Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Semi-supervised D-Learning for Optimal Individual Treatment Regimes

Xintong Li[1] | **Shuyi Zhang**[1,2,3] | **Yong Zhou**[1,2,3]

[1]School of Statistics, East China Normal University, Shanghai, China

[2]Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, East China Normal University, Shanghai, China

[3]Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai, China

Correspondence
Shuyi Zhang, 3663 North Zhongshan Road, Shanghai, China. 200062
Email: syzhang@fem.ecnu.edu.cn
Yong Zhou, 3663 North Zhongshan Road, Shanghai, China. 200062
Email: yzhou@fem.ecnu.edu.cn

Abstract

In precision medicine, linear treatment decision class have attracted widespread attention due to their simple structure and good interpretability. However, linear decision models may often be misspecified in practical applications. To address this issue, we develop an imputation-based semi-supervised D-learning method that leverages information from unlabeled data to enhance the efficiency of estimating optimal individual treatment regimes (ITRs) when the linear decision model is misspecified. Specifically, we estimate the imputation function using a projection-based dimension reduction approach, adjust for the bias in imputation estimates via a residual refitting step, and estimate decision function parameters based on the debiased imputation function. To mitigate bias from potential overfitting, cross-validation is incorporated. Theoretical results show that the semi-supervised D-learning method achieves $\sqrt{n}$-consistent parameter estimates with asymptotic normality. Numerical experiments on both simulated and real datasets demonstrate the superior performance of our proposed semi-supervised D-learning approach under model misspecification.

KEYWORDS
Precision medicine, Direct learning, Semi-supervised inference, Least square estimation

1 | INTRODUCTION

Precision medicine, also known as personalized medicine, has gained widespread attention in recent years. Due to patient heterogeneity, traditional 'one-size-fits-all' approaches often fall short in adequately addressing the diverse needs of individuals, thereby limiting their effectiveness in real-world settings [1,2]. Precision medicine is a healthcare model that tailors health management and clinical decisions to individual patients based on their intrinsic biological information and clinical presentations. Precision medicine aims to leverage abundant electronic medical records (EMRs) data to recommend optimal individualized treatment regime (ITR) for patients, thereby achieving better expected outcomes across the population.

In the literature, various traditional statistical methods have been developed for learning optimal individualized treatment rules (ITRs), including indirect methods such as Q-learning, A-learning, and direct methods such as outcome-weighted learning (OWL), D-learning, along with their extensions. Q-learning [3–6] estimates the expected outcomes associated with different treatment actions at each decision point. The optimal treatment ITR is derived by selecting the treatment that maximizes the Q-function. While Q-learning is intuitive and widely applicable, it is sensitive to model misspecification due to its reliance on correctly specifying the outcome regression models. A-learning [7–10] improves upon Q-learning by directly modeling the contrast function, which quantifies the differential treatment effect between options. This approach avoids the unnecessary modeling of treatment-free effect and improves its reliability in practical applications. Outcome-weighted learning [11–14] approaches the estimation of optimal ITRs as a weighted classification problem. By assigning weights based on patients' observed outcomes, OWL prioritizes individuals who demonstrate larger treatment benefits. D-learning [15–18], a one-step estimation method combining the advantage of model-based methods and classification-based methods, targets directly on decision functions which maximizing the value function and is robust to the misspecification of treatment-free effect model. Unlike Q-learning

Abbreviations: ITR, individualized treatment regime; CV, cross-validation; KRLS, kernel-regularized least squares.

and A-learning, direct methods focus solely on treatment allocation rules, making it particularly advantageous in settings with complex or difficult-to-model outcome distributions.

In the study of estimating optimal ITR, linear decision classes are particularly favored by researchers due to their simple structure and good interpretability[19–21]. However, linear decision model can suffer from bias due to misspecified relationships between covariates and treatment-related interaction effect, leading to suboptimal decisions. To address model misspecification, some efforts have been made for nonlinear decision-making[11,22]. For example, Song et al.[23] proposed a novel method to estimate optimal ITR under a semiparametric additive single-index model and the link function was estimated by B-spline. But it suffers from model mis-specification. Qi and Liu[15] handled the nonlinear ITR by kenel-based and machine learning methods. But these methods potentially require full observation of individual-level data. However, collecting labeled data of interest is often costly and labor-intensive, resulting in a scarcity of labeled data within EMRs, which poses challenges for effective modeling.

This paper concerns the challenge of the underutilization of large amounts of unlabeled data. Semi-supervised learning (SSL) has garnered significant attention for tackling this issue by leveraging information from both labeled and unlabeled data to enhance estimation efficiency and robustness. While SSL methods have been increasingly applied in precision medicine for instance, in treatment effect estimation[24–28] their application to the estimation of optimal ITR remains relatively limited. For example, Sonabend-W et al.[29] introduced a semi-supervised Q-learning method along with a doubly robust off-policy reinforcement learning framework for evaluating optimal dynamic treatment regime (DTR). Gunn et al.[30] proposed a kernel-based semi-supervised approach to estimate contrast functions and thereby learn optimal ITR. However, this approach may be suffered by the curse of dimensionality inherent to nonparametric methods. These existing methods focus primarily on accurate estimation of the Q-function or contrast functions derived from it, then induce the optimal ITR by maximizing the estimated value function. In contrast, our work directly targets the decision function itself, providing a more focused framework for optimal treatment decision-making.

In this paper, we first review the traditional supervised D-learning method for estimating optimal ITR based on fully observed data. Building on this foundation, we propose several imputation-based semi-supervised D-learning estimators to address the challenges posed by partially observed data. To mitigate the curse of dimensionality commonly associated with nonparametric kernel methods, we introduce a projection step for dimensionality reduction, followed by a refitting step to correct the bias in the imputation-based estimators. Furthermore, we incorporate $\mathbb{K}$-fold cross-validation (CV) to reduce overfitting-related bias. Our proposed semi-supervised D-learning framework is flexible and accommodates various approaches for estimating the imputation function. For example, the kernel-regularized least squares (KRLS) method proposed by Hainmueller and Hazlett[31] can be seamlessly integrated. We establish the asymptotic properties of the proposed estimators, demonstrating $\sqrt{n}$-consistency and asymptotic normality. Under certain regularity conditions, the proposed framework achieves efficiency gains when the linear decision function is misspecified. In specific cases, it attains asymptotic optimality. For inference, we compare the variance estimates derived from direct plug-in consistent estimators with those obtained using a double CV procedure. These comparisons are supported by numerical simulations under various settings. To extend the frameworks applicability, we briefly discuss the use of weighting schemes to construct efficient estimators and its potential application to observational studies.

The remainder of this paper is structured as follows. Section 2 introduces the semi-supervised D-learning framework for estimating optimal ITR. Section 3 establishes the theoretical properties of the proposed estimators and highlights their advantages over supervised methods under linear model misspecification. Section 4 presents Monte Carlo simulation results and an application to the MIMIC-IV dataset. Section 5 explores potential extensions for improving estimator efficiency and discusses the methods applicability in observational study settings. Finally, Section 6 concludes the paper with a discussion. Detailed proofs of the theoretical results are provided in supporting information.

2 | METHODOLOGY

2.1 | Preliminaries

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}$ be the outcome variable which is assumed that larger values are better without loss of generality. Let $\mathbf{X} = (1, X_1, \ldots, X_p) \in \mathcal{X} \subseteq \mathbb{R}^{p+1}$ be the individual level covariates with bounded support $\mathcal{X}$, and $\mathbf{X}^- = (X_1, \ldots, X_p)$ positive definite variance $\mathrm{Var}(\mathbf{X}^-) > 0$ and fixed dimension. Note that $\mathbf{X}$ potentially includes the interception term. Assume that $(Y, \mathbf{X})$ has finite second moment and denote the corresponding joint distribution as $\mathbb{P}_{(Y,\mathbf{X})} = \mathbb{P}_{Y|\mathbf{X}}\mathbb{P}_{\mathbf{X}}$. Let $A \in \mathcal{A} = \{-1, 1\}$ be

the binary treatment indicator. In this paper, we consider the random clinic trial (RCT) setting with known propensity score $\pi(a, \mathbf{x}) = P(A = a|\mathbf{X} = \mathbf{x})$ to demonstrate the main ideas.

The observations are from two data sources that one is the labeled class $\mathcal{L} = \{\mathbf{Z}_i = (\mathbf{X}_i, A_i, Y_i) : i = 1, 2, \ldots, n\}$ and the other is the unlabeled data class $\mathcal{U} = \{\mathbf{X}_i : i = n+1, n+2, \ldots, n+N\}$. Note that all the samples in the same source are iid and $\mathcal{L} \perp \mathcal{U}$. Assume that $N \gg n$, i.e. as $n, N \to \infty$, $\frac{n}{N} \to 0$, and assume that $\mathcal{L}$ and $\mathcal{U}$ follow the same underlying distribution $\mathbb{P}_{(Y,\mathbf{X})}$ as in traditional semi-supervised framework.

Let $Y^*(a)$ be the potential outcome that would be observed if the patient received treatment a. The standard stable unit treatment value assumption (SUTVA) that $Y = Y^*(A)$, the no unmeasured confounding (NUC) assumption that $A \perp \{Y^*(1), Y^*(-1)\}|\mathbf{X}$ and the positivity assumption that $0 < \pi(A, \mathbf{X}) < 1$ for $A \in \mathcal{A}$ and $\mathbf{X} \in \mathcal{X}$ are also made here. Our goal is to find the optimal individualized treatment regimes (ITR) $d(\mathbf{X}) : \mathcal{X} \to \mathcal{A}$, a mapping from the covariates space to the treatment space, that maximize the average potential outcome $V(d) = E[Y^*(d)]$ that $d^{\text{opt}}(\mathbf{X}) = \arg\max_d V(d)$. Qian and Murphy[5] represented the value function as

$$V(d) = E\left[\frac{YI(A = d(\mathbf{X}))}{\pi(A, \mathbf{X})}\right], \tag{1}$$

where $I(\cdot)$ represent the indicator function. Song et al.[23] reformulated this problem and similarly Qi and Liu[15] proposed the following D-Learning method that the maximizer of (1) could be rewritten as

$$\begin{aligned} d^{\text{opt}}(\mathbf{X}) =& \text{sign}\{E[Y|\mathbf{X}, A = 1] - E[Y|\mathbf{X}, A = -1]\} \\ =& \text{sign}\left\{E\left[\frac{AY}{\pi(A, \mathbf{X})}\middle|\mathbf{X}\right]\right\} := \text{sign}\{f^{\text{opt}}(\mathbf{X})\} \end{aligned}$$

where $d^{\text{opt}}(\mathbf{X})$ is the corresponding optimal ITR defined as the sign of the conditional average treatment effect (CATE). Assume that the outcome can be generally expressed by

$$Y = \mu_0(\mathbf{X}) + A\delta(\mathbf{X}) + e,$$

where $\mu_0(\mathbf{X})$ is the treatment-free effect, $\delta(\mathbf{X})$ is the interaction effect and e is the mean zero random error term. Thus the optimal decision function $f^{\text{opt}}(\mathbf{X}) = E\left[\frac{AY}{\pi(A,\mathbf{X})}\middle|\mathbf{X}\right] = 2\delta(\mathbf{X})$. For simplicity and interpretability the linear decision class $f \in \mathcal{F} = \{f(\mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^{p+1}\}$ is often adopted. Naturally, the supervised estimator for $\boldsymbol{\beta} = (c_0, \beta_1, \ldots, \beta_p)$ can be obtained by the ordinary least squares (OLS) method that

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} P_n\left\{\frac{AY}{\pi(A, \mathbf{X})} - \mathbf{X}^T\boldsymbol{\beta}\right\}^2, \tag{2}$$

where $P_n f(x) = \frac{1}{n}\sum_{i=1}^n f(x_i)$. This method offers a simple way to obtain the estimator without requirement of modeling $\mu_0(\mathbf{X})$. In this paper, we mainly focus on the known propensity score case in a RCT setting with $\pi(1, \mathbf{X}) = \pi(-1, \mathbf{X}) = 0.5$, then the problem (2) could be simplified as the solution of

$$P_n\mathbf{X}(2AY - \mathbf{X}^T\boldsymbol{\beta}) = 0. \tag{3}$$

Define $\boldsymbol{\beta}_0$ as the solution to the equation $E[\mathbf{X}(2AY - \mathbf{X}^T\boldsymbol{\beta})] = 0$ in $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$. Then the OLS estimator $\hat{\boldsymbol{\beta}}$ is asymptotic linear with influence function $\varphi_0(\mathbf{Z}_i) = E[\mathbf{X}\mathbf{X}^T]^{-1}\{\mathbf{X}_i(2A_iY_i - \mathbf{X}_i^T\boldsymbol{\beta}_0)\}$[32].

2.2 | Semi-supervised D-Learning

The OLS estimator is optimal within the linear decision class and efficiency could not gain from additional information about $\mathbb{P}_{\mathbf{X}}$, the marginal distribution of $\mathbf{X}$[33]. However, when the decision model $f(\mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta}$ is mis-specified, which is common, the OLS estimator $\hat{\boldsymbol{\beta}}$ obtained from labeled data $\mathcal{L}$ only may suffer from downgraded efficiency and the induced ITR is suboptimal in general sense. In order to improve the efficiency of estimator by sufficiently and safely utilizing the auxiliary information in $\mathcal{U}$, we consider the semi-supervised learning method based on imputation. Obviously, accurate imputation is of great importance, otherwise it will introduce non-negligible bias to the estimate. A natural idea is to impute based on the nonparametric kernel

approach. Let $m(\mathbf{X}) = E[2AY|\mathbf{X}] = E[2AY|\mathbf{X}^-]$, and its corresponding nonparametric kernel estimator is given by

$$\hat{m}(\mathbf{X}_j) = \frac{(nh^p)^{-1}\sum_{i=1}^{n} H_h(\mathbf{X}_i^-, \mathbf{X}_j^-) * 2A_iY_i}{(nh^p)^{-1}\sum_{i=1}^{n} H_h(\mathbf{X}_i^-, \mathbf{X}_j^-)},$$

where $H_h(u,v) = H(\frac{u-v}{h})$ with kernel function $H : \mathbb{R}^p \to \mathbb{R}$ and for simplicity, we take the same bandwidth h for each component in $H_h(u,v)$. Thus the Semi-supervised D-Learning estimator $\hat{\boldsymbol{\beta}}_{np}$ based on fully nonparametric imputation [30] could be obtained by the solution of

$$\frac{1}{N}\sum_{j=n+1}^{n+N} \mathbf{X}_j(\hat{m}(\mathbf{X}_j) - \mathbf{X}_j^T\boldsymbol{\beta}) = 0.$$

As $\hat{\boldsymbol{\beta}}_{np}$ has been shown to be asymptotically optimal with influence function $\varphi_{\text{eff}}(\mathbf{Z}_i) = E[\mathbf{X}\mathbf{X}^T]^{-1}\{\mathbf{X}_i(2A_iY_i - m(\mathbf{X}_i))\}$ among the class of all regular and asymptotically linear (RAL) estimators under some regularity conditions [34]. However, the p-dimensional nonparametric kernel method often becomes impractical because of the curse of dimensionality caused by larger p. Thus we conduct a dimension reduction step by projection and then a refitting step as in the literature [29,30,34] to remove the bias of the imputation estimator. Define $m(\mathbf{X}^T\boldsymbol{\beta}) = E[2AY|\mathbf{X}^T\boldsymbol{\beta}]$ and denote its corresponding nonparametric kernel estimator as

$$\hat{m}(\mathbf{X}_j^T\boldsymbol{\beta}) = \frac{(nh)^{-1}\sum_{i=1}^{n} K_h(\mathbf{X}_i^T\boldsymbol{\beta}, \mathbf{X}_j^T\boldsymbol{\beta}) * 2A_iY_i}{(nh)^{-1}\sum_{i=1}^{n} K_h(\mathbf{X}_i^T\boldsymbol{\beta}, \mathbf{X}_j^T\boldsymbol{\beta})}. \tag{4}$$

where $K_h(u,v) = K(\frac{u-v}{h})$ with kernel function $K : \mathbb{R} \to \mathbb{R}$ and bandwidth h. Note that $m(\mathbf{X}^T\boldsymbol{\beta})$ is not necessarily equal to $m(\mathbf{X})$. Then the corresponding parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ of refitting step could be estimated by regressing the residual on $\mathbf{X}$. Specifically, define $\boldsymbol{\theta}_0$ as the solution of $E[\mathbf{X}(2AY - m(\mathbf{X}^T\boldsymbol{\beta}_0) - \mathbf{X}^T\boldsymbol{\theta})] = 0$, and $\hat{\boldsymbol{\theta}}$ is estimated by the estimating equation that

$$P_n\mathbf{X}(2AY - \hat{m}(\mathbf{X}^T\hat{\boldsymbol{\beta}}) - \mathbf{X}^T\boldsymbol{\theta}) = 0. \tag{5}$$

Denote the imputation function after the refitting step as $\nu(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\theta}) = m(\mathbf{X}^T\boldsymbol{\beta}) + \mathbf{X}^T\boldsymbol{\theta}$, and its semiparametric (SP) estimator is

$$\hat{\nu}(\mathbf{X}; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \hat{m}(\mathbf{X}^T\hat{\boldsymbol{\beta}}) + \mathbf{X}^T\hat{\boldsymbol{\theta}}.$$

Thus, the Semi-supervised D-learning estimator $\hat{\boldsymbol{\beta}}_{sp}$ based on semiparametric imputation function estimation can be obtained by solving the following estimation equation

$$\frac{1}{N}\sum_{j=n+1}^{n+N} \mathbf{X}_j(\hat{\nu}(\mathbf{X}_j; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) - \mathbf{X}_j^T\boldsymbol{\beta}) = 0.$$

To overcome the bias caused by over-fitting, we consider the $\mathbb{K}$-fold cross-validation (CV) technique here. Let $\mathcal{L}_k$ be the k-th random disjoint partition of $\mathcal{L}$ with sample size $n_{\mathbb{K}} = \frac{n}{\mathbb{K}}$ and denote its corresponding index set as $\mathcal{I}_k$ for $k \in \{1, \ldots, \mathbb{K}\}$. Denote the set excluding the k-th partition as $\mathcal{L}_k^- = \mathcal{L} - \mathcal{L}_k$ with corresponding sample size $n_{\mathbb{K}}^- = n - n_{\mathbb{K}}$ and index set $\mathcal{I}_k^-$. And denote the OLS estimator from (3) and nonparametric imputation estimator under dimension reduction from (4) based on $\mathcal{L}_k^-$ as $\hat{\boldsymbol{\beta}}_k$ and $\hat{m}_k(\mathbf{X}^T\boldsymbol{\beta})$ respectively. Then we could rewrite the equation (5) based on CV as

$$\frac{1}{n}\sum_{k=1}^{\mathbb{K}}\sum_{i\in\mathcal{I}_k} \mathbf{X}_i(2A_iY_i - \hat{m}_k(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_k) - \mathbf{X}_i^T\boldsymbol{\theta}) = 0, \tag{6}$$

and denote the solution of (6) as $\hat{\boldsymbol{\theta}}_{\mathbb{K}}$. Next, the semiparametric imputation function estimation based on CV could be obtained by

$$\hat{\nu}(\mathbf{X}; \hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\theta}}_{\mathbb{K}}) = \frac{1}{\mathbb{K}}\sum_{k=1}^{\mathbb{K}} \hat{m}_k(\mathbf{X}^T\hat{\boldsymbol{\beta}}_k) + \mathbf{X}^T\hat{\boldsymbol{\theta}}_{\mathbb{K}}.$$

Thus the Semi-supervised D-Learning estimator $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ based on semiparametric imputation with $\mathbb{K}$-fold CV could be obtained by the solution of

$$\frac{1}{N}\sum_{j=n+1}^{n+N} \mathbf{X}_j(\hat{\nu}(\mathbf{X}_j; \hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\theta}}_{\mathbb{K}}) - \mathbf{X}_j^T\boldsymbol{\beta}) = 0.$$

When $\mathbb{K} = 1$, the semiparametric imputation function $\nu(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\theta})$ is estimated by the entire $\mathcal{L}$ and the corresponding parameter estimator is $\hat{\boldsymbol{\beta}}_{sp}$.

Remark 1. Alternatively, the KRLS method proposed by Hainmueller and Hazlett[31] can also be employed for estimating the imputation function. This method maintains both model flexibility and interpretability by integrating kernel methods with Tikhonov regularization. Following a procedure similar to that used for estimating $\hat{\boldsymbol{\beta}}_{sp}$, we first use the KRLS method (without the need for dimensionality reduction) to obtain $\hat{m}_{KRLS}(\mathbf{X})$, the estimator of $m(\mathbf{X})$, and then perform a refitting step to remove the bias before obtaining the final parameter estimation $\hat{\boldsymbol{\beta}}_{KRLS}$. Similarly, the $\mathbb{K}$-fold CV could also be conducted here as before to obtain $\hat{\boldsymbol{\beta}}_{KRLS,\mathbb{K}}$. We will demonstrate the performance of this approach then in the numerical simulations.

Remark 2. Although our work and Chakrabortty and Cai[34] both employ semi-supervised learning to leverage unlabeled data, they differ significantly in both research focus and methodology. First, in terms of research objectives, Chakrabortty and Cai[34] aimed to enhance the estimation efficiency of linear regression parameters in biomarker-disease association studies, such as rheumatoid arthritis. In contrast, our work is centered on optimizing personalized treatment decision-making in precision medicine. While linear decision models are widely used for their simplicity and interpretability, they are susceptible to model misspecification in real-world applications. To address this issue, we develop an innovative semi-supervised D-learning framework designed to improve the estimation efficiency of decision parameters, ultimately leading to more accurate treatment decisions. Second, from a methodological perspective, our approach to semi-supervised learning is specifically tailored to the structure of decision functions. Unlike Chakrabortty and Cai[34], which employed sufficient dimension reduction techniques such as sliced inverse regression (SIR) to determine projection directions, we directly project onto the single index $\mathbf{X}^T\boldsymbol{\beta}$. This not only enhances the estimation efficiency of linear decision parameters under potential model misspecification but also circumvents the computational burden associated with selecting projection directions. Furthermore, we will rigorously validate the effectiveness of our method through comprehensive theoretical analysis and extensive numerical experiments.

3 | THEORETICAL RESULTS

In this section, we will establish the asymptotic normality of our semi-supervised estimator $\hat{\boldsymbol{\beta}}_{sp}$ and $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ under some reasonable assumptions.

Assumption 1. (i) The kernel function $K(\cdot) : \mathbb{R} \to \mathbb{R}$ is a symmetric q-th order kernel with finite q-th moment for some integer $q \geqslant 2$. (ii) $K(\cdot)$ is bounded, integrable and Lipschitz continuous with a compact support. (iii) $K(\cdot)$ has a bounded, integrable and Lipschitz continuous derivative $\nabla K(\cdot)$ on a compact support. (iv) For any $z_1, z_2 \in \mathbb{R}$ that $|z_1 - z_2|$ is bounded, there exists some bounded and integrable function $\phi(\cdot)$, such that $|\nabla K(z_1) - \nabla K(z_2)| \leqslant \phi(z_1)|z_1 - z_2|$. (v) Denote $s = \mathbf{x}^T\boldsymbol{\beta}$ with a compact support $\mathcal{S} \subseteq \mathbb{R}$ and denote its density function as $f(s)$. $f(s)$ is bounded and away from zero on $\mathcal{S}$. $m(s)$ and $f(s)$ are both q times continuous differentiable with bounded q-th derivative on some open set containing $\mathcal{S}$. (vi) $E[\mathbf{X}|S = s]$ and $E[\mathbf{X}Y|S = s]$ are both continuous differentiable with bounded first derivative on some open set containing $\mathcal{S}$. (vii) $E[|Y|^\alpha] < \infty$ for some $\alpha > 2$ and $E[|Y|^\alpha|S = s]f(s)$ is bounded on $\mathcal{S}$.

Before giving the main theorem, we first give the following two useful lemmas.

Lemma 1. *Under Assumption 1 and the the optimal bandwidth order* $h_{opt} = O_p\left(n^{-\frac{1}{2q+1}}\right)$*, we have*

$$\sup_{\mathbf{x}\in\mathcal{X}} |\hat{m}(\mathbf{x}^T\hat{\boldsymbol{\beta}}) - m(\mathbf{x}^T\boldsymbol{\beta}_0)| = O_p(a_n),$$

where $a_n = n^{-\frac{q}{2q+1}}\sqrt{\log n}$.

Lemma 2. *Denote* $G(\mathbf{X}_i) = \mathbf{X}_i(\hat{m}(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}) - m(\mathbf{X}_i^T\boldsymbol{\beta}_0)) - E_{\mathbf{X}}[\mathbf{X}(\hat{m}(\mathbf{X}^T\hat{\boldsymbol{\beta}}) - m(\mathbf{X}^T\boldsymbol{\beta}_0))]$. *Under Assumption 1 and the the optimal bandwidth order as in Lemma 1, we have*

$$\sqrt{n}P_nG(\mathbf{X}) = O_p(b_n),$$

where $b_n = n^{-\frac{2q-3}{2(2q+1)}}$.

Remark 3. Lemma 1 establishes the uniform consistency of nonparametric kernel-based estimator $\hat{m}(\mathbf{x}^T\hat{\boldsymbol{\beta}})$ with L_∞ error rate $O_p\left(n^{-\frac{q}{2q+1}}\sqrt{\log n}\right)$ under the optimal bandwidth order $h_{\text{opt}} = O_p\left(n^{-\frac{1}{2q+1}}\right)$. And Lemma 2 establishes the convergence rate of the

centered empirical process of $\hat{m}(\mathbf{x}^T\hat{\boldsymbol{\beta}})$. These two lemmas will be employed in the proof of the upcoming main theorem. And the proofs of both lemmas are provided in the supporting information.

Theorem 1. *Under Assumption 1 and the the optimal bandwidth order as in Lemma 1, we have*

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{sp} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(\mathbf{Z}_i) + O_p(r_{n,N}),$$

where the influence function $\varphi(\mathbf{Z}_i) = E[\mathbf{X}\mathbf{X}^T]^{-1}\{\mathbf{X}_i[2A_iY_i - \nu(\mathbf{X}_i;\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]\}$ *and* $r_{n,N} = O_p\left(\frac{n}{N}\right)^{\frac{1}{2}} + O_p(b_n)$ *with* b_n *was defined in Lemma 2. Thus* $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{sp} - \boldsymbol{\beta}_0) \xrightarrow{d} N_{p+1}(\boldsymbol{\theta}, \Sigma)$ *with positive definite* $(p+1)\times(p+1)$ *matrix* $\Sigma = E[\mathbf{X}\mathbf{X}^T]^{-1}E\{\mathbf{X}\mathbf{X}^T[2AY - \nu(\mathbf{X};\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)]^2\}E[\mathbf{X}\mathbf{X}^T]^{-1}$.

Theorem 2. *When the fold of CV is fixed and satisfies* $\mathbb{K} \geqslant 2$*, under the conditions same as Theorem 1, we have*

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(\mathbf{Z}_i) + O_p(\gamma_{n,N}),$$

where $\gamma_{n,N} = O_p\left(\frac{n}{N}\right)^{\frac{1}{2}} + O_p(a_{n_{\mathbb{K}^-}})$ *with* a_n *was defined in Lemma 1. Thus* $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}} - \boldsymbol{\beta}_0) \xrightarrow{d} N_{p+1}(\boldsymbol{\theta}, \Sigma)$ *as in Theorem 1.*

Remark 4. Theorem 1 establishes the asymptotic normality of Semi-supervised parameter estimator $\hat{\boldsymbol{\beta}}_{sp}$ without CV and Theorem 2 establishes that with $\mathbb{K}$-fold CV. When $\nu(\mathbf{X};\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) = m(\mathbf{X})$ (e.g., when $\delta(\mathbf{X}) = g(\mathbf{X}^T\boldsymbol{\beta})$ is derived from a single-index model with any link function $g(\cdot) : \mathbb{R} \to \mathbb{R}$), our Semi-supervised D-learning method based on semi-parametric imputation can induce asymptotically optimal estimators $\hat{\boldsymbol{\beta}}_{sp}$ and $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$. When $\nu(\mathbf{X};\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) \neq m(\mathbf{X})$, we could construct a weighted version estimator to guarantee the efficiency of our Semi-supervised method, which will be discussed in Section 5.1 more detailed. It is worth noting that the kernel method with dimension reduction are only one option for estimating the imputation function. Other methods, such as the KRLS method mentioned in Remark 1, can also be used to estimate the imputation function. Hainmueller and Hazlett[31] proved that the KRLS estimator is consistent and asymptotically normal under some regularity conditions. Therefore, the asymptotic normality described by this theorem still holds, with a little difference in the order of the remainder term $r_{n,N}$ under the KRLS method. So our proposed semi-supervised method is flexible and adaptive of the imputation function estimation method.

Then we provide the variance estimation procedure of $\hat{\boldsymbol{\beta}}_{sp}$ and $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ for inference. By the influence function derived before, we can simply estimate Σ by its consistent estimator that

$$\hat{\Sigma} = \hat{E}[\mathbf{X}\mathbf{X}^T]^{-1}\hat{E}\{\mathbf{X}\mathbf{X}^T[2AY - \hat{\nu}]^2\}\hat{E}[\mathbf{X}\mathbf{X}^T]^{-1}.$$

Note that $\hat{E}[\mathbf{X}\mathbf{X}^T]$ could be estimated based on $\mathcal{L}$, $\mathcal{U}$ or $\mathcal{L}\cup\mathcal{U}$ while $\hat{E}\{\mathbf{X}\mathbf{X}^T[2AY - \hat{\nu}]^2\}$ is estimated only based on $\mathcal{L}$. For example, $\hat{E}[\mathbf{X}\mathbf{X}^T] = P_n\mathbf{X}\mathbf{X}^T$ based on the labeled data $\mathcal{L}$. Here the estimator of imputation function $\hat{\nu}$ is equal to $\hat{\nu}(\cdot;\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\theta}})$ for $\hat{\boldsymbol{\beta}}_{sp}$ and $\hat{\nu}(\cdot;\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\theta}}_{\mathbb{K}})$ for $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ respectively.

We also can adopt the double CV procedure[30,34] to estimate the variance of $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ which aimed to overcome the over-fitting problem in variance estimation. We continue to use the notation of sample splitting for CV in Section 2.2 here. Specifically, we first obtain $\hat{\boldsymbol{\theta}}_k$ by solve the equation $\sum_{k'\neq k}\sum_{i\in\mathcal{I}_{k'}}\mathbf{X}_i(2A_iY_i - \hat{m}_{k'}(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_{k'}) - \mathbf{X}_i^T\boldsymbol{\theta}) = 0$ for $k, k' \in \{1,\ldots,\mathbb{K}\}$. Then naturally $\hat{\nu}_k(\mathbf{X}_i;\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\theta}}_k)$ is estimated by $\hat{m}_k(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_k) + \mathbf{X}_i^T\boldsymbol{\theta}_k$. Then the variance estimation based on double CV is obtained by

$$\hat{\Sigma}_{\mathbb{K}} = \hat{E}[\mathbf{X}\mathbf{X}^T]^{-1}\left\{\frac{1}{n}\sum_{k=1}^{\mathbb{K}}\sum_{i\in\mathcal{I}_k}\mathbf{X}_i\mathbf{X}_i^T[2A_iY_i - \hat{\nu}_k(\mathbf{X}_i;\hat{\boldsymbol{\beta}}_k,\hat{\boldsymbol{\theta}}_k)]^2\right\}\hat{E}[\mathbf{X}\mathbf{X}^T]^{-1}.$$

For comparison, we will report both $\hat{\Sigma}$ and $\hat{\Sigma}_{\mathbb{K}}$ for $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ in the next numerical simulation section.

4 | NUMERICAL RESULTS

4.1 | Monte Carlo Simulations

In this section, we conduct a series of simulation studies to evaluate the finite-sample performance of the estimators introduced in Section 2, including 'SUP' (the supervised OLS estimator $\hat{\boldsymbol{\beta}}$), 'NP' (Semi-supervised estimator $\hat{\boldsymbol{\beta}}_{np}$ based on fully non-parametric imputation), 'SP' (Semi-supervised estimator $\hat{\boldsymbol{\beta}}_{sp}$ based on semiparametric imputation), 'SP.CV' ('SP' estimator with 10-fold CV), 'KRLS' (Semi-supervised estimator $\hat{\boldsymbol{\beta}}_{KRLS}$ based on KRLS imputation) and 'KRLS.CV' ('KRLS' estimator with 10-fold CV). The tuning parameters are chosen via least square CV such as the bandwidth selection in imputation function estimation by the R function 'npregbw' in 'np' package. We generate data by

$$Y = \mu_0(\mathbf{X}) + A\delta(\mathbf{X}) + e,$$

with the error term $e \sim N(0, 1)$. And treatment A is generated as -1 and 1 with equal probability 0.5 under the RCT scenario. We compare the methods based on 500 simulation runs with varying covariates dimension $p = 4$ and 20 respectively. Covariates $X_1, \ldots, X_p$ are generated independently and identically distributed from a uniform distribution on [–5, 5]. We fix the sample size of $\mathcal{L}$ as $n = 500$ and $\mathcal{U}$ as $N = 10000$. Set the true value of parameter α as $\alpha = \alpha^{(a)} = (0, -\mathbf{1}_{p/2}^T, \mathbf{1}_{p/2}^T)^T$ or $\alpha = \alpha^{(b)} = (0, \mathbf{1}_p^T)^T$, and consider the following cases for $\delta(\mathbf{X})$ that:

- Linear (Lin): $\delta(\mathbf{X}) = 20\mathbf{X}^T\alpha$,
- Nonlinear 1 (NL1): $\delta(\mathbf{X}) = 0.2(\mathbf{X}^T\alpha)^3$,
- Nonlinear 2 (NL2): $\delta(\mathbf{X}) = \mathbf{X}^T\alpha + 0.2(\mathbf{X}^T\alpha)^3 + \sin(\mathbf{X}^T\alpha)$.

Under the linear scenario 'Lin', the optimal decision function $f^{\text{opt}}(\mathbf{X}) = 2\delta(\mathbf{X}) \in \mathcal{F}$, the linear decision class, thus $\boldsymbol{\beta}_0 = 40\alpha$. Otherwise, under the nonlinear scenarios 'NL1' and 'NL2', the true value of the parameter $\boldsymbol{\beta}$ indexing the linear ITR are estimated by simulating a fully observed Monte Carlo data set of size 1,000,000. To verify that our Semi-supervised D-learning methods are insensitive to the form of treatment-free effect, we study different settings of $\mu_0(\mathbf{X})$ for each scenario that

- Quadratic: $\mu_0^Q(\mathbf{X}) = \mathbf{X}^T\omega_1 + (\mathbf{X}^T\omega_2)^2$,
- Cubic: $\mu_0^C(\mathbf{X}) = 0.1(\mathbf{X}^T\omega_2)^3$,

with $\omega_1 = (0, \mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T)^T$ and $\omega_2 = (0, \mathbf{1}_p^T)^T$.

We next provide a detailed report on the bias (Bias), the standard deviation (SD), and relative efficiency (RE) of the parameter estimates for each dimension under different methods. The RE refers to the ratio of the mean squared error (MSE) of the OLS estimator 'SUP' to that of the comparative method, i.e.,

$$\text{RE} = \frac{\text{MSE of SUP}}{\text{MSE of comparative method}},$$

across various scenarios. The percentage of making correct decision (PCD) is calculated by

$$\text{PCD} = \frac{1}{n}\sum_{i=1}^{n} I\left(\text{sign}(\delta(\mathbf{X}_i))\text{sign}(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_\cdot) > 0\right),$$

where $\hat{\boldsymbol{\beta}}_\cdot$ represents the estimator such as $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{np}, \hat{\boldsymbol{\beta}}_{sp}$ and $\hat{\boldsymbol{\beta}}_{KRLS}$ that corresponding to the specific method introduced at the beginning of this section.

When $p = 4$, we report the Bias, SD and RE of the parameter estimators for each dimension, as well as the PCD, averaged over 500 simulation runs. The results are presented in Tables 1 through 3. As shown in Table 1, when $\delta(\mathbf{X})$ follows a linear structure, all the Semi-supervised methods do not demonstrate significant advantages over the supervised method. In this case, the REs are close to 1, except for the estimator based on fully non-parametric imputation. In contrast, when $\delta(\mathbf{X})$ follows a non-linear structure, indicating model misspecification for the linear decision function class, the proposed Semi-supervised methods based on semiparametric or KRLS imputation show considerable improvement, as shown in Table 2 and Table 3. Specifically, these methods double the REs compared to the supervised method. Among these, the Semi-supervised method

based on semiparametric imputation yields a slightly higher REs than the KRLS-based method. Notably, the estimators based on fully non-parametric imputation perform poorly across all three scenarios, due to the significant bias introduced by the imputation estimator obtained using a multi-dimensional kernel approach. Additionally, as the dimensionality increases, the issue of the 'curse of dimensionality' becomes more pronounced, as will be further demonstrated in the following simulations.

We then verify the performance of the variance estimation procedures proposed in section 3. Recall that $\hat{\Sigma} = \hat{E}[\mathbf{X}\mathbf{X}^T]^{-1}\hat{E}\{\mathbf{X}\mathbf{X}^T[2AY-\hat{\nu}]^2\}\hat{E}[\mathbf{X}\mathbf{X}^T]^{-1}$, where $\hat{E}[\mathbf{X}\mathbf{X}^T]$ is estimated by $P_n\mathbf{X}\mathbf{X}^T$. Denote the variance estimation method as 'SP', 'SP.CV' and 'SP.DCV' when the imputation function $\hat{\nu}$ is estimated without CV, with 10-fold CV and with 10-fold double CV respectively. Table 4 reports the average of the estimated standard error (SE) corresponding 95% empirical coverage probability (CP) for each dimension under all the 12 settings when $p = 4$.

As shown in Table 4, compared to the estimators obtained by 'SP' and 'SP.CV', that by the double CV method 'SP.DCV' performs better in the sense that the SEs are closer to the SDs and the CPs are closer to the nominal level. And the same performance is also seen in the following simulation with p=10. Therefore, we recommend using the double CV procedure for inference in applications.

We next show the similar results when $p = 10$. Due to space limitations, we no longer present the results for each dimension separately, but instead present the average of the statistics for each dimension, including RE and PCD in Table 5, Bias and SD in Table 6 as well as SE and CP in Table 7.

As the results shown in Table 5 to Table 7, when the sample size is fixed, the performance of the 'NP' estimators deteriorates as the dimension p of the covariates increases. Specifically, when p increases from 4 to 10, the REs of the 'NP' estimators decrease from approximately half of the 'SUP' estimators' REs to about one-fifth, and the PCDs become significantly lower than those of the 'SUP' estimators. Furthermore, when $p = 10$, the Semi-supervised methods based on KRLS imputation performs notably worse than the Semi-supervised methods based on semiparametric imputation, which we propose. In the case with model misspecification, the REs of the 'KRLS' estimators without CV even fall below that of the supervised method, despite having higher PCDs. Even with CV technique incorporated, the 'KRLS.CV' estimators' REs are at most 1.5 times that of the supervised method. In contrast, our semiparametric-based estimation methods, both with and without CV, perform well. Their REs are consistently higher than that of 'KRLS.CV', even three times higher than that of supervised methods in some settings, and their PCDs are also significantly improved compared to the supervised method. Overall, the Semi-supervised D-learning methods based on semiparametric imputation that we propose remain robust even when the covariates are multi-dimensional. When $\delta(\mathbf{X})$ follows a linear structure, our methods achieve performance comparable to the supervised method, while in the case of a non-linear structure, our methods are notably more accurate and effective.

4.2 | A Real Data Application

The data utilized in this analysis were extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database, a comprehensive collection of de-identified health-related information from patients admitted to the emergency department or intensive care units (ICUs) at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, MA, between 2008 and 2022 [35,36]. The MIMIC-IV database includes data on over 65,000 ICU patients and more than 200,000 emergency department patients. The version used in this study, MIMIC-IV Version 3.1, was made publicly available on October 11, 2024. The dataset encompasses a wide range of clinical variables, including patient demographics, diagnoses, laboratory results, vital signs, medications, and outcomes, making it an invaluable resource for precision medicine research. Data extraction for this study was performed following the completion of the online training course required for access to MIMIC-IV. Upon successful completion of the course, the author obtained access to the database (Record ID: 67005348) with the approval of the Institutional Review Board.

Sepsis is a life-threatening condition caused by a dysregulated host response to infection and remains a leading cause of morbidity and mortality in critically ill patients [37,38]. Assessing tissue perfusion and guiding resuscitation are key challenges in sepsis management, with lactate serving as an important biomarker for hypoxia and circulatory failure [39,40]. Among the MIMIC-IV database, 9,052 adult patients with sepsis who were first admitted to the ICU were selected for this study. The extraction depended on the International Classification of Diseases, Ninth revision (ICD-9) and International Classification of Diseases, Tenth revision (ICD-10) codes recorded in the database (ICD-9 codes 99591, 99592 or ICD-10 codes A419, R6520 and R6521). The outcome we interested in is lactate clearance within 48 hours of ICU admission, calculated as

$$\text{Lactate Clearance Rate} = \frac{\text{Initial Lactate Level} - \text{Final Lactate Level}}{\text{Initial Lactate Level}} \times 100\%,$$

where 'Initial Lactate Level' is the first lactate measurement after the patient enters the hospital and before entering the ICU, and 'Final Lactate Level' is the last lactate measurement within 48 hours after the patient enters the ICU. Elevated lactate levels, particularly when not cleared within the first 48 hours, are strongly linked to poor outcomes[41]. Lactate clearance, as a dynamic marker, offers a better indication of recovery compared to static lactate levels. Thus we focus on lactate clearance 48 hours after ICU admission, a critical period for assessing resuscitation effectiveness[42,43]. The two treatment strategies compared in this analysis are vasopressor therapy and intravenous fluid resuscitation[44]. Vasopressor therapy, which includes vasopressin (itemid=222315), epinephrine (itemid=221289), dopamine (itemid=221662), phenylephrine (itemid=221749) and norepinephrine (itemid=221906), is coded as A = 1. In contrast, intravenous fluid resuscitation, which consists of 0.9% saline (itemid=220954), Ringer's lactate (itemid=220955) and albumin (itemid=220861, 220862, 220863, 220864), is coded as A = -1. The main covariates we selected in the analysis includes age (years), admission weight (kilograms), blood urea nitrogen (BUN) amount (mg/dL), creatinine (mg/dL), white blood cell count (WBC) count (K/μL), and heart rate (HR) (bpm). After excluding incomplete and abnormal records, such as the patient's weight less than 30 kg or greater than 300 kg, WBC less than 0.5 K/μL or greater than 200 K/μL, HR less than 30 bpm or greater than 200 bpm, and balancing the treatment groups, the labeled dataset finally includes n=184 samples, while the unlabeled dataset contains N=7,623 samples.

The KolmogorovSmirnov (KS) test is employed to examine whether the underlying probability distributions of covariates in the labeled and unlabeled datasets differ. Table 8 presents the median, mean, and standard deviation (SD) of each covariate, and demonstrates that the MCAR assumption holds in this analysis, as the p-values from the KS test are all greater than 0.05 for the six covariates.

In the numerical simulations, the covariates are generated from a $U[-5, 5]$ distribution. Accordingly, we transform the covariates to lie within the interval $[-5, 5]$ and asymptotically follow the uniform distribution before proceeding with the estimation procedure. As in the numerical simulations, a Gaussian kernel is chosen for kernel smoothing estimators and the CV-based methods employ 10-fold cross-validation, with tuning parameters selected via least squares cross-validation using the R function 'npregbw' from the 'np' package. Subsequently, we estimate the parameters of the linear decision function using the methods compared in Section 4.1. The point estimators are presented in Table 9, with the corresponding estimated standard errors (SE) for the supervised estimator and the semiparametric-based semi-supervised estimator shown in Table 10. And Table 11 summarizes the treatment recommendations generated by various methods.

From these tables, we observe that the point estimates from the supervised and semi-supervised estimators are comparable, and these methods recommend similar treatment decisions for the vast majority of patients. In contrast, the SP.CV, KRLS and KRLS.CV methods tend to recommend intravenous fluid resuscitation for a greater number of spesis patients to enhance lactate clearance. However, the parameter estimates obtained using the semi-supervised D-learning method, which leverages our proposed semi-parametric estimation of the imputation function, exhibit smaller estimated SE, which is consistent with the theoretical results and simulation performance.

5 | EXTENSIONS

5.1 | Efficiency

As described in Remark 4, our proposed estimator is asymptotically optimal when $\nu(\mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0) = m(\mathbf{X})$. Thus for the sake of efficiency, we could construct a weighted estimator $\hat{\boldsymbol{\beta}}_w$ by combining the supervised estimator $\hat{\boldsymbol{\beta}}$ and the Semi-supervised estimator $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ for fixed $\mathbb{K} \geqslant 1$ with a tuning diagonal weights matrix $\Lambda = diag(\lambda_0, \ldots, \lambda_p)$ that

$$\hat{\boldsymbol{\beta}}_w = \Lambda \hat{\boldsymbol{\beta}}_{sp,\mathbb{K}} + (I - \Lambda)\hat{\boldsymbol{\beta}},$$

with every element $\lambda_\ell \in [0, 1]$ for $\ell \in \{0, 1, \ldots, p\}$ which balancing the contributions of the labeled data and unlabeled data. Since all the estimators for $\boldsymbol{\beta}$ we derived above, including $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}_{np}$, $\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ and $\hat{\boldsymbol{\beta}}_w$, are asymptotically unbiased to $\boldsymbol{\beta}_0$, thus the optimal Λ could be selected by the criteria minimizing the asymptotic variance of $\hat{\boldsymbol{\beta}}_w$ that

$$\Lambda^{\text{opt}} = \arg\min_{\Lambda} Var(\hat{\boldsymbol{\beta}}_w). \tag{7}$$

Based on the basic knowledge of quadratic functions, we can get the explicit solution of optimal problem (7) for $\ell \in \{0, 1, \ldots, p\}$ as

$$\lambda_\ell^{\text{opt}} = -\frac{Cov(\hat{\boldsymbol{\beta}}_{[\ell]}, \hat{\boldsymbol{\beta}}_{sp,\mathbb{K}[\ell]} - \hat{\boldsymbol{\beta}}_{[\ell]})}{Var(\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}[\ell]} - \hat{\boldsymbol{\beta}}_{[\ell]})}.$$

Here for identification, we additional define $\lambda_\ell^{\text{opt}} = 0$ when $Var(\hat{\boldsymbol{\beta}}_{sp,\mathbb{K}[\ell]} - \hat{\boldsymbol{\beta}}_{[\ell]}) = 0$. Hence the weighted estimator $\hat{\boldsymbol{\beta}}_w$ is always more effective than $\hat{\boldsymbol{\beta}}$ or at least as effective as $\hat{\boldsymbol{\beta}}$, which demonstrates the superiority of our Semi-supervised approach.

5.2 | Observational Studies

In the framework presented in this paper, we assume that the data are derived from a randomized clinical trial, where the true propensity score is predetermined by the study design. However, in practical applications, data may also come from observational studies, where the true propensity score is typically unknown but can be estimated using the available data. Commonly, Probit and logistic regression models are employed to estimate the propensity score[45,46], where treatment assignment is modeled as a function of observed baseline covariates. The estimated propensity score corresponds to the predicted probability of treatment from the fitted regression model. Beyond these parametric approaches, more flexible nonparametric methods are also available, such as boosting[47,48], random forests[49,50], and neural networks[51,52]. Thus we can use the consistent estimator $\hat{\pi}(A, \mathbf{X})$ to replace the unknown $\pi(A, \mathbf{X})$ in observational studies.

6 | DISCUSSION

In summary, we propose a novel semi-supervised D-learning framework for estimating optimal individualized treatment regime. focusing directly on the decision function that maximizing the value function rather than minimizing the prediction error of value function. Unlike indirect approaches that minimize the prediction error of the value function, our method focus directly on the decision function that maximizing the value function. Theoretical analysis shows that the proposed parameter estimators converge asymptotically to a normal distribution at the $\sqrt{n}$ rate, where the rate depends solely on the labeled sample size. Notably, under misspecification of the linear decision model, the semi-supervised estimator outperforms the supervised estimator by achieving a lower asymptotic variance. Extensive simulations in Section 4 indicate that, under a correctly specified linear decision model, incorporating unlabeled samples does not improve estimation efficiency. However, when the linear decision model is misspecified, the semi-supervised estimator shows significant improvement, consistent with theoretical results. Furthermore, as the covariate dimension p increases, fully nonparametric imputation methods fail to maintain stable performance, whereas our proposed semiparametric imputation-based estimator remains robust and demonstrates superior efficiency. In particular, our method outperforms KRLS-based estimators, which exhibit inferior efficiency in lager dimensional settings. These results highlight the ability of our framework to mitigate the curse of dimensionality associated with nonparametric estimation and maintain robustness to increasing covariate dimension.

Our framework is naturally extensible to multi-class treatment settings and more complex data structures, such as censored survival data. While this study focuses on single-stage optimal ITR, future research could explore extensions to dynamic treatment regimes (DTRs) for chronic diseases. Developing semi-supervised frameworks for these scenarios presents a promising direction for further investigation.

ACKNOWLEDGMENTS

The authors thank the co-editor, the associate editor and reviewers for their helpful suggestions. Zhou's work was supported by the National Key R&D Program of China (2021YFA1000100, 2021YFA1000101, 2021YFA1000104) and Science and Technology Commission of Shanghai Municipality (23JS1400500). Zhang's work was supported by the National Natural Science Foundation of China (72201101, 72331005, 12271171), and Youth Fund of General Project of MoE (Ministry of Education) of Humanities and Social Sciences (22YJC910013).

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publishers website.

REFERENCES

1. Kosorok MR, Laber EB. Precision Medicine. *Annual Review of Statistics and Its Application.* 2019;6(1):263–286.
2. Mo W, Liu Y. A Selective Review of Individualized Decision Making. *Statistics in Precision Health: Theory, Methods and Applications.* 2024:13–39.
3. Watkins C. Learning from Delayed Rewards. *PhD thesis, Cambridge University, Cambridge, England.* 1989.
4. Watkins C, Dayan P. Q-learning. *Machine Learning.* 1992;8(3):279–292.
5. Qian M, Murphy SA. Performance guarantees for individualized treatment rules. *The Annals of statistics.* 2011;39(2):1180–1210.
6. Song R, Wang W, Zeng D, Kosorok MR. Penalized q-learning for dynamic treatment regimens. *Statistica Sinica.* 2015;25(3):901.
7. Murphy SA. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 2003;65(2):331–355.
8. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11(5):550–560.
9. Kang S, Lu W, Zhang J. On estimation of the optimal treatment regime with the additive hazards model. *Statistica Sinica.* 2018;28(3):1539.
10. Shi C, Fan A, Song R, Lu W. High-dimensional A-learning for optimal dynamic treatment regimes. *The Annals of Statistics.* 2018;46(3):925–957.
11. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association.* 2012;107(499):1106–1118.
12. Chen G, Zeng D, Kosorok MR. Personalized dose finding using outcome weighted learning. *Journal of the American Statistical Association.* 2016;111(516):1509–1521.
13. Liu Y, Wang Y, Kosorok MR, Zhao Y, Zeng D. Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in medicine.* 2018;37(26):3776–3788.
14. Huang X, Goldberg Y, Xu J. Multicategory individualized treatment regime using outcome weighted learning. *Biometrics.* 2019;75(4):1216–1227.
15. Qi Z, Liu Y. D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics.* 2018;12(2):3601–3638.
16. Qi Z, Liu D, Fu H, Liu Y. Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *Journal of the American Statistical Association.* 2020;115(530):678–691.
17. Meng H, Qiao X. Augmented direct learning for conditional average treatment effect estimation with double robustness. *Electronic Journal of Statistics.* 2022;16(1):3523–3560.
18. Shah KS, Fu H, Kosorok MR. Stabilized direct learning for efficient estimation of individualized treatment rules. *Biometrics.* 2023;79(4):2843–2856.
19. Fan C, Lu W, Song R, Zhou Y. Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 2017;79(5):1565–1582.
20. Chu J, Lu W, Yang S. Targeted optimal treatment regime learning using summary statistics. *Biometrika.* 2023;110(4):913–931.
21. Li C, Zeng D, Zhu W. A robust covariate-balancing method for learning optimal individualized treatment regimes. *Biometrika.* 2025;112(1):asae036.
22. Maronge JM, D HULING J, Chen G. A reluctant additive model framework for interpretable nonlinear individualized treatment rules. *The annals of applied statistics.* 2023;17(4):3384.
23. Song R, Luo S, Zeng D, Zhang HH, Lu W, Li Z. Semiparametric single-index model for estimating optimal individualized treatment strategy. *Electronic journal of statistics.* 2017;11(1):364.
24. Zhang A, Brown LD, Cai TT. Semi-supervised inference: General theory and estimation of means. *The Annals of Statistics.* 2019;47(5):2538–2566.
25. Cheng D, Ananthakrishnan AN, Cai T. Robust and efficient semi-supervised estimation of average treatment effects with application to electronic health records data. *Biometrics.* 2021;77(2):413–423.
26. Tony Cai T, Guo Z. Semisupervised inference for explained variance in high dimensional linear regression and its applications. *Journal of the Royal Statistical Society Series B: Statistical Methodology.* 2020;82(2):391–419.
27. Chakrabortty A, Dai G, Tchetgen ET. A general framework for treatment effect estimation in semi-supervised and high dimensional settings. *arXiv preprint arXiv:2201.00468.* 2022.
28. Zhang Y, Bradic J. High-dimensional semi-supervised learning: in search of optimal inference of the mean. *Biometrika.* 2022;109(2):387–403.
29. Sonabend-W A, Laha N, Ananthakrishnan AN, Cai T, Mukherjee R. Semi-supervised off-policy reinforcement learning and value estimation for dynamic treatment regimes. *Journal of Machine Learning Research.* 2023;24(323):1–86.
30. Gunn K, Lu W, Song R. Adaptive semi-supervised inference for optimal treatment decisions with electronic medical record data. In: , , Springer, 2024:229–246.
31. Hainmueller J, Hazlett C. Kernel regularized least squares: Reducing misspecification bias with a flexible and interpretable machine learning approach. *Political Analysis.* 2014;22(2):143–168.
32. Tsiatis AA. *Semiparametric Theory and Missing Data*. Springer Series in StatisticsSpringer New York, 2007.
33. Seeger M. Learning with labeled and unlabeled data. *Technical Report, The University of Edinburgh.* 2000.
34. Chakrabortty A, Cai T. Efficient and adaptive linear regression in semi-supervised settings. *The Annals of Statistics.* 2018;46(4):1541–1572.
35. Johnson AE, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific data.* 2023;10(1):1.
36. Johnson A, Bulgarelli L, Pollard T, et al. MIMIC-IV (version 3.1). PhysioNet. 2024.
37. Fleischmann C, Scherag A, Adhikari NK, et al. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *American journal of respiratory and critical care medicine.* 2016;193(3):259–272.
38. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA: Journal of the American Medical Association.* 2016;315(8):801–810.

39. Jones AE, Shapiro NI, Trzeciak S, et al. Lactate Clearance vs Central Venous Oxygen Saturation as Goals of Early Sepsis Therapy: A Randomized Clinical Trial. *JAMA: Journal of the American Medical Association.* 2010;303(8):739–746.
40. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Critical care medicine.* 2013;41(2):580–637.
41. Nguyen HB, Rivers EP, Knoblich BP, et al. Early lactate clearance is associated with improved outcome in severe sepsis and septic shock. *Critical care medicine.* 2004;32(8):1637–1642.
42. Chertoff J, Chisum M, Simmons L, King B, Walker M, Lascano J. Prognostic utility of plasma lactate measured between 24 and 48 h after initiation of early goal-directed therapy in the management of sepsis, severe sepsis, and septic shock. *Journal of intensive care.* 2016;4:1–8.
43. Lu S, Sun J, Tong C. Prognostic value of lactic acid and 48-hour lactate clearance rate in patients with sepsis. *Chinese Journal of Clinic Medicine.* 2023;30(1):91-96. doi: 10.12025/j.issn.1008-6358.2023.20220972
44. Zampieri FG, Bagshaw SM, Semler MW. Fluid Therapy for Critically Ill Adults With Sepsis: A Review.. *JAMA: Journal of the American Medical Association.* 2023;329(22).
45. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys.* 2008;22(1):31–72.
46. Mo W, Liu Y. 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.* 2022;84(2):440–472.
47. McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies.. *Psychological methods.* 2004;9(4):403.
48. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Statistics in medicine.* 2010;29(3):337–346.
49. Breiman L. Random forests. *Machine learning.* 2001;45:5–32.
50. Athey S, Tibshirani J, Wager S. Generalized random forests. *The Annals of Statistics.* 2019;47(2):1148–1178.
51. Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and drug safety.* 2008;17(6):546–555.
52. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of clinical epidemiology.* 2010;63(8):826–833.

TABLE 1 Simulation results for Lin scenario with $p = 4$

Method	$\alpha = \alpha^{(a)}$						$\alpha = \alpha^{(b)}$					
	Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$			Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$		
SUP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1	0.304	4.971	1	-0.171	5.698	1	0.304	4.971	1	-0.171	5.698
β_1	1	0.009	1.954	1	0.097	2.409	1	0.009	1.954	1	0.097	2.409
β_2	1	0.006	2.071	1	0.144	2.522	1	0.006	2.071	1	0.144	2.533
β_3	1	-0.183	2.017	1	0.090	2.600	1	-0.183	2.017	1	0.090	2.600
β_4	1	-0.073	2.051	1	0.209	2.550	1	-0.073	2.051	1	0.209	2.550
PCD		98.43%			98.01%			98.91%			98.88%	
NP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	0.97	0.245	5.055	1.05	-0.195	5.555	1.01	0.171	4.960	1.06	-0.204	5.528
β_1	0.51	1.708	2.132	0.49	2.195	2.637	0.52	-1.656	2.144	0.47	-2.067	2.860
β_2	0.55	1.698	2.218	0.47	2.155	2.970	0.56	-1.695	2.198	0.57	-1.919	2.731
β_3	0.49	-1.835	2.232	0.52	-2.009	2.985	0.51	-1.833	2.174	0.56	-2.014	2.850
β_4	0.49	-1.779	2.318	0.51	-2.012	2.965	0.52	-1.735	2.242	0.53	-1.998	2.876
PCD		98.22%			97.65%			98.65%			98.35%	
SP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	0.98	0.324	5.018	0.98	-0.133	5.763	0.92	0.369	5.165	1.01	-0.163	5.656
β_1	1.00	0.141	1.954	0.99	0.224	2.414	1.01	-0.124	1.942	0.99	0.031	2.428
β_2	0.98	0.120	2.092	0.99	0.267	2.523	1.02	-0.118	2.050	1.03	0.049	2.484
β_3	0.98	-0.309	2.028	0.99	-0.034	2.619	1.02	-0.288	1.986	1.00	-0.006	2.605
β_4	0.98	-0.210	2.068	0.99	0.075	2.571	0.98	-0.196	2.063	1.01	0.114	2.543
PCD		98.42%			98.00%			98.90%			98.90%	
SP.CV	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	0.97	0.332	5.049	0.96	-0.139	5.807	0.88	0.475	5.289	0.91	-0.234	5.975
β_1	1.00	0.013	1.951	0.99	0.090	2.425	0.92	-0.002	2.035	0.84	0.262	2.620
β_2	0.97	-0.009	2.101	1.00	0.136	2.520	0.92	0.008	2.156	0.88	0.281	2.679
β_3	0.99	-0.184	2.028	0.98	0.106	2.623	0.94	-0.166	2.083	0.86	0.230	2.803
β_4	0.98	-0.077	2.075	0.97	0.214	2.590	0.91	-0.062	2.154	0.85	0.354	2.755
PCD		98.42%			98.01%			98.86%			98.85%	
KRLS	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	0.86	0.296	5.349	0.93	-0.294	5.890	0.91	0.376	5.206	0.94	-0.208	5.862
β_1	0.93	0.266	2.012	0.97	0.402	2.416	0.92	-0.214	2.024	0.96	-0.094	2.457
β_2	0.97	0.228	2.087	0.95	0.419	2.558	0.96	-0.226	2.107	0.98	-0.048	2.558
β_3	0.93	-0.408	2.064	0.94	-0.139	2.673	0.94	-0.346	2.063	0.96	-0.065	2.658
β_4	0.95	-0.319	2.082	0.95	-0.067	2.627	0.95	-0.302	2.088	1.00	-0.035	2.558
PCD		98.39%			97.97%			98.84%			98.84%	
KRLS.CV	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	0.84	0.358	5.431	0.88	-0.306	6.075	0.88	0.439	5.284	0.88	-0.214	6.056
β_1	0.89	0.038	2.075	0.91	0.185	2.516	0.88	0.012	2.084	0.89	0.165	2.554
β_2	0.92	0.002	2.153	0.89	0.204	2.673	0.90	0.007	2.183	0.89	0.221	2.671
β_3	0.91	-0.185	2.113	0.87	0.144	2.780	0.92	-0.130	2.112	0.88	0.211	2.762
β_4	0.92	-0.079	2.138	0.87	0.222	2.737	0.91	-0.073	2.145	0.91	0.244	2.677
PCD		98.37%			97.93%			98.83%			98.79%	

TABLE 2 Simulation results for NL1 scenario with $p = 4$

Method	$\alpha = \alpha^{(a)}$						$\alpha = \alpha^{(b)}$					
	Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$			Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$		
SUP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1	0.470	8.314	1	-0.078	8.429	1	0.186	8.167	1	-0.362	8.627
β_1	1	0.314	3.240	1	0.391	3.474	1	-0.026	3.319	1	0.051	3.611
β_2	1	0.189	3.326	1	0.255	3.492	1	0.094	3.395	1	0.160	3.496
β_3	1	-0.314	3.433	1	-0.031	3.855	1	-0.125	3.247	1	0.158	3.604
β_4	1	-0.353	3.513	1	-0.007	3.919	1	-0.041	3.478	1	0.304	3.683
PCD		97.61%			97.27%			97.99%			98.01%	
NP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1.96	0.287	5.941	1.90	-0.212	6.110	1.83	0.005	6.032	1.82	-0.507	6.380
β_1	0.68	3.191	2.326	0.64	3.477	2.664	0.65	-3.225	2.540	0.63	-3.399	3.021
β_2	0.70	3.214	2.375	0.61	3.453	2.863	0.71	-3.103	2.559	0.62	-3.328	2.950
β_3	0.70	-3.329	2.448	0.74	-3.322	3.017	0.59	-3.287	2.650	0.62	-3.332	3.170
β_4	0.73	-3.347	2.443	0.76	-3.325	3.006	0.69	-3.291	2.613	0.71	-3.212	2.989
PCD		97.94%			97.50%			98.11%			98.01%	
SP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	2.47	0.256	5.288	2.00	-0.253	5.963	2.26	0.065	5.434	2.15	-0.553	5.865
β_1	2.14	0.701	2.114	1.72	0.780	2.551	2.38	-0.604	2.063	1.88	-0.417	2.604
β_2	2.04	0.743	2.211	1.60	0.810	2.650	2.32	-0.502	2.172	1.78	-0.364	2.595
β_3	2.14	-0.842	2.198	1.79	-0.585	2.817	2.07	-0.781	2.118	1.69	-0.458	2.736
β_4	2.13	-0.883	2.253	1.93	-0.556	2.767	2.14	-0.716	2.265	1.85	-0.307	2.696
PCD		98.12%			97.66%			98.69%			98.68%	
SP.CV	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	2.31	0.249	5.475	1.88	-0.279	6.146	2.04	0.160	5.724	1.76	-0.710	6.464
β_1	2.29	-0.006	2.150	1.86	0.070	2.559	2.15	0.114	2.259	1.50	0.438	2.920
β_2	2.15	0.040	2.271	1.74	0.094	2.650	2.02	0.228	2.382	1.41	0.498	2.910
β_3	2.31	-0.140	2.263	1.83	0.150	2.845	1.94	-0.072	2.335	1.36	0.399	3.072
β_4	2.31	-0.172	2.317	1.91	0.164	2.832	1.95	0.029	2.488	1.44	0.571	3.032
PCD		98.20%			97.75%			98.64%			98.59%	
KRLS	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1.93	0.297	5.981	1.80	-0.301	6.279	2.00	0.096	5.779	1.92	-0.478	6.213
β_1	1.97	0.624	2.233	1.69	0.750	2.580	2.14	-0.478	2.220	1.78	-0.347	2.682
β_2	2.01	0.641	2.263	1.55	0.742	2.712	2.13	-0.377	2.294	1.66	-0.264	2.704
β_3	2.01	-0.731	2.320	1.66	-0.438	2.961	1.85	-0.607	2.310	1.63	-0.298	2.808
β_4	2.01	-0.809	2.356	1.78	-0.494	2.897	2.03	-0.598	2.366	1.78	-0.260	2.758
PCD		98.13%			97.69%			98.57%			98.62%	
KRLS.CV	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1.85	0.374	6.116	1.64	-0.318	6.573	1.90	0.178	5.917	1.75	-0.507	6.501
β_1	1.95	0.065	2.328	1.64	0.234	2.719	2.04	0.099	2.320	1.64	0.270	2.805
β_2	1.98	0.085	2.368	1.47	0.213	2.885	1.99	0.216	2.397	1.47	0.354	2.863
β_3	2.02	-0.183	2.420	1.52	0.180	3.124	1.83	-0.038	2.401	1.48	0.325	2.953
β_4	2.01	-0.237	2.479	1.63	0.124	3.071	1.98	-0.008	2.473	1.56	0.363	2.934
PCD		98.08%			97.61%			98.55%			98.57%	

TABLE 3 Simulation results for NL2 scenario with $p = 4$

Method	$\alpha = \alpha^{(a)}$						$\alpha = \alpha^{(b)}$					
	Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$			Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$		
SUP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1	0.470	8.320	1	-0.077	8.433	1	0.187	8.169	1	-0.360	8.625
β_1	1	0.313	3.239	1	0.391	3.475	1	-0.027	3.319	1	0.050	3.610
β_2	1	0.189	3.326	1	0.255	3.493	1	0.094	3.395	1	0.159	3.496
β_3	1	-0.315	3.433	1	-0.032	3.853	1	-0.124	3.245	1	0.160	3.604
β_4	1	-0.352	3.515	1	-0.007	3.920	1	-0.041	3.479	1	0.304	3.682
PCD		97.73%			97.42%			98.10%			98.12%	
NP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1.95	0.290	5.957	1.90	-0.244	6.110	1.83	0.007	6.043	1.81	-0.490	6.402
β_1	0.65	3.285	2.342	0.61	3.560	2.693	0.63	-3.310	2.572	0.61	-3.488	3.032
β_2	0.67	3.299	2.389	0.58	3.556	2.899	0.68	-3.183	2.594	0.59	-3.434	2.995
β_3	0.67	-3.426	2.463	0.70	-3.448	3.050	0.57	-3.385	2.670	0.59	-3.419	3.218
β_4	0.70	-3.433	2.456	0.74	-3.409	3.006	0.65	-3.393	2.645	0.67	-3.329	3.026
PCD		98.04%			97.63%			98.20%			98.10%	
SP	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	2.48	0.263	5.289	2.00	-0.248	5.961	2.26	0.055	5.433	2.15	-0.552	5.866
β_1	2.14	0.702	2.112	1.73	0.780	2.546	2.39	-0.604	2.062	1.88	-0.421	2.601
β_2	2.04	0.744	2.209	1.60	0.812	2.649	2.33	-0.502	2.170	1.78	-0.369	2.594
β_3	2.15	-0.844	2.194	1.79	-0.584	2.817	2.08	-0.780	2.115	1.69	-0.461	2.735
β_4	2.14	-0.885	2.247	1.94	-0.554	2.763	2.15	-0.715	2.262	1.86	-0.309	2.694
PCD		98.23%			97.79%			98.76%			98.75%	
SP.CV	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	2.31	0.255	5.477	1.88	-0.276	6.145	2.04	0.143	5.715	1.76	-0.706	6.463
β_1	2.29	-0.004	2.151	1.87	0.074	2.557	2.15	0.110	2.260	1.50	0.434	2.917
β_2	2.15	0.043	2.272	1.74	0.098	2.650	2.01	0.225	2.384	1.41	0.493	2.908
β_3	2.32	-0.142	2.260	1.83	0.148	2.847	1.93	-0.075	2.335	1.36	0.395	3.072
β_4	2.32	-0.173	2.312	1.92	0.165	2.828	1.95	0.025	2.492	1.44	0.569	3.030
PCD		98.28%			97.87%			98.72%			98.67%	
KRLS	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1.93	0.296	5.999	1.79	-0.305	6.290	1.99	0.098	5.790	1.92	-0.470	6.216
β_1	1.96	0.637	2.237	1.68	0.762	2.585	2.12	-0.490	2.225	1.78	-0.360	2.685
β_2	2.00	0.653	2.266	1.54	0.753	2.716	2.12	-0.390	2.297	1.65	-0.278	2.708
β_3	1.99	-0.745	2.326	1.65	-0.452	2.965	1.84	-0.619	2.312	1.63	-0.307	2.813
β_4	2.00	-0.822	2.362	1.77	-0.505	2.901	2.02	-0.610	2.372	1.77	-0.273	2.760
PCD		98.23%			97.80%			98.66%			98.69%	
KRLS.CV	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD	RE	Bias	SD
c_0	1.84	0.373	6.135	1.64	-0.321	6.585	1.90	0.181	5.927	1.75	-0.497	6.506
β_1	1.94	0.066	2.333	1.64	0.235	2.724	2.04	0.100	2.324	1.64	0.270	2.809
β_2	1.97	0.084	2.373	1.46	0.212	2.889	1.99	0.215	2.399	1.47	0.353	2.867
β_3	2.01	-0.184	2.426	1.51	0.178	3.129	1.83	-0.039	2.402	1.47	0.329	2.959
β_4	2.00	-0.237	2.486	1.62	0.125	3.075	1.97	-0.009	2.477	1.56	0.363	2.937
PCD		98.18%			97.74%			98.64%			98.65%	

TABLE 4 SE and CP of proposed estimators with $p = 4$

SE(CP %)	Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$		
	SP	SP.CV	SP.DCV	SP	SP.CV	SP.DCV
			Lin $\alpha^{(a)}$			
c_0	4.849 (93.8)	5.001 (95.2)	5.087 (95.6)	5.581 (94.2)	5.749 (95.0)	5.853 (95.6)
β_1	1.962 (95.2)	2.026 (96.2)	2.062 (96.2)	2.382 (93.2)	2.455 (94.6)	2.499 (95.4)
β_2	1.960 (93.2)	2.024 (93.8)	2.060 (94.0)	2.376 (94.0)	2.449 (94.6)	2.494 (94.8)
β_3	1.964 (92.4)	2.028 (93.2)	2.064 (93.4)	2.382 (92.4)	2.456 (93.8)	2.501 (94.2)
β_4	1.966 (92.0)	2.030 (93.4)	2.066 (94.4)	2.386 (93.4)	2.459 (93.4)	2.504 (93.6)
			Lin $\alpha^{(b)}$			
c_0	4.565 (90.6)	5.012 (92.8)	5.095 (93.4)	5.037 (90.6)	5.710 (92.8)	5.810 (93.0)
β_1	1.818 (92.8)	2.036 (93.6)	2.072 (94.4)	2.105 (87.2)	2.435 (89.6)	2.479 (90.4)
β_2	1.818 (91.2)	2.032 (93.0)	2.068 (93.6)	2.102 (89.0)	2.428 (91.2)	2.471 (91.6)
β_3	1.819 (91,6)	2.037 (93.4)	2.073 (93.6)	2.103 (84.4)	2.436 (88.0)	2.480 (88.4)
β_4	1.817 (90.6)	2.034 (92.8)	2.070 (93.0)	2.104 (86.8)	2.434 (89.8)	2.478 (90.6)
			NL1 $\alpha^{(a)}$			
c_0	4.882 (93.0)	5.234 (94.0)	5.324 (94.0)	5.632 (93.2)	5.973 (93.8)	6.080 (94.0)
β_1	1.976 (92.6)	2.150 (95.6)	2.189 (95.6)	2.402 (92.0)	2.567 (95.0)	2.614 (95.4)
β_2	1.975 (89.4)	2.151 (94.6)	2.189 (95.0)	2.397 (91.0)	2.563 (94.0)	2.610 (94.2)
β_3	1.977 (91.0)	2.156 (93.4)	2.196 (94.0)	2.402 (90.8)	2.571 (92.6)	2.619 (93.6)
β_4	1.979 (89.0)	2.156 (94.0)	2.195 (94.4)	2.406 (92.2)	2.573 (92.0)	2.620 (92.6)
			NL1 $\alpha^{(b)}$			
c_0	4.488 (88.0)	5.328 (92.2)	5.420 (92.2)	4.923 (87.4)	6.133 (90.8)	6.244 (91.4)
β_1	1.777 (89.6)	2.195 (93.4)	2.235 (93.6)	2.042 (83.6)	2.648 (87.4)	2.697 (88.0)
β_2	1.778 (87.0)	2.197 (92.2)	2.237 (92.4)	2.039 (84.6)	2.642 (90.4)	2.691 (91.0)
β_3	1.779 (86.2)	2.199 (91.6)	2.240 (92.0)	2.036 (80.2)	2.644 (86.6)	2.694 (87.4)
β_4	1.777 (85.2)	2.197 (90.6)	2.238 (91.0)	2.040 (82.2)	2.646 (88.6)	2.696 (89.2)
			NL2 $\alpha^{(a)}$			
c_0	4.878 (93.2)	5.231 (94.0)	5.321 (94.0)	5.625 (93.4)	5.969 (93.8)	6.076 (94.0)
β_1	1.974 (92.6)	2.149 (95.4)	2.189 (95.6)	2.399 (92.0)	2.565 (95.0)	2.613 (95.4)
β_2	1.973 (89.6)	2.150 (94.6)	2.188 (95.0)	2.394 (91.0)	2.562 (94.0)	2.609 (94.0)
β_3	1.975 (90.8)	2.155 (93.6)	2.195 (94.0)	2.399 (90.4)	2.570 (92.8)	2.617 (93.4)
β_4	1.977 (88.8)	2.155 (94.0)	2.195 (94.4)	2.403 (92.2)	2.571 (91.8)	2.618 (92.8)
			NL2 $\alpha^{(b)}$			
c_0	4.487 (88.0)	5.328 (92.2)	5.421 (92.2)	4.919 (87.4)	6.134 (90.8)	6.245 (91.6)
β_1	1.776 (89.6)	2.195 (93.4)	2.236 (93.6)	2.040 (83.6)	2.648 (87.4)	2.697 (88.0)
β_2	1.778 (87.2)	2.198 (92.2)	2.238 (92.8)	2.038 (84.4)	2.642 (90.4)	2.692 (91.0)
β_3	1.779 (86.2)	2.200 (91.6)	2.240 (92.2)	2.034 (80.2)	2.645 (86.8)	2.694 (87.4)
β_4	1.776 (85.4)	2.198 (90.6)	2.238 (91.0)	2.038 (82.0)	2.646 (88.6)	2.696 (89.2)

TABLE 5 The average of RE and PCD for $p = 10$

Method	$\alpha = \alpha^{(a)}$				$\alpha = \alpha^{(b)}$			
	Quadratic $\mu_0^Q(\mathbf{X})$		Cubic $\mu_0^C(\mathbf{X})$		Quadratic $\mu_0^Q(\mathbf{X})$		Cubic $\mu_0^C(\mathbf{X})$	
Lin	RE	PCD	RE	PCD	RE	PCD	RE	PCD
SUP	1	96.20%	1	92.69%	1	96.85%	1	94.28%
NP	0.19	92.64%	0.38	87.75%	0.19	92.99%	0.38	88.45%
SP	0.99	96.16%	1.00	92.64%	1.06	96.94%	1.10	94.55%
SP.CV	0.98	96.14%	0.98	92.64%	0.94	96.81%	0.92	94.18%
KRLS	1.01	96.20%	1.07	92.79%	1.02	96.80%	1.06	94.30%
KRLS.CV	0.94	96.10%	0.95	92.54%	0.94	96.71%	0.95	94.14%
NL1	RE	PCD	RE	PCD	RE	PCD	RE	PCD
SUP	1	96.60%	1	95.60%	1	96.84%	1	96.24%
NP	0.20	92.51%	0.24	91.24%	0.19	91.87%	0.21	89.81%
SP	3.05	97.97%	1.76	96.40%	3.12	98.46%	2.10	97.63%
SP.CV	3.24	97.99%	1.89	96.57%	2.94	98.23%	1.62	97.27%
KRLS	0.91	97.28%	0.98	96.06%	0.86	97.56%	0.94	96.62%
KRLS.CV	1.54	97.08%	1.18	95.76%	1.52	97.34%	1.16	96.40%
NL2	RE	PCD	RE	PCD	RE	PCD	RE	PCD
SUP	1	96.67%	1	95.69%	1	96.90%	1	96.32%
NP	0.20	92.27%	0.24	91.36%	0.18	91.99%	0.20	90.01%
SP	3.07	98.02%	1.76	96.47%	3.13	98.50%	2.11	97.69%
SP.CV	3.24	98.04%	1.89	96.64%	2.94	98.28%	1.63	97.33%
KRLS	0.89	97.33%	0.97	96.14%	0.85	97.61%	0.93	96.69%
KRLS.CV	1.53	97.13%	1.18	95.84%	1.51	97.38%	1.15	96.47%

TABLE 6 The average of Bias and SD for $p = 10$

Method	$\alpha = \alpha^{(a)}$				$\alpha = \alpha^{(b)}$			
	Quadratic $\mu_0^Q(\mathbf{X})$		Cubic $\mu_0^C(\mathbf{X})$		Quadratic $\mu_0^Q(\mathbf{X})$		Cubic $\mu_0^C(\mathbf{X})$	
Lin	Bias	SD	Bias	SD	Bias	SD	Bias	SD
SUP	0.191	5.583	0.174	11.01	0.191	5.583	0.174	11.01
NP	0.046	7.316	0.420	11.85	-10.94	7.304	-12.45	11.89
SP	0.192	5.604	0.162	11.03	-0.077	5.472	-0.228	10.53
SP.CV	0.195	5.647	0.171	11.10	0.257	5.773	0.473	11.46
KRLS	0.229	5.528	0.192	10.67	-0.309	5.523	-0.424	10.71
KRLS.CV	0.249	5.748	0.197	11.30	0.289	5.752	0.254	11.29
NL1	Bias	SD	Bias	SD	Bias	SD	Bias	SD
SUP	-0.099	14.35	-0.321	17.40	-0.890	14.22	-1.103	16.91
NP	-0.274	16.72	-0.394	18.57	-34.01	17.40	-35.14	21.52
SP	0.143	7.021	-0.162	12.29	-4.106	6.615	-4.178	10.73
SP.CV	0.082	8.010	-0.219	12.65	-0.281	8.339	-0.185	13.35
KRLS	0.019	9.451	-0.160	13.26	-10.84	9.393	-10.36	12.90
KRLS.CV	0.118	11.58	-0.131	16.07	-0.382	11.59	-0.612	15.82
NL2	Bias	SD	Bias	SD	Bias	SD	Bias	SD
SUP	-0.098	14.35	-0.311	17.40	-0.890	14.22	-1.103	16.91
NP	-0.304	16.82	-0.437	18.80	-34.55	17.55	-35.67	21.60
SP	0.144	7.006	-0.159	12.27	-4.092	6.609	-4.170	10.73
SP.CV	0.083	8.005	-0.217	12.64	-0.286	8.333	-0.174	13.34
KRLS	0.019	9.477	-0.162	13.29	-10.95	9.417	-10.45	12.92
KRLS.CV	0.119	11.61	-0.130	16.10	-0.382	11.61	-0.612	15.84

TABLE 7 The average of SE and CP of proposed estimators with $p = 10$

SE(CP %)	Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$		
	SP	SP.CV	SP.DCV	SP	SP.CV	SP.DCV
Lin $\alpha^{(a)}$	5.300 (93.5)	5.539 (94.6)	5.728 (95.2)	10.48 (93.8)	10.90 (94.6)	11.28 (95.6)
Lin $\alpha^{(b)}$	4.931 (91.8)	5.525 (93.5)	5.711 (94.1)	9.297 (89.5)	10.69 (92.2)	11.06 (93.0)
NL1 $\alpha^{(a)}$	5.590 (81.1)	7.214 (92.5)	7.473 (93.5)	10.75 (89.1)	12.10 (94.3)	12.53 (95.1)
NL1 $\alpha^{(b)}$	4.827 (73.2)	7.415 (90.4)	7.682 (91.4)	8.494 (81.6)	11.98 (90.4)	12.41 (91.2)
NL2 $\alpha^{(a)}$	5.579 (81.1)	7.204 (92.5)	7.462 (93.5)	10.74 (89.1)	12.09 (94.4)	12.52 (95.1)
NL2 $\alpha^{(b)}$	4.820 (73.2)	7.406 (90.4)	7.671 (91.4)	8.490 (81.6)	11.98 (90.4)	12.40 (91.2)

TABLE 8 Testing of the MCAR assumption

Predictors	Labeled Data			Unlabeled Data			P-value of Test
	Median	Mean	SD	Median	Mean	SD	
age	65	64.84	15.64	68	67.10	16.34	0.1229
weight	75.85	80.22	21.77	78.4	82.80	26.15	0.0921
BUN	45.5	59.01	42.56	39	50.47	36.62	0.0646
creatinine	1.9	2.743	2.451	1.7	2.497	2.229	0.2489
WBC	20.35	23.12	12.13	18.9	21.46	13.06	0.1522
HR	92.66	92.52	14.33	89.87	90.29	14.73	0.1220

TABLE 9 Parameter estimation

Predictors	Methods					
	SUP	NP	SP	SP.CV	KRLS	KRLS.CV
intercept	-4.502	-4.502	-4.495	-4.616	-4.614	-4.718
age	0.093	0.093	0.094	0.090	0.082	0.100
weight	-2.520	-2.520	-2.521	-2.518	-2.542	-2.571
BUN	-1.302	-1.302	-1.301	-1.173	-1.318	-1.205
creatinine	1.721	1.721	1.721	1.608	1.735	1.656
WBC	-0.216	-0.216	-0.214	-0.172	-0.225	-0.212
HR	-1.758	-1.758	-1.758	-1.824	-1.762	-1.830

TABLE 10 SE estimation

Predictors	Methods			
	SUP	SP	SP.CV	SP.DCV
intercept	7.559	7.292	7.332	7.498
age	2.647	2.482	2.496	2.569
weight	2.809	2.642	2.665	2.764
BUN	4.455	4.071	4.110	4.288
creatinine	4.988	4.509	4.549	4.746
WBC	2.851	2.525	2.541	2.625
HR	2.737	2.553	2.566	2.623

TABLE 11 Treatment recommendation

Treatment	Methods					
	SUP	NP	SP	SP.CV	KRLS	KRLS.CV
A=-1: IV Fluid Resuscitation	5113	5113	5112	5132	5150	5142
A=1: Vasopressors	2694	2694	2695	2675	2657	2665