularization in DE-Learning from (4.14), we solve a penalized minimization problem:

$$\widehat{\boldsymbol{\beta}}_{t,n} \in \underset{\boldsymbol{\beta}_t \in \mathscr{B}_t}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbb{E}_n \left[\boldsymbol{\phi}_{\text{eff},t} \left(\boldsymbol{\beta}_t; \widehat{\boldsymbol{\beta}}_{(t+1):T,n}, \widehat{\eta}_{t:T,n} \right) \right] \right\|^2_{\mathsf{W}} + \lambda_t J_t(\boldsymbol{\beta}_t) \right\}; \quad t = T, T-1, \cdots, 1, \quad (4.15)$$

where W can be a general positive definite weighting matrix $\mathsf{W} \in \mathbb{R}^{p_t \times p_t}$. A typical choice of W can be $\mathsf{I}_{p_t \times p_t}$ or the inverse of the empirical information matrix $\left\{ \mathbb{E}_n \left[-(\partial / \partial \boldsymbol{\beta}_t^{\intercal}) \boldsymbol{\phi}_{\text{eff},t} \left(\boldsymbol{\beta}_t; \widehat{\boldsymbol{\beta}}_{(t+1):T,n}, \widehat{\eta}_{t:T,n} \right) \right] \right\}^{-1}$. Problem (4.15) can be solved by the accelerated proximal gradient method (Nesterov, 2013) with the gradient $\boldsymbol{\beta}_t \mapsto \mathbb{E}_n \left[\boldsymbol{\phi}_{\text{eff},t} \left(\boldsymbol{\beta}_t; \widehat{\boldsymbol{\beta}}_{(t+1):T,n}, \widehat{\eta}_{t:T,n} \right) \right]$. For a fixed tuning parameter λ_t, the estimation procedure follows **Steps 1-4** in Section 4.3.1. The parameter λ_t can be further tuned by cross validation. The IPWE of the value function is used as the tuning criteria. Denote $\widehat{\boldsymbol{\beta}}_{t,n}(\lambda_t)$ as the solution to (4.15). The corresponding decision rule is $\widehat{d}_{t,n}(\boldsymbol{H}_t; \lambda_t) := \operatorname{argmax}_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \widehat{\boldsymbol{\beta}}_{t,n}(\lambda_t))$.

Let $\{(\boldsymbol{H}_{it}, A_{it}, Y_{\mathrm{ort},it})\}_{i=1}^{n_{\mathrm{valid}}}$ be the stage-t validation dataset. Then the criteria for tuning λ_t is $\frac{1}{n_{\mathrm{valid}}}\sum_{i=1}^{n_{\mathrm{valid}}} \frac{\mathbb{1}[\widehat{d}_{t,n}(\boldsymbol{H}_{it};\lambda_t)=A_{it}]}{\widehat{p}_{\mathscr{A},t,n}(A_{it}|\boldsymbol{H}_{it})} Y_{\mathrm{ort},it}$, which is larger the better.

4.4 Simulation Studies

We compare the proposed DE-Learning with serveral existing methods via simulation studies in this section. Consider the following data generation process. First we generate the stagewise covariates, pre-treatment histories and treatments from:

$$\{Z_{tj} : 1 \leqslant t \leqslant T+1, 1 \leqslant j \leqslant p\} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1);$$

$$X_{1j} = H_{1j} = Z_{1j}; \quad \mathbb{P}(A_1 = k|\boldsymbol{X}_1) = \frac{e^{X_{1k}/2}}{\sum_{k'=1}^{K} e^{X_{1k'}/2}};$$

$$\begin{aligned} &X_{tj} = Z_{tj} + \mathbb{1}(A_{t-1} = j) - \tfrac{1}{K-1}\mathbb{1}(A_{t-1} \neq j, 1 \leqslant j \leqslant K); \\ &\boldsymbol{H}_t = (\boldsymbol{H}_{t-1}^{\intercal}, A_{t-1}, \boldsymbol{X}_t^{\intercal})^{\intercal}; \quad \mathbb{P}(A_t = k|\boldsymbol{H}_t) = \tfrac{e^{X_{tk}/2}}{\sum_{k'=1}^{K} e^{X_{tk'}/2}}; \end{aligned} \qquad 2 \leqslant t \leqslant T.$$

Then we generate outcome according to Corollary 4.5 as follows. Consider the coefficient vector for the stage-t interaction effect $\gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}^{(t)})$ at the k-th treatment as:

$$\boldsymbol{\beta}_k^{(t)} = (\beta_{00k}^{(t)}; \underbrace{\beta_{11k}^{(t)}, \beta_{12k}^{(t)}, \overbrace{0, \cdots, 0}^{p-2}}_{\boldsymbol{X}_1}; \underbrace{\beta_{1\mathscr{A}k}^{(t)}}_{A_1}; \cdots; \underbrace{\beta_{t-1,1k}^{(t)}, \beta_{t-1,2k}^{(t)}, \overbrace{0, \cdots, 0}^{p-2}}_{\boldsymbol{X}_{t-1}}; \underbrace{\beta_{t-1,\mathscr{A}k}^{(t)}}_{A_{t-1}}; \underbrace{\beta_{t1k}^{(t)}, \beta_{t2k}^{(t)}, \overbrace{0, \cdots, 0}^{p-2}}_{\boldsymbol{X}_t}),$$

where we first randomly generate $(\widetilde{\beta}_{00k}^{(t)}; \widetilde{\beta}_{11k}^{(t)}, \widetilde{\beta}_{12k}^{(t)}; \widetilde{\beta}_{1\mathscr{A}k}^{(t)}; \cdots; \widetilde{\beta}_{t-1,1k}^{(t)}, \widetilde{\beta}_{t-1,2k}^{(t)}; \widetilde{\beta}_{t-1,\mathscr{A}k}^{(t)}; \widetilde{\beta}_{t1k}^{(t)}, \widetilde{\beta}_{t2k}^{(t)})^{\intercal} \in \mathbb{R}^{3t}$ from the unit sphere $\{\boldsymbol{u} \in \mathbb{R}^{3t} : \|\boldsymbol{u}\|_2 = 1\}$, and then let $\beta_{sjk}^{(t)} := \sqrt{1-1/K}[\widetilde{\beta}_{sjk}^{(t)} - (1/K)\sum_{k'=1}^{K} \widetilde{\beta}_{sjk'}^{(t)}]$ for $0 \leqslant s \leqslant t \leqslant T$, $0 \leqslant j \leqslant p$, $j = \mathscr{A}$ and $1 \leqslant k \leqslant K$. Then we define the $(\boldsymbol{H}_t, A_t)_{t=1}^{T}$-martingale-difference sequence as: for $1 \leqslant t \leqslant T$,

$$\Delta\mathcal{M}_1(\boldsymbol{H}_1) = \frac{1}{\sqrt{K}}\sum_{k=1}^{K} Z_{1k}; \quad \Delta\mathcal{M}_{t+1}(\boldsymbol{H}_{t+1}) = e^{X_{t,A_t}} \times \frac{1}{\sqrt{K}}\sum_{k=1}^{K} Z_{t+1,k}; \quad \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}^{(t)}) = \boldsymbol{\beta}_{A_t}^{(t)\intercal}(1, \boldsymbol{H}_t^{\intercal})^{\intercal}.$$

The outcome according to (4.6) is finally given by

$$Y = \underbrace{1}_{\text{optimal value}} + \sum_{t=1}^{T} \Delta\mathcal{M}_t(\boldsymbol{H}_t) - \sum_{t=1}^{T}\left\{\max_{1\leqslant k\leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}^{(t)}) - \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}^{(t)})\right\} + \Delta\mathcal{M}_{T+1}(\boldsymbol{H}_{T+1}).$$

To implement penalized Q-Learning (Zhu et al., 2019), penalized stagewise A-Learning (Shi et al., 2018a), and our proposed DE-Learning, we consider the q-outcomes (4.8), the g-outcomes (4.3) and the ort-outcomes (4.12) respectively for stagewise working outcomes. The nuisance functions utilized in Q-Learning, A-Learning and DE-Learning are estimated according to the implementation details discussed in Section 4.3.2. In particular, the stagewise treatment-free effects are estimated from linear working models with the ℓ_1-penalty, and the stagewise propensity scores are estimated from multinomial logistic working models with the ℓ_1-penalty. Both Q-Learning and A-Learning utilize 1 as the working variance function and 0 as the working covariance function. Our proposed DE-Learning with the ort-outcomes estimates the variance and covariance functions using regression forest. In order to demonstrate the performance improvement of cross-stage orthogonalization, we also consider DE-Learning with the g-outcomes, in which case the working covariance function is set to 0.

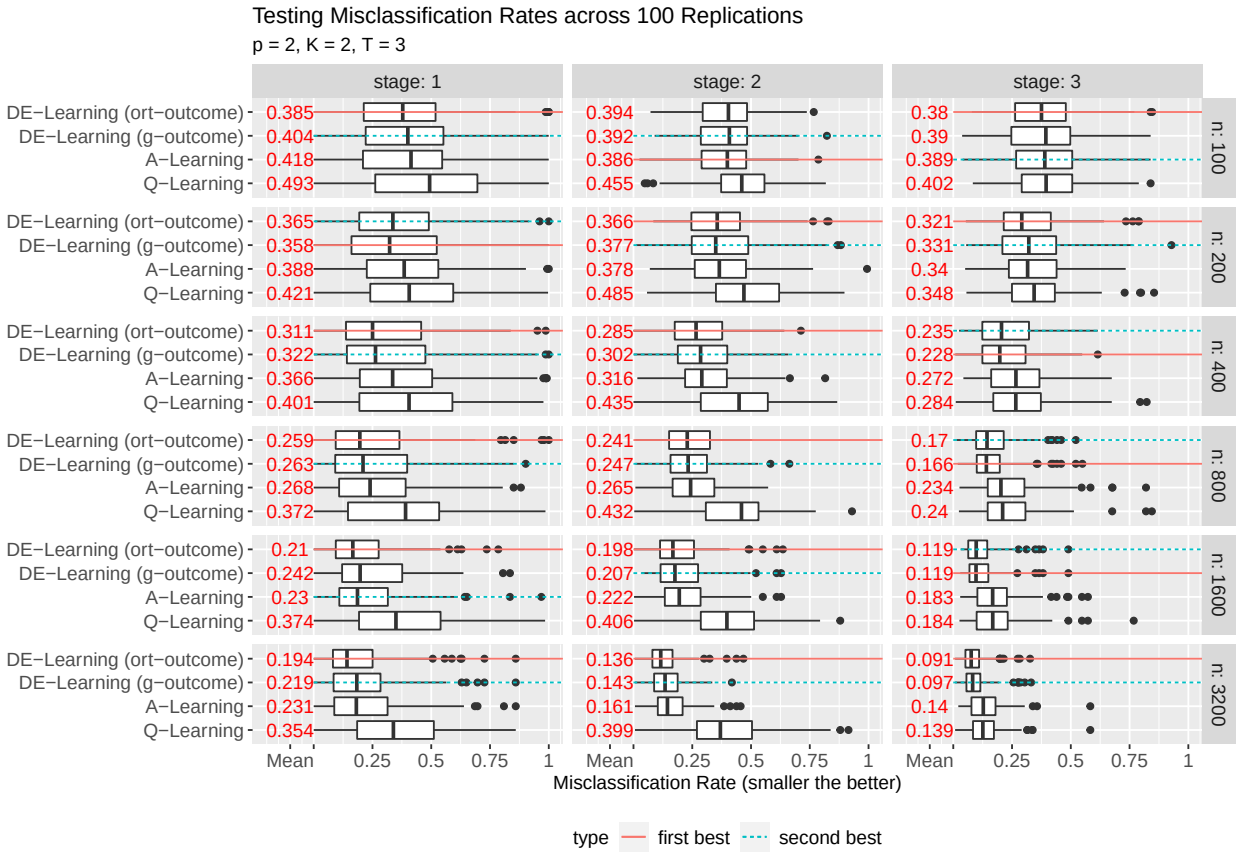


Figure 4.1: Testing misclassification rates in the simulation studies.

During the training stage, we consider the training sample size $n \in \{100, 200, 400, 800, 1600, 3200\}$, with the number of variables $p = 2$, the number of treatments $K = 2$ and the number of stages $T = 3$. In Figure 4.1, we report the misclassification rates on the testing sample of size 10,000 across 100 replications. Among the comparing methods, DE-Learning demonstrates the best testing performance, and the superiority is more evident as the training sample size increases. Q-Learning generally has the worst performance other than the final stage, since the working treatment-free effect functions are misspecified, and the q-outcomes for stages 1 and 2 are incorrect. A-Learning demonstrates its robustness in presence of misspecified treatment-free effects. However, since A-Learning is suboptimal in presence of treatment-free effect misspecification and heteroscedasticity, the testing misclassification rates are generally inferior to our proposed DE-Learning. When comparing DE-Learning with the ort-outcomes and the g-outcomes, the ort-outcomes can generally help to improve the testing performance, especially when the training sample size increases. This confirms the optimality of DE-Learning based on the efficient estimating procedure.

4.5 Discussion

In this chapter, we introduce a general class of semiparametric models, the BCP-SNMM, that incorporates the standard SNMM and enjoys more robustness. The class of semiparametric estimates for the BCP-SNMM can be a suitable framework for theoretically studying backward stagewise estimates. We also propose DE-Learning that solves the semiparametric efficient estimating equations under the BCP-SNMM and the multiple treatment setting. In particular, DE-Learning is optimal among a class of regular estimates of the BCP-SNMM even when the treatment-free effects are misspecified. It enjoys stagewise double robustness and the robustness with respect to backward model misspecifications. Compared with G-Estimation, DE-Learning is more tractable with much fewer nuisance functions to estimate and can be carried out in a backward stagewise fashion, which allows implementable rigorous semiparametric efficient estimation. Our simulation studies also demonstrate the superiority of DE-Learning in presence of stagewise misspecified treatment-free effects and heteroscedasticity.

There are some important future work for this chapter. First of all, we can explore more on the connections of DE-Learning with existing methods from literature under the BCP-SNMM framework. More comprehensive numerical studies are also needed to demonstrate the superiority of DE-Learning, including the cases with increasing numbers of variables and treatments, the existence of a backward change point, and the comparisons with other nonparametric methods. Furthermore, we can establish more theoretical properties for DE-Learning, including the stagewise double robustness and the optimality in presence of misspecified treatment-free effect models. Last but not least, the high-dimensional estimation properties as in Shi et al. (2018a); Zhu et al. (2019) can be established for DE-Learning.

4.6 Appendix

4.6.1 Pseudo Outcome

The pseudo outcome Y_t^* is defined as

$$Y_t^* = Y_t^*(\vec{\boldsymbol{A}}_1^t) := Y\Big[\underbrace{\Big(A_1, \cdots, A_t, A_{t+1}^*(\vec{\boldsymbol{A}}_1^t), A_{t+2}^*(\vec{\boldsymbol{A}}_1^t), \cdots, A_T^*(\vec{\boldsymbol{A}}_1^t)\Big)^\intercal}_{:=\vec{\boldsymbol{A}}_1^{*T}(\vec{\boldsymbol{A}}_1^t)}\Big]; \quad 1 \leqslant t \leqslant T. \tag{4.16}$$

Here, $\vec{\boldsymbol{A}}_1^{*T}(\vec{\boldsymbol{A}}_1^T) = \vec{\boldsymbol{A}}_1^T$, and hence $Y_T^*(\vec{\boldsymbol{A}}_1^T) = Y(\vec{A}_1^T) = Y$ from Assumption 4.1. For $1 \leqslant t \leqslant T-1$, $\{A_u^*(\vec{\boldsymbol{A}}_1^t)\}_{u=t+1}^T$ are pseudo treatments obtained from the following algorithm:

$$\begin{aligned}
\boldsymbol{H}_{t+1}^*(\vec{\boldsymbol{A}}_1^t) &:= \boldsymbol{H}_{t+1}(\vec{\boldsymbol{A}}_1^t); \\
A_{t+1}^*(\vec{\boldsymbol{A}}_1^t) &:= d_{t+1}^\star\Big[\boldsymbol{H}_{t+1}^*(\vec{\boldsymbol{A}}_1^t)\Big]; \\
\boldsymbol{H}_u^*(\vec{\boldsymbol{A}}_1^t) &:= \boldsymbol{H}_u\Big[\Big(A_1, \cdots, A_t, A_{t+1}^*(\vec{\boldsymbol{A}}_1^t), \cdots, A_{u-1}^*(\vec{\boldsymbol{A}}_1^t)\Big)^\intercal\Big]; \\
A_u^*(\vec{\boldsymbol{A}}_1^t) &:= d_u^\star\Big[\boldsymbol{H}_u^*(\vec{\boldsymbol{A}}_1^t)\Big]; \qquad\qquad u = t+2, t+3, \cdots, T.
\end{aligned} \tag{4.17}$$

Then $\vec{\boldsymbol{A}}_1^{T*}(\vec{\boldsymbol{A}}_1^t)$ can be interpreted as the treatment assignment trajectory that follows the observed treatments up to stage t and then follows the optimal treatments up to the end. Such a treatment trajectory corresponds to the potential pre-treatment histories $\boldsymbol{H}_1, \cdots, \boldsymbol{H}_t, \boldsymbol{H}_{t+1}^*(\vec{\boldsymbol{A}}_1^t), \cdots, \boldsymbol{H}_T^*(\vec{\boldsymbol{A}}_1^t)$

defined from (4.17).The pseudo outcome (4.16) is the potential outcomes under such a treatment trajectory and the resulting pre-treatment histories.

4.6.2 Semiparametric Efficient Estimate under the SNMM

Theorem 4.14 (Semiparametric Efficient Score under SNMM)**.** *Consider* $(\text{SNMM})_1^T$ *and the* ort*-residuals* $\{e_t^{(\text{ort})}\}_{t=1}^T$ *in Lemma 4.10. The semiparametric efficient score is* $\boldsymbol{S}_{\text{eff}} = \left(\boldsymbol{S}_{\text{eff},1}^\intercal, \boldsymbol{S}_{\text{eff},2}^\intercal, \cdots, \boldsymbol{S}_{\text{eff},T}^\intercal\right)^\intercal$*, where*

$$\boldsymbol{S}_{\text{eff},t} := \sum_{s=1}^{t} \left\{ \sum_{k=1}^{K} \mathbb{E}\left[-\frac{\partial e_s^{(\text{ort})}(\boldsymbol{\beta}_{s:T})}{\partial \boldsymbol{\beta}_t} \middle| \boldsymbol{H}_s, A_s = k \right] \boldsymbol{\omega}_k^\intercal \right\} \mathsf{V}_s(\boldsymbol{H}_s)^{-1} \frac{\vec{\boldsymbol{A}}_s r_s}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)}; \quad 1 \leqslant t \leqslant T.$$

The semiparametric Fisher information matrix for $\boldsymbol{\beta}_{1:T}$ *is* $\mathcal{I}(\boldsymbol{\beta}_{1:T}) = [\mathcal{I}_{ut}(\boldsymbol{\beta}_{1:T}) : 1 \leqslant u, t \leqslant T]$*, where for* $1 \leqslant u, t \leqslant T$*,*

$$\mathcal{I}_{ut}(\boldsymbol{\beta}_{1:T}) := \sum_{s=1}^{u\wedge t} \sum_{k,k'=1}^{K} \boldsymbol{\omega}_k^\intercal \mathsf{V}_s(\boldsymbol{H}_s)^{-1} \boldsymbol{\omega}_{k'} \mathbb{E}\left[-\frac{\partial e_s^{(\text{ort})}(\boldsymbol{\beta}_{s:T})}{\partial \boldsymbol{\beta}_u} \middle| \boldsymbol{H}_s, A_s = k \right] \mathbb{E}\left[-\frac{\partial e_s^{(\text{ort})}(\boldsymbol{\beta}_{s:T})}{\partial \boldsymbol{\beta}_t} \middle| \boldsymbol{H}_s, A_s = k' \right]^\intercal.$$

As a consequence of Theorem 4.14, the semiparametric efficient estimate for $\boldsymbol{\beta}_{1:T}$ in $(\text{SNMM})_1^T$ can be obtained as follows. Consider the angle-based decision functions $\{\vec{\boldsymbol{f}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)\}$ in Lemma 4.8. Denote $\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t) := (\partial/\partial\boldsymbol{\beta}_t^\intercal)\vec{\boldsymbol{f}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)$, $\vec{\boldsymbol{A}}_t = \boldsymbol{\omega}_{A_t}$ and $\vec{\boldsymbol{d}}_t^\star(\boldsymbol{H}_t) = \boldsymbol{\omega}_{d_t^\star(\boldsymbol{H}_t)}$. Then $(\partial/\partial\boldsymbol{\beta}_t)\gamma_t(\boldsymbol{H}_t, A_t;\boldsymbol{\beta}_t) = (1-1/K)\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^\intercal\vec{\boldsymbol{A}}_t$, and $-(\partial/\partial\boldsymbol{\beta}_t)e_s^{(\text{g})}(\boldsymbol{\beta}_{s:T}) = (1-1/K)\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^\intercal[\vec{\boldsymbol{A}}_t\mathbb{1}(s\leqslant t) - \vec{\boldsymbol{d}}_t^\star(\boldsymbol{H}_t)\mathbb{1}(s<t)]$ for $1\leqslant s\leqslant t\leqslant T$. By definition, the stage-s ort-residual can be represented as $e_s^{(\text{ort})} = \sum_{t=s}^T \nu_{st} e_t^{(\text{g})}$, where $\nu_{st} = \nu_{st}(\boldsymbol{H}_t, A_t) = 1$ if $1\leqslant s = t\leqslant T$; and $\prod_{u=s+1}^t[-\tau_u(\boldsymbol{H}_u, A_u)]$ if $1\leqslant s<t\leqslant T$. Define $\bar{\nu}_{st} = \bar{\nu}_{st}(\boldsymbol{H}_t, A_t) := \sum_{u=s}^t \nu_{su}(\boldsymbol{H}_u, A_u)$ if $s\leqslant t$; and 0 if $s>t$. Then for $1\leqslant s\leqslant t\leqslant T$, we have

$$-\frac{\partial e_s^{(\text{ort})}(\boldsymbol{\beta}_{s:T})}{\partial\boldsymbol{\beta}_t} = \left(1-\frac{1}{K}\right)\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^\intercal\left[\bar{\nu}_{st}(\boldsymbol{H}_t, A_t)\vec{\boldsymbol{A}}_t - \bar{\nu}_{s,t-1}(\boldsymbol{H}_{t-1}, A_{t-1})\vec{\boldsymbol{d}}_t^\star(\boldsymbol{H}_t)\right].$$

Therefore, for $1\leqslant t\leqslant T$, the stage-t semiparametric efficient score can be expressed as

$$\begin{aligned}\boldsymbol{S}_{\text{eff},t} = \sum_{s=1}^{t-1}\sum_{k=1}^{K} & \overbrace{\left(1-\frac{1}{K}\right)\mathbb{E}\left\{\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^\intercal\left[\bar{\nu}_{st}(\boldsymbol{H}_t, A_t)\vec{\boldsymbol{A}}_t - \bar{\nu}_{s,t-1}(\boldsymbol{H}_{t-1}, A_{t-1})\vec{\boldsymbol{d}}_t^\star(\boldsymbol{H}_t)\right]\middle|\boldsymbol{H}_s, A_s = k\right\}}^{\text{an } \mathbb{R}^{pt}\text{-valued nuisance function of } \boldsymbol{H}_s \mapsto \mathbb{E}[-(\partial/\partial\boldsymbol{\beta}_t)e_s^{(\text{ort})}(\boldsymbol{\beta}_{s:T})|\boldsymbol{H}_s, A_s = k]}\boldsymbol{\omega}_k^\intercal \\ & \times \mathsf{V}_s(\boldsymbol{H}_s)^{-1}\frac{\vec{\boldsymbol{A}}_s e_s^{(\text{ort})}(\boldsymbol{\beta}_{s:T})}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)} + \dot{\mathsf{F}}(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^\intercal\Omega^\intercal\Omega\mathsf{V}_t(\boldsymbol{H}_t)^{-1}\frac{\vec{\boldsymbol{A}}_t e_t^{(\text{ort})}(\boldsymbol{\beta}_{t:T})}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}. \end{aligned} \tag{4.18}$$

In order to implement the semiparametric efficient procedure, we need to estimate the $\mathbb{R}^{p_t}$-valued nuisance functions in (4.18). If we assume the additional condition: $\mathbb{E}(e_t^{(\mathrm{g})2}|\boldsymbol{H}_t, A_t) = \mathbb{E}(e_t^{(\mathrm{g})2}|\boldsymbol{H}_t)$ for $1 \leqslant t \leqslant T$, then $\bar{\nu}_{st} = \mathbb{1}(s \leqslant t)$. The required nuisance functions are $\boldsymbol{H}_s \mapsto (1-1/K)\mathbb{E}\{\dot{\mathsf{F}}_t(\boldsymbol{H}_t;\boldsymbol{\beta}_t)^\intercal[\vec{\boldsymbol{A}}_t - \vec{\boldsymbol{d}}_t^\star(\boldsymbol{H}_t)]|\boldsymbol{H}_s, A_s = k\}$ for $1 \leqslant k \leqslant K$ and $1 \leqslant s \leqslant t-1 \leqslant T$. For the linear decision function $\vec{\boldsymbol{f}}_t(\boldsymbol{H}_t;\mathsf{B}_t) = \mathsf{B}_t^\intercal \boldsymbol{H}_t$ with $\mathsf{B}_t \in \mathbb{R}^{\dim(\boldsymbol{H}_t)\times(K-1)}$, it further reduces to $\boldsymbol{H}_s \mapsto (1-1/K)\mathbb{E}\{[\vec{\boldsymbol{A}}_t - \vec{\boldsymbol{d}}_t^\star(\boldsymbol{H}_t)] \otimes \boldsymbol{H}_t|\boldsymbol{H}_s, A_s = k\}$ (Almirall et al., 2010, Section 3.3.1). Such an $\mathbb{R}^{(K-1)\dim(\boldsymbol{H}_t)}$-valued nuisance function is generally nonlinear can be hard to estimate well in practice. The total number of such nuisance functions are $KT(T-1)/2$. Therefore, the semiparametric efficient G-Estimation is rarely used in practice (Vansteelandt and Joffe, 2014; Wallace et al., 2019; Liu et al., 2021).

4.6.3 Technical Proofs

4.6.3.1 Proof of Lemma 4.1

Proof of Lemma 4.1. By Assumption 4.1, the algorithm (4.17) implies that $\boldsymbol{H}_u = \boldsymbol{H}_u^*(\vec{\boldsymbol{A}}_1^t)$ and $A_u = A_u^*(\vec{\boldsymbol{A}}_1^t)$ for $u = t+1, t+2, \cdots, T$ on $\{d_u^\star(\boldsymbol{H}_u) = A_u \ (t+1 \leqslant u \leqslant T)\}$. Then, we further have

$$Y = Y\left[\vec{\boldsymbol{A}}_1^{*T}(\vec{\boldsymbol{A}}_1^t)\right] = Y_t^* \text{ on } \left\{d_u^\star(\boldsymbol{H}_u) = A_u \ (t+1 \leqslant u \leqslant T)\right\}; \quad 1 \leqslant t \leqslant T. \tag{4.19}$$

Therefore, for $1 \leqslant t \leqslant T$, we have

$$\begin{aligned}
&\mathbb{E}\left\{\mathcal{V}_{t+1}\left[\left(\boldsymbol{H}_t^\intercal, A_t, \vec{\boldsymbol{X}}_{t+1}^\intercal\right)^\intercal\right]\Big|\boldsymbol{H}_t, A_t\right\} \\
=&\mathbb{E}\Big\{\mathbb{E}[Y|\underbrace{\boldsymbol{H}_t, A_t, \boldsymbol{X}_{t+1}}_{\boldsymbol{H}_{t+1}}, d_u^\star(\boldsymbol{H}_u) = A_u \ (t+1 \leqslant u \leqslant T)]\Big|\boldsymbol{H}_t, A_t\Big\} && \text{(by definition (4.1))} \\
=&\mathbb{E}[Y|\boldsymbol{H}_t, A_t, d_u^\star(\boldsymbol{H}_u) = A_u \ (t+1 \leqslant u \leqslant T)] \\
=&\mathbb{E}[Y_t^*|\boldsymbol{H}_t, A_t, d_u^\star(\boldsymbol{H}_u) = A_u \ (t+1 \leqslant u \leqslant T)] && \text{(by (4.19))} \\
=&\mathbb{E}(Y_t^*|\boldsymbol{H}_t, A_t). && \text{(by Assumption 4.2)}
\end{aligned}$$

□

4.6.3.2 Proof of Lemma 4.3

Proof of Lemma 4.3. For $t = T-1$, we have

$$\begin{aligned}
Y_{T-1}^{(\mathrm{g})} &= Y - \gamma_T(\boldsymbol{H}_T, A_T; \boldsymbol{\beta}_T) + \max_{1 \leqslant k \leqslant K} \gamma_T(\boldsymbol{H}_T, k; \boldsymbol{\beta}_T) && \text{(by definition)} \\
&= Y_T^* - \gamma_T(\boldsymbol{H}_T, A_T; \boldsymbol{\beta}_T) + \max_{1 \leqslant k \leqslant K} \gamma_T(\boldsymbol{H}_T, k; \boldsymbol{\beta}_T) && \text{(by definition)} \\
&= \mu_T(\boldsymbol{H}_T) + \max_{1 \leqslant k \leqslant K} \gamma_T(\boldsymbol{H}_T, k; \boldsymbol{\beta}_T) + \epsilon_T^* && \text{(by (SNMM)}_T\text{)} \\
&= \max_{1 \leqslant k \leqslant K} \mathbb{E}(Y_T^* | \boldsymbol{H}_T, A_T = k) + \epsilon_T^*. && \text{(by (SNMM)}_T\text{)}
\end{aligned}$$

In particular, $\mathbb{E}(\epsilon_T^* | \boldsymbol{H}_{T-1}, A_{T-1}) = \mathbb{E}[\mathbb{E}(\epsilon_T^* | \boldsymbol{H}_T, A_T) | \boldsymbol{H}_{T-1}, A_{T-1}] = 0$ by $(\mathrm{SNMM})_T$. Then

$$\begin{aligned}
\mathbb{E}(Y_{T-1}^{(\mathrm{g})} | \boldsymbol{H}_{T-1}, A_{T-1}) &= \mathbb{E}\Bigg\{ \underbrace{\max_{1 \leqslant k \leqslant K} \mathbb{E}(Y_T^* | \boldsymbol{H}_T, A_T = k)}_{= \mathcal{V}_T(\boldsymbol{H}_T)} \Bigg| \boldsymbol{H}_{T-1}, A_{T-1} \Bigg\} \\
&= \mathbb{E}(Y_{T-1}^* | \boldsymbol{H}_{T-1}, A_{T-1}). && \text{(by Lemma 4.1)}
\end{aligned}$$

For Mathematical Induction on stage $t = T-1, T-2, \cdots, 1$, we assume

$$\mathbb{E}(Y_t^{(\mathrm{g})} | \boldsymbol{H}_t, A_t) = \mathbb{E}(Y_t^* | \boldsymbol{H}_t, A_t), \tag{4.20}$$

which we have proven for stage $t = T-1$ as above.

Next, for stage $t-1$, by definition,

$$\begin{aligned}
Y_{t-1}^{(\mathrm{g})} &= Y - \sum_{u=t}^{T} \left\{ \gamma_u(\boldsymbol{H}_u, A_u; \boldsymbol{\beta}_u) - \max_{1 \leqslant k \leqslant K} \gamma_u(\boldsymbol{H}_u, k; \boldsymbol{\beta}_u) \right\} \\
&= Y_t^{(\mathrm{g})} - \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) + \max_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t).
\end{aligned}$$

Conditional on $(\boldsymbol{H}_{t-1}, A_{t-1})$, we further have

$$
\begin{aligned}
&\mathbb{E}(Y_{t-1}^{(\mathrm{g})}|\boldsymbol{H}_{t-1}, A_{t-1}) \\
=&\mathbb{E}\left\{Y_t^{(\mathrm{g})} - \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) + \max_{1\leqslant k\leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t)\middle|\boldsymbol{H}_{t-1}, A_{t-1}\right\} \\
=&\mathbb{E}\left\{Y_t^{*} - \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) + \max_{1\leqslant k\leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t)\middle|\boldsymbol{H}_{t-1}, A_{t-1}\right\} && \text{(by induction hypothesis (4.20))} \\
=&\mathbb{E}\left\{\mu_t(\boldsymbol{H}_t) + \max_{1\leqslant k\leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) + \epsilon_t^{*}\middle|\boldsymbol{H}_{t-1}, A_{t-1}\right\} && \text{(by } (\mathrm{SNMM})_t\text{)} \\
=&\mathbb{E}\Big\{\underbrace{\max_{1\leqslant k\leqslant K} \mathbb{E}(Y_t^{*}|\boldsymbol{H}_t, A_t = k)}_{\mathcal{V}_t(\boldsymbol{H}_t)}\Big|\boldsymbol{H}_{t-1}, A_{t-1}\Big\} && \text{(by } (\mathrm{SNMM})_t\text{)} \\
=&\mathbb{E}(Y_{t-1}^{*}|\boldsymbol{H}_{t-1}, A_{t-1}). && \text{(by Lemma 4.1)}
\end{aligned}
$$

This proves the induction hypothesis (4.20) at stage $t-1$.

By Mathematical Induction, the induction hypothesis (4.20) holds for $1 \leqslant t \leqslant T-1$. □

4.6.3.3 Proof of Corollary 4.5

Proof of Corollary 4.5. First, we assume (4.6). By Lemma 4.3, $(\mathrm{SNMM})_1^T(\boldsymbol{\beta}_{1:T})$ is equivalent to: there exists some treatment-free effect functions $\{\mu_t(\boldsymbol{H}_t)\}_{t=1}^T$ such that

$$\mathbb{E}(Y_t^{(\mathrm{g})}|\boldsymbol{H}_t, A_t) = \mu_t(\boldsymbol{H}_t) + \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t); \quad 1 \leqslant t \leqslant T.$$

Under (4.6), for $1 \leqslant t \leqslant T$, we have

$$
\begin{aligned}
Y_t^{(\mathrm{g})} =& Y - \sum_{u=t+1}^{T}\left[\gamma_u(\boldsymbol{H}_u, A_u; \boldsymbol{\beta}_u) - \max_{1\leqslant k\leqslant K}\gamma_u(\boldsymbol{H}_u, k; \boldsymbol{\beta}_u)\right] \\
=& \underbrace{\mathcal{V}_0 + \sum_{s=1}^{t-1}\left\{\Delta\mathcal{M}_s(\boldsymbol{H}_s) - \max_{1\leqslant k\leqslant K}\gamma_s(\boldsymbol{H}_s, k; \boldsymbol{\beta}_s) + \gamma_s(\boldsymbol{H}_s, A_s; \boldsymbol{\beta}_s)\right\} + \Delta\mathcal{M}_t(\boldsymbol{H}_t) - \max_{1\leqslant k\leqslant K}\gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t)}_{:=\mu_t(\boldsymbol{H}_t)} \\
&+ \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) + \underbrace{\sum_{u=t+1}^{T+1}\Delta\mathcal{M}_u(\boldsymbol{H}_u)}_{:=e_t^{(\mathrm{g})}},
\end{aligned}
$$

where $\mathbb{E}(e_t^{(\mathrm{g})}|\boldsymbol{H}_t, A_t) = \sum_{u=t+1}^{T+1}\mathbb{E}[\Delta\mathcal{M}_u(\boldsymbol{H}_u)|\boldsymbol{H}_t, A_t] = 0$. Therefore, (4.6) implies $(\mathrm{SNMM})_1^T$. In particular, $\mu_t(\boldsymbol{H}_t)$ and $e_t^{(\mathrm{g})}$ can be defined from above.

Next, we assume Model $(\text{SNMM})_1^T$. Define

$$Y_0^{(\mathrm{g})} := Y - \sum_{t=1}^{T} \left\{ \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) - \max_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) \right\}.$$

Then $Y_0^{(\mathrm{g})}$ can be further decomposed as:

$$Y_0^{(\mathrm{g})} = \underbrace{\mathbb{E}\left(Y_0^{(\mathrm{g})}\right)}_{:=\mathcal{V}_0} + \sum_{t=1}^{T} \Big[\underbrace{\mathbb{E}\left(Y_0^{(\mathrm{g})} \Big| \boldsymbol{H}_t, A_t\right) - \mathbb{E}\left(Y_0^{(\mathrm{g})} \Big| \boldsymbol{H}_{t-1}, A_{t-1}\right)}_{:=\Delta\mathcal{M}_t(\boldsymbol{H}_t, A_t)} \Big] + \underbrace{Y_0^{(\mathrm{g})} - \mathbb{E}\left(Y_0^{(\mathrm{g})} \Big| \boldsymbol{H}_T, A_T\right)}_{:=\Delta\mathcal{M}_{T+1}(\boldsymbol{H}_{T+1})}.$$

For $1 \leqslant t \leqslant T$, the stage-$t$ working pseudo outcome can be represented as:

$$\begin{aligned}
Y_t^{(\mathrm{g})} =& Y - \sum_{u=t+1}^{T} \left\{ \gamma_u(\boldsymbol{H}_u, A_u; \boldsymbol{\beta}_u) - \max_{1 \leqslant k \leqslant K} \gamma_u(\boldsymbol{H}_u, k; \boldsymbol{\beta}_u) \right\} \\
=& Y_0^{(\mathrm{g})} + \sum_{s=1}^{t-1} \left\{ \gamma_s(\boldsymbol{H}_s, A_s; \boldsymbol{\beta}_s) - \max_{1 \leqslant k \leqslant K} \gamma_s(\boldsymbol{H}_s, k; \boldsymbol{\beta}_s) \right\} - \max_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t) + \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) \\
=& \underbrace{m_0 + \sum_{s=1}^{t-1} \left\{ \Delta\mathcal{M}_s(\boldsymbol{H}_s, A_s) + \gamma_s(\boldsymbol{H}_s, A_s; \boldsymbol{\beta}_s) - \max_{1 \leqslant k \leqslant K} \gamma_s(\boldsymbol{H}_s, k; \boldsymbol{\beta}_s) \right\} + \Delta\mathcal{M}_t(\boldsymbol{H}_t, A_t) - \max_{1 \leqslant k \leqslant K} \gamma_t(\boldsymbol{H}_t, k; \boldsymbol{\beta}_t)}_{\text{function of } (\boldsymbol{H}_t, A_t), \text{ while depending on } A_t \text{ only through } \Delta\mathcal{M}_t(\boldsymbol{H}_t, A_t)} \\
& + \gamma_t(\boldsymbol{H}_t, A_t; \boldsymbol{\beta}_t) + \underbrace{\sum_{u=t+1}^{T+1} \Delta\mathcal{M}_u(\boldsymbol{H}_u, A_u)}_{:=e_t^{(\mathrm{g})}}.
\end{aligned}$$

By construction, we have $\mathbb{E}[\Delta\mathcal{M}_t(\boldsymbol{H}_t, A_t) | \boldsymbol{H}_{t-1}, A_{t-1}] = 0$ for $1 \leqslant t \leqslant T$, which implies that $\mathbb{E}(e_t^{(\mathrm{g})} | \boldsymbol{H}_t, A_t) = 0$ for $1 \leqslant t \leqslant T$. Then $(\text{SNMM})_1^T$ and Theorem 4.4 together imply that

$$\Delta\mathcal{M}_t(\boldsymbol{H}_t, A_t) = \Delta\mathcal{M}_t(\boldsymbol{H}_t); \quad 1 \leqslant t \leqslant T.$$

Therefore, $Y_0^{(\mathrm{g})}$ satisfies the following restrictions:

$$Y_0^{(\mathrm{g})} = \mathcal{V}_0 + \sum_{t=1}^{T+1} \Delta\mathcal{M}_t(\boldsymbol{H}_t) \quad \text{subject to} \quad \mathbb{E}[\Delta\mathcal{M}_t(\boldsymbol{H}_t) | \boldsymbol{H}_{t-1}, A_{t-1}] = 0; \quad 1 \leqslant t \leqslant T+1,$$

which gives (4.6).

□

4.6.3.4 Proof of Lemma 4.10

Proof of Lemma 4.10. By definition,

$$\mathring{\Lambda}_t^{\perp} = \left\{ \frac{\mathsf{G}_t(\boldsymbol{H}_t)\vec{\boldsymbol{A}}_t e_t^{(\mathrm{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \middle| \mathsf{G}_t : \mathcal{H}_t \to \mathbb{R}^{p\times(K-1)} \right\}.$$

For any $\boldsymbol{G} \in \mathcal{G}$, we have

$$\begin{aligned}
\mathbb{E}(\boldsymbol{G}|\mathring{\Lambda}_t^{\perp}) &= \mathbb{E}\left[\frac{\boldsymbol{G}\vec{\boldsymbol{A}}_t^{\intercal} e_t^{(\mathrm{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \middle| \boldsymbol{H}_t \right] \mathbb{E}\left[\left(\frac{\vec{\boldsymbol{A}}_t e_t^{(\mathrm{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \right)^{\otimes 2} \middle| \boldsymbol{H}_t \right]^{-1} \frac{\vec{\boldsymbol{A}}_t e_t^{(\mathrm{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)} \\
&= \left[\sum_{k=1}^{K} \mathbb{E}\left(\boldsymbol{G} e_t^{(\mathrm{ort})} \middle| \boldsymbol{H}_t, A_t = k \right) \boldsymbol{\omega}_k^{\intercal} \right] \frac{\mathsf{V}_t(\boldsymbol{H}_t)^{-1}\vec{\boldsymbol{A}}_t e_t^{(\mathrm{ort})}}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}. \qquad (4.21)
\end{aligned}$$

For Mathematical Induction on stage $t = T, T-1, T-2, \cdots, 1$, we assume

$$\{\mathring{\Lambda}_u^{\perp}\}_{u=t}^{T} \text{ are mutually orthogonal and } \Lambda_{t:T}^{\perp} = \bigoplus_{u=t}^{T} \mathring{\Lambda}_u^{\perp}. \qquad (4.22)$$

Then the induction hypothesis (4.22) holds for $t = T$ by definition.

Consider the induction hypothesis (4.22) at stage $t-1$. Fix $1 \leqslant s \leqslant t-1$. Let $\mathsf{G}_s : \mathcal{H}_s \to \mathbb{R}^{p\times(K-1)}$ be an arbitrary function. First, we consider $\boldsymbol{G}_s^{(\mathrm{g})} := \frac{\mathsf{G}_s(\boldsymbol{H}_s)\vec{\boldsymbol{A}}_s e_s^{(\mathrm{g})}}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)} \in \Lambda_s^{\perp}$. For $t \leqslant u \leqslant T$, we have

$$\begin{aligned}
\mathbb{E}(\boldsymbol{G}_s^{(\mathrm{g})}|\mathring{\Lambda}_u^{\perp}) &= \mathbb{E}\left[\frac{\mathsf{G}_s(\boldsymbol{H}_s)\vec{\boldsymbol{A}}_s e_s^{(\mathrm{g})}}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)} \middle| \mathring{\Lambda}_u^{\perp} \right] \\
&= \left\{ \sum_{k=1}^{K} \mathbb{E}\left[\frac{\mathsf{G}_s(\boldsymbol{H}_s)\vec{\boldsymbol{A}}_s e_s^{(\mathrm{g})} e_u^{(\mathrm{ort})}}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)} \middle| \boldsymbol{H}_u, A_u = k \right] \boldsymbol{\omega}_k^{\intercal} \right\} \frac{\mathsf{V}_u(\boldsymbol{H}_u)^{-1}\vec{\boldsymbol{A}}_u e_u^{(\mathrm{ort})}}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)} \qquad \text{(by (4.21))} \\
&= \frac{\mathsf{G}_s(\boldsymbol{H}_s)\vec{\boldsymbol{A}}_s}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)} \left\langle \sum_{k=1}^{K} \mathbb{E}\left(e_s^{(\mathrm{g})} e_u^{(\mathrm{ort})} \middle| \boldsymbol{H}_u, A_u = k \right) \boldsymbol{\omega}_k, \frac{\mathsf{V}_u(\boldsymbol{H}_u)^{-1}\vec{\boldsymbol{A}}_u}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)} \right\rangle e_u^{(\mathrm{ort})} \\
&= \frac{\mathsf{G}_s(\boldsymbol{H}_s)\vec{\boldsymbol{A}}_s}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)} \left\langle \sum_{k=1}^{K} \rho_u(\boldsymbol{H}_u, k)\boldsymbol{\omega}_k, \frac{\mathsf{V}_u(\boldsymbol{H}_u)^{-1}\vec{\boldsymbol{A}}_u}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)} \right\rangle e_u^{(\mathrm{ort})} \\
&= \frac{\mathsf{G}_s(\boldsymbol{H}_s)\vec{\boldsymbol{A}}_s}{p_{\mathscr{A},s}(A_s|\boldsymbol{H}_s)} \tau_u e_u^{(\mathrm{ort})}.
\end{aligned}$$

Here, the second last equality follows from Proposition 4.5 that for $1 \leqslant s \leqslant u \leqslant T$, we have $e_s^{(\mathrm{g})} = \sum_{t=s+1}^{T+1} \Delta\mathcal{M}_t(\boldsymbol{H}_t) = \sum_{t=s+1}^{u} \Delta\mathcal{M}_t(\boldsymbol{H}_t) + e_u^{(\mathrm{g})}$, so that

$$\mathbb{E}\left(e_s^{(\mathrm{g})} e_u^{(\mathrm{ort})} \middle| \boldsymbol{H}_u, A_u\right) = \sum_{t=s+1}^{u} \Delta\mathcal{M}_t(\boldsymbol{H}_t) \underbrace{\mathbb{E}\left(e_u^{(\mathrm{ort})} \middle| \boldsymbol{H}_u, A_u\right)}_{=0} + \mathbb{E}\left(e_u^{(\mathrm{g})} e_u^{(\mathrm{ort})} \middle| \boldsymbol{H}_u, A_u\right) = \rho_u(\boldsymbol{H}_u, A_u).$$

Next, we specifically consider $s = t-1$ and $\boldsymbol{G}_{t-1}^{(\mathrm{ort})} := \boldsymbol{G}_{t-1}^{(\mathrm{g})} - \sum_{u=t}^{T} \mathbb{E}(\boldsymbol{G}_{t-1}^{(\mathrm{g})} | \mathring{\Lambda}_u^{\perp}) = \frac{\mathsf{G}_{t-1}(\boldsymbol{H}_{t-1}) \vec{\boldsymbol{A}}_{t-1} e_{t-1}^{(\mathrm{ort})}}{p_{\mathscr{A},t-1}(A_{t-1}|\boldsymbol{H}_{t-1})}$. It can be clear that $\boldsymbol{G}_{t-1}^{(\mathrm{ort})} \perp \bigoplus_{u=t}^{T} \mathring{\Lambda}_u^T$. By G_{t-1} is arbitrary, we further have $\mathring{\Lambda}_{t-1}^{\perp} \perp \bigoplus_{u=t}^{T} \mathring{\Lambda}_u^T$, and $\Lambda_{t-1} \subseteq \bigoplus_{u=t-1}^{T} \mathring{\Lambda}_u^T$. Then by induction hypothesis (4.22) that $\Lambda_{t:T}^{\perp} = \bigoplus_{u=t}^{T} \mathring{\Lambda}_u^{\perp}$, we have $\Lambda_{(t-1):T}^{\perp} \subseteq \bigoplus_{u=t-1}^{T} \mathring{\Lambda}_u^{\perp}$. Conversely, $\mathring{\Lambda}_{t-1}^{\perp} \subseteq \overline{\mathsf{span}}\{\Lambda_{t-1}^{\perp}, \mathring{\Lambda}_{t:T}^{\perp}\} = \Lambda_{(t-1):T}^{\perp}$. Therefore, $\Lambda_{(t-1):T}^{\perp} = \bigoplus_{u=t-1}^{T} \mathring{\Lambda}_u^{\perp}$, and the induction hypothesis (4.22) at stage $t-1$ is proved.

By Mathematical Induction, the induction hypothesis (4.22) holds for $1 \leqslant t \leqslant T$,

Assume the additional assumption:

$$\mathbb{E}\left(e_t^{(\mathrm{g})2} \middle| \boldsymbol{H}_t, A_t\right) = \mathbb{E}\left(e_t^{(\mathrm{g})2} \middle| \boldsymbol{H}_t\right); \quad 1 \leqslant t \leqslant T. \tag{4.23}$$

We aim to show the following by Mathematical Induction:

$$e_t^{(\mathrm{ort})} = e_t^{(\mathrm{g})}; \quad \rho_t(\boldsymbol{H}_t, 1) = \cdots = \rho_t(\boldsymbol{H}_t, K) = \mathbb{E}\left(e_t^{(\mathrm{g})2} \middle| \boldsymbol{H}_t\right). \tag{4.24}$$

By definition, $e_T^{(\mathrm{ort})} = e_T^{(\mathrm{g})}$ and $\rho_T(\boldsymbol{H}_T, k) = \mathbb{E}(e_T^{(\mathrm{g})2} | \boldsymbol{H}_T, A_T = k)$. Then by (4.23), we have $\rho_T(\boldsymbol{H}_T, 1) = \cdots = \rho_T(\boldsymbol{H}_T, K) = \mathbb{E}(e_T^{(\mathrm{g})2} | \boldsymbol{H}_T)$, which proves the induction hypothesis (4.24) at $t = T$.

Assume the induction hypothesis (4.24) holds for stages $t, t+1, \cdots, T$. Notice that the second part of the induction hypothesis (4.24) implies that $\tau_u = 0$ for $t \leqslant u \leqslant T$. Then for stage $t-1$, we have by definition $e_{t-1}^{(\mathrm{ort})} = e_{t-1}^{(\mathrm{g})} - \sum_{u=t}^{T} \tau_u e_u^{(\mathrm{ort})} = e_{t-1}^{(\mathrm{g})}$, and $\rho_{t-1}(\boldsymbol{H}_{t-1}, k) = \mathbb{E}(e_{t-1}^{(\mathrm{g})2} | \boldsymbol{H}_{t-1}, A_{t-1} = k)$. By (4.23), we further have $\rho_{t-1}(\boldsymbol{H}_{t-1}, 1) = \cdots = \rho_{t-1}(\boldsymbol{H}_{t-1}, K) = \mathbb{E}(e_{t-1}^{(\mathrm{g})2} | \boldsymbol{H}_{t-1})$, which proves the induction hypothesis (4.24) at stage $t-1$.

By Mathematical induction, the induction hypothesis (4.24) holds for $1 \leqslant t \leqslant T$. Combining (4.24) for $1 \leqslant t \leqslant T$ with (4.22) at $t = 1$, we have $\{\Lambda_t^{\perp}\}_{t=1}^{T}$ are mutually orthogonal and $\Lambda_{1:T}^{\perp} = \bigoplus_{t=1}^{T} \Lambda_t^{\perp}$. □

4.6.3.5 Proof of Theorem 4.11

Proof of Theorem 4.11. The general characterization of the G-Estimation IF in Robins (2004, Equation (3.10)) is

$$\sum_{t=1}^{T}\Big\{Y_t^{(\mathrm{g})}-\gamma_t^{(\mathrm{g})}(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)-\mathbb{E}[Y_t^{(\mathrm{g})}-\gamma_t^{(\mathrm{g})}(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)|\boldsymbol{H}_t]\Big\}\Big\{\boldsymbol{G}_t(\boldsymbol{H}_t,A_t)-\mathbb{E}[\boldsymbol{G}_t(\boldsymbol{H}_t,A_t)|\boldsymbol{H}_t]\Big\} \tag{4.25}$$

for some instrument functions $\boldsymbol{G}_t : \mathcal{H}_t \to \mathbb{R}^p$ $(1 \leqslant t \leqslant T)$. This can be related to the form in our Theorem 4.11 as follows. We first replace $\boldsymbol{G}_t(\boldsymbol{H}_t,A_t)-\mathbb{E}[\boldsymbol{G}_t(\boldsymbol{H}_t,A_t)|\boldsymbol{H}_t]$ by $\frac{\mathsf{G}_t(\boldsymbol{H}_t)\vec{\boldsymbol{A}}_t}{p_{\mathscr{A},t}(A_t|\boldsymbol{H}_t)}$ with some $\mathsf{G}_t : \mathcal{H}_t \to \mathbb{R}^{p\times(K-1)}$, and then consider the block representation $\mathsf{G}_t(\boldsymbol{H}_t) := [\mathsf{G}_{1t}(\boldsymbol{H}_t)^\intercal,\cdots,\mathsf{G}_{Tt}(\boldsymbol{H}_t)^\intercal]^\intercal$ for $\mathsf{G}_{st} : \mathcal{H}_t \to \mathbb{R}^{p_s\times(K-1)}$ $(1 \leqslant s \leqslant T)$. The function $\boldsymbol{H}_t \mapsto \mathbb{E}[Y_t^{(\mathrm{g})}-\gamma_t^{(\mathrm{g})}(\boldsymbol{H}_t,A_t;\boldsymbol{\beta}_t)|\boldsymbol{H}_t]$ at the true parameter $\boldsymbol{\beta}_t$ is further replaced by the treatment-free effect function $\mu_t(\boldsymbol{H}_t)$. Then (4.25) is equivalent to the block matrix form

$$\begin{bmatrix}\mathsf{G}_{11}(\boldsymbol{H}_1) & \mathsf{G}_{12}(\boldsymbol{H}_2) & \cdots & \mathsf{G}_{1T}(\boldsymbol{H}_1)\\ \mathsf{G}_{21}(\boldsymbol{H}_1) & \mathsf{G}_{22}(\boldsymbol{H}_2) & \cdots & \mathsf{G}_{2T}(\boldsymbol{H}_T)\\ \vdots & \vdots & \ddots & \vdots\\ \mathsf{G}_{T1}(\boldsymbol{H}_1) & \mathsf{G}_{T2}(\boldsymbol{H}_2) & \cdots & \mathsf{G}_{TT}(\boldsymbol{H}_T)\end{bmatrix}\begin{bmatrix}\frac{\vec{\boldsymbol{A}}_1 e_1^{(\mathrm{g})}}{p_{\mathscr{A},1}(A_1|\boldsymbol{H}_1)}\\ \frac{\vec{\boldsymbol{A}}_2 e_2^{(\mathrm{g})}}{p_{\mathscr{A},2}(A_2|\boldsymbol{H}_2)}\\ \vdots\\ \frac{\vec{\boldsymbol{A}}_T e_T^{(\mathrm{g})}}{p_{\mathscr{A},T}(A_T|\boldsymbol{H}_T)}\end{bmatrix}.$$

By $\mathsf{G} = [\mathsf{G}_{st} : 1 \leqslant s,t \leqslant T]$ is arbitrary, (4.25) is equivalent to (4.25) with $\{e_t^{(\mathrm{g})}\}_{t=1}^T$ replaced by $\{e_t^{(\mathrm{ort})}\}_{t=1}^T$. The restriction $\mathbb{E}[\mathsf{IF}(\mathsf{G})\boldsymbol{S}^\intercal] = \mathbf{I}_{p\times p}$ follows from (Tsiatis, 2007, Theorem 4.2 (i)).

Next, we consider $\mathsf{IF}(\mathsf{L})$. In order to show $\mathsf{IF}(\mathsf{L})$ is also an IF, it suffices to show $\mathbb{E}[\mathsf{IF}(\mathsf{L})\boldsymbol{S}^\intercal] = \mathbf{I}_{p\times p}$. Denote $st(\mathsf{G}) = (\mathsf{IF}_1^\intercal,\mathsf{IF}_2^\intercal,\cdots,\mathsf{IF}_T^\intercal)^\intercal(\mathsf{G})$ with $\mathsf{IF}_s(\mathsf{G}) := \sum_{t=1}^T \mathsf{G}_{st}(\boldsymbol{H}_t)\frac{\vec{\boldsymbol{A}}_t e_t^{(\mathrm{ort})}}{p_{\mathscr{A}}(A_t|\boldsymbol{H}_t)}$, and $\boldsymbol{S} = (\boldsymbol{S}_1^\intercal,\boldsymbol{S}_2^\intercal\cdots,\boldsymbol{S}_T^\intercal)^\intercal$ with $\boldsymbol{S}_t := (\partial/\partial\boldsymbol{\beta}_t)\log[\texttt{likelihood}(\boldsymbol{\beta}_{1:T})]$. Consider some $1 \leqslant u,t \leqslant T$. By $\mathbb{E}_{\boldsymbol{\beta}_1^T}[e_u^{(\mathrm{ort})}(\boldsymbol{\beta}_{u:T})|\boldsymbol{H}_u,A_u] = 0$ for any $\boldsymbol{\beta}_t \in \mathscr{B}_t$, we have $\mathbf{0} = (\partial/\partial\boldsymbol{\beta}_t)\mathbb{E}_{\boldsymbol{\beta}_1^T}[e_u^{(\mathrm{ort})}(\boldsymbol{\beta}_{u:T})|\boldsymbol{H}_u,A_u] = \mathbb{E}_{\boldsymbol{\beta}_{1:T}}[(\partial/\partial\boldsymbol{\beta}_t)e_u^{(\mathrm{ort})}(\boldsymbol{\beta}_{t:T})|\boldsymbol{H}_u,A_u]+\mathbb{E}_{\boldsymbol{\beta}_{1:T}}[e_u^{(\mathrm{ort})}(\boldsymbol{\beta}_{u:T})\boldsymbol{S}_t|\boldsymbol{H}_u,A_u]$. In particular, $\mathbb{E}[e_u^{(\mathrm{ort})}\boldsymbol{S}_t|\boldsymbol{H}_u,A_u] = \mathbb{E}[-(\partial/\partial\boldsymbol{\beta}_t)e_u^{(\mathrm{ort})}(\boldsymbol{\beta}_{u:T})|\boldsymbol{H}_u,A_u] = 0$ for $1 \leqslant t < u \leqslant T$. Then

$$\mathbb{1}(s=t)\mathbf{I} = \mathbb{E}[\mathsf{IF}_s(\mathsf{G})\boldsymbol{S}_t^\intercal] = \mathbb{E}\left[\sum_{u=1}^{T}\frac{\mathsf{G}_{su}(\boldsymbol{H}_u)\vec{\boldsymbol{A}}_u}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}e_u^{(\mathrm{ort})}\boldsymbol{S}_t^\intercal\right] = \mathbb{E}\left[\sum_{u=1}^{t}\frac{\mathsf{G}_{su}(\boldsymbol{H}_u)\vec{\boldsymbol{A}}_u}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}e_u^{(\mathrm{ort})}\boldsymbol{S}_t^\intercal\right] \quad (1 \leqslant s,t \leqslant T)$$

$$\Rightarrow \quad \mathbb{1}(s=t)\mathbf{I} = \mathbb{E}[\mathsf{IF}_s(\mathsf{L})\boldsymbol{S}_t^\intercal] = \mathbb{E}\left[\sum_{u=1}^{s}\frac{\mathsf{G}_{su}(\boldsymbol{H}_u)\vec{\boldsymbol{A}}_u}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}e_u^{(\mathrm{ort})}\boldsymbol{S}_t^\intercal\right] = \mathbb{E}\left[\sum_{u=1}^{s\wedge t}\frac{\mathsf{G}_{su}(\boldsymbol{H}_u)\vec{\boldsymbol{A}}_u}{p_{\mathscr{A},u}(A_u|\boldsymbol{H}_u)}e_u^{(\mathrm{ort})}\boldsymbol{S}_t^\intercal\right] \quad (1 \leqslant s,t \leqslant T).$$

Finally, we compare $\mathbb{E}[\mathsf{IF}(\mathsf{L})^{\otimes 2}] = \mathbb{E}(\mathsf{LDL}^\intercal)$ with $\mathbb{E}[\mathsf{IF}(\mathsf{G})^{\otimes 2}] = \mathbb{E}(\mathsf{GDG}^\intercal)$, where $\mathsf{D} := \mathsf{diag}\{\mathsf{V}_t(\boldsymbol{H}_t)\}_{t=1}^T$. For ease of notation, we suppress the dependency on $\boldsymbol{H}_t$. Then we have

$$\mathbb{E}[\mathsf{IF}(\mathsf{G})^{\otimes 2}] - \mathbb{E}[\mathsf{IF}(\mathsf{L})^{\otimes 2}] = \sum_{t=1}^{T} (\mathsf{G}_t \mathsf{V}_t \mathsf{G}_t^\intercal - \mathsf{L}_t \mathsf{V}_t \mathsf{L}_t^\intercal) = \sum_{t=1}^{T} \mathsf{U}_t \mathsf{V}_t \mathsf{U}_t^\intercal \geqslant 0,$$

where $\mathsf{L}_t := [\mathbf{O}, \cdots, \mathbf{O}, \mathsf{G}_{tt}^\intercal, \cdots, \mathsf{G}_{Tt}^\intercal]^\intercal$ and $\mathsf{U}_t := \mathsf{G}_t - \mathsf{L}_t = [\mathsf{G}_{1t}, \cdots, \mathsf{G}_{t-1,t}, \mathbf{O}, \cdots, \mathbf{O}]^\intercal$. This concludes $\mathbb{E}[\mathsf{IF}(\mathsf{L})^{\otimes 2}] \leqslant \mathbb{E}[\mathsf{IF}(\mathsf{G})^{\otimes 2}]$. □

BIBLIOGRAPHY

Abrevaya, J., Hsu, Y.-C., and Lieli, R. P. (2015). Estimating conditional average treatment effects. *Journal of Business & Economic Statistics*, 33(4):485–505.

Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Springer, Cham.

Alaa, A. M. and van der Schaar, M. (2017). Bayesian inference of individualized treatment effects using multi-task gaussian processes. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Alaa, A. M. and van der Schaar, M. (2018). Limits of estimating heterogeneous treatment effects: Guidelines for practical algorithm design. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 129–138. PMLR.

Allen, G. I. (2013). Automatic feature selection via weighted kernels and regularization. *Journal of Computational and Graphical Statistics*, 22(2):284–299.

Almirall, D., Griffin, B. A., McCaffrey, D. F., Ramchand, R., Yuen, R. A., and Murphy, S. A. (2014). Time-varying effect moderation using the structural nested mean model: estimation using inverse-weighted regression with residuals. *Statistics in Medicine*, 33(20):3466–3487.

Almirall, D., Ten Have, T., and Murphy, S. A. (2010). Structural nested mean models for assessing time-varying effect moderation. *Biometrics*, 66(1):131–139.

Alyass, A., Turcotte, M., and Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Medical Genomics*, 8(33):1–12.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

Bacon, M. C., Von Wyl, V., Alden, C., Sharp, G., Robison, E., Hessol, N., Gange, S., Barranday, Y., Holman, S., and Weber, K. (2005). The women's interagency HIV study: An observational cohort brings clinical sciences to the bench. *Clinical and Diagnostic Laboratory Immunology*, 12(9):1013–1019.

Bartlett, P. L. and Mendelson, S. (2002). Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482.

Bellman, R. (1966). Dynamic programming. *Science*, 153(3731):34–37.

Bembom, O. and van der Laan, M. J. (2008). Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Statistics in Medicine*, 27(19):3689–3716.

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

Benkeser, D., Carone, M., van der Laan, M. J., and Gilbert, P. B. (2017). Doubly robust nonparametric inference on the average treatment effect. *Biometrika*, 104(4):863–880.

Bennett, A. and Kallus, N. (2020a). Efficient policy learning from surrogate-loss classification reductions. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 788–798. PMLR.

Bennett, A. and Kallus, N. (2020b). The variational method of moments. *arXiv preprint arXiv:2012.09422*.

Bennett, M., Vielma, J. P., and Zubizarreta, J. R. (2020). Building representative matched samples with multi-valued treatments in large observational studies. *Journal of Computational and Graphical Statistics*, 29(4):744–757.

Bertsimas, D., Dunn, J., and Mundru, N. (2019). Optimal prescriptive trees. *INFORMS Journal on Optimization*, 1(2):164–183.

Bertsimas, D. and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044.

Bertsimas, D., Kallus, N., Weinstein, A. M., and Zhuo, Y. D. (2017). Personalized diabetes management using electronic medical records. *Diabetes Care*, 40(2):210–217.

Bertsimas, D. and Koduri, N. (2021). Data-driven optimization: A reproducing kernel Hilbert space approach. *Operations Research*. To appear.

Beygelzimer, A. and Langford, J. (2009). The offset tree for learning with partial labels. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, page 129–138, New York, NY, USA. Association for Computing Machinery.

Bickel, P. J. and Kwon, J. (2001). Inference for semiparametric models: some questions and an answer. *Statistica Sinica*, 11(4):863–886.

Blatt, D., Murphy, S. A., and Zhu, J. (2004). A-learning for approximate planning. Technical Report 04-63, The Methodology Center, Pennsylvania State University. `http://people.seas.harvard.edu/~samurphy/papers/Alearning2004.pdf`.

Boucheron, S., Bousquet, O., and Lugosi, G. (2005). Theory of classification: A survey of some recent advances. *ESAIM: Probability and Statistics*, 9:323–375.

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Mollan, K. R., Sax, P. E., Daar, E. S., Adimora, A. A., Eron, J. J., and Mugavero, M. J. (2018). Generalizing evidence from randomized trials using inverse probability of sampling weights. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4):1193–1209.

Cao, W., Tsiatis, A. A., and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96(3):723–734.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *The Annals of Statistics*, 10(4):1224–1233.

Chakraborty, B., Laber, E. B., and Zhao, Y.-Q. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*, 69(3):714–723.

Chakraborty, B., Murphy, S., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19(3):317–343.

Chen, G., Zeng, D., and Kosorok, M. R. (2016). Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association*, 111(516):1509–1521.

Chen, J., Fu, H., He, X., Kosorok, M. R., and Liu, Y. (2018). Estimating individualized treatment rules for ordinal treatments. *Biometrics*, 74(3):924–933.

Chen, S., Tian, L., Cai, T., and Yu, M. (2017). A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4):1199–1209.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018a). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2018b). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. Working Paper 24678, National Bureau of Economic Research. `http://www.nber.org/papers/w24678`.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. (2018c). Locally robust semiparametric estimation. cemmap working paper CWP30/18, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, London, UK.

Cressie, N. and Read, T. R. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 46(3):440–464.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2006). Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand. Working Paper 0330, National Bureau of Economic Research. `http://www.nber.org/papers/t0330`.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199.

Cui, Y., Zhu, R., and Kosorok, M. (2017). Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic Journal of Statistics*, 11(2):3927–3953.

Curth, A., Alaa, A. M., and van der Schaar, M. (2020). Estimating structural target functions using machine learning and influence functions. *arXiv preprint arXiv:2008.06461*.

Curth, A. and van der Schaar, M. (2021). Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms. In Banerjee, A. and Fukumizu, K., editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1810–1818. PMLR.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association*, 82(400):1079–1091.

Ding, P. and Li, F. (2018). Causal inference: A missing data perspective. *Statistical Science*, 33(2):214–237.

Dorie, V., Hill, J., Shalit, U., Scott, M., and Cervone, D. (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science*, 34(1):43–68.

Duchi, J. C., Hashimoto, T., and Namkoong, H. (2019). Distributionally robust losses against mixture covariate shifts. *arXiv preprint arXiv:2007.13982*.

Duchi, J. C. and Namkoong, H. (2018). Learning models with uniform performance via distributionally robust optimization. *arXiv preprint arXiv:1810.08750*.

Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. In Getoor, L. and Scheffer, T., editors, *Proceedings of the 28th International Conference on Machine Learning*, ICML '11, pages 1097–1104, New York, NY, USA. ACM.

Ertefaie, A., McKay, J. R., Oslin, D., and Strawderman, R. L. (2021). Robust Q-learning. *Journal of the American Statistical Association*, 116(533):368–381.

Ertefaie, A. and Strawderman, R. L. (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika*, 105(4):963–977.

Fan, J., Imai, K., Liu, H., Ning, Y., and Yang, X. (2020). *Optimal Covariate Balancing Conditions in Propensity Score Estimation*. Working Paper. `https://cpb-us-w2.wpmucdn.com/sites.coecis.cornell.edu/dist/3/72/files/2020/09/CBPStheory.pdf`.

Fong, C., Hazlett, C., Imai, K., et al. (2018). Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. *The Annals of Applied Statistics*, 12(1):156–177.

Foster, J. C., Taylor, J. M., and Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24):2867–2880.

Friedberg, R., Tibshirani, J., Athey, S., and Wager, S. (2020). Local linear forests. *Journal of Computational and Graphical Statistics*. To appear.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):1–67.

Fu, S., He, Q., Zhang, S., and Liu, Y. (2019). Robust outcome weighted learning for optimal individualized treatment rules. *Journal of Biopharmaceutical Statistics*, 29(4):606–624.

Gandhi, M., Ameli, N., Bacchetti, P., Sharp, G. B., French, A. L., Young, M., Gange, S. J., Anastos, K., Holman, S., Levine, A., and Greenblatt, R. M. (2005). Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols. *AIDS*, 19(16):1885–1896.

Gatsonis, C. and Morton, S. C. (2017). *Methods in Comparative Effectiveness Research*. Chapman & Hall/CRC Biostatistics Series. Chapman and Hall/CRC, New York, NY, USA.

Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *The Annals of Statistics*, 40(1):529–560.

Goldberg, Y., Song, R., and Kosorok, M. R. (2013). Adaptive Q-learning. In Banerjee, M., Bunea, F., Huang, J., Koltchinskii, V., and Maathuis, M. H., editors, *From Probability to Statistics and Back: High-Dimensional Models and Processes – A Festschrift in Honor of Jon A. Wellner*, volume 9 of *Institute of Mathematical Statistics Collections*, pages 150–162. Institute of Mathematical Statistics.

Gu, C. (2013). *Smoothing Spline ANOVA Models*. Springer, Verlag, NY, USA, second edition.

Guo, R., Cheng, L., Li, J., Hahn, P. R., and Liu, H. (2020). A survey of learning causality with data: Problems and methods. *ACM Computing Surveys*, 53(4). Article 75.

Guo, W., Zhou, X.-H., and Ma, S. (2021). Estimation of optimal individualized treatment rules using a covariate-specific treatment effect curve with high-dimensional covariates. *Journal of the American Statistical Association*, 116(533):309–321.

Hahn, P. R., Murray, J. S., and Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis*, 15(3):965–1056.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S., and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090.

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science*, 21(1):1–14.

Henderson, R., Ansell, P., and Alshibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics*, 66(4):1192–1201.

Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., and Smeeth, L. (2015). Data resource profile: Clinical practice research datalink (CPRD). *International Journal of Epidemiology*, 44(3):827–836.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.

Hirshberg, D. A. and Wager, S. (2017). Augmented minimax linear estimation. *arXiv preprint arXiv:1712.00038*.

Hotz, V. J., Imbens, G. W., and Mortimer, J. H. (2005). Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics*, 125(1-2):241–270.

Huang, X., Goldberg, Y., and Xu, J. (2019). Multicategory individualized treatment regime using outcome weighted learning. *Biometrics*, 75(4):1216–1227.

Huang, X., Ning, J., and Wahed, A. S. (2014). Optimization of individualized dynamic treatment regimes for recurrent diseases. *Statistics in Medicine*, 33(14):2363–2378.

Huang, Y. and Fong, Y. (2014). Identifying optimal biomarker combinations for treatment selection via a robust kernel method. *Biometrics*, 70(4):891–901.

Huling, J. D. and Yu, M. (2018). Subgroup identification using the personalized package. *arXiv preprint arXiv:1809.07905*.

Imai, K. and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.

Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263.

Imai, K. and Ratkovic, M. (2015). Robust estimation of inverse probability weights for marginal structural models. *Journal of the American Statistical Association*, 110(511):1013–1023.

Jeng, X. J., Lu, W., and Peng, H. (2018). High-dimensional inference for personalized treatment decision. *Electronic Journal of Statistics*, 12(1):2074–2089.

Jiang, B., Song, R., Li, J., and Zeng, D. (2019). Entropy learning for dynamic treatment regimes. *Statistica Sinica*, 29(4):1633–1655.

Jiang, N. and Li, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In Balcan, M. F. and Weinberger, K. Q., editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 652–661, New York, NY, USA. PMLR.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics*, 43(6):2451–2483.

Johansson, F., Shalit, U., and Sontag, D. (2016). Learning representations for counterfactual inference. In Balcan, M. F. and Weinberger, K. Q., editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 3020–3029, New York, NY, USA. PMLR.

Johansson, F. D., Kallus, N., Shalit, U., and Sontag, D. (2018). Learning weighted representations for generalization across designs. *arXiv preprint arXiv:1802.08598*.

Johansson, F. D., Shalit, U., Kallus, N., and Sontag, D. (2020). Generalization bounds and representation learning for estimation of potential outcomes and causal effects. *arXiv preprint arXiv:2001.07426*.

Johnson, B. A., Lin, D., and Zeng, D. (2008). Penalized estimating functions and variable selection in semiparametric regression models. *Journal of the American Statistical Association*, 103(482):672–680.

Josey, K. P., Juarez-Colunga, E., Yang, F., and Ghosh, D. (2020). A framework for covariate balance using Bregman distances. *Scandinavian Journal of Statistics*. To appear.

Kallus, N. (2017). Recursive partitioning for personalization using observational data. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1789–1798. PMLR.

Kallus, N. (2018). Balanced policy evaluation and learning. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Kallus, N. (2020). Generalized optimal matching methods for causal inference. *Journal of Machine Learning Research*, 21(62):1–54.

Kallus, N., Pennicooke, B., and Santacatterina, M. (2021). More robust estimation of average treatment effects using kernel optimal matching in an observational study of spine surgical interventions. *Statistics in Medicine*. To appear.

Kallus, N. and Uehara, M. (2020). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. *Journal of Machine Learning Research*, 21(167):1–63.

Kallus, N. and Zhou, A. (2018). Policy evaluation and optimization with continuous treatments. In Storkey, A. and Perez-Cruz, F., editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1243–1251. PMLR.

Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539.

Kennedy, E. H. (2020). Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1229–1245.

Kitagawa, T. and Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616.

Krokhmal, P. A. (2007). Higher moment coherent risk measures. *Quantitative Finance*, 7(4):373–387.

Kube, A., Das, S., and Fowler, P. J. (2019). Allocating interventions based on predicted outcomes: A case study on homelessness services. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1):622–629.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165.

Laber, E. B., Linn, K. A., and Stefanski, L. A. (2014a). Interactive model building for Q-learning. *Biometrika*, 101(4):831–847.

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E., and Murphy, S. A. (2014b). Dynamic treatment regimes: Technical challenges and applications. *Electronic Journal of Statistics*, 8(1):1225.

Laber, E. B. and Zhao, Y.-Q. (2015). Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514.

Lafferty, J. and Wasserman, L. (2008). Rodeo: Sparse, greedy nonparametric regression. *The Annals of Statistics*, 36(1):28–63.

Li, F. and Li, F. (2019). Propensity score weighting for causal inference with multiple treatments. *The Annals of Applied Statistics*, 13(4):2389–2415.

Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400.

Li, S., Cai, T. T., and Li, H. (2020). Transfer learning for high-dimensional linear regression: Prediction, estimation, and minimax optimality. *arXiv preprint arXiv:2006.10593*.

Lian, H., Liang, H., and Carroll, R. J. (2015). Variance function partially linear single-index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):171–194.

Liang, M., Choi, Y.-G., Ning, Y., Smith, M. A., and Zhao, Y.-Q. (2020). Estimation and inference on high-dimensional individualized treatment rule in observational data using split-and-pooled de-correlated score. *arXiv preprint arXiv:2007.04445*.

Liang, M., Ye, T., and Fu, H. (2018). Estimating individualized optimal combination therapies through outcome weighted deep learning algorithms. *Statistics in Medicine*, 37(27):3869–3886.

Liang, M. and Yu, M. (2020). A semiparametric approach to model effect modification. *Journal of the American Statistical Association*. To appear.

Liao, P., Klasnja, P., and Murphy, S. (2021). Off-policy estimation of long-term average outcomes with applications to mobile health. *Journal of the American Statistical Association*, 116(533):382–391.

Liao, P., Qi, Z., and Murphy, S. (2020). Batch policy learning in average reward markov decision processes. *arXiv preprint arXiv:2007.11771*.

Lin, Y. and Zhang, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *The Annals of Statistics*, 34(5):2272–2297.

Linn, K. A., Laber, E. B., and Stefanski, L. A. (2017). Interactive Q-learning for quantiles. *Journal of the American Statistical Association*, 112(518):638–649.

Liu, B., Zhou, C., Zhang, X., and Liu, Y. (2020). A unified data-adaptive framework for high dimensional change point detection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):933–963.

Liu, L., Shahn, Z., Robins, J. M., and Rotnitzky, A. (2021). Efficient estimation of optimal regimes under a no direct effect assumption. *Journal of the American Statistical Association*. To appear.

Liu, Y., Wang, Y., Kosorok, M. R., Zhao, Y.-Q., and Zeng, D. (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in Medicine*, 37(26):3776–3788.

Lou, Z., Shao, J., and Yu, M. (2018). Optimal treatment assignment to maximize expected outcome with multiple treatments. *Biometrics*, 74(2):506–516.

Louizos, C., Shalit, U., Mooij, J. M., Sontag, D., Zemel, R., and Welling, M. (2017). Causal effect inference with deep latent-variable models. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Lu, W., Zhang, H. H., and Zeng, D. (2013). Variable selection for optimal treatment decision. *Statistical Methods in Medical Research*, 22(5):493–504.

Luckett, D. J., Laber, E. B., Kahkoska, A. R., Maahs, D. M., Mayer-Davis, E., and Kosorok, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *Journal of the American Statistical Association*, 115(530):692–706.

Luedtke, A. and Chambaz, A. (2020). Performance guarantees for policy learning. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2162–2188.

Luedtke, A. R. and van der Laan, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, 44(2):713.

Luo, J., Schumacher, M., Scherer, A., Sanoudou, D., Megherbi, D., Davison, T., Shi, T., Tong, W., Shi, L., Hong, H., Zhao, C., Elloumi, F., Shi, W., Thomas, R., Lin, S., Tillinghast, G., Liu, G., Zhou, Y., Herman, D., Li, Y., Deng, Y., Fang, H., Bushel, P., Woods, M., and Zhang, J. (2010). A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *The Pharmacogenomics Journal*, 10(4):278–291.

Luo, W., Li, B., and Yin, X. (2014). On efficient dimension reduction with respect to a statistical functional of interest. *The Annals of Statistics*, 42(1):382–412.

Ma, Y., Chiou, J.-M., and Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika*, 93(1):75–84.

Ma, Y. and Zhu, L. (2019). Semiparametric estimation and inference of variance function with large dimensional covariates. *Statistica Sinica*, 29(2):567–588.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246.

Meng, H. and Qiao, X. (2020). A robust method for estimating individualized treatment effect. *arXiv preprint arXiv:2004.10108*.

Meng, H., Zhao, Y.-Q., Fu, H., and Qiao, X. (2020). Near-optimal individualized treatment recommendations. *Journal of Machine Learning Research*, 21(183):1–28.

Mi, X., Zou, F., and Zhu, R. (2019). Bagging and deep learning in optimal individualized treatment rules. *Biometrics*, 75(2):674–684.

Mo, W. and Liu, Y. (2021). Supervised learning. In *Wiley StatsRef: Statistics Reference Online*. Wiley Online Library. To appear.

Mo, W., Qi, Z., and Liu, Y. (2021a). Learning optimal distributionally robust individualized treatment rules. *Journal of the American Statistical Association*. To appear.

Mo, W., Qi, Z., and Liu, Y. (2021b). Rejoinder: Learning optimal distributionally robust individualized treatment rules. *Journal of the American Statistical Association*. To appear.

Moodie, E. E. M., Dean, N., and Sun, Y. R. (2014). Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences*, 6(2):223–243.

Moodie, E. E. M. and Richardson, T. S. (2010). Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics*, 37(1):126–146.

Moodie, E. E. M., Richardson, T. S., and Stephens, D. A. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics*, 63(2):447–455.

Muller, S. (2014). Randomised trials for policy: A review of the external validity of treatment effects. Technical report, The Southern Africa Labour and Development Research Unit, University of Cape Town. Working Paper Number 127. `http://www.opensaldru.uct.ac.za/bitstream/handle/11090/691/2014_127_Saldruwp.pdf`.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355.

Murphy, S. A. (2005). A generalization error for Q-learning. *Journal of Machine Learning Research*, 6(Jul):1073–1097.

Murphy, S. A., van der Laan, M. J., Robins, J. M., and Conduct Problems Prevention Research Group (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423.

Murray, T. A., Yuan, Y., and Thall, P. F. (2018). A bayesian machine learning approach for optimizing dynamic treatment regimes. *Journal of the American Statistical Association*, 113(523):1255–1267.

Nesterov, Y. (2013). Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):1349–1382.

Nie, X., Brunskill, E., and Wager, S. (2021). Learning when-to-treat policies. *Journal of the American Statistical Association*, 116(533):392–409.

Nie, X. and Wager, S. (2020). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*. To appear.

Ning, Y., Sida, P., and Imai, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. *Biometrika*, 107(3):533–554.

O'Muircheartaigh, C. and Hedges, L. V. (2014). Generalizing from unrepresentative experiments: a stratified propensity score approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(2):195–210.

Orellana, L., Rotnitzky, A., and Robins, J. M. (2010a). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: Main content. *The International Journal of Biostatistics*, 6(2). Article 8.

Orellana, L., Rotnitzky, A., and Robins, J. M. (2010b). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part II: Proofs of results. *The International Journal of Biostatistics*, 6(2). Article 9.

Ostroff, J. L. (2016). GLP-1 receptor agonists: An alternative for rapid-acting insulin. *U.S. Pharmacist*, 41(10):3–6.

Pan, Y. and Zhao, Y.-Q. (2021). Improved doubly robust estimation in learning optimal individualized treatment rules. *Journal of the American Statistical Association*, 116(533):283–294.

Pang, J.-S., Razaviyayn, M., and Alvarado, A. (2016). Computing B-stationary points of nonsmooth DC programs. *Mathematics of Operations Research*, 42(1):95–118.

Pardo, L. (2005). *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC, New York, NY, USA.

Pearl, J. and Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, 29(4):579–595.

Powers, S., Qian, J., Jung, K., Schuler, A., Shah, N. H., Hastie, T., and Tibshirani, R. (2018). Some methods for heterogeneous treatment effect estimation in high dimensions. *Statistics in Medicine*, 37(11):1767–1787.

Qi, Z., Cui, Y., Liu, Y., and Pang, J.-S. (2019a). Estimation of individualized decision rules based on an optimized covariate-dependent equivalent of random outcomes. *SIAM Journal on Optimization*, 29(3):2337–2362.

Qi, Z., Liu, D., Fu, H., and Liu, Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *Journal of the American Statistical Association*, 115(530):678–691.

Qi, Z. and Liu, Y. (2018). D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics*, 12(2):3601–3638.

Qi, Z., Pang, J.-S., and Liu, Y. (2019b). Estimating individualized decision rules with tail controls. *arXiv preprint arXiv:1903.04367*.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210.

Qiu, X., Zeng, D., and Wang, Y. (2018). Estimation and evaluation of linear individualized treatment rules to guarantee performance. *Biometrics*, 74(2):517–528.

Rahimian, H. and Mehrotra, S. (2019). Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*.

Razaviyayn, M., Hong, M., and Luo, Z.-Q. (2013). A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153.

Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics - Theory and Methods*, 23(8):2379–2412.

Robins, J. M. (1998). Marginal structural models. In *1997 Proceedings of the Section on Bayesian Statistical Science*, pages 1–10, Alexandria, VA, USA. American Statistical Association.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In Halloran, M. E. and Berry, D., editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, volume 116 of *The IMA Volumes in Mathematics and its Applications*, pages 95–133. Springer, New York, NY, USA.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In Lin, D. Y. and Heagerty, P. J., editors, *Proceedings of the Second Seattle Symposium in Biostatistics*, volume 179 of *Lecture Notes in Statistics*, pages 189–326, New York, NY, USA. Springer.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, 56(4):931–954.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–42.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rotnitzky, A., Lei, Q., Sued, M., and Robins, J. M. (2012). Improved double-robust estimation in missing data and causal inference models. *Biometrika*, 99(2):439–456.

Rotnitzky, A., Smucler, E., and Robins, J. M. (2021). Characterization of parameters with a mixed bias property. *Biometrika*, 108(1):231–238.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014). Q-and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science*, 29(4):640–661.

Schulz, J. and Moodie, E. E. M. (2021). Doubly robust estimation of optimal dosing strategies. *Journal of the American Statistical Association*, 116(533):256–268.

Semenova, V. and Chernozhukov, V. (2017). Debiased machine learning of conditional average treatment effects and other causal functions. *arXiv preprint arXiv:1702.06240*.

Shalit, U., Johansson, F. D., and Sontag, D. (2017). Estimating individual treatment effect: generalization bounds and algorithms. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3076–3085. PMLR.

Shen, X., Tseng, G. C., Zhang, X., and Wong, W. H. (2003). On ψ-learning. *Journal of the American Statistical Association*, 98(463):724–734.

Shi, C., Fan, A., Song, R., and Lu, W. (2018a). High-dimensional A-learning for optimal dynamic treatment regimes. *The Annals of Statistics*, 46(3):925–957.

Shi, C., Lu, W., and Song, R. (2020a). Breaking the curse of nonregularity with subagging – inference of the mean outcome under optimal treatment regimes. *Journal of Machine Learning Research*, 21(176):1–67.

Shi, C., Song, R., and Lu, W. (2016). Robust learning for optimal treatment decision with NP-dimensionality. *Electronic Journal of Statistics*, 10(2):2894–2921.

Shi, C., Song, R., Lu, W., and Fu, B. (2018b). Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):681–702.

Shi, C., Zhang, S., Lu, W., and Song, R. (2020b). Statistical inference of the value function for reinforcement learning in infinite horizon settings. *arXiv preprint arXiv:2001.04515*.

Simoneau, G., Moodie, E. E. M., Nijjar, J. S., Platt, R. W., and Investigators, S. E. R. A. I. C. (2020). Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115(531):1531–1539.

Simoneau, G., Moodie, E. E. M., Platt, R. W., and Chakraborty, B. (2018). Non-regular inference for dynamic weighted ordinary least squares: Understanding the impact of solid food intake in infancy on childhood weight. *Biostatistics*, 19(2):233–246.

Song, R., Kosorok, M., Zeng, D., Zhao, Y.-Q., Laber, E., and Yuan, M. (2015a). On sparse representation for optimal individualized treatment selection with penalized outcome weighted learning. *Stat*, 4(1):59–68.

Song, R., Luo, S., Zeng, D., Zhang, H. H., Lu, W., and Li, Z. (2017). Semiparametric single-index model for estimating optimal individualized treatment strategy. *Electronic Journal of Statistics*, 11(1):364–384.

Song, R., Wang, W., Zeng, D., and Kosorok, M. R. (2015b). Penalized Q-learning for dynamic treatment regimens. *Statistica Sinica*, 25(3):901–920.

Steinwart, I. and Scovel, C. (2007). Fast rates for support vector machines using gaussian kernels. *The Annals of Statistics*, 35(2):575–607.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2):369–386.

Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M., and Li, B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, 10(2):141–158.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, MA, USA, second edition.

Sverdrup, E., Kanodia, A., Zhou, Z., Athey, S., and Wager, S. (2020). policytree: Policy learning via doubly robust empirical welfare maximization over trees. *Journal of Open Source Software*, 5(50):2232.

Swaminathan, A. and Joachims, T. (2015a). Batch learning from logged bandit feedback through counterfactual risk minimization. *Journal of Machine Learning Research*, 16(1):1731–1755.

Swaminathan, A. and Joachims, T. (2015b). The self-normalized estimator for counterfactual learning. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*, 97(3):661–682.

Tao, Y. and Wang, L. (2017). Adaptive contrast weighted learning for multi-stage multi-treatment decision-making. *Biometrics*, 73(1):145–155.

Tao, Y., Wang, L., and Almirall, D. (2018). Tree-based reinforcement learning for estimating optimal dynamic treatment regimes. *The Annals of Applied Statistics*, 12(3):1914–1938.

Thomas, P. and Brunskill, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In Balcan, M. F. and Weinberger, K. Q., editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2139–2148, New York, New York, USA. PMLR.

Tian, L., Alizadeh, A. A., Gentles, A. J., and Tibshirani, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*, 109(508):1517–1532.

Tsiatis, A. (2007). *Semiparametric Theory and Missing Data*. Springer Series in Statistics. Springer, New York, NY, USA.

van der Laan, M. J. and Luedtke, A. R. (2014). Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome. Technical report, U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 329. `https://biostats.bepress.com/ucbbiostat/paper329`.

van der Laan, M. J. and Luedtke, A. R. (2015). Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of Causal Inference*, 3(1):61–95.

van der Laan, M. J. and Rose, S. (2018). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics. Springer, Cham.

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1). Article 11.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, Verlag, NY, USA.

Vansteelandt, S. and Joffe, M. (2014). Structural nested models and G-estimation: The partially realized promise. *Statistical Science*, 29(4):707–731.

Vermeulen, K. and Vansteelandt, S. (2015). Bias-reduced doubly robust estimation. *Journal of the American Statistical Association*, 110(511):1024–1036.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wager, S. and Walther, G. (2015). Adaptive concentration of regression trees, with application to random forests. *arXiv preprint arXiv:1503.06388*.

Wallace, M. P. and Moodie, E. E. M. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71(3):636–644.

Wallace, M. P., Moodie, E. E. M., and Stephens, D. A. (2017). Dynamic treatment regimen estimation via regression-based techniques: Introducing R package DTRreg. *Journal of Statistical Software*, 80(2):1–20.

Wallace, M. P., Moodie, E. E. M., and Stephens, D. A. (2019). Model selection for G-estimation of dynamic treatment regimes. *Biometrics*, 75(4):1205–1215.

Wang, T. and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology*, 80(1):57–83.

Wang, Y. and Zubizarreta, J. R. (2020). Minimal dispersion approximately balancing weights: asymptotic properties and practical considerations. *Biometrika*, 107(1):93–105.

Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK.

Wong, R. K. W. and Chan, K. C. G. (2018). Kernel-based covariate functional balancing for observational studies. *Biometrika*, 105(1):199–213.

Wu, Y. and Liu, Y. (2007). Robust truncated hinge loss support vector machines. *Journal of the American Statistical Association*, 102(479):974–983.

Xiao, W., Zhang, H. H., and Lu, W. (2019). Robust regression for optimal individualized treatment rules. *Statistics in Medicine*, 38(11):2059–2073.

Xu, Y., Yu, M., Zhao, Y.-Q., Li, Q., Wang, S., and Shao, J. (2015). Regularized outcome weighted subgroup identification for differential treatment effects. *Biometrics*, 71(3):645–653.

Xue, F., Zhang, Y., Zhou, W., Fu, H., and Qu, A. (2021). Multicategory angle-based learning for estimating optimal dynamic treatment regimes with censored data. *Journal of the American Statistical Association*. To appear.

Yao, L., Li, S., Li, Y., Huai, M., Gao, J., and Zhang, A. (2018). Representation learning for treatment effect estimation from observational data. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Yoon, J., Jordon, J., and van der Schaar, M. (2018). GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *International Conference on Learning Representations*.

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

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012b). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3):681–694.

Zhang, B. and Zhang, M. (2018). C-learning: A new classification framework to estimate optimal dynamic treatment regimes. *Biometrics*, 74(3):891–899.

Zhang, C., Chen, J., Fu, H., He, X., Zhao, Y., and Liu, Y. (2020). Multicategory outcome weighted margin-based learning for estimating individualized treatment rules. *Statistica Sinica*, 30(4):1857–1879.

Zhang, C. and Liu, Y. (2014). Multicategory angle-based large-margin classification. *Biometrika*, 101(3):625–640.

Zhang, H. H., Cheng, G., and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of the American Statistical Association*, 106(495):1099–1112.

Zhang, Y., Laber, E. B., Davidian, M., and Tsiatis, A. A. (2018). Interpretable dynamic treatment regimes. *Journal of the American Statistical Association*, 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, Q. (2019). Covariate balancing propensity score by tailored loss functions. *The Annals of Statistics*, 47(2):965–993.

Zhao, Q., Small, D. S., and Ertefaie, A. (2017). Selective inference for effect modification via the LASSO. *arXiv preprint arXiv:1705.08020*.

Zhao, Y., Kosorok, M. R., and Zeng, D. (2009). Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, 28(26):3294–3315.

Zhao, Y.-Q., Laber, E. B., Ning, Y., Saha, S., and Sands, B. E. (2019a). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *Journal of Machine Learning Research*, 20(48):1–23.

Zhao, Y.-Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015a). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598.

Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2015b). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1):151–168.

Zhao, Y.-Q., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

Zhao, Y.-Q., Zeng, D., Tangen, C. M., and Leblanc, M. L. (2019b). Robustifying trial-derived optimal treatment rules for a target population. *Electronic Journal of Statistics*, 13(1):1717–1743.

Zheng, W. and van der Laan, M. J. (2010). Asymptotic theory for cross-validated targeted maximum likelihood estimation. Technical report, U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 273. `https://biostats.bepress.com/ucbbiostat/paper273`.

Zhou, X. and Kosorok, M. R. (2017). Augmented outcome-weighted learning for optimal treatment regimes. *arXiv preprint arXiv:1711.10654*.

Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112(517):169–187.

Zhou, X., Wang, Y., and Zeng, D. (2018a). Outcome-weighted learning for personalized medicine with multiple treatment options. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 565–574. IEEE.

Zhou, Z., Athey, S., and Wager, S. (2018b). Offline multi-action policy learning: Generalization and optimization. *arXiv preprint arXiv:1810.04778*.

Zhu, L., Dong, Y., and Li, R. (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling. *Statistica Sinica*, 23(3):1235–1255.

Zhu, L. and Zhu, L. (2009). Dimension reduction for conditional variance in regressions. *Statistica Sinica*, 19(2):869–883.

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.

Zhu, W., Zeng, D., and Song, R. (2019). Proper inference for value function in high-dimensional Q-learning for dynamic treatment regimes. *Journal of the American Statistical Association*, 114(527):1404–1417.