Statistica Sinica Preprint No: SS-2025-0168	
Title	Doubly Robust Estimation of Optimal Individual Treatment Regime in A Semi-supervised Framework
Manuscript ID	SS-2025-0168
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0168
Complete List of Authors	Xintong Li, Mengjiao Peng and Yong Zhou
Corresponding Authors	Mengjiao Peng
E-mails	mjpeng@fem.ecnu.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Statistica Sinica

DOUBLY ROBUST ESTIMATION OF OPTIMAL INDIVIDUAL TREATMENT REGIME IN A SEMI-SUPERVISED FRAMEWORK

Xintong Li, Mengjiao Peng and Yong Zhou

East China Normal University

Abstract: In many health-care datasets like the electronic health record (EHR) dataset, collecting labeled data can be a laborious and expensive task, resulting in a scarcity of labeled data while unlabeled data is already available. This has sparked a growing interest in developing methods to leverage the abundant unlabeled data. We thus develop several types of semi-supervised (SS) methods for estimating optimal individulized treatment regime (ITR) that utilize both labeled and unlabeled data in a general model-free framework, with efficiency gains compared to supervised estimation methods. Our proposed method first utilizes a flexible imputation technique through single index kernel smoothing to exploit the unlabeled data, which performs well even in cases of multidimensional covariates, with a follow-up estimation to determine the optimal ITR by directly optimizing the imputed value function. Additionally, in cases where the propensity score function is unknown like in observational studies, we also develop a doubly robust SS estimation method based on a class of monotonic index models. Our estimators are shown to be consistent with the cube root convergence

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

rate and exhibit a nonstandard asymptotic distribution characterized as the maximizer of a centered Gaussian process with a quadratic drift. Simulation studies demonstrate the efficiency and robustness of the proposed methods compared to supervised approach in finite samples. Additionally, a practical example from the ACTG 175 study illustrates its real-world application.

Key words and phrases: Optimal treatment regime, Semi-supervised inference, Doubly robustness, Precision medicine.

1. Introduction

Precision medicine has emerged as a promising field, aiming to provide tailored medical treatments for individual patients based on their unique characteristics. One main purpose of precision medicine is to find the optimal individualized treatment regime (ITR) mapping from the individual characteristics or contextual information to the treatment assignment, that maximizes the expected outcome, known as the value function (Manski, 2004; Qian and Murphy, 2011). ITRs apply to a wide range of areas, including disease management, recommender systems and public policy evaluation. In disease management, the physician needs to decide the optimizing drug dosages based on patients' characteristics in order to optimize his/her clinical outcome (Correa et al., 2024). In a context-based recommender system, contextual information such as time, location, and social connection can be

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

incorporated to increase the effectiveness of the recommendation (Aggarwal, 2016). In the realm of public management, the process of policy learning and evaluation encounters challenges posed by clustered network interference. Addressing influential individuals with extensive social connections can yield positive spillover effects, ultimately enhancing the overall efficacy (Zhang and Imai, 2023). Several methods have been developed for estimating the optimal individualized treatment regime, which can be broadly classified into two main strategies: the model-based approach which estimates the mean outcome model given treatment and covariates, and the direct-search approach that non-parametrically estimates the value function and maximizes the estimated value function over a pre-specified ITRs class. Methods falling into the model-based approach category include Q-Learning (posits regression models for the outcome of interest, e.g., Watkins, 1989; Watkins and Dayan, 1992; Chakraborty et al., 2010; Qian and Murphy, 2011; Wang et al., 2018), and A-Learning (builds models for the contrast functions, e.g., Robins et al., 2000; Murphy, 2003; Robins, 2004; Blatt et al., 2004; Lu et al., 2013; Shi et al., 2018). The direct-search approach involves inverse-probability weighted estimation (IPWE) (Kitagawa and Tetenov, 2018; Liu et al., 2018; Zhao et al., 2012; Zhou et al., 2017). While most of the model-based approach relies on correctly specified outcome models,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

the direct-search approach based on IPWE necessitates accurate estimation of the propensity score (PS) function. The concept of double robustness is fundamental in the field of causal inference (Ding and Li, 2018; Robins et al., 1994, 1995), especially concerning the impact of model misspecification on estimation results. Various approaches have been proposed to combine the strengths of the model-based and direct-search methods and therefore enhance the robustness of the estimation process. Notably, Zhang et al. (2012), Zhao et al. (2019), and Athey and Wager (2021), augmented IPWE with the outcome model to obtain the augmented IPWE (AIPWE) of the value function, which could be robust even if the outcome model or PS model is misspecified.

Recently, large unlabeled datasets generated electronically are becoming increasingly accessible, but few studies have investigated safe and effective ways to leverage this wealth of abundant auxiliary information. In biomedical applications, for example, electronic medical record (EMR) data often remain underutilized due to difficulties in obtaining accurate clinical data (Liao et al., 2010). To address this, semi-supervised learning (SSL) has attracted significant attention. In traditional SSL, more information from the distribution of covariates $\mathbf{X}$, obtained from unlabeled individuals, is utilized to enhance the inference of the conditional distribution of out-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

come Y given $\mathbf{X}$ (Chapelle et al., 2006; Chakrabortty and Cai, 2018; Song et al., 2023). Although SSL methods have begun to be used in the field of precision medicine such as estimating treatment effects (Zhang et al., 2019; Cheng et al., 2021; Chakrabortty et al., 2022), there is still limited literature on using these methods to estimate optimal treatment rules. Sonabend-W et al. (2023) introduced a semi-supervised off-policy reinforcement learning framework for optimizing and evaluating dynamic treatment regimes. Their proposed SSL estimator enhances efficiency by leveraging both labeled and unlabeled data, along with outcome surrogates, to estimate the value function. They constructed a doubly robust value function estimator based on AIPW, which ensures consistency if either the Q-function or the PS is correctly specified. Gunn et al. (2024) used covariate information from unlabeled data to estimate the contrast function, which improves the estimation of the linear decision rule based on a semiparametric working model. Their method directly imputes the contrast function using kernel methods with the covariate vector $\mathbf{X}$, which may be slow in multidimensional cases.

This paper aims to develop efficient and robust estimators for determining optimal ITR using semi-supervised (SS) techniques within a model-free framework. We propose a flexible imputation approach based on single-index kernel smoothing, which performs well even with multidimensional

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

covariates. The optimal ITR is then estimated by directly optimizing the value function. Additionally, we introduce a doubly robust estimation method for cases where the PS function is unknown. The remainder of the paper is organized as follows. We introduce the data and notations in Section 2.1 and formally describe the proposed SS estimators with known and unknown PS in Section 2.2 and 2.3, respectively. Asymptotic properties of the proposed estimator are provided in Section 3. To facilitate inference, a perturbation resampling procedure is proposed in Section 4 for inference. Section 5 presents simulation results showing the robustness and efficiency of the proposed estimators, followed by an application to AIDS clinical trial data in Section 6. Some concluding discussions and extensions are given in Section 7. Theoretical proofs and additional numerical results can be found in the Supplementary Material.

2. Methodology

2.1 Notations and Data representation

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}$ be the outcome variable which is assumed that a larger value of Y implies a better response without loss of generality. Denote $\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ as the p-dimensional predictor vector with bounded support $\mathcal{X}$. Let A, taking values in $\mathcal{A} = \{0, 1\}$, be the treatment indica-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

tor. As in traditional SS framework (Chapelle et al., 2006), the available data consists of two independent data sources $\mathcal{L}$ and $\mathcal{U}$, where $\mathcal{L} = \{(Y_i, \mathbf{X}_i, A_i) : i = 1, 2, \ldots, n\}$ consist n iid labeled observations and $\mathcal{U} = \{(\mathbf{X}_j, A_j) : j = n+1, n+2, \ldots, n+N\}$ consist N iid unlabeled observations. Assuming that the observations in both $\mathcal{L}$ and $\mathcal{U}$ follow the same potential distribution and for some constant $\rho \in [0, \infty)$, $\rho_n = n/N \to \rho$ as $n, N \to \infty$. And assuming that observations in $\mathcal{L}$ were randomly selected from $\mathcal{L} \cup \mathcal{U}$ for labeling so that Y is essentially missing completely at random (MCAR) (Chakrabortty and Cai, 2018). Note that the major difference between the SS framework and the MCAR assumption is that the SS setting allows $n/N \to 0$ (Song et al., 2023) while the latter may require $n/N \to c$ for some $c > 0$.

Let $Y^*(a)$ denote the potential outcome that would result if the subject were given treatment $a \in \mathcal{A}$ (Rubin, 1974). Let '$\perp$' represent independence. Three identification assumptions are typically made in potential outcome framework: (A1) SUTVA: $Y = Y^*(1)A + Y^*(0)(1-A)$; (A2) Strong ignorability: $A \perp \{Y^*(0), Y^*(1)\} | \mathbf{X}$; (A3) Positivity: $0 < P(A = 1|\mathbf{X}) < 1$.

The ITR $d(\mathbf{X})$ is defined as a decision function that maps $\mathbf{X} \in \mathcal{X}$ to $a \in \mathcal{A}$. For any ITR $d(\mathbf{X})$, the potential outcome $Y^*(d(\mathbf{X}))$ can be written by $Y^*(d(\mathbf{X})) = Y^*(1)d(\mathbf{X}) + Y^*(0)\{1 - d(\mathbf{X})\}$. Then, the outcome $Y^*(d(\mathbf{X}))$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

would be observed if a randomly chosen subject from the population were to be assigned treatment according to ITR $d(\mathbf{X})$. The optimal ITR is defined as $d^{\text{opt}}(\mathbf{X}) = \arg\max_{d\in\mathcal{D}} E\left[Y^*\{d(\mathbf{X})\}\right]$, where $\mathcal{D}$ is some decision class contains all possible ITRs of interest and $E\left[Y^*\{d(\mathbf{X})\}\right]$ is called the value function of a given ITR $d(\mathbf{X})$. For simplicity and interpretability, we will focus on the linear decision class $\mathcal{D} = \{d_{\boldsymbol{\beta}}(\mathbf{X}) = I(\boldsymbol{\beta}'\mathbf{X} \geqslant 0) : \boldsymbol{\beta} \in \mathcal{B}\}$ where $\mathcal{B} = \{\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^p, \|\boldsymbol{\beta}\| = 1\}$. We assume that $\|\boldsymbol{\beta}\| = 1$ for identifiability and $\|\mathbf{a}\|$ represents the Euclidean norm of a vector $\mathbf{a}$.

Remark 1. The linear ITR with intercept term that $d_{\tilde{\boldsymbol{\beta}}} = I\left(\tilde{\boldsymbol{\beta}}'\tilde{\mathbf{X}} \geqslant c_0\right)$ are considered as in literature (Fan et al., 2017; Chu et al., 2023), which is equavilent to $d_{\boldsymbol{\beta}} = I\left(\boldsymbol{\beta}'\mathbf{X} \geqslant 0\right)$ with $\boldsymbol{\beta} = (c_0, \tilde{\boldsymbol{\beta}}')'$ and $\mathbf{X} = (1, \tilde{\mathbf{X}}')'$. Thus, we assume the intercept term is contained in the covariate $\mathbf{X}$ throughout this paper for notation simplicity without loss of generality.

2.2 Supervised and semi-supervised estimation

In this section, we establish the main framework of our SS methodology when the PS is known. This methodological framework can be easily extended to the case of unknown PS in observational studies, and the details will be discussed in the following section 2.3. Denote the conditional average treatment effect (CATE) as $D(\mathbf{X}) = E(Y|\mathbf{X}, A = 1) - E(Y|\mathbf{X}, A = 0)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

To illustrate our basic idea, we begin with a lemma.

Lemma 1. *Under identification assumptions made in Section 2.1, we have*

$$E[\{Y^*(1) - Y^*(0)\}d_{\boldsymbol{\beta}}(\mathbf{X})] = E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})].$$

From Lemma 1 and the definition of optimal ITR in Section 2.1 that $d^{\text{opt}}_{\boldsymbol{\beta}} = \arg\max_{d_{\boldsymbol{\beta}}\in\mathcal{D}} E[Y^*(d_{\boldsymbol{\beta}}(\mathbf{X}))] = \arg\max_{d_{\boldsymbol{\beta}}\in\mathcal{D}} E[\{Y^*(1)-Y^*(0)\}d_{\boldsymbol{\beta}}(\mathbf{X})]$, we have $d^{\text{opt}}_{\boldsymbol{\beta}} = \arg\max_{d_{\boldsymbol{\beta}}\in\mathcal{D}} E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})]$. Let $E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})]$ be the value function and $\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta}\in\mathcal{B}} E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})]$, then the induced optimal ITR in $\mathcal{D}$ is $I(\boldsymbol{\beta}_0'\mathbf{X} \geqslant 0)$.

Combining the ideas of the direct search method from Zhang et al. (2012) and robust A-learning (Murphy, 2003), we can construct the consistent estimator of the value function as follows:

$$E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})] = E[V(\mathbf{Z},\boldsymbol{\theta})d_{\boldsymbol{\beta}}(\mathbf{X})] := \Delta(\boldsymbol{\beta},\boldsymbol{\theta}),$$

where $\mathbf{Z} = (\mathbf{X}, Y, A)$, $V(\mathbf{Z},\boldsymbol{\theta}) = \frac{\{Y-\nu(\mathbf{X},\boldsymbol{\theta})\}\{A-\pi(\mathbf{X})\}}{\pi(\mathbf{X})\{1-\pi(\mathbf{X})\}}$ satisfies $E[V(\mathbf{Z},\boldsymbol{\theta})|\mathbf{X}] = D(\mathbf{X})$ (Fan et al., 2017), $\pi(\mathbf{X}) = P(A = 1|\mathbf{X})$ is the PS function, and $\nu(\mathbf{X},\boldsymbol{\theta})$ is a model parameterized by $\boldsymbol{\theta}$ for $\nu(\mathbf{X})$, an arbitrary function of $\mathbf{X}$. Let $\hat{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}$, such as the least squares estimator for the linear model $\nu(\mathbf{X},\boldsymbol{\theta}) = \boldsymbol{\theta}'\mathbf{X}$. As a prelude to Section 2.3, we assume that $\nu(\mathbf{X})$ is the baseline treatment-free effect $\mu_0(\mathbf{X}) = E(Y|\mathbf{X}, A = 0)$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

without loss of generality. Thus we can obtain the supervised estimator of the value function based only on $\mathcal{L}$ that

$$\hat{\Delta}_{\text{sup}}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^{n} V(\mathbf{Z}_i, \hat{\boldsymbol{\theta}}) I\left(\boldsymbol{\beta}' \mathbf{X}_i \geqslant 0\right),$$

and the supervised estimator of the optimal ITR parameter $\boldsymbol{\beta}_0$ is then obtained by $\hat{\boldsymbol{\beta}}_{\text{sup}} = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \hat{\Delta}_{\text{sup}}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$.

Due to the absence of labels in the unlabeled data $\mathcal{U}$, we cannot estimate $V(\mathbf{Z}, \boldsymbol{\theta})$ directly. Therefore, to leverage information from the unlabeled data, we consider imputing the value function estimation by appropriately training on $\mathcal{L}$. For the linear ITR $d_{\boldsymbol{\beta}}(\mathbf{X}) = I(\boldsymbol{\beta}' \mathbf{X} \geqslant 0)$, we have

$$\Delta(\boldsymbol{\beta}, \boldsymbol{\theta}) = E[E[V(\mathbf{Z}, \boldsymbol{\theta}) | \boldsymbol{\beta}' \mathbf{X}] d_{\boldsymbol{\beta}}(\mathbf{X})]$$

by the law of iterated expectations. Let $m(\boldsymbol{\beta}' \mathbf{X}, \boldsymbol{\theta}) = E[V(\mathbf{Z}, \boldsymbol{\theta}) | \boldsymbol{\beta}' \mathbf{X}]$. Notably, we implicitly utilize a single-index projection for dimensionality reduction, which alleviates the 'curse of dimensionality' problem when estimating $m(\boldsymbol{\beta}' \mathbf{X}, \boldsymbol{\theta})$ using nonparametric methods. The accuracy of imputation estimation is crucial for the effectiveness of SS methods based on imputation (Chakrabortty and Cai, 2018; Wang et al., 2023). Therefore, we employ a nonparametric kernel smoothing method to estimate $m(\boldsymbol{\beta}' \mathbf{X}, \boldsymbol{\theta})$ to avoid model misspecification, that is

$$\hat{m}(\boldsymbol{\beta}' \mathbf{X}_j, \boldsymbol{\theta}) = \frac{n^{-1} \sum_{i=1}^{n} K_h(\boldsymbol{\beta}' \mathbf{X}_i - \boldsymbol{\beta}' \mathbf{X}_j) V(\mathbf{Z}_i, \boldsymbol{\theta})}{n^{-1} \sum_{i=1}^{n} K_h(\boldsymbol{\beta}' \mathbf{X}_i - \boldsymbol{\beta}' \mathbf{X}_j)}, \tag{2.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

where $K_h(u-v) = \frac{1}{h}K\left(\frac{u-v}{h}\right)$ with K: $\mathbb{R} \to \mathbb{R}$ being some suitable kernel function and $h = h(n) > 0$ being the bandwidth.

Next, we establish our SS method based on the above imputation estimation. We introduce a weight parameter $\lambda \in [0,1]$ to balance the contributions of labeled and unlabeled data in value function estimation. Specifically, since $\Delta(\boldsymbol{\beta}, \boldsymbol{\theta}) = \lambda E[V(\mathbf{Z}, \boldsymbol{\theta}) d_{\boldsymbol{\beta}}(\mathbf{X})] + (1-\lambda) E[m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}) d_{\boldsymbol{\beta}}(\mathbf{X})]$, we construct the following SS estimator of the value function that

$$\hat{\Delta}_\lambda(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = \frac{\lambda}{n}\sum_{i=1}^{n} V(\mathbf{Z}_i, \hat{\boldsymbol{\theta}}) I(\boldsymbol{\beta}'\mathbf{X}_i \geqslant 0) + \frac{1-\lambda}{N}\sum_{j=n+1}^{n+N} \hat{m}(\boldsymbol{\beta}'\mathbf{X}_j, \hat{\boldsymbol{\theta}}) I(\boldsymbol{\beta}'\mathbf{X}_j \geqslant 0),$$

and the corresponding SS estimator of the optimal ITR parameter $\boldsymbol{\beta}_0$ is given by $\hat{\boldsymbol{\beta}}_\lambda = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \hat{\Delta}_\lambda(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$. Section 3 establishes $\lambda = \frac{\rho^2}{1+\rho^2}$ as the optimal weight, which we recommend for applications.

Furthermore, based on the kernel estimator of the imputation function, we propose the following pooled estimator of the value function that

$$\hat{\Delta}_{\text{pl}}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = \frac{1}{n+N}\sum_{j=1}^{n+N} \hat{m}(\boldsymbol{\beta}'\mathbf{X}_j, \hat{\boldsymbol{\theta}}) I\left(\boldsymbol{\beta}'\mathbf{X}_j \geqslant 0\right),$$

and the pooled estimator of the optimal ITR parameter $\boldsymbol{\beta}_0$ is given by $\hat{\boldsymbol{\beta}}_{\text{pl}} = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \hat{\Delta}_{\text{pl}}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$.

Notably, since we do not require a specific model for $V(\mathbf{Z}, \boldsymbol{\theta})$, the proposed framework for learning the optimal ITR is model-free. Unlike the method proposed by Gunn et al. (2024), which is only applicable to low-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

dimensional covariates, our method utilizes projection-based dimensionality reduction in imputation estimation. This allows us to employ a one-dimensional kernel function, avoiding the 'curse of dimensionality' issue when p is large.

Remark 2. Since the objective functions $\hat{\Delta}_\lambda(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ and $\hat{\Delta}_{\text{pl}}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ involve kernel estimators with an unknown parameter, we simplify the optimization procedure by employing a two-step iterative algorithm for implementation. See Section S4.1 of the Supplementary Material for algorithm details.

2.3 Doubly robust estimation with unknown propensity score

Section 2.2 clearly establishes a framework of SS estimation methods when $\pi(\mathbf{X})$ is known. However, in practice, when data come from observational studies, $\pi(\mathbf{X})$ is typically unknown and therefore needs to be estimated, using parametric methods, such as logistic regression (Mo and Liu, 2022), or flexible nonparametric methods, such as regression forests (Athey et al., 2019). Let $\pi(\mathbf{X}, \boldsymbol{\alpha})$ denote the model posited for $\pi(\mathbf{X})$ with parameter $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ has a compact support $\boldsymbol{\alpha}_{\text{supp}}$. Let $\hat{\boldsymbol{\alpha}}$ be the estimate of $\boldsymbol{\alpha}$ obtained based on either $\mathcal{L}$ or $\mathcal{L} \cup \mathcal{U}$. In this section, building on the SS framework from the Section 2.2, we similarly construct a doubly robust value function estimation method. This method guarantees the consistency

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

of the value function estimation when either the PS model $\pi(\mathbf{X}, \boldsymbol{\alpha})$ or the baseline treatment-free effect model $\nu(\mathbf{X}, \boldsymbol{\theta})$ is correctly specified.

Next, we will elaborate on the doubly robust value function estimation method. Define $V(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \frac{\{Y-\nu(\mathbf{X},\boldsymbol{\theta})\}\{A-\pi(\mathbf{X},\boldsymbol{\alpha})\}}{\pi(\mathbf{X},\boldsymbol{\alpha})\{1-\pi(\mathbf{X},\boldsymbol{\alpha})\}}$ and the imputation function $m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = E[V(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha})|\boldsymbol{\beta}'\mathbf{X}]$. When $\pi(\mathbf{X}, \boldsymbol{\alpha})$ is correctly specified, it is similar to the case discussed in Section 2.2. When $\nu(\mathbf{X}, \boldsymbol{\theta})$ is correctly specified, we have

$$m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = E\left[\frac{AD(\mathbf{X})\{A-\pi(\mathbf{X},\boldsymbol{\alpha})\}}{\pi(\mathbf{X},\boldsymbol{\alpha})\{1-\pi(\mathbf{X},\boldsymbol{\alpha})\}}\middle|\boldsymbol{\beta}'\mathbf{X}\right] = E\left[D(\mathbf{X})\frac{\pi(\mathbf{X})}{\pi(\mathbf{X},\boldsymbol{\alpha})}\middle|\boldsymbol{\beta}'\mathbf{X}\right].$$

It is worth noting that when $D(\mathbf{X})$ is a monotonic increasing index model, for any positive function $g(\cdot)$, we have

$$d^{\text{opt}}_{\boldsymbol{\beta}} = \arg\max_{d_{\boldsymbol{\beta}}\in\mathcal{D}} E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})] = \arg\max_{d_{\boldsymbol{\beta}}\in\mathcal{D}} E[D(\mathbf{X})g(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})].$$

Due to the positivity assumption, $\frac{\pi(\mathbf{X})}{\pi(\mathbf{X},\boldsymbol{\alpha})}$ is always a positive function, hence

$$d^{\text{opt}}_{\boldsymbol{\beta}} = \arg\max_{d_{\boldsymbol{\beta}}\in\mathcal{D}} E[D(\mathbf{X})d_{\boldsymbol{\beta}}(\mathbf{X})] = \arg\max_{d_{\boldsymbol{\beta}}\in\mathcal{D}} E[m(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\alpha})d_{\boldsymbol{\beta}}(\mathbf{X})].$$

Similar to Section 2.2, based on $\mathcal{L}$, the doubly robust supervised estimator of the value function can be obtained as

$$\hat{\Delta}^{DR}_{sup}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n} V(\mathbf{Z}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})I\left(\boldsymbol{\beta}'\mathbf{X}_i \geqslant 0\right),$$

and correspondingly, the doubly robust supervised estimate of the optimal ITR parameter is $\hat{\boldsymbol{\beta}}^{DR}_{sup} = \arg\max_{\boldsymbol{\beta}\in\mathcal{B}} \hat{\Delta}^{DR}_{sup}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

The doubly robust semi-supervised estimate of the value function based on weighting is given by $\hat{\Delta}_{\lambda}^{DR}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}) = \frac{\lambda}{n} \sum_{i=1}^{n} V(\mathbf{Z}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}) I(\boldsymbol{\beta}'\mathbf{X}_i \geqslant 0) + \frac{1-\lambda}{N} \sum_{i=n+1}^{M} \hat{m}(\boldsymbol{\beta}'\mathbf{X}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}) I(\boldsymbol{\beta}'\mathbf{X}_i \geqslant 0)$, where $\lambda \in [0,1]$ and $\hat{m}(\boldsymbol{\beta}'\mathbf{x}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}) = \frac{n^{-1}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{x}) V(\mathbf{Z}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})}{n^{-1}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{x})}$. The corresponding estimator of optimal ITR parameter is $\hat{\boldsymbol{\beta}}_{\lambda}^{DR} = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \hat{\Delta}_{\lambda}^{DR}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$.

Furthermore, the doubly robust pooled estimate of the value function is $\hat{\boldsymbol{\beta}}_{pl}^{DR} = \frac{1}{n+N} \sum_{i=1}^{n+N} \hat{m}(\boldsymbol{\beta}'\mathbf{X}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}) I(\boldsymbol{\beta}'\mathbf{X}_i \geqslant 0)$, and the corresponding estimator of the optimal ITR parameter is $\hat{\boldsymbol{\beta}}_{pl}^{DR} = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \hat{\Delta}_{pl}^{DR}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$.

3. Asymptotic Properties

In this section, we will mainly study the asymptotic properties of the parameter estimators we proposed in Section 2.2, and that of the estimators in Section 2.3 can be obtained similarly and will be also given in Section S1.2 of the Supplementary Material. In order to establish asymptotic results, some regularity conditions need to be assumed, which can be found in Section S1.1 of the Supplementary Material.

Theorem 1. *Let $G(t)$, $G_{\lambda}(t)$ and $G_{pl}(t)$ be the mean-zero Gaussian process with continuous sample paths. Under conditions C1-C7, and as $n, N \to \infty$, $\frac{n}{N} = \rho_n \to \rho \in [0, \infty)$, $\lambda \in [0,1]$, we have:*

(a1) $\hat{\boldsymbol{\beta}}_{sup} \xrightarrow{p} \boldsymbol{\beta}_0$. *(a2)* $\hat{\boldsymbol{\beta}}_{pl} \xrightarrow{p} \boldsymbol{\beta}_0$. *(a3)* $\hat{\boldsymbol{\beta}}_{\lambda} \xrightarrow{p} \boldsymbol{\beta}_0$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

(b1) $n^{\frac{1}{3}}(\hat{\boldsymbol{\beta}}_{sup} - \boldsymbol{\beta}_0) \xrightarrow{d} \arg\max_t Z(t)$, *where the process* $Z(t) = G(t) - \frac{1}{2}t'Vt$. *Here* $G(t)$ *has the covariance kernel function* $Cov(\cdot,\cdot)$ *and* $-V$ *is the second derivative matrix of* $E[V(\mathbf{Z}, \boldsymbol{\theta}_0)I(\boldsymbol{\beta}'\mathbf{X} \geqslant 0)]$ *with respect to* $\boldsymbol{\beta}$ *at* $\boldsymbol{\beta}_0$.

(b2) $n^{\frac{1}{3}}(\hat{\boldsymbol{\beta}}_{pl} - \boldsymbol{\beta}_0) \xrightarrow{d} \arg\max_t Z_{pl}(t)$, *where the process* $Z_{pl}(t) = G_{pl}(t) - \frac{1}{2}t'Vt$. *Here* $G_{pl}(t)$ *has the covariance kernel function* $(\frac{\rho}{1+\rho})^2 Cov(\cdot,\cdot)$.

(b3) $n^{\frac{1}{3}}(\hat{\boldsymbol{\beta}}_{\lambda} - \boldsymbol{\beta}_0) \xrightarrow{d} \arg\max_t Z_{\lambda}(t)$, *where the process* $Z_{\lambda}(t) = G_{\lambda}(t) - \frac{1}{2}t'Vt$. *Here* $G_{\lambda}(t)$ *has the covariance kernel function* $[\lambda^2 + (1-\lambda)^2\rho^2]Cov(\cdot,\cdot)$.

Note that $Cov(\cdot,\cdot)$ in this theorem is calculated in the proof of Theorem 1 that $Cov(C_1, C_2) = \frac{1}{2}(L(C_1) + L(C_2) - L(C_1 - C_2))$ for $C_1, C_2 \in \mathbb{R}^p$, where $L(C) := \int |C'\mathbf{v}| q(\mathbf{v}) p(0, \mathbf{v}) d\mathbf{v}$, $q(\mathbf{X}) = E[V^2(\mathbf{Z}, \boldsymbol{\theta}_0)|\mathbf{X}]$, $p(r, \mathbf{v})$ is the joint density function of $(r, \mathbf{v})$, and other specific definitions and proof details can be found in the Supplementary Material. We can see from this theorem that the convergence rates of $\hat{\boldsymbol{\beta}}_{\text{sup}}$, $\hat{\boldsymbol{\beta}}_{\text{pl}}$ and $\hat{\boldsymbol{\beta}}_{\lambda}$ are all the cube root of n, which implies that unlabeled data does not improve the convergence rate of the estimators. Although SS methods may not significantly reduce the bias of the estimator, their asymptotic variance will decrease substantially as the size of the unlabeled data increases. Denote the covariance of $\hat{\boldsymbol{\beta}}_{\text{sup}}$, $\hat{\boldsymbol{\beta}}_{\lambda}$ and $\hat{\boldsymbol{\beta}}_{\text{pl}}$ are Σ_{sup}, Σ_{λ} and Σ_{pl} respectively. Theorem 1 shows that Σ_{λ} is minimized when the weight $\lambda = \frac{\rho^2}{1+\rho^2}$ (See more details in Section S1.3 of the Supplementary Material). Accordingly, in our numerical simulations,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

we set the tuning parameter λ to this optimal value. With this choice, a comparison of Σ_{sup}, Σ_λ, and Σ_{pl} reveals that $\Sigma_{\text{sup}} \succcurlyeq \Sigma_\lambda \succcurlyeq \Sigma_{\text{pl}}$, since $1 \geqslant \frac{\rho^2}{1+\rho^2} \geqslant \frac{\rho^2}{(1+\rho)^2}$ holds for all $\rho \in [0, \infty)$. This implies that our SS estimators, $\hat{\boldsymbol{\beta}}_\lambda$ and $\hat{\boldsymbol{\beta}}_{\text{pl}}$, are more efficient than or at least as effective as $\hat{\boldsymbol{\beta}}_{\text{sup}}$. The reduction in asymptotic variance leads to a significant improvement in efficiency, which we will visually demonstrate through numerical results in Section 5.

4. Variance Estimation

Since the asymptotic variance is challenging to compute directly, we employ a simple resampling approach based on repeatedly perturbing the value function, as proposed in Jin et al. (2001), to estimate the variance of our estimators for inference. Here, we provide a detailed description of the perturbation resampling procedure for estimating Σ_λ. A similar approach can be applied to perturb the corresponding value function for estimating other asymptotic variances, such as Σ_{sup}, Σ_{pl} and $\Sigma_{\cdot}^{DR}$, with the only distinction being the specific form of the value function undergoing perturbation. Other variance estimation perturbation steps are provided in the Supplementary Material. Let ξ_i $(i = 1, \ldots, n)$ be n iid copies of a random variable ξ following a Beta distribution, $Beta(\sqrt{2} - 1, 1)$, which is assumed to be

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

independent of the observed data $\mathcal{L} \cup \mathcal{U}$. Notably, the variance estimation is generally robust to the choice of ξ's distribution (Jin et al., 2001), and alternative choices such as $\Gamma(1,1)$ can also be used (Peng and Huang, 2008; Fan et al., 2017). The resampling procedure is outlined as follows:

1. Generate iid perturbation ξ_i from $Beta(\sqrt{2}-1,1)$ for $i = 1, \ldots, n+N$.

2. Perturb the value function. Let $\hat{\boldsymbol{\theta}}^b = \arg\min_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} \xi_i (1-A_i)[Y_i - \nu(\mathbf{X}_i, \boldsymbol{\theta})]^2$ and $\hat{m}^b(\boldsymbol{\beta}'\mathbf{X}_j, \boldsymbol{\theta}) = \frac{\sum_{i=1}^{n} \xi_i K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{X}_j) V(\mathbf{Z}_i, \boldsymbol{\theta})}{\sum_{i=1}^{n} \xi_i K_h(\boldsymbol{\beta}'\mathbf{X}_i - \boldsymbol{\beta}'\mathbf{X}_j)}$, then for linear decision $d_{\boldsymbol{\beta}}(\mathbf{X}) = I(\boldsymbol{\beta}'\mathbf{X} \geqslant 0)$, we perturb the value function by

$$\hat{\Delta}_\lambda^b\left(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}^b\right) = \frac{\lambda}{n}\sum_{i=1}^{n} \xi_i V\left(\mathbf{Z}_i, \hat{\boldsymbol{\theta}}^b\right) d_{\boldsymbol{\beta}}(\mathbf{X}_i) + \frac{1-\lambda}{N}\sum_{j=n+1}^{n+N} \xi_j \hat{m}^b\left(\boldsymbol{\beta}'\mathbf{X}_j, \hat{\boldsymbol{\theta}}^b\right) d_{\boldsymbol{\beta}}(\mathbf{X}_j).$$

3. Re-estimate $\boldsymbol{\beta}$. We use the iterative algorithm derived in Remark 2 to obtain the new estimator that $\hat{\boldsymbol{\beta}}_\lambda^b = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} \hat{\Delta}_\lambda^b\left(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}^b\right)$.

4. Estimate the variance. Repeat the above steps for B times and compute the empirical variance matrix $\hat{\Sigma}_\lambda$ of $\{\hat{\boldsymbol{\beta}}_\lambda^b, b = 1, \ldots, B\}$ to estimate the population variance Σ_λ.

Note that the nusaince estimators $\hat{\boldsymbol{\theta}}$ and $\hat{m}(\boldsymbol{\beta}'\mathbf{X}, \boldsymbol{\theta})$ don't need to be perturbed technically. However, in order to make more accurate variance estimates with finite samples, we therefore perturb it as in Cheng et al. (2021). The above variance estimation procedure ensures that $n^{\frac{1}{3}}\left(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}_0\right)$ and $n^{\frac{1}{3}}\left(\hat{\boldsymbol{\beta}}_\lambda^b - \hat{\boldsymbol{\beta}}_\lambda\right)$ have the same asymptotic distribution, so we denote the empirical variance of $\left\{\hat{\boldsymbol{\beta}}_\lambda^b : b = 1, \ldots, B\right\}$ as an estimator of the population

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

asymptotic variance. The relevant theoretical proof will be given in the Supplementary Material.

5. Simulation Results

To evaluate the finite sample performance of the proposed estimators, we consider a class of monotonic index models with different types of outcomes and decision rules that

$$Y = \nu(\mathbf{X}) + AD(\mathbf{X}) + \epsilon,$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_6)'$, A is generated from Bernoulli $\{\pi(\mathbf{X})\}$ and ϵ is generated from $N(0, 0.5^2)$. Six cases are studied:

- S1, $\nu(\mathbf{X}) = 1 + \boldsymbol{\gamma}_1'\mathbf{X}$ and $D(\mathbf{X}) = 2\boldsymbol{\beta}_0'\mathbf{X}$;
- S2, $\nu(\mathbf{X}) = 1 + \boldsymbol{\gamma}_1'\mathbf{X}$ and $D(\mathbf{X}) = \exp(0.5\boldsymbol{\beta}_0'\mathbf{X}) - 1$;
- S3, $\nu(\mathbf{X}) = 1 + \sin(\boldsymbol{\gamma}_1'\mathbf{X}) + 0.5(\boldsymbol{\gamma}_2'\mathbf{X})^2$ and $D(\mathbf{X}) = 10\boldsymbol{\beta}_0'\mathbf{X}$;
- S4, $\nu(\mathbf{X}) = 1 + X_1X_2 + 0.5X_3^2$ and $D(\mathbf{X}) = 10\boldsymbol{\beta}_0'\mathbf{X}$;
- S5, $\nu(\mathbf{X}) = 1 + \sin(\boldsymbol{\gamma}_1'\mathbf{X}) + 0.5(\boldsymbol{\gamma}_2'\mathbf{X})^2$ and $D(\mathbf{X}) = 2(\boldsymbol{\beta}_0'\mathbf{X})^3$;
- S6, $\nu(\mathbf{X}) = 1 + X_1X_2 + 0.5X_3^2$ and $D(\mathbf{X}) = 2(\boldsymbol{\beta}_0'\mathbf{X})^3$.

The linear model $\nu(\mathbf{X}, \boldsymbol{\theta}) = \boldsymbol{\theta}'(1, \mathbf{X}')'$ is posited for $\nu(\mathbf{X}) = \mu_0(\mathbf{X})$, thus case S1 and case S2 represents the model $\nu(\mathbf{X}, \boldsymbol{\theta})$ correctly specified , while

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

case S3 to case S6 are misspecified. For all the cases, we independently generate the covariates $(X_1, \ldots, X_4)$ from the multivariate standard normal distribution $N_4(0, I_4)$, X_5 from Bernoulli(0.5) and X_6 from uniform distribution U(0,1). The true parameters are set as $\boldsymbol{\beta}_0 = (1, -1, 2, 1, 2, 1)'$, $\boldsymbol{\gamma}_1 = (1, -1, 1, 1, -1, 1)'$ and $\boldsymbol{\gamma}_2 = (1, 0, -1, 0, 1, -1)'$. To evaluate the proposed SS and pooled estimatior with a known $\pi(\mathbf{X})$ in Section 2.2, we set $\pi(\mathbf{X}) = 0.5$. Due to space limitations, we provide a detailed description of the numerical simulation in the Supplementary Material. The simulation results with known $\pi(\mathbf{X})$ for case S1 is shown in Table 1 and for case S2 through case S6 are presented in Tables S1 to Table S5 in the Supplementary Material, respectively.

In all scenarios, the proposed estimators exhibit negligible bias relative to their SEs. The performance of our variance estimation method is also satisfactory, as the estimated SDs align closely with the SEs. The coverage probabilities (CPs) of confidence intervals based on adaptive skewness-adjusted quantiles are close to the nominal level of 95%, and more details are given in the Supplementary Material. Generally, for a fixed labeled sample size of $n = 200$, all estimators show comparable bias. As the size of the unlabeled sample N increases from 200 to 500, the SEs of the SS and pl estimators decrease, leading to improved efficiency. This indicates that the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Table 1: Results under case S1 with known propensity score

Method	N	Statistics	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
sup		Bias	0.000	0.013	- 0.009	- 0.001	- 0.006	- 0.004
		SE	0.051	0.049	0.047	0.049	0.061	0.115
		SD	0.048	0.046	0.047	0.049	0.060	0.111
		CP(%)	97.5	96.4	97.1	97.1	96.8	92.7
SS	200	Bias	0.005	0.021	- 0.005	0.005	- 0.002	0.007
		SE	0.040	0.034	0.034	0.040	0.036	0.037
		SD	0.035	0.032	0.032	0.035	0.034	0.037
		CP(%)	97.9	94.3	97.4	97.8	98.8	99.6
		Eff	**1.644**	**1.808**	**1.886**	**1.512**	**2.896**	**9.744**
	500	Bias	0.004	0.016	- 0.006	0.006	- 0.006	0.010
		SE	0.028	0.024	0.029	0.028	0.033	0.037
		SD	0.029	0.026	0.029	0.029	0.032	0.037
		CP(%)	97.5	95.8	96.7	98.2	99.0	97.7
		Eff	**3.237**	**3.498**	**2.618**	**3.086**	**3.440**	**9.429**
pl	200	Bias	0.004	0.017	- 0.005	0.005	- 0.006	0.013
		SE	0.020	0.019	0.023	0.021	0.024	0.037
		SD	0.024	0.023	0.027	0.024	0.028	0.039
		CP(%)	98.9	95.0	96.8	99.2	99.6	97.6
		Eff	**6.252**	**5.013**	**4.226**	**5.477**	**6.435**	**9.192**
	500	Bias	0.003	0.014	- 0.004	0.005	- 0.007	0.012
		SE	0.015	0.013	0.019	0.015	0.018	0.033
		SD	0.019	0.018	0.023	0.019	0.023	0.036
		CP(%)	99.2	94.8	95.5	99.2	99.5	95.4
		Eff	**11.543**	**8.900**	**6.244**	**9.878**	**10.831**	**11.369**

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

SS and pl methods enhance accuracy and efficiency as the size of the unlabeled sample grows. Compared to the sup estimator, the proposed SS and pl estimators demonstrate superior performance with the relative efficiency (Eff) gains across all cases. Notably, in most cases, the reduction in SE of the pl method relative to the sup method is more pronounced than that of the SS method, resulting in higher efficiency for the pl method relative to the SS method, which is consistent with the theoretical results discussed in Section 3.

When the PS is unknown, we estimate it using a logistic regression model and apply the doubly robust SS method outlined in Section 2.3 to estimate the parameters indexing optimal ITR. The relevant details and related simulation results are also shown in the Supplementary Material through Tables S6 to S10, respectively. Tables S6 and S7 show the results that both models are correctly specified, Tables S8 shows results under misspecified baseline treatment-free effect model and correctly specified PS model, and Tables S9 and S10 show results under misspecified PS model and correctly specified baseline treatment-free effect model. These results demonstrate that our method exhibits similar superiority to the supervised method when the PS is known, and validate the doubly robust property of our proposed methods against model misspecification.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

6. Real Data Analysis

In this section, we apply the proposed methods to analyze the 'ACTG 175' dataset. Due to space limitations, data descriptions can be found in the Supplementary Material. Here we focus on a subset of patients who received combination antiretroviral therapy with ZDV+ddI or ZDV+ddC. The subset comprises 1046 participants, with 522 receiving ZDV+ddI (denoted as A=1) and 524 receiving ZDV+ddC (denoted as A=0). The primary outcome Y of interest is the the CD4 T cell count (cells per cubic millimeter) at 96 ± 5 weeks, a critical marker of immune function (Phillips and Lundgren, 2006). In the subset of ACTG175 dataset, outcome CD496 is missing for 376 out of the 1046 total samples. We here focus on the SS setting, thus we randomly selecte $n = 532$ entries from the 670 samples with observed outcome values to form our labeled dataset $\mathcal{L}$ like Gunn et al. (2024) did. The remaining $N = 532$ entries are naturally designated as the unlabeled dataset $\mathcal{U}$. Baseline characteristics $\mathbf{X}$ of participants include seven binary variables and four continuous variables. These variables are essential for assessing the efficacy of the optimal ITR and understanding the impact of individual patient characteristics on treatment outcomes. The results in Table S11 of the Supplementary Material indicate that the MCAR assumption is appropriate in this study.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Table 2 reports the estimated coefficients for $\boldsymbol{\beta}$ along with the standard deviation (SD) estimated by perturbation resampling bootstrap with 500 bootstrap samples. The results indicate that the SS estimator and pl estimator yield estimates that are relatively close to those obtained from the

Table 2: Estimated parameters of optimal ITR for ACTG 175 study

Mehods	sup		SS		pl	
Predictors	Est	SD	Est	SD	Est	SD
intercept	0.024	0.032	0.026	0.049	0.138	0.104
hemo	-0.839	0.359	-0.821	0.069	-0.809	0.077
homo	-0.211	0.284	-0.237	0.074	-0.207	0.037
drugs	-0.171	0.290	-0.096	0.103	-0.173	0.037
race	0.270	0.217	0.327	0.077	0.285	0.076
gender	-0.024	0.322	-0.022	0.084	-0.010	0.056
str2	0.189	0.184	0.208	0.059	0.213	0.033
symptom	-0.190	0.206	0.087	0.097	-0.220	0.031
age	0.098	0.115	0.110	0.051	0.092	0.041
weight	0.060	0.107	0.068	0.046	0.048	0.044
cd40	-0.219	0.132	-0.251	0.060	-0.229	0.020
cd80	0.125	0.107	0.145	0.053	0.124	0.030

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

sup estimator. However, they exhibit significantly smaller standard deviations for all covariates except for the intercept term, which is consistent with our theoretical findings in Section 3 and simulation conclusions in Section 5. This suggests that the SS methods, by leveraging auxiliary information from the unlabeled data, can improve the efficiency and enhance the stability of the estimates. Table S12 and Table S13 in the Supplementary Material presents the 95% and 90% quantile-based confidence intervals (CIs) for the covariates and displays the lengths of the corresponding CIs respectively. In Table S12, CIs that are significant at the 0.05 or 0.1 level are highlighted in bold. In many studies across various medical fields, the impact of clinical, demographic, and behavioral variables on individuals with AIDS has been investigated. Ragni et al. (1995) showed that patients with hemophilia exhibited a higher incidence of severe hepatotoxicity and a shorter time to onset of this toxicity compared to non-hemophilic patients when treated with ZDV+ddI. Furthermore, asymptomatic patients demonstrated a better CD4 response to the treatment. As commonly known, the risk of HIV infection is higher among homosexual or bisexual individuals compared to heterosexual individuals (Carré et al., 1994). Similarly, individuals who inject drugs have a higher risk of HIV infection compared to those who do not use drugs (Schoenbaum et al., 1990). The significance of CD8 T-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

lymphocyte function in HIV progression has been established in studies such as Langford et al. (2007). Activated by CD4+ T-helper cells, anti-HIV CD8 T-cells assume a pivotal role in controlling viremia, as demonstrated by research like Ogg et al. (1998), responding to ongoing viral replication by increasing CD8 T-cells (Keoshkerian et al., 2003). Additionally, Friedland et al. (1991) highlighted the influence of age, race, and risk behaviors on AIDS progression. These findings indicate that our statistical analysis results align with established clinical research outcomes. From Table S13, it is evident that the confidence interval lengths of pl estimator are shorter than those of SS estimator, which in turn are shorter than those estimated using the supervised method. This observation holds true for all covariates, with the exception of the intercept term that is not focused on. Notably, traditional supervised methods failed to identify statistically significant covariate coefficients in this data analysis, neither at the 0.05 level nor at the 0.1 level. This lack of significance may be attributed to the relatively larger variance, which leads to unstable estimates and consequently wider confidence intervals. In contrast, both the SS estimation and the pl estimation demonstrated superior performance. To further validate the performance of our method, we conducted additional analyses using a train-test splitting approach. More details are described in Section S4.3 of the Supplementary

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Material. Consequently, the optimal individualized treatment recommendations derived from these methods are presented in the following Table 3. The sup, SS, and pl estimators generally recommend similar treatments for

Table 3: Treatment recommendation for ACTG 175 study

ITR	sup	SS	pl
ZDV+ddI	457	545	621
ZDV+ddC	589	501	425

most patients. However, the SS and pl estimators suggest the ZDV+ddI regime for a larger number of patients, while the recommendation for the ZDV+ddC regime is relatively smaller. And in medical research, treatment with ZDV+ddI has demonstrated a more pronounced efficacy in improving patient outcomes and slowing the progression of disease in individuals with HIV/AIDS compared to the treatment with ZDV+ddC (Darbyshire et al., 1996; Hammer et al., 1996; Mauss et al., 1996). Therefore, analysis of ACTG 175 study suggests that our SS methods are more effective than supervised method in treatment assignment and are more likely to recommend appropriate treatments to patients, which again demonstrates the superiority of our SS methods in learning the optimal ITR.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

7. Conclusion

This paper introduces a novel method for estimating optimal individualized treatment regime (ITR) in a SS setting, where the true outcome variable, denoted as Y, is observed for only a small portion of the data. Our proposed estimators use a kernel smoothing imputation technique to estimate the value function, effectively leveraging the unlabeled data $\mathcal{U}$. We then directly optimize the estimated value function to obtain parameter estimates that index the optimal ITR. To address the multidimensional covariates $\mathbf{X}$, we use a dimensionality reduction approach by projecting $\mathbf{X}$ onto a one-dimensional index $\boldsymbol{\beta}'\mathbf{X}$. This technique helps mitigate the 'curse of dimensionality' associated with Nadaraya-Watson kernel regression.

Two main types of estimators are developed that utilize both labeled and unlabeled data to enhance estimation efficiency compared to traditional supervised estimators. The first is the SS estimator, which adjusts the contribution of labeled and unlabeled data using a tuning parameter λ. The SS estimator achieves minimum asymptotic variance when $\lambda = \frac{\rho^2}{1+\rho^2}$, where ρ represents the limiting ratio of labeled to unlabeled samples as the sample size goes to infinity. The second is the pooled (pl) estimator, which impute the value function for all subjects, including those with labels. Furthermore, we propose a doubly robust estimation method for situations

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

when the PS $\pi(\mathbf{X})$ is unknown, as often occurs in observational studies. Our approach allows for potential misspecification in either the baseline treatment-free effect model or the PS model. This doubly robust property make our method more broadly applicable. We show that all the proposed estimators provide improvements in efficiency and effectiveness compared to the supervised estimators. Our results also indicate that as the amount of unlabeled data increases, the efficiency of both the SS and pl estimators improves correspondingly.

We evaluated the performance of our estimators through theoretical analysis and simulation studies under four different true model settings, highlighting their practical advantages. Our findings demonstrate that the proposed SS methods are as effective as or more effective than traditional supervised methods in estimating the optimal ITR, suggesting promising prospects for real-world applications. Overall, this paper contributes to the development of robust and efficient methods for estimating the optimal ITR in a SS framework. By effectively integrating labeled and unlabeled data, we improve estimation accuracy and effectiveness. The proposed estimators, along with the theoretical and practical advantages demonstrated under various cases, provide a solid foundation for future research and practical applications in learning optimal ITR.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Several directions are worth considering for further research. Firstly, it is natural to extend our fixed-dimensional results to high-dimensional settings where p grows with sample size. Secondly, the proposed estimation methods could be further developed to handle more complex data structures, such as survival outcome with censoring or truncation, or outcomes are missing under missing at random mechanisms or in blocks. Thirdly, addressing potential heterogeneity between labeled and unlabeled populations through methods such as transfer learning and federated learning warrants investigation. Methodologically, while our current approach uses the Nelder-Mead algorithm to handle non-smoothness, our estimation procedure can be easily extended to incorporate the smoothing techniques, such as those proposed in Feng et al. (2022, 2024) to facilitate gradient-based optimization and improve inferential properties, albeit at the cost of bandwidth selection. Furthermore, extending our framework to accommodate nonlinear decision rules and more complex policy learning problems would provide additional insights. These challenges merit further investigation.

Supplementary Material

The online Supplementary Material contains additional asymptotic results, theoretical proofs and additional numerical descriptions and results.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Acknowledgments

The authors thank the co-editor, the associate editor and reviewers for their helpful suggestions. Zhou and Peng' s work was supported by the National Key R&D Program of China (2021YFA1000100, 2021YFA1000101 and 2021YFA1000104) and Shanghai Key Program of Computational Biology (23JS1400500). Zhou' s work was supported by the National Natural Science Foundation of China (71931004). Peng' s work was supported by the National Natural Science Foundation of China (12301337, 72331005).

Conflict of interest

The authors declare that there are no conflicts of interest.

References

Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Springer, Cham.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47:1148–1178.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89:133–161.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

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

Carré, N., Deveau, C., Belanger, F., Boufassa, F., Persoz, A., Jadand, C., Rouzioux, C., Delfraissy, J.-F., Bucquet, D., Group, S. S., et al. (1994). Effect of age and exposure group on the onset of aids in heterosexual and homosexual hiv-infected patients. *Aids*, 8(6):797–802.

Chakrabortty, A. and Cai, T. (2018). Efficient and adaptive linear regression in semi-supervised settings. *Ann. Statist.*, 46:1541–1572.

Chakrabortty, A., Dai, G., and Tchetgen, E. T. (2022). A general framework for treatment effect estimation in semi-supervised and high dimensional settings. *arXiv preprint arXiv:2201.00468*.

Chakraborty, B., Murphy, S., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical methods in medical research*, 19(3):317–343.

Chapelle, O., Scholkopf, B., and Zien, A. (2006). Semi-supervised learning. 2006. *Cambridge, Massachusettes: The MIT Press View Article*, 2:1.

Cheng, D., Ananthakrishnan, A. N., and Cai, T. (2021). Robust and effi-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

cient semi-supervised estimation of average treatment effects with application to electronic health records data. *Biometrics*, 77(2):413–423.

Chu, J., Lu, W., and Yang, S. (2023). Targeted optimal treatment regime learning using summary statistics. *Biometrika*, 110(4):913–931.

Correa, N., Cerquides, J., Vassena, R., Popovic, M., and Arcos, J. L. (2024). Idoser: Improving individualized dosing policies with clinical practice and machine learning. *Expert Systems with Applications*, 238:121796.

Darbyshire, J., Foulkes, M., Peto, R., Duncan, W., Babiker, A., Collins, R., Hughes, M., Peto, T. E., Walker, S. A., and Group, C. H. (1996). Zidovudine (azt) versus azt plus didanosine (ddi) versus azt plus zalcitabine (ddc) in hiv infected adults. *Cochrane database of systematic reviews*, 2010(3).

Ding, P. and Li, F. (2018). Causal inference: a missing data perspective. *Statistical Science*, 33:214–237.

Fan, C., Lu, W., Song, R., and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(5):1565–1582.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Feng, H., Duan, J., Ning, Y., and Zhao, J. (2024). Test of significance for high-dimensional thresholds with application to individualized minimal clinically important difference. *Journal of the American Statistical Association*, 119(546):1396–1408.

Feng, H., Ning, Y., and Zhao, J. (2022). Nonregular and minimax estimation of individualized thresholds in high dimension with binary responses. *The Annals of Statistics*, 50(4):2284–2305.

Friedland, G. H., Saltzman, B., Vileno, J., Freeman, K., Schrager, L. K., and Klein, R. S. (1991). Survival differences in patients with aids. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 4(2):144–153.

Gunn, K., Lu, W., and Song, R. (2024). Adaptive semi-supervised inference for optimal treatment decisions with electronic medical record data. In *Statistics in Precision Health: Theory, Methods and Applications*, pages 229–246. Springer.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090.

Jin, Z., Ying, Z., and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika*, 88(2):381–390.

Keoshkerian, E., Ashton, L. J., Smith, D. G., Ziegler, J. B., Kaldor, J. M., Cooper, D. A., Stewart, G. J., and Ffrench, R. A. (2003). Effector hiv-specific cytotoxic t-lymphocyte activity in long-term nonprogressors: Associations with viral replication and progression. *Journal of medical virology*, 71(4):483–491.

Kitagawa, T. and Tetenov, A. (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86:591–616.

Langford, S. E., Ananworanich, J., and Cooper, D. A. (2007). Predictors of disease progression in hiv infection: a review. *AIDS research and therapy*, 4:1–14.

Liao, K. P., Cai, T., Gainer, V., Goryachev, S., Zeng-treitler, Q., Raychaudhuri, S., Szolovits, P., Churchill, S., Murphy, S., Kohane, I., et al. (2010). Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis care & research*, 62(8):1120–1127.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Liu, Y., Wang, Y., Kosorok, M., Zhao, Y.-Q., and Zeng, D. (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in Medicine*, 37(22):3776–3788.

Lu, W., Zhang, H., and Zeng, D. (2013). Variable selection for optimal treatment decision. *Statistical Methods in Medical Research*, 22(5):493–504.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246.

Mauss, S., Adams, O., Willers, R., and Jablonowski, H. (1996). Combination therapy with zdv+ ddi versus zdv+ ddc in patients with progression of hiv-infection under treatment with zdv. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 11(5):469–477.

Mo, W. and Liu, Y. (2022). Efficient learning of optimal individualized treatment rules for heteroscedastic or misspecified treatment-free effect models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):440–472.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Ogg, G. S., Jin, X., Bonhoeffer, S., Dunbar, P. R., Nowak, M. A., Monard, S., Segal, J. P., Cao, Y., Rowland-Jones, S. L., Cerundolo, V., et al. (1998). Quantitation of hiv-1-specific cytotoxic t lymphocytes and plasma load of viral rna. *Science*, 279(5359):2103–2106.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649.

Phillips, A. N. and Lundgren, J. D. (2006). The cd4 lymphocyte count and risk of clinical progression. *Current opinion in HIV and AIDS*, 1(1):43–49.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180.

Ragni, M. V., Amato, D. A., LoFaro, M. L., DeGruttola, V., Van Der Horst, C., Eyster, M. E., Kessler, C. M., Gjerset, G. F., Ho, M., Parenti, D. M., et al. (1995). Randomized study of didanosine monotherapy and combination therapy with zidovudine in hemophilic and nonhemophilic subjects with asymptomatic human immunodeficiency virus-1 infection. *Blood*, 85(9):2337–2346.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In Lin, D. Y. and Heagerty, P. J., editors, *Proceedings of the Second Seattle Symposium in Biostatistics*, volume 179 of *Lecture Notes in Statistics*, pages 189–326, New York, NY, USA. Springer.

Robins, J. M., Hernán, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *American journal of epidemiology*, 152(4):327–333.

Robins, J. M., Rotnitzky, A., and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866.

Robins, J. M., Rotnitzky, A., and Zhao, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90:106–121.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Schoenbaum, E. E., Hartel, D., and Friedland, G. (1990). Hiv infection and intravenous drug use. *Current Opinion in Infectious Diseases*, 3(1):80–93.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Shi, C., Fan, A., Song, R., and Lu, W. (2018). High-dimensional a-learning for optimal dynamic treatment regimes. *The Annals of Statistics*, 46:925–957.

Sonabend-W, A., Laha, N., Ananthakrishnan, A. N., Cai, T., and Mukherjee, R. (2023). Semi-supervised off-policy reinforcement learning and value estimation for dynamic treatment regimes. *Journal of Machine Learning Research*, 24(323):1–86.

Song, S., Lin, Y., and Zhou, Y. (2023). A general m-estimation theory in semi-supervised framework. *Journal of the American Statistical Association*, pages 1–11.

Wang, L., Zhou, Y., Song, R., and Sherwood, B. (2018). Quantile-optimal treatment regimes. *Journal of the American Statistical Association*, 113(523):1243–1254.

Wang, Y., Zhou, Q., Cai, T., and Wang, X. (2023). Semi-supervised estimation of event rate with doubly-censored survival data. *arXiv preprint arXiv:2311.02574*.

Watkins, C. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

Watkins, C. J. H. (1989). *Learning from delayed rewards*. PhD thesis, King's College.

Zhang, A., Brown, L. D., and Cai, T. T. (2019). Semi-supervised inference: General theory and estimation of means. *The Annals of Statistics*, 47(5):2538 – 2566.

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

Zhang, Y. and Imai, K. (2023). Individualized policy evaluation and learning under clustered network interference. *arXiv preprint arXiv:2311.02467*.

Zhao, Y.-Q., Laber, E., Ning, Y., Saha, S., and Sands, B. (2019). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *Journal of Machine Learning Research*, 20:1–23.

Zhao, Y.-Q., Zeng, D., Rush, A., and Kosorok, M. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107:1106–1118.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0168

REFERENCES

Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112:169–187.

Xintong Li

School of Statistics, East China Normal University, Shanghai, China

E-mail: xtli2022@163.com

Mengjiao Peng

School of Statistics, Academy of Statistics and Interdisciplinary Sciences, KLATASDS-MOE, East China Normal University, Shanghai, China

E-mail: mjpeng@fem.ecnu.edu.cn

Yong Zhou

School of Statistics, Academy of Statistics and Interdisciplinary Sciences, KLATASDS-MOE, East China Normal University, Shanghai, China

E-mail: yzhou@fem.ecnu.edu.cn