ABSTRACT

RUAN, SHUPING. Optimal Treatment Regimes under Constraints. (Under the direction of Dr. Eric Laber.)

Precision medicine aims to improve disease interventions by incorporating the variability of patients. Clinicians often need to make sequences of decisions for patients with chronic conditions, such as diabetes, cancer, HIV, etc. This sequential decision making problem in precision medicine is mathematically formalized as a dynamic treatment regime (DTR). DTRs are defined as a sequence of decision rules, one for each decision point, that take patient cumulative information as input and output recommended treatment assignments. In most cases, a treatment regime is considered to be optimal, if it optimizes the expected value of a single scalar potential outcome in a population of interest. However, this framework neglects the practical clinical need of balancing several competing outcomes such as, treatment effectiveness, side effect burden, cost, etc. To handle the trade-off among multiple competing outcomes, we propose a new framework where the primary potential outcome of interest is optimized, subject to constraints on secondary outcomes. In Chapter 1, we introduce dynamic treatment regimes, and develop a new method to construct a constrained optimal regime with a single decision point. In Chapter 2, we extend the method into the multiple decision point setting. In Chapter 3, we consider the infinite horizon setting, which is suitable for life-long clinical conditions, and discuss some potential future research directions.

Optimal Treatment Regimes under Constraints

by
Shuping Ruan

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

Statistics

Raleigh, North Carolina

2018

APPROVED BY:

Dr. Anastasios Tsiatis	Dr. Marie Davidian
Dr. Leonard Stefanski	Dr. Krishna Pacifici

Dr. Eric Laber
Chair of Advisory Committee

DEDICATION

To my parents.

BIOGRAPHY

The author was born in Shanghai, China. She obtained her bachelors degree in Biotechnology from Fudan University in Spring, 2011. Her thesis research has been focusing on dynamic treatment regimes in the precision medicine paradigm, under the guidance of her advisor Dr. Eric Laber. She is broadly interested in science and technology, and is possessed with an insatiable curiosity. Her interests include, but are not limited to, statistics, bioinformatics, machine learning, artificial intelligence, quantum computing and so on. She will graduate with her doctoral degree in Statistics in December, 2017. She has also earned masters of Bioinformatics and Statistics while pursuing her doctoral degree.

ACKNOWLEDGEMENTS

I would like to thank my advisor and committee members for their guidance and support. I would like to thank professors, staff and friends from the Department of Statistics and the Bioinformatics Research Center at North Carolina State University for their endless support along the way. I also would like to thank my family for their understanding and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Single-stage Constrained Optimal Treatment Regimes

1.1 Introduction

Precision medicine tailors medical treatments to each patient's own characteristics. It categorize individuals into subpopulations based on, for example, their response to a specific treatment, or their susceptibility to a certain disease, etc. Hence, it targets therapeutic or preventive interventions to those who may benefit, and save those who may not benefit from unnecessary side effects and costs. Given a patient state, such as genetic information, demographics, results of diagnostic test, and so on, dynamic treatment regimes determine what treatment should be assigned next. These are data-driven decision rules that map patient characteristics to recommended treatments.

There is a rich body of research on estimating optimal treatment regimes using data from randomized clinical trials or observational studies. In most cases, a dynamic treatment regime is defined to be optimal if it maximizes the expected value of a certain cumulative clinical outcome when applied to a population of interest. Methods to estimate an optimal treatment regime include Q-learning [35], penalized Q-learning [46], interactive Q-learning [25], A-learning [44], regret-regression [14], g-estimation [41], and policy search methods [37, 49–52, 52]. However, these estimators seek to maximize the expectation of a single scalar outcome, and therefore, neglect the clinical need to balance several competing outcomes. For example, a clinician may have to balance treatment effectiveness,

side-effect burden, and cost while developing a treatment strategy for a patient with a chronic disease; or maximize the expected time to an adverse event while controlling the variance of the time to the adverse event.

Despite its practical importance, very little work has been done on handling multiple competing outcomes. Lizotte et al. considered linear combinations of two competing outcomes indexed by a trade-off parameter and compute the optimal treatment regime for all combination [26]. However, it may not be realistic to assume that a linear trade-off is sufficient to describe all possible patient preferences [19]. Wang et al. used a compound score or "expert score" by numerically combining information on treatment efficacy, toxicity, and the risk of disease progression [48]. Unfortunately, the elicitation of a good composite outcome can be difficult and the misspecification of a composite outcome may severely affect the quality of the estimated treatment regime [20]. There are also some methods to avoid formation of composite outcomes. Laber et al. proposed set-valued dynamic treatment regimes [19]. This method inputs current patient information and outputs a set of recommended treatments. This set contains multiple treatments unless there exists a treatment that is best across all outcomes. This method may not be able to recommend a single treatment and needs expertise for tie breaking when a set of several treatments are recommended. Also, it needs to specify "clinically significant differences" for competing outcomes. Linn at el. proposed constrained interactive Q-learning algorithm [24], which provides an algorithm to find the optimal regime under constraints in the two-stage setting.

We propose a new statistical framework to tackle the problem of balancing multiple competing outcomes using constrained estimation. By constraining the values of secondary outcomes, we search for the optimal feasible regimes for the primary outcome, there by finding constrained optimal regimes. This type of framework is useful in scenarios such as where the clinicians desire to find a treatment strategy that maximize the effectiveness of a treatment regime while controls the side-effect burden and cost. In this chapter, we consider the single-stage scenario. The constrained optimal regime estimator is developed and demonstrated through simulations. Its consistency and asymptotic normality are proven. For demonstration, data from single-stage randomized trials are assumed. Observational data also fit in our framework provided additional assumptions

about the treatment assignment mechanism are reasonable, specifically the no unmeasured confounder assumptions. However, data from observational studies should be used with caution, as the no unmeasured confounder assumption is often unverified [8].

1.2 Methodology

1.2.1 Define single-stage constrained optimal regimes

Dataset

There is only one decision point in the single stage setting. The data from a randomized trial are denoted as

$$\left\{\left(\boldsymbol{X}^i, A^i, \boldsymbol{Y}^i\right)\right\}_{i=1}^{n},$$

consisting of n identically, independently distributed trajectories of $(\boldsymbol{X}, A, \boldsymbol{Y})$, whose distribution are often unknown. Capital letters, $\boldsymbol{X}$, A, $\boldsymbol{Y}$, are used to denote the random variables; lower case letters $\boldsymbol{x}$, a, $\boldsymbol{y}$ to denote realized values of these random variables. $\boldsymbol{X} \in \boldsymbol{\mathcal{X}}$ represents the patient information collected up to the decision point, where $\boldsymbol{\mathcal{X}} \subseteq \mathbb{R}^p$ is the support of $\boldsymbol{X}$. $A \in \mathcal{A}$ represents the treatment assignment, where $\mathcal{A} = \{1, 2, \cdots, m\}$ is the set of all possible treatments. The vector variable $\boldsymbol{Y} \in \mathbb{R}^J$ denotes the outcomes of interest. Let Y_1, the first component of $\boldsymbol{Y}$, be the primary outcome of interest. It is coded so that higher values are desirable. Meanwhile, $Y_2, \cdots, Y_J$ are the secondary outcomes of interest, coded so that the lower values are better.

Potential outcome framework

To identify the causal effect of a certain regime, we take on the potential outcome or counter-factual framework established by Neyman, Rubin and Robins for assessing treatment effects from either randomized or observational studies [15,17,40,42,43]. The set of potential outcomes is $\boldsymbol{W}^* = \left\{\boldsymbol{Y}^*(a), \text{for all } a \in \mathcal{A}\right\}$, where $\boldsymbol{Y}^*(a)$ is the vector-valued outcome that would have been observed if the subject was assigned treatment a. The assumptions made in this framework are as follows.

- *A1. Consistency:*

$$\boldsymbol{Y} = \boldsymbol{Y}^*(A).$$

This means that actual observed outcome vector $\boldsymbol{Y}$ for an individual who received treatment A is the same as the potential outcome for that individual assigned with the same treatment, regardless of the experimental conditions used to assign treatment. It also implies that there is no interference among individuals [42].

- *A2. No unmeasured confounders:*

$$\boldsymbol{W}^* \perp\!\!\!\perp A \mid \boldsymbol{X}.$$

This means that the set of potential outcomes, $\left\{\boldsymbol{Y}^*(a), \text{for all } a \in \mathcal{A}\right\}$, are conditionally independent of treatment assignment A given patient information $\boldsymbol{X}$. In randomized study, this condition is satisfied by construction in randomized studies. However, it can not be verified in observational studies [40].

- *A3. Positivity assumption:* There exists $\epsilon > 0$, so that

$$\Pr(A = a \mid \boldsymbol{X}) > \epsilon, \text{ for all } a \in \mathcal{A}$$

with probability one [15]. This ensures that there is a positive probability of receiving every possible treatment assignment for every value of patient covariates in the population. This assumption is satisfied in well-designed randomized studies. It can also be empirically verified in observational studies. Yet, if it is violated, estimating of regimes for certain subsets of patients can be impossible.

Under A1-A3, $\Pr\left(\boldsymbol{Y}^*(a) \leq \boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x}\right) = \Pr\left(\boldsymbol{Y} \leq \boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x}, A = a\right)$. This implies that the value for a regime can be estimated using the observed data.

Define constrained optimal regimes

In the single stage setting, a treatment regime $\pi : \boldsymbol{\mathcal{X}} \to \mathcal{A}$ is a function that maps the support of patient information $\boldsymbol{X}$ to the set of all possible treatments. Hence, under a regime π, a patient with $\boldsymbol{X} = \boldsymbol{x}$ is recommended to receive treatment $\pi(\boldsymbol{x})$. The vector-valued potential outcome of the regime π is $\boldsymbol{Y}^*(\pi) = \sum_{a \in \mathcal{A}} \boldsymbol{Y}^*(a)\, \mathbb{I}\left\{\pi(\boldsymbol{X}) = a\right\}$. The value vector of a regime π is defined as the expected outcome if every patient in the population of interest is assigned treatment according to π. Mathematically, the value vector of the regime π is $\boldsymbol{V}(\pi) = \mathbb{E}\boldsymbol{Y}^*(\pi)$, of which each component is $V_j(\pi) = \mathbb{E}Y_j^*(\pi)$,

$j = 1, \cdots, J$.

The goal is to find a constrained optimal treatment regime, defined in terms of potential outcomes, that maximizes the expectation of the primary outcome over the space of all the possible regimes under consideration, say Π, and meanwhile satisfies the upper-bound constraints on the expectations of the secondary outcomes. Let the constraint upper-bounds be $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_{J-1})^\intercal$, which can be specified based on patient preference and/or expert domain knowledge. Therefore, estimating a single-stage constrained optimal regime is equivalent to solving

$$\begin{aligned} &\max_{\pi \in \Pi} \; V_1(\pi) \\ &\text{subject to } V_j(\pi) \leq \nu_{j-1}, \end{aligned} \tag{1.1}$$

where $j = 2, \cdots, J$. Hence, a single-stage constrained optimal regime is defined as $\pi^*_{\boldsymbol{\nu}} = \text{argmax}_{\pi \in \Pi} \, V_1(\pi)$, subject to $V_j(\pi) - \nu_{j-1} \leq 0$, where $j = 2, \cdots, J$. Denote the feasible regime space $\mathcal{F}(\Pi)$, which is the set of all regimes satisfying the constraints, i.e., for each $\pi \in \mathcal{F}(\Pi)$, $V_j(\pi) \leq \nu_{j-1}$, where $j = 2, \cdots, J$. Then, a single-stage constrained optimal regime can also be written as $\pi^*_{\boldsymbol{\nu}} = \text{argmax}_{\pi \in \mathcal{F}(\Pi)} V_1(\pi)$.

The class of regimes considered, Π, is restricted to be a family of policy approximation functions parameterized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Denote the regime approximation function as $\pi(\boldsymbol{x}; \boldsymbol{\theta})$, and $\boldsymbol{V}(\pi) = \mathbb{E}\boldsymbol{Y}^*(\pi)$ can be represented as $\boldsymbol{V}(\boldsymbol{\theta}) = \mathbb{E}\boldsymbol{Y}^*(\boldsymbol{\theta})$. Hence, the policy search over the space of regimes in the considered class is turned into a constrained optimization problem over the parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^q$. Problem (1.1) can be represented as

$$\begin{aligned} &\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \; V_1(\boldsymbol{\theta}) \\ &\text{subject to } V_j(\boldsymbol{\theta}) \leq \nu_{j-1}, \end{aligned} \tag{1.2}$$

for $j = 2, \cdots, J$. Moreover, a single-stage constrained optimal regime can be re-written as $\pi^*_{\boldsymbol{\nu}} = \text{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \, V_1(\boldsymbol{\theta})$, subject to $V_j(\boldsymbol{\theta}) - \nu_{j-1} \leq 0$, where $j = 2, \cdots, J$. Denote the feasible parameter space $\mathcal{F}(\boldsymbol{\Theta})$ which is the set of every $\boldsymbol{\theta}$ satisfying the constraints, i.e., for each $\boldsymbol{\theta} \in \mathcal{F}(\boldsymbol{\Theta})$, $V_j(\boldsymbol{\theta}) \leq \nu_{j-1}$, where $j = 2, \cdots, J$. Then, the parameter indexing a true single-stage constrained optimal regime is $\boldsymbol{\theta}^*_{\boldsymbol{\nu}} = \text{argmax}_{\boldsymbol{\theta} \in \mathcal{F}(\boldsymbol{\Theta})} V_1(\boldsymbol{\theta})$.

For computational simplicity, we focus on linear decision rules, so that $\pi(\boldsymbol{x};\boldsymbol{\theta}) = \text{sgn}(\boldsymbol{x}^\intercal\boldsymbol{\theta})$, where we define the sgn function to be

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0. \end{cases}$$

As only the sign of $\boldsymbol{x}^\intercal\boldsymbol{\theta}$ matters for the treatment decision, we restrict the Euclidean norm of $\boldsymbol{\theta}$ to be one, i.e., $\|\boldsymbol{\theta}\|_2^2 = 1$. In this case, problem (1.2) becomes

$$\begin{aligned} &\max_{\boldsymbol{\theta}\in\mathbb{R}^q} \; V_1(\boldsymbol{\theta}) \\ &\text{subject to } V_j(\boldsymbol{\theta}) - \nu_{j-1} \leq 0,\, \boldsymbol{\theta}^\intercal\boldsymbol{\theta} - 1 = 0, \end{aligned} \tag{1.3}$$

for $j = 2, \cdots, J$. The solution to problem (1.3), the indexing parameter for a true constrained optimal regime, is denoted by $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$. The corresponding true constrained optimal regime is denoted by $\pi^*_{\boldsymbol{\nu}} = \text{sgn}(\boldsymbol{x}^\intercal\boldsymbol{\theta}^*_{\boldsymbol{\nu}})$.

1.2.2 Re-define constrained optimal regimes via penalization

Interior-point methods are adopted to solve problem (1.2), a nonlinear constrained continuous optimization problem. To fit in the framework of interior point methods, we reformalize Problem (1.2). Let $v_1(\boldsymbol{\theta}) = -V_1(\boldsymbol{\theta})$ and $v_j(\boldsymbol{\theta}) = V_j(\boldsymbol{\theta}) - \nu_j$, for $j = 2, \cdots, J$. Also, let $h(\boldsymbol{\theta}) = \boldsymbol{\theta}^\intercal\boldsymbol{\theta} - 1$. Hence, problem (1.2) is simplified as

$$\begin{aligned} &\min_{\boldsymbol{\theta}\in\mathbb{R}^q} \; v_1(\boldsymbol{\theta}) \\ &\text{subject to } v_j(\boldsymbol{\theta}) \leq 0,\, h(\boldsymbol{\theta}) = 0, \end{aligned} \tag{1.4}$$

where $j = 2, \cdots, J$. The interior point method solves a sequence of approximate minimization problem (1.4), where μ is always positive and approaches to zero in the limit. For each $\mu > 0$, the approximate problem is

$$\min_{\boldsymbol{\theta},\boldsymbol{z}} \phi_\mu(\boldsymbol{\theta},\boldsymbol{z}) = \min \; v_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J} \ln z_j, \text{ subject to } v_j(\boldsymbol{\theta}) + z_j = 0, h(\boldsymbol{\theta}) = 0 \tag{1.5}$$

where $j = 2, \cdots, J$. The number of slack variables z_j are the number of the inequality constraints ν_j. The z_j are always positive due to the restriction of $\ln z_j$. As μ decreases

to zero, the minimums of ϕ_μ form a trajectory path that approaches the minimum of $v_1(\boldsymbol{\theta})$ in the limit. The extra logarithmic terms $\ln z_j$, named barrier functions, force the trajectory path to be within the feasible region of the problem.

Problem (1.5) forms a sequence of equality constrained problems to approximate problem (1.4) which is a harder inequality-equality mixed constrained problem. An interior point method solves the approximate problem (1.5) iteratively using mainly a Newton step and/or a conjugate gradient step. By default, the algorithm first tries a Newton step which solve the KKT equations for the approximate problem (1.5) through a linear approximation. If this attempt is rejected based on the reduction obtained in a merit function specified for this problem, the algorithm then tries a conjugate gradient step using a trust region. For instance, when the local convexity near the current iterate is not satisfied in the approximate problem, the Newton step is not accepted and the algorithm switches to a conjugate gradient step [6, 11, 47].

1.2.3 Convergence of penalty-barrier trajectory $\left\{\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right\}_{\mu\to 0+}$

The sequence of solutions to Problem (1.5) forms a trajectory path that converges locally to a solution $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$ to the original problem (1.4) from the barrier-penalty method perspective. Interior methods have been identified with barrier methods theoretically. Interior methods use a set of perturbed KKT equations that is connected with the KKT conditions of the barrier method. In this subsection, the conditions for local convergence are examined. Relevant conditions are listed in Appendix A.1.

Solutions to problem (1.5) is equivalent to minimizers to the following penalty-barrier problem (1.6).

$$\min_{\boldsymbol{\theta}} \phi_\mu^{PB}(\boldsymbol{\theta}) = \min \; v_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln(-v_j(\boldsymbol{\theta})) + \frac{1}{2\mu} h^2(\boldsymbol{\theta}) \tag{1.6}$$

where μ is a sequence of decreasing constants approaching zero from the right. The logarithmic terms ensure the inequality constraints hold. The quadratic term penalizes the violation of the equality constraint. The barrier terms and quadratic penalty term provide a smooth function for inference later on. Denote a minimizer to problem (1.6) $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$.

The sequence of minimizers forms a barrier-penalty/central path trajectory $\left\{\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right\}_{\mu\to 0+}$ which converges locally to the minimizer of the original problem (1.4). Here, we specify the conditions needed for local convergence.

Theorem 1.2.1 (Conditions for the penalty-barrier trajectory $\left\{\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right\}_{\mu\to 0+}$ converging to $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$ [3, 11, 36])**.** *Assume:*

1. *the objective and constraint functions $v_j(\boldsymbol{\theta})$, for $j = 1, \cdots, J$, and $h(\boldsymbol{\theta})$ are twice continuously differentiable with respect to $\boldsymbol{\theta}$;*

2. *the gradients of constraints, $\nabla v_j(\boldsymbol{\theta})$, for $j = 2, \cdots, J$ and $\nabla h(\boldsymbol{\theta})$ are linearly independent, where the gradients are taken with respect to $\boldsymbol{\theta}$;*

3. *strict complementarity holds for $\boldsymbol{\lambda}^*_{\mathcal{I}}\boldsymbol{v}(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}) = 0$, where $\boldsymbol{\lambda}^*_{\mathcal{I}}$ are the Lagrangian multipliers of the inequality constraints $\boldsymbol{v} = (v_2, \cdots, v_J)$; Strict complementarity means that the multipliers for inequality constraints $\boldsymbol{\lambda}^*_{\mathcal{I}}$ have the property that $\lambda^*_i > 0$, for all $i \in \mathcal{A}_{\mathcal{I}}(\boldsymbol{\theta}^*_{\boldsymbol{\nu}})$, the set of indices of active inequality constraints at $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$;*

4. *the sufficient conditions under which $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$ is an isolated local constrained minimizer of the original problem (1.4) are satisfied by $\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}, \boldsymbol{\lambda}^*_{\mathcal{I}}, \lambda^*_{\mathcal{E}}\right)$, where $\boldsymbol{\lambda}^*_{\mathcal{I}}$ is the Lagrangian multiplier for the equality constraint $h(\boldsymbol{\theta})$. The sufficient conditions for optimality are:*

   (a) *$\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$ is feasible and the LICQ (Linear Independence Constraint Qualification) holds at $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$, i.e., the Jacobian matrix of active constraints at $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$, $J_{\mathcal{A}}(\boldsymbol{\theta}^*_{\boldsymbol{\nu}})$, has full row rank;*

   (b) *$\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$ is a KKT point and strict complementarity holds, i.e, the (necessarily unique) multipliers $\boldsymbol{\lambda}^{*\intercal} = \left(\boldsymbol{\lambda}^{*\intercal}_{\mathcal{I}}, \lambda^*_{\mathcal{E}}\right)$ have the property that $\boldsymbol{\lambda}^*_i > 0$, for all $i \in \mathcal{A}_{\mathcal{I}}(\boldsymbol{\theta}^*_{\boldsymbol{\nu}})$, the set of indices of active inequality constraints at $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$;*

   (c) *for all nonzero vectors p, there exists $\omega > 0$ such that $\boldsymbol{p}^\intercal H(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}, \boldsymbol{\lambda}^*)\boldsymbol{p} \geq \omega\|p\|^2$., where $H(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}, \boldsymbol{\lambda}^*)$ is the hessian of the Lagrangian at $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}$ and $\boldsymbol{\lambda}^*$, where $\boldsymbol{\lambda}^*$ is the vector of the Lagrangian multipliers, $\boldsymbol{\lambda}^{*\intercal} = \left(\boldsymbol{\lambda}^{*\intercal}_{\mathcal{I}}, \lambda^*_{\mathcal{E}}\right)$.*

   then there is a positive neighborhood about $\mu = 0$ for which a unique-isolated differentiable function $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$ exists. It describes a unique isolated trajectory of local minima of $\phi^{PB}_{\mu}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu) \to \boldsymbol{\theta}^*_{\boldsymbol{\nu}}$ as $\mu \to 0+$.

To find $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$, we need to examine the stationarity of $\phi^{PB}_{\mu}(\boldsymbol{\theta})$. That is $\nabla\phi^{PB}_{\mu}(\boldsymbol{\theta}) = 0$ is satisfied at $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$. Its equivalent system of non-linear equations is

$$F_{\mu}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \begin{pmatrix} g(\boldsymbol{\theta}) - J(\boldsymbol{\theta})\boldsymbol{\lambda} \\ \tilde{\boldsymbol{v}}(\boldsymbol{\theta})\boldsymbol{\lambda}_{\mathcal{I}} - \mu \\ h(\boldsymbol{\theta}) + \mu\lambda_{\mathcal{E}} \end{pmatrix} = 0, \tag{1.7}$$

where $g(\boldsymbol{\theta}) = \nabla v_1(\boldsymbol{\theta})$, and $J(\boldsymbol{\theta})$ is the Jacobian matrix of the constraints.

Together with $\boldsymbol{\lambda} > \mathbf{0}$, the non-linear system (1.7) forms the KKT condtions of $\phi^{PB}_{\mu}(\boldsymbol{\theta})$, the penalty-barrier problem (1.6). If we define $\chi_1 \triangleq {}^{\mu}/_{\tilde{\boldsymbol{v}}(\boldsymbol{\theta})}$ and $\chi_2 \triangleq -{}^{h(\boldsymbol{\theta})}/_{\mu}$, then χ_1 and χ_2 are considered as approximates of the Lagrangian multipliers under μ-perturbed KKT conditions of the interior-point problem (1.5). This shows the connection between interior methods and barrier methods. More details can be found in reference [3, 11, 36].

Moreover, the log barrier implies that the inequality constraint is strictly satisfied at $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$, i.e., $v_j(\boldsymbol{\theta}) = V_j(\boldsymbol{\theta}) - \nu_j < 0$, for $j = 2, \cdots, J$. For a minimizer of $\phi^{PB}_{\mu}(\boldsymbol{\theta})$ to exists, the strict feasible set, strict($\mathcal{F}(\boldsymbol{\Theta})$), of the original constrained problem (1.4) is assumed to be non-empty.

1.2.4 Consistency of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$

Let $\widehat{\boldsymbol{V}}(\pi)$ be a consistent estimator of the value of a regime π, and each component is denoted by $\widehat{V}_j(\pi)$, for $j = 1, \cdots, J$. As a regime function π is parameterized by index $\boldsymbol{\theta}$, $\widehat{v}_1(\boldsymbol{\theta}) = -\widehat{V}_1(\boldsymbol{\theta})$ and $\widehat{v}_j(\boldsymbol{\theta}) = \widehat{V}_j(\boldsymbol{\theta}) - \nu_j$, for $j = 2, \cdots, J$. Then, problem (1.6) with the plugin estimators, which is the formalization to be solved numerically, is

$$\min_{\boldsymbol{\theta}, \boldsymbol{z}} \widehat{\phi}_{\mu}(\boldsymbol{\theta}, \boldsymbol{z}) = \min\ \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln z_j,\ \text{subject to } \widehat{v}_j(\boldsymbol{\theta}) + z_j = 0, \boldsymbol{\theta}^{\intercal}\boldsymbol{\theta} - 1 = 0 \tag{1.8}$$

where z_j's are the slack variables. The solution to (1.8) is theoretically equivalent to the solution to

$$\min_{\boldsymbol{\theta}} \widehat{\phi}^{PB}_{\mu}(\boldsymbol{\theta}) = \min\ \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln \widehat{v}_j(\boldsymbol{\theta}) + \frac{1}{2\mu}\left(\boldsymbol{\theta}^{\intercal}\boldsymbol{\theta} - 1\right)^2 \tag{1.9}$$

Denote a solution to (1.9) $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$. It is proven that $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ is a consistent estimator of $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$, when $\widehat{\boldsymbol{V}}(\pi)$ is a co nsistent estimator of the value of a regime π.

Theorem 1.2.2. *For any fixed μ, assume*

1. *Point-wise convergence of $\widehat{v}_j(\boldsymbol{\theta})$ in probability:*
   For every $\boldsymbol{\theta} \in \mathcal{F}(\boldsymbol{\Theta})$, we have $\lim_{n\to\infty} Pr\{\mid v_j(\boldsymbol{\theta}) - \widehat{v}_j(\boldsymbol{\theta}) \mid \leq \epsilon_j\} = 1$, $\forall \epsilon_j > 0$, where $j = 1, \cdots, J$;

2. *Existence of a strict local minimizers of $\phi^{PB}_{\mu}(\boldsymbol{\theta})$:*
   There exists a neighborhood of $\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$, denoted $\mathcal{N}\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)$ such that $\phi^{PB}_{\mu}\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right) < \phi^{PB}_{\mu}(\boldsymbol{\theta})$, for any $\boldsymbol{\theta} \in \mathcal{N}\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)$;

3. *Existence of strict local minimizer $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ of $\widehat{\phi}^{PB}_{\mu}(\boldsymbol{\theta})$ in the neighborhood $\mathcal{N}\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)$:*
   $\widehat{\phi}^{PB}_{\mu}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)\right) < \widehat{\phi}^{PB}_{\mu}(\boldsymbol{\theta})$, for any $\boldsymbol{\theta} \in \mathcal{N}\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)$, where $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) \in \mathcal{N}\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)$;

then

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) \xrightarrow{p} \boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu).$$

See Appendix A.2 for proof.

1.2.5 Estimation of the values of a regime

Modeling the value functions

Under the three assumptions of the potential outcome framework *A1, A2*, and *A3* mentioned in Subsection 1.1.1, it can be shown that for any $\boldsymbol{x}$ such that $\Pr(\boldsymbol{X} = \boldsymbol{x}) > 0$, $\mathbb{E}\left\{\boldsymbol{Y}^*(a) \mid \boldsymbol{X} = \boldsymbol{x}\right\} = \mathbb{E}\left(\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x}, A = a\right)$. Define $\boldsymbol{Q}(\boldsymbol{x}, a) = \mathbb{E}\left(\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x}, A = a\right)$. and $\boldsymbol{Q}^{\pi}(\boldsymbol{x}) = \mathbb{E}\left\{\boldsymbol{Y} \mid \boldsymbol{X} = \boldsymbol{x}, A = \pi(\boldsymbol{x})\right\}$.These are the $\boldsymbol{Q}$ functions for measuring the quality of a treatment assignment and a regime for a given $\boldsymbol{x}$. The $\boldsymbol{Q}$ function has the same dimension as the outcome vector $\boldsymbol{Y}$. Then the value for a regime π is $\boldsymbol{V}(\pi) = \mathbb{E}\boldsymbol{Y}^*(\pi) = \mathbb{E}\left\{\boldsymbol{Q}^{\pi}(\boldsymbol{X})\right\}$.

To model each component of $\boldsymbol{Q}(\boldsymbol{x}, a)$, a linear working model of the forms $Q_j(\boldsymbol{x}, a) = \boldsymbol{x}_0^{\intercal}\boldsymbol{\alpha}_j + a \cdot \boldsymbol{x}_1^{\intercal}\boldsymbol{\beta}_j$ is used, where $\boldsymbol{x}^{\intercal} = \left(\boldsymbol{x}_0^{\intercal}, \boldsymbol{x}_1^{\intercal}\right)$. A regime is approximated using the function $\pi(\boldsymbol{x}) = \text{sgn}(\boldsymbol{x}^{\intercal}\boldsymbol{\theta})$, and an optimal regime is searched over this class of function. Then, $V_j(\boldsymbol{\theta}) = \mathbb{E}\left(\boldsymbol{x}_0^{\intercal}\boldsymbol{\alpha}_j + \text{sgn}(\boldsymbol{x}^{\intercal}\boldsymbol{\theta}) \cdot \boldsymbol{x}_1^{\intercal}\boldsymbol{\beta}_j\right)$, for $j = 1, \cdots, J$. Let $m_{\boldsymbol{\alpha}_j} = \boldsymbol{x}_0^{\intercal}\boldsymbol{\alpha}_j$, which is the

part not related to $\boldsymbol{\theta}$. Also, let $z_1 = \boldsymbol{x}^\intercal\boldsymbol{\theta}$ and $z_2 = \boldsymbol{x}_1^\intercal\boldsymbol{\beta}_j$. Let $f_{\boldsymbol{\beta}_j}(\boldsymbol{z};\boldsymbol{\theta})$ be the joint distribution of $\boldsymbol{z} = (z_1, z_2)^\intercal$. Assuming all the models are correctly specified, we denote the true parameter values in the working models $\left(\boldsymbol{\alpha}_j^*, \boldsymbol{\beta}_j^*\right)$. Hence, the j-th value function is modeled as

$$V_j(\boldsymbol{\theta}) = m_{\boldsymbol{\alpha}_j^*} + \iint \operatorname{sgn}(z_1)\, z_2 f_{\boldsymbol{\beta}_j^*}(z_1, z_2;\boldsymbol{\theta})\, dz_1\, dz_2.$$

Estimating the value functions

Denote the corresponding least-squared estimators $\left(\widehat{\boldsymbol{\alpha}}_j^\intercal, \widehat{\boldsymbol{\beta}}_j^\intercal\right)$. Then the estimated Q_j functions is $\widehat{Q}_j(\boldsymbol{x}, a) = \boldsymbol{x}_0^\intercal\widehat{\boldsymbol{\alpha}}_j + a \cdot \boldsymbol{x}_1^\intercal\widehat{\boldsymbol{\beta}}_j$, for $j = 1, \cdots, J$. The estimated values of a regime π are

$$\widehat{V}_j(\boldsymbol{\theta}) = m_{\widehat{\boldsymbol{\alpha}}_j} + \iint \operatorname{sgn}(z_1)\, z_2 \widehat{f}_{\widehat{\boldsymbol{\beta}}_j}(z_1, z_2;\boldsymbol{\theta})\, dz_1\, dz_2,$$

where $\widehat{f}_{\widehat{\boldsymbol{\beta}}_j}(\boldsymbol{z};\boldsymbol{\theta})$ is a kernel density estimator (KDE) of the joint distribution of $(\boldsymbol{x}^\intercal\boldsymbol{\theta}, \boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}}_j)$. This approach only requires two dimensional density estimation, contrasting with estimating the entire density of $\boldsymbol{X}$ which could potentially be high dimensional. For any fixed $\boldsymbol{\theta}$ and $\boldsymbol{\beta}_j$, a KDE $\widehat{f}_{\boldsymbol{\beta}_j}(z_1, z_2;\boldsymbol{\theta})$ is used to estimate the distribution of $(Z_1, Z_2) = (\boldsymbol{X}^\intercal\boldsymbol{\theta}, \boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j)$, where

$$\widehat{f}_{\boldsymbol{\beta}_j}(z_1, z_2;\boldsymbol{\theta}) = (nh_1h_2)^{-1}\sum_{i=1}^{n} k\left({}^{(z_1 - Z_1^i)}/_{h_1}\right) k\left({}^{(z_2 - Z_2^i)}/_{h_2}\right).$$

For instance, a Gaussian kernel can be used such that $k(x) = {}^1/_{\sqrt{2\pi}}\exp\left(-{}^{x^2}/_2\right)$. Moreover, the marginal density of Z_2 is $f_{\boldsymbol{\beta}_j}(z_2)$ is estimated by $\widehat{f}_{\boldsymbol{\beta}_j}(z_2) = (nh_2)^{-1}\sum_{i=1}^{n} k\left({}^{(z_2 - Z_2^i)}/_{h_2}\right)$. For exposition, let $h = h_n = h_1 = h_2$. h is a function of sample size n. For KDEs to be consistent, we need $h \to 0$ and $nh \to \infty$, as $n \to \infty$ (see Appendix A.3 for details on conditions for consistency of KDEs). Also, let $K(x) = \int_{-\infty}^{x} k(x)\, dx$. After some algebra (see Appendix A.4), we can derive that

$$\widehat{V}_j(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{X}_0^{i\intercal}\widehat{\boldsymbol{\alpha}}_j + \boldsymbol{X}_1^{i\intercal}\widehat{\boldsymbol{\beta}}_j\left\{1 - 2K\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\right\}\right].$$

Assuming the model is correctly specified, estimators $\widehat{\boldsymbol{\alpha}}_j$ and $\widehat{\boldsymbol{\beta}}_j$ are consistent, along with the KDEs. Therefore, $\widehat{V}_j(\boldsymbol{\theta})$, which are used to construct $\widehat{\phi}_{\boldsymbol{\nu}}^{PB}(\mu)$, are point-wise

consistent. Additionally, if we assume isolated local minima exist for $\phi_{\mu}^{PB}(\boldsymbol{\theta})$ and $\widehat{\phi}_{\mu}^{PB}(\boldsymbol{\theta})$ respectively, then $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ is consistent based on Theorem 1.1.2.

Note, for any fixed value $\boldsymbol{\alpha}_j$ and $\boldsymbol{\beta}_j$, we use notation

$$\widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_j, \boldsymbol{\beta}_j\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{X}_0^{i\intercal}\boldsymbol{\alpha}_j + \boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j\left\{1 - 2K\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\right\}\right],$$

Moreover, $\boldsymbol{\alpha}_j$ may be dropped in the gradient $\nabla\widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right)$, as it becomes irrelevant.

1.2.6 Asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$

Limiting distribution of $\nabla\widehat{V}_j(\boldsymbol{\theta})$

Before deriving the limiting distribution of the estimator $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$, we examine the limiting distribution of $\nabla\widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right)$ for any fixed value of $\boldsymbol{\theta} \in \mathcal{F}(\boldsymbol{\Theta})$ and $\boldsymbol{\beta}_j$, where

$$\nabla\widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}^i,$$

for $j = 1, \cdots, J$. Notation ∇ denotes the first-order derivatives with respect to $\boldsymbol{\theta}$.

Lemma 1.2.3. *Suppose the following conditions hold*

1. $\forall \boldsymbol{a} \in \mathbb{R}^p$, $\exists\, \delta > 0$ *,such that*
   - *(a)* $\mathbb{E}\left|\boldsymbol{a}^{\intercal}\frac{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X} - \mu_n\right|^{2+\delta} < \infty$, *where* $\mu_n = \mathbb{E}\left\{\boldsymbol{a}^{\intercal}\frac{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}$;
   - *(b)* $\boldsymbol{a}^{\intercal}V\left\{\frac{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j}{h}k(-\frac{\boldsymbol{X}^{\intercal}\boldsymbol{\theta}}{h})\boldsymbol{X}\right\}\boldsymbol{a}^{1+\frac{\delta}{2}} < \infty$.

Then, for any fixed $\boldsymbol{\theta}$ *and* $\boldsymbol{\beta}_j$,

$$\sqrt{n}\left(\nabla\widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right) - \mathbb{E}\left\{\frac{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right) \xrightarrow{d} N\left(0, Avar\left\{\frac{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right),$$

where $j = 1, \cdots, J$.

Notation ∇ denotes the first-order derivatives with respect to $\boldsymbol{\theta}$. Avar stands for asymptotic variance. See appendix A.5 for proof of Lemma 1.1.3.

Parameters $\boldsymbol{\beta}_j^*$'s are unknown, and are estimated by consistent least square estimators $\widehat{\boldsymbol{\beta}}_j$'s. Moreover, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ is proven to be a consistent estimator for $\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)$ in Theorem 1.1.2 as well. The following corollary shows that the estimation does not effect the limiting distribution obtained above.

Corollary 1.2.4. *Suppose all the assumptions in Lemma 1.1.3 hold. Also, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ and $\widehat{\boldsymbol{\beta}}_j$ are consistent estimators of $\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)$ and $\boldsymbol{\beta}_j^*$, respectively. Then,*

$$\sqrt{n}\left(\nabla\widehat{V}_j\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu),\widehat{\boldsymbol{\beta}}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)}{h}\right)\boldsymbol{X}\right\}\right)\xrightarrow{d}N\left(0,Avar\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)}{h}\right)\boldsymbol{X}\right\}\right).$$

See Appendix A.6 for the proof.

Limiting distribution of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$

Based on the limiting distribution of $\nabla\widehat{V}_j\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu),\widehat{\boldsymbol{\beta}}_j\right)$ and Taylor expansion, we derive the limiting distribution of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$.

Theorem 1.2.5. *Suppose all the assumptions in Lemma 1.1.4 and Corollary 1.1.5 hold. Then we have, as $n\to\infty$*

$$\sqrt{n}\left\{\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)-\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right\}\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^*=\boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$, where $\boldsymbol{\Sigma}^*=\boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$, $\boldsymbol{C}^*=\mathbb{E}\left\{\nabla v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\nabla^\intercal v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right\}-\mathbb{E}\left\{\nabla v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right\}\mathbb{E}\left\{\nabla^\intercal v_1\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu)\right)\right\}$, and $\boldsymbol{D}^*=\nabla^2\phi_\mu^{BP}(\boldsymbol{\theta}_{\boldsymbol{\nu}}^*(\mu))$.

Proof and related limits are provided in Appendix A.7. Due to the complexity of the variance matrix formulation, the Bootstrap is recommended for variance estimation.

1.3 Simulation

Simulated experiments are carried out to examine the finite sample performance of the proposed method.

1.3.1 Simulation design

The generative model for simulation is

$$
\begin{aligned}
&\boldsymbol{X} \sim MVN(\boldsymbol{0}, \boldsymbol{I}),\\
&A \sim \text{Uniform}\{-1, 1\},\\
&Y_1 = \bar{\boldsymbol{X}}^\intercal \boldsymbol{\alpha}_1 + A \cdot (\bar{\boldsymbol{X}}^\intercal \boldsymbol{\beta}_1) + \epsilon_1,\\
&\epsilon_{Y1} \sim N(0, \sigma_1^2),\\
&Y_2 = \bar{\boldsymbol{X}}^\intercal \boldsymbol{\alpha}_2 + A \cdot (\bar{\boldsymbol{X}}^\intercal \boldsymbol{\beta}_1) + \epsilon_2,\\
&\epsilon_2 \sim N(0, \sigma_2^2),
\end{aligned}
$$

where $\boldsymbol{I}$ is a 2×2 identity matrix and $\bar{\boldsymbol{X}}^\intercal = (1, \boldsymbol{X}^\intercal)$. The values of these parameters are discussed shortly. For simplicity, we consider two competing outcomes, i.e., $J = 2$. Also, let $\boldsymbol{X}_0 = \boldsymbol{X}_1 = \boldsymbol{X}$. All the parameters in the generative model are set based on the two factors mentioned in below and R-squares.

As nonlinear constrained optimization is expensive to carry out, the number of Monte Carlo iteration is set to $M = 200$. Sample size of the training set in each iteration is set to $N_{train} = 1000$. For a sequence of upper bounds on $\mathbb{E}Y_2(\pi)$, say v_k, $k = 1, ..., K$, we use training data to estimate a constrained optimal regime. The estimated regime is then applied to test data generated from the same model to estimate the values of that regime. The sample size of the test set is set to $N_{test} = 10000$.

Moreover, because larger values of Y_1 are more desirable, and it is modeled that $Q_1(\boldsymbol{x}, a) = \boldsymbol{x}^\intercal \boldsymbol{\alpha}_1^* + a \cdot (\boldsymbol{x}^\intercal \boldsymbol{\beta}_1^*)$. Therefore, $\max_a Q_1(\boldsymbol{x}, a) = \boldsymbol{x}^\intercal \boldsymbol{\alpha}_1^* + |\boldsymbol{x}^\intercal \boldsymbol{\beta}_1^*|$. The true unconstrained optimal regime for the primary outcome Y_1 is $\pi_1^*(\boldsymbol{x}) = \text{sgn}(\boldsymbol{x}^\intercal \boldsymbol{\beta}_1^*)$. Meanwhile, smaller values of Y_2 are more desirable, and it is modeled that $Q_2(\boldsymbol{x}, a) = \boldsymbol{x}^\intercal \boldsymbol{\alpha}_2^* + a \cdot (\boldsymbol{x}^\intercal \boldsymbol{\beta}_2^*)$. Thus, $\min_a Q_2(\boldsymbol{x}, a) = \boldsymbol{x}^\intercal \boldsymbol{\alpha}_2^* - |\boldsymbol{x}^\intercal \boldsymbol{\beta}_2^*|$. The true unconstrained optimal regime for the secondary outcome Y_2 is $\pi_2^*(\boldsymbol{x}) = -\text{sgn}(\boldsymbol{x}^\intercal \boldsymbol{\beta}_2^*)$.

Two major factors are considered in the simulation. To examine how constraints affect an estimated constrained optimal treatment regime, two factors are considered for the simulation setting. First define $\Omega_1 = \mathbb{E}\big(\mathbb{I}\{(\boldsymbol{X}^\intercal \boldsymbol{\beta}_1^*)(\boldsymbol{X}^\intercal \boldsymbol{\beta}_2^*) > 0\}\big)$ to be the probability of optimal regimes disagree $\pi_1^*(\boldsymbol{x}) \neq \pi_2^*(\boldsymbol{x})$ (abbreviated by Prob. DIS). Three

levels are set for Ω_1: slightly disagree 0.3, moderate disagree 0.5, and strongly disagree 0.7. Second is $\Omega_2 = {}^{\Omega_{21}}\!/_{\Omega_{22}}$, where $\Omega_{21} = {}^{\mathbb{E}(|\boldsymbol{X}^\intercal\boldsymbol{\beta}_1^*|\mathbb{I}\{(\boldsymbol{X}^\intercal\boldsymbol{\beta}_1^*)(\boldsymbol{X}^\intercal\boldsymbol{\beta}_2^*)>0\})}\!/_{\mathbb{E}|\boldsymbol{X}^\intercal\boldsymbol{\beta}_1^*|}$ and $\Omega_{22} = {}^{\mathbb{E}(|\boldsymbol{X}^\intercal\boldsymbol{\beta}_2^*|\mathbb{I}\{(\boldsymbol{X}^\intercal\boldsymbol{\beta}_1^*)(\boldsymbol{X}^\intercal\boldsymbol{\beta}_2^*)>0\})}\!/_{\mathbb{E}|\boldsymbol{X}^\intercal\boldsymbol{\beta}_2^*|}$. Ω_{21} defines the relative expected treatment effect with respect to Y_1 when π_1^* and π_2^* disagree. Ω_{22} is defined analogously. Therefore, Ω_2 is the ratio between the two relative treatment effects when when π_1^* and π_2^* disagree (abbreviated by RRTE). It is set to low ratio 0.5, medium ratio 1.0 and high ratio 1.5. Additionally, the R-squares for the regression of Y_1 on $\boldsymbol{X}$ and A and the regression of Y_2 on $\boldsymbol{X}$ and A, respectively. Both are set to be 0.6. Table 1.1 summarize the 9 settings. Appendix A.8 describes the details on specifying the parameters values for these 9 settings.

Table 1.1: 9 Settings for Monte Carlo Simulations

Setting	Ω_1 Prob.	DIS	Ω_2	RRTE.
1	Slight	0.3	Low	0.5
2	Slight	0.3	Medium	1.0
3	Slight	0.3	High	1.5
4	Moderate	0.5	Low	0.5
5	Moderate	0.5	Medium	1.0
6	Moderate	0.5	High	1.5
7	Strong	0.7	Low	0.5
8	Strong	0.7	Medium	1.0
9	Strong	0.7	High	1.5

1.3.2 Summary of simulation results

We summarize the simulation results here. The complete results are summarized in appendix A.8, along with the details of the simulations. Table 1.2 below shows the estimated optimal regime values for setting 1 and their standard deviation. The corresponding index parameter estimates are also included along with their standard deviation. Figure 1.1 is the efficient frontier plot for setting 1. The red dashed line represents $\widehat{V}_1$ under estimated constrained optimal regime, and the blue dash-dotted line represents $\widehat{V}_2$ under that regime. These plots borrow the concept of efficient frontier in modern portfolio

theory [30]. It represents the best possible value of the primary potential outcome for its level of risk, which is the value of the secondary potential outcome. In the plot, the value of the primary outcome increases as the constraint bound gets looser. Meanwhile the value of the secondary outcome keep up with the constraint, until the constraint is not active. Once the constraint gets larger than the maximum value of the secondary potential outcome, the constrained problem becomes an unconstrained problem.

Table 1.2: Simulation Result for Setting 1

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
0.23	0.55	0.29	0.22	0.08	0.34	0.21	-0.91	0.11
0.28	0.74	0.27	0.28	0.08	0.46	0.19	-0.86	0.10
0.33	0.90	0.26	0.34	0.08	0.56	0.20	-0.79	0.17
0.38	1.05	0.24	0.39	0.08	0.65	0.17	-0.73	0.13
0.43	1.15	0.34	0.45	0.08	0.69	0.29	-0.62	0.23
0.48	1.25	0.38	0.50	0.08	0.73	0.34	-0.52	0.28
0.53	1.44	0.20	0.56	0.08	0.85	0.15	-0.46	0.20
0.59	1.50	0.31	0.60	0.08	0.86	0.27	-0.35	0.25
0.64	1.61	0.30	0.65	0.08	0.90	0.25	-0.25	0.24
0.69	1.67	0.32	0.70	0.09	0.91	0.28	-0.13	0.27
0.74	1.74	0.35	0.75	0.09	0.92	0.30	-0.01	0.26
0.79	1.81	0.26	0.80	0.08	0.94	0.23	0.10	0.24
0.84	1.84	0.26	0.84	0.06	0.92	0.25	0.20	0.23
0.89	1.87	0.21	0.87	0.04	0.92	0.20	0.28	0.20
0.94	1.89	0.18	0.88	0.03	0.92	0.17	0.32	0.15
0.99	1.91	0.14	0.89	0.02	0.93	0.13	0.33	0.12

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

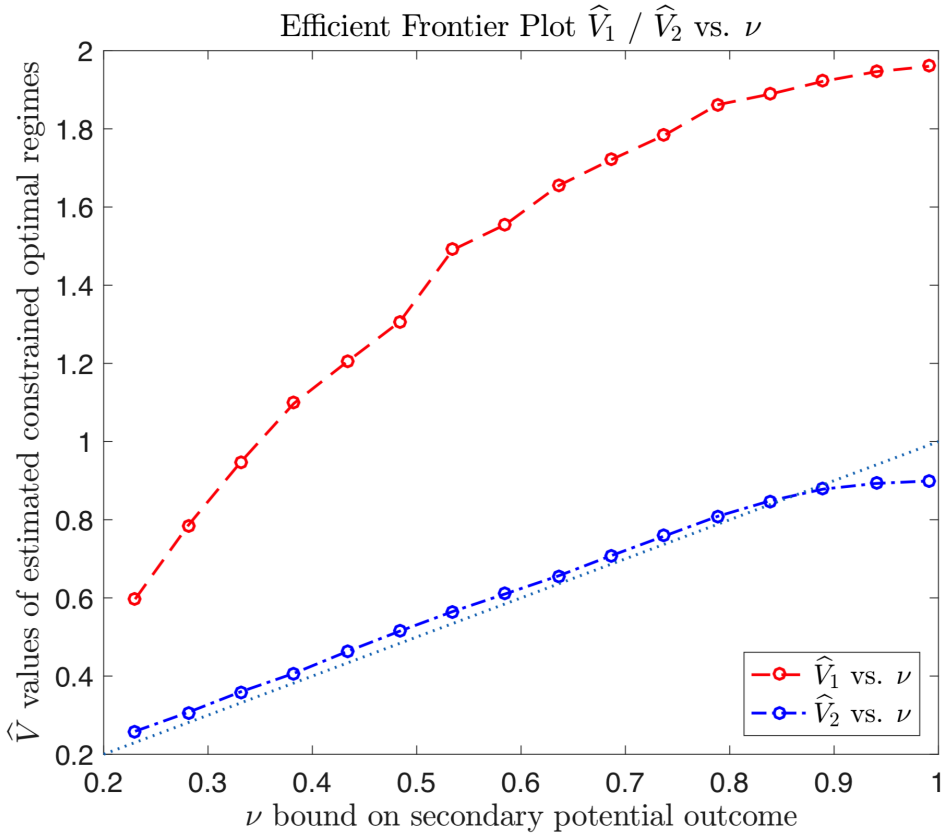

Figure 1.1: Efficient frontier for estimated constrained optimal regimes (single-stage) for Setting 1.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

1.4 Conclusion

Most of the research in dynamics treatment regimes has been focusing on optimizing a single scalar outcome. However, it may be an oversimplification of the goals of practical clinical decision making. In this chapter, a new method is proposed to handle multiple competing outcomes. We cast estimation of an optimal treatment regime with competing outcomes as a constrained optimization problem, which maximizes the primary outcome of interest, subject to the constraints on the secondary outcomes of interest. We prove that our estimator of a constrained optimal treatment regime is consistent under mild regularity conditions. The asymptotic limiting distribution is derived for the estimated indexing parameter for the estimated optimal regimes. Our efficient frontier plots provide an intuitive way for clinicians to examine the trade-off between multiple competing outcomes.

Chapter 2

Multi-stage Constrained Optimal Treatment Regimes

2.1 Introduction

Dynamic treatment regimes (DTRs), also known as adaptive treatment strategies or policies, are sequences of decision rules of which input is time-varying patient information and output is a recommended treatment at each intervention point [9, 31, 32]. These decision rules can be used to inform treatment decision for chronic conditions, e.g., depression, alcohol and drug abuse, HIV infection, cancer, diabetes etc., where clinicians have to make decisions at each stage based on evolving patient histories. A handful of methods have been developed to estimate the optimal treatment regimes. For example, indirect methods include Q-learning [35], penalized Q-learning [46], interactive Q-learning [25], A-learning [44], regret-regression [14], g-estimation [41] and so on. Policy search methods include marginal structural mean models [37, 39], outcome weighted learning [49, 51, 52], doubly robust estimators [50], and so forth. However, these methods only take a single clinical outcome into consideration, and neglect the clinical need to balance several competing outcomes. For example, a clinician may have to balance treatment effectiveness, side-effect burden, and cost while developing a treatment strategy for a patient with a chronic disease; or maximize the expected time to an adverse event while controlling the variance of the time to the adverse event.

Although handling the trade-off among multiple competing outcomes is important in

practice, there has been little work done on this issue. Lizotte et al. proposed to compute the optimal treatment regimes of all the possible linear combinations of two competing outcomes [26]. However, only considering linear trade-off between two competing outcomes may not be sufficient to describe all possible patient preferences [19]. Wang et al. considered a compound score or "expert score" by numerically combining information on treatment efficacy, toxicity, and the risk of disease progression [48]. Unfortunately, it can be difficult to elicit a good composite outcome, and the quality of the estimated treatment regime maybe severely affect by the misspecification of a composite outcome [20]. Some methods do not require the formation of composite outcomes. For example, set-valued dynamic treatment regimes proposed by Laber et al. inputs current patient information and outputs a set of recommended treatments. Multiple treatments may included in the set recommended, unless there exists a treatment that is best across all outcomes. Domain expertise is needed for tie breaking when a set of several treatments are recommended. Also, it needs to specify "clinically significant differences" for competing outcomes [19].

In this chapter, we continue the work previously done by Linn at el. [18], and propose a new statistical framework to tackle the problem of balancing multiple competing outcomes using constrained estimation. By restricting the values of secondary ones, we search for the feasible regimes with the maximized value of the primary outcome, ie., constrained optimal regimes. This method is useful, for example, when the clinicians need to find an adaptive intervention strategy that maximize the effectiveness and controls the side-effect burden simultaneously. This chapter focuses on constrained optimal regimes under the multiple stage setting. Data are assumed to be from Sequential, Multiple Assignment, Randomized Trials [23]. Observational data can also fit in our framework if the additional assumptions about the treatment assignment mechanism are tenable. However, precaution is needed when using data from observational studies, as one key assumption, the no unmeasured confounder assumption, can not be verified [8].

2.2 Methodology

2.2.1 Define multi-stage constrained optimal treatment regimes

Dataset

The dataset is denoted by

$$\{(\boldsymbol{X}_1^i, A_1^i, \boldsymbol{X}_2^i, A_2^i, \cdots, \boldsymbol{X}_T^i, A_T^i, \boldsymbol{Y}^i)\}_{i=1}^n,$$

which is composed of n identically, independently distributed patient trajectories $\{(\boldsymbol{X}_1, A_1, \boldsymbol{X}_2, A_2, \cdots, \boldsymbol{X}_T, A_T, \boldsymbol{Y})\}$. Capital letters denote random variables; lower case letters denote realized values of these random variables. Let $\boldsymbol{X}_1$ be a patient baseline covariate, A_1 be the first-stage treatment variable, $\boldsymbol{X}_2$ be the patient covariate collected between first decision point and second decision point, A_2 be the second-stage treatment variable. So on, and so forth. Finally, $\boldsymbol{X}_T$ is the patient intermediate outcomes collected at the final decision point T, A_T is the treatment assignment at that time point, and $\boldsymbol{Y}$ is the final outcome vector. For $t = 1, \cdots, T$, $\boldsymbol{X}_t \in \boldsymbol{\mathcal{X}}_t \subseteq \mathbb{R}^{p_t}$, $A_t \in \mathcal{A}_t = \{1, 2, \cdots, m_t\}$, and $\boldsymbol{Y} \in \mathbb{R}^J$. The first component Y_1 denote the primary outcome of interest, which is coded so that larger values are more desirable. Meanwhile, $Y_2, \cdots, Y_J$ are the secondary outcomes of interest, which are coded so that lower is better. Let $\boldsymbol{H}_t$ denote the patient history information up to the decision point t, i.e., $\boldsymbol{H}_1^\intercal = (1, \boldsymbol{X}_1^\intercal)$, $\boldsymbol{H}_2^\intercal = (\boldsymbol{H}_1^\intercal, A_1, \boldsymbol{X}_2^\intercal)$, $\cdots$, $\boldsymbol{H}_t^\intercal = (\boldsymbol{H}_{t-1}^\intercal, A_{t-1}, \boldsymbol{X}_t^\intercal)$, $\cdots$, $\boldsymbol{H}_T^\intercal = (\boldsymbol{H}_{T-1}^\intercal, A_{T-1}, \boldsymbol{X}_T^\intercal)$. Besides, let $\bar{A}_t = (A_1, A_2, \cdots, A_t)$ denotes a sequence of treatment history up to time point t, and $\bar{A}_t \in \bar{\mathcal{A}}_t$, where $\bar{\mathcal{A}}_t = \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_t$, $t = 2, \cdots, T$.

Potential outcomes

The potential outcome or counter-factual framework by Neyman, Rubin and Robins are adopted to identify the causal effect of a regime. The set of potential outcomes is $\boldsymbol{W}^* = \{\boldsymbol{X}_2^*(a_1), \boldsymbol{X}_3^*(\bar{a}_2), \cdots, \boldsymbol{X}_T^*(\bar{a}_{T-1}), \boldsymbol{Y}_T^*(\bar{a}_T), \text{for all } \bar{a}_t \in \bar{\mathcal{A}}_t, t = 1, 2, \cdots, T\}$, where $\boldsymbol{X}_t^*(\bar{a}_{t-1})$ is the potential outcome that would have been observed if the patient followed the treatment history sequence $\bar{a}_{t-1}$. The following three necessary assumptions are necessary to connect observed data with potential outcomes [15, 17, 40, 42, 43].

- *B1. Consistency:* $\boldsymbol{Y} = \boldsymbol{Y}^*(\bar{A}_T)$, and $\boldsymbol{X}_t = \boldsymbol{X}_t^*(\bar{A}_{t-1})$, $t = 2, \cdots, T$.

- *B2. Sequential randomization assumption:* $A_t \perp\!\!\!\perp \boldsymbol{W}^* \mid \boldsymbol{H}_t$ for $t = 1, 2, \cdots, T$.
- *B3. Positivity assumption:* $\exists\, \epsilon_t > 0$, such that $\Pr(A_t = a_t \mid \boldsymbol{H}_t = \boldsymbol{h}_t) > \epsilon_t$, for all $a_t \in \mathcal{A}_t$, $t = 1, 2, \cdots, T$.

B1) states that the intermediate and final outcomes observed equal to the patient's intermediate and final potential outcomes under the sequence of treatment actually assigned. It also implies no interference among individuals. B2) mean that conditional on the observed patient history $\boldsymbol{H}_t$, the treatment at time point t is assigned independently of the his or her potential outcomes. B3) guarantees a positive possibility for any $a_t \in \mathcal{A}_t$ having been assigned to patients with $\boldsymbol{H}_t = \boldsymbol{h}_t$. These assumptions imply that $\Pr(\boldsymbol{Y}^*(\bar{a}_T) \leq \boldsymbol{y} \,|\, \boldsymbol{H}_T^*(\bar{a}_{T-1}) = \boldsymbol{h}_T) = \Pr(\boldsymbol{Y} \leq \boldsymbol{y} \,|\, \boldsymbol{H}_T = \boldsymbol{h}_T, A_T = a_T)$ and $\Pr(\boldsymbol{X}_{t+1}^*(\bar{a}_t) \leq \boldsymbol{x}_{t+1} \,|\, \boldsymbol{H}_t^*(\bar{a}_{t-1}) = \boldsymbol{h}_t^*) = \Pr(\boldsymbol{X}_{t+1} \leq \boldsymbol{x}_{t+1} \mid \boldsymbol{H}_t = \boldsymbol{h}_t, A_t = a_t)$ for $t = 1, 2, \cdots, T-1$. Hence, we can estimate the values of a regime using the observed dataset.

Define constrained optimal dynamic treatment regimes

A dynamic treatment regime, $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_T)$, is a sequence of decision rules. Each decision rule, $\pi_t : \text{supp}(\boldsymbol{H}_t) \to \mathcal{A}_t$, is a function that maps the support of patient history information $\boldsymbol{H}_t$ to the set of all possible treatments at time point t. The final potential outcome under the regime $\boldsymbol{\pi}$ is $\boldsymbol{Y}^*(\boldsymbol{\pi}) = \sum_{\bar{a}_T \in \bar{\mathcal{A}}_T} \boldsymbol{Y}^*(\bar{a}_T)\mathbb{I}(\boldsymbol{\pi} = \bar{a}_T)$, and the intermediate potential outcome under that regime is $\boldsymbol{X}_{t+1}^*(\boldsymbol{\pi}_t) = \sum_{\bar{a}_t \in \bar{\mathcal{A}}_t} \boldsymbol{X}_{t+1}^*(\bar{a}_t)\mathbb{I}(\boldsymbol{\pi}_t = \bar{a}_t)$, where $\boldsymbol{\pi}_t = (\pi_1, \pi_2, \cdots, \pi_t)$. The value of a dynamic treatment regime, $\boldsymbol{V}(\boldsymbol{\pi}) = \mathbb{E}\boldsymbol{Y}^*(\boldsymbol{\pi})$, is defined as the expected final outcome if each patient in the population of interest is treated according to $\boldsymbol{\pi}$. Each component of $\boldsymbol{V}(\boldsymbol{\pi})$ is denoted by $V_j(\boldsymbol{\pi}) = \mathbb{E}Y_j^*(\boldsymbol{\pi})$, for $j = 1, \cdots, J$. Our goal is to find a constrained optimal regime, $\boldsymbol{\pi}_{\boldsymbol{\nu}}^*$, that maximizes the expectation of the primary final potential outcome $V_1(\boldsymbol{\pi})$, subject to an upper bound constraints on the expectation of the secondary final potential outcomes $V_j(\boldsymbol{\pi})$, for $j = 2, \cdots, J$. The $J-1$ dimensional vector of upper bounds is denoted as $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_{J-1})$, which can be determined by the preference of clinicians or patients. Therefore, a multi-stage constrained optimal regime problem is defined as

$$\begin{aligned} &\max_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} V_1(\boldsymbol{\pi}) \\ &\text{subject to } V_j(\boldsymbol{\pi}) \leq \nu_{j-1}, \end{aligned} \tag{2.1}$$

where $j = 2, 3, \cdots, J$ and $\mathbf{\Pi}$ is the class of dynamic treatment regimes under consideration. The feasible space of the class of regimes, $\mathcal{F}(\mathbf{\Pi})$, is the set of all regimes satisfying the constraints. For each $\boldsymbol{\pi} \in \mathcal{F}(\mathbf{\Pi})$, $V_j(\boldsymbol{\pi}) \leq \nu_{j-1}$, for $j = 2, \cdots, J$. Then, a multi-stage constrained optimal regime can also be written as $\boldsymbol{\pi}^*_{\boldsymbol{\nu}} = \text{argmax}_{\boldsymbol{\pi} \in \mathcal{F}(\mathbf{\Pi})} V_1(\boldsymbol{\pi})$.

We choose the class of regime to be the class of linear decision rules, where each mapping function π_t at time point t is indexed by $\boldsymbol{\theta}_t$. More specifically, $\pi_t(\boldsymbol{h}_t) = \text{sgn}(\boldsymbol{h}_t^\intercal \boldsymbol{\theta}_t)$. Hence, all $V_j(\boldsymbol{\pi})$'s can be considered as functions of $\boldsymbol{\theta}$ and can be exchangeably denoted $V_j(\boldsymbol{\theta})$'s. As only the directions of $\boldsymbol{h}_t^\intercal \boldsymbol{\theta}_t$ matters, we restrict $\boldsymbol{\theta}_t$ to be unit vectors, i.e., $\boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t = 1$, for $t = 1, \cdots, T$. Then, problem (2.1) above can be written as

$$\begin{aligned} &\max_{\boldsymbol{\theta} \in \mathbf{\Theta}} V_1(\boldsymbol{\theta}) \\ &\text{subject to} \, V_j(\boldsymbol{\theta}) - \nu_j \leq 0, \\ &\qquad \boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t^\intercal - 1 = 0. \end{aligned} \tag{2.2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_T)$, $j = 2, \cdots, J$ and $t = 1, \cdots, T$. Let the feasible set of $\mathbf{\Theta}$ be $\mathcal{F}(\mathbf{\Theta})$, such that $V_j(\boldsymbol{\theta}) \leq \nu_j$, for any $\boldsymbol{\theta} \in \mathcal{F}(\mathbf{\Theta})$ and $j = 2, \cdots, J$. Then, the corresponding index parameter of a constrained optimal dynamic treatment regime is $\boldsymbol{\theta}^*_{\boldsymbol{\nu}} = \text{argmax}_{\boldsymbol{\theta} \in \mathcal{F}(\mathbf{\Theta})} V_1(\boldsymbol{\theta})$, and $\boldsymbol{\theta}^*_{\boldsymbol{\nu}} = (\boldsymbol{\theta}^*_{\boldsymbol{\nu},1}, \boldsymbol{\theta}^*_{\boldsymbol{\nu},2}, \cdots, \boldsymbol{\theta}^*_{\boldsymbol{\nu},T})$.

2.2.2 Re-define constrained optimal regimes via penalization

Problem (2.2) is a nonlinear constrained continuous optimization task. It is solved via interior point method, where we re-formalize the problem via quadratic-barrier penalization. To re-formalize the problem, we let $v_1(\boldsymbol{\theta}) = -V_1(\boldsymbol{\theta})$ and $v_j(\boldsymbol{\theta}) = V_j(\boldsymbol{\theta}) - \nu_j$ for $j = 2, \cdots, J$. Moreover, $h_t(\boldsymbol{\theta}_t) = \boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t - 1$. Hence, we have problem (2.2) equivalent to the following

$$\begin{aligned} &\min_{\boldsymbol{\theta} \in \mathbf{\Theta}} v_1(\boldsymbol{\theta}) \\ &\text{subject to} \, v_j(\boldsymbol{\theta}) \leq 0, \\ &\qquad h_t(\boldsymbol{\theta}_t) = 0. \end{aligned} \tag{2.3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_T)$, $j = 1, \cdots, J$ and $t = 1, \cdots, T$. Interior point method approximate the solution of problem (2.3) by solving a sequence of the following problem (2.4),

where μ is positive and approaches to zero in the limit. For each $\mu > 0$, the approximate problem is

$$\min_{\boldsymbol{\theta},\boldsymbol{z}} \phi_\mu(\boldsymbol{\theta}, \boldsymbol{z}) = \min v_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln z_j, \text{subject to } v_j(\boldsymbol{\theta}) + z_j = 0, h_t(\boldsymbol{\theta}_t) = 0 \tag{2.4}$$

where $j = 2, \cdots, J$ and $t = 1, \cdots, T$. z_j's are the slack variables, which are restricted to be positive due the ln operator. The logarithmic terms, $\ln z_j$'s, are the barrier functions, which enforce the solution path to be within the feasible region of the problem (2.3). More details on interior points method can be found at section 2.1.2.

The sequence of solutions to problem (2.4) forms a trajectory path $\left\{\boldsymbol{\theta}^*_\nu(\mu)\right\}_{\mu\to 0+}$ that convergence to the solution to problem (2.3) as $\mu \to 0$, i.e., $\lim_{\mu\to 0} \boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu) = \boldsymbol{\theta}^*_{\boldsymbol{\nu}}$. The conditions for its convergence can be found at section 2.1.3. Let $\widehat{\boldsymbol{V}}(\theta)$ be a consistent estimator of the values of a regime π. Then, correspondingly, $\widehat{v}_1(\boldsymbol{\theta}) = -\widehat{V}_1(\boldsymbol{\theta})$ and $\widehat{v}_j(\boldsymbol{\theta}) = \widehat{\boldsymbol{V}}_j(\boldsymbol{\theta}) - \nu_j$ for $j = 2, \cdots, J$. Then, problem (2.4) with the plugin estimator is

$$\min_{\boldsymbol{\theta},\boldsymbol{z}} \widehat{\phi}_\mu(\boldsymbol{\theta}, \boldsymbol{z}) = \min \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln z_j, \text{ subject to } \widehat{v}_j(\boldsymbol{\theta}) + z_j = 0, h_t(\boldsymbol{\theta}_t) = 0, \tag{2.5}$$

for $j = 2, \cdots, J$ and $t = 1, \cdots, T$. The solution to problem (2.5) is equivalent to the solution to penalty-barrier function below.

$$\min_{\boldsymbol{\theta}} \widehat{\phi}^{BP}_\mu(\boldsymbol{\theta}) = \min \widehat{v}_1(\boldsymbol{\theta}) - \mu \sum_{j=2}^{J} \ln(-\widehat{v}_j(\boldsymbol{\theta})) + \frac{1}{2\mu} \sum_{t=1}^{T} (\boldsymbol{\theta}_t^\intercal \boldsymbol{\theta}_t - 1)^2 \tag{2.6}$$

Denote the solution to (2.5/2.6) as $\widehat{\boldsymbol{\theta}}_\nu(\mu)$.We have proven that $\widehat{\boldsymbol{\theta}}_\nu(\mu) \xrightarrow{p} \boldsymbol{\theta}^*_\nu(\mu)$. For the consistency of this estimator, the details and proof are provided in section 2.1.4 and appendix A.2.

2.2.3 Estimation of the values of a regime

To estimate the values of a regime, we use the G-computation formula by Robins, etc [13]. For any arbitrary regime $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_T)$, assume the three causal assumptions B1)-B3)

are satisfied, then for each component of $\boldsymbol{Y}$

$$\begin{aligned}
&\Pr\left(Y_j^*(\boldsymbol{\pi}) \leq y_j\right) = F_{Y_j^*(\boldsymbol{\pi})}\left(y_j\right) \\
&= \int \cdots \int F_{Y_j|\boldsymbol{H}_T,A_T}\left(y_j|\boldsymbol{h}_T, \pi_T(\boldsymbol{h}_T)\right)\, dF_{\boldsymbol{H}_T|\boldsymbol{H}_{T-1},A_{T-1}}\left(\boldsymbol{h}_T|\boldsymbol{h}_{T-1}, \pi_{T-1}(\boldsymbol{h}_{T-1})\right) \\
&\quad dF_{\boldsymbol{H}_{T-1}|\boldsymbol{H}_{T-2},A_{T-2}}\left(\boldsymbol{h}_{T-1}|\boldsymbol{h}_{T-2}, \pi_{T-2}(\boldsymbol{h}_{T-2})\right) \cdots dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left(\boldsymbol{h}_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right)\, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1)
\end{aligned}$$

where $F_{Y_j|\boldsymbol{H}_T,A_T}\left(\cdot \mid \cdot,\cdot\right)$ is the conditional cumulative density function of Y_j conditioning on $\boldsymbol{H}_T$ and A_T, $F_{\boldsymbol{H}_t|\boldsymbol{H}_{t-1},A_{t-1}}\left(\cdot \mid \cdot,\cdot\right)$ the conditional cumulative density function of $\boldsymbol{H}_t$ conditioning on $\boldsymbol{H}_{t-1}, A_{t-1}$, and $F_{\boldsymbol{H_1}}(\cdot)$ the cumulative density function of $\boldsymbol{H}_1$. Thus, the marginal distribution of the potential outcomes under any regime $\boldsymbol{\pi}$ can be estimated from observed data, if we can estimate the conditional distributions involved. However, the estimation of the sequence of conditional distribution could being a daunting task. Linn et al. used a two-step estimator via mean and variance modeling to construct two-stage constrained optimal dynamic treatment regimes [18], and it is demonstrated in the following simulation studies. However, the modeling becomes complex rapidly as the number of stages increases. Note the regime is index by $\boldsymbol{\theta}$, we use $F_{Y_j^*(\boldsymbol{\pi})}\left(y_j\right)$ and $F_{Y_j^*(\boldsymbol{\theta})}\left(y_j\right)$ interchangeably.

2.2.4 Asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$

The asymptotic properties of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ here is similar to the corresponding part for one-stage problem in Chapter 1 (Section 1.1.6).

Limiting distribution of $\nabla\widehat{V}_j(\boldsymbol{\theta})$

Before we derive the limiting distribution of the estimator $\widehat{\boldsymbol{\theta}}_{\kappa}(\mu)$, we need to examine, for any fixed value of $\boldsymbol{\theta} : \boldsymbol{\theta}_1^{\intercal}\boldsymbol{\theta}_1 = 1$ and $\boldsymbol{\theta}_2^{\intercal}\boldsymbol{\theta}_2 = 1$, the limiting distribution of $\nabla\widehat{V}_j(\boldsymbol{\theta})$,

where

$$\begin{aligned}\nabla \widehat{V}_j(\boldsymbol{\theta}) =& \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y) \\ =& \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d \left(\frac{1}{n} \sum_{i=1}^{n} \widehat{F}_{Y^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) \right) \\ =& \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}),\end{aligned}$$

where $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(\cdot)$ denote the estimator of $F_{Y_j^*(\boldsymbol{\theta})}(\cdot)$

Lemma 2.2.1. *Suppose the following conditions hold.*

1. $\forall \boldsymbol{a} \in \mathbb{R}^p$,$\exists \delta > 0$,*such that*
   (a) $\mathbb{E} \left| \boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) \right|^{2+\delta} < \infty$
   (b) $\left\{ \boldsymbol{a}^\intercal Var \left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) \right] \boldsymbol{a} \right\}^{1+\frac{\delta}{2}} < \infty.$

Then, we have, for any fixed $\boldsymbol{\theta}$*,*

The proof of this is similar to the proof of Lemma 1.1.3 and is shown in Appendix B.1. Assume $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ is consistent, and the following corollary shows that the estimations do not effect the limiting distribution obtained above.

Corollary 2.2.2. *Suppose all the assumptions in Lemma 2.2.1 hold, and* $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ *is a consistent estimator of* $F_{Y_j^*(\boldsymbol{\theta})}(y_j|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$*. Then, we have*

$$\sqrt{n} \left(\nabla \widehat{V}_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)) - \nabla V_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)) \right) \xrightarrow{d} \mathcal{N} \left(0, Avar \left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y_j \, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)} \right) \right)$$

See Appendix B.2 for proof.

Limiting distribution of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$

Now, we investigate the limiting distribution of $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$.

Theorem 2.2.3. *Suppose all the assumptions above hold. Then we have, as* $n \to \infty$

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) - \boldsymbol{\theta}_{\boldsymbol{\nu}}(\mu)^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^*),$$

where $\boldsymbol{\Sigma}^* = \boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$,
$\boldsymbol{C}^* = \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\nabla^\intercal v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right) - \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right)\mathbb{E}\left(\nabla^\intercal v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right)$,
and $\boldsymbol{D}^* = \nabla^2\phi^{BP}_\mu(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu))$.

The proof is similar to the proof of Theorem 1.1.5, and is presented in Appendix B.3.

2.3 Simulation

2.3.1 Simulation design

We demonstrate our proposed method using the toy example presented by Linn et al [18], where there are two competing outcomes Y and Z. The goal is to maximize the mean of Y, subject to an upper bound on the mean of Z. Y is coded so that the higher the value the better, such as the effectiveness of the treatment regimes. Meanwhile, Z is coded the lower the better, such as the side-effect burden. The model for generating the patient trajectories $(X_1, A_1, X_2, A_2, Y, Z)$ are as follow:

$$\begin{aligned}
&X_1 \sim \text{Normal}(1,1),\\
&\boldsymbol{H}_1 = (1, X_1)^\intercal,\\
&A_1 \sim \text{Uniform}\{-1,1\},\\
&X_2 = \boldsymbol{H}_1^\intercal\boldsymbol{\beta}_{1,0} + A_1\boldsymbol{H}_1^\intercal\boldsymbol{\beta}_{1,1} + \epsilon,\\
&\epsilon \sim \text{Normal}(0,1),\\
&\boldsymbol{H}_2 = (1, X_2)^\intercal,\\
&A_2 \sim \text{Uniform}\{-1,1\},\\
&Y = \boldsymbol{H}_2^\intercal\boldsymbol{\beta}_{2,0,Y} + A_2\boldsymbol{H}_2^\intercal\boldsymbol{\beta}_{2,1,Y} + \epsilon_Y\\
&Z = \boldsymbol{H}_2^\intercal\boldsymbol{\beta}_{2,0,Z} + A_2\boldsymbol{H}_2^\intercal\boldsymbol{\beta}_{2,1,Z} + \epsilon_Z\\
&(\epsilon_Y, \epsilon_Z)^\intercal \sim \text{Normal}(\mathbf{0}_2, \Sigma_{Y,Z})
\end{aligned}$$

This model is a simple representation of the data from a two-stage randomized SMART. Variable X_1 represents the summary of patient status before the first treatment as-

signment A_1. Variable X_2 represents the summary of patient status before the second treatment assignment A_2. The parameters involved are set to the following,

$$\begin{aligned}
\boldsymbol{\beta}_{1,0} &= (0.5, 0.75)^\intercal \\
\boldsymbol{\beta}_{1,1} &= (0.25, 0.5)^\intercal \\
\boldsymbol{\gamma}_0 &= (0.25, -0.05)^\intercal \\
\boldsymbol{\gamma}_1 &= (0.1, -0.05)^\intercal \\
\boldsymbol{\beta}_{2,0,Y} &= (30, 2)^\intercal \\
\boldsymbol{\beta}_{2,1,Y} &= (5, -1.5)^\intercal \\
\boldsymbol{\beta}_{2,0,Z} &= (15, 1)^\intercal \\
\boldsymbol{\beta}_{2,1,Z} &= (3, -0.5)^\intercal \\
\Sigma_{Y,Z} &= \begin{bmatrix} 1.0, & 0.7 \\ 0.7, & 1.0 \end{bmatrix}
\end{aligned}$$

The class of regimes under consideration is restricted to linear decision rules at each stage. That is $\pi_1 = \text{sgn}(\boldsymbol{h}_1^\intercal \boldsymbol{\theta}_1)$ and $\pi_2 = \text{sgn}(\boldsymbol{h}_2^\intercal \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the index parameters for the regimes. The true optimal regimes are denoted by $\pi_1^* = \text{sgn}(\boldsymbol{h}_1^\intercal \boldsymbol{\theta}_1^*)$ and $\pi_2^* = \text{sgn}(\boldsymbol{h}_2^\intercal \boldsymbol{\theta}_2^*)$. The estimated optimal regimes are denoted by $\widehat{\pi}_1 = \text{sgn}(\boldsymbol{h}_1^\intercal \widehat{\boldsymbol{\theta}}_1)$ and $\widehat{\pi}_2 = \text{sgn}(\boldsymbol{h}_2^\intercal \widehat{\boldsymbol{\theta}}_2)$. Here, the sgn function is defined as

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0. \end{cases}$$

2.3.2 Modeling and estimation

Modeling for and estimation of the distributions of potential outcomes

The distribution of potential outcomes are unknown. The two major quantities under an arbitrary regime $\boldsymbol{\pi}$ involved, $\mathbb{E}Y^*(\boldsymbol{\pi})$ and $\mathbb{E}Z^*(\boldsymbol{\pi})$, need to be estimated from the observed data. Our strategy for estimating these two quantities is to model the marginal distribution of each potential outcome, and then draw random samples from the estimated marginal distributions to calculate their expectations numerically. To connect observed data with potential outcomes, three necessary causal inference assumptions *B1)-B3).* are assumed to hold.

Following the G-computation formula [13], we have, for any arbitrary regime $\boldsymbol{\pi} = (\pi_1, \pi_2)$, that

$$\Pr\{Y^*(\boldsymbol{\pi})) \leq y\} = \mathbb{E}_{\boldsymbol{H}_1}\left\{\mathbb{E}_{\boldsymbol{H}_2}\left[\Pr\left\{Y \leq y \mid \boldsymbol{H}_2, A_2 = \pi_2(\boldsymbol{H}_2), \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right\} \mid \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right]\right\}$$

and similarly,

$$\Pr\{Z^*(\boldsymbol{\pi}) \leq z\} = \mathbb{E}_{\boldsymbol{H}_1}\left\{\mathbb{E}_{\boldsymbol{H}_2}\left[\Pr\left\{Z \leq z \mid \boldsymbol{H}_2, A_2 = \pi_2(\boldsymbol{H}_2), \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right\} \mid \boldsymbol{H}_1, A_1 = \pi_1(\boldsymbol{H}_1)\right]\right\}$$

Hence, we can estimate the probabilty function of the potential outcomes under a regime $\boldsymbol{\pi}$, $\Pr\{Y^*(\boldsymbol{\pi}) \leq y\}$ and $\Pr\{Z^*(\boldsymbol{\pi}) \leq z\}$, using observed data by modeling and estimating the conditional distributions involved, and hence, $\mathbb{E}Z^*(\boldsymbol{\pi})$ and $\mathbb{E}Z^*(\boldsymbol{\pi})$.

Following the modeling tactic in "Constrained estimation for competing outcomes" by Linn et al [18]. We assume the following model,

$$\begin{aligned}
&Y = \mathbb{E}(Y|\boldsymbol{H}_2, A_2) + \varepsilon_Y, \\
&\mathbb{E}(Y|\boldsymbol{H}_2, A_2) = m_Y(\boldsymbol{H}_2) + A_2 c_Y(\boldsymbol{H}_2), \\
&\text{where } \mathbb{E}(\varepsilon_Y) = 0, \mathrm{Var}(\varepsilon_Y) = \sigma^2, \text{and } \varepsilon_Y \perp\!\!\!\perp (\boldsymbol{H}_2, A_2).
\end{aligned}$$

Define $F_{\varepsilon_Y}(\cdot)$ to be the distribution of ε_Y; $F_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}(\cdot \mid \boldsymbol{h}_1, a_1)$ to be the conditional distribution of $\boldsymbol{H}_2$ given $\boldsymbol{H}_1 = \boldsymbol{h}_1$ and $A_1 = a_1$; $F_{\boldsymbol{H}_1}(\cdot)$ to be the distribution of $\boldsymbol{H}_1$. Again, we have $\boldsymbol{H}_1^\intercal = (1, \boldsymbol{X}_1^\intercal)$, $\pi_1(\boldsymbol{H}_1)$, $\boldsymbol{H}_2 = \{\boldsymbol{H}_1^\intercal, \pi_1(\boldsymbol{H}_1), \boldsymbol{X}_2^\intercal\}^\intercal$.

$$\begin{aligned}
&\Pr\left\{Y \leq y \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\} \\
=&\Pr\left\{m(\boldsymbol{H}_2) + \pi_2(\boldsymbol{H}_2)c_Y(\boldsymbol{H}_2) + \varepsilon_Y \leq y \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\} \\
=&\Pr\left\{\varepsilon_Y \leq y - m(\boldsymbol{H}_2) - \pi_2(\boldsymbol{H}_2)c_Y(\boldsymbol{H}_2) \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\} \\
=&F_{\varepsilon_Y}\left\{y - m(\boldsymbol{h}_2) - \pi_2(\boldsymbol{h}_2)c_Y(\boldsymbol{h}_2)\right\} \\
=&F_{\varepsilon_Y}\left[y - m(\boldsymbol{h}_2) - \mathrm{sgn}\left\{r_2(\boldsymbol{h}_2; \boldsymbol{\theta}_2)\right\} c_Y(\boldsymbol{h}_2)\right]
\end{aligned}$$

Hence, we have

$$
\begin{aligned}
&\Pr\left\{Y^*(\boldsymbol{\pi}) \le y\right\} \\
&= \iint \Pr\left\{Y \le y \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, A_2 = \pi_2(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2|\pi_1(\boldsymbol{h_1}), \boldsymbol{h}_1\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Y}\left\{y - m(\boldsymbol{h}_2) - \pi_2(\boldsymbol{h}_2)c_Y(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2 \mid \boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Y}\left[y - m(\boldsymbol{h}_2) - \operatorname{sgn}\left\{r_2(\boldsymbol{h}_2;\boldsymbol{\theta}_2)\right\} c_Y(\boldsymbol{h}_2)\right] \, dG_Y\left\{m_Y, c_Y, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Y}\left[y - m(\boldsymbol{h}_2) - \operatorname{sgn}(r_2)c_Y(\boldsymbol{h}_2)\right] \, dG_Y\left\{m_Y, c_Y, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1)
\end{aligned}
$$

where $G_Y\left\{m_Y, c_Y, r_2 \mid \boldsymbol{h}_1, a_1\right\}$ is the joint conditional distribution of $m_Y(\boldsymbol{H}_2)$, $c_Y(\boldsymbol{H}_2)$ and $r_2(\boldsymbol{H}_2;\boldsymbol{\theta}_2)$ given $\boldsymbol{H}_1 = \boldsymbol{h}_1$ and $A_1 = a_1$. The second equality is due to $\int z(x,y)\, dF_{X|Y}(x|y) = \mathbb{E}(z|y) = \int z\, dF_{Z|Y}(z|y)$.

Same applies to Z:

$$
\begin{aligned}
&Z = \mathbb{E}(Z|\boldsymbol{H}_2, A_2) + \epsilon, \\
&\text{where } \mathbb{E}(\epsilon) = 0, \operatorname{Var}(\epsilon) = \sigma^2, \text{and } \epsilon \perp\!\!\!\perp (\boldsymbol{H}_2, A_2) \\
&\mathbb{E}(Z|\boldsymbol{H}_2, A_2) = m_Z(\boldsymbol{H}_2) + A_2 c_Z(\boldsymbol{H}_2)
\end{aligned}
$$

$$
\begin{aligned}
&\Pr\left\{Z^*(\boldsymbol{\pi}) \le z\right\} \\
&= \iint \Pr\left\{Z \le z \mid \boldsymbol{H}_2 = \boldsymbol{h}_2, \pi_2(\boldsymbol{H}_2) = \pi_2(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2|d_1(\boldsymbol{h_1}), \boldsymbol{h}_1\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Z}\left\{z - m(\boldsymbol{h}_2) - \pi_2(\boldsymbol{h}_2)c_Z(\boldsymbol{h}_2)\right\} \, dF_{\boldsymbol{H}_2|\boldsymbol{H}_1,A_1}\left\{\boldsymbol{h}_2 \mid \boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Z}\left[z - m(\boldsymbol{h}_2) - \operatorname{sgn}\left\{r_2(\boldsymbol{h}_2;\boldsymbol{\theta}_2)\right\} c_Z(\boldsymbol{h}_2)\right] \, dG_Z\left\{m_Z, c_Z, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1) \\
&= \iint F_{\varepsilon_Z}\left[z - m(\boldsymbol{h}_2) - \operatorname{sgn}(r_2)c_Z(\boldsymbol{h}_2)\right] \, dG_Z\left\{m_Z, c_Z, r_2|\boldsymbol{h}_1, \pi_1(\boldsymbol{h}_1)\right\} \, dF_{\boldsymbol{H}_1}(\boldsymbol{h}_1)
\end{aligned}
$$

where $G_Z\left\{m_Z, c_Z, r_2 \mid \boldsymbol{h}_1, a_1\right\}$ is the joint conditional distribution of $m_Z(\boldsymbol{H}_2)$, $c_Z(\boldsymbol{H}_2)$ and $r_2(\boldsymbol{H}_2;\boldsymbol{\theta}_2)$ given $\boldsymbol{H}_1 = \boldsymbol{h}_1$ and $A_1 = a_1$.

We model the joint distribution of $\left\{m_Y(\boldsymbol{H}_2), c_Y(\boldsymbol{H}_2), m_Z(\boldsymbol{H}_2), c_Z(\boldsymbol{H}_2)\right\}$ by modeling

the joint distribution of the standardized residuals obtained from the mean and variance modeling of each component for given $\boldsymbol{H}_1$ and A_1

$$
\begin{aligned}
e_Y^m &= \frac{m_Y(\boldsymbol{H}_2) - \mu_Y^m(\boldsymbol{H}_1, A_1)}{\sigma_Y^m(\boldsymbol{H}_1, A_1)} \\
e_Y^c &= \frac{c_Y(\boldsymbol{H}_2) - \mu_Y^c(\boldsymbol{H}_1, A_1)}{\sigma_Y^c(\boldsymbol{H}_1, A_1)} \\
e_Z^m &= \frac{m_Z(\boldsymbol{H}_2) - \mu_Z^m(\boldsymbol{H}_1, A_1)}{\sigma_Z^m(\boldsymbol{H}_1, A_1)} \\
e_Z^c &= \frac{c_Z(\boldsymbol{H}_2) - \mu_Z^c(\boldsymbol{H}_1, A_1)}{\sigma_Z^c(\boldsymbol{H}_1, A_1)} \\
e_{f_2} &= \frac{f_2(\boldsymbol{H}_2) - \mu_{f_2}(\boldsymbol{H}_1, A_1)}{\sigma_{f_2}(\boldsymbol{H}_1, A_1)}
\end{aligned}
$$

The mean functions are defined as

$$
\begin{aligned}
\mu_Y^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{m_Y(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_Y^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{c_Y(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_Z^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{m_Z(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_Z^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{c_Z(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\} \\
\mu_{f_2}(\boldsymbol{H}_1, A_1) &= \mathbb{E}\{f_2(\boldsymbol{H}_2) \mid \boldsymbol{H}_1, A_1\}
\end{aligned}
$$

and the standard deviation functions are defined as

$$
\begin{aligned}
\sigma_Y^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{m_Y(\boldsymbol{H}_2) - \mu_Y^m(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_Y^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{c_Y(\boldsymbol{H}_2) - \mu_Y^c(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_Z^m(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{m_Z(\boldsymbol{H}_2) - \mu_Z^m(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_Z^c(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{c_Z(\boldsymbol{H}_2) - \mu_Z^c(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2} \\
\sigma_{f_2}(\boldsymbol{H}_1, A_1) &= \mathbb{E}\big[\{f_2(\boldsymbol{H}_2) - \mu_{f_2}(\boldsymbol{H}_1, A_1)\}^2 \mid \boldsymbol{H}_1, A_1\big]^{1/2}
\end{aligned}
$$

Therefore, we model the joint distribution of the standardized residuals $(e_Y^m, e_Y^c, e_Z^m, e_Z^c, e_{f_2})$ to obtain an estimator of $G_{Y,Z}^{\pi}(\cdot, \cdot, \cdot, \cdot, \cdot \mid x_1, a_1)$.

Due to the cost of clinical data, sample sizes are usually small. We consider parametric

models for $m_Y(\boldsymbol{H}_2), c_Y(\boldsymbol{H}_2), m_Z(\boldsymbol{H}_2), c_Z(\boldsymbol{H}_2)$ and $f_2(\boldsymbol{H}_2)$. Here, we model $m_Y(\boldsymbol{H}_2) = \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_1 + A_1 \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_2 + \varepsilon$, where ε is a mean-zero error term. Then, $\mu_Y^m(\boldsymbol{H}_1, A_1) = \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_1 + A_1 \boldsymbol{H}_1^\intercal \boldsymbol{\alpha}_2$.

To estimate, we fit the corresponding least squares regressions, and estimate the residuals empirically. For more details, see reference [18].

2.3.3 Summary of simulation results

We summarize the simulation results here. Figure 2. below shows the estimated optimal regime values and their standard deviation. Figure 2.1 is the efficient frontier plot. The red dashed line represents $\widehat{V}_1$ under estimated constrained optimal regime, and the blue dash-dotted line represents $\widehat{V}_2$ under that regime. The plot represents the best possible value of the primary potential outcome for its level of risk, which is the value of the secondary potential outcome. In the plot, the value of the primary outcome increases as the constraint bound gets looser. Meanwhile the value of the secondary outcome keep up with the constraint, until the constraint is not active. Once the constraint gets larger than the maximum value of the secondary potential outcome, the constrained problem becomes an unconstrained problem.

Table 2.1: Simulation results

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$
12.86	27.48	0.46	12.86	0.23
13.45	28.90	0.42	13.37	0.15
14.03	30.38	0.64	13.89	0.17
14.62	31.85	0.83	14.46	0.16
15.21	33.53	0.64	15.05	0.13
15.79	34.47	0.70	15.64	0.14
16.38	35.47	0.94	16.15	0.29
16.97	36.33	0.93	16.68	0.40
17.55	37.08	0.87	17.31	0.41
18.14	37.62	0.51	17.89	0.31
18.72	37.79	0.72	18.32	0.44
19.31	38.03	0.17	18.91	0.10
19.90	38.03	0.17	18.91	0.10
20.48	38.03	0.17	18.91	0.10
21.07	38.03	0.17	18.91	0.10
21.66	38.03	0.17	18.91	0.10
22.24	38.03	0.17	18.91	0.10
22.83	38.03	0.17	18.91	0.10
23.41	38.03	0.17	18.91	0.10
24.00	38.03	0.17	18.91	0.10

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest.

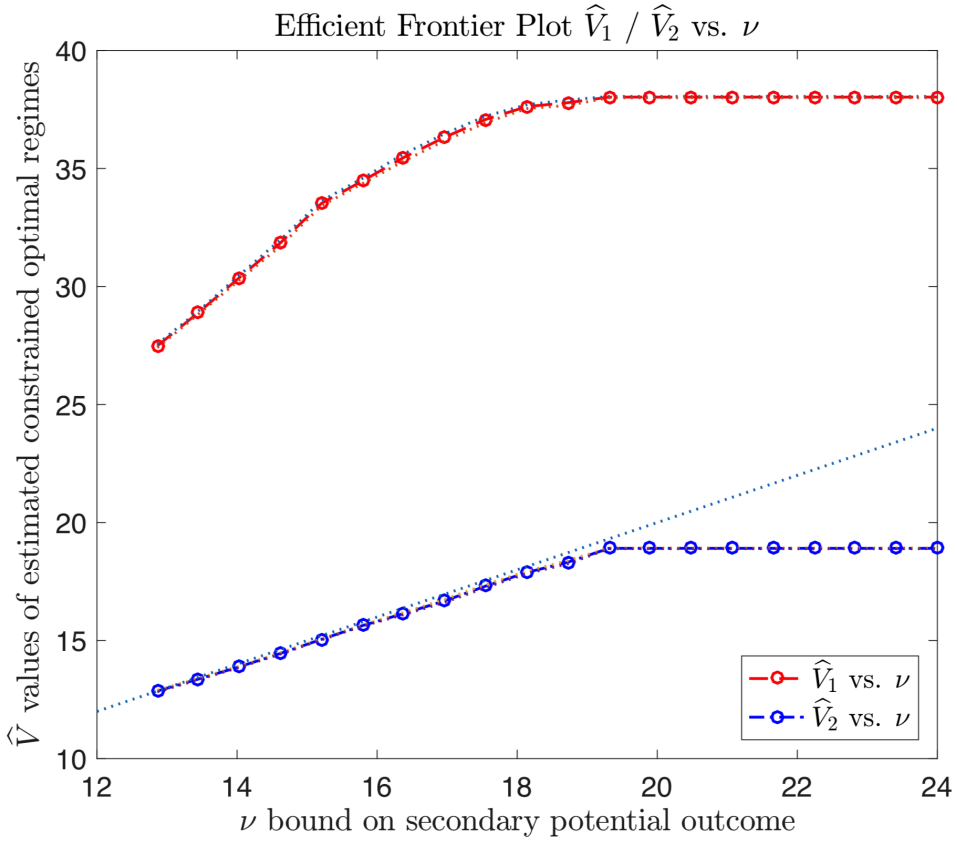

Figure 2.1: Efficient frontier for estimated constrained optimal regimes (multi-stage)

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

2.4 Conclusion

Focusing on optimizing a single scalar outcome may be an oversimplification of the goals of practical clinical decision making. In this chapter, a new method is proposed to handle multiple competing outcomes in the multi-stage setting. Estimating an optimal treatment regime with competing outcomes is cast as a constrained optimization problem. We maximize the primary outcome of interest, subject to the constraints on the secondary

outcomes of interest. Our estimator of a constrained optimal treatment regime has the properties of consistency and asymptotic normality under mild regularity conditions. The efficient frontier plots provide an intuitive visualization for clinicians to examine the trade-off between two competing outcomes.

Chapter 3

Infinite-stage Constrained Optimal Treatment Regimes

3.1 Introduction

Precision medicine aims to accommodate interventions to individual patient attributes. For chronic illnesses, such as cancer, diabetes and so on, clinicians often need to make sequences of treatment decisions based on the evolving status of the patient's condition. To personalize this multi-stage intervention process, researchers have been developing dynamic treatment regimes (DTRs) to inform clinicians of treatment decisions adaptively. DTRs are a sequential decision making process, of which each decision is made based on the evolving patient status with the goal of maximizing the overall long-term treatment efficacy. It is well-studied in the statistical and biomedical literature [20,25,31,32,34,49]. From the standpoint of Markov decision process, reinforcement learning algorithms, such as Q-learning [34], A-learning [5], V-learning [29] etc., are developed to estimate optimal treatment regimes.

Most previous methods for construct optimal dynamic treatment regimes have focused on optimizing a scalar measurement of the long-term efficacy over a fixed time period (finite stage). However, in practice, the clinical situations are more complex. First, they often require consideration of the trade-off among multiple objectives, e.g., effectiveness, side-effect, cost, and so on. The preference of those objectives varies among people and changes over time, thus a single scalar reward or value can not represent the qual-

ity of a policy well enough. Thus, considering multiple rewards are necessary. Previous works by Lizotte et al. learn the value function and optimal policy for all preferences, i.e., all the possible convex combination of all the rewards [26, 27]. More recent works adopt multi-objective Markov decision processes (MOMDPs) framework with finite stage. Practical domination is proposed for flexibility based on Pareto domination, and a set of policies that are maximal based on the partial order are treated as indistinguishably optimal [19,28]. Secondly, patients with chronic diseases are often monitored and treated throughout their life. It often requires taking real-time actions and has no a-priori fixed end point (infinite stage), and progress is made in infinite stage reinforcement learning for health applications [10, 29, 33].

To deal with the trade-off between multiple objectives, we take a different perspective and adopt the constrained Markov decision processes (CMDPs) framework with infinite horizon. CMDPs are a well-studied framework for reinforcement learning under constraints [2]. The goal is to find the optimal policy, while satisfying constraints on expectations of secondary costs. For many applications of reinforcement learning, the constrained approach is more intuitive and more practical than eliciting a single reward function in order to achieve desirable results. For instance, satisfying safety constraints is necessary for systems that physically interact with humans. Previously, linear programming is used to seek constrained optimal policies in the setting of finite CMDPs with known models. However, few methods have been proposed for high-dimensional constrained reinforcement learning problems without modeling the underlying dynamics. Recently, Achiam et al [1] proposed constrained policy optimization, a general-purpose policy search algorithm for constrained reinforcement learning guaranteeing near-constraint satisfaction at each iteration. Taking into consideration properties of clinical applications, such as data scarcity and off-policy learning, we develop an algorithm by taking advantage of least-squares policy evaluation and interior-point methods for estimating constrained optimal dynamic treatment regimes. Our method is applied to a simulated cancer trial dataset based on a chemotherapy mathematical model.

3.2 Methodology

3.2.1 Set-up

Observed Data

We use dataset observed over a finite length of time steps to construct a regime in the setting of infinite horizon Markov decision process. The structure of the available data is $\boldsymbol{D} = \left\{(\boldsymbol{S}_0^i, A_0^i, \boldsymbol{R}_0^i, \boldsymbol{S}_1^i, \cdots, \boldsymbol{S}_{T_i-1}^i, A_{T_i-1}^i, \boldsymbol{R}_{T_i-1}^i, \boldsymbol{S}_{T_i}^i)\right\}_{i=1}^n$, a set of n independent, identically distributed trajectories of $(\boldsymbol{S}_0, A_0, \boldsymbol{R}_0, \boldsymbol{S}_1, \cdots, \boldsymbol{S}_{T-1}, A_{T-1}, \boldsymbol{R}_{T-1}, \boldsymbol{S}_T)$. Note $T \in \mathbb{N}$ denotes the total number of follow-up time steps for a patient. For each patient, his or her follow up time length T_i may be different. $\boldsymbol{S}_t \in \boldsymbol{\mathcal{S}}$ denotes a vector of patient clinical information recorded up to and including time point t, referred as *state* in the reinforcement learning vocabulary. $\boldsymbol{\mathcal{S}} \subseteq \mathbb{R}^m$ denotes the support for state variable. Adopting the time homogeneous Markov decision process model, we assume $\boldsymbol{\mathcal{S}}$ is the same across all time points t. $A_t \in \mathcal{A}$ denotes the treatment assignment at time point t after measuring $\boldsymbol{S}_t$, referred as *action* in reinforcement learning. $\mathcal{A}$ denotes the support for treatment assignment, a finite set of all possible treatment options, and is assumed to be the same across all time points t. $\boldsymbol{R}_t \in \mathbb{R}^J$ is the reward obtained after treatment A_t is assigned. We assume the reward, possibly defined based on domain expertise, is a known vector-valued function $\boldsymbol{r}$: $\boldsymbol{\mathcal{S}} \times \mathcal{A} \times \boldsymbol{\mathcal{S}} \to \mathbb{R}^J$, so that $\boldsymbol{R}_t = \boldsymbol{r}(\boldsymbol{S}_t, A_t, \boldsymbol{S}_{t+1})$. The vector-valued reward function is the same across all the time points t as well. Moreover, if a patient dies during the follow-up, say at decision point t, we set $\boldsymbol{S}_t = \boldsymbol{\emptyset}$, referred to as the absorbing state in reinforcement learning. Then, the patient's treatment assignment at time t is $A_t = \emptyset$, and his/er length of follow-up $T = t$.

Potential outcomes

In reality, a patient can only be assigned to one of the sequences of treatment assignments. Hence, we can only observe the consequence of that treatment sequence, while the others remain unobserved. To identify the average causal effect of a certain regime, we adopt the counter-factual or potential outcomes framework, established by Neyman, Rubin and Robins for assessment of the time-dependent treatment effect from either randomized or observational studies [17, 40, 42]. Let $\bar{\boldsymbol{a}}_t = (a_0, a_1, \cdots, a_t) \in \bar{\boldsymbol{\mathcal{A}}}_t$ be a possible treatment assignment sequence up to time point t, $t \geq 0$, where $\bar{\boldsymbol{\mathcal{A}}}_t = \mathcal{A} \times \cdots \times \mathcal{A}$

is the set of all possible the treatment assignment sequences up to time point t. Let $\bar{\boldsymbol{s}}_t = (\boldsymbol{s}_0, \boldsymbol{s}_1, \cdots, \boldsymbol{s}_t) \in \bar{\boldsymbol{\mathcal{S}}}_t$ be a possible state sequences up to time point t, $t \geq 0$, where $\bar{\boldsymbol{\mathcal{S}}}_t = \boldsymbol{\mathcal{S}} \times \cdots \times \boldsymbol{\mathcal{S}}$ is the set of all possible state sequences up to time point t. The set of potential outcomes is $\boldsymbol{W}^* = \left\{\boldsymbol{S}_1^*(a_0), \boldsymbol{S}_2^*(\bar{\boldsymbol{a}}_1), \cdots, \boldsymbol{S}_{t+1}^*(\bar{\boldsymbol{a}}_t), \cdots, \text{ for all } \bar{\boldsymbol{a}}_\infty \in \bar{\boldsymbol{\mathcal{A}}}_\infty\right\}$, where $\boldsymbol{S}_{t+1}^*(\bar{\boldsymbol{a}}_t)$ is the potential state at $(t+1)$-th time point that would have been observed if the individual had been assigned the treatment sequence $\bar{\boldsymbol{a}}_t$, $t \geq 0$. Moreover, if $\boldsymbol{S}_{t+1}^*(\bar{\boldsymbol{a}}_t) = \boldsymbol{\emptyset}$ happens at time point $t+1$, then $\boldsymbol{S}_{t+2}^*(\bar{\boldsymbol{a}}_{t+1}) = \boldsymbol{S}_{t+3}^*(\bar{\boldsymbol{a}}_{t+2}) = \cdots = \boldsymbol{\emptyset}$, which indicates the patient, if followed treatment assignment sequence $\bar{\boldsymbol{a}}_t$, would have died at time point $t+1$ in the counter-factual world. Rewards with respect to potential states are $\boldsymbol{R}_t^* = \boldsymbol{r}(\boldsymbol{S}_t^*, A_t, \boldsymbol{S}_{t+1}^*)$. The following assumptions are made in the potential outcome framework [15, 17, 40, 42, 43].

- *A1. Consistency*: $\boldsymbol{S}_{t+1} = \boldsymbol{S}_{t+1}^*(\bar{\boldsymbol{A}}_t)$, for all $t \geq 0$.
- *A2. Sequential randomization assumption*: $A_{t+1} \perp\!\!\!\perp \boldsymbol{W}^* \mid \bar{\boldsymbol{S}}_{t+1}, \bar{\boldsymbol{A}}_t$, for all $t \geq 0$.
- *A3. Positivity*: there exists $\epsilon_0 > 0$, so that $P\left(A_{t+1} = a_{t+1} \mid \bar{\boldsymbol{S}}_{t+1} = \bar{\boldsymbol{s}}_{t+1}, \bar{\boldsymbol{A}}_t = \bar{\boldsymbol{a}}_t\right) > \epsilon_0$, for all $a_{t+1} \in \mathcal{A}$, $\bar{\boldsymbol{a}}_t \in \bar{\boldsymbol{\mathcal{A}}}_t$ and $\bar{\boldsymbol{s}}_{t+1} \in \bar{\boldsymbol{\mathcal{S}}}_{t+1}$, and all $t \geq 0$.

These assumptions link the potential outcome and the observed data, and are guaranteed in well-designed Sequential, Multiple Assignment, Randomized Trials (SMARTs). Therefore, the observed data from those trials are used to infer the average causal effect of a regime of interest.

Markov Decision Processes

To construct a regime in the infinite-horizon setting using a dataset observed over a finite number of time steps, we assume that the underlying dynamics is a time homogeneous Markov Decision Processes (MDPs). In infinite-horizon setting, MDP is considered as a 5-tuple of $(\boldsymbol{\mathcal{S}}, \mathcal{A}, \mathbb{P}, \boldsymbol{R}, \gamma)$, where $\boldsymbol{\mathcal{S}}$, $\mathcal{A}$ and $\boldsymbol{R}$ is as described above. Additionally, $\mathbb{P}$ is a markovian transition model in which $p(\boldsymbol{s}' \mid \boldsymbol{s}, a)$ denotes the probability density of a transition to state $\boldsymbol{s}'$ when taking action a in state $\boldsymbol{s}$. A discount factor for future reward, $\gamma \in [0, 1)$, is also introduced to form total discounted rewards, which are the value functions to operate constrained optimization on. The following assumptions are made for infinite-stage time homogeneous Markov decision process.

- *A4. Markov assumption*: $\boldsymbol{S}_{t+1} \perp\!\!\!\perp \left(\bar{\boldsymbol{A}}_{t-1}, \bar{\boldsymbol{S}}_{t-1}\right) \mid (A_t, \boldsymbol{S}_t)$, for all $t \geq 1$.

- *A5. Time homogeneity*: the conditional density $P_t(\boldsymbol{S}_{t+1} = \boldsymbol{s}' \mid A_t = a, \boldsymbol{S}_t = \boldsymbol{s}) = p(\boldsymbol{s}' \mid a, \boldsymbol{s})$ for all $\boldsymbol{s}, \boldsymbol{s}' \in \boldsymbol{\mathcal{S}}$ and $a \in \mathcal{A}$ and $t \geq 0$, where $\boldsymbol{s}$ and $\boldsymbol{s}'$ denote the current state and the next state, respectively.

As the time homogeneity is assumed in infinite-stage(3. setting, time step subscripts maybe dropped for simplicity.

Values of dynamic treatment regimes

A dynamic treatment regime is equivalent to a *policy* in reinforcement learning vocabulary, which is mostly to be considered deterministic. Thus, a dynamic treatment regime, $\pi : \boldsymbol{\mathcal{S}} \to \mathcal{A}$, is defined as a function which maps the support of the state variable to the set of the possible treatment assignments. As time homogeneity is assumed, we consider only stationary deterministic regimes, where this mapping function $\pi(\boldsymbol{s})$ does not change over time. Hence, a patient with state $\boldsymbol{S}_t = \boldsymbol{s}$ at time point t will be assigned with treatment $A_t = \pi(\boldsymbol{s})$ for all t. The value function $\boldsymbol{V}^{\pi}(\boldsymbol{s})$ of a state under a certain policy π is defined as the expected total discounted rewards when the process begins in state $\boldsymbol{s}$ and all decisions are made according to policy π. Mathematically, $\boldsymbol{V}^{\pi}(\boldsymbol{s}) = \mathbb{E}^{\pi}_{\boldsymbol{s}} \sum_{t=0}^{\infty} \gamma^t \boldsymbol{r}(\boldsymbol{s}_t, a_t, \boldsymbol{s}_{t+1})$, where $\mathbb{E}^{\pi}_{\boldsymbol{s}}$ is the expectation when the initial state is $\boldsymbol{s}$ and a policy π is followed. The value function can also be defined recursively via the bellman equation, $\boldsymbol{V}^{\pi}(\boldsymbol{s}) = \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}' \mid \boldsymbol{s}, \pi(\boldsymbol{s})) \left(\boldsymbol{R}(\boldsymbol{s}, \pi(\boldsymbol{s}), \boldsymbol{s}') + \gamma \boldsymbol{V}^{\pi}(\boldsymbol{s}') \right)$. Here, $\boldsymbol{V}^{\pi}(\boldsymbol{s}) \in \mathbb{R}^J$ has the same dimensionality as the reward vector $\boldsymbol{R}$, as we are considering multiple reward functions instead of a scalar reward function. Moreover, the state-action value function under policy π, $\boldsymbol{Q}^{\pi}(\boldsymbol{s}, a) = \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p\left(\boldsymbol{s}' \mid \boldsymbol{s}, a\right) \left(\boldsymbol{R}\left(\boldsymbol{s}, a, \boldsymbol{s}'\right) + \gamma \, \boldsymbol{Q}^{\pi}(\boldsymbol{s}', \pi(\boldsymbol{s}')) \right)$, is defined similar but the first step takes action a and $\boldsymbol{Q}^{\pi}(\boldsymbol{s}, a) \in \mathbb{R}^J$. In clinical cases, the transition model $\mathbb{P}$ is unknown, optimal regimes must be learn from observed dataset. In infinite horizon setting, as time steps are dropped, we break n observed trajectories into 4-tuple of $(\boldsymbol{s}, a, \boldsymbol{r}, \boldsymbol{s}')$ for estimating value functions. The counter-factual assumptions *(A1-A3)* rule out the confounding phenomena and guarantee the identifiability of the average causal effect of a regime.

Define infinite-stage constrained optimal dynamic treatment regimes

Our strategy to cope with multiple competing outcomes is constrained optimization. We optimize the primary outcome of interest, subject to the constraints on the secondary outcomes, over the space of all the possible regimes under consideration, Π. Here, the average

of value functions is referred as competing outcomes, denoted as $\boldsymbol{V}(\pi) = \mathbb{E}\boldsymbol{V}^{\pi}(\boldsymbol{s})$. The space of regimes under consideration may be crafted by experts with domain knowledge via policy function approximation. As We have $\boldsymbol{V}^{\pi}(\boldsymbol{s}) = \Big(V_1(\pi), V_2^{\pi}(\boldsymbol{s}), \cdots, V_q^{\pi}(\boldsymbol{s})\Big)^{\intercal}$, Let $V_1(\pi) = \mathbb{E}V_1(\pi)$ be the primary outcome of interest, and $V_j(\pi) = \mathbb{E}V_j^{\pi}(\boldsymbol{s})$, $j = 2, 3, \cdots, J$ be the secondary outcomes. Mathematically,

$$\begin{aligned} &\max_{\pi \in \Pi} \; V_1(\pi), \\ &\text{subject to } V_j(\pi) \leq \nu_{j-1}, \end{aligned} \tag{3.1}$$

where $j = 2, \cdots, J$. The constraint upper-bounds $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_{J-1})^{\intercal}$ can be specified based on patient preference and/or expert domain knowledge. Therefore, we define an infinite-stage constrained optimal regime as $\pi^*_{\boldsymbol{\nu}} = \text{argmax}_{\pi \in \Pi} \, V_1(\pi)$, subject to $V_j(\pi) - \nu_{j-1} \leq 0$, where $j = 2, \cdots, J$. Denote the feasible policy space, which is the set of all policy satisfying the constraints, as $\mathcal{F}(\Pi)$. For all $\pi \in \mathcal{F}(\Pi)$ and all $j = 2, \cdots, J$, $V_j(\pi) \leq \nu_{j-1}$. Then, an infinite-stage constrained optimal regime can also be written as $\pi^*_{\boldsymbol{\nu}} = \text{argmax}_{\pi \in \mathcal{F}(\Pi)} V_1(\pi)$. To search over a feasible policy space with manageable computation complexity, we use policy function approximation such that $\pi(\boldsymbol{s}) = \pi(\boldsymbol{s}; \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathbb{R}^q$ is the indexing parameter for policies. Hence, we use $V_j(\pi)$ and $V_j(\boldsymbol{\theta})$ interchangeably, for all j. The search space is reduced from the set of all feasible policies to the feasible space of the indexing parameter $\boldsymbol{\theta}$, denoted as $\mathcal{F}(\boldsymbol{\Theta}) = \{\boldsymbol{\theta} \in \mathbb{R}^q : V_j(\boldsymbol{\theta}) \leq \nu_{j-1}, j = 2, \cdots, J\}$.

To carry out policy search, we need to solve the constrained optimization problem (3.1). This is done using interior-point methods, which are constrained nonlinear optimization methods for finding local optimums, implemented in Matlab fmincon [6, 47]. We also need a method to estimate the value functions using observed dataset. This is done by least-square policy evaluation (LSQ) , a part of the least-squares policy iteration (LSPI) algorithm. [21, 22].

3.2.2 Interior point method for constrained optimization

To solve our constrained optimization problem (3.1) above, interior point algorithm is used. As the optimization softwares often implemented as minimization instead of maximization, we denote $v_1(\boldsymbol{\theta}) = -V_1(\boldsymbol{\theta})$ and $v_j(\boldsymbol{\theta}) = V_j(\boldsymbol{\theta}) - \nu_{j-1}$, for $j = 2, \cdots, J$. Hence,

problem (3.1) notation is simplified as

$$\begin{aligned} &\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \; v_1(\boldsymbol{\theta}) \\ &\text{subject to } v_j(\boldsymbol{\theta}) \leq 0, \end{aligned} \tag{3.2}$$

where $j = 2, \cdots, J$. The interior point method solves a following sequence of approximate minimization problem (2), where ρ is always positive and approaches to zero in the limit. For each $\rho > 0$, the approximate problem is

$$\min_{\boldsymbol{\theta},\boldsymbol{z}} \; f_\rho(\boldsymbol{\theta}, \boldsymbol{z}) = \min \; v_1(\boldsymbol{\theta}) - \rho \sum_{j=2}^{J} \ln(z_j), \text{ subject to } v_j(\boldsymbol{\theta}) + z_j = 0, \tag{3.3}$$

where $j = 2, \cdots, J$. There are as many slack variables z_j as there are inequality constraints ν_j. The z_j are restricted to be positive to keep $\ln(z_j)$ bounded. As ρ decreases to zero, the minimum of f_ρ should approach the minimum of ν_1. The added logarithmic term is called a barrier function [6, 47].

3.2.3 Least-squares policy evaluation

Least-squares policy evaluation, is adopted to approximate the state-action value function of a fixed regime/policy. As it is the state-action value function being approximated, instead of the state value function, changing policy is allowed without a model for the underlying dynamics. The exact $\boldsymbol{Q}^\pi$ values for all state-action pairs can be found by solving the linear system of the Bellman equations,

$$\boldsymbol{Q}^\pi(\boldsymbol{s}, a) = \sum_{\boldsymbol{s}'\in\mathcal{S}} p(\boldsymbol{s}'|\boldsymbol{s}, a) \left\{ \boldsymbol{R}(\boldsymbol{s}, a, \boldsymbol{s}') + \gamma \sum_{a'\in\mathcal{A}} \pi(a' \mid \boldsymbol{s}')\, \boldsymbol{Q}^\pi(\boldsymbol{s}', a') \right\},$$

for any $\boldsymbol{s} \in \mathcal{S}$ and $a' \in \mathcal{A}$. Thus, the state-action value function $\boldsymbol{Q}^\pi$ is considered the fixed point of the Bellman operator: $\boldsymbol{Q}^\pi = T^\pi \boldsymbol{Q}^\pi$, where the Bellman operator defined as $T^\pi \boldsymbol{Q}^\pi = \sum_{\boldsymbol{s}'\in\mathcal{S}} p(\boldsymbol{s}'|\boldsymbol{s}, a) \left(\boldsymbol{R}(\boldsymbol{s}, a, \boldsymbol{s}') + \gamma \sum_{a'\in\mathcal{A}} \pi(a' \mid \boldsymbol{s}')\, \boldsymbol{Q}^\pi(\boldsymbol{s}', a') \right)$.

Linear approximation is a common way for estimating value functions, so that each component of the vector $\boldsymbol{Q}^\pi(\boldsymbol{s}, a; w)$ are approximated by a linear parametric combination of K basis functions (features). As $\boldsymbol{Q}(\boldsymbol{s}, a) = (Q_1(\boldsymbol{s}, a), Q_2(\boldsymbol{s}, a), \cdots, Q_J(\boldsymbol{s}, a))^\intercal$, the ap-

proximation for each component is $Q_j^\pi(\boldsymbol{s}, a; w) = \sum_{k=1}^K \phi_{j,k}(\boldsymbol{s}, a) w_{j,k} = \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}, a)\boldsymbol{w}_j$, where $\boldsymbol{w}_j = (w_{j,1}, \cdots, w_{j,K})^\intercal$ are the parameters to estimate. Moreover, the basis functions $\boldsymbol{\phi}_j(\boldsymbol{s}, a) = (\phi_{j,1}(\boldsymbol{s}, a), \cdots, \phi_{j,K}(\boldsymbol{s}, a))^\intercal$ are arbitrary and fixed, which are often non-linear functions of $\boldsymbol{s}$ and a . It is also required that the basis functions $\phi_{j,k}$ are linearly independent to ensure that there are no redundant parameters and that the matrices involved in the computations are full rank.

Substituting each component of the Q function vector with the linear approximator, we get, for $j = 1, \cdots, J$,

$$\boldsymbol{\phi}_j^\intercal(\boldsymbol{s}, a)\boldsymbol{w}_j = \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}, a) R_j(\boldsymbol{s}, a, \boldsymbol{s}') + \gamma \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}, a) \sum_{a' \in \mathcal{A}} \pi(a' \mid \boldsymbol{s}') \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}', a') \boldsymbol{w}_j.$$

This fixed point equation then can be rearranged as

$$\left\{ \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}, a) - \gamma \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}, a) \sum_{a' \in \mathcal{A}} \pi(a' \mid \boldsymbol{s}') \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}', a') \right\} \boldsymbol{w}_j = \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}, a) R_j(\boldsymbol{s}, a, \boldsymbol{s}'). \tag{3.4}$$

Given a sample set of 4-tuples $\boldsymbol{\mathcal{D}} = \{\boldsymbol{s}^i, a^i, \boldsymbol{s}^{i\prime}, \boldsymbol{r}^i\}_{i=1}^N$, the equation (3.4) above becomes a over-constrained/overdetermined linear system over the parameter vector $\boldsymbol{w}_j$. The linear system can be written as

$$\boldsymbol{B}_j \boldsymbol{w}_j = \boldsymbol{b}_j,$$

where $\boldsymbol{B}_j = \boldsymbol{\Phi}_j^\intercal(\boldsymbol{\Phi}_j - \gamma \boldsymbol{P}^\pi \boldsymbol{\Phi}_j)$ and $\boldsymbol{b}_j = \boldsymbol{\Phi}_j^\intercal \boldsymbol{R}_j$. Moreover,

$$\boldsymbol{\Phi}_j = \begin{pmatrix} \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^1, a^1) \\ \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^2, a^2) \\ \cdots \\ \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^n, a^n) \end{pmatrix}, \; \boldsymbol{P}^\pi \boldsymbol{\Phi}_j = \begin{pmatrix} \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}^1, a^1) \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}', \pi(\boldsymbol{s}')) \\ \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}^2, a^2) \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}', \pi(\boldsymbol{s}')) \\ \cdots \\ \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}^n, a^n) \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}', \pi(\boldsymbol{s}')) \end{pmatrix},$$

$$\text{and } \boldsymbol{R}_j = \begin{pmatrix} \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}^1, a^1) R_j(\boldsymbol{s}^1, a^1, \boldsymbol{s}') \\ \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}^2, a^2) R_j(\boldsymbol{s}^2, a^2, \boldsymbol{s}') \\ \cdots \\ \sum_{\boldsymbol{s}' \in \boldsymbol{\mathcal{S}}} p(\boldsymbol{s}'|\boldsymbol{s}^n, a^n) R_j(\boldsymbol{s}^n, a^n, \boldsymbol{s}') \end{pmatrix},$$

where $\boldsymbol{\phi}_j^\intercal(\boldsymbol{s}', \pi(\boldsymbol{s}')) = \sum_{a' \in \mathcal{A}} \pi(a' \mid \boldsymbol{s}') \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}', a')$. Since the transition probability function $p(\boldsymbol{s}' \mid \boldsymbol{s}, a)$ and reward functions $R_j(\boldsymbol{s}, a, \boldsymbol{s}')$ may be unknown, we can construct approximators for $\boldsymbol{B}$ and $\boldsymbol{b}$ using samples. More precisely, we have approximated versions of $\boldsymbol{\Phi}_j$, $\boldsymbol{P}^\pi \boldsymbol{\Phi}_j$ and $\boldsymbol{R}_j$ based on the sample set as follows:

$$\widehat{\boldsymbol{\Phi}}_j = \begin{pmatrix} \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^1, a^1) \\ \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^2, a^2) \\ \cdots \\ \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^n, a^n) \end{pmatrix}, \widehat{\boldsymbol{P^\pi \Phi}}_j = \begin{pmatrix} \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^{1\prime}, \pi(\boldsymbol{s}^{1\prime})) \\ \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^{2\prime}, \pi(\boldsymbol{s}^{2\prime})) \\ \cdots \\ \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^{n\prime}, \pi(\boldsymbol{s}^{n\prime})) \end{pmatrix}, \text{ and } \widehat{\boldsymbol{R}}_j = \begin{pmatrix} r_1 \\ r_2 \\ \cdots \\ r_n \end{pmatrix}.$$

Given $\widehat{\boldsymbol{\Phi}}_j$, $\widehat{\boldsymbol{P^\pi \Phi}}_j$, and $\widehat{\boldsymbol{R}}_j$, $\boldsymbol{B}_j$ and $\boldsymbol{b}_j$ can be approximated as $\widehat{\boldsymbol{B}}_j = n^{-1} \widehat{\boldsymbol{\Phi}}_j^\intercal (\widehat{\boldsymbol{\Phi}}_j - \gamma \widehat{\boldsymbol{P^\pi \Phi}}_j) = n^{-1} \sum_{i=1}^n \boldsymbol{\phi}_j(\boldsymbol{s}^i, a^i) \left(\boldsymbol{\phi}_j(\boldsymbol{s}^i, a^i) - \gamma \boldsymbol{\phi}_j \left(\boldsymbol{s}^{i\prime}, \pi \left(\boldsymbol{s}^{i\prime} \right) \right) \right)$ and $\widehat{\boldsymbol{b}}_j = n^{-1} \widehat{\boldsymbol{\Phi}}_j^\intercal \widehat{\boldsymbol{R}}_j = n^{-1} \sum_{i=1}^n \boldsymbol{\phi}_j(\boldsymbol{s}^i, a^i) r_j^i$. It is shown in the least-squares policy iteration paper that $\lim_{n \to \infty} \widehat{\boldsymbol{B}}_j = \boldsymbol{B}_j$ and $\lim_{n \to \infty} \widehat{\boldsymbol{b}}_j = \boldsymbol{b}_j$, if the samples are uniformly distributed over the state space. Moreover, the Markov property ensures that the solution $\widehat{\boldsymbol{w}}^\pi = \widehat{\boldsymbol{B}}^{-1} \widehat{\boldsymbol{b}}$ will converge to the true solution $\boldsymbol{w}^\pi$ for sufficiently large n whenever $\boldsymbol{w}^\pi$ exists [21, 22].The least-squares policy evaluation algorithm is listed in Algorithm 2 below.

Equipped with least-squares policy evaluation, we can hence calculate the values of any arbitrary regime/policy π. For $j = 1, \cdots, J$, $V(\pi)$ is estimated by $n^{-1} \sum_{i=1}^n \widehat{Q}_j^\pi(\boldsymbol{s}^i, \pi(\boldsymbol{s}^i)) = n^{-1} \sum_{i=1}^n \boldsymbol{\phi}_j^\intercal(\boldsymbol{s}^i, \pi(\boldsymbol{s}^i)) \widehat{\boldsymbol{w}}_j$.

Algorithm

Putting together in the following box, our algorithm uses interior point method for policy search in terms policy indexing parameters $\boldsymbol{\theta}$, and least-squares policy evaluation for policy evaluation.

Algorithm 1: Constrained optimal regime with least-squares policy evaluation [21, 22] for policy evaluation and interior-point method [6, 47] for policy search.

Input : A sample set $\boldsymbol{D}$ of 4 tuples $(\boldsymbol{s}', a, \boldsymbol{s}, \boldsymbol{r})$

Output: Estimated constrained optimal regime $\widehat{\pi}_{\boldsymbol{\nu}}$ indexed by $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}$

$\pi \leftarrow$ random initialization

until converge

$\widehat{\boldsymbol{Q}}^{\pi}(\boldsymbol{s}, a) \leftarrow$ least-squares policy evaluation $(\boldsymbol{D}, \pi)$

$\widehat{\boldsymbol{V}}(\pi) \leftarrow \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{Q}}^{\pi}(\boldsymbol{s}^i, \pi(\boldsymbol{s}^i))$

$\pi \leftarrow \underset{\pi \in \mathcal{F}(\Pi)}{\operatorname{argmax}} \boldsymbol{V}(\pi)$

end

$\widehat{\pi}_{\boldsymbol{\nu}} \leftarrow \pi$

return $\widehat{\pi}_{\boldsymbol{\nu}}$

Algorithm 2: Least-squares policy evaluation (LSQ). [21, 22]

Input : A sample set $\boldsymbol{D}$ of 4 tuples $(\boldsymbol{s}', a, \boldsymbol{s}, \boldsymbol{r})$

k: Number of basis functions

ϕ: Basis functions

γ: Discount factor

π: policy whose value function is sought

Output: Weights $\widehat{\boldsymbol{w}}^{\pi}$

$\widehat{\boldsymbol{B}} \leftarrow 0$ // $(k \times k)$ matrix

$\widehat{\boldsymbol{b}} \leftarrow 0$ // $(k \times 1)$ vector

for each $(\boldsymbol{s}, a, \boldsymbol{r}, \boldsymbol{s}') \in \boldsymbol{D}$

$\widehat{\boldsymbol{B}} \leftarrow \widehat{\boldsymbol{B}} + \phi(\boldsymbol{s}, a)\left(\phi(\boldsymbol{s}, a) - \gamma\phi(\boldsymbol{s}', \pi(\boldsymbol{s}'))\right)^{\intercal}$

$\widehat{\boldsymbol{b}} \leftarrow \widehat{\boldsymbol{b}} + \phi(\boldsymbol{s}, a)\boldsymbol{r}$

$\widehat{\boldsymbol{w}}^{\pi} \leftarrow \widehat{\boldsymbol{B}}^{-1}\widehat{b}$

return $\widehat{\boldsymbol{w}}^{\pi}$

3.3 Simulation

3.3.1 Chemotherapy mathematical model

The chemotherapy mathematical model, a system of ordinary differential equations (ODE), proposed by Zhao et al [53]. is modified and used to generate a hypothetical clinical trial data. Their model reflects the capability of the drug to suppress tumor growth, as well as its negative impact on patient wellness due to the toxicity of chemotherapy. The dose assignment is discretized to $L = 5$ levels, $\mathcal{A} = \{0.00, 0.25, 0.50, 0.75, 1.00\}$. Two state variables W_t and M_t are considered. W_t denotes the patient negative wellness (toxcity) at time point t. M_t denotes the tumor size observed at time point t. A_t denotes chemotherapy agent dose at time point t. The ODE system is modeled as [53]

$$\dot{W}_t = b_1 \max(M_t, M_0) + c_1(A_t - d_1),$$

$$\dot{M}_t = \big(b_2 \max(W_t, W_0) - c_2(A_t - d_2)\big) \times \mathbb{I}(M_t > 0),$$

where decision points are $t = 0, 1, \cdots, T-1$. Moreover, $\dot{W}_t$ and $\dot{M}_t$ are the transition functions. The indicator function term $\mathbb{I}(M_t > 0)$ means tumor size is absorbed at 0, the patient has been cured and no future tumor recurrence considered. These changing rate yields a piece-wise linear model over time. Constants value are set as $b_1 = 0.1, b_2 = 0.15, c_1 = 1.2, c_2 = 1.2, d_1 = 0.5$ and $d_2 = 0.5$. The initial states are draw as $M_0 \sim \text{Uniform}(0, 2)$ and $W_0 \sim \text{Uniform}(0, 2)$. The initial dose level assignment is draw as $A_0 \sim$ Discrete Uniform$(0.25, 0.50, 0.75, 1.00)$. The state variables for the next time point can be obtained via $W_{t+1} = \max(W_t + \dot{W}_t, 0)$, $M_{t+1} = \max(M_t + \dot{M}_t, 0)$. The dose level assignment is drawn as $A_t \sim$ Discrete Uniform$(0.00, 0.25, 0.50, 0.75, 1.00)$, for $t = 1, \cdots, T-1$.

The survival status of the patient, denoted by F_t, is also modeled. We assume everyone is alive at the initial decision point $t = 0$, that is $p_0 = 0$ and $F_0 = 0$. Death events occur during time interval $(t-1, t]$, $t = 1, 2, \cdots, 6$, and are recorded at the end of each interval as variable F_t, $t = 1, 2, \cdots, 6$. Assume that survival status depends on both toxicity and tumor size. For each time interval $(t-1, t]$, define the hazard function as $\lambda(t)$, where $\log \lambda(t) = \mu_0 + \mu_1 W_t + \mu_2 M_t$, $\mu_1 = \mu_2 = 1$ and $\mu_0 = -8.5$. This again is a piece-wise linear approximation with $\lambda(t) = \exp(\mu_0 + \mu_1 W_t + \mu_2 M_t)$. Then, the

cumulative hazard function during time interval $(t-1,t]$ is $\Delta\Lambda(t) = \sum_{t-1}^{t} \lambda(s)\, ds = \sum_{t-1}^{t} \exp\left(\mu_0 + \mu_1 W_t + \mu_2 M_t\right)\, ds = \exp\left(\mu_0 + \mu_1 W_t + \mu_2 M_t\right)$. The survival function is $\Delta F(t) = \exp\big(-\Delta\Lambda(t)\big) = \exp\big(-\exp\left(\mu_0 + \mu_1 W_t + \mu_2 M_t\right)\big)$. The random event of death during time interval $(t-1,t]$ is drawn as $F_t \sim \text{Bernoulli}(p_t)$, where $p_t = 1 - \exp\big(-\exp\left(\mu_0 + \mu_1 W_t + \mu_2 M_t\right)\big)$. If $F_{t-1} = 1$, then $F_t = 1$. Also, as long as death occurred, all the other state variables at the following decision points are all set to null.

The reward functions here is a bivariate vector, consisting of positive reward and negative reward, denoted as $\boldsymbol{R}_t = (R_t^+, R_t^-)^\intercal$. The positive reward function is used to assess tumor size reduction, while the negative reward to assess the increase of patient negative wellness (toxicity). Specifically, the reward functions are defined as follow.

$$\begin{aligned} R^+(:,t) = &- 15 \times \mathbb{I}(F_{t+1} = 1) \\ &+ 5 \times \mathbb{I}(F_{t+1} \neq 1, M_{t+1} = 0) \\ &+ 5 \times |M_{t+1} - M_t| \times \mathbb{I}(F_{t+1} \neq 1, M_{t+1} - M_t <= 0) \\ &+ 5 \times |M_{t+1} - M_t| \times \mathbb{I}(F_{t+1} \neq 1, M_{t+1} - M_t \leq -0.5), \end{aligned} \tag{3.5}$$

$$\begin{aligned} R_t^- = &5 \times |W_{t+1} - W_t| \times \mathbb{I}(W_{t+1} - W_t) \geq -0.5) + \\ &5 \times |W_{t+1} - W_t|\, \mathbb{I}(W_{t+1} - W_t \geq 0.5). \end{aligned} \tag{3.6}$$

To sum up, the trajectories / training data generated according to the ODE model, where with $N = 1000$ and $T = 6$, are as follow

$$\boldsymbol{S}_0 \xrightarrow{A_0} \underset{\curvearrowright \boldsymbol{R}_0}{\boldsymbol{S}_1} \xrightarrow{A_1} \underset{\curvearrowright \boldsymbol{R}_1}{\boldsymbol{S}_2} \xrightarrow{A_2} \underset{\curvearrowright \boldsymbol{R}_2}{\boldsymbol{S}_3} \xrightarrow{A_3} \underset{\curvearrowright \boldsymbol{R}_3}{\boldsymbol{S}_4} \xrightarrow{A_4} \underset{\curvearrowright \boldsymbol{R}_4}{\boldsymbol{S}_5} \xrightarrow{A_5} \underset{\curvearrowright \boldsymbol{R}_5}{\boldsymbol{S}_6},$$

where $\boldsymbol{S}_t = (M_t, W_t, F_t)$, $t = 0, 1, \cdots, 6$. Moreover, $\boldsymbol{R}_t = (R_t^+, R_t^-), t = 0, 1, \cdots, 5$. There are 7 decision points. The last decision point $T = 6$ has only states $\boldsymbol{S}_6 = (M_6, W_6, F_6)$, without following action nor reward. The trajectories is then broken down into 4-tuples of $(\boldsymbol{s}, a, \boldsymbol{s}', \boldsymbol{r})$ with the time stamps dropped.

3.3.2 Function approximation

Q function approximation

To construct linear approximators for Q functions [12], we use $K = 4$ Gaussian radial basis functions and an intercept of one. The Gaussian radial basis function has the form of $\phi(x) = \exp\left(\|x-\mu\|^2/2\sigma^2\right)$, where μ and σ^2 are the parameters to be specified. Denote the Q function for positive rewards as $Q^+(\boldsymbol{s}, a)$, and the one for negative rewards as $Q^-(\boldsymbol{s}, a)$. As the positive reward function is a function of M_t, we only incorporate M_t in the basis functions for estimating $Q^+(\boldsymbol{s}, a)$. Hence, we can rewrite $\widehat{Q}^+(\boldsymbol{s}, a)$ as $\widehat{Q}^+(m, a)$. Specifically, $\widehat{Q}^+(m, a) = \widehat{w}_0^+ + \sum_{k=1}^4 \widehat{w}_k^+ \exp\left(-\|m-\mu_k^+\|^2/2(\sigma_k^+)^2\right)$, where $\widehat{w}_k^+$ are the weights to be estimated via least-squares policy evaluation. μ_k^+'s are set as the 20, 40, 60, 80 percentiles of the states, and σ_k^+'s the average distance of the percentiles to all the sample points. $\widehat{Q}^-(\boldsymbol{s}, a)$ is constructed similarly.

Policy function approximation

For policy function approximation [12], we focus on simple decision rule to reduce the search space. Let $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5)^\intercal$ be the index parameters for a policy. The policy function is defined as $\pi(\boldsymbol{s}; \boldsymbol{\theta}) = 0.00 \times \mathbb{I}(f(\boldsymbol{s}, \boldsymbol{\theta}) < 1) + 0.25 \times \mathbb{I}(1 \leq f(\boldsymbol{s}, \boldsymbol{\theta}) < 2) + 0.50 \times \mathbb{I}(2 \leq f(\boldsymbol{s}, \boldsymbol{\theta}) < 3) + 0.75 \times \mathbb{I}(3 \leq f(\boldsymbol{s}, \boldsymbol{\theta}) < 4) + 1 \times \mathbb{I}(f(\boldsymbol{s}, \boldsymbol{\theta}) > 4)$, where $f(\boldsymbol{s}, \boldsymbol{\theta}) = \theta_0 + \theta_1 m + \theta_2 m^2 + \theta_3 w + \theta_4 w^2 + \theta_5 mw$.

3.3.3 Simulation results

The goal here is to maximize $V^+(\pi)$, subject to $V^-(\pi) \leq \nu$, where ν is the bound on the secondary outcome. We applied our method to the simulated dataset. Table 3.1. shows the values of primary and secondary outcomes of the estimated constrained optimal regimes, along with their standard deviations. Table 3.2. shows the estimated indexing parameters of the estimated regimes, along with their standard deviations. Figure 3.1. shows the values of the primary objective(red) / secondary objective (blue) vs. constraint ν. Figure 3.2-3.6 shows the actions of the estimated regime for each state under different constraint values. As the ν increases, we start to observe more higher dosages being assigned to patients. Higer dosage leads to better treatment effect (more reduced tumor size), but more toxicity on patient's wellness.

Table 3.1: Values of estimated optimal regimes under different constraint bounds.

ν	$\widehat{V}^+$	std^+	$\widehat{V}^-$	std^-
5.49	0.39	0.36	5.42	0.16
6.85	1.35	0.29	6.62	0.43
8.21	3.58	1.06	7.66	0.30
9.57	4.07	0.89	7.77	0.35
10.93	4.97	0.56	9.53	1.42
12.29	5.98	1.10	11.43	0.69
13.65	6.13	0.99	11.90	1.46
15.01	7.08	0.93	14.88	0.42
16.37	7.98	1.31	14.69	2.15
17.73	8.89	0.61	16.84	0.57
19.09	9.15	0.30	16.93	0.71
20.45	9.85	1.16	17.86	2.33
21.81	9.76	1.01	18.18	2.88
23.18	9.76	1.01	18.18	2.88
24.54	11.62	1.38	21.88	0.63
25.89	12.02	1.17	23.45	2.47
27.25	12.04	1.17	23.57	2.69
28.61	12.86	0.54	28.27	1.06
29.98	13.69	0.57	30.25	1.04
31.34	14.53	0.97	31.11	1.10

The constraint bounds are denoted by ν. $\widehat{V}^+$ denotes the primary outcome values of the estimated regimes. $\widehat{V}^-$ denotes the secondary outcome values of the estimated regimes. Standard deviations of those estimated regime values are reported as well.

Table 3.2: The estimated indexing parameters of estimated regimes under different constraint bounds.

ν	$\widehat{\theta}_{\nu,1}$	std_1	$\widehat{\theta}_{\nu,2}$	std_2	$\widehat{\theta}_{\nu,3}$	std_3	$\widehat{\theta}_{\nu,4}$	std_4	$\widehat{\theta}_{\nu,5}$	std_5	$\widehat{\theta}_{\nu,6}$	std_6
5.49	0.36	0.79	-0.38	0.76	0.13	0.48	0.03	0.84	-0.29	0.71	-0.06	0.76
6.85	0.11	1.47	-1.55	1.10	0.29	0.67	0.35	1.44	-1.06	1.19	-0.20	1.41
8.21	-0.00	1.66	-2.02	1.09	0.58	0.80	0.27	1.48	-1.26	1.40	-0.41	1.53
9.57	0.29	1.65	-2.38	1.16	0.58	0.84	0.32	1.71	-1.43	1.37	0.02	1.66
10.93	0.49	1.64	-2.61	1.13	0.67	0.90	0.18	1.72	-1.47	1.57	0.23	1.64
12.29	0.46	1.70	-2.90	1.16	0.71	0.89	0.08	1.90	-1.27	1.58	0.26	1.59
13.65	0.56	1.76	-3.20	1.12	0.70	0.92	0.01	1.70	-1.06	1.66	0.32	1.58
15.01	0.67	1.74	-3.46	1.08	0.65	0.90	-0.02	1.87	-0.83	1.77	0.14	1.55
16.37	0.89	1.81	-3.49	1.10	0.48	0.95	-0.05	1.95	-0.43	1.81	0.16	1.61
17.73	0.88	1.92	-3.54	1.17	0.39	0.99	-0.21	2.03	-0.07	1.70	0.05	1.56
19.09	1.20	2.08	-3.36	1.54	0.47	1.17	-0.41	2.14	0.11	1.56	0.15	1.52
20.45	1.69	2.18	-3.06	1.79	0.72	1.44	-0.84	2.28	0.13	1.52	0.16	1.61
21.81	1.95	2.20	-2.78	2.06	1.05	1.64	-0.88	2.40	0.12	1.48	0.27	1.88
23.18	2.21	2.21	-2.64	2.19	1.30	1.76	-1.09	2.32	0.15	1.47	0.03	1.93
24.54	2.46	2.27	-2.46	2.23	1.46	1.86	-1.32	2.31	0.36	1.43	-0.14	2.13
25.89	2.84	2.14	-2.14	2.32	1.70	2.03	-1.57	2.12	0.61	1.45	-0.19	2.31
27.25	3.25	1.82	-1.95	2.26	1.66	2.14	-1.60	2.05	1.01	1.60	-0.12	2.41
28.61	3.41	1.73	-1.86	2.26	1.58	2.08	-1.10	2.06	1.46	1.65	-0.05	2.53
29.98	3.65	1.42	-1.53	2.27	2.00	2.06	-0.98	2.04	1.99	1.68	-0.07	2.65
31.34	3.89	1.14	-1.07	2.18	2.23	1.91	-0.60	2.15	2.64	1.61	-0.13	2.70

Here, ν denotes the constraint bounds. $\widehat{\theta}_{\nu,j}$ denotes the j-th component of the estimated parameter vector $\widehat{\boldsymbol{\theta}}_{\nu}$ of the estimated regimes. Standard deviations of those estimated regime values are reported as well.

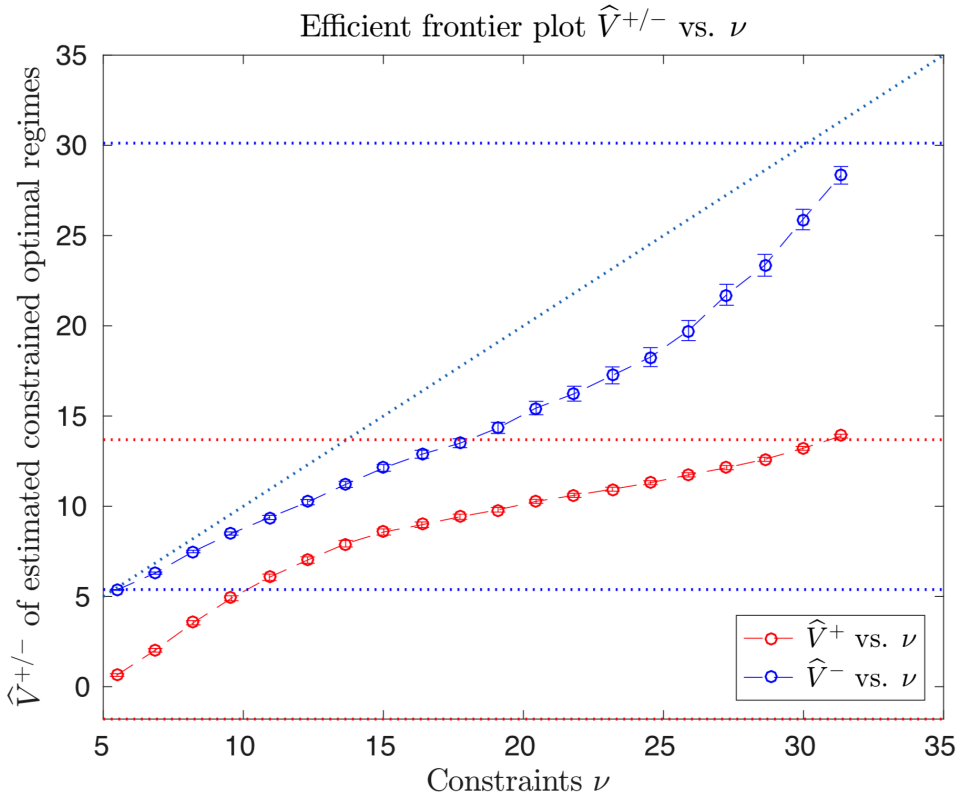

Figure 3.1: Efficient frontier for estimated constrained optimal regimes (infinite-stage).

The red dashed line is for the primary outcome to maximized. The blue dashed line is for the secondary outcome to be constrained. The red dotted lines are the minima and maxima for unconstrained optimization of the primary objective. The blue dotted lines are the minimal and maximal for unconstrained optimization of the secondary objective.

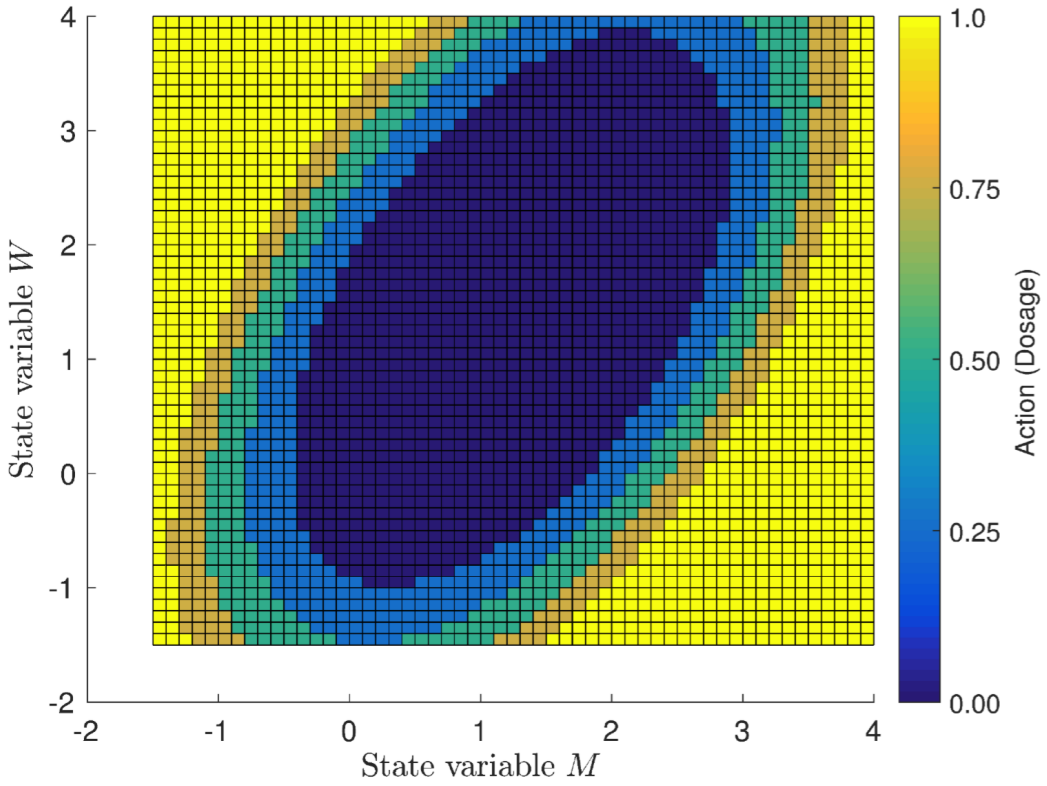

Figure 3.2: Action for each state under constraint $\nu = 10.93$

Yellow represents high dosage treatment assignment. Blue represents low dosage assignment. As the constraint bound gets loose, more higher dosage treatments are assigned to patients.

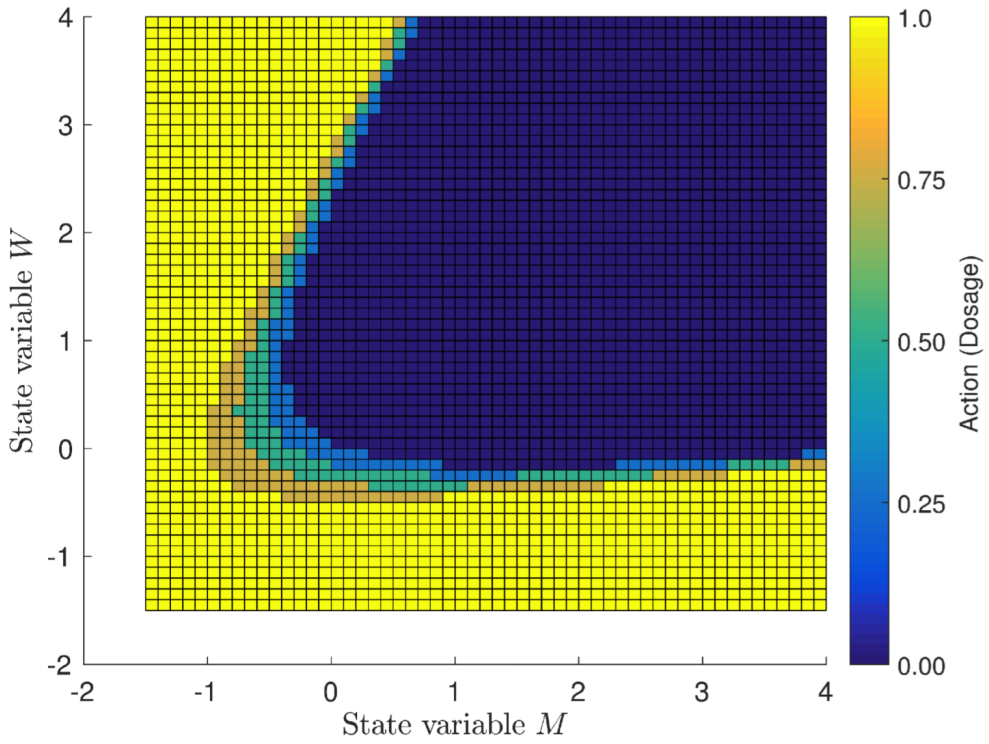

Figure 3.3: Action for each state under constraint $\nu = 17.73$

Yellow represents high dosage treatment assignment. Blue represents low dosage assignment. As the constraint bound gets loose, more higher dosage treatments are assigned to patients.

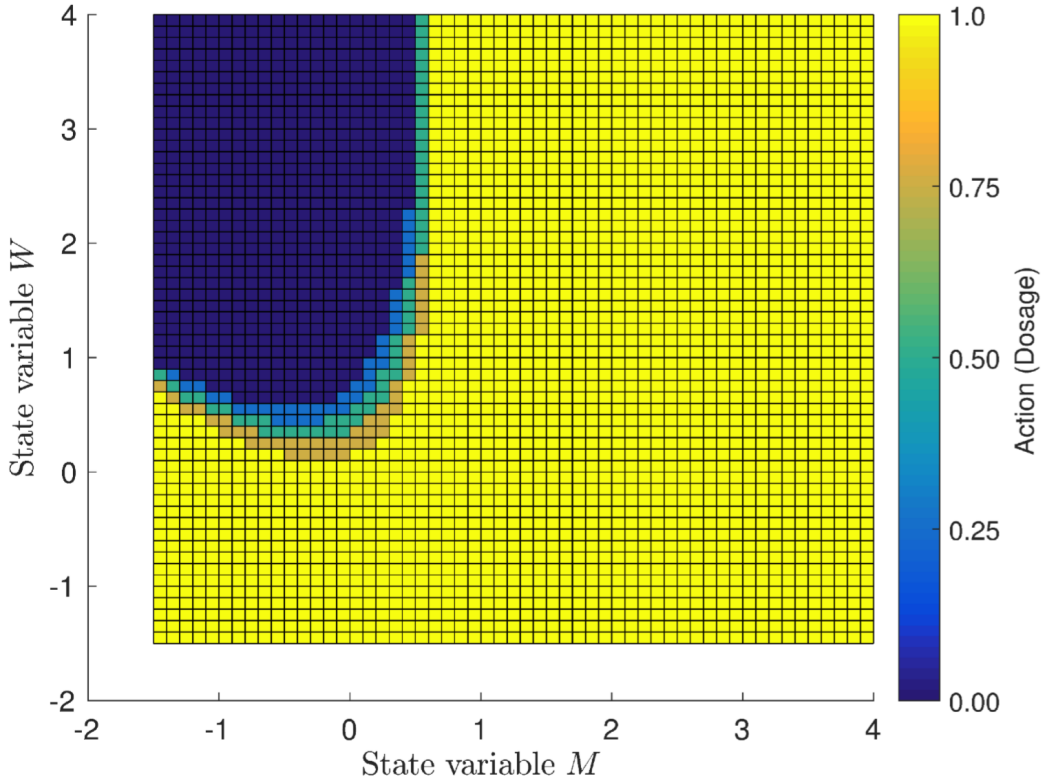

Figure 3.4: Action for each state under constraint $\nu = 24.54$

Yellow represents high dosage treatment assignment. Blue represents low dosage assignment. As the constraint bound gets loose, more higher dosage treatments are assigned to patients.

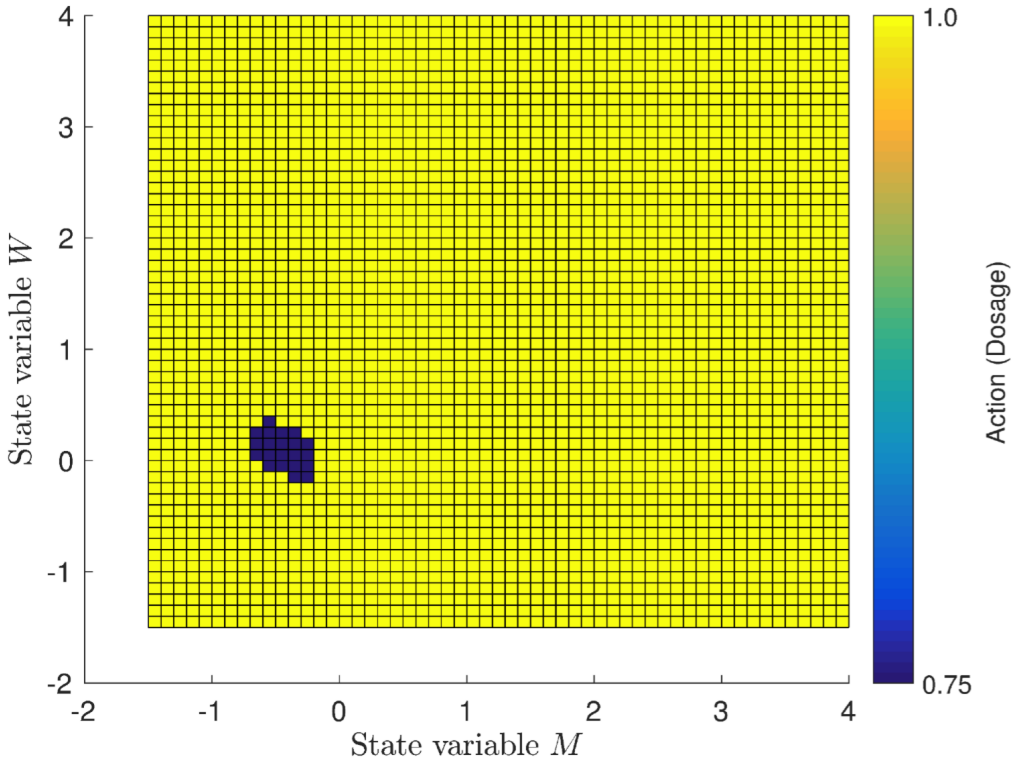

Figure 3.5: Action for each state under constraint $\nu = 31.34$

Yellow represents high dosage treatment assignment. Blue represents low dosage assignment. As the constraint bound gets loose, more higher dosage treatments are assigned to patients.

3.4 Conclusion and Future

We propose a framework for constrained optimal dynamic treatment regimes to handle the trade-off between the primary objective and all other secondary objectives in infinite stage setting. The simulation results based on the chemotherapy ODE system are presented and visualized. This framework offers an intuitive way for clinicians to exam the trade-off and make treatment decisions based on patient's preference. Different from CPO, our method takes into consideration that clinical data is expensive and scarce. Borrowing strength from least-squares policy evaluation, our method is able to learn from data efficiently. Moreover, least-squares policy evaluation is an iterative method for

policy evaluation, and has the advantage of being able to learn both offline and online. Hence, our method can fits in not only the situation where clinical policies needed to be learned after data collect (offline, off-policy batched), but also the situation where online real-time policy learning is needed. Interior-point method is also a well-studies optimization method for constrained estimation. Its theoretical guarantees assure us of good enough optimal solutions. However, it is obvious that the choice of policy function approximation may have impact on the decision. So does the choice of Q function approximation. Clinical domain expertise may required. Alternatively, automated feature learning techniques for function approximation from the machine learning community can be incorporated. More complex dataset maybe collected, such as text, image, speech and so on, considering the recent technology advancement in mobile devices. How to incorporate those complex information to better describe an individual's state of health is challenging. Rigorous theoretical work for our method is also under investigation.

Besides constraints on the expected value of a policy, we can also consider risk constraints, where the probabilities of adverse events occurring are restricted. Although reinforcement learning is a powerful technique to find optimal treatment regimes for clinical practice, designing appropriate reward functions is crucial for serving the desired clinical purpose, but very difficult. Current approach may not scale well in complex clinical situations or preventive healthcare where multiple subgoals may be involved. How to automatic generate rewards and objectives in complex clinical situations can be an interesting direction for investigation.

Nowadays, many aspects of the clinical practice have been transformed by mobile devices, such as smart phones, tablets, wearable sensors etc. It allows clinicians remotely monitor and intervene patients chronic conditions in real-time. It also allows for adaptive preventive interventions for motivating and maintaining healthy behaviors, such as physical exercise, diets, and so on. To better understand adaptive interventions, interdisciplinary collaborations becomes a necessity among clinicians, medical researchers, behavioral scientists, statisticians, and computer scientists.

REFERENCES

[1] Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. Constrained Policy Optimization. *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[2] E. Altman. *Constrained Markov Decision Processes*. Stochastic Modeling Series. Taylor & Francis, 1999.

[3] Garth P. McCormick Anthony V. Fiacco. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*.

[4] Patrick Billingsley. Probability & Measure. page 362, 1995.

[5] D Blatt, SA Murphy, and J Zhu. A-learning for approximate planning. *Ann Arbor*, 1001:48109–2122, 2004.

[6] Richard H. Byrd, Mary E. Hribar, and Jorge Nocedal. An Interior Point Algorithm for Large-Scale Nonlinear Programming. *SIAM Journal on Optimization*, 9(4):877–900, 1999.

[7] Theophilos Cacoullos. Estimation of a Multivariate Density. pages 251–255, 1964.

[8] Bibhas Chakraborty and Erica E.M. Moodie. *Statistical Methods for Dynamic Treatment Regimes*. 2013.

[9] Bibhas Chakraborty and Susan A. Murphy. Dynamic Treatment Regimes. *Annual Review of Statistics and Its Application*, 1(1):447–464, 2014.

[10] Ashkan Ertefaie. Constructing Dynamic Treatment Regimes in Infinite-Horizon Settings. pages 1–39, 2014.

[11] Anders Forsgren, PE Gill, and MH Wright. *Interior Methods for Nonlinear Optimization*, volume 44. 2002.

[12] Alborz Geramifard. A Tutorial on Linear Function Approximators for Dynamic Programming and Reinforcement Learning. *Foundations and Trends® in Machine Learning*, 6(4):375–451, 2013.

[13] Richard D. Gill and James M. Robins. Causal inference for complex longitudinal data: The continuous case. *Annals of Statistics*, 29(6):1785–1811, 2001.

[14] Robin Henderson, Phil Ansell, and Deyadeen Alshibani. Regret-regression for optimal dynamic treatment regimes. *Biometrics*, 66(4):1192–201, December 2010.

[15] Miguel A Hernan and James M Robins. Estimating causal effects from epidemiological data. *Journal of epidemiology and community health*, (7):578–86, July.

[16] David R Hunter. Notes for a graduate-level course in asymptotics for statisticians. page 97, 2014.

[17] Jerzy Splawa-Neyman, D. M. Dabrowska and T. P. Speed. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472, 1990.

[18] Linn KA, Laber EB, and Stefanski LA. Constrained estimation for competing outcomes. *Chapter in Adaptive Treatment Strategies In Practice, ASA-SIAM Statistics and Applied Probability Series, 2015*, 29, 2001.

[19] Eric B. Laber, Daniel J. Lizotte, and Bradley Ferguson. Set-valued dynamic treatment regimes for competing outcomes. *Biometrics*, 70(1):53–61, 2014.

[20] Eric B Laber, Daniel J Lizotte, Min Qian, William E Pelham, and Susan A Murphy. Dynamic treatment regimes : technical challenges and applications. 2014.

[21] Michail G. Lagoudakis and Ronald Parr. Model-Free Least-Squares Policy Iteration. In *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, pages 1547–1554, 2001.

[22] Michail G. Lagoudakis and Ronald Parr. Least-squares policy iteration. *The Journal of Machine Learning Research*, 4:1107–1149, 2003.

[23] H. Lei, I. Nahum-Shani, K. Lynch, D. Oslin, and S.A. Murphy. A "SMART" Design for Building Individualized Treatment Sequences. *Annual Review of Clinical Psychology*, 8(1):21–48, 2012.

[24] Kristin A. Linn, Eric B. Laber, and Leonard A. Stefanski. Constrained estimation for competing outcomes. 2014.

[25] Kristin A Linn, Eric B Laber, and Leonard A Stefanski. Interactive Q-learning for Probabilities and Quantiles. 2014.

[26] Daniel J Lizotte, Michael H. Bowling, and Susan A. Murphy. Efficient reinforcement learning with multiple reward functions for randomized controlled trial analysis. *in Proc. of Int. Conf. on Machine Learning*, pages 695–702, 2010.

[27] Daniel J. Lizotte, Michael H. Bowling, and Susan A. Murphy. Linear Fitted-Q Iteration with Multiple Reward Functions. *Journal of Machine Learning Research*, 13:3253–3295, 2012.

[28] Daniel J. Lizotte and Eric B. Laber. Multi-objective markov decision processes for data-driven decision support. *J. Mach. Learn. Res.*, 17(1):7378–7405, January 2016.

[29] Daniel J. Luckett, Eric B. Laber, Anna R. Kahkoska, David M. Maahs, Elizabeth Mayer-Davis, and Michael R. Kosorok. Estimating Dynamic Treatment Regimes in Mobile Health Using V-learning. 2016.

[30] Harry Markowitz. PORTFOLIO SELECTION. *The Journal of Finance*, 7(1):77–91, March 1952.

[31] Erica Moodie. Dynamic treatment regimes. *Clinical trials (London, England)*, 1(5):471, 2004.

[32] S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, May 2003.

[33] S A Murphy, Y Deng, E B Laber, H R Maei, R S Sutton, and K. Witkiewitz. A Batch, Off-Policy, Actor-Critic Algorithm for Optimizing the Average Reward. *arXiv*, pages 1–18, 2016.

[34] Susan A Murphy. A Generalization Error for Q-Learning. *Journal of machine learning research : JMLR*, 6:1073–1097, July 2005.

[35] Inbal Nahum-Shani, Min Qian, Daniel Almirall, William E. Pelham, Beth Gnagy, Gregory A. Fabiano, James G. Waxmonsky, Jihnhee Yu, and Susan A. Murphy. Q-learning: A data analysis method for constructing adaptive interventions. *Psychological Methods*, 17(4):478–494, 2012.

[36] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, second edition, 2006.

[37] Liliana Orellana, Andrea Rotnitzky, and James M Robins. 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, January 2010.

[38] A. R. Pagan and Aman. Ullah. *Nonparametric econometrics / Adrian Pagan, Aman Ullah.* Themes in modern econometrics. Cambridge University Press,, Cambridge, 1999.

[39] J M Robins, M A Hernán, and B Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

[40] James M. Robins. Causal Inference from Complex Longitudinal Data, 1997.

[41] James M. Robins, Donald Blevins, Grant Ritter, and Michael Wulfsohn. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of aids patients. *Epidemiology*, 3(4):319–336, 1992.

[42] D. B. Rubin. Discussion of Randomized analysis of experimental data: The Fisher randomization test by D. Basu. *Journal of the American Statistical Association*, (75):591–593, 1980.

[43] Donald B. Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[44] Phillip J. Schulte, Anastasios A. Tsiatis, Eric B. Laber, and Marie Davidian. Q- and A-Learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Statistical Science*, 29(4):640–661, 2014.

[45] Bernard W Silverman. Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives. *The Annals of Statistics*, 6(1):177–184, 1978.

[46] Rui Song, Weiwei Wang, Donglin Zeng, and Michael R. Kosorok. Penalized Q-Learning for Dynamic Treatment Regimes. August 2011.

[47] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107(3):391–408, 2006.

[48] Lu Wang, Andrea Rotnitzky, Xihong Lin, Randall E Millikan, and Peter F Thall. Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer. *Journal of the American Statistical Association*, 107(498):493–508, June 2012.

[49] Baqun Zhang, Anastasios A. Tsiatis, Marie Davidian, Min Zhang, and Eric Laber. Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114, 2012.

[50] Baqun Zhang, Anastasios a. Tsiatis, Eric B. Laber, and Marie Davidian. A Robust Method for Estimating Optimal Treatment Regimes. *Biometrics*, 68(4):1010–1018, 2012.

[51] Ying-Qi Zhao, Donglin Zeng, Eric B. Laber, and Michael R. Kosorok. New Statistical Learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015.

[52] Yingqi Zhao, Donglin Zeng, a John Rush, and Michael R Kosorok. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*, 107(449):1106–1118, 2012.

[53] Yufan Zhao, Michael R. Kosorok, and Donglin Zeng. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, 28(26):3294–3315, 2010.

APPENDICES

Appendix A

Supplement materials for Chapter 1

A.1 Conditions for convergence of the penalty-barrier trajectory for mixed constraints

We revisit the conditions under which the penalty-barrier trajectory converging to the solution to the original mixed-constraint problem. The original inequality-equality constrained problem is

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}}\ f(\boldsymbol{x}) \\ &\text{subject to } c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I}, \text{ and } c_j(\boldsymbol{x}) = 0, j \in \mathcal{E}, \end{aligned} \tag{A.1}$$

where $\mathcal{I}$ is the set of the indices for inequality constraints, and $\mathcal{E}$ is the set of the indices for equality constraints. Let $\boldsymbol{x}^*$ denote a solution to the original problem (A.1). A classical strategy to solve this mixed constraint problem is to find an unconstrained minimizer of a composite function that consists of the objective function, the barrier penalty for the inequality constraints, and the quadratic penalty for the equality constraints, i.e., a penalty-barrier function. It is defined as

$$\Phi_{PB}(\boldsymbol{x}, \mu) \triangleq f(\boldsymbol{x}) - \mu \sum_{i\in\mathcal{I}} \log\, c_i(\boldsymbol{x}) + \frac{1}{2\mu} \sum_{j\in\mathcal{E}} c_j^2(\boldsymbol{x}), \tag{A.2}$$

where μ is a sequence of sufficiently small, positive decreasing constants. Let $\boldsymbol{x}(\mu)$ denote an unconstrained minimizer of $\Phi_{PB}(\boldsymbol{x}, \mu)$. The following theorem gives the conditions that

ensure the convergence of the differentiable penalty-barrier trajectory sequence $\{\boldsymbol{x}(\mu)\}$ to the original solution $\boldsymbol{x}^*$.

Theorem A.1.1 (Second-Order Sufficient Conditions for Problem (A.1) [3, 11])**.** *Sufficient conditions that a point $\boldsymbol{x}^*$ be an isolated (uniquely) local minimum of Problem (A.1), where f, $c_i, \forall i \in \mathcal{I}$, and $c_j, \forall j \in \mathcal{E}$ are twice-differentiable functions, are that there exist vectors $\lambda_{\mathcal{I}}^*$ and $\lambda_{\mathcal{E}}^*$ such that $\left(\boldsymbol{x}^*, \lambda_{\mathcal{I}}^*, \lambda_{\mathcal{E}}^*\right)$ satisfies*

1. *$\boldsymbol{x}^*$ is feasible and the LICQ (Linear Independence Constraint Qualification) holds at $\boldsymbol{x}^*$, i.e., the Jacobian matrix of active constraints at $\boldsymbol{x}^*$, $J_{\mathcal{A}}(\boldsymbol{x}^*)$, has full row rank;*

2. *$\boldsymbol{x}^*$ is a KKT point and strict complementarity holds, i.e, the (necessarily unique) multiplier λ^* has the property that $\lambda_i^* > 0$, for all $i \in \mathcal{A}_{\mathcal{I}}(\boldsymbol{x}^*)$, the set of indices of active inequality constraints at $\boldsymbol{x}^*$;*

3. *for all nonzero vectors p satisfying $J_{\mathcal{A}}(\boldsymbol{x}^*)\boldsymbol{p} = 0$, there exists $\omega > 0$ such that $\boldsymbol{p}^{\intercal} H(\boldsymbol{x}^*, \lambda^*)\boldsymbol{p} \geq \omega \|\boldsymbol{p}\|^2$., where $H(\boldsymbol{x}^*, \lambda^*)$ is the hessian of the Lagrangian at $\boldsymbol{x}^*$ and λ^*.*

Theorem A.1.2 (Isolated Trajectory for $\Phi_{PB}(\boldsymbol{x}, \mu)$ Function [3,11])**.** *If (a) the functions f, $c_i, \forall i \in \mathcal{I}$, and c_j, $\forall j \in \mathcal{E}$ are twice differentiable, (b) the gradients ∇c_i, $\forall i \in \mathcal{I}$, and $\nabla c_j, \forall j \in \mathcal{E}$ are linearly independent, (c) strict complementarity holds for $u_i^* c_i(\boldsymbol{x}^*) = 0, \forall i \in \mathcal{I}$, and (d) the sufficient conditions stated above under which $\boldsymbol{x}^*$ be an isolated local constrained minimum of Problem (A.1) are satisfied by $\left(\boldsymbol{x}^*, \lambda_{\mathcal{I}}^*, \lambda_{\mathcal{E}}^*\right)$, then there is a positive neighborhood about $\mu = 0$ for which a unique-isolated differentiable function $\boldsymbol{x}(\mu)$ exists that describes a unique isolated trajectory of local minima of $\Phi_{PB}(\boldsymbol{x}, \mu)$, where $\boldsymbol{x}(\mu) \to \boldsymbol{x}^*$ as $\mu \to 0$.*

Note that $c_i(\boldsymbol{x}), \forall i \in \mathcal{I}$ is embedded in the log operator, $c_i(\boldsymbol{x}_\mu) > 0$ is enforced implicitly.

A.2 Proof of Theorem 1.1.2

Theorem A.2.1. *For any fixed μ, assume*

1. *Point-wise convergence of $\widehat{v}_j(\boldsymbol{\theta})$ in probability:*
   For every $\boldsymbol{\theta} \in \mathcal{F}(\boldsymbol{\Theta})$, we have $\lim_{n\to\infty} Pr\left\{ \mid v_j(\boldsymbol{\theta}) - \widehat{v}_j(\boldsymbol{\theta}) \mid \leq \epsilon_j \right\} = 1$, $\forall \epsilon_j > 0$, where $j = 1, \cdots, J$;

2. *Existence of a strict local minimizers of* $\phi_{\mu}^{PB}(\boldsymbol{\theta})$*:*
   There exists a neighborhood of $\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)$*, denoted* $\mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)\right)$ *such that* $\phi_{\mu}^{PB}\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)\right) < \phi_{\mu}^{PB}(\boldsymbol{\theta})$*, for any* $\boldsymbol{\theta} \in \mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)\right)$*;*

3. *Existence of strict local minimizer* $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ *of* $\widehat{\phi}_{\mu}^{PB}(\boldsymbol{\theta})$ *in the neighborhood* $\mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)\right)$*:*
   $\widehat{\phi}_{\mu}^{PB}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)\right) < \widehat{\phi}_{\mu}^{PB}(\boldsymbol{\theta})$*, for any* $\boldsymbol{\theta} \in \mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)\right)$*, where* $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) \in \mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)\right)$*;*

then

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) \xrightarrow{p} \boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu).$$

Proof. In this part, we simplify the notations locally just for this proof. Suppose there exists a local minimum $\boldsymbol{\theta}^{*} = \boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)$. Let its estimator be $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu)$ and its neighborhood $\mathcal{N}^{*} = \mathcal{N}\left(\boldsymbol{\theta}_{\boldsymbol{\nu}}^{*}(\mu)\right)$. Also, let $\phi(\boldsymbol{\theta}) = \phi_{\mu}^{PB}(\boldsymbol{\theta})$ and $\widehat{\phi}(\boldsymbol{\theta}) = \widehat{\phi}_{\mu}^{PB}(\boldsymbol{\theta})$. By assumption 1, $|\phi(\boldsymbol{\theta}^{*}) - \widehat{\phi}(\boldsymbol{\theta}^{*})| = o_p(1)$, as $n \to \infty$; $|\phi(\widehat{\boldsymbol{\theta}}) - \widehat{\phi}(\widehat{\boldsymbol{\theta}})| = o_p(1)$, as $n \to \infty$. Both $\boldsymbol{\theta}^{*} \in \mathcal{N}^{*}$ and $\widehat{\boldsymbol{\theta}} \in \mathcal{N}^{*}$.

$$\begin{aligned}
\phi(\boldsymbol{\theta}^{*}) &= \widehat{\phi}(\widehat{\boldsymbol{\theta}}) + \left\{\phi(\boldsymbol{\theta}^{*}) - \widehat{\phi}(\widehat{\boldsymbol{\theta}})\right\} \\
&> \widehat{\phi}(\widehat{\boldsymbol{\theta}}) + \left\{\phi(\boldsymbol{\theta}^{*}) - \widehat{\phi}(\boldsymbol{\theta}^{*})\right\} \text{ (by assumption 3)} \\
&\geq \widehat{\phi}(\widehat{\boldsymbol{\theta}}) - |\phi(\boldsymbol{\theta}^{*}) - \widehat{\phi}(\boldsymbol{\theta}^{*})| \\
&= \phi(\widehat{\boldsymbol{\theta}}) + \left\{\widehat{\phi}(\widehat{\boldsymbol{\theta}}) - \phi(\widehat{\boldsymbol{\theta}})\right\} - |\phi(\boldsymbol{\theta}^{*}) - \widehat{\phi}(\boldsymbol{\theta}^{*})| \\
&\geq \phi(\widehat{\boldsymbol{\theta}}) - |\widehat{\phi}(\widehat{\boldsymbol{\theta}}) - \phi(\widehat{\boldsymbol{\theta}})| - |\phi(\boldsymbol{\theta}^{*}) - \widehat{\phi}(\boldsymbol{\theta}^{*})| \\
&\geq \phi(\widehat{\boldsymbol{\theta}}) + o_p(1) \text{ (implied by assumption 1)}
\end{aligned}$$

Suppose $\widehat{\boldsymbol{\theta}} \not\to \boldsymbol{\theta}^{*}$, and then $\phi(\boldsymbol{\theta}^{*}) > \liminf \phi(\widehat{\boldsymbol{\theta}})$. This is opposed to assumption 2, which claims $\boldsymbol{\theta}^{*}$ to be a strict local minimizer. By contradictory, it is proven that $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^{*}$, as $n \to \infty$. ■

A.3 Consistency of Kernel Density Estimators

As kernel density estimators (KDEs) are used to estimate values of regimes, we review the necessary asymptotic properties of Kernel Density Estimators briefly here.

A.3.1 Consistency of univariate Kernel Density Estimator

We review uniform consistency of Kernel Density Estimators for a univariate distribution $g(x)$ [38, 45]. Consider the kernel estimate $\widehat{g}_n(x)$ of a real univariate density $g(x)$ introduced by Rosenblatt (1956) [38, 45], and defined as

$$\widehat{g}_n(x) = \sum_{i=1}^{n} \frac{1}{nh} k\left(\frac{x - X_i}{h}\right),$$

where $X_1, \cdots, X_n$ are identically independent observations from the distribution $g(x)$; k is a kernel function satisfying suitable conditions given below; $h = h_n$ is the bandwidth which is also a function of sample size n.

Theorem A.3.1 (Uniform consistency of univariate Kernel Density Estimators)**.** *[38,45] If all the following assumptions hold,*

1. *If the kernel density function $k(s)$ satisfies*
   - *(a)* $\int k(s)\, ds = 1$;
   - *(b)* $\int \mid k(s) \mid \, ds < \infty$;
   - *(c)* $|s| \left|k(s)\right| \to 0$, *as* $s \to \infty$;
   - *(d)* $\sup \left|k(s)\right| < \infty$.
2. *The bandwidth h satisfies that $h \to 0$ and $nh^2 \to \infty$, as $n \to \infty$;*
3. *$g(x)$ is uniformly continuous on $\mathbb{R}$;*
4. *The characteristic function $\phi(t)$ of a random variable s with the density $k(s)$, $\psi(t) = \int e^{its} k(s)\, ds$, is absolutely integrable,*

and then we have that $\widehat{g}_n(x)$ is uniformly weak consistent, that is,

$$\underset{n\to\infty}{p\lim} \left[\sup_x \mid \widehat{g}_n(x) - g(x) \mid\right] = 0,$$

where $\underset{n\to\infty}{p\lim}$ *denotes convergence in probability.*

A.3.2 Consistency of multivariate Kernel Density Estimator

The uniform convergence theorem of univariate Kernel Density Estimators above is extended to multivariate case by Cacoullos (1964) [7]. Consider an estimator of a d-dimensional density function $g(\boldsymbol{x})$ of the following form:

$$\widehat{g}_n(\boldsymbol{x}) = \frac{1}{h^d} \sum_{i=1}^{n} \bar{k} \left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h} \right)$$

where $\bar{k}(\boldsymbol{s})$ is a multivariate kernel of choice satisfying suitable conditions given below, and $h = h_n$ is the bandwidth.

Theorem A.3.2 (Uniform consistency of multivariate Kernel Density Estimators)**.** *[7] Assume:*

1. *$\bar{k}(\boldsymbol{s})$ is a Borel scalar function on $\mathbb{R}^d$, where $\boldsymbol{s} := (s_1, \cdots, s_d)$ such that*
   - *(a) $\int \cdots \int \bar{k}(\boldsymbol{s})\, ds_1 \cdots ds_d = 1$;*
   - *(b) $\int \cdots \int |\bar{k}(\boldsymbol{s})|\, ds_1 \cdots ds_d < \infty$;*
   - *(c) $|\boldsymbol{s}|^d |\bar{k}(\boldsymbol{s})| \to 0$, as $\boldsymbol{s} \to \infty$, where $|\boldsymbol{s}|$ is the length of $\boldsymbol{s}$;*
   - *(d) $\sup_{\boldsymbol{s}} |\bar{k}(\boldsymbol{s})| < \infty$.*
2. *$h \to 0$ and $nh^{2d} \to \infty$, as $n \to \infty$;*
3. *$g(\boldsymbol{x})$ is uniformly continuous in $\mathbb{R}^d$;*
4. *The characteristic function of a random vector $\boldsymbol{s}$ with the density of $\bar{k}(\boldsymbol{s})$, $\psi(\boldsymbol{t}) = \int \cdots \int e^{i\boldsymbol{t}^\intercal \boldsymbol{s}} \bar{k}(\boldsymbol{s})\, d\boldsymbol{s}$, is absolutely integrable,*

and then, $\widehat{g}_n(\boldsymbol{x})$ is uniform consistent, that is,

$$\operatorname*{p\,lim}_{n\to\infty} \left[\sup_{\boldsymbol{x}} \left| \widehat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x}) \right| \right] = 0.$$

Usually, we use a product kernel for multivariate distributions. For random vector $\boldsymbol{S} \in \mathbb{R}^d$, $\boldsymbol{S} := (S_1, \cdots, S_d)$,

$$\frac{1}{h^d} \bar{k} \left(\frac{\boldsymbol{s}}{h} \right) = \frac{1}{h^d} \prod_{j=1}^{d} k \left(\frac{s_j}{h} \right),$$

where $k(s)$ is a suitable univariate kernel function. Here, we exposit the bandwidths for each component with the same magnitude, $h_n = h$. which is also inferred by optimal bandwidth choice.

A.4 Estimating the value functions via KDE

We derive the value function estimator using KDE. Recall the j-th value function is modeled as

$$V_j(\boldsymbol{\theta}) = m_{\boldsymbol{\alpha}_j^*} + \iint \operatorname{sgn}(z_1)\, z_2 f_{\boldsymbol{\beta}_j^*}(z_1, z_2; \boldsymbol{\theta})\, dz_1\, dz_2,$$

Note, for any fixed $\boldsymbol{\beta}_j$, $\iint \operatorname{sgn}(z_1) z_2\, f_{\boldsymbol{\beta}_j}(z_1, z_2; \boldsymbol{\theta})\, dz_1\, dz_2 =$
$2\iint z_2\, \mathbb{I}(z_1 \geq 0)\, f_{\boldsymbol{\beta}_j}(z_1, z_2; \boldsymbol{\theta})\, dz_1\, dz_2 - \int z_2\, f_{\boldsymbol{\beta}_j}(z_2)\, dz_2$. To estimate this quantity, we plug in kernel density estimators for $f_{\boldsymbol{\beta}_j}(z_1, z_2; \boldsymbol{\theta})$ and $f_{\boldsymbol{\beta}_j}(z_2)$, and get

$$\begin{aligned}
&\iint \operatorname{sgn}(z_1)\, z_2 \widehat{f}_{\boldsymbol{\beta}_j}(z_1, z_2; \boldsymbol{\theta})\, dz_1\, dz_2 \\
=&2\iint z_2\, \mathbb{I}(z_1 \geq 0)\, \widehat{f}_{\boldsymbol{\beta}_j}(z_1, z_2; \boldsymbol{\theta})\, dz_1\, dz_2 - \int z_2\, \widehat{f}_{\boldsymbol{\beta}_j}(z_2)\, dz_2 \\
=&2\iint z_2\, \mathbb{I}(z_1 \geq 0) \left\{ \frac{1}{nhh} \sum_{i=1}^{n} k\left(\frac{z_1 - Z_1^i}{h}\right) k\left(\frac{z_2 - Z_2^i}{h}\right) \right\} dz_1\, dz_2 - \int z_2 \left\{ \frac{1}{nh2} \sum_{i=1}^{n} k\left(\frac{z_2 - Z_2^i}{h}\right) \right\} dz_2 \\
=&\frac{2}{n} \sum_{i=1}^{n} \boldsymbol{X}_1^{i\intercal} \boldsymbol{\beta}_j \left\{ 1 - K\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right) \right\} - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_1^{i\intercal} \boldsymbol{\beta}_j \\
=&\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_1^{i\intercal} \boldsymbol{\beta}_j \left\{ 1 - 2K\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right) \right\},
\end{aligned} \tag{A.3}$$

where $K(s)$ is the corresponding CDF of the kernel function $k(s)$.The third equality is derived in the following. As we use the Gaussian kernel for $k(s)$, it satisfies the following

1. $\int_{-\infty}^{\infty} k(s)\, ds = 1$;
2. $k(s) > 0$ for all s;
3. $k(-s) = k(s)$ for all s;
4. The first order derivative of the kernel, $k'(s)$, exists and is bounded.

To calculate the first term on the right hand side, let $s = \frac{z_1 - Z_1^i}{h}$ and $t = \frac{z_2 - Z_2^i}{h}$. Then, $z_1 = Z_1^i + sh$ and $z_2 = Z_2^i + th$. Also, $dz_1 = h\,ds$ and $dz_2 = h\,dt$. Then,

$$
\begin{aligned}
&\frac{2}{hh}\iint z_2\,\mathbb{I}\left(z_1 \geq 0\right) k\left(\frac{z_1 - Z_1^i}{h}\right) k\left(\frac{z_2 - Z_2^i}{h}\right)\, dz_1\, dz_2 \\
=&2\iint \left(Z_2^i + th\right)\,\mathbb{I}\left(Z_1^i + sh \geq 0\right) k\left(s\right) k\left(t\right)\, ds\, dt \\
=&2\iint Z_2^i\,\mathbb{I}\left(Z_1^i + sh \geq 0\right) k\left(s\right) k\left(t\right)\, ds\, dt + 2\iint th\,\mathbb{I}\left(Z_1^i + sh \geq 0\right) k\left(s\right) k\left(t\right)\, ds\, dt \\
=&2\int Z_2^i\,\mathbb{I}\left(s \geq -Z_1^i/h\right) k\left(s\right)\, ds + 0 \\
=&2Z_2^i\left\{1 - K\left(-Z_1^i/h\right)\right\} \\
=&2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j\left\{1 - K\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\right\},
\end{aligned}
$$

where $K\left(s\right) = \int k\left(s\right)\, ds + c$. The third equality holds, as $\int k(t)\,dt = 1$ and $\int t\,k(t)\,dt = 0$. The fourth equality holds as $\int \mathbb{I}\left(s \geq -Z_i/h\right) k\left(s\right)\, ds = 1 - \int_{-\infty}^{-Z_1^i/h} k\left(s\right)\, ds = 1 - K\left(-Z_1^i/h\right)$, where $Z_1^i = \boldsymbol{X}^{i\intercal}\boldsymbol{\theta}$ and $Z_2^i = \boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j$.

To calculate the second term on the right hand side, we derive $\frac{1}{h}\int z_2 k(\frac{z_2 - Z_2^i}{h})\, dz_2$ by changing variable similarly. Let $t = \frac{z_2 - Z_2^i}{h}$,and we get $z_2 = Z_2^i + th$ and $dz_2 = h\,dt$. Then,

$$
\frac{1}{h}\int z_2 k\left(\frac{z_2 - Z_2^i}{h}\right)\, dz_2 = \int \left(Z_2^i + th\right) k\left(t\right)\, dt = Z_2^i = \boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j.
$$

Again, the second equality holds as $\int k(t)\,dt = 1$, and $\int t\,k(t)\,dt = 0$. Together, we complete the derivation for (A.3).

A.5 Proof of Lemma 1.1.3

Lemma A.5.1. *Suppose the following conditions hold*

1. $\forall \boldsymbol{a} \in \mathbb{R}^p$, $\exists\, \delta > 0$ *,such that*

   (a) $\mathbb{E}\left|\boldsymbol{a}^\intercal \frac{2\boldsymbol{X}_1^\intercal \boldsymbol{\beta}_j}{h} k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X} - \mu_n\right|^{2+\delta} < \infty$, *where* $\mu_n = \mathbb{E}\left\{\boldsymbol{a}^\intercal \frac{2\boldsymbol{X}_1^\intercal \boldsymbol{\beta}_j}{h} k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\tau}}{h}\right)\boldsymbol{X}\right\}$

(b) $\boldsymbol{a}^\intercal V\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h})\boldsymbol{X}\right\}\boldsymbol{a}^{1+\frac{\delta}{2}}<\infty.$

Then, for any fixed $\boldsymbol{\theta}$ *and* $\boldsymbol{\beta}_j$,

$$\sqrt{n}\left(\nabla\widehat{V}_j\left(\boldsymbol{\theta},\boldsymbol{\beta}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right)\xrightarrow{d}N\left(0,AV\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right),$$

where $j=1,\cdots,J$.

Notation ∇ denotes the first-order derivatives with respect to $\boldsymbol{\theta}$. AV stands for asymptotic variance. Moreover, recall that $\nabla\widehat{V}_j\left(\boldsymbol{\theta},\boldsymbol{\beta}_j\right)=\frac{1}{n}\sum_{i=1}^n\frac{2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}^i$.

Proof. For any $\boldsymbol{a}\in\mathbb{R}^p$, we let $W_{ni}=\boldsymbol{a}^\intercal\frac{2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}_i$. For each value of n, $w_{n1},w_{n2},\cdots,w_{nn}$ are i.i.d, and functions of the sample size n. This is because that $\boldsymbol{X}_i$ are assumed to be i.i.d., and h is a function of sample size n. Then, we have

$$\mu_n:=\mathbb{E}W_{ni}=\mathbb{E}\left\{\boldsymbol{a}^\intercal\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\},$$

and

$$\sigma_n^2:=V(W_{ni})=\boldsymbol{a}^\intercal V\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\boldsymbol{a}.$$

We let $G_{ni}=W_{ni}-\mu_n$, and $T_n=\sum_{i=1}^n G_{ni}$. Also, we let $s_n^2=V(T_n)=\sum_{i=1}^n V(G_{ni})=\sum_{i=1}^n\sigma_n^2=n\sigma_n^2$, where the second equality is because of independence, and the last equality is due to identicalness. Therefore, T_n/s_n has mean 0, and variance 1. If we can show G_{ni} satisfying the Lyapunov condition, then we have

$$\frac{T_n}{s_n}\xrightarrow{d}N(0,1),$$

as $n\to\infty$.
Now, we check the Lyapunov condition, that is, [4, 16]

$$\exists\delta>0,\ \text{such that}\ \frac{1}{s_n^{2+\delta}}\sum_{i=1}^n\mathbb{E}\mid G_{n,i}\mid^{2+\delta}\to 0,\ \text{as}\ n\to 0.$$

We define, for any $\boldsymbol{a}$,

$$C_1 \triangleq \mathbb{E}\left|G_{ni}\right|^{2+\delta} = \mathbb{E}\left|W_{ni}-\mu_n\right|^{2+\delta} = \mathbb{E}\left|\boldsymbol{a}^\intercal \frac{2\boldsymbol{X}_1^\intercal \boldsymbol{\beta}_j}{h} k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}-\mu_n\right|^{2+\delta},$$

and

$$C_2 \triangleq s_n^{2+\delta} = n^{1+\frac{\delta}{2}}\sigma_n^{2+\delta} = n^{1+\frac{\delta}{2}}\left\{\boldsymbol{a}^\intercal V\left[\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right]\boldsymbol{a}\right\}^{1+\frac{\delta}{2}}.$$

Then, we have

$$\begin{aligned}
&\frac{1}{s_n^{2+\delta}}\sum_{i=1}^{n}\mathbb{E}\mid G_{n,i}\mid^{2+\delta}\\
=&\frac{n\mathbb{E}\left|\boldsymbol{a}^\intercal\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h})\boldsymbol{X}-\mu_n\right|^{2+\delta}}{n^{1+\frac{\delta}{2}}\left\{\boldsymbol{a}^\intercal V\left[\frac{2\boldsymbol{X}_1^\intercal\beta}{h}k(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h})\boldsymbol{X}\right]\boldsymbol{a}\right\}^{1+\frac{\delta}{2}}}\\
=&\frac{\mathbb{E}\left|\boldsymbol{a}^\intercal\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h})\boldsymbol{X}-\mu_n\right|^{2+\delta}}{n^{\frac{\delta}{2}}\left\{\boldsymbol{a}^\intercal V\left[\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h})\boldsymbol{X}\right]\boldsymbol{a}\right\}^{1+\frac{\delta}{2}}}\\
=&\frac{C_1}{n^{\frac{\delta}{2}}C_2}.
\end{aligned}$$

As long as $\delta > 0$, for finite C_1 and finite C_2, we have ${}^{C_1}\!/_{n^{\frac{\delta}{2}}C_2} \to 0$, as $n \to \infty$. This means that the Lyapunov condition is satisfied, if $\mathbb{E}\left|G_{ni}\right|^{2+\delta}$ and $s_n^{2+\delta}$ are finite. Then, by Lyapunov Central Limit Theorem, we have

$$\frac{T_n}{s_n} \xrightarrow{d} N(0,1).$$

As this hold for any arbitary non-random vector $\boldsymbol{a} \in \mathbb{R}^p$, we have, by Cramer-Wold Theorem, that

$$\sqrt{n}\left[\sum_{i=1}^{n}\frac{2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}_i - \mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right] \xrightarrow{d} N\left(0, V\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right),$$

as $n \to \infty$. Denote $\boldsymbol{L}_{ni} = \frac{2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j}{h} k\left(-\frac{\boldsymbol{X}_i^{\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}^i$, then this is written as

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}\right) \xrightarrow{d} N\left(0, V\left(\boldsymbol{L}_{n1}\right)\right).$$

Then, we have

$$\frac{1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}}{[V(\boldsymbol{L}_{n1})/n]^{1/2}} \frac{[V(\boldsymbol{L}_{n1})/n]^{1/2}}{\left[AV(\boldsymbol{L}_{n1})/n\right]^{1/2}} \xrightarrow{d} N(0,1).$$

As $n \to \infty$,

$$\frac{V(\boldsymbol{L}_{n1})^{1/2}}{AV(\boldsymbol{L}_{n1})^{1/2}} \to 1,$$

then we have

$$\frac{1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}}{[AV(\boldsymbol{L}_{n1})/n]^{1/2}} \xrightarrow{d} N(0,1),$$

i.e.,

$$\sqrt{n}\left[1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}\right] \xrightarrow{d} N\left(0, AV(\boldsymbol{L}_{n1})\right).$$

As $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{L}_{ni} = \frac{1}{n}\sum_{i=1}^{n}\frac{2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}^i = \nabla\widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right)$, we have

$$\sqrt{n}\left[\nabla\widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right) - \mathbb{E}\left\{\frac{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right] \xrightarrow{d} N\left(0, AV\left\{\frac{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j}{h}k\left(-\frac{\boldsymbol{X}^{\intercal}\boldsymbol{\theta}}{h}\right)\boldsymbol{X}\right\}\right).$$

■

A.6 Proof of Corollary 1.1.4

Corollary A.6.1. *Suppose all the assumptions in lemma 3 hold. Also, $\widehat{\boldsymbol{\theta}}_\nu(\mu)$ and $\widehat{\boldsymbol{\beta}}_j$ are consistent estimators of $\boldsymbol{\theta}^*_\nu(\mu)$ and $\boldsymbol{\beta}^*_j$, respectively. Then,*

$$\sqrt{n}\left(\nabla\widehat{V}_j\left(\boldsymbol{\theta}^*_\nu(\mu),\widehat{\boldsymbol{\beta}}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}^*_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*_\nu(\mu)}{h}\right)\boldsymbol{X}\right\}\right)\xrightarrow{d}N\left(0,AV\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}^*_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*_\nu(\mu)}{h}\right)\boldsymbol{X}\right\}\right).$$

Proof. For notation simplicity, again let $\boldsymbol{\theta}^*=\boldsymbol{\theta}^*_\nu(\mu)$ and $\widehat{\boldsymbol{\theta}}=\widehat{\boldsymbol{\theta}}_\nu(\mu)$ here. Write

$$\begin{aligned}
&\nabla\widehat{V}_j\left(\boldsymbol{\theta}^*,\widehat{\boldsymbol{\beta}}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}^*_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\\
=&\left(\nabla\widehat{V}_j\left(\boldsymbol{\theta}^*,\widehat{\boldsymbol{\beta}}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\widehat{\boldsymbol{\beta}}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\right)+\\
&\left(\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\widehat{\boldsymbol{\beta}}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}^*_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\right)\\
=&\nabla\widehat{V}_j\left(\boldsymbol{\theta}^*,\widehat{\boldsymbol{\beta}}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\widehat{\boldsymbol{\beta}}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}+o_p(1)
\end{aligned}$$

For the second equality, as $\widehat{\boldsymbol{\beta}}_j$is consistent, $\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\widehat{\boldsymbol{\beta}}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}$
$-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}^*_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}=o_p(1)$ which can be proven by Taylor expansions. Let $\boldsymbol{\theta}=\boldsymbol{\theta}^*$ and $\boldsymbol{\beta}_j=\widehat{\boldsymbol{\beta}}_j$ in lemma 1.1.3, and then

$$\sqrt{n}\left(\nabla\widehat{V}_j\left(\boldsymbol{\theta}^*,\widehat{\boldsymbol{\beta}}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\widehat{\boldsymbol{\beta}}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\right)\xrightarrow{d}N\left(0,AV\left\{\frac{2\boldsymbol{X}_1^\intercal\widehat{\boldsymbol{\beta}}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\right)$$

As $\widehat{\boldsymbol{\beta}}_j$ are consistent estimators,

$$\frac{AV\left\{\frac{2\boldsymbol{X}_1^\intercal\widehat{\boldsymbol{\beta}}_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}}{AV\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}^*_j}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}}\xrightarrow{p}1.$$

Together, it is proven that

$$\sqrt{n}\left(\nabla\widehat{V}_j\left(\boldsymbol{\theta}^*,\widehat{\boldsymbol{\beta}}_j\right)-\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\right)\xrightarrow{d}N\left(0,AV\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\right)$$

■

A.7 Proof of Theorem 1.1.5

Theorem A.7.1. *Suppose all the assumptions in Lemma 1.1.3 and Corollary 1.1.4 hold. Then we have, as* $n\to\infty$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\nu(\mu)-\boldsymbol{\theta}_{\boldsymbol{\nu}}(\mu^*)\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^*=\boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\nu(\mu)-\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^*=\boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$,
$\boldsymbol{C}^*=\mathbb{E}\left\{\nabla v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\nabla v_1^\intercal\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right\}-\mathbb{E}\left\{\nabla v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right\}\mathbb{E}\left\{\nabla v_1^\intercal\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right\}$, *and* $\boldsymbol{D}^*=\nabla^2\phi(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu))$.

A.7.1 Related Limits

Here, we exposit the terms that are involved in deriving the limiting distribution of $\widehat{\boldsymbol{\theta}}_\nu(\mu)$.

$V_j(\boldsymbol{\theta})$ and $\widehat{V}_j(\boldsymbol{\theta})$

Recall $V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)=\mathbb{E}\left\{\boldsymbol{X}_0^\intercal\boldsymbol{\alpha}_j+\operatorname{sgn}\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}\right)\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j\right\}$
$=m_{\boldsymbol{\alpha}_j}+\iint\operatorname{sgn}(z_1)\,z_2 f_{\boldsymbol{\beta}_j}(z_1,z_2;\boldsymbol{\theta})\,dz_1\,dz_2$, and
$\widehat{V}_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)=\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{X}_0^{i\intercal}\boldsymbol{\alpha}_j+\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j\left\{1-2K\left(-\frac{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}{h}\right)\right\}\right)$. As stated before, due the the consistency of and the KDEs, we have $p\lim_{n\to\infty}\widehat{V}_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)=V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)$, where $p\lim$ means converging in probability. As $\widehat{\boldsymbol{\alpha}}_j$ and $\widehat{\boldsymbol{\beta}}_j$ are consistent estimators of $\boldsymbol{\alpha}_j^*$

and $\boldsymbol{\beta}_j^*$, we have $p\lim_{n\to\infty}\widehat{V}_j\left(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}_j,\widehat{\boldsymbol{\beta}}_j\right)=V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j^*,\boldsymbol{\beta}_j^*\right)$. As $\widehat{V}_j\left(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}_j,\widehat{\boldsymbol{\beta}}_j\right)$ is denoted by $\widehat{V}_j\left(\boldsymbol{\theta}\right)$ and $V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j^*,\boldsymbol{\beta}_j^*\right)$ is denoted by $V_j\left(\boldsymbol{\theta}\right)$, we have

$$p\lim_{n\to\infty}\widehat{V}_j\left(\boldsymbol{\theta}\right)=V_j\left(\boldsymbol{\theta}\right) \tag{A.4}$$

$\nabla V_j(\boldsymbol{\theta})$ and $\nabla\widehat{V}_j(\boldsymbol{\theta})$

The gradient of the value function with respect to $\boldsymbol{\theta}$ is
$\nabla V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)=\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left\{\operatorname{sgn}\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}\right)\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j\right\}=\frac{\partial}{\partial\boldsymbol{\theta}}\iint\operatorname{sgn}\left(z_1\right)z_2f_{\boldsymbol{\beta}_j}\left(z_1,z_2;\boldsymbol{\theta}\right)\,dz_1\,dz_2$. The interchangeability between integral and differentiation is assumed to hold. Then, we can write $\nabla V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)=\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left\{\operatorname{sgn}\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}\right)\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j\right\}=\int_{\boldsymbol{\mathcal{X}}}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\operatorname{sgn}\left(\boldsymbol{x}^\intercal\boldsymbol{\theta}\right)\boldsymbol{x}^\intercal\boldsymbol{\beta}_j\right\}f(\boldsymbol{x})\,d\boldsymbol{x}=\int_{\boldsymbol{\mathcal{X}}}2\delta\left(\boldsymbol{x}^\intercal\boldsymbol{\theta}\right)\boldsymbol{x}^\intercal\boldsymbol{\beta}_jf(\boldsymbol{x})\,d\boldsymbol{x}=\mathbb{E}\left\{2\delta\left(\boldsymbol{x}^\intercal\boldsymbol{\theta}\right)\boldsymbol{x}^\intercal\boldsymbol{\beta}_j\right\}$, where $\delta(x)=\frac{\partial}{\partial\boldsymbol{\theta}}\operatorname{sgn}(x)$ is the Dirac delta function. Our kernel $k(x)$ is the Gaussian Kernel, where $k(x)={}^1\!/\!_{\sqrt{2\pi}}\exp\left(-{}^{x^2}\!/\!_2\right)$. Then, $\frac{1}{h}k\left(\frac{x}{h}\right)=\frac{1}{h\sqrt{2\pi}}\exp\left(-\frac{x^2}{2h^2}\right)$. It is defined the Dirac delta function $\delta(x)$ to be the limit (in the sense of distributions) of the sequence of zero-centered normal distributions, i.e., $\delta(x)=\lim_{h\to0}{}^1\!/\!_{h\sqrt{2\pi}}\exp\left(-{}^{x^2}\!/\!_{2h^2}\right)=\lim_{h\to0}{}^1\!/\!_{hk}\left({}^x\!/\!_h\right)$. It is an even distribution, such that $\delta(x)=\delta(-x)$.Moreover, $\nabla\widehat{V}_j\left(\boldsymbol{\theta},\boldsymbol{\beta}_j\right)={}^1\!/\!_n\sum_{i=1}^n{}^{2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j}\!/\!_{hk}\left(-{}^{\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}}\!/\!_h\right)\boldsymbol{X}^i$. Its expectation is $\mathbb{E}\left\{\nabla\widehat{V}_j\left(\boldsymbol{\theta},\boldsymbol{\beta}_j\right)\right\}=\mathbb{E}\left\{{}^{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j}\!/\!_{hk}\left(-{}^{\boldsymbol{X}^\intercal\boldsymbol{\theta}}\!/\!_h\right)\boldsymbol{X}\right\}=\mathbb{E}\left\{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j\cdot\delta\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}\right)\right\}$. This is because, as $n\to\infty$, $h\to0$ and $nh\to\infty$, we have $\lim_{n\to\infty}{}^{2\boldsymbol{x}_1^\intercal\boldsymbol{\beta}_j}\!/\!_{hk}\left(-{}^{\boldsymbol{x}^\intercal\boldsymbol{\theta}}\!/\!_h\right)=2\boldsymbol{x}_1^\intercal\boldsymbol{\beta}_j\cdot\delta\left(-\boldsymbol{x}^\intercal\boldsymbol{\theta}\right)=2\boldsymbol{x}_1^\intercal\boldsymbol{\beta}_j\cdot\delta\left(\boldsymbol{x}^\intercal\boldsymbol{\theta}\right)$. Thus, by weak law of large numbers, $p\lim_{n\to\infty}\nabla\widehat{V}_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)=\nabla V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)$. Together, with the consistency of $\widehat{\boldsymbol{\alpha}}_j$ and $\widehat{\boldsymbol{\beta}}_j$, we have
$p\lim_{n\to\infty}\nabla\widehat{V}_j\left(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}_j,\widehat{\boldsymbol{\beta}}_j\right)=\nabla V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j^*,\boldsymbol{\beta}_j^*\right)$. Using the simplified notation, that is

$$p\lim_{n\to\infty}\nabla\widehat{V}_j\left(\boldsymbol{\theta}\right)=\nabla V_j\left(\boldsymbol{\theta}\right) \tag{A.5}$$

$\nabla^2V_j(\boldsymbol{\theta})$ and $\nabla^2\widehat{V}_j(\boldsymbol{\theta})$

The second order derivative of value function, or Hessian, is
$\nabla^2V_j\left(\boldsymbol{\theta},\boldsymbol{\alpha}_j,\boldsymbol{\beta}_j\right)=\nabla^2\mathbb{E}\left\{\operatorname{sgn}\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}\right)\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j\right\}=\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\iint\operatorname{sgn}\left(z_1\right)z_2f_{\boldsymbol{\beta}_j}\left(z_1,z_2;\boldsymbol{\theta}\right)\,dz_1\,dz_2$. Again, the interchangeability between integral and differentiation is assumed to hold.
$\nabla^2\mathbb{E}\{\operatorname{sgn}(\boldsymbol{X}^\intercal\boldsymbol{\theta})\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j\}=\int_{\boldsymbol{\mathcal{X}}}\left\{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\operatorname{sgn}\left(\boldsymbol{x}^\intercal\boldsymbol{\theta}\right)\boldsymbol{x}_1^\intercal\boldsymbol{\beta}_j\right\}f\left(\boldsymbol{x}\right)\,d\boldsymbol{x}$
$=\int_{\boldsymbol{\mathcal{X}}}2\boldsymbol{x}_1^\intercal\boldsymbol{\beta}_j\delta'\left(\boldsymbol{x}^\intercal\boldsymbol{\theta}\right)\boldsymbol{x}\boldsymbol{x}^\intercal f(\boldsymbol{x})\,d\boldsymbol{x}=\mathbb{E}\left\{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_j\delta'\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}\right)\boldsymbol{X}\boldsymbol{X}^\intercal\right\}$, where $\delta'(x)$ is the distri-

butional derivative of the Dirac function. $\delta'(x) = \lim_{h\to 0} {}^{1}\!/_{h^2} k'\left(\frac{x}{h}\right)$
$= \lim_{h\to 0} {}^{-x}\!/_{h^3\sqrt{2\pi}} \exp\left(-{}^{x^2}\!/_{2h^2}\right)$. Moreover,
$\nabla^2 \widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right) = {}^{1}\!/_{n} \sum_{i=1}^{n} \left\{-2\boldsymbol{X}_1^{i\intercal}\boldsymbol{\beta}_j/h^2 k'\left(-\boldsymbol{X}^{i\intercal}\boldsymbol{\theta}/h\right)\boldsymbol{X}^i\boldsymbol{X}^{i\intercal}\right\}$. Its expectation is
$\mathbb{E}\left\{\nabla^2 \widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right)\right\} = \mathbb{E}\left\{-2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j/h^2 k'\left(-\boldsymbol{X}^{\intercal}\boldsymbol{\theta}/h\right)\boldsymbol{X}\boldsymbol{X}^{\intercal}\right\} = \mathbb{E}\left\{2\boldsymbol{X}_1^{\intercal}\boldsymbol{\beta}_j \delta'\left(\boldsymbol{X}^{\intercal}\boldsymbol{\theta}\right)\boldsymbol{X}\boldsymbol{X}^{\intercal}\right\}$.
Thus, by weak law of large numbers, $p\lim_{n\to\infty} \nabla^2 \widehat{V}_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right) = \nabla^2 V_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j\right)$. Together, with the consistency of $\widehat{\boldsymbol{\beta}}_j$, $p\lim_{n\to\infty} \nabla^2 \widehat{V}_j\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}}_j\right) = \nabla^2 V_j\left(\boldsymbol{\theta}, \boldsymbol{\beta}_j^*\right)$. Using simplified notation, we have

$$p \lim_{n\to\infty} \nabla^2 \widehat{V}_j\left(\boldsymbol{\theta}\right) = \nabla^2 V_j\left(\boldsymbol{\theta}\right) \tag{A.6}$$

A.7.2 Proof

We derive the limiting distribution of $\widehat{\boldsymbol{\theta}}_\nu(\mu)$, and prove Theorem 1.1.5.

Proof. For notation simplicity in this proof, let $\phi\left(\boldsymbol{\theta}\right) = \phi_\mu^{PB}\left(\boldsymbol{\theta}\right)$ and $\widehat{\phi}(\boldsymbol{\theta}) = \widehat{\phi}_\mu^{PB}\left(\boldsymbol{\theta}\right)$ for this proof. Also, let $\boldsymbol{\theta}^* = \boldsymbol{\theta}_\nu^*(\mu)$ and $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_\nu(\mu)$ here. Recall $\widehat{\phi}\left(\boldsymbol{\theta}\right) = \widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J} \ln \widehat{v}_j\left(\boldsymbol{\theta}\right) + {}^{1}\!/_{2\mu}\left(\boldsymbol{\theta}^{\intercal}\boldsymbol{\theta} - 1\right)^2$. As $\boldsymbol{\theta}^{\intercal}\boldsymbol{\theta} - 1 = 0$ is always satisfied as a constraint, the gradient is $\nabla\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J} {}^{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}\!/_{\widehat{v}_j\left(\boldsymbol{\theta}\right)}$. Taylor expansion of $\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right)$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ shows that

$$\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = \nabla\widehat{\phi}(\widehat{\boldsymbol{\theta}}) - \nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + o_p(1),$$

where $\tilde{\boldsymbol{\theta}}$ is between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$. As $\widehat{\boldsymbol{\theta}}$ is the maximizer of $\widehat{\phi}\left(\boldsymbol{\theta}\right)$, it satisfies the first order condition that $\nabla\widehat{\phi}(\widehat{\boldsymbol{\theta}}) = 0$. Therefore,

$$\sqrt{n}\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = -\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \tag{A.7}$$

where $\nabla\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J} {}^{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}\!/_{\widehat{v}_j\left(\boldsymbol{\theta}\right)}$. Recall $v_1\left(\boldsymbol{\theta}\right) = -V_1\left(\boldsymbol{\theta}\right)$ and $v_j\left(\boldsymbol{\theta}\right) = V_j\left(\boldsymbol{\theta}\right) - \nu_j$, for $j = 2, \cdots, J$. Due to Corollary 1.1.4, together with (A.4) and (A.5),

$$\sqrt{n}\left(\nabla\widehat{v}_1(\boldsymbol{\theta}^*) - \nabla v_1(\boldsymbol{\theta}^*)\right) \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right), \tag{A.8}$$

where $\boldsymbol{C}^* = AV\left(\nabla v_1(\boldsymbol{\theta}^*)\right) = AV\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}$. That is,

$$
\begin{aligned}
\boldsymbol{C}^* \triangleq & AV\left(\nabla v_1(\boldsymbol{\theta}^*)\right) = AV\left[\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right] = p\lim_{n\to\infty} V\left[\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right] \\
= & p\lim_{n\to\infty}\left[\mathbb{E}\left\{\frac{4\boldsymbol{\beta}_1^{*\intercal}\boldsymbol{X}_1\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*}{h^2}k^2\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\boldsymbol{X}^\intercal\right\} - \mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}\mathbb{E}\left\{\frac{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*}{h}k\left(-\frac{\boldsymbol{X}^\intercal\boldsymbol{\theta}^*}{h}\right)\boldsymbol{X}\right\}^\intercal\right] \\
= & \mathbb{E}\left\{4\left(\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*\delta\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}^*\right)\right)^2\boldsymbol{X}\boldsymbol{X}^\intercal\right\} - \mathbb{E}\left\{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*\delta\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}^*\right)\boldsymbol{X}\right\}\mathbb{E}\left\{2\boldsymbol{X}_1^\intercal\boldsymbol{\beta}_1^*\delta\left(\boldsymbol{X}^\intercal\boldsymbol{\theta}^*\right)\boldsymbol{X}\right\}^\intercal \\
= & \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\right\}\mathbb{E}\left\{\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\}.
\end{aligned}
$$

Then, due to (A.4) and (A.5), we have

$$\sum_{j=2}^{J}\frac{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}{\widehat{v}_j\left(\boldsymbol{\theta}\right)} - \sum_{i=2}^{J}\frac{\nabla v_j(\boldsymbol{\theta})}{v_j(\boldsymbol{\theta})} = o_p(1). \tag{A.9}$$

Note $v_j(\boldsymbol{\theta}) > 0$, for $j = 2, \cdots, J$, is implied by the log barrier operator. Put (A.8) and (A.9) together by Slutsky's theorem, we have

$$\sqrt{n}\left\{\left(\nabla\widehat{v}_1(\boldsymbol{\theta}^*) - \mu\sum_{j=2}^{J}\frac{\nabla\widehat{v}_j\left(\boldsymbol{\theta}^*\right)}{\widehat{v}_j(\boldsymbol{\theta}^*)}\right) - \left(\nabla v_1(\boldsymbol{\theta}^*) - \mu\sum_{i=2}^{J}\frac{\nabla v_j(\boldsymbol{\theta}^*)}{v_j(\boldsymbol{\theta}^*)}\right)\right\} \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right),$$

Due to the stationarity of $\boldsymbol{\theta}^*$, $\nabla\phi(\boldsymbol{\theta}^*) = \nabla v_1(\boldsymbol{\theta}^*) - \mu\sum_{i=2}^{J}{}^{\nabla v_j(\boldsymbol{\theta}^*)}/_{v_j(\boldsymbol{\theta}^*)} = 0$. Together with Sluskty's theorem, we have

$$\sqrt{n}\nabla\widehat{\phi}(\boldsymbol{\theta}^*) \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right),$$

where $\boldsymbol{C}^* = \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\right\}\mathbb{E}\left\{\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\}$.

As $\sqrt{n}\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = -\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ stated in (A.7),we have

$$\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N(0, \boldsymbol{C}^*) \tag{A.10}$$

The Hessian is $\nabla^2\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla^2\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J}\left(\nabla^2\widehat{v}_j\left(\boldsymbol{\theta}\right)\widehat{v}_j\left(\boldsymbol{\theta}\right) - \left(\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)\right)^2\right)/\widehat{v}_j^2\left(\boldsymbol{\theta}\right)$. Based on (A.4)

and (A.5), we have

$$\boldsymbol{D}^* \triangleq_p \lim_{n\to\infty} \nabla^2\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = \nabla^2\phi(\boldsymbol{\theta}^*) = \nabla^2 v_1(\boldsymbol{\theta}^*) - \mu\sum_{j=2}^{J}\frac{\nabla^2 v_j\left(\boldsymbol{\theta}^*\right) v_j\left(\boldsymbol{\theta}^*\right) - \left\{\nabla v_j\left(\boldsymbol{\theta}^*\right)\right\}^2}{v_j^2\left(\boldsymbol{\theta}^*\right)}. \tag{A.11}$$

As $\tilde{\boldsymbol{\theta}}$ is a vector in-between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}$, we have $\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}}) = \nabla^2\widehat{\phi}(\boldsymbol{\theta}^*) + o_p(1)$. Therefore, based on (A.10) and (A.11), we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^* = \boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$, $\boldsymbol{C}^* = \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\nabla v_1(\boldsymbol{\theta}^*)\mathbb{E}\nabla^\intercal v_1(\boldsymbol{\theta}^*)$ and $\boldsymbol{D}^* = \nabla^2\phi(\boldsymbol{\theta}^*)$. ■

A.8 Details on simulation

A.8.1 Parameters

To find parameter values in the generative model that satisfy the levels of these two factors, we use the solver `fmincon` in Matlab to minimize the sum of two empirical quadratic loss functions for Ω_1 and Ω_2 and find a set of possible solution.

A.8.2 Details about simulation studies with kernel density estimation

A.8.3 Simulation results

Table A.1: Simulation Result for Setting 2

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-2.62	3.94	0.16	-2.61	0.10	0.14	0.07	0.99	0.01
-2.26	4.32	0.07	-2.24	0.10	-0.06	0.05	1.00	0.00
-1.89	4.48	0.03	-1.89	0.12	-0.22	0.04	0.98	0.01
-1.52	4.54	0.01	-1.53	0.10	-0.35	0.04	0.94	0.01
-1.16	4.55	0.00	-1.35	0.10	-0.41	0.04	0.91	0.02
-0.79	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
-0.42	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
-0.06	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
0.31	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
0.68	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
1.04	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
1.41	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
1.78	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
2.14	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
2.51	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
2.88	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
3.24	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02
3.61	4.44	0.89	-1.26	0.63	-0.41	0.04	0.88	0.23
3.97	4.55	0.00	-1.34	0.11	-0.42	0.04	0.91	0.02

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

Table A.2: Simulation Result for Setting 3

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-2.50	1.19	0.19	-2.55	0.15	-0.05	0.10	0.99	0.01
-2.15	1.33	0.05	-2.18	0.16	-0.24	0.07	0.97	0.02
-1.80	1.42	0.04	-1.82	0.17	-0.39	0.06	0.92	0.03
-1.45	1.50	0.04	-1.48	0.17	-0.51	0.06	0.86	0.03
-1.09	1.56	0.03	-1.15	0.18	-0.60	0.05	0.79	0.04
-0.74	1.61	0.02	-0.83	0.20	-0.69	0.05	0.72	0.04
-0.39	1.62	0.02	-0.53	0.26	-0.75	0.06	0.66	0.06
-0.04	1.63	0.01	-0.30	0.32	-0.79	0.06	0.61	0.07
0.32	1.63	0.01	-0.18	0.43	-0.81	0.07	0.58	0.10
0.67	1.63	0.01	-0.06	0.47	-0.82	0.08	0.55	0.11
1.02	1.63	0.01	-0.08	0.51	-0.82	0.08	0.55	0.12
1.37	1.63	0.01	-0.03	0.52	-0.83	0.08	0.54	0.13
1.73	1.62	0.02	-0.04	0.53	-0.83	0.08	0.54	0.13
2.08	1.63	0.01	-0.04	0.50	-0.83	0.08	0.54	0.12
2.43	1.63	0.01	-0.03	0.51	-0.83	0.08	0.54	0.12
2.78	1.63	0.01	-0.03	0.51	-0.83	0.08	0.54	0.12
3.13	1.63	0.01	-0.02	0.50	-0.83	0.08	0.54	0.12
3.49	1.62	0.13	0.01	0.57	-0.83	0.10	0.53	0.17
3.84	1.63	0.01	0.00	0.52	-0.83	0.08	0.53	0.13

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

Table A.3: Simulation Result for Setting 4

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-1.57	1.62	0.01	-1.82	0.03	0.14	0.12	-0.98	0.02
-1.32	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
-1.08	1.62	0.02	-1.82	0.03	0.15	0.13	-0.98	0.02
-0.84	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
-0.59	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
-0.35	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
-0.10	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
0.14	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
0.38	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
0.63	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
0.87	1.62	0.02	-1.81	0.04	0.14	0.14	-0.98	0.02
1.11	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
1.36	1.62	0.02	-1.81	0.04	0.15	0.13	-0.98	0.02
1.60	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
1.84	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
2.09	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
2.33	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
2.57	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02
2.82	1.62	0.02	-1.81	0.04	0.14	0.13	-0.98	0.02

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

Table A.4: Simulation Result for Setting 5

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-2.58	0.64	0.04	-2.62	0.15	0.80	0.09	-0.59	0.10
-2.22	0.70	0.03	-2.29	0.20	0.90	0.07	-0.42	0.11
-1.85	0.74	0.04	-1.97	0.24	0.95	0.07	-0.27	0.12
-1.49	0.76	0.04	-1.68	0.33	0.98	0.05	-0.16	0.14
-1.13	0.79	0.04	-1.37	0.39	0.99	0.04	-0.05	0.15
-0.77	0.80	0.04	-1.12	0.46	0.98	0.03	0.04	0.17
-0.40	0.80	0.05	-0.91	0.59	0.97	0.06	0.10	0.22
-0.04	0.80	0.05	-0.67	0.63	0.95	0.13	0.18	0.22
0.32	0.80	0.06	-0.44	0.75	0.92	0.15	0.24	0.27
0.68	0.80	0.06	-0.23	0.82	0.91	0.15	0.29	0.26
1.05	0.80	0.07	-0.08	0.93	0.88	0.19	0.33	0.28
1.41	0.80	0.07	0.10	0.97	0.86	0.20	0.37	0.28
1.77	0.80	0.05	0.15	1.04	0.86	0.17	0.39	0.29
2.13	0.80	0.05	0.24	1.10	0.84	0.18	0.41	0.30
2.50	0.80	0.05	0.29	1.14	0.83	0.19	0.42	0.31
2.86	0.80	0.03	0.38	1.20	0.82	0.17	0.44	0.31
3.22	0.80	0.03	0.44	1.21	0.81	0.19	0.46	0.31
3.58	0.80	0.03	0.34	1.19	0.83	0.18	0.44	0.31
3.94	0.80	0.04	0.48	1.29	0.80	0.22	0.46	0.32

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

Table A.5: Simulation Result for Setting 6

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-0.50	0.59	0.11	-0.55	0.41	-0.83	0.31	-0.44	0.13
-0.38	0.73	0.08	-0.36	0.07	-0.96	0.03	-0.26	0.11
-0.26	0.85	0.07	-0.25	0.08	-0.99	0.01	-0.11	0.10
-0.14	0.94	0.06	-0.13	0.09	-1.00	0.01	0.02	0.09
-0.03	1.01	0.05	-0.00	0.09	-0.99	0.01	0.15	0.08
0.09	1.07	0.04	0.12	0.09	-0.96	0.02	0.26	0.08
0.21	1.12	0.03	0.24	0.09	-0.93	0.03	0.36	0.07
0.33	1.15	0.14	0.36	0.09	-0.87	0.17	0.45	0.13
0.44	1.18	0.14	0.48	0.08	-0.82	0.17	0.54	0.13
0.56	1.21	0.14	0.60	0.09	-0.76	0.17	0.62	0.13
0.68	1.20	0.25	0.71	0.09	-0.65	0.30	0.67	0.22
0.80	1.23	0.19	0.80	0.09	-0.61	0.23	0.74	0.17
0.91	1.23	0.18	0.88	0.10	-0.56	0.23	0.78	0.16
1.03	1.25	0.09	0.92	0.12	-0.54	0.15	0.82	0.10
1.15	1.26	0.01	0.95	0.14	-0.52	0.12	0.84	0.08
1.27	1.25	0.01	0.96	0.15	-0.51	0.13	0.85	0.08
1.38	1.25	0.01	0.96	0.15	-0.51	0.12	0.85	0.08
1.50	1.25	0.01	0.96	0.15	-0.51	0.13	0.85	0.08
1.62	1.25	0.01	0.96	0.15	-0.51	0.13	0.85	0.08

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

Table A.6: Simulation Result for Setting 7

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-3.01	1.12	0.01	-3.28	0.12	-0.70	0.10	0.70	0.10
-2.60	1.12	0.01	-3.27	0.14	-0.69	0.11	0.71	0.10
-2.19	1.12	0.01	-3.26	0.16	-0.69	0.11	0.71	0.11
-1.77	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
-1.36	1.12	0.01	-3.27	0.14	-0.69	0.11	0.70	0.11
-0.95	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
-0.54	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
-0.12	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
0.29	1.12	0.01	-3.25	0.17	-0.69	0.11	0.71	0.11
0.70	1.12	0.01	-3.26	0.17	-0.69	0.12	0.71	0.11
1.11	1.12	0.01	-3.25	0.17	-0.68	0.11	0.71	0.11
1.53	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
1.94	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
2.35	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
2.76	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
3.18	1.12	0.01	-3.26	0.17	-0.69	0.11	0.71	0.11
3.59	1.12	0.01	-3.25	0.17	-0.68	0.11	0.71	0.11
4.00	1.12	0.01	-3.25	0.17	-0.69	0.11	0.71	0.11
4.41	1.11	0.08	-3.22	0.57	-0.68	0.15	0.70	0.14

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

Table A.7: Simulation Result for Setting 8

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-2.80	-0.36	0.33	-2.85	0.12	0.78	0.05	-0.63	0.06
-2.41	0.54	0.26	-2.44	0.14	0.64	0.04	-0.77	0.04
-2.02	1.24	0.22	-2.05	0.14	0.51	0.05	-0.86	0.03
-1.64	1.82	0.21	-1.66	0.15	0.39	0.04	-0.92	0.02
-1.25	2.35	0.19	-1.24	0.16	0.27	0.04	-0.96	0.01
-0.86	2.82	0.17	-0.84	0.16	0.15	0.04	-0.99	0.01
-0.47	3.18	0.50	-0.44	0.15	0.05	0.05	-0.99	0.13
-0.08	3.51	0.51	-0.08	0.15	-0.06	0.05	-0.99	0.13
0.31	3.75	0.86	0.30	0.17	-0.16	0.07	-0.96	0.24
0.70	3.92	1.21	0.71	0.16	-0.26	0.09	-0.90	0.33
1.09	4.16	1.19	1.10	0.16	-0.36	0.08	-0.87	0.33
1.47	4.30	1.32	1.51	0.16	-0.45	0.09	-0.81	0.37
1.86	4.22	1.74	1.88	0.15	-0.53	0.12	-0.69	0.48
2.25	3.73	2.35	2.26	0.15	-0.59	0.15	-0.45	0.65
2.64	4.44	1.50	2.65	0.11	-0.71	0.09	-0.56	0.42
3.03	4.79	0.64	2.74	0.06	-0.75	0.04	-0.63	0.19
3.42	4.67	0.92	2.76	0.15	-0.75	0.03	-0.59	0.30
3.81	4.81	0.44	2.75	0.14	-0.75	0.01	-0.64	0.16
4.20	4.85	0.15	2.74	0.12	-0.75	0.02	-0.65	0.07

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

Table A.8: Simulation Result for Setting 9

ν	$\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_1)$	$\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$	$std(\widehat{V}_2)$	$\widehat{\theta}_{\nu,1}$	$std(\widehat{\theta}_{\nu,1})$	$\widehat{\theta}_{\nu,2}$	$std(\widehat{\theta}_{\nu,2})$
-0.70	3.32	1.24	-0.97	0.63	0.75	0.04	-0.66	0.05
-0.56	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
-0.41	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
-0.27	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
-0.13	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
0.01	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
0.15	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
0.29	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
0.43	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
0.58	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
0.72	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
0.86	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
1.00	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
1.14	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
1.28	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
1.43	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
1.57	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
1.71	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04
1.85	3.72	0.01	-0.77	0.02	0.73	0.04	-0.68	0.04

Here, ν denotes the values of the constraint; $\widehat{V}_1(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of primary outcome of interest; $std(\widehat{V}_1)$ denotes the standard deviation of the estimated regime values in terms of primary outcome of interest; $\widehat{V}_2(\widehat{\boldsymbol{\theta}}_\nu)$ denotes the values of estimated regimes in terms of secondary outcome of interest; $std(\widehat{V}_2)$ denotes the standard deviation of the estimated regime values in terms of secondary outcome of interest; $\widehat{\theta}_{\nu,1}$ and $\widehat{\theta}_{\nu,2}$ denote the estimated index parameters of the regimes; $std(\widehat{\theta}_{\nu,1})$ and $std(\widehat{\theta}_{\nu,2})$ denote the standard deviations of those estimated index parameters.

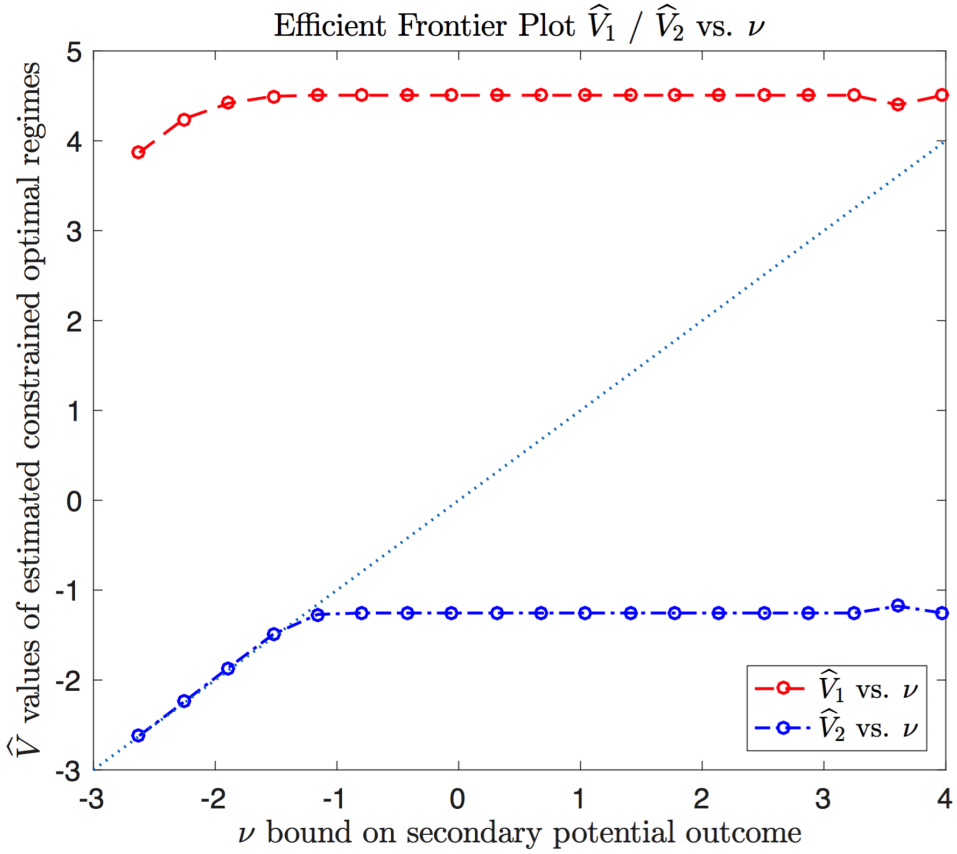

Figure A.1: Efficient frontier for estimated constrained optimal regimes for Setting 2.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

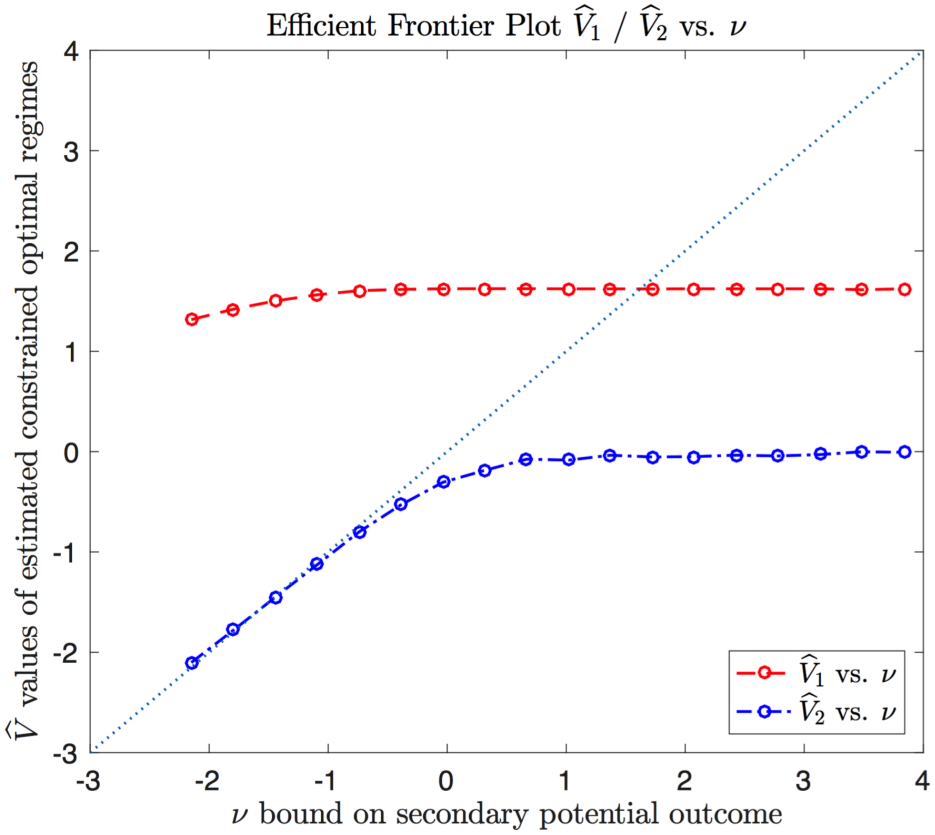

Figure A.2: Efficient frontier for estimated constrained optimal regimes for Setting 3.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

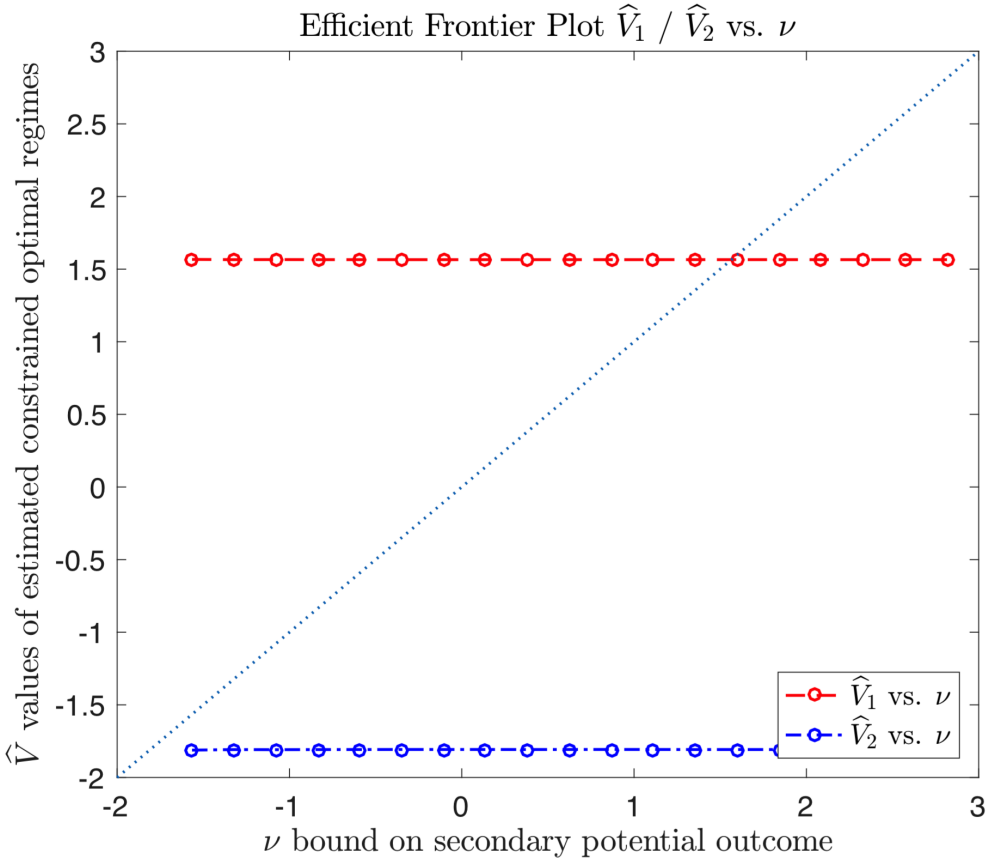

Figure A.3: Efficient frontier for estimated constrained optimal regimes for Setting 4.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

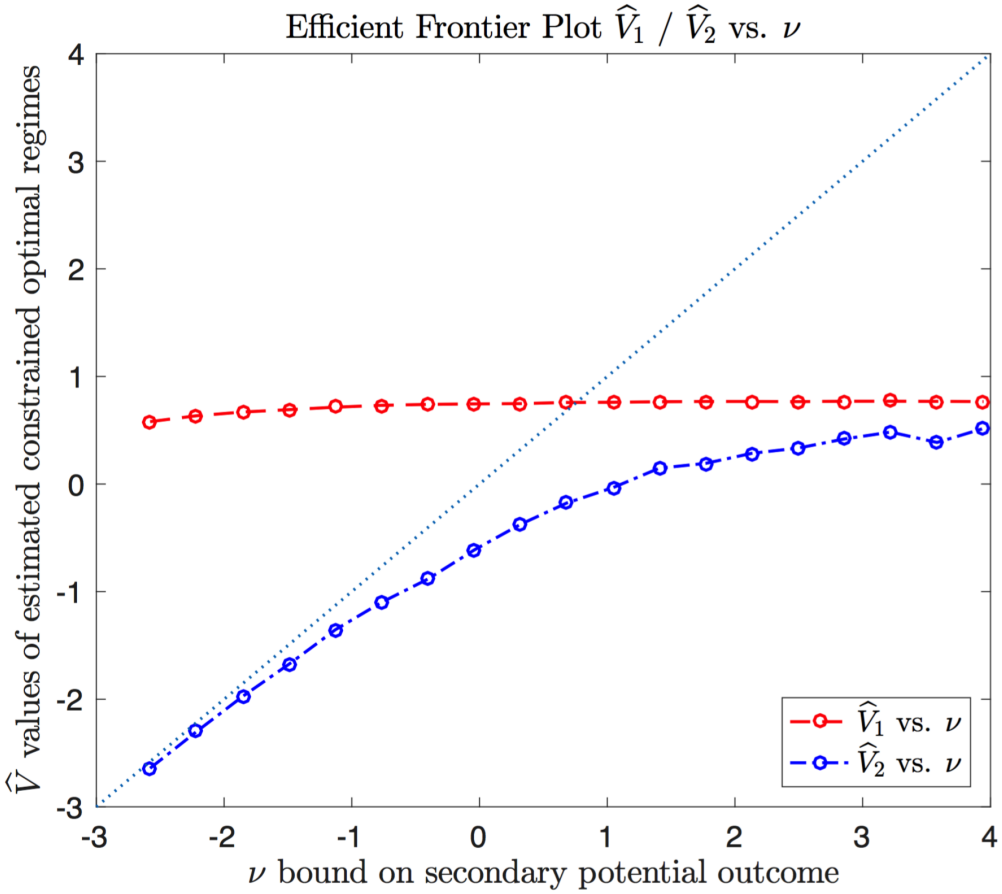

Figure A.4: Efficient frontier for estimated constrained optimal regimes for Setting 5.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

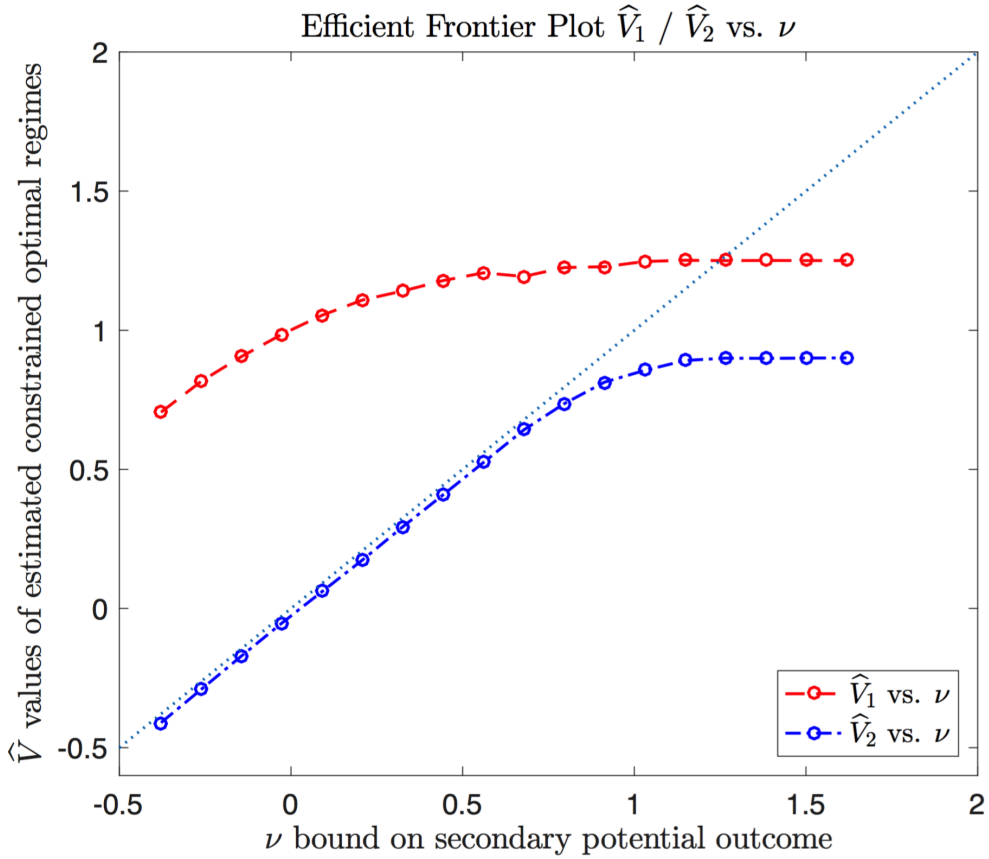

Figure A.5: Efficient frontier for estimated constrained optimal regimes for Setting 6.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

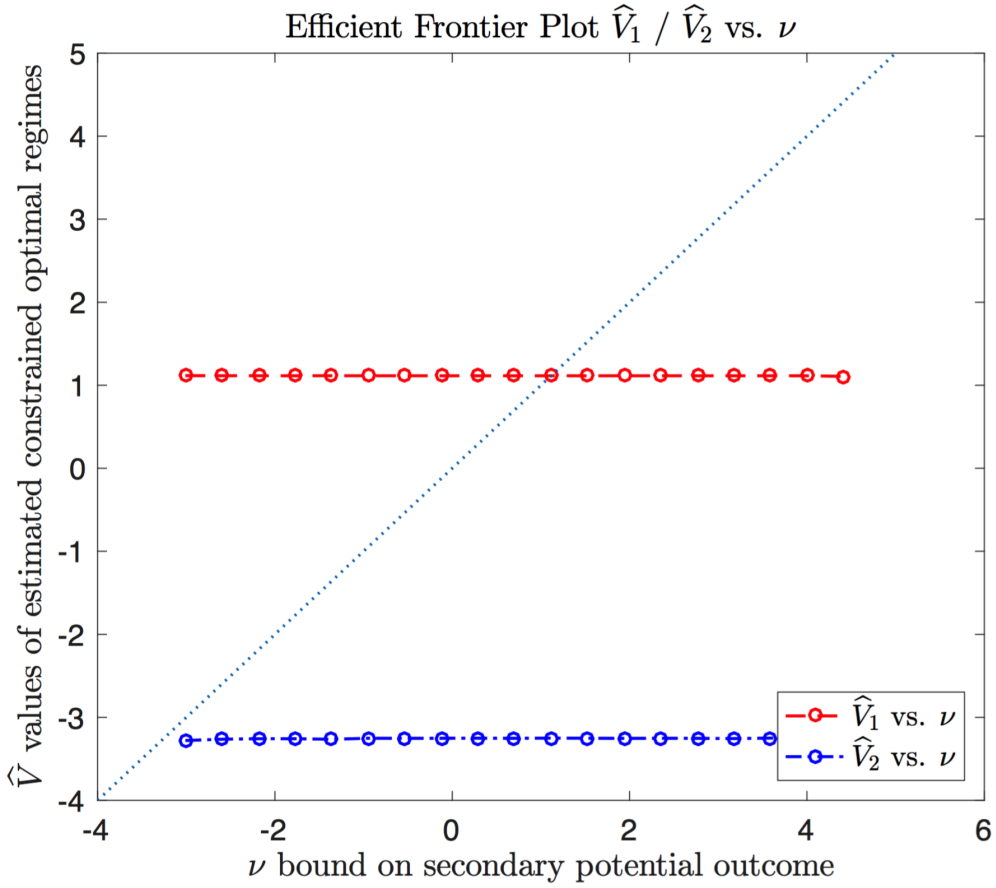

Figure A.6: Efficient frontier for estimated constrained optimal regimes for Setting 7.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

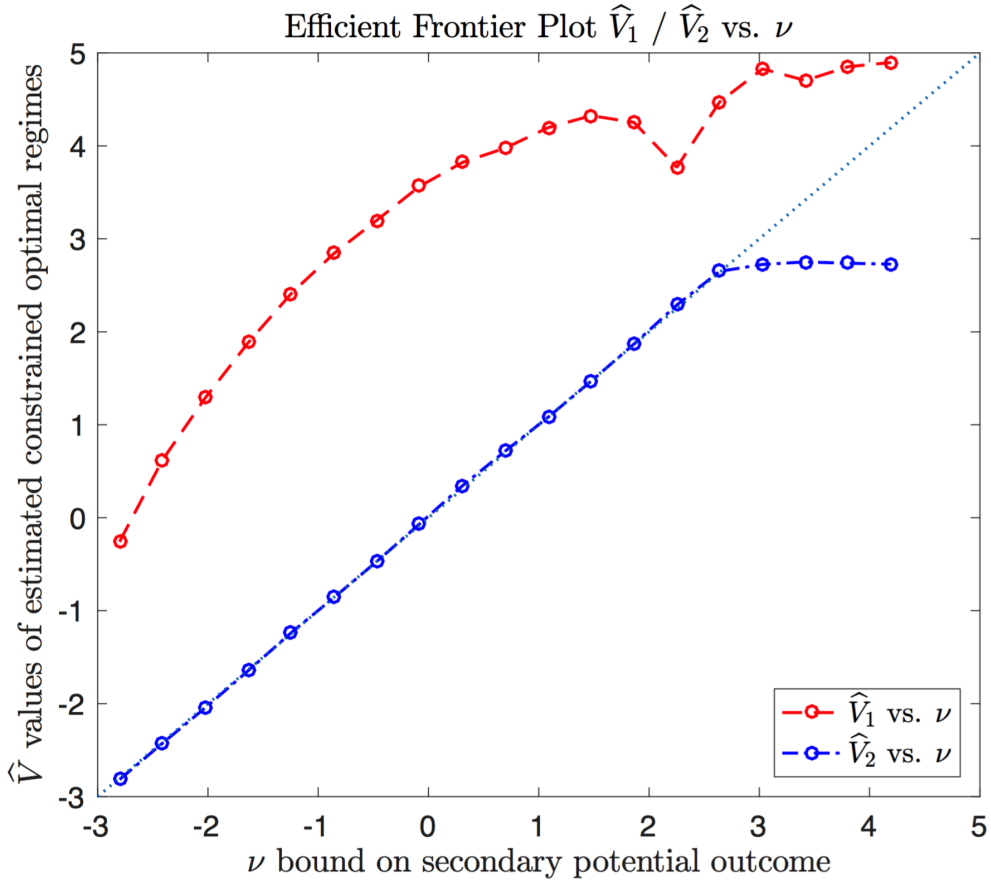

Figure A.7: Efficient frontier for estimated constrained optimal regimes for Setting 8.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

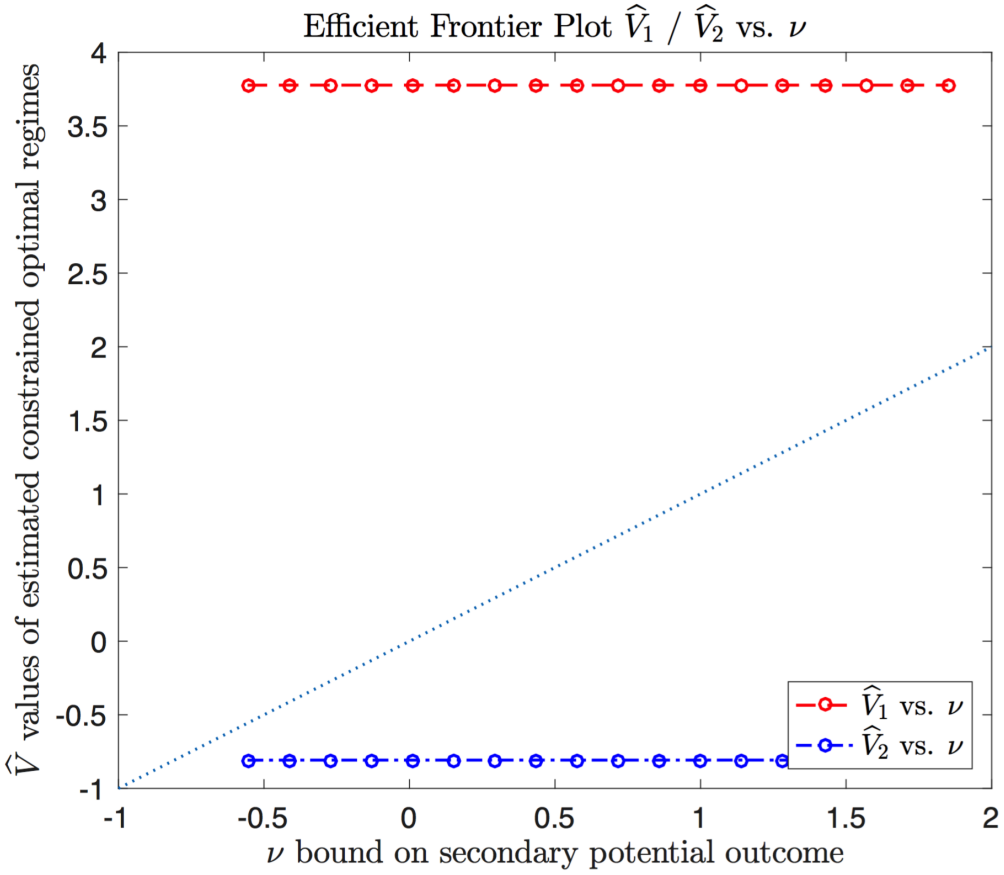

Figure A.8: Efficient frontier for estimated constrained optimal regimes for Setting 9.

X-axis is for the values for the constraints ν; Y-axis is for the values of estimated regimes. Red dashed line is for the values in terms of the primary outcome of interest. Blue dashed line is for the values in terms of the secondary outcome of interest.

Appendix B

Supplement materials for Chapter 3

B.1 Proof of Lemma 2.1.1

Lemma B.1.1. *Suppose the following conditions hold.*

1. $\forall \boldsymbol{a} \in \mathbb{R}^p$,$\exists \delta > 0$,*such that*

   (a) $\mathbb{E}\left|\boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right|^{2+\delta} < \infty$

   (b) $\left\{\boldsymbol{a}^\intercal V\left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right]\boldsymbol{a}\right\}^{1+\frac{\delta}{2}} < \infty.$

Then, we have, for any fixed $\boldsymbol{\theta}$*,*

$$\sqrt{n}\left(\nabla \widehat{V}_j(\boldsymbol{\theta}) - \mathbb{E}\left(\nabla \widehat{V}_j(\boldsymbol{\theta})\right)\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right)\right)$$

The proof of this is similar to the proof of Lemma 1.1.3 and is shown in APPENDIX.

Proof. For any $\boldsymbol{a} \in \mathbb{R}^p$, we let $W_{ni} = \boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$. For each value of n, $w_{n1}, w_{n2}, \cdots, w_{nn}$ are i.i.d, and functions of the sample size n. This is because that $\boldsymbol{X}_i$ are assumed to be i.i.d., and h is a function of sample size n. Then, we have

$$\mu_n := \mathbb{E}W_{ni} = \mathbb{E}\left(\boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})\right),$$

and

$$\sigma_n^2 := V(W_{ni}) = \boldsymbol{a}^\intercal V \left(\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) \right) \boldsymbol{a}$$

We let $G_{ni} = W_{ni} - \mu_n$, and $T_n = \sum_{i=1}^n G_{ni}$. Also, we let $s_n^2 = V(T_n) = \sum_{i=1}^n V(G_{ni}) = \sum_{i=1}^n \sigma_n^2 = n\sigma_n^2$, where the second equality is because of independence, and the last equality is due to identicalness. Therefore, T_n/s_n has mean 0, and variance 1. If we can show G_{ni} satisfying the Lyapunov condition, then we have

$$\frac{T_n}{s_n} \xrightarrow{d} \mathcal{N}(0,1), \text{ as } n \to \infty$$

,

Now, we check the Lyapunov condition, that is, [4, 16]

$$\exists \delta > 0, \text{ such that } \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} \mid G_{n,i} \mid^{2+\delta} \to 0, \text{ as } n \to 0.$$

We define, for any $\boldsymbol{a}$,

$$C_1 \triangleq \mathbb{E} \left| G_{ni} \right|^{2+\delta} = \mathbb{E} \left| W_{ni} - \mu_n \right|^{2+\delta} = \mathbb{E} \left| \boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) - \mu_n \right|^{2+\delta},$$

and

$$C_2 \triangleq s_n^{2+\delta} = n^{1+\frac{\delta}{2}} \sigma_n^{2+\delta} = n^{1+\frac{\delta}{2}} \left\{ \boldsymbol{a}^\intercal V \left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) \right] \boldsymbol{a} \right\}^{1+\frac{\delta}{2}}.$$

Then, we have

$$\begin{aligned}
&\frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} \mid G_{n,i} \mid^{2+\delta} \\
=&\frac{\mathbb{E} \left| \boldsymbol{a}^\intercal \frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i}) - \mu_n \right|^{2+\delta}}{n^{\frac{\delta}{2}} \left\{ \boldsymbol{a}^\intercal V \left[\frac{\partial}{\partial \boldsymbol{\theta}} \int y \, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})} \left(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i} \right) \right] \boldsymbol{a} \right\}^{1+\frac{\delta}{2}}} \\
=&\frac{C_1}{n^{\frac{\delta}{2}} C_2}.
\end{aligned}$$

As long as $\delta > 0$, for finite C_1 and finite C_2, we have ${}^{C_1}/_{n^{\frac{\delta}{2}}C_2} \to 0$, as $n \to \infty$. This means that the Lyapunov condition is satisfied, if $\mathbb{E}\left|G_{ni}\right|^{2+\delta}$ and $s_n^{2+\delta}$ are finite. Then, by Lyapunov Central Limit Theorem, we have

$$\frac{T_n}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

As this hold for any arbitary non-random vector $\boldsymbol{a} \in \mathbb{R}^p$, we have, by Cramer-Wold Theorem, that

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i}) - \mathbb{E}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right\}\right] \xrightarrow{d} \mathcal{N}\left(0, V\left[\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right]\right),$$

as $n \to \infty$. We denote $\boldsymbol{L}_{ni} = \frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})$, then this is written as

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}\right] \xrightarrow{d} \mathcal{N}\left(0, V\left[\boldsymbol{L}_{n1}\right]\right).$$

Then, we have

$$\frac{1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}}{\left[V(\boldsymbol{L}_{n1})/n\right]^{1/2}}\frac{\left[V(\boldsymbol{L}_{n1})/n\right]^{1/2}}{\left[AV(\boldsymbol{L}_{n1})/n\right]^{1/2}} \xrightarrow{d} \mathcal{N}(0,1).$$

As $n \to \infty$,

$$\frac{V(\boldsymbol{L}_{n1})^{1/2}}{AV(\boldsymbol{L}_{n1})^{1/2}} \to 1,$$

then we have

$$\frac{1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}}{\left[AV(\boldsymbol{L}_{n1})/n\right]^{1/2}} \xrightarrow{d} \mathcal{N}(0,1),$$

i.e.,

$$\sqrt{n}\left[1/n\sum_{i=1}^{n}\boldsymbol{L}_{ni} - \mathbb{E}\boldsymbol{L}_{n1}\right] \xrightarrow{d} N\left(0, AV(\boldsymbol{L}_{n1})\right).$$

As $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{L}_{ni} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i}) = \nabla\widehat{V}_j(\boldsymbol{\theta})$, we have

$$\sqrt{n}\left[\nabla\widehat{V}_j(\boldsymbol{\theta}) - \mathbb{E}\left\{\nabla\widehat{V}_j(\boldsymbol{\theta})\right\}\right] \xrightarrow{d} \mathcal{N}\left(0, AV\left[\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right]\right)$$

■

B.2 Proof of Corollary 2.1.2

Corollary B.2.1. *Suppose all the assumptions in Lemma 3 hold, and* $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ *is a consistent estimator of* $F_{Y_j^*(\boldsymbol{\theta})}(y_j|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})$*. Then, we have*

$$\sqrt{n}\left(\nabla\widehat{V}_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)) - \nabla V_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu))\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)}\right)\right)$$

Proof. We write

$$\begin{aligned}&\nabla\widehat{V}_j(\boldsymbol{\theta}) - \nabla V_j^*(\boldsymbol{\theta})\\ =&\nabla\widehat{V}_j(\boldsymbol{\theta}) - \mathbb{E}\left(\nabla\widehat{V}_j(\boldsymbol{\theta})\right) + \mathbb{E}\left(\nabla\widehat{V}_j(\boldsymbol{\theta})\right) - \nabla V_j^*(\boldsymbol{\theta}),\end{aligned}$$

where $\mathbb{E}\left(\nabla\widehat{V}_j(\boldsymbol{\theta})\right) - \nabla V_j^*(\boldsymbol{\theta}) = \mathbb{E}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right) -$
$\mathbb{E}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, dF_{Y_j^* x(\boldsymbol{\theta})}(y_j|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i}\right) = o_p(1)$, due to the consistency of $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i} = \boldsymbol{h}_{1,i})$ and dominated convergence theorem.

In lemma 2.1.1, let $\boldsymbol{\theta} = \boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)$ and then

$$\sqrt{n}\left(\nabla\widehat{V}_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)) - \nabla\mathbb{E}\left(\widehat{V}_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu))\right)\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)}\right)\right).$$

As $\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})$ is consistent, we have

$$\frac{AV\left[\frac{\partial}{\partial\boldsymbol{\theta}}\int y\, d\widehat{F}_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right]}{AV\left[\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, dF_{Y^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\right]} \xrightarrow{p} 1.$$

Then, we have

$$\sqrt{n}\left(\nabla\widehat{V}_j(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)) - \nabla V_j\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right) \xrightarrow{d} \mathcal{N}\left(0, AV\left(\frac{\partial}{\partial\boldsymbol{\theta}}\int y_j\, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i}=\boldsymbol{h}_{1,i})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)}\right)\right).$$

■

B.3 Proof of Theorem 2.1.3

Theorem B.3.1. *Suppose all the assumptions above hold. Then we have, as* $n \to \infty$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\nu}}(\mu) - \boldsymbol{\theta}_{\boldsymbol{\nu}}(\mu)^*\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right),$$

where $\boldsymbol{\Sigma}^* = \boldsymbol{D}^{*-1}\boldsymbol{C}^*\boldsymbol{D}^{*-1}$,
$\boldsymbol{C}^* = \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\nabla^\intercal v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right) - \mathbb{E}\left(\nabla v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right)\mathbb{E}\left(\nabla^\intercal v_1\left(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu)\right)\right)$,
and $\boldsymbol{D}^* = \nabla^2\phi^{BP}_\mu(\boldsymbol{\theta}^*_{\boldsymbol{\nu}}(\mu))$.

Proof. For notation simplicity in this proof, let $\phi\left(\boldsymbol{\theta}\right) = \phi^{PB}_\mu\left(\boldsymbol{\theta}\right)$ and $\widehat{\phi}(\boldsymbol{\theta}) = \widehat{\phi}^{PB}_\mu\left(\boldsymbol{\theta}\right)$ for this proof. Also, let $\boldsymbol{\theta}^* = \boldsymbol{\theta}^*_\nu(\mu)$ and $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_\nu(\mu)$ here. Recall $\widehat{\phi}(\boldsymbol{\theta}) = \widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^J \ln\widehat{v}_j(\boldsymbol{\theta}) + \frac{1}{2\mu}\sum_{t=1}^T(\boldsymbol{\theta}_t^\intercal\boldsymbol{\theta}_t - 1)^2$. As $\boldsymbol{\theta}_t^\intercal\boldsymbol{\theta}_t - 1 = 0$ is always satisfied as a constraint, the gradient is $\nabla\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^J {}^{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}\!/_{\widehat{v}_j\left(\boldsymbol{\theta}\right)}$. Taylor expansion of $\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right)$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ shows that

$$\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = \nabla\widehat{\phi}(\widehat{\boldsymbol{\theta}}) - \nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + o_p(1),$$

where $\tilde{\boldsymbol{\theta}}$ is between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$. As $\widehat{\boldsymbol{\theta}}$ is the maximizer of $\widehat{\phi}\left(\boldsymbol{\theta}\right)$, it satisfies the first order condition that $\nabla\widehat{\phi}(\widehat{\boldsymbol{\theta}}) = 0$. Therefore,

$$\sqrt{n}\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = -\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \tag{B.1}$$

where $\nabla\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^J {}^{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}\!/_{\widehat{v}_j\left(\boldsymbol{\theta}\right)}$. Recall $v_1\left(\boldsymbol{\theta}\right) = -V_1\left(\boldsymbol{\theta}\right)$ and $v_j\left(\boldsymbol{\theta}\right) = V_j\left(\boldsymbol{\theta}\right) - \nu_j$, for $j = 2, \cdots, J$. Due to Corollary 2.1.2, together with (A.4) and (A.5),

$$\sqrt{n}\Big(\nabla\widehat{v}_1(\boldsymbol{\theta}^*) - \nabla v_1(\boldsymbol{\theta}^*)\Big) \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right), \tag{B.2}$$

where $\boldsymbol{C}^* = AV\left(\nabla v_1(\boldsymbol{\theta}^*)\right) = \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\nabla v_1(\boldsymbol{\theta}^*)\mathbb{E}\nabla^\intercal v_1(\boldsymbol{\theta}^*)$
$= AV\left(\frac{\partial}{\partial \boldsymbol{\theta}}\int y\, dF_{Y_j^*(\boldsymbol{\theta})}(y|\boldsymbol{H}_{1,i})\right)$. That is, Then, due to (B.1) and (B.2), we have

$$\sum_{j=2}^{J}\frac{\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)}{\widehat{v}_j\left(\boldsymbol{\theta}\right)} - \sum_{i=2}^{J}\frac{\nabla v_j(\boldsymbol{\theta})}{v_j(\boldsymbol{\theta})} = o_p(1). \tag{B.3}$$

Note $v_j(\boldsymbol{\theta}) > 0$, for $j = 2, \cdots, J$, is implied by the log barrier operator. Put (B.2) and (B.3) together by Slutsky's theorem, we have

$$\sqrt{n}\left\{\left(\nabla\widehat{v}_1(\boldsymbol{\theta}^*) - \mu\sum_{j=2}^{J}\frac{\nabla\widehat{v}_j\left(\boldsymbol{\theta}^*\right)}{\widehat{v}_j(\boldsymbol{\theta}^*)}\right) - \left(\nabla v_1(\boldsymbol{\theta}^*) - \mu\sum_{i=2}^{J}\frac{\nabla v_j(\boldsymbol{\theta}^*)}{v_j(\boldsymbol{\theta}^*)}\right)\right\} \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right),$$

Due to the stationarity of $\boldsymbol{\theta}^*$, $\nabla\phi(\boldsymbol{\theta}^*) = \nabla v_1(\boldsymbol{\theta}^*) - \mu\sum_{i=2}^{J}{}^{\nabla v_j(\boldsymbol{\theta}^*)}/_{v_j(\boldsymbol{\theta}^*)} = 0$. Together with Sluskty's theorem, we have

$$\sqrt{n}\nabla\widehat{\phi}(\boldsymbol{\theta}^*) \xrightarrow{d} N\left(0, \boldsymbol{C}^*\right),$$

where $\boldsymbol{C}^* = \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\} - \mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^*)\right\}\mathbb{E}\left\{\nabla^\intercal v_1(\boldsymbol{\theta}^*)\right\}$.

As $\sqrt{n}\nabla\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = -\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ stated in (A.7),we have

$$\sqrt{n}\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N(0, \boldsymbol{C}^*) \tag{B.4}$$

The Hessian is $\nabla^2\widehat{\phi}\left(\boldsymbol{\theta}\right) = \nabla^2\widehat{v}_1(\boldsymbol{\theta}) - \mu\sum_{j=2}^{J}\left(\nabla^2\widehat{v}_j\left(\boldsymbol{\theta}\right)\widehat{v}_j\left(\boldsymbol{\theta}\right) - \left(\nabla\widehat{v}_j\left(\boldsymbol{\theta}\right)\right)^2\right)/\widehat{v}_j^2\left(\boldsymbol{\theta}\right)$. Based on (A.4) and (A.5), we have

$$\boldsymbol{D}^* \triangleq p\lim_{n\to\infty}\nabla^2\widehat{\phi}\left(\boldsymbol{\theta}^*\right) = \nabla^2\phi(\boldsymbol{\theta}^*) = \nabla^2 v_1(\boldsymbol{\theta}^*) - \mu\sum_{j=2}^{J}\frac{\nabla^2 v_j\left(\boldsymbol{\theta}^*\right)v_j\left(\boldsymbol{\theta}^*\right) - \left\{\nabla v_j\left(\boldsymbol{\theta}^*\right)\right\}^2}{v_j^2\left(\boldsymbol{\theta}^*\right)}. \tag{B.5}$$

As $\tilde{\boldsymbol{\theta}}$ is a vector in-between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}$, we have $\nabla^2\widehat{\phi}(\tilde{\boldsymbol{\theta}}) = \nabla^2\widehat{\phi}(\boldsymbol{\theta}^*) + o_p(1)$. Therefore,

based on (A.10) and (A.11), we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{*}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}^{*}\right),$$

where $\boldsymbol{\Sigma}^{*}=\boldsymbol{D}^{*-1}\boldsymbol{C}^{*}\boldsymbol{D}^{*-1}$, $\boldsymbol{C}^{*}=\mathbb{E}\left\{\nabla v_1(\boldsymbol{\theta}^{*})\nabla^{\intercal} v_1(\boldsymbol{\theta}^{*})\right\}-\mathbb{E}\nabla v_1(\boldsymbol{\theta}^{*})\mathbb{E}\nabla^{\intercal} v_1(\boldsymbol{\theta}^{*})$ and $\boldsymbol{D}^{*}=\nabla^{2}\phi(\boldsymbol{\theta}^{*})$. ■