Adaptive Semi-Supervised Inference for Optimal Treatment Decisions with Electronic Medical Record Data

By Kevin Gunn, Wenbin Lu, Rui Song
Department of Statistics, North Carolina State University
email: kpgunn@ncsu.edu

SUMMARY. A treatment decision is a rule that assigns a treatment to a patient based on their observed clinical information. The treatment decision that yields the greatest overall expected clinical benefit to the entire patient population is called the optimal treatment decision. We consider estimation of the optimal treatment decision within the restricted class of linear decision rules under semi-supervised settings, where the data consists of a set of 'labeled' patients and a much larger set of 'unlabeled' patients. This paper proposes an imputation-based semi-supervised method, SSQ-SNP, utilizing 'unlabeled' individuals into the linear decision rule to offer a more efficient estimator. An inference procedure and results for asymptotic normality and consistency are provided. We present simulation studies to assess its performance relative to a fully supervised method and the effects of misspecification for the proposed linear decision rule. Afterward, application to an EMR study on the treatment of hypotensive episodes during an ICU stay is discussed.

1 Introduction

Precision medicine, which is focused on creating treatment decisions for a patient based on his/her clinical information, has earned considerable interest. A treatment decision is a rule which, given a patient's observed clinical information, assigns a treatment to the patient. The objective is to choose the treatment, from the set of all possible treatments, that maximizes the patient's expected outcome. This creates a treatment tailored to each patient in the population of interest. The treatment decision that yields the greatest overall expected clinical benefit to the entire patient population is called the optimal treatment decision (OTD).

A recent area of interest in optimal treatment decision estimation has been the utilization of Electronic Medical Record (EMR) data. These data allow researchers to explore optimal treatment decisions in specific clinical scenarios not feasible in a randomized clinical trial (RCT) such as patients suffering from sepsis (Raghu et al. 2017). EMR data can also provide guidance where there is little previous research conducted, including second line treatment choices for type 2 diabetes episodes (Wang et al. 2016). The recent shift in hospitals and other healthcare organizations to store clinical information in EMRs provide an immense amount of detailed information on the clinical process of patients and their health status during interactions with their healthcare system. With databases such as MIMIC-III, an openly available critical care dataset (Johnson et al. 2016), these data are more readily accessible. MIMIC-III encompasses medical record chart data recorded by clinicians, laboratory measurements, progress notes, imaging results, and billing information. The wealth of detailed clinical data provides the opportunity to enhance clinical decisions for accurate personalized medicine.

There is a great deal of work on statistical techniques to estimate the OTD from RCT data or observational study data under fully supervised settings, where a single decision or a series of sequential decisions may be of interest (Murphy 2003, Moodie et al. 2007, Robins 2004, Zhao et al. 2012, Zhang et al. 2012). The purpose of this paper is to address the estimation of the OTD under an assumed linear *working* model, where the majority of individuals are missing response or treatment information. In other words, a semi-supervised learning (SSL) problem. In SSL, the knowledge gained through $\mathbb{P}_{\mathbf{X}}$ (distribution of $\mathbf{X}$) from the 'unlabeled' individuals is incorporated to improve inference on $\mathbb{P}_{\mathbf{Y}|\mathbf{X}}$ (Chapelle et al. 2006). In the OTD setting, we extend this to include treatment information, $\mathbf{A}$, into the inference procedure by trying to estimate $\mathbb{P}_{\mathbf{Y}|\mathbf{X},\mathbf{A}}$ using $\mathbb{P}_{\mathbf{A}|\mathbf{X}}$ and $\mathbb{P}_{\mathbf{X}}$. The proposed semi-supervised (SS) method is a three-step semi-parametric estimator to find the target parameter, $\boldsymbol{\beta}$, from the assumed linear *working* model. We combine SSL with Q-learning (Schulte et al. 2014, Nahum-Shani et al. 2012) to estimate the OTD at one decision point. Q-learning posits a model, denoted as the Q-function, for the outcome of interest given the subject's information for each available treatment. The OTD finds the treatment that maximizes the Q-function. The proposed method is an imputation based semi-nonparametric kernel regression Q-function (SSQ-SNP) that is flexible to the underlying distribution, $\mathbb{P}_{\mathbf{Y}|\mathbf{A},\mathbf{X}}$. This is necessary when the assumed *working* model is misspecified. The 'unlabeled' subjects have their outcome imputed for each treatment, before a linear regression of the contrast between imputed outcomes between treatments against the covariates to obtain the OTD. SSQ-SNP's incorporation of all available data into the estimation procedure for the OTD intends to reduce the bias and the standard error for the parameters of interest.

Our work also discusses a fully supervised linear estimator, titled the transformed response ordinary least squares (TR-OLS), to estimate the OTD. TR-OLS combines inverse propensity score weighted estimators with a linear model to estimate $\boldsymbol{\beta}$. The focus with both approaches is on providing linear decision rules in low-dimensional settings due to their simple interpretations. Our method, SSQ-SNP, is intended to be adaptive to the true underlying distribution making it robust to model misspecification, and the TR-OLS estimator measures the benefit of including the covariate information from the 'unlabeled' subjects into the estimation procedure.

This article is organized as follows; Section 2 discusses the assumptions and framework for optimal treatment decisions. The estimation procedure and asymptotic theory results are presented in Section 3 for TR-OLS and Section 4 for SSQ-SNP, respectively. Section 5 provides simulation studies to compare the empirical performances of both methods. Section 6 demonstrates an application to a study of patients in the ICU undergoing a hypotensive episode. Afterward, a discussion about future work within the area of optimal treatment decision estimation with EMR data is given.

2 Framework and Assumptions

Treatments received by patients will be denoted $A \in \mathcal{A}$, where $\mathcal{A}$ is the set of possible treatments. Let $\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^p$ be a vector of subject characteristics ascertained prior to treatment. We will further assume $\mathcal{X}$ is a compact set, and $Var(\mathbf{X})$ is positive definite. Let $Y \in \mathbb{R}$ denote the observed response variable of interest and consider a larger value of Y as the better outcome. Suppose we have a fully 'observed' subject group, $\mathbf{O} = [\mathbf{O}_i = \{Y_i, A_i, \mathbf{X}_i\} : i = 1, ...n$ are i.i.d.], and an 'unobserved' subject group where the treatment and/or response are not available, $\mathbf{U} = [\mathbf{X}_j : j = n+1, ...N$ are i.i.d.], where $N >> n$. Assume $\mathcal{O}$ has finite 2^{nd} moments and all observations, $\mathcal{O} \cup \mathcal{L}$, are drawn from the same underlying distribution, $\mathbb{P} = \left(\mathbb{P}_{\mathbf{Y}|\mathbf{A},\mathbf{X}}, \mathbb{P}_{\mathbf{A}|\mathbf{X}}, \mathbb{P}_{\mathbf{X}}\right)$, i.e. the response and treatment information is missing completely at random (MCAR). The mathematical formulation of the optimal treatment decision will be defined to clarify the estimation procedure.

Consider a disease with two treatment options, each patient is assigned one treatment and the patient's response to that treatment is observed. Assume the patient's response to the given treatment is constructed as, $Y_i = A_i Y^*(1) + (1 - A_i) Y^*(0)$ (Rubin 1978). Furthermore, it is assumed that there are *no unmeasured confounders*, i.e., $(Y^*(1), Y^*(0)) \perp A \mid \mathbf{X}$ (Rosenbaum & Rubin 1983). The last assumption made is the *positivity* assumption, which states, $0 < \pi(\mathbf{X}) < 1$ where $\pi(\mathbf{X})$

denotes the propensity score, i.e. $\pi(\mathbf{X}) = P(A = 1 \mid \mathbf{X})$.

Formally, a treatment decision is a rule that maps the covariates onto the treatment space, $\delta : \mathcal{X} \to \mathcal{A}$. A decision rule is implemented as a treatment, $Y^*(\delta) = \delta(\mathbf{X})Y^*(1) + (1 - \delta(\mathbf{X}))Y^*(0)$. The optimal treatment decision is denoted as δ^{opt}, and is defined as $\delta^{opt} = \arg\max_{\delta \in \Delta} E\left[Y^*(\delta)\right]$, where Δ is the class of all treatment decisions. Under this framework, $E\left[Y^*(\delta)\right] = E_{\mathbf{X}}\left\{\delta(\mathbf{X})E\left[Y \mid \mathbf{X}, A = 1\right] + (1 - \delta(\mathbf{X}))E\left[Y \mid \mathbf{X}, A = 0\right]\right\}$. Hence, the OTD becomes, $\delta^{opt} = I\left(E\left[Y \mid \mathbf{X}, A = 1\right] - E\left[Y \mid \mathbf{X}, A = 0\right] \geq 0\right)$.

Suppose the following model for the rest of the article, $E\left[Y \mid \mathbf{X}, A\right] = \mu(\mathbf{X}) + AC(\mathbf{X})$. The function, $\mu(\mathbf{X})$, represents the baseline effects of $\mathbf{X}$ on Y, and the contrast function, $C(\mathbf{X}) = E\left[Y \mid \mathbf{X}, A = 1\right] - E\left[Y \mid \mathbf{X}, A = 0\right]$, is the expected difference between treatments on the individual given his/her covariates. We further consider a linear contrast function as our assumed *working* model, i.e. $E\left[Y \mid \mathbf{X}, A\right] = \mu(\mathbf{X}) + A(\boldsymbol{\beta}'\widetilde{\mathbf{X}})$, where $\widetilde{\mathbf{X}} = (1, \mathbf{X}')'$. The OTD, δ^{opt}, arising from a posited linear model with parameter, $\boldsymbol{\beta}$, will belong to the class of estimators, $\Delta_{\boldsymbol{\beta}}$, and be denoted as $\delta^{opt}_{\boldsymbol{\beta}}$. Our estimator for the OTD within $\Delta_{\boldsymbol{\beta}}$ can be deduced to the function, $\widehat{\delta}^{opt}_{\boldsymbol{\beta}}(\mathbf{X}) = I(\widehat{\boldsymbol{\beta}}'\widetilde{\mathbf{X}} > 0)$, where $\widehat{\boldsymbol{\beta}}$ is the solution to the normal equations. This class of estimators is chosen to produce decision rules that are easily interpretable due to their simplicity.

3 Transformed Response Ordinary Least Squares

The OTD is a function of $C(\mathbf{X})$, and as a result it is not necessary to estimate the baseline effect, $\mu(\mathbf{X})$. To reduce the possibility of model misspecification, we estimate $\delta^{opt}_{\boldsymbol{\beta}}$ without fully specifying $E\left[Y \mid \mathbf{X}, A = a\right]$ for $a \in \{0, 1\}$. The transformed response ordinary least squares (TR-OLS) method regresses $C(\mathbf{X})$ directly onto the covariates to provide a fully supervised linear decision rule for the OTD.

TR-OLS uses an inverse propensity weighted (IPW) estimator of $C(\mathbf{X})$ as responses in a linear model. An unbiased IPW estimator for $C(\mathbf{X})$ is, $\widetilde{Y} = \frac{Y(A-\pi(\mathbf{X}))}{\pi(\mathbf{X})(1-\pi(\mathbf{X}))}$. Therefore, the transformed response model is, $\widetilde{Y} = C\left(\mathbf{X}\right) + \epsilon$, whereby under the assumed *working* model $C\left(\mathbf{X}\right)$ is equivalent to $\boldsymbol{\beta}'\widetilde{\mathbf{X}}$ and $E\left[\epsilon \mid \mathbf{X}\right] = 0$. The least squares is minimized to obtain $\widehat{\boldsymbol{\beta}}_{TR}$, the regression parameters for the observed cases,

$$\widehat{\boldsymbol{\beta}}_{TR} = \arg\min_{\beta} \sum_{i=1}^{n} \left[\widetilde{Y}_i - \boldsymbol{\beta}'\widetilde{\mathbf{X}}_i\right]^2 . \tag{3.1}$$

Under the assumptions given in section 2 and as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{TR} - \boldsymbol{\beta}\right) = n^{-1/2} \sum_{i=1}^{n} \Psi_{TR}\left(\mathbf{O}_i\right) + o_p(1) \xrightarrow{d} N_{p+1}\left(0, V_{\boldsymbol{\beta}_{TR}}\right), \tag{3.2}$$

where $\Psi_{TR}\left(\mathbf{O}\right) = \Lambda^{-1}\widetilde{\mathbf{X}}\left(\widetilde{Y} - \widetilde{\boldsymbol{\beta}}'_{TR}\widetilde{\mathbf{X}}\right)$ and $\Lambda = E[\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}']$. The asymptotic variance, $V_{\boldsymbol{\beta}_{TR}}$, is equivalent to $E\left[\Psi_{TR}\left(\mathbf{O}\right)\Psi'_{TR}\left(\mathbf{O}\right)\right]$. Consistent estimators for $V_{\boldsymbol{\beta}_{TR}}$ and Λ are given by $\widehat{V}_{\boldsymbol{\beta}_{TR}} = \frac{1}{n}\sum_{i=1}^{n} \Psi_{TR}\left(\mathbf{O}_i\right)\Psi'_{TR}\left(\mathbf{O}_i\right)$ and $\Lambda_n = n^{-1}\sum_{i=1}^{n} \widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}'_i$, respectively.

The supervised estimator, $\widehat{\boldsymbol{\beta}}_{TR}$, uses data only from $\mathcal{O}$. By including the data from $\mathcal{U}$, the SS estimator intends to create a more efficient estimator for $\boldsymbol{\beta}$ relative to $\widehat{\boldsymbol{\beta}}_{TR}$, especially under model misspecification. The parameter, $\widehat{\boldsymbol{\beta}}_{TR}$, gauges the improvement offered by incorporating $\mathcal{U}$ into the estimation procedure discussed in Sections 4.1 and 4.2. A discussion of Q-learning for a single decision point is needed to establish the difference in the SS approach.

4 Semi-Supervised Q-Learning for a Single Decision Point

Q-learning at a single decision point aims to estimate the expected outcome of interest conditioned on treatment and covariates, $Q(\mathbf{x}, a) = E\left[Y \mid \mathbf{X} = \mathbf{x}, A = a\right]$. Q-functions typically model $Q(\mathbf{x}, a)$ using all the observed data, $\mathcal{O}$, with a parametric or semi-parametric estimator, $Q(\mathbf{x}, a, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a finite-dimensional parameter (Schulte et al. 2014). Using the estimator, $Q(\mathbf{x}, a, \boldsymbol{\theta})$, the OTD is defined as $I(Q(\mathbf{x}, 1, \boldsymbol{\theta}) - Q(\mathbf{x}, 0, \boldsymbol{\theta}) \geq 0)$. Given the assumed *working* model, the Q-functions are expressed as, $Q(\mathbf{x}, a) = \mu(\mathbf{x}) + a(\boldsymbol{\beta}'\widetilde{\mathbf{x}})$, and the OTD is equivalent to $\delta^{opt}_{\boldsymbol{\beta}}(\mathbf{x}) = I(\boldsymbol{\beta}'\widetilde{\mathbf{x}} \geq 0)$. If the data available consists of sets $\mathcal{O}$ and $\mathcal{U}$, then Q-functions should use all available data to make treatment decisions. We propose a three-step SS imputation procedure as follows: (i) impute $Q\left(\mathbf{X}_j, a\right)$ for $a \in \{0, 1\}$ and $\forall j = n+1, ..., N$, (ii) compute the imputed contrast functions, $C(\mathbf{X}_j) = Q\left(\mathbf{X}_j, 1\right) - Q\left(\mathbf{X}_j, 0\right)$ for $\forall j = n+1, ..., N$, (iii) regress the imputed contrast functions on $\mathbf{X}$ to obtain $\widehat{\boldsymbol{\beta}}$ for $\widehat{\delta}^{opt}_{\boldsymbol{\beta}}$.

4.1 Fully Nonparametric Imputation of the Q-Function

If all the Y and A in $\mathcal{U}$ were actually observed, we could obtain an estimate for $\boldsymbol{\beta}$ utilizing the entire data set with the TR-OLS approach discussed in Section 3. Instead, we must use a different approach. We present a fully non-parametric estimator based on kernel regression (KR). KR is a locally weighted average function with no parametric assumptions. The kernel function is denoted as $W(\cdot)$ with $W : \mathbb{R}^p \to \mathbb{R}$. The bandwidth, h, is a function of n and greater than zero. Q-functions as KR estimators are defined as,

$$Q^{(np)}(\mathbf{x},1) = \frac{\sum_{i=1}^{n} W\left(\frac{\mathbf{x}-\mathbf{X_i}}{h}\right) A_i Y_i}{\sum_{i=1}^{n} W\left(\frac{\mathbf{x}-\mathbf{X_i}}{h}\right) A_i} \qquad \text{and} \qquad Q^{(np)}(\mathbf{x},0) = \frac{\sum_{i=1}^{n} W\left(\frac{\mathbf{x}-\mathbf{X_i}}{h}\right)(1-A_i)Y_i}{\sum_{i=1}^{n} W\left(\frac{\mathbf{x}-\mathbf{X_i}}{h}\right)(1-A_i)}. \tag{4.1}$$

The first step is to impute both $Q^{(np)}(\mathbf{x},0)$ and $Q^{(np)}(\mathbf{x},1)$, followed by the computation of $\widehat{C}^{(np)}(\mathbf{x})$. The set $\mathcal{U}$ with imputed data is $\left[\left\{\widehat{C}^{(np)}(\mathbf{X}_j), \mathbf{X}_j\right\} : j = n+1, ..., N\right]$, and the solution the normal equations is obtained by minimizing the argument, $\widehat{\boldsymbol{\beta}}_{np} = \arg\min_{\beta} \sum_{j=n+1}^{N} \left[\widehat{C}^{(np)}(\mathbf{X}_j) - \boldsymbol{\beta}' \widetilde{\mathbf{X}}_j\right]^2$.The new nonparametric SS linear decision rule is $\widehat{\delta}^{opt}_{\boldsymbol{\beta}_{np}} = I(\boldsymbol{\beta}'_{np}\widetilde{\mathbf{X}} > 0)$, and utilizes $\mathbb{P}_{\mathbf{X}}$ from $\mathcal{U}$.

The subsequent assumptions are needed for Theorem 1 given below. Most of these assumptions are fairly standard (Fan 1992, Hansen 2008, Newey 1994), but are adapted slightly to the framework needed for OTDs. (1) $h \equiv h(n) = o(1)$, (2) $Q^{(np)}(\mathbf{x},a)$, $\pi(\mathbf{x})$, and $f(\cdot)$ are r times continuously differentiable with bounded r^{th} derivatives on some open set within $\mathcal{X}$. (3) $W(\cdot)$ is a symmetric r^{th} order kernel for some integer $r \geq 2$. $W(\cdot)$ is Lipschitz continuous and has bounded support that it shares with $\pi(\cdot)$, $\mathcal{W} \subseteq \mathbb{R}^p$. (4) $E(|AY|^s) < \infty$ and $E(|(1-A)Y|^s) < \infty$ for some $s > 2$. (5) $E(|AY|^s \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x}) = E(|Y|^s \mid \mathbf{X} = \mathbf{x})\pi(\mathbf{x}) f(\mathbf{x})$ and $f(\mathbf{x})$ are bounded on $\mathcal{X}$ and $\inf_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) > 0$.

Theorem 1. *Suppose* $\mathrm{n}^{1/2}\mathrm{h}^{\mathrm{r}} \to 0$ *and* $\sqrt{\frac{\ln n}{nh^p}} \to 0$ *as* $n \to \infty$ *and* $N >> n$ *such that* $n/N \to 0$. *Then, under the required assumptions (1)-(5),*

$$n^{1/2}\left(\widehat{\boldsymbol{\beta}}_{np} - \boldsymbol{\beta}\right) = n^{-1/2}\sum_{i=1}^{n} \Psi_{np}(\mathbf{O}_i) + o_p(1) \xrightarrow{d} N_{p+1}\left(0, V_{\boldsymbol{\beta}_{np}}\right) \tag{4.2}$$

where $\Psi_{np}(\mathbf{O}) = \left\{\frac{A}{\pi(\mathbf{X})} - \frac{1-A}{1-\pi(\mathbf{X})}\right\}\Lambda^{-1}\mathbf{X}(Y - Q^{(np)}(\mathbf{X},A))$ *and asymptotic variance,* $V_{\boldsymbol{\beta}_{np}} = E\left[\Psi_{np}(\mathbf{O})\Psi'_{np}(\mathbf{O})\right]$.

There are drawbacks to using KR owing to the *curse of dimensionality* and inherent *finite sample bias.* Fan (1992) discusses *finite sample bias* for KR has order of h^2 for 2^{nd} order kernels, and through Taylor expansion it is clear the bias has order of h^r for r^{th} order kernels. For this article, the goal is developing an efficient low-dimensional linear decision rule, so the *curse of dimensionality* is not of concern. However, if the sample size within $\mathcal{O}$ is not large enough, $\boldsymbol{\beta}_{np}$ is a biased estimator. In Section 4.2, we discuss a cross-validation technique to correct for bias with a linear term by regressing the residuals, $Y - Q^{(np)}(\mathbf{x}, a)$, against the covariates.

4.2 Semi-Nonparametric Imputation of the Q-Function

We overcome the issue of *finite sample bias* in nonparametric estimation by adapting the work of Chakrabortty & Cai (2018) on semi-nonparametric imputation based estimators into the SS OTD framework. The new semi-nonparametric (SSQ-SNP) estimator, $Q^{(SS)}(\mathbf{x}, a, \boldsymbol{\theta}_a)$, builds upon the estimator in Section 4.1, but produces a more efficient estimator for $\boldsymbol{\beta}$ by reducing bias. SSQ-SNP achieves this through a *linear refitting* step.

SSQ-SNP is composed of two steps, where the first step is to fit the KR estimator, $\widehat{Q}^{(np)}(\mathbf{x}, a)$, on the data in $\mathcal{O}$, and the second step (*linear refitting* step) corrects for the bias of $\widehat{Q}^{(np)}(\mathbf{x}, a)$ with a linear model. The *linear refitting* step applies an inverse propensity score weighted least squares regression of the residuals, $Y - Q^{(np)}(\mathbf{X}, a)$, on $\mathbf{X}$ to obtain estimates for the parameters of interest, $\boldsymbol{\theta}_a$ for $a \in \{0, 1\}$. Namely, $\boldsymbol{\theta}_a$ for $a \in \{0, 1\}$, are the solutions to minimizing the weighted least squares equations, $\widehat{\boldsymbol{\theta}}_1 = \arg\min_{\boldsymbol{\theta}_1} \sum_{k=1}^{\mathcal{K}} \sum_{i \in \mathcal{O}_k} \frac{A_i}{\pi(\mathbf{X}_i)} \left(Y_i - \widehat{Q}_k^{(np)}(\mathbf{X}_i, A_i) - \boldsymbol{\theta}_1' \widetilde{\mathbf{X}}_i\right)^2$ and, $\widehat{\boldsymbol{\theta}}_0 = \arg\min_{\boldsymbol{\theta}_0} \sum_{k=1}^{\mathcal{K}} \sum_{i \in \mathcal{O}_k} \frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} \left(Y_i - \widehat{Q}_k^{(np)}(\mathbf{X}_i, A_i) - \boldsymbol{\theta}_0' \widetilde{\mathbf{X}}_i\right)^2$. The *linear refitting* step utilizes $\mathcal{K}$-fold cross-validation to obtain an unbiased estimator of $\boldsymbol{\theta}_a$ for $a \in \{0, 1\}$. Cross-validation separates $\mathcal{O}$ into $\mathcal{K}$ partitions of equal size, where both treatment cohorts have their populations split equally among the folds. The partitions are denoted as $\mathcal{O}_k$ for $k \in \{1, 2, .., \mathcal{K}\}$. Cross-validation overcomes underestimation of the true residuals caused by over-fitting during the kernel regression step. The final imputation function is $Q^{(SS)}(\mathbf{x}, a, \boldsymbol{\theta}_a) = Q^{(np)}(\mathbf{x}, a) + \boldsymbol{\theta}_a' \mathbf{x}$, which is estimated as: $\widehat{Q}^{(SS)}\left(\mathbf{x}, a, \widehat{\boldsymbol{\theta}}_a\right) = \mathcal{K}^{-1} \sum_{k=1}^{\mathcal{K}} \widehat{Q}_k^{(np)}(\mathbf{x}, a) + \widehat{\boldsymbol{\theta}}_a' \widetilde{\mathbf{x}}$, for $a \in \{0, 1\}$. We compute $\widehat{C}^{(SS)}(\mathbf{x}) = \widehat{Q}^{(SS)}\left(\mathbf{x}, 1, \widehat{\boldsymbol{\theta}}_1\right) - \widehat{Q}^{(SS)}\left(\mathbf{x}, 0, \widehat{\boldsymbol{\theta}}_0\right)$, and obtain $\widehat{\boldsymbol{\beta}}_{SS}$ similarly to Section 4.1. The OTD based on SSQ-SNP is $\widehat{\delta}_{\boldsymbol{\beta}_{SS}}^{opt} = I\left(\widehat{\boldsymbol{\beta}}_{SS}' \widetilde{\mathbf{x}} > 0\right)$.

We need an additional condition for the influence function given in Theorem 2 to be unbiased and consistent: *condition* (1) let $\widehat{\mathbf{R}}_{k,a}(\mathbf{X}) = \mathbf{X}\left[\widehat{Q}_k^{(np)}(\mathbf{X}, a) - Q^{(np)}(\mathbf{X}, a)\right]$, $\bar{\mathbf{R}}_{k,a}(\mathbf{X}) = E[\widehat{\mathbf{R}}_{k,a}] - \widehat{\mathbf{R}}_{k,a}$, and $\mathcal{R}_{n,\mathcal{K},a} = n^{-1/2} \sum_{k=1}^{\mathcal{K}} \sum_{i \in \mathcal{O}_k} \bar{\mathbf{R}}_{k,a}(\mathbf{X}_i)$. Then for $\mathcal{K} \geq 2$, $\mathcal{R}_{n,\mathcal{K},a} = o_p(1)$, $\forall a \in \mathcal{A}$ by *Lemma A.1* in Chakrabortty & Cai (2018).

Theorem 2. *Suppose the KR estimator, $Q^{(np)}(\mathbf{X}, a)$, satisfies assumptions (1) through (5) and* *condition (1) holds,*

$$n^{1/2}\left(\widehat{\boldsymbol{\beta}}_{SS}-\boldsymbol{\beta}\right)=n^{-1/2}\sum_{i=1}^{n}\Psi_{SS}(\mathbf{O}_i)+o_p(1)\xrightarrow{d}N_{p+1}\left(0,V_{\boldsymbol{\beta}_{SS}}\right) \qquad (4.3)$$

where $\Psi_{SS}(\mathbf{O})=\left\{\frac{A}{\pi(\mathbf{X})}-\frac{1-A}{1-\pi(\mathbf{X})}\right\}\Lambda^{-1}\mathbf{X}(Y-Q^{(SS)}(\mathbf{X},A,\boldsymbol{\theta}_A))$ *and asymptotic variance,* $V_{\boldsymbol{\beta}_{SS}}=E\left[\Psi_{SS}\left(\mathbf{O}\right)\Psi'_{SS}\left(\mathbf{O}\right)\right]$.

Both SSQ-SNP and TR-OLS incorporate propensity scores into their estimation procedure, but SSQ-SNP is much less reliant on correct specification of the propensity score in situations where it needs to be estimated from the data. SSQ-SNP is an adaptive imputation estimator since it does not assume a fully parametric model for $Q(\mathbf{x},a)$, which allows flexibility during the imputation procedure of $\mathcal{U}$.

4.3 Inference for Semi-Supervised Q-Functions

The SSQ-SNP Q-functions, $Q^{(SS)}\left(\mathbf{x},a,\widehat{\boldsymbol{\theta}}_a\right)$, over-fit Y in $\mathcal{O}$ due to the *linear refitting* step. Due to this issue, $\widehat{V}_{\boldsymbol{\beta}_{SS}}=\frac{1}{n}\sum_{i=1}^{n}\Psi_{SS}\left(\mathbf{O}_i\right)\Psi'_{SS}\left(\mathbf{O}_i\right)$ underestimates $V_{\boldsymbol{\beta}_{SS}}$ in practice. A 'double-CV' step is proposed to reduce bias during the variance estimation procedure. We create $\mathcal{K}$ folds, $\mathcal{O}_k$ for $k\in\{1,..,\mathcal{K}\}$, then construct $\mathcal{K}$ distinct estimates of $\boldsymbol{\theta}_a$, $\left\{\boldsymbol{\theta}_{a,(k)}:k=1,..,\mathcal{K}\right\}$. Let $\mathcal{O}_k^*$ represent the data not included in $\mathcal{O}_k$, where $\mathcal{O}_k^*\perp\mathcal{O}_k$ and $\mathcal{O}_k^*\bigcup\mathcal{O}_k=\mathcal{O}$. Fold $\mathcal{O}_k$ is not involved in estimating $\left\{\widehat{Q}^{(np)}\left(\mathbf{X}_i,A_i\right)\right\}_{i\in\mathcal{O}_k^*}$ as well as the parameter, $\boldsymbol{\theta}_{a,(k)}$, to mirror imputation of missing response information in the SS setting. The cross-validated estimate $\boldsymbol{\theta}_{a,(k)}$ is the solution to: $\widehat{\boldsymbol{\theta}}_{1,(k)}=\arg\min_{\boldsymbol{\theta}_1}\sum_{i\notin\mathcal{O}_k}\frac{A_i}{\pi(\mathbf{X}_i)}\left(Y_i-\widehat{Q}_k^{(np)}\left(\mathbf{X}_i,A_i\right)-\boldsymbol{\theta}_1'\mathbf{X}_i\right)^2\ ,\forall k\in\{1,...,\mathcal{K}\}$, where $\boldsymbol{\theta}_{0,(k)}$ is estimated analogously. The 'double-CV' imputation function becomes, $\widehat{Q}_k^{(SS)}\left(\mathbf{x},a,\boldsymbol{\theta}_{a,(k)}\right)=\widehat{Q}_k^{(np)}\left(\mathbf{x},a\right)+\widehat{\boldsymbol{\theta}}'_{a,(k)}\mathbf{x}$. We substitute $\widehat{Q}_k^{(SS)}\left(\mathbf{x},a,\boldsymbol{\theta}_{a,(k)}\right)$ for $\widehat{Q}^{(SS)}\left(\mathbf{x},a,\boldsymbol{\theta}_a\right)$ in the corresponding influence functions from Theorem 2. The estimators $\Lambda_N=N^{-1}\sum_{j=n+1}^{N}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i'$ and Λ_{n+N} are consistent estimators of Λ using the available data in $\mathcal{U}$. The asymptotic variance, $V_{\boldsymbol{\beta}_{SS}}$, has the consistent 'double-CV' estimator, $\widehat{V}_{\boldsymbol{\beta}_{SS,(\mathcal{K})}}=\frac{1}{n}\sum_{i=1}^{n}\Psi_{SS,(k)}\left(\mathbf{O}_i\right)\Psi'_{SS,(k)}\left(\mathbf{O}_i\right)$, which allows us to establish standard error and confidence intervals estimates for $\widehat{\boldsymbol{\beta}}_{SS}$.

5 Simulation Analysis

We present the percent of correct decisions (PCD) and value function to illustrate the benefits of semi-supervised prediction on assessing $\delta_{\boldsymbol{\beta}}^{opt}$. Furthermore, we study the bias, empirical standard error (ESE), asymptotic standard error (ASE), and the relative efficiency of $\widehat{\boldsymbol{\beta}}_{SS}$ with respect to $\widehat{\boldsymbol{\beta}}_{TR}$, and demonstrate the results.

In this simulation, n is the amount of subjects in $\mathcal{O}$ and N is the amount of subjects in $\mathcal{U}$. The N unobserved responses were simulated to be missing completely at random (MCAR), where the response and treatment are not observed. Let $\mathbf{X} \sim N\left(\mathbf{0}_p, I_p\right)$, where $\mathbf{X}$ is contained in $[-5,5]^p$ to ensure it is in a compact set, and $\mathcal{A} \in \{0,1\}$. The following models are tested with $p = 2$, $n = 500$ and $N = 5000$;

- *Model 1 (Linear)*: $\mathbf{Y} = \mu(\mathbf{X}) + \mathbf{A}\left(\boldsymbol{\eta}'\mathbf{X}\right) + \epsilon$
- *Model 2 (Cubic)*: $\mathbf{Y} = \mu(\mathbf{X}) + \mathbf{A}\left(\boldsymbol{\gamma}'\mathbf{X}\right)^{\mathbf{3}} + \epsilon$
- *Model 3 (Sine)*: $\mathbf{Y} = \mu(\mathbf{X}) + \mathbf{A}\left(\sin(\boldsymbol{\eta}'\mathbf{X})\right) + \epsilon$,

where $\boldsymbol{\eta} = \mathbf{1_2}$, $\boldsymbol{\gamma} = (0.3, 0.6)'$, and $\epsilon \sim \mathcal{N}(0,1)$. Two baseline functions were studied. The first one simulates a situation where the baseline effect for subjects is small and the second function proposes a larger baseline effect,

1. $\mu(\mathbf{X}) = \left(\boldsymbol{\alpha}'\mathbf{X}\right)\left(1 + \boldsymbol{\omega}'\mathbf{X}\right)$
2. $\mu(\mathbf{X}) = \left(\boldsymbol{\omega}'\mathbf{X}\right)^3$

The models were studied with $\boldsymbol{\omega} = (0.5, 0.5)'$ and $\boldsymbol{\alpha} = (0.75, 0.75)'$. We simulate a propensity score of $\pi(\mathbf{X}) = logit(0.5X_1 - 0.5X_2)$, and an allocation of treatment as $\mathbf{A} \sim bernoulli(p = \pi(\mathbf{X}))$. The true values, $\boldsymbol{\beta}_0$, are estimated by simulating a fully observed Monte Carlo dataset of size 500,000.

The results of interest are percent of correct decisions (PCD), the value function (V), and relative efficiency (RE) over 500 replications. The component-wise RE for each estimator was calculated as $\sum_{i=1}^{500} \|\widehat{\boldsymbol{\beta}}_{TR-OLS,i,j} - \boldsymbol{\beta}_{0,j}\|^2 / \sum_{k=1}^{500} \|\widehat{\boldsymbol{\beta}}_{SSQ-SNP,j} - \boldsymbol{\beta}_{0,j}\|^2$, where $j = 1, .., p+1$. During each replication, new vectors $\mathbf{Y}$, $\mathbf{A}$ and matrix $\mathbf{X}$ are simulated. The percent of correct decisions (PCD) was calculated for each method, $PCD_i = 1 - \sum_{k=1}^{5500} \left| I(\widehat{\boldsymbol{\beta}}'\mathbf{X}_k > 0) - I(\boldsymbol{\beta}_{\mathbf{0}}'\mathbf{X}_k > 0) \right| / 5500$, and then averaged over all 500 simulations, $PCD = (1/500) \sum_{i=1}^{500} PCD_i$. The true value function is calculated with the Monte Carlo data set of sample size 500,000 as, $V_0 =$

$\sum_{m=1}^{500{,}000} \left\{\mu(\mathbf{X}_m) + I(C(\mathbf{X}_m) > 0)C(\mathbf{X}_m)\right\}/500,000$. The value function with the estimated regression coefficients is, $\widehat{V} = \sum_{m=1}^{500{,}000} \left\{\mu(\mathbf{X}_m) + \widehat{\delta}_{\boldsymbol{\beta}}^{opt} C(\mathbf{X}_m)\right\}/500,000$. Since we assume a larger response to treatment is preferred, a larger value function signifies an OTD that on average provides the best treatment allocation to the population of interest. A Gaussian kernel was chosen for $W(\cdot)$, and a value of $\mathcal{K} = 5$ was chosen for cross-validation.

Table 1 summarizes the value function and PCD results under all 6 settings. SSQ-SNP outperforms TR-OLS in every setting according to both value function and PCD results. Less reliance on $\pi(\mathbf{X})$ and inclusion of 'unlabeled' subjects indicates SSQ-SNP is a robust estimator for $\boldsymbol{\beta}$.

			TR-OLS		SSQ-SNP	
$\mu(\mathbf{X})$	Model	V_0	V	PCD	V	PCD
$(\boldsymbol{\omega}'\mathbf{X})^3$	Linear	0.56	0.54 (0.04)	0.92 (0.06)	0.56 (0.01)	0.96 (0.02)
	Cubic	0.24	0.22 (0.06)	0.87 (0.12)	0.24 (0.01)	0.93 (0.05)
	Sine	0.32	0.21 (0.12)	0.80 (0.17)	0.26 (0.06)	0.88 (0.09)
$(\boldsymbol{\alpha}'\mathbf{X})(1+\boldsymbol{\omega}'\mathbf{X})$	Linear	1.31	1.29 (0.05)	0.91 (0.06)	1.31 (0.01)	0.96 (0.02)
	Cubic	0.99	0.98 (0.05)	0.86 (0.11)	0.99 (<0.01)	0.94 (0.04)
	Sine	1.06	0.96 (0.11)	0.80 (0.16)	1.02 (0.04)	0.89 (0.06)

Table 1: Mean PCD and Mean Value Function over 500 replications for $\widehat{\delta}_{\boldsymbol{\beta}_{TR}}^{opt}$ and $\widehat{\delta}_{\boldsymbol{\beta}_{SS}}^{opt}$. Empirical SEs are provided in parentheses.

Table 2 presents the bias, empirical SE, asymptotic SE, component-wise RE, and component-wise coverage probabilities of the 95% confidence intervals for SSQ-SNP and TR-OLS when $\mu(\mathbf{X}) = (\boldsymbol{\alpha}'\mathbf{X})(1+\boldsymbol{\omega}'\mathbf{X})$. SSQ-SNP demonstrates its ability to reduce the SE in the regression model regardless of the true underlying model. This translates to a more precise estimator for $\delta_{\boldsymbol{\beta}}^{opt}$.

		TR-OLS				SSQ-SNP				
Model	$\boldsymbol{\beta}$	Bias	ESE	ASE	CP	Bias	ESE	ASE	CP	RE
Linear	0	-0.017	0.242	0.230	0.94	-0.007	0.114	0.116	0.95	1.06
	1	-0.012	0.348	0.326	0.94	0.011	0.150	0.151	0.93	5.37
	1	-0.021	0.352	0.347	0.94	-0.011	0.163	0.154	0.91	4.66
Cubic	0	-0.005	0.236	0.223	0.93	-0.026	0.121	0.122	0.95	1.07
	0.41	-0.014	0.341	0.313	0.93	0.003	0.155	0.154	0.92	1.57
	0.81	-0.004	0.432	0.392	0.91	-0.008	0.193	0.186	0.92	4.99
Sine	0	0.003	0.210	0.202	0.94	0.009	0.117	0.113	0.95	1.24
	0.37	0.002	0.296	0.289	0.94	-0.004	0.146	0.146	0.94	4.12
	0.37	-0.005	0.300	0.290	0.93	0.001	0.136	0.140	0.95	4.90

Table 2: Component-wise Bias, Empirical SE, Asymptotic SE, Coverage Probability (CP), and RE when $\mu(\mathbf{X}) = (\boldsymbol{\alpha}'\mathbf{X})\,(1 + \boldsymbol{\omega}'\mathbf{X})$.

The OTD estimator, $\widehat{\delta}^{opt}_{\boldsymbol{\beta}_{SS}}$, obtained through SSQ-SNP is particularly advantageous when model misspecification occurs according to Table 1. The results suggest $\widehat{\delta}^{opt}_{\boldsymbol{\beta}_{SS}}$ is the preferred estimator for $\delta^{opt}_{\boldsymbol{\beta}}$, especially under varying degrees of model misspecification for the assumed *working* model.

6 Application to an EMR Study

We apply our proposed method, SSQ-SNP, to an EMR study on patients undergoing a hypotensive episode in the ICU within the MIMIC-III database. It is important to treat Hypotensive Episodes (HE's) in ICU patients to minimize end-organ damage. A marker of end-organ damage is a rise in serum creatinine post-hypotensive episode. An initial serum creatinine measurement was taken prior to and within 24 hours of the episode and a post-episode serum creatinine value measurement was performed within 72 hours after the HE. Two treatments for HE's include IV fluid resuscitation and vasopressors. The objective is to determine the optimal treatment between IV fluid resuscitation and vasoactive agent interventions to minimize the rise in serum creatinine post hypotensive episode for each patient.

To formulate this problem into the OTD framework, we must define Y as the negative of the difference between pre-HE and post-HE serum creatinine measurements. The treatment space is defined as $A = 1$ if the patient was given vasopressor treatment, or $A = 0$ if they received IV fluid resuscitation. The beginning of an HE was defined as the time of the first of two consecutive mean arterial pressure (MAP) measurements ≤ 60 *mm Hg*, preceded by two consecutive MAP values > 60 *mm Hg*. The end of an HE was defined as the time of the first of two consecutive MAP measurements > 60 *mm Hg*, preceded by two consecutive MAP values ≤ 60 *mm Hg* (Lee et al. 2012). The predictors, $\mathbf{X}$, include normalized versions of *baseline serum creatinine* measurement prior to and within 24 hours of the HE and *age* of the subject. Patients receiving medical and surgical services during their ICU stay were included in the study, which provided a total of 3,316 individuals. The number of patients with treatment and response available amounted to 1,243 and the number of patients missing treatment and/or response information was 2,073. For this data analysis we are focusing on the semi-supervised setting, therefore we randomly choose $n = 300$ out of 1,243 subjects to create the fully observed set, $\mathcal{O}$.

Propensity scores were modeled with a logistic regression model that included the following covariates; *baseline creatinine, age, gender, service type (medical or surgical)*, *total urine output, mean blood oxygen saturation*, and *average mean arterial pressure.* The last three covariates use recorded information from the 4 hours prior to the HE. We perform TR-OLS and SSQ-SNP on the data set. A Gaussian kernel was chosen for $W(\cdot)$ in Equation 4.1 and $\mathcal{K} = 5$ similar to Section 5. Table 3 presents the point estimates, estimated SE, and p-values for testing null effects. SSQ-SNP reduces the estimated ASE for each predictor.

	TR-OLS			SSQ-SNP		
Predictors	$\boldsymbol{\beta}_{TR}$	ASE	P-Value	$\boldsymbol{\beta}_{SS}$	ASE	P-Value
Intercept	-0.102	0.111	0.355	-0.119	0.102	0.246
Baseline Creatinine	-0.371	0.256	0.147	-0.508	0.197	0.010
Age	0.196	0.146	0.178	0.114	0.133	0.392

Table 3: Estimated regression coefficients and ASE. P-values for testing $H_0 : \boldsymbol{\beta}_{TR} = 0$ and $\boldsymbol{\beta}_{SS} = 0$.

In Table 4, we demonstrate treatment allocation given by the decision rules produced by the two methods on all subjects in the data set. The OTD's estimated by SSQ-SNP and TR-OLS produce similar treatment decisions for the vast majority of subjects. SSQ-SNP assigns more patients to a vasopressor treatment than TR-OLS and is less likely to assign IV fluid resuscitation. The point estimates are relatively close, but the smaller ASE for SSQ-SNP suggests it is the more reliable estimator.

		SSQ-SNP	
	Treatment	IV Fluid	Vasopressors
TR-OLS	IV Fluid	1553	314
	Vasopressors	119	1330

Table 4: Treatment Allocation Given by SSQ-SNP and TR-OLS.

7 Discussion

We proposed a new method for estimating the optimal treatment decision within a specified class of rules at a single decision point, where the class is chosen based on considerations of simplicity and interpretation. Extensive simulation shows the proposed method, SSQ-SNP, outperforms a supervised regression method under the correct specification and misspecification of the assumed *working* model. An application to an EMR study illustrates SSQ-SNP improvements in efficiency of the linear regression coefficients involved in the optimal treatment decision.

This method does have limitations. It is not designed directly for high dimensional data as the *curse of dimensionality* is a problem with kernel regression estimators. Dimension reduction techniques, such as principal component analysis or sliced inverse regression, can be utilized to reduce information provided by all the covariates into a lower-dimensional subspace for accurate prediction with kernel regression estimators, but this eliminates the ability to make simple interpretations from the decision rules created by SSQ-SNP. An appropriate incorporation of high-dimensional data is necessary to fully utilize the available information from EMR studies. This work may be extended to create dynamic optimal treatment regimes utilizing Electronic Health Record data that has information on patients with extended periods of stay or multiple visits to the ICU.

8 Appendix

Proof of Theorem 1. Let $\widehat{C}(\mathbf{x}) = \widehat{Q}^{(np)}(\mathbf{x},1) - \widehat{Q}^{(np)}(\mathbf{x},0)$.

$$\begin{aligned}
\left(\widehat{\boldsymbol{\beta}}_{np} - \boldsymbol{\beta}\right) &= \Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{\widehat{C}(\mathbf{X}_j) - \boldsymbol{\beta}'\widetilde{\mathbf{X}}_j\right\}\right] = \Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{\widehat{C}(\mathbf{X}_j) - C(\mathbf{X}_j)\right\}\right] \\
&+ \Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{C(\mathbf{X}_j) - \boldsymbol{\beta}'\widetilde{\mathbf{X}}_j\right\}\right] = \Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{X}) - C(\mathbf{X})\right\}\right] + O_p(N^{-1/2}).
\end{aligned}$$

The first step follows from the normal equations. The last step is due to the fact

$\Lambda_N^{-1}\left[N^{-1}\sum_{j=n+1}^{n+N}\mathbf{X}_j\left\{\widehat{C}(\mathbf{X}_j)-C(\mathbf{X}_j)\right\}\right] = \Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{X})-C(\mathbf{X})\right\}\right] + o_p(1)$ by standard arguments involving the weak law of large numbers, and according to the central limit theorem, $N^{-1/2}\left[N^{-1/2}\sum_{j=n+1}^{N}\Lambda_n^{-1}\left\{C(\mathbf{X})_j-\boldsymbol{\beta}'\widetilde{\mathbf{X}}_j\right\}\right] = O_p(N^{-1/2})$. Multiplying both sides by $n^{1/2}$ we have,

$$\begin{aligned}
n^{1/2}\left(\widehat{\boldsymbol{\beta}}_{np}-\boldsymbol{\beta}\right) &= n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{x})-C(\mathbf{X})\right\}\right]+O_p\left((n/N)^{\frac{1}{2}}\right)\\
&= n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},1)-Q^{(np)}(\mathbf{X},1)\right\}\right]\\
&- n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},0)-Q^{(np)}(\mathbf{X},0)\right\}\right]+O_p\left((n/N)^{\frac{1}{2}}\right).
\end{aligned}$$

Note that $n/N \rightarrow 0$ implying $O_p\left((n/N)^{\frac{1}{2}}\right) \equiv o_p(1)$. Next, let $\tau(\mathbf{X}) = \pi(\mathbf{X})f(\mathbf{X})$ and $\widehat{\tau}(\mathbf{X}) = \frac{1}{nh^p}\sum_{i=1}^{n}A_iW_h(\mathbf{X}_i-\mathbf{X})$, where $W_h(\mathbf{X}_i-\mathbf{X}) = W(\frac{\mathbf{X}_i-\mathbf{X}}{h})$. Let's rewrite $E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},1)-Q^{(np)}(\mathbf{X},1)\right\}\right]$ as,

$$= E\left\{\frac{\frac{1}{nh^p}\sum_{i=1}^{n}A_i\mathbf{X}W_h(\mathbf{X}_i-\mathbf{X})\left\{Y_i-Q^{(np)}(\mathbf{X},1)\right\}}{\tau(\mathbf{X})}\right\} \tag{8.1}$$

$$+E\left\{\mathbf{X}\left(\widehat{Q}^{(np)}(\mathbf{X},1)-Q^{(np)}(\mathbf{X},1)\right)\left\{\frac{\tau(\mathbf{X})-\widehat{\tau}(\mathbf{X})}{\tau(\mathbf{X})}\right\}\right\} = H_{n,1}^{(1)}+H_{n,2}^{(1)}. \tag{8.2}$$

Then $H_{n,1}^{(1)}$ is equivalent to:

$$\begin{aligned}
&= \frac{1}{nh^p}\sum_{i=1}^{n}A_i\int\mathbf{X}\left\{Y_i-Q^{(np)}(\mathbf{X},1)\right\}\frac{W_h(\mathbf{X}-\mathbf{X}_i)}{\tau(\mathbf{X})}f(\mathbf{X})d\mathbf{X}\\
&= \frac{1}{nh^p}\sum_{i=1}^{n}A_i\int\mathbf{X}\left\{Y_i-Q^{(np)}(\mathbf{X},1)\right\}\frac{W_h(\mathbf{X}-\mathbf{X}_i)}{\pi(\mathbf{X})}d\mathbf{X}\\
&= \frac{1}{n}\sum_{i=1}^{n}A_i\int(\mathbf{X}_i+h\mathbf{t}_i)\left\{Y_i-Q^{(np)}(\mathbf{X}_i+h\mathbf{t}_i,1)\right\}\frac{W(\mathbf{t}_i)}{\pi\left(\mathbf{X}_i+h\mathbf{t}_i\right)}d\mathbf{t}_i
\end{aligned}$$

By *assumptions* (1) and (2), Taylor expansion in $h\mathbf{t}_i$ for sufficiently small h leads to,

$$H_{n,1}^{(1)} = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{\pi\left(\mathbf{X}_i\right)}\mathbf{X}_i\left\{Y_i-Q^{(np)}(\mathbf{X}_i,1)\right\}+O_p(h^r).$$

Since $n^{1/2}h^r \rightarrow 0$ as $n\rightarrow\infty$,

$$n^{1/2}\Lambda H_{n,1}^{(1)} = \widetilde{H}_{n,1}^{(1)} = n^{-1/2}\sum_{i=1}^{n}\frac{A_i}{\pi\left(\mathbf{X}_i\right)}\Lambda\mathbf{X}_i\left\{Y_i-Q^{(np)}(\mathbf{X}_i,1)\right\}+o_p(1). \tag{8.3}$$

Let $q(\mathbf{X}) = \widehat{Q}^{(np)}(\mathbf{X},1)-Q^{(1)}(\mathbf{X},1)$, $l(\mathbf{X}) = \dfrac{\tau(\mathbf{X})-\widehat{\tau}(\mathbf{X})}{\tau(\mathbf{X})} = 1-\dfrac{\widehat{\tau}(\mathbf{X})}{\tau(\mathbf{X})}$, and $Q^{(np)}(\mathbf{X},1) =$

$\alpha(\mathbf{X})/\tau(\mathbf{X})$, where $\alpha(\mathbf{X})$ is the numerator in Equation 8.1. It follows that $H^{(1)}_{n,2} = E\left\{\mathbf{X}q(\mathbf{X})l(\mathbf{X})\right\}$ and,

$$E\left\{\mathbf{X}q(\mathbf{X})l(\mathbf{X})\right\} \leq \sup_{\mathbf{x}\in\mathcal{X}}\left\{\|\mathbf{X}\|\ |q(\mathbf{X})|\ |l(\mathbf{X})|\right\} = o_p(1). \tag{8.4}$$

Equation 8.4 requires that $\mathbf{X}$ is bounded, $\sqrt{\frac{\ln n}{nh^p}} \to 0$ as $n \to \infty$, as well as *assumptions* (3) and (5). It follows by similar argument of *Lemma B.1* in Newey (1994) combined with Taylor series expansion that $\sup_{\mathbf{x}\in\mathcal{X}}|\widehat{\tau}(\mathbf{X}) - \tau(\mathbf{X})| = \sup_{\mathbf{x}\in\mathcal{X}}|\widehat{\alpha}(\mathbf{X}) - \alpha(\mathbf{X})| = o_p(1)$. Afterwards, it holds that $\sup_{\mathbf{x}\in\mathcal{X}}|q(\mathbf{X})| = o_p(1)$ through similar reasoning given by *Theorem 8* of Hansen (2008). The same technique proves,

$$n^{1/2}\Lambda E\left[\mathbf{X}\left\{\widehat{Q}^{(np)}(\mathbf{X},0) - Q^{(np)}(\mathbf{X},0)\right\}\right] = \widetilde{H}^{(0)}_{n,1} + o_p(1).$$

This leaves us with,

$$n^{1/2}\Lambda^{-1}E\left[\mathbf{X}\left\{\widehat{C}(\mathbf{x}) - C(\mathbf{X})\right\}\right] = \widetilde{H}^{(1)}_{n,1} - \widetilde{H}^{(0)}_{n,1} + o_p(1) = n^{-1/2}\sum_{i=1}^{n}\Psi_{np}(\mathbf{O}_i) + o_p(1).$$

Since $\Psi_{np}(\mathbf{O})$ is the influence function for $\widehat{\boldsymbol{\beta}}_{np}$ with $E\left[\Psi_{np}(\mathbf{O})\right] = 0$ and variance $V_{\boldsymbol{\beta}_{np}} = E\left[\Psi_{np}(\mathbf{O})\Psi'_{np}(\mathbf{O})\right]$, the *CLT* states it will converge to $\mathcal{N}_{p+1}(0, V_{\boldsymbol{\beta}_{np}})$. □

Proof of Theorem 2. The proof of Theorem 2 follows similar logic to the proof of *Theorem 3.2* from Chakrabortty & Cai (2018). □

References

Chakrabortty, A. & Cai, T. (2018), 'Efficient and adaptive linear regression in semi-supervised settings', *The Annals of Statistics* **46**(4), 1541–1572.

Chapelle, O., Scholkopf, B. & Zien, A. (2006), *Semi-Supervised Learning*, MIT Press, Cambridge, MA, USA.

Fan, J. (1992), 'Design-adaptive nonparametric regression', *Journal of the American statistical Association* **87**(420), 998–1004.

Hansen, B. E. (2008), 'Uniform convergence rates for kernel estimation with dependent data', *Econometric Theory* **24**(3), 726–748.

Johnson, A. E. et al. (2016), 'Mimic-iii, a freely accessible critical care database', *Scientific data* **3**, 160035.

Lee, J., Kothari, R., Ladapo, J. A., Scott, D. J. & Celi, L. A. (2012), 'Interrogating a clinical database to study treatment of hypotension in the critically ill', *BMJ Open* **2**(3).

Moodie, E. E., Richardson, T. S. & Stephens, D. A. (2007), 'Demystifying optimal dynamic treatment regimes', *Biometrics* **63**(2), 447–455.

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

Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., Waxmonsky, J. G., Yu, J. & Murphy, S. A. (2012), 'Q-learning: A data analysis method for constructing adaptive interventions.', *Psychological Methods* **17**(4), 478.

Newey, W. K. (1994), 'Kernel estimation of partial means and a general variance estimator', *Econometric Theory* **10**(2), 1–21.

Raghu, A., Komorowski, M., Celi, L. A., Szolovits, P. & Ghassemi, M. (2017), 'Continuous state-space models for optimal sepsis treatment-a deep reinforcement learning approach', *arXiv preprint arXiv:1705.08422* .

Robins, J. M. (2004), Optimal Structural Nested Models for Optimal Sequential Decisions, *in* D. Y. Lin & P. J. Heagerty, eds, 'Proceedings of the Second Seattle Symposium in Biostatistics', New York: Springer.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.

Rubin, D. B. (1978), 'Bayesian inference for causal effects: The role of randomization', *The Annals of Statistics* **6**(1), 34–58.

Schulte, P., Tsiatis, A., Laber, E. & Davidian, M. (2014), 'Q-and A-learning methods for estimating optimal dynamic treatment regimes', *Statistical Science: a Review Journal of the Institute of Mathematical Statistics* **29**(4), 640–661.

Wang, Y., Wu, P., Liu, Y., Weng, C. & Zeng, D. (2016), 'Learning optimal individualized treatment rules from electronic health record data', *2016 IEEE International Conference on Healthcare Informatics (ICHI)* pp. 65–71.

Zhang, B., Tsiatis, A. A., Laber, E. B. & Davidian, M. (2012), 'A robust method for estimating optimal treatment regimes', *Biometrics* **68**(4), 1010–1018.

Zhao, Y., Zeng, D., Rush, A. J. & Kosorok, M. R. (2012), 'Estimating individualized treatment rules using outcome weighted learning', *Journal of the American Statistical Association* **107**(499), 1106–1118.