Journal of Machine Learning Research 24 (2023) 1-86 Submitted 2/21; Revised 8/23; Published 10/23

Semi-Supervised Off-Policy Reinforcement Learning and Value Estimation for Dynamic Treatment Regimes

Aaron Sonabend-W ASONABEND@GOOGLE.COM
Department of Biostatistics
Harvard University
Boston, USA

Nilanjana Laha NLAHA@TAMU.EDU
Department of Statistics
Texas A& M
Texas, USA

Ashwin N. Ananthakrishnan AANANTHAKRISHNAN@MGH.HARVARD.EDU
Division of Gastroenterology
Massachusetts General Hospital
Boston, USA

Tianxi Cai TCAI@HSPH.HARVARD.EDU
Rajarshi Mukherjee RAM521@MAIL.HARVARD.EDU
Department of Biostatistics
Harvard University
Boston, USA

Editor: David Sontag

Abstract

Reinforcement learning (RL) has shown great promise in estimating dynamic treatment regimes which take into account patient heterogeneity. However, health-outcome information, used as the reward for RL methods, is often not well coded but rather embedded in clinical notes. Extracting precise outcome information is a resource-intensive task, so most of the available well-annotated cohorts are small. To address this issue, we propose a semi-supervised learning (SSL) approach that efficiently leverages a small-sized labeled data set with actual outcomes observed and a large unlabeled data set with outcome surrogates. In particular, we propose a semi-supervised, efficient approach to Q-learning and doubly robust off-policy value estimation. Generalizing SSL to dynamic treatment regimes brings interesting challenges: 1) Feature distribution for Q-learning is unknown as it includes previous outcomes. 2) The surrogate variables we leverage in the modified SSL framework are predictive of the outcome but not informative of the optimal policy or value function. We provide theoretical results for our Q function and value function estimators to understand the degree of efficiency gained from SSL. Our method is at least as efficient as the supervised approach, and robust to bias from mis-specification of the imputation models.

Keywords: semi-supervised learning, Q-learning, reinforcement-learning, dynamical treatment regime, doubly robust value function, off-policy learning

1. Introduction

Finding optimal treatment strategies incorporating patient heterogeneity is a cornerstone of personalized medicine. When treatment options change over time, optimal dynamic treatment regimes (DTR) can be learned using longitudinal patient data. With the increasing availability of large-scale longitudinal data such as electronic health records (EHR) data in recent years, reinforcement learning (RL) has shown promise in estimating such optimal DTR (Kosorok and Laber, 2019; Schulte et al., 2014; Sonabend-W et al., 2020b; Zhou et al., 2023; Chakraborty and Moodie, 2013). Existing RL methods include G-estimation (Robins, 2004), Q-learning (Watkins, 1989; Murphy, 2005), A-learning (Murphy, 2003) and directly maximizing the value function (Zhao et al., 2015). G-estimation and A-learning attempt to model only the component of the outcome regression relevant to the treatment contrast, while Q-learning posits complete models for the outcome regression. Although G-estimation and A-learning models can be more efficient and robust to mis-specification, Q-learning is widely adopted due to its ease of implementation, flexibility, and interpretability (Watkins, 1989; Chakraborty and Moodie, 2013; Schulte et al., 2014).

However, learning DTR with EHR data often faces an additional challenge of whether outcome information is readily available. Outcome information, such as the development of a clinical event or whether a patient is considered to have responded, is often not well coded but rather embedded in clinical notes. Proxy variables, such as diagnostic codes or mentions of relevant clinical terms in clinical notes via natural language processing (NLP), while predictive of the true outcome, are often not sufficiently accurate to be used directly in place of the outcome (Hong et al., 2019; Zhang et al., 2019; Sonabend W. et al., 2020; Cheng et al., 2020). On the other hand, extracting precise outcome information often requires manual chart review, which is resource intensive, particularly when the outcome needs to be annotated over time. This challenge indicates the need for a semi-supervised learning (SSL) approach that can efficiently leverage a small-sized labeled data $\mathcal{L}$ with true outcome observed, and a large-sized unlabeled data $\mathcal{U}$ for predictive modeling. It is worthwhile to note that the SSL setting differs from the standard missing data setting in that the probability of missing tends to 1 asymptotically, which violates the positivity assumption required by the classical missing data methods (Chakrabortty et al., 2018).

While SSL methods have been well developed for prediction, classification, and regression tasks (e.g. Chapelle et al., 2006; Zhu, 2008; Blitzer and Zhu, 2008; Zhixing and Shaohong, 2011; Qiao et al., 2018; Chakrabortty et al., 2018), there is a paucity of literature on SSL methods for estimating optimal treatment rules. Recently, Cheng et al. (2020) and Kallus and Mao (2020) proposed SSL methods for estimating an average causal treatment effect. Finn et al. (2016) proposed a semi-supervised RL method that achieves impressive empirical results and outperforms simple approaches such as direct imputation of the reward. However, there are no theoretical guarantees, and the approach lacks causal validity and interpretability within a domain context. Additionally, this method does not leverage available surrogates. In this work, we fill this gap by proposing a theoretically justified SSL approach to Q-learning using a large set of unlabeled data $\mathcal{U}$, which contains sequential observations on features $\mathbf{O}$, treatment assignment A, and surrogates $\mathbf{W}$ that are imperfect proxies of Y, as well as a small set of labeled data $\mathcal{L}$ which contains true outcome Y at multiple stages along with $\mathbf{O}$, A and $\mathbf{W}$. We will also develop a robust and efficient

SSL approach to estimating the value function of the derived optimal DTR, defined as the expected counterfactual outcome under the derived DTR.

To describe the main contributions of our proposed SSL approach to RL, we first note two crucial distinctions between the proposed framework and classical SSL methods. First, existing SSL literature often assumes that $\mathcal{U}$ is large enough that the feature distribution is known (Wasserman and Lafferty, 2008). However, under the RL setting, the outcome of the stage $t-1$, denoted by Y_{t-1}, becomes a feature of stage t for predicting Y_t. As such, the feature distribution for predicting Y_t can not be viewed as known in the Q-learning procedure. Our methods for estimating an optimal DTR and its associated value function carefully adapt to this sequentially missing data structure. We do not make a Markov decision process (MDP) assumption or have a partially observed MDP framework (Kaelbling et al., 1998), as biological mechanisms of the patient's features or treatment may take more than one time step to manifest in the outcome. Second, we modify the SSL framework to handle the use of surrogate variables $\mathbf{W}$, which are predictive of the outcome through the joint law $\mathbb{P}_{Y,\mathbf{O},A,\mathbf{W}}$, but are not part of the conditional distribution of interest $\mathbb{P}_{Y|\mathbf{O},A}$. To address these issues, we propose a two-step fitting procedure for finding an optimal DTR and estimating its value function in the SSL setting. Our method consists of using the outcome surrogates ($\mathbf{W}$) and features ($\mathbf{O}, A$) for non-parametric estimation of the missing outcomes (Y). We subsequently use these imputations to estimate Q functions, learn the optimal treatment rule and estimate its associated value function. We provide theoretical results to understand when and to what degree efficiency can be gained from $\mathbf{W}$ and $\mathbf{O}, A$.

We further show that our approach is robust to mis-specification of the imputation models. To account for potential mis-specification in the models for the Q function, we provide a double robust value function estimator for the derived DTR. If either the regression models for the Q functions or the propensity score functions are correctly specified, our value function estimators are consistent for the true value function.

We organize the rest of the paper as follows. In Section 2 we formalize the problem mathematically and provide some notation to be used in developing and analyzing the methods. In Section 3, we discuss traditional Q-learning and propose an SSL estimation procedure for the optimal DTR. Section 4 details an SSL doubly robust estimator of the value function for the derived DTR. In Section 5, we provide theoretical guarantees for our approach and discuss the implications of our assumptions and results. Section 6 is devoted to numerical experiments and real data analysis with an inflammatory bowel disease (IBD) data set. We end with a discussion of the methods and possible extensions in Section 7. The proposed method has been implemented in R, and the code can be found at `github.com/asonabend/SSOPRL`. A reference table with the main notation used can be found in Appendix A. An extension of the SSL algorithms to a general time horizon and simulations can be found in appendices B and C. Finally, all the technical proofs and supporting lemmas are collected in Appendices D and E.

2. Problem Setup

We consider a longitudinal observational study with outcomes, confounders and treatment indices potentially available over multiple stages. The proposed semi-supervised methods

are valid for a general time horizon, we describe them in detail in Appendix B and show empirical results in Section 6. However, for ease of presentation, in the main text we will use two time points of (binary) treatment allocation as follows. For time point $t \in \{1, 2\}$, let $\mathbf{O}_t \in \mathbb{R}^{d_t^o}$ denote the vector of covariates measured prior at stage t of dimension d_t^o; $A_t \in \{0, 1\}$ a treatment indicator variable; and $Y_{t+1} \in \mathbb{R}$ the outcome observed at stage $t + 1$, for which higher values of Y_{t+1} are considered beneficial. Additionally we observe surrogates $\mathbf{W}_t \in \mathbb{R}^{d_t^\omega}$, a d_t^ω-dimensional vector of post-treatment covariates potentially predictive of Y_{t+1}. In the labeled data where $\mathbf{Y} = (Y_2, Y_3)^\intercal$ is annotated, we observe a random sample of n independent and identically distributed (iid) random vectors, denoted by

$$\mathcal{L} = \{\mathbf{L}_i = (\vec{\mathbf{U}}_i^\intercal, \mathbf{Y}_i^\intercal)^\intercal\}_{i=1}^n, \quad \text{where } \mathbf{U}_{ti} = (\mathbf{O}_{ti}^\intercal, A_{ti}, \mathbf{W}_{ti}^\intercal)^\intercal \text{ and } \vec{\mathbf{U}}_i = (\mathbf{U}_{1i}^\intercal, \mathbf{U}_{2i}^\intercal)^\intercal.$$

We additionally observe an unlabeled set consisting of N iid random vectors,

$$\mathcal{U} = \{\vec{\mathbf{U}}_j\}_{j=1}^N$$

with $N \gg n$. We denote the entire data as $\mathbb{S} = (\mathcal{L} \cup \mathcal{U})$. To operationalize our statistical arguments we denote the joint distribution of the observation vector $\mathbf{L}_i$ in $\mathcal{L}$ as $\mathbb{P}$. In order to connect to the unlabeled set, we assume that any observation vector $\vec{\mathbf{U}}_j$ in $\mathcal{U}$ has the distribution induced by $\mathbb{P}$.

We are interested in finding the optimal DTR and estimating its *value function* to be defined as expected counterfactual outcomes under the derived regime. To this end, let $Y_{t+1}^{(a)}$ be the potential outcome for a patient at time $t + 1$ had the patient been assigned at time t to treatment $a \in \{0, 1\}$. A dynamic treatment regime is a set of functions $\mathcal{D} = (d_1, d_2)$, where $d_t(\cdot) \in \{0, 1\}$, $t = 1, 2$ map from the patient's history up to time t to the treatment choice $\{0, 1\}$. We define the patient's history as $\mathbf{H}_1 \equiv [\mathbf{H}_{10}^\intercal, \mathbf{H}_{11}^\intercal]^\intercal$ with $\mathbf{H}_{1k} = \boldsymbol{\phi}_{1k}(\mathbf{O}_1)$, $\mathbf{H}_2 = [\mathbf{H}_{20}^\intercal, \mathbf{H}_{21}^\intercal]^\intercal$ with $\mathbf{H}_{2k} = \boldsymbol{\phi}_{2k}(\mathbf{O}_1, A_1, \mathbf{O}_2)$, where $\{\boldsymbol{\phi}_{tk}(\cdot), t = 1, 2, k = 0, 1\}$ are pre-specified basis functions. We then define features derived from patient history for regression modeling as $\mathbf{X}_1 \equiv [\mathbf{H}_{10}^\intercal, A_1\mathbf{H}_{11}^\intercal]^\intercal$ and $\mathbf{X}_2 \equiv [\mathbf{H}_{20}^\intercal, A_2\mathbf{H}_{21}^\intercal]^\intercal$. For ease of presentation, we also use the check symbol (i.e., $\check{\mathbf{H}}_2$) to denote vectors that contain outcome Y_2 when applicable. Hence, we let $\check{\mathbf{H}}_2 = (Y_2, \mathbf{H}_2^\intercal)^\intercal$, $\check{\mathbf{X}}_2 = (Y_2, \mathbf{X}_2^\intercal)^\intercal$, for consistency we also write $\check{\mathbf{H}}_1 = \mathbf{H}_1$, $\check{\mathbf{X}}_1 = \mathbf{X}_1$, and finally we define $\boldsymbol{\Sigma}_t = \mathbb{E}[\check{\mathbf{X}}_t\check{\mathbf{X}}_t^\intercal]$. We collect this and the main notation used throughout the paper in Table 8, Appendix A.

Let $\mathbb{E}_\mathcal{D}$ be the expectation with respect to the measure that generated the data under regime $\mathcal{D}$. Then these sets of rules $\mathcal{D}$ have an associated value function which we can write as $V(\mathcal{D}) = \mathbb{E}_\mathcal{D}\left[Y_2^{(d_1)} + Y_3^{(d_2)}\right]$. Thus, an optimal dynamic treatment regime is a rule $\bar{\mathcal{D}} = (\bar{d}_1, \bar{d}_2)$ such that $\bar{V} = V\left(\bar{\mathcal{D}}\right) \geq V\left(\mathcal{D}\right)$ for all $\mathcal{D}$ in a suitable class of admissible decisions (Chakraborty and Moodie, 2013). To identify $\bar{\mathcal{D}}$ and $\bar{V}$ from the observed data we will require the following sets of standard assumptions (Robins, 1997; Schulte et al., 2014): (i) consistency – $Y_{t+1} = Y_{t+1}^{(0)}I(A_t = 0) + Y_{t+1}^{(1)}I(A_t = 1)$ for $t = 1, 2$, (ii) sequential ignorability, also known as no unmeasured confounding – for outcomes, intermediate covariates and surrogates: $Y_{t+1}^{(a)}, \mathbf{O}_t^{(a)}, \mathbf{W}_t^{(a)} \perp\!\!\!\perp A_t|\mathbf{H}_t$ for $a \in \{0, 1\}$, $t = 1, 2$ (iii) positivity – $\mathbb{P}(A_t|\mathbf{H}_t) > \nu$, for $t = 1, 2$, $A_t \in \{0, 1\}$, for some fixed $\nu > 0$.

We will develop SSL inference methods to derive optimal DTR $\bar{\mathcal{D}}$ as well the associated value function $\bar{V}$ by leveraging the richness of the unlabeled data and the predictive

power of surrogate variables which allows us to gain crucial statistical efficiency. Our main contributions in this regard can be described as follows. First, we provide a systematic generalization of the Q-learning framework, with theoretical guarantees to the semi-supervised setting with improved efficiency. Second, we provide a doubly robust estimator of the value function in the semi-supervised setup. Third, our Q-learning procedure and value function estimator are flexible enough to allow for standard off-the-shelf machine learning tools and are shown to perform well in finite-sample numerical examples.

3. Semi-Supervised Q-Learning

In this section we propose a semi-supervised Q-learning approach to derive an optimal DTR. To this end, we first recall the basic mechanism of traditional linear parametric Q-learning (Chakraborty and Moodie, 2013) and then detail our proposed method. We defer the theoretical guarantees to Section 5.

3.1 Traditional Q-Learning

Q-learning is a backward recursive algorithm that identifies optimal DTR by optimizing two stage Q functions. Under the consistency, sequential ignorability and positivity assumptions (i)-(iii), the optimal treatment is identifiable and the Q functions can be expressed as:

$$Q_2(\check{\mathbf{H}}_2, A_2) \equiv \mathbb{E}[Y_3|\check{\mathbf{H}}_2, A_2], \quad \text{and} \quad Q_1(\check{\mathbf{H}}_1, A_1) \equiv \mathbb{E}[Y_2 + \max_{a_2} Q_2(\check{\mathbf{H}}_2, a_2)|\check{\mathbf{H}}_1, A_1]$$

(Sutton, 2018; Murphy, 2005). In order to perform inference one typically proceeds by positing models for the Q functions. In its simplest form one assumes a (working) linear model for some parameters $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t^\intercal, \boldsymbol{\gamma}_t^\intercal)^\intercal$, $t = 1, 2$, as follows:

$$\begin{aligned} Q_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\theta}_1^0) =& \check{\mathbf{X}}_1^\intercal \boldsymbol{\theta}_1^0 = \mathbf{H}_{10}^\intercal \boldsymbol{\beta}_1^0 + A_1(\mathbf{H}_{11}^\intercal \boldsymbol{\gamma}_1^0), \\ Q_2(\check{\mathbf{H}}_2, A_2; \boldsymbol{\theta}_2^0) =& \check{\mathbf{X}}_2^\intercal \boldsymbol{\theta}_2^0 = Y_2\beta_{21}^0 + \mathbf{H}_{20}^\intercal \boldsymbol{\beta}_{22}^0 + A_2(\mathbf{H}_{21}^\intercal \boldsymbol{\gamma}_2^0). \end{aligned} \tag{1}$$

Typical Q-learning consists of performing a least squares regression for the second stage to estimate $\widehat{\boldsymbol{\theta}}_2$ followed by defining the stage 1 pseudo-outcome for $i = 1, ..., n$ as

$$\widehat{Y}_{2i}^* = Y_{2i} + \max_{a_2} Q_2(\check{\mathbf{H}}_{2i}, a_2; \widehat{\boldsymbol{\theta}}_2) = Y_{2i}(1 + \hat{\beta}_{21}) + \mathbf{H}_{20i}^\intercal \widehat{\boldsymbol{\beta}}_{22} + [\mathbf{H}_{21i}^\intercal \widehat{\boldsymbol{\gamma}}_2]_+,$$

where $[x]_+ = xI(x > 0)$. One then proceeds to estimate $\widehat{\boldsymbol{\theta}}_1$ using least squares again, with $\widehat{Y}_2^*$ as the outcome variable. Indeed, valid inference on $\widehat{\mathcal{D}}$ using the method described above crucially depends on the validity of the model assumed. However as we shall see, even without validity of this model we will be able to provide valid inference on suitable analogues of the Q function working model parameters, and on the value function using a double robust type estimator. To that end it will be instructive to define the least square projections of Y_3 and Y_2^* onto $\check{\mathbf{X}}_2$ and $\check{\mathbf{X}}_1$ respectively. The linear regression working models given by (1) have $\boldsymbol{\theta}_1^0$, $\boldsymbol{\theta}_2^0$ as unknown regression parameters. To account for the potential mis-specification of the working models in (1), we define the target population parameters $\bar{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2$ as the population solutions to the expected normal equations

$$\mathbb{E}\left\{\check{\mathbf{X}}_1(\bar{Y}_2^* - \check{\mathbf{X}}_1^\intercal \bar{\boldsymbol{\theta}}_1)\right\} = \mathbf{0}, \quad \text{and} \quad \mathbb{E}\left\{\check{\mathbf{X}}_2\left(Y_3 - \check{\mathbf{X}}_2^\intercal \bar{\boldsymbol{\theta}}_2\right)\right\} = \mathbf{0},$$

where $\bar{Y}_2^* = Y_2 + \max_{a_2} Q_2(\check{\mathbf{H}}_2, a_2; \bar{\boldsymbol{\theta}}_2)$. As these are linear in the parameters, uniqueness and existence for $\bar{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2$ are well defined. In fact, $Q_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\theta}}_1) = \check{\mathbf{X}}_1^\intercal \bar{\boldsymbol{\theta}}_1, Q_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\theta}}_2) = \check{\mathbf{X}}_2^\intercal \bar{\boldsymbol{\theta}}_2$ are the L_2 projection of $\mathbb{E}(Y_2^*|\check{\mathbf{X}}_1) \in \mathcal{L}_2\left(\mathbb{P}_{\check{\mathbf{X}}_1}\right)$, $\mathbb{E}(Y_3|\check{\mathbf{X}}_2) \in \mathcal{L}_2\left(\mathbb{P}_{\check{\mathbf{X}}_2}\right)$ onto the subspace of all linear functions of $\check{\mathbf{X}}_1, \check{\mathbf{X}}_2$ respectively. Therefore, Q functions in (1) are the best linear predictors of $\bar{Y}_2^*$ conditional on $\check{\mathbf{X}}_1$ and Y_3 conditional on $\check{\mathbf{X}}_2$.

Traditionally, one only has access to labeled data $\mathcal{L}$, and hence proceeds to estimate $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in (1) by solving the following sample version set of normal equations:

$$\begin{aligned} \mathbb{P}_n\left[\check{\mathbf{X}}_2(Y_3 - \check{\mathbf{X}}_2^\intercal \boldsymbol{\theta}_2)\right] \equiv \mathbb{P}_n \begin{bmatrix} Y_2\{Y_3 - (Y_2, \mathbf{X}_2^\intercal)\boldsymbol{\theta}_2\} \\ \mathbf{X}_2\{Y_3 - (Y_2, \mathbf{X}_2^\intercal)\boldsymbol{\theta}_2\} \end{bmatrix} &= \mathbf{0}, \\ \mathbb{P}_n\left[\mathbf{X}_1\{Y_2(1+\beta_{21}) + \mathbf{H}_{20}^\intercal \boldsymbol{\beta}_{22} + [\mathbf{H}_{21}^\intercal \boldsymbol{\gamma}_2]_+ - \mathbf{X}_1^\intercal \boldsymbol{\theta}_1\}\right] &= \mathbf{0}. \end{aligned} \tag{2}$$

(Chakraborty and Moodie, 2013), where $\mathbb{P}_n$ denotes the empirical measure: i.e. for a measurable function $f : \mathbb{R}^p \mapsto \mathbb{R}$ and random sample $\{\mathbf{L}_i\}_{i=1}^n$, $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(\mathbf{L}_i)$. The asymptotic distribution for the Q function parameters in the fully-supervised setting has been well studied (see Laber et al., 2014). It is worth recalling that we use the check symbol: "$\check{}$" to denote the entire set of features available, including outcomes when available (for $t = 2$), this symbol is used for both the patient history $\check{\mathbf{H}}_t$, and the linear features for the $Q-$functions $\check{\mathbf{X}}_t$. We also use *bar* (i.e., $\bar{\boldsymbol{\theta}}$) over the variables to denote population parameters, and we use *hat* (i.e., $\widehat{\boldsymbol{\theta}}$) over the variables to denote estimated parameters. For more detail on notation see Table 8.

3.2 Semi-Supervised Q-Learning

We next detail our robust imputation-based semi-supervised Q-learning approach that leverages the unlabeled data $\mathcal{U}$ to replace the unobserved Y_t in (2) with their properly imputed values for subjects in $\mathcal{U}$. Our SSL procedure includes three key steps: (i) imputation, (ii) refitting, and (iii) projection to the unlabeled data. In step (i), we develop flexible imputation models for the conditional mean functions $\{\mu_t(\cdot), \mu_{2t}(\cdot), t = 2, 3\}$, where $\mu_t(\vec{\mathbf{U}}) = \mathbb{E}(Y_t|\vec{\mathbf{U}})$ and $\mu_{2t}(\vec{\mathbf{U}}) = \mathbb{E}(Y_2 Y_t|\vec{\mathbf{U}})$. The refitting in step (ii) will ensure the validity of the SSL estimators under potential mis-specifications of the imputation models.

Step I: Imputation

Our first imputation step involves weakly parametric or non-parametric prediction modeling to approximate the conditional mean functions $\{\mu_t(\cdot), \mu_{2t}(\cdot), t = 2, 3\}$. Commonly used models such as non-parametric kernel smoothing, basis function expansion or kernel machine regression can be used. We denote the corresponding estimated mean functions as $\{\widehat{m}_t(\cdot), \widehat{m}_{2t}(\cdot), t = 2, 3\}$ under the corresponding imputation models $\{m_t(\vec{\mathbf{U}}), m_{2t}(\vec{\mathbf{U}}), t = 2, 3\}$. Theoretical properties of our proposed SSL estimators on specific choices of the imputation models are provided in Section 5. We also provide additional simulation results comparing different imputation models in Section 6.

Step II: Refitting

To overcome the potential bias in the fitting from the imputation model, especially under model mis-specification, we update the imputation model with an additional refitting step by expanding it to include linear effects of $\{\mathbf{X}_t, t = 1, 2\}$ with cross-fitting to control overfitting bias. Specifically, to ensure the validity of the SSL algorithm from the refitted imputation model, we note that the final imputation models for $\{Y_t, Y_2Y_t, t = 2, 3\}$, denoted by $\{\bar{\mu}_t(\vec{\mathbf{U}}), \bar{\mu}_{2t}, t = 2, 3\}$, need to satisfy

$$\mathbb{E}\left[\vec{\mathbf{X}}\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\}\right] = \mathbf{0}, \quad \mathbb{E}\left\{Y_2^2 - \bar{\mu}_{22}(\vec{\mathbf{U}})\right\} = 0,$$
$$\mathbb{E}\left[\mathbf{X}_2\{Y_3 - \bar{\mu}_3(\vec{\mathbf{U}})\}\right] = \mathbf{0}, \quad \mathbb{E}\left\{Y_2Y_3 - \bar{\mu}_{23}(\vec{\mathbf{U}})\right\} = 0,$$

where $\vec{\mathbf{X}} = (1, \mathbf{X}_1^\intercal, \mathbf{X}_2^\intercal)^\intercal$. We thus propose a refitting step that expands $\{m_t(\vec{\mathbf{U}}), m_{2t}(\vec{\mathbf{U}}), t = 2, 3\}$ to additionally adjust for linear effects of $\mathbf{X}_1$ and/or $\mathbf{X}_2$ to ensure the subsequent projection step is unbiased. To this end, let $\{\mathcal{I}_k, k = 1, ..., K\}$ denote K random equal sized partitions of the labeled index set $\{1, ..., n\}$, and let $\{\widehat{m}_t^{(-k)}(\vec{\mathbf{U}}), \widehat{m}_{2t}^{(-k)}(\vec{\mathbf{U}}), t = 2, 3\}$ be the counterpart of $\{\widehat{m}_t(\vec{\mathbf{U}}), \widehat{m}_{2t}(\vec{\mathbf{U}}), t = 2, 3\}$ with labeled observations in $\{1, .., n\} \setminus \mathcal{I}_k$. We then obtain $\widehat{\boldsymbol{\eta}}_2$, $\widehat{\eta}_{22}$, $\widehat{\boldsymbol{\eta}}_3$, $\widehat{\eta}_{23}$ respectively as the solutions to

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} \vec{\mathbf{X}}_i \left\{Y_{2i} - \widehat{m}_2^{(-k)}(\vec{\mathbf{U}}_i) - \boldsymbol{\eta}_2^\intercal \vec{\mathbf{X}}_i\right\} = \mathbf{0}, \quad \sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} \left\{Y_{2i}^2 - \widehat{m}_{22}^{(-k)}(\vec{\mathbf{U}}_i) - \eta_{22}\right\} = 0, \\
&\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} \mathbf{X}_{2i} \left\{Y_{3i} - \widehat{m}_3^{(-k)}(\vec{\mathbf{U}}_i) - \boldsymbol{\eta}_3^\intercal \mathbf{X}_{2i}\right\} = \mathbf{0}, \sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} \left\{Y_{2i}Y_{3i} - \widehat{m}_{23}^{(-k)}(\vec{\mathbf{U}}_i) - \eta_{23}\right\} = 0.
\end{aligned} \tag{3}$$

Finally, we impute Y_2, Y_3, Y_2^2 and Y_2Y_3 respectively as $\widehat{\mu}_2(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^{K} \widehat{m}_2^{(-k)}(\vec{\mathbf{U}}) + \widehat{\boldsymbol{\eta}}_2^\intercal \vec{\mathbf{X}}$, $\widehat{\mu}_3(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^{K} \widehat{m}_3^{(-k)}(\vec{\mathbf{U}}) + \widehat{\boldsymbol{\eta}}_3^\intercal \mathbf{X}_2$, $\widehat{\mu}_{22}(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^{K} \widehat{m}_{22}^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_{22}$, and $\widehat{\mu}_{23}(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^{K} \widehat{m}_{23}^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_{23}$.

Step III: Projection

In the last step, we proceed to estimate $\widehat{\boldsymbol{\theta}}$ by replacing $\{Y_t, Y_2Y_t, t = 2, 3\}$ in (2) with their the imputed values $\{\widehat{\mu}_t(\vec{\mathbf{U}}), \widehat{\mu}_{2t}(\vec{\mathbf{U}}), t = 2, 3\}$ and project to the unlabeled data. Specifically, we obtain the final SSL estimators for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ via the following steps:

1. Stage 2 regression: we obtain the SSL estimator for $\boldsymbol{\theta}_2$ as

$$\widehat{\boldsymbol{\theta}}_2 = (\widehat{\boldsymbol{\beta}}_2^\intercal, \widehat{\boldsymbol{\gamma}}_2^\intercal)^\intercal : \text{the solution to} \quad \mathbb{P}_N \begin{bmatrix} \widehat{\mu}_{23}(\vec{\mathbf{U}}) - [\widehat{\mu}_{22}(\vec{\mathbf{U}}), \widehat{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal]\boldsymbol{\theta}_2 \\ \mathbf{X}_2\{\widehat{\mu}_3(\vec{\mathbf{U}}) - [\widehat{\mu}_2(\vec{\mathbf{U}}), \mathbf{X}_2^\intercal]\boldsymbol{\theta}_2\} \end{bmatrix} = \mathbf{0}.$$

2. We compute the imputed pseudo-outcome:

$$\widetilde{Y}_2^* = \widehat{\mu}_2(\vec{\mathbf{U}}) + \max_{a\in\{0,1\}} Q_2\left(\mathbf{H}_2, \widehat{\mu}_2(\vec{\mathbf{U}}), a; \widehat{\boldsymbol{\theta}}_2\right).$$

3. Stage 1 regression: we estimate $\widehat{\boldsymbol{\theta}}_1 = (\widehat{\boldsymbol{\beta}}_1^\intercal, \widehat{\boldsymbol{\gamma}}_1^\intercal)^\intercal$ as the solution to:

$$\mathbb{P}_N\left\{\mathbf{X}_1(\widetilde{Y}_2^* - \mathbf{X}_1^\intercal\boldsymbol{\theta}_1)\right\} = \mathbf{0}.$$

Based on the SSL estimator for the Q-learning model parameters, we can then obtain an estimate for the optimal treatment protocol as:

$$\widehat{d}_t \equiv \widehat{d}_t(\mathbf{H}_t) \equiv d_t(\mathbf{H}_t; \widehat{\boldsymbol{\theta}}_t), \text{ where } d_t(\mathbf{H}_t, \boldsymbol{\theta}_t) = \underset{a\in\{0,1\}}{\operatorname{argmax}} Q_t(\mathbf{H}_t, a; \boldsymbol{\theta}_t) = I\left(\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t > 0\right), t = 1, 2.$$

Theorems 2 and 3 of Section 5 demonstrate the consistency and asymptotic normality of the SSL estimators $\{\widehat{\boldsymbol{\theta}}_t, t = 1, 2\}$ for their respective population parameters $\{\bar{\boldsymbol{\theta}}_t, t = 1, 2\}$, even in the case where (1) is mis-specified. As we explain next, this in turn yields desirable statistical results for evaluating the derived policy $\bar{d}_t \equiv \bar{d}_t(\mathbf{H}_t) \equiv d_t(\mathbf{H}_t, \bar{\boldsymbol{\theta}}_t) = \operatorname{argmax}_{a\in\{0,1\}} Q_t(\check{\mathbf{H}}_t, a; \bar{\boldsymbol{\theta}}_t)$ for $t = 1, 2$.

4. Semi-Supervised Off-Policy Policy Evaluation

To evaluate the performance of the optimal policy $\bar{\mathcal{D}} = \{\bar{d}_t(\mathbf{H}_t), t = 1, 2\}$, derived under the Q-learning framework, one may estimate the expected population outcome under the policy $\bar{\mathcal{D}}$:

$$\bar{V} \equiv \mathbb{E}\left[\mathbb{E}\{Y_2 + \mathbb{E}\{Y_3|\check{\mathbf{H}}_2, A_2 = \bar{d}_2(\mathbf{H}_2)\}|\mathbf{H}_1, A_1 = \bar{d}_1(\mathbf{H}_1)\}\right].$$

If models in (1) are correctly specified, then under standard causal assumptions (i)-(iii) (consistency, sequential ignorability, and positivity), an asymptotically consistent supervised estimator for the value function can be obtained as

$$\widehat{V}_Q = \mathbb{P}_n\left[Q_1^o(\check{\mathbf{H}}_1; \widehat{\boldsymbol{\theta}}_1)\right],$$

where $Q_t^o(\check{\mathbf{H}}_t; \boldsymbol{\theta}_t) \equiv Q_t\left(\check{\mathbf{H}}_t, d_t(\mathbf{H}_t; \boldsymbol{\theta}_t); \boldsymbol{\theta}_t\right)$. However, $\widehat{V}_Q$ is likely to be biased when the outcome models in (1) are mis-specified. This occurs frequently in practice since $Q_1(\check{\mathbf{H}}_1, A_1)$ is especially difficult to specify correctly.

To improve the robustness to model mis-specification, we augment $\widehat{V}_Q$ via propensity score weighting. This gives us an SSL doubly robust (SSL_{DR}) estimator for $\bar{V}$. To this end, we define propensity scores:

$$\pi_t(\check{\mathbf{H}}_t) = \mathbb{P}\{A_t = 1|\check{\mathbf{H}}_t\}, \quad t = 1, 2.$$

To estimate $\{\pi_t(\cdot), t = 1, 2\}$, we impose the following generalized linear models (GLMs):

$$\pi_t(\check{\mathbf{H}}_t; \boldsymbol{\xi}_t) = \sigma\left(\check{\mathbf{H}}_t^\intercal \boldsymbol{\xi}_t\right), \quad \text{with} \quad \sigma(x) \equiv 1/(1 + e^{-x}) \quad \text{for} \quad t = 1, 2. \tag{4}$$

We use the logistic model with potentially non-linear basis functions $\check{\mathbf{H}}$ for simplicity of presentation, but one may choose other GLMs or alternative basis expansions to incorporate non-linear effects in the propensity score models. One can estimate $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^\intercal, \boldsymbol{\xi}_2^\intercal)^\intercal$ based on the standard maximum likelihood estimators using labeled data, denoted by $\widehat{\boldsymbol{\xi}} = (\widehat{\boldsymbol{\xi}}_1^\intercal, \widehat{\boldsymbol{\xi}}_2^\intercal)^\intercal$. We denote the limit of $\widehat{\boldsymbol{\xi}}$ as $\bar{\boldsymbol{\xi}} = (\bar{\boldsymbol{\xi}}_1^\intercal, \bar{\boldsymbol{\xi}}_2^\intercal)^\intercal$. Note that this is not necessarily equal to the true model parameter under correct specification of (4), but corresponds to the population solution of the fitted models.

Our framework is flexible to allow an SSL approach to estimate the propensity scores. As these are nuisance parameters needed for estimation of the value function, and SSL for GLMs has been widely explored (See Chakrabortty, 2016, Ch. 2), we proceed with the usual GLM estimation to keep the discussion focused. However, SSL for propensity scores can be beneficial in certain cases, as we show in Proposition 9.

4.1 SUP_{DR} Value Function Estimation

To derive a supervised doubly robust (SUP_{DR}) estimator for $\bar{V}$ overcoming confounding in the observed data, we let $\boldsymbol{\Theta} = (\boldsymbol{\theta}^{\mathsf{T}}, \boldsymbol{\xi}^{\mathsf{T}})^{\mathsf{T}}$ and define the inverse probability weights (IPW) using the propensity scores as

$$\omega_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta}) \equiv \frac{d_1(\mathbf{H}_1;\boldsymbol{\theta}_1)A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{\{1-d_1(\mathbf{H}_1;\boldsymbol{\theta}_1)\}\{1-A_1\}}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}, \quad \text{and}$$
$$\omega_2(\check{\mathbf{H}}_2, A_2, \boldsymbol{\Theta}) \equiv \omega_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta}) \left(\frac{d_2(\mathbf{H}_2;\boldsymbol{\theta}_2)A_2}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} + \frac{\{1-d_2(\mathbf{H}_2;\boldsymbol{\theta}_2)\}\{1-A_2\}}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} \right).$$

Then we augment $Q_1^o(\mathbf{H}_1;\widehat{\boldsymbol{\theta}}_1)$ based on the estimated propensity scores via

$$\mathcal{V}_{\text{SUP}_{\text{DR}}}(\mathbf{L};\widehat{\boldsymbol{\Theta}}) = Q_1^o(\mathbf{H}_1;\widehat{\boldsymbol{\theta}}_1) + \omega_1(\check{\mathbf{H}}_1, A_1, \widehat{\boldsymbol{\Theta}}) \left[Y_2 - \left\{ Q_1^o(\mathbf{H}_1,\widehat{\boldsymbol{\theta}}_1) - Q_2^o(\check{\mathbf{H}}_2;\widehat{\boldsymbol{\theta}}_2) \right\} \right]$$
$$+ \omega_2(\check{\mathbf{H}}_2, A_2, \widehat{\boldsymbol{\Theta}}) \left\{ Y_3 - Q_2^o(\check{\mathbf{H}}_2;\widehat{\boldsymbol{\theta}}_2) \right\}$$

and estimate $\bar{V}$ as

$$\widehat{V}_{\text{SUP}_{\text{DR}}} = \mathbb{P}_n \left\{ \mathcal{V}_{\text{SUP}_{\text{DR}}}(\mathbf{L};\widehat{\boldsymbol{\Theta}}) \right\}. \tag{5}$$

Remark 1 *Importance-sampling value function estimators employ similar augmentation strategies. See, for example, the estimators shown in (10) in Jiang and Li (2016a) and (3) in Thomas and Brunskill (2016b). However, these consider a fixed policy, and we account for the fact that we estimate the DTR with the same data set. The construction of augmentation in $\widehat{V}_{\text{SUP}_{\text{DR}}}$ also differs from the usual augmented IPW estimators (Chakraborty and Moodie, 2013). As we are interested in the value had the population been treated with policy $\bar{\mathcal{D}}$ and not a fixed sequence (A_1, A_2), we augment the weights for a fixed treatment (i.e., $A_t = 1$ for $t = 1, 2$) with the propensity score weights for the estimated regime $I(A_t = \bar{d}_t)$, $t = 1, 2$. Finally, we note that this estimator can easily be extended to incorporate non-binary treatments.*

The supervised value function estimator $\widehat{V}_{\text{SUP}_{\text{DR}}}$ is doubly robust in the sense that if either the outcome, or the propensity score models are correctly specified, then $\widehat{V}_{\text{SUP}_{\text{DR}}} \xrightarrow{\mathbb{P}} \bar{V}$ in probability. Moreover, under certain reasonable assumptions, $\widehat{V}_{\text{SUP}_{\text{DR}}}$ is asymptotically normal. Theoretical guarantees and proofs for this procedure are shown in Appendix F.1.

4.2 SSL_{DR} Value Function Estimation

Analogous to semi-supervised Q-learning, we propose a procedure for adapting the augmented value function estimator to leverage $\mathcal{U}$, by imputing suitable functions of the unobserved outcome in (5). Note that since $\check{\mathbf{H}}_2$ involves Y_2, both $\omega_2(\check{\mathbf{H}}_2, A_2;\boldsymbol{\Theta})$ and $Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2) = Y_2\beta_{21} + Q_{2-}^o(\mathbf{H}_2;\boldsymbol{\theta}_2)$ are not available in the unlabeled set, where $Q_{2-}^o(\mathbf{H}_2;\boldsymbol{\theta}_2) = \mathbf{H}_{20}^{\mathsf{T}}\boldsymbol{\beta}_{22} + [\mathbf{H}_{21}^{\mathsf{T}}\boldsymbol{\gamma}_2]_+$. By writing $\mathcal{V}_{\text{SUP}_{\text{DR}}}(\mathbf{L};\widehat{\boldsymbol{\Theta}})$ as

$$\mathcal{V}_{\text{SUP}_{\text{DR}}}(\mathbf{L};\widehat{\boldsymbol{\Theta}}) = Q_1^o(\mathbf{H}_1;\widehat{\boldsymbol{\theta}}_1) + \omega_1(\check{\mathbf{H}}_1, A_1, \widehat{\boldsymbol{\Theta}}) \left\{ (1+\widehat{\beta}_{21})Y_2 - Q_1^o(\mathbf{H}_1,\widehat{\boldsymbol{\theta}}_1) + Q_{2-}^o(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2) \right\}$$
$$+ \omega_2(\check{\mathbf{H}}_2, A_2, \widehat{\boldsymbol{\Theta}}) \left\{ Y_3 - \widehat{\beta}_{21}Y_2 - Q_{2-}^o(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2) \right\},$$

we note that to impute $\mathcal{V}_{\mathsf{SUP_{DR}}}(\mathbf{L};\widehat{\bar{\boldsymbol{\Theta}}})$ for subjects in $\mathcal{U}$, we need to impute Y_2, $\omega_2(\check{\mathbf{H}}_2, A_2;\widehat{\bar{\boldsymbol{\Theta}}})$, and $Y_t\omega_2(\check{\mathbf{H}}_2, A_2;\widehat{\bar{\boldsymbol{\Theta}}})$ for $t=2,3$. We define the conditional mean functions

$$\mu_2^v(\vec{\mathbf{U}}) \equiv \mathbb{E}[Y_2|\vec{\mathbf{U}}], \quad \mu_{\omega_2}^v(\vec{\mathbf{U}}) \equiv \mathbb{E}[\omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}})|\vec{\mathbf{U}}], \quad \mu_{t\omega_2}^v(\vec{\mathbf{U}}) \equiv \mathbb{E}[Y_t\omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}})|\vec{\mathbf{U}}],$$

for $t=2,3$, where $\bar{\boldsymbol{\Theta}} = (\bar{\boldsymbol{\theta}}^\mathsf{T}, \bar{\boldsymbol{\xi}}^\mathsf{T})^\mathsf{T}$. As in Section 3.2 we approximate these expectations using a flexible imputation model followed by a refitting step for bias correction under possible mis-specification of the imputation models.

Step I: Imputation

We fit flexible weakly parametric or non-parametric models to the labeled data to approximate the functions $\{\mu_2^v(\vec{\mathbf{U}}), \mu_{\omega_2}^v(\vec{\mathbf{U}}), \mu_{t\omega_2}^v(\vec{\mathbf{U}}),\ t=2,3\}$ with unknown parameter $\boldsymbol{\Theta}$, estimated via the SSL Q-learning as in Section 3.2 and the propensity score modeling as discussed above. Denote the respective imputation models as $\{m_2(\vec{\mathbf{U}}), m_{\omega_2}(\vec{\mathbf{U}}), m_{t\omega_2}(\vec{\mathbf{U}}),\ t=2,3\}$ and their fitted values as $\{\widehat{m}_2(\vec{\mathbf{U}}), \widehat{m}_{\omega_2}(\vec{\mathbf{U}}), \widehat{m}_{t\omega_2}(\vec{\mathbf{U}}),\ t=2,3\}$.

Step II: Refitting

To correct for potential biases arising from finite sample estimation and model misspecifications, we perform refitting to obtain final imputed models for $\{Y_2, \omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}}), Y_t\omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}}),\ t=2,3\}$ as $\{\bar{\mu}_2^v(\vec{\mathbf{U}}) = m_2(\vec{\mathbf{U}}) + \eta_2^v, \bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}}) = m_{\omega_2}(\vec{\mathbf{U}}) + \eta_{\omega_2}^v, \bar{\mu}_{t\omega_2}^v(\vec{\mathbf{U}}) = m_{t\omega_2}(\vec{\mathbf{U}}) + \eta_{t\omega_2}^v, t=2,3\}$. As for the estimation of $\boldsymbol{\theta}$ for Q-learning case, these refitted models are not required to be correctly specified but need to satisfy the following constraints:

$$\begin{aligned}
\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1;\bar{\boldsymbol{\Theta}})\left\{Y_2 - \bar{\mu}_2^v(\vec{\mathbf{U}})\right\}\right] &= 0,\\
\mathbb{E}\left[Q_{2-}^o(\vec{\mathbf{U}};\boldsymbol{\theta}_2)\left\{\omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}}) - \bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}})\right\}\right] &= 0,\\
\mathbb{E}\left[\omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}})Y_t - \bar{\mu}_{t\omega_2}^v(\vec{\mathbf{U}})\right] &= 0,\ t=2,3.
\end{aligned}$$

To estimate η_2^v $\eta_{\omega_2}^v$, and $\eta_{t\omega_2}^v$ under these constraints, we again employ cross-fitting and obtain $\widehat{\eta}_2^v$ $\widehat{\eta}_{\omega_2}^v$, and $\widehat{\eta}_{t\omega_2}^v$ as the solution to the following estimating equations

$$\begin{aligned}
\sum_{k=1}^K \sum_{i\in\mathcal{I}_k} \omega_1(\check{\mathbf{H}}_{1i}, A_{1i};\widehat{\bar{\boldsymbol{\Theta}}})\left\{Y_2 - \widehat{m}_2^{(-k)}(\vec{\mathbf{U}}_i) - \widehat{\eta}_2^v\right\} &= 0,\\
\sum_{k=1}^K \sum_{i\in\mathcal{I}_k} Q_{2-}^o(\vec{\mathbf{U}}_i;\widehat{\boldsymbol{\theta}}_2)\left\{\omega_2(\check{\mathbf{H}}_{2i}, A_{2i};\widehat{\bar{\boldsymbol{\Theta}}}) - \widehat{m}_{\omega_2}^{(-k)}(\vec{\mathbf{U}}_i) - \widehat{\eta}_{\omega_2}^v\right\} &= 0, \qquad (6)\\
\sum_{k=1}^K \sum_{i\in\mathcal{I}_k} \left\{\omega_2(\check{\mathbf{H}}_{2i}, A_{2i};\widehat{\bar{\boldsymbol{\Theta}}})Y_{ti} - \widehat{m}_{t\omega_2}^{(-k)}(\vec{\mathbf{U}}_i) - \widehat{\eta}_{t\omega_2}^v\right\} &= 0,\ t=2,3.
\end{aligned}$$

The resulting imputation functions for $Y_2, \omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}})$ and $Y_t\omega_2(\check{\mathbf{H}}_2, A_2;\bar{\boldsymbol{\Theta}})$ are respectively constructed as $\widehat{\mu}_2^v(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^K \widehat{m}_2^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_2^v$, $\widehat{\mu}_{\omega_2}^v(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^K \widehat{m}_{\omega_2}(\vec{\mathbf{U}}) +$

$\widehat{\eta}^v_{\omega_2}$, and $\widehat{\mu}^v_{t\omega_2}(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^K \widehat{m}^{(-k)}_{t\omega_2}(\vec{\mathbf{U}}) + \widehat{\eta}^v_{t\omega_2}$, for $t = 2, 3$.

Step III: Projection, Semi-Supervised Augmented Value Function Estimator

Finally, we proceed to estimate the value of the policy $\bar{V}$, using the following semi-supervised augmented estimator:

$$\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}} = \mathbb{P}_N \left\{ \mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}; \widehat{\boldsymbol{\Theta}}, \widehat{\mu}) \right\}, \tag{7}$$

where $\widehat{\mathcal{V}}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}})$ is the semi-supervised augmented estimator for observation $\vec{\mathbf{U}}$ defined as:

$$\begin{aligned}\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}; \widehat{\boldsymbol{\Theta}}, \widehat{\mu}) =& Q^o_1(\check{\mathbf{H}}_1; \widehat{\boldsymbol{\theta}}_1) + \omega_1(\check{\mathbf{H}}_1, A_1, \widehat{\boldsymbol{\Theta}}) \left[(1 + \widehat{\beta}_{21})\widehat{\mu}^v_2(\vec{\mathbf{U}}) - Q^o_1(\check{\mathbf{H}}_1; \widehat{\boldsymbol{\theta}}_1) + Q^o_{2-}(\mathbf{H}_2; \widehat{\boldsymbol{\theta}}_2) \right] \\ &+ \widehat{\mu}_{3\omega_2}(\vec{\mathbf{U}}) - \widehat{\beta}_{21}\widehat{\mu}_{2\omega_2}(\vec{\mathbf{U}}) - Q^o_{2-}(\mathbf{H}_2; \widehat{\boldsymbol{\theta}}_2)\widehat{\mu}_{\omega_2}(\vec{\mathbf{U}}).\end{aligned}$$

The above SSL estimator uses both labeled and unlabeled data along with outcome surrogates to estimate the value function, which yields a gain in efficiency as we show in Proposition 9. As its supervised counterpart, $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$ is doubly robust in the sense that if either the Q functions or the propensity scores are correctly specified, the value function will converge in probability to the true value $\bar{V}$. Additionally, it does not assume that the estimated treatment regime was derived from a different sample. These properties are summarized in Theorem 7 and Proposition 8 of the following section.

5. Theoretical Results

In this section we discuss our assumptions and theoretical results for the semi-supervised Q-learning and value function estimators. Throughout, we define the norm $\|g(x)\|_{L_2(\mathbb{P})} \equiv \sqrt{\int g(x)^2 d\mathbb{P}(x)}$ for any real valued function $g(\cdot)$. Additionally, let $\{U_n\}$, and $\{V_n\}$ be two sequences of random variables. We will use $U_n = O_{\mathbb{P}}(V_n)$ to denote stochastic boundedness of the sequence $\{U_n/V_n\}$, that is, for any $\epsilon > 0$, $\exists M_\epsilon, n_\epsilon \in \mathbb{R}$ such that $\mathbb{P}\left(|U_n/V_n| > M_\epsilon\right) < \epsilon$ $\forall n > n_\epsilon$. We use $U_n = o_{\mathbb{P}}(V_n)$ to denote that $U_n/V_n \xrightarrow{\mathbb{P}} 0$.

5.1 Theoretical Results for SSL Q-Learning

Assumption 1 *(a) Sample size for $\mathcal{U}$, and $\mathcal{L}$, are such that $n/N \longrightarrow 0$ as $N, n \longrightarrow \infty$, (b) $\check{\mathbf{H}}_t \in \mathcal{H}_t$, $\check{\mathbf{X}}_t \in \mathcal{X}_t$ have finite second moments and compact support in $\mathcal{H}_t \subset \mathbb{R}^{q_t}$, $\mathcal{X}_t \subset \mathbb{R}^{p_t}$ $t = 1, 2$ respectively (c) $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ are nonsingular.*

Assumption 2 *Functions m_s, $s \in \{2, 3, 22, 23\}$ are such that (i) $\sup_{\vec{\mathbf{U}}} |m_s(\vec{\mathbf{U}})| < \infty$, and (ii) the estimated functions $\hat{m}_s$ satisfy (ii) $\sup_{\vec{\mathbf{U}}} |\widehat{m}_s(\vec{\mathbf{U}}) - m_s(\vec{\mathbf{U}})| = o_{\mathbb{P}}(1)$.*

Assumption 3 *Suppose Θ_1, Θ_2 are open bounded sets, and p_1, p_2 fixed under (1). We define the following class of functions:*

$$\mathcal{Q}_t \equiv \{Q_t : \mathcal{X}_1 \mapsto \mathbb{R} | \boldsymbol{\theta}_1 \in \Theta_1 \subset \mathbb{R}^{p_t}\}, \; t = 1, 2.$$

Further suppose for $t = 1, 2$, the solutions for $\mathbb{E}[S_t^\theta(\boldsymbol{\theta}_t)] = \mathbf{0}$, i.e., $\bar{\boldsymbol{\theta}}_1$ and $\bar{\boldsymbol{\theta}}_2$ satisfy

$$S_2^\theta(\boldsymbol{\theta}_2) = \frac{\partial}{\partial \boldsymbol{\theta}_2^\intercal}\|Y_3 - Q_2(\check{\mathbf{X}}_2; \boldsymbol{\theta}_2)\|_2^2,\ S_1^\theta(\boldsymbol{\theta}_1) = \frac{\partial}{\partial \boldsymbol{\theta}_1^\intercal}\|Y_2^* - Q_1(\check{\mathbf{X}}_1; \boldsymbol{\theta}_1)\|_2^2.$$

The target parameters satisfy $\bar{\boldsymbol{\theta}}_t \in \Theta_t\ , t = 1, 2$. We write $\bar{\boldsymbol{\beta}}_t, \bar{\boldsymbol{\gamma}}_t$ as the components of $\bar{\boldsymbol{\theta}}_t$, according to (2).

Assumption 1 (a) distinguishes our setting from the standard missing data context. Theoretical results for the missing completely at random (MCAR) setting generally assume that the missingness probability is bounded away from zero (Tsiatis, 2006), which enables the use of standard semiparametric theory. However, in our setting one can intuitively consider the probability of observing an outcome being $\frac{n}{n+N}$ which converges to 0.

Assumption 2 usually follows when imputation functions are bounded– which is natural to expect from the boundedness of the covariates. This is the case for many practical settings, including clinical history variables. We also require uniform convergence of the estimated functions to their limit. This allows for the normal equations targeting the imputation residuals in (6) and Appendix (B) (for the $T > 2$ case) to be well defined. Moreover, several off-the-shelf flexible imputation models for estimation can satisfy these conditions. See for example, local polynomial estimators, basis expansion regression like natural cubic splines or wavelets (Tsybakov, 2009). In particular, it is worth noting that we do not require any specific rate of convergence. As a result, the required condition is typically much easier to verify for many off-the-shelf algorithms. It is likely that other classes of models such as random forests can satisfy Assumption 2. Recent work suggests that it is plausible to use the existing point-wise convergence results to show uniform convergence. (see Scornet et al., 2015; Biau et al., 2008). Using some L_2 -type guarantee might be possible with extra care in the analysis, but that is out of the scope of the present paper.

Assumption 3 is fairly standard in the literature and ensures well-defined population level solutions for Q-learning regressions $\bar{\boldsymbol{\theta}}$ exist, and belong to that parameter space. In this regard, we differentiate between population solutions $\bar{\boldsymbol{\theta}}$ and true model parameters $\boldsymbol{\theta}^0$ shown in equation (1). If the working models are mis-specified, Theorems 2 and 3 still guarantee that $\widehat{\boldsymbol{\theta}}$ is consistent and asymptotically normal centered at the population solution $\bar{\boldsymbol{\theta}}$. However, when equation (1) is correct, $\widehat{\boldsymbol{\theta}}$ is asymptotically normal and consistent for the true parameter $\boldsymbol{\theta}^0$. Now we are ready to state the theoretical properties of the semi-supervised Q-learning procedure described in Section 3.2.

Theorem 2 (Distribution of $\widehat{\boldsymbol{\theta}}_2$) *Under Assumptions 1-3, $\widehat{\boldsymbol{\theta}}_2$ satisfies*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_2 - \bar{\boldsymbol{\theta}}_2) = \boldsymbol{\Sigma}_2^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}_2(\mathbf{L}_i; \bar{\boldsymbol{\theta}}_2) + o_{\mathbb{P}}(1) \xrightarrow{d} \mathcal{N}\Big(\mathbf{0}, \mathbf{V}_{2\mathsf{SSL}}(\bar{\boldsymbol{\theta}}_2)\Big),$$

where $\boldsymbol{\Sigma}_2 = \mathbb{E}[\check{\mathbf{X}}_2\check{\mathbf{X}}_2^\intercal]$ is defined in Section 2, the influence function $\boldsymbol{\psi}_2$ is given by

$$\boldsymbol{\psi}_2(\mathbf{L}; \bar{\boldsymbol{\theta}}_2) = \begin{bmatrix} \{Y_2Y_3 - \bar{\mu}_{23}(\vec{\mathbf{U}})\} - \bar{\beta}_{21}\{Y_2^2 - \bar{\mu}_{22}(\vec{\mathbf{U}})\} - Q_{2-}(\mathbf{H}_2, A_2; \bar{\boldsymbol{\theta}}_2)\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\} \\ \mathbf{X}_2\{Y_3 - \bar{\mu}_3(\vec{\mathbf{U}})\} - \bar{\beta}_{21}\mathbf{X}_2\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\} \end{bmatrix},$$

and $\mathbf{V}_{2\mathsf{SSL}}(\bar{\boldsymbol{\theta}}_2) = \boldsymbol{\Sigma}_2^{-1}\mathbb{E}\left[\boldsymbol{\psi}_2(\mathbf{L}; \bar{\boldsymbol{\theta}}_2)\boldsymbol{\psi}_2(\mathbf{L}; \bar{\boldsymbol{\theta}}_2)^\intercal\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^\intercal$.

We hold off remarks until the end of the results for the Q-learning parameters. Since the first stage regression depends on the second stage regression through a non-smooth maximum function, we make the following standard assumption (Laber et al., 2014) in order to provide valid statistical inference.

Assumption 4 *Non-zero estimable population treatment effects $\bar{\gamma}_t$, $t = 1, 2$: i.e., the population solution to (2), is such that (a) $\mathbf{H}_{21}^{\intercal}\bar{\gamma}_2 \neq 0$ for all $\mathbf{H}_{21} \neq \mathbf{0}$, and (b) $\bar{\gamma}_1$ is such that $\mathbf{H}_{11}^{\intercal}\bar{\gamma}_1 \neq 0$ for all $\mathbf{H}_{11} \neq \mathbf{0}$ almost everywhere.*

Assumption 4 yields regular estimators for the stage one regression and the value function, which depend on non-smooth components of the form $[x]_+$. This property is needed to achieve asymptotic normality of the Q-learning parameters for the first stage regression. Note that the estimating equation for the stage one regression in Section 3.2 includes $[\mathbf{H}_{21}^{\intercal}\widehat{\gamma}_2]_+$. Thus, for the asymptotic normality of $\widehat{\boldsymbol{\theta}}_1$, we require $\sqrt{n}\mathbb{P}_n\left([\mathbf{H}_{21}^{\intercal}\widehat{\gamma}_2]_+ - [\mathbf{H}_{21}^{\intercal}\bar{\gamma}_2]_+\right)$ to be asymptotically normal.

We also note that this requirement is automatically guaranteed as long as $\mathbf{H}_{11}$ contains at least one continuous covariate, as this implies $\mathbb{P}\left(\mathbf{H}_{21}^{\intercal}\bar{\gamma}_2 = 0\right) = 0$ (analogous for $\bar{\gamma}_1$). Several feature covariates in the DTR context are continuous, e.g., blood pressure, blood sugar level, blood cholesterol levels, body temperature, weight, etc. Violation of Assumption 4 will yield asymptotically biased and non-regular Q-learning estimates, which translate into poor coverage of the confidence intervals (see Laber et al. (2014) for a thorough discussion on this topic). This is why Assumption 4 is fairly standard in Q-learning, A-learning, etc. (Chakraborty and Moodie, 2013; Schulte et al., 2014; Laber et al., 2014; Robins, 2004; Tsiatis et al., 2019).

Theorem 3 (Distribution of $\hat{\boldsymbol{\theta}}_1$) *Under Assumptions 1-3, and 4 (a), $\widehat{\boldsymbol{\theta}}_1$ satisfies*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \bar{\boldsymbol{\theta}}_1) = \boldsymbol{\Sigma}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}_1(\mathbf{L}_i;\bar{\boldsymbol{\theta}}_1) + o_{\mathbb{P}}(1) \xrightarrow{d} \mathcal{N}\Big(\mathbf{0}, \mathbf{V}_{1\text{SSL}}(\bar{\boldsymbol{\theta}}_1)\Big)$$

where $\boldsymbol{\Sigma}_1^{-1} = \mathbb{E}[\check{\mathbf{X}}_1\check{\mathbf{X}}_1^{\intercal}]$, the influence function $\boldsymbol{\psi}_1$ is given by

$$\begin{aligned}\boldsymbol{\psi}_1(\mathbf{L};\bar{\boldsymbol{\theta}}_1) =& \mathbf{X}_1(1+\bar{\beta}_{21})\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\} + \mathbb{E}\left[\mathbf{X}_1\left(Y_2, \mathbf{H}_{20}^{\intercal}\right)\right]\boldsymbol{\psi}_{\beta_2}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)\\ &+\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^{\intercal}|\mathbf{H}_{21}^{\intercal}\bar{\gamma}_2 > 0\right]\mathbb{P}\left(\mathbf{H}_{21}^{\intercal}\bar{\gamma}_2 > 0\right)\boldsymbol{\psi}_{\gamma_2}(\mathbf{L};\bar{\boldsymbol{\theta}}_2),\end{aligned}$$

$\mathbf{V}_{1\text{SSL}}(\bar{\boldsymbol{\theta}}_1) = \boldsymbol{\Sigma}_1^{-1}\mathbb{E}\left[\boldsymbol{\psi}_1(\mathbf{L};\bar{\boldsymbol{\theta}}_1)\boldsymbol{\psi}_1(\mathbf{L};\bar{\boldsymbol{\theta}}_1)^{\intercal}\right]\left(\boldsymbol{\Sigma}_1^{-1}\right)^{\intercal}$, *and $\boldsymbol{\psi}_{\beta_2}$, $\boldsymbol{\psi}_{\gamma_2}$ are the elements corresponding to $\bar{\boldsymbol{\beta}}_2$, $\bar{\gamma}_2$ of the influence function $\boldsymbol{\psi}_2$ defined in Theorem 2.*

Remark 4 *1) Theorems 2 and 3 establish the $\sqrt{n}$-consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2$ for any $K \geq 2$. Beyond asymptotic normality at $\sqrt{n}$ scale, these theorems also provide an asymptotic linear expansion of the estimators with influence functions $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ respectively. For extension and discussion of the method for $T > 2$ stages see Appendix B.*

2) $\mathbf{V}_{1\text{SSL}}(\bar{\boldsymbol{\theta}})$, $\mathbf{V}_{2\text{SSL}}(\bar{\boldsymbol{\theta}})$ reflect an efficiency gain over the fully supervised approach due to sample $\mathcal{U}$ and the surrogates contribution to prediction performance. This gain is formalized in Proposition 5 which quantifies how correlation between surrogates and outcome increases

efficiency.
3) Let $\boldsymbol{\psi} = [\boldsymbol{\psi}_1^\intercal, \boldsymbol{\psi}_2^\intercal]^\intercal$*, we collect the vector of estimated Q-learning parameters* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\intercal, \boldsymbol{\theta}_2^\intercal)^\intercal$*, then under Assumptions 1-3, 4 (a), we have*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\psi}(\mathbf{L}_i; \bar{\boldsymbol{\theta}}) + o_{\mathbb{P}}(1) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{V}_{\mathsf{SSL}}\left(\bar{\boldsymbol{\theta}}\right)\right)$$

with $\mathbf{V}_{\mathsf{SSL}}(\bar{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}^{-1}\mathbb{E}\left[\boldsymbol{\psi}(\mathbf{L}; \bar{\boldsymbol{\theta}})\boldsymbol{\psi}(\mathbf{L}; \bar{\boldsymbol{\theta}})^\intercal\right]\left(\boldsymbol{\Sigma}^{-1}\right)^\intercal$.
4) Theorems 2 and 3 hold even when the Q functions are mis-specified, that is, $\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2$ *are consistent and asymptotically normal for* $\bar{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2$*. Furthermore, if model* (1) *is correctly specified then we can simply replace* $\bar{\boldsymbol{\theta}}$ *with* $\boldsymbol{\theta}^0$ *in the above result.*
5) We estimate $\mathbf{V}_{\mathsf{SSL}}(\bar{\boldsymbol{\theta}})$ *via sample-splitting as*

$$\begin{aligned}
\widehat{\mathbf{V}}_{\mathsf{SSL}}(\widehat{\boldsymbol{\theta}}) &= \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{A}}(\boldsymbol{\theta})\left(\widehat{\boldsymbol{\Sigma}}^{-1}\right)^\intercal, \text{ where} \\
\widehat{\mathbf{A}}(\widehat{\boldsymbol{\theta}}) &= n^{-1}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} \boldsymbol{\psi}^{(-k)}\left(\mathbf{L}_i; \widehat{\boldsymbol{\theta}}\right)\boldsymbol{\psi}^{(-k)}\left(\mathbf{L}_i; \widehat{\boldsymbol{\theta}}\right)^\intercal, \\
\widehat{\boldsymbol{\Sigma}}_t &= \mathbb{P}_n\left\{\mathbf{X}_t\mathbf{X}_t^\intercal\right\}, \ t = 1, 2.
\end{aligned}$$

Note that we can decompose $\boldsymbol{\psi}$ into the influence function for each set of parameters, for example, we have $\boldsymbol{\psi}_2 = \left(\boldsymbol{\psi}_{\beta_2}^\intercal, \boldsymbol{\psi}_{\gamma_2}^\intercal\right)^\intercal$ where

$$\boldsymbol{\psi}_{\gamma_2}(\mathbf{L}; \bar{\boldsymbol{\theta}}_2) = \mathbf{H}_{21}A_2\left[\{Y_3 - \bar{\mu}_3(\vec{\mathbf{U}})\} - \bar{\beta}_{21}\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\}\right].$$

Thus, we can decompose the variance-covariance matrix into a component for each parameter, the variance-covariance for the treatment effect for stage 2 regression $\boldsymbol{\gamma}_2$ is

$$\mathbb{E}\left[\boldsymbol{\psi}_{\gamma_2}(\mathbf{L}; \bar{\boldsymbol{\theta}}_2)\boldsymbol{\psi}_{\gamma_2}(\mathbf{L}; \bar{\boldsymbol{\theta}}_2)^\intercal\right] = \mathbb{E}\left[\mathbf{H}_{21}\mathbf{H}_{21}^\intercal A_2^2\left\{Y_3 - \bar{\mu}_3(\vec{\mathbf{U}}) - \beta_{21}\left(Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\right)\right\}^2\right].$$

This gives us some insight into how the predictive power of $\vec{\mathbf{U}}$, which contains surrogates $\mathbf{W}_1, \mathbf{W}_2$, decreases uncertainty of parameter estimates, yielding smaller standard errors. This is the case in general for the influence functions of estimators for $\bar{\boldsymbol{\theta}}_1$, $\bar{\boldsymbol{\theta}}_2$. We formalize this result with the following proposition. Let $\widehat{\boldsymbol{\theta}}_{\mathsf{SUP}}$ be the estimator for the fully supervised Q-learning procedure (i.e. only using labeled data), with influence function and asymptotic variance denoted as $\boldsymbol{\psi}_{\mathsf{SUP}}$ and $\mathbf{V}_{\mathsf{SUP}}$ respectively (see Appendix D.1 for the derivation and exact form of $\boldsymbol{\psi}_{\mathsf{SUP}}$ and $\mathbf{V}_{\mathsf{SUP}}$).

For the following proposition we need the imputation models $\bar{\mu}_s$, $s \in \{2, 3, 22, 23\}$ to satisfy additional constraints of the form $\mathbb{E}\left[\mathbf{X}_2\mathbf{X}_2^\intercal\{Y_2Y_3 - \bar{\mu}_{23}(\vec{\mathbf{U}})\}\right] = \mathbf{0}$. We list them in Assumption 7, Appendix D.1. One can construct estimators which satisfy such conditions by simply augmenting $\boldsymbol{\eta}_2$, η_{22}, $\boldsymbol{\eta}_3$, η_{23} in (3) with additional terms in the refitting step.

Proposition 5 *Under Assumptions 1-3, 4 (a), and 7 then*

$$\mathbf{V}_{\mathsf{SSL}}(\bar{\boldsymbol{\theta}}) = \mathbf{V}_{\mathsf{SUP}}(\bar{\boldsymbol{\theta}}) - \boldsymbol{\Sigma}^{-1}\, Var\left[\boldsymbol{\psi}_{\mathsf{SUP}}(\mathbf{L}; \bar{\boldsymbol{\theta}}) - \boldsymbol{\psi}_{\mathsf{SSL}}(\mathbf{L}; \bar{\boldsymbol{\theta}})\right]\left(\boldsymbol{\Sigma}^{-1}\right)^\intercal.$$

Remark 6 *Proposition 5 illustrates how the estimates for the semi-supervised Q-learning parameters are at least as efficient, if not more so, than the supervised ones. Intuitively, the difference in efficiency is explained by how much information is gained by incorporating the surrogates* $\mathbf{W}_1, \mathbf{W}_2$ *into the estimation procedure. If there is no new information in the surrogate variables, then residuals found in* $\boldsymbol{\psi}_{\mathsf{SSL}}(\mathbf{L};\boldsymbol{\theta})$ *will be of similar magnitude to those in* $\boldsymbol{\psi}_{\mathsf{SUP}}(\mathbf{L};\boldsymbol{\theta})$*, and thus the difference in efficiency will be small: Var* $\left[\boldsymbol{\psi}_{\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}) - \boldsymbol{\psi}_{\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\theta}})\right] \approx 0$*. In this case both methods will yield equally efficient parameters. The gain in precision is especially relevant for the treatment interaction coefficients* $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ *used to learn the dynamic treatment rules. Finally, note that for Proposition 5, we do not need the correct specification of Q functions or imputation models.*

5.2 Theoretical Results for SSL Estimation of the Value Function

If model (1) is correct, one only needs to add Assumption 4 (b) for $\mathbb{P}_N\{Q_1^o(\mathbf{H}_1;\widehat{\boldsymbol{\theta}}_1)\}$ to be a consistent estimator of the value function $\bar{V}$ (Zhu et al., 2019). However, as we discussed earlier, (1) is likely mis-specified. This motivates the use of our doubly robust semi-supervised value function estimator. We also show our estimator is asymptotically normal and achieves efficiency equal to or exceeding that of the corresponding supervised estimator. To that end, define the following class of functions:

$$\mathcal{W}_t \equiv \{\pi_t : \mathcal{H}_t \mapsto \mathbb{R} | \boldsymbol{\xi}_t \in \Omega_t\}, \; t = 1, 2,$$

under propensity score models π_1, π_2 in (4).

Assumption 5 *Let the score population equations* $\mathbb{E}\left[S_t^{\xi}(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)\right] = \mathbf{0}, t = 1, 2$ *have solutions* $\bar{\boldsymbol{\xi}}_1, \bar{\boldsymbol{\xi}}_2$*, where*

$$S_t^{\xi}(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t) = \frac{\partial}{\partial \boldsymbol{\xi}_t} \log\left[\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)^{A_t}\{1-\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)\}^{(1-A_t)}\right], \; t = 1, 2,$$

(i) Ω_1, Ω_2 *are open, bounded sets and the population solutions satisfy* $\bar{\boldsymbol{\xi}}_t \in \Omega_t, t = 1, 2$*,*
(ii) for $\bar{\boldsymbol{\xi}}_t, t = 1, 2,$ $\inf_{\check{\mathbf{H}}_t \in \mathcal{H}_1} \pi_1(\check{\mathbf{H}}_t;\bar{\boldsymbol{\xi}}_t) > 0$*,*
(iii) Finite second moment: $\mathbb{E}\left[S_t^{\xi}(\check{\mathbf{H}}_t;\boldsymbol{\Theta}_t)^2\right] \leq \infty$*, and Fisher information matrix* $\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\xi}_t} S_t^{\xi}(\check{\mathbf{H}}_t;\boldsymbol{\Theta}_t)\right]$ *exists and is non singular,*
(iv) Second-order partial derivatives of $S_t^{\xi}(\check{\mathbf{H}}_t;\boldsymbol{\Theta}_t)$ *with respect to* $\boldsymbol{\xi}$ *exist and for every* $\check{\mathbf{H}}_t$*, and satisfy* $|\partial^2 S_t^{\xi}(\check{\mathbf{H}}_t;\boldsymbol{\Theta}_t)/\partial \boldsymbol{\xi}_i \partial \boldsymbol{\xi}_j| \leq \tilde{S}_t(\check{\mathbf{H}}_t)$ *for some integrable measurable function* $\tilde{S}_t$ *in a neighborhood of* $\bar{\boldsymbol{\xi}}$*.*

Assumption 6 *Functions* $m_2, m_{\omega_2}, m_{t\omega_2}$ $t = 2, 3$ *are such that (i)* $\sup_{\vec{\mathbf{U}}} |m_s(\vec{\mathbf{U}})| < \infty$*, and (ii) the estimated functions* $\hat{m}_s$ *satisfy (ii)* $\sup_{\vec{\mathbf{U}}} |\widehat{m}_s(\vec{\mathbf{U}}) - m_s(\vec{\mathbf{U}})| = o_{\mathbb{P}}(1)$*,*
$s \in \{2, \omega_2, 2\omega_2, 3\omega_2\}$*.*

Assumption 5 is standard for Z-estimators (see Vaart, 1998, Ch. 5.6). Finally, we use $\boldsymbol{\psi}^{\xi}$ and $\boldsymbol{\psi}^{\theta}$ to denote the influence function for $\widehat{\boldsymbol{\xi}}$, and $\widehat{\boldsymbol{\theta}}$ respectively. We are now ready to state our theoretical results for the value function estimator in equation (7). The proof, and the exact form of $\boldsymbol{\psi}^{\xi}$ can be found in Appendix D.2.

Theorem 7 (Asymptotic Normality for $\widehat{V}_{\mathsf{SSL_{DR}}}$) *Under Assumptions 1-6, $\widehat{V}_{\mathsf{SSL_{DR}}}$ defined in (7) satisfies*

$$\sqrt{n}\left\{\widehat{V}_{\mathsf{SSL_{DR}}} - \mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\mathsf{SSL_{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}},\bar{\mu})\right]\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{v}_{\mathsf{SSL_{DR}}}(\mathbf{L}_i;\bar{\mathbf{\Theta}}) + o_{\mathbb{P}}(1),$$

where

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{v}_{\mathsf{SSL_{DR}}}(\mathbf{L}_i;\bar{\mathbf{\Theta}}) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2_{\mathsf{SSL_{DR}}}\right).$$

Here

$$\begin{aligned}
\psi^{v}_{\mathsf{SSL_{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}}) =& \nu_{\mathsf{SSL_{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}}) + \boldsymbol{\psi}^{\theta}(\mathbf{L})^{\mathsf{T}}\frac{\partial}{\partial\boldsymbol{\theta}}\int \mathcal{V}_{\mathsf{SUP_{DR}}}(\mathbf{L};\mathbf{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} \\
&+ \boldsymbol{\psi}^{\xi}(\mathbf{L})^{\mathsf{T}}\frac{\partial}{\partial\boldsymbol{\xi}}\int \mathcal{V}_{\mathsf{SUP_{DR}}}(\mathbf{L};\mathbf{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}}, \\
\nu_{\mathsf{SSL_{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}}) =& \omega_1(\check{\mathbf{H}}_1, A_1;\bar{\mathbf{\Theta}}_1)(1+\bar{\beta}_{21})\left\{Y_2 - \bar{\mu}^{v}_2(\vec{\mathbf{U}})\right\} + \omega_2(\check{\mathbf{H}}_2, A_2, \bar{\mathbf{\Theta}}_2)Y_3 - \bar{\mu}_{3\omega_2}(\vec{\mathbf{U}}) \\
&-\bar{\beta}_{21}\left\{\omega_2(\check{\mathbf{H}}_2, A_2, \bar{\mathbf{\Theta}}_2)Y_2 - \bar{\mu}_{2\omega_2}(\vec{\mathbf{U}})\right\} \\
&-Q^{o}_{2-}(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\left\{\omega_2(\check{\mathbf{H}}_2, A_2, \bar{\mathbf{\Theta}}_2) - \bar{\mu}_{\omega_2}(\vec{\mathbf{U}})\right\},
\end{aligned}$$

$\sigma^2_{\mathsf{SSL_{DR}}} = \mathbb{E}\left[\psi^{v}_{\mathsf{SSL_{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})^2\right]$, *and $\mathcal{V}_{\mathsf{SUP_{DR}}}(\mathbf{L};\mathbf{\Theta})$ is as defined in (5).*

Proposition 8 (Double robustness of $\widehat{V}_{\mathsf{SSL_{DR}}}$ as an estimator of $\bar{V}$) *(a) If either $\|Q_t(\check{\mathbf{H}}_t, A_t;\widehat{\boldsymbol{\theta}}_t) - Q_t(\check{\mathbf{H}}_t, A_t)\|_{L_2(\mathbb{P})} \to 0$, or $\|\pi_t(\check{\mathbf{H}}_t;\widehat{\boldsymbol{\xi}}_t) - \pi_t(\check{\mathbf{H}}_t)\|_{L_2(\mathbb{P})} \to 0$ for $t = 1, 2$, then under Assumptions 1-6, $\widehat{V}_{\mathsf{SSL_{DR}}}$ satisfies*

$$\widehat{V}_{\mathsf{SSL_{DR}}} \xrightarrow{\mathbb{P}} \bar{V}.$$

(b) If $\|Q_t(\check{\mathbf{H}}_t, A_t;\widehat{\boldsymbol{\theta}}_t) - Q_t(\check{\mathbf{H}}_t, A_t)\|_{L_2(\mathbb{P})}\|\pi_t(\check{\mathbf{H}}_t;\widehat{\boldsymbol{\xi}}_t) - \pi_t(\check{\mathbf{H}}_t)\|_{L_2(\mathbb{P})} = o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ for $t = 1, 2$, then under Assumptions 1-6, $\widehat{V}_{\mathsf{SSL_{DR}}}$ satisfies

$$\sqrt{n}\left(\widehat{V}_{\mathsf{SSL_{DR}}} - \bar{V}\right) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2_{\mathsf{SSL_{DR}}}\right).$$

Note that our positivity assumptions guarantee overlap of the propensity scores. In particular, the positivity assumption (iii) in Section 2 guarantees that $\pi_t(\mathbf{H}_t)$, $t = 1, 2$ are bounded away from zero almost everywhere; this translates into having distribution overlap between both treatments. As for the models, if they are correctly specified, then from assumption (iii), the logistic estimators $\widehat{\pi}$ can not be too small. If, in the other case, the propensity score models are mis-specified, Assumption 5(ii) guarantees that the target parameters of $\widehat{\pi}$'s are bounded away from zero, which translates into the limits of $\widehat{\pi}$'s being bounded as well.

Next we define the supervised influence function of estimator $\widehat{V}_{\mathsf{SUP_{DR}}}$ (see Theorem 19 and corresponding proof in Appendix F.1). Let $\boldsymbol{\psi}^{\theta}_{\mathsf{SUP}}$, be the influence function of the supervised

estimator $\widehat{\boldsymbol{\theta}}_{\mathsf{SUP}}$ for model (1). The influence function, and variance of the Value Function $\widehat{V}_{\mathsf{DR}}$ defined in (5) are

$$\begin{aligned}\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) =&\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) - \mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})\right]\\ &+\boldsymbol{\psi}^{\theta}_{\mathsf{SUP}}(\mathbf{L})^{\mathsf{T}}\frac{\partial}{\partial\boldsymbol{\theta}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}} + \boldsymbol{\psi}^{\xi}(\mathbf{L})^{\mathsf{T}}\frac{\partial}{\partial\boldsymbol{\xi}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}},\\ \sigma^2_{\mathsf{SUP}_{\mathsf{DR}}} =&\mathbb{E}\left[\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})^2\right].\end{aligned}$$

The flexibility of our SSL value function estimator $V_{\mathsf{SSL}_{\mathsf{DR}}}$ allows the use of either supervised or SSL approach for estimation of propensity score nuisance parameters $\boldsymbol{\xi}$. For SSL estimation, we can use an approach similar to Section 3.2, (see Chakrabortty et al., 2018, Ch. 2 for details). This allows us to quantify the efficiency gain of $V_{\mathsf{SSL}_{\mathsf{DR}}}$ vs. $V_{\mathsf{SUP}_{\mathsf{DR}}}$ by comparing the asymptotic variances. In light of this, we assume SSL is used for $\boldsymbol{\xi}$ when estimating $V_{\mathsf{SSL}_{\mathsf{DR}}}$.

Before stating the result we discuss an additional requirement for the imputation models. As for Proposition 5, models $\bar{\mu}^v_2(\vec{\mathbf{U}})$, $\bar{\mu}^v_{\omega_2}(\vec{\mathbf{U}})$, $\bar{\mu}^v_{t\omega_2}(\vec{\mathbf{U}})$, $t=2,3$ need to satisfy a few additional constraints of the form

$$\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\boldsymbol{\Theta}}_1)Q^o_{2-}(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_1)\{Y_2-\bar{\mu}^v_2(\vec{\mathbf{U}})\}\right]=\mathbf{0}.$$

As there are several constraints, we list them in Appendix D.2, and condense them in Assumption 8, Appendix D.2. Again, one can construct estimators which satisfy such conditions by simply augmenting η^v_2, $\eta^v_{\omega_2}$, $\eta^v_{t\omega_2}$, $t=2,3$ in (6) with additional terms in the refitting step.

Proposition 9 *Under Assumptions 1-6, and 8, asymptotic variances $\sigma^2_{\mathsf{SSL}_{\mathsf{DR}}}$, $\sigma^2_{\mathsf{SUP}_{\mathsf{DR}}}$ satisfy*

$$\sigma^2_{\mathsf{SSL}_{\mathsf{DR}}} = \sigma^2_{\mathsf{SUP}_{\mathsf{DR}}} - Var\left[\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) - \psi^v_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})\right].$$

Remark 10 *1) Proposition 8 illustrates how $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$ is asymptotically unbiased if either the Q functions or the propensity scores are correctly specified.*
2) An immediate consequence of Proposition 9 is that the semi-supervised estimator is at least as efficient (or more) than its supervised counterpart, that is $Var\left[\psi_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})\right] \leq Var\left[\psi_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})\right]$. As with Proposition 5, the difference in efficiency is explained by the information gain from incorporating surrogates.
3) To estimate standard errors for $V_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\boldsymbol{\Theta}})$, we will approximate the derivatives of the expectation terms $\frac{\partial}{\partial\bar{\boldsymbol{\Theta}}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})d\mathbb{P}_{\mathbf{L}}$ using kernel smoothing to replace the indicator functions. In particular, let $\mathbb{K}_h(x)=\frac{1}{h}\sigma(x/h)$, σ defined as in (4), we approximate $d_t(\mathbf{H}_t,\boldsymbol{\theta}_2)=I(\mathbf{H}^{\mathsf{T}}_{t1}\boldsymbol{\gamma}_t>0)$ with $\mathbb{K}_h(\mathbf{H}^{\mathsf{T}}_{t1}\boldsymbol{\gamma}_t)$ $t=1,2$, and define the smoothed propensity score weights as

$$\begin{aligned}\tilde{\omega}_1(\check{\mathbf{H}}_1,A_1,\boldsymbol{\Theta}) &\equiv \frac{A_1\mathbb{K}_h(\mathbf{H}^{\mathsf{T}}_{11}\boldsymbol{\gamma}_1)}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{\{1-A_1\}\{1-\mathbb{K}_h(\mathbf{H}^{\mathsf{T}}_{11}\boldsymbol{\gamma}_1)\}}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}, \quad \textit{and}\\ \tilde{\omega}_2(\check{\mathbf{H}}_2,A_2,\boldsymbol{\Theta}) &\equiv \tilde{\omega}_1(\check{\mathbf{H}}_1,A_1,\boldsymbol{\Theta})\left[\frac{A_2\mathbb{K}_h(\mathbf{H}^{\mathsf{T}}_{21}\boldsymbol{\gamma}_2)}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} + \frac{\{1-A_2\}\{1-\mathbb{K}_h(\mathbf{H}^{\mathsf{T}}_{21}\boldsymbol{\gamma}_2)\}}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right].\end{aligned}$$

We simply replace the propensity score functions with these smooth versions in $\psi^{v}_{\mathsf{SSL_{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})$*, detail is given in Appendix D.2.1. To estimate the variance we use a sample-split estimator:*

$$\hat{\sigma}^2_{\mathsf{SSL_{DR}}} = n^{-1}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\psi^{v(-k)}_{\mathsf{SSL_{DR}}}(\vec{\mathbf{U}}_i;\widehat{\boldsymbol{\Theta}})^2.$$

5.3 Related Literature

There is a significant amount of work that uses surrogate variables for estimating treatment effects. Applications range from economics and education to healthcare. Prentice (1989); Begg and Leung (2000) propose to use surrogate endpoints for randomized trials. These surrogates are proxies of the missing outcomes and usually require the outcomes to be independent of treatment when conditioned on the surrogate. While helpful in establishing a theoretical framework, this assumption is usually hard to validate in practice. For example, unmeasured confounding between a long-term outcome and the chosen surrogate would invalidate this assumption. Athey et al. (2019) provide a solution in this problem space by combining several short-term outcomes into a surrogate index, their work provides an introduction to using surrogates for treatment effect estimation and further development of the surrogate index. This surrogate index assimilates to our approach because it is a good proxy of the outcome. However, the short-term outcomes used in the index are not only proxies but outcomes relevant as direct effects of the intervention. In our case, the surrogates are only correlated with the outcome of interest through the joint probability law $\mathbb{P}_{\mathbf{Y},\mathbf{O},A,\mathbf{W}}$ but are not relevant to the conditional law of interest $\mathbb{P}_{\mathbf{Y}|\mathbf{O},\mathbf{A}}$ used to find the optimal policy.

In line with our approach, Pepe (1992) proposes using labeled and unlabeled data to estimate regression parameters, a valuable framework for average treatment effect estimation. More recently, Kallus and Mao (2020) proposed SSL methods for estimating an average causal treatment effect and using semi-parametric efficiency to show efficiency bounds on the ATE. However, both methods focus on a single time step. A key difference is that we are interested in the conditional treatment effect, which is needed to learn an optimal policy function. Also, the approach of Kallus and Mao (2020) assumes a missing at random setting, whereas we have a missing completely at random setting by design, as we first sample health records randomly and then label them. Alternatively, Chapter 4 of Van der Laan and Rose (2018) gives an overview of DTR using targeted learning theory. Both frameworks differ from our approach as we require that the probability of missingness goes to one as both labeled and unlabeled samples increase in size.

Davidian et al. (2005) use a semi-parametric approach to propose a consistent estimator for the average treatment effect under missing follow-up data. Chakrabortty et al. (2022) provide a survey of semi-supervised causal inference for estimating average and quantile treatment effects. To the best of our knowledge, this is the first work to propose estimating the time-dependent conditional treatment effect for policy learning in the missing outcome space. As previously discussed, Finn et al. (2016) proposed a semi-supervised RL method that achieves good empirical results and outperforms simple approaches such as direct imputation of the reward. However, the method does not leverage surrogates; there are no theoretical guarantees, and the approach lacks causal validity.

As to doubly robust estimation of DTRs, Dudík et al. (2011); Thomas and Brunskill (2016a); Kallus and Uehara (2020b) consider doubly robust policy evaluation; see also Murphy et al. (2001). A key difference between these approaches and our work is that the policy they evaluate is a pre-existing fixed function.

Another line of research focuses on doubly robust policy learning. They optimize a doubly robust value function estimator over a rich class of functions to estimate the optimal policy. To the best of our knowledge, existing work in this direction either takes $T = 1$ (Athey and Wager, 2021; Zhou et al., 2022a; Zhao et al., 2019; Bennett and Kallus, 2020), or does not provide theoretical guarantees (Zhang et al., 2013; Sonabend-W et al., 2020a). The associated optimization is generally complicated unless the search space for the policies is sufficiently small (Zhou et al., 2022b). Their estimate is doubly robust in the sense that if either the Q function or the propensity scores are correct, their value estimate is consistent for the optimal value function. See also Tsiatis et al. (2019) Ch. 2 for a discussion of doubly robust estimation of causal effects. Our work differs from these types of approaches because we use the Q functions to estimate the optimal policy and then estimate the value function of *such* policy using a doubly robust estimator. Therefore, our target policy is not the underlying optimal policy but the best one attainable by linear approximations of our Q functions. This target policy only matches the optimal treatment policy if the used models are correctly specified. This characteristic is expected because ours is a purely model-based approach not depending on complicated optimization, likely necessary for doubly robust estimation of the true value function.

6. Simulations and Application to Electronic Health Record Data:

We perform extensive simulations to evaluate the finite sample performance of our methods for $T = 2, 3, 5, 7$. Additionally we apply our methods to an EHR study of treatment response for patients with inflammatory bowel disease to identify the optimal treatment sequence for each patient. These data have treatment response outcomes available for a small subset of patients only.

6.1 Simulation Results

We compare our SSL Q-learning methods to fully supervised Q-learning using labeled data sets of different sizes and settings. We focus on the efficiency gains of our approach. First we discuss our simulation settings, then go on to show results for the Q function parameters under correct and incorrect working models for (1). We then show value function summary statistics under correct models, mis-specification for the Q models in (1), and the propensity score function π_2 in (4). Finally we show the correct and mis-specified results for a general $T > 2$ time horizon setting (see Appendix B for the extension of the methods to a general time horizon).

Following a similar set-up as in Schulte et al. (2014), we first consider a simple scenario with a single confounder variable at each stage with $\mathbf{H}_{10} = \mathbf{H}_{11} = (1, O_1)^\intercal$, $\check{\mathbf{H}}_{20} = (Y_2, 1, O_1, A_1, O_1A_1, O_2)^\intercal$, and $\mathbf{H}_{21} = (1, A_1, O_2)^\intercal$. Specifically, recalling that $\sigma(x) \equiv$

$1/(1+e^{-1})$, we sequentially generate

$$\begin{aligned} &O_1 \sim \mathsf{Bern}(0.5), \quad && A_1 \sim \mathsf{Bern}(\sigma\left\{\mathbf{H}_{10}^\intercal \boldsymbol{\xi}_1^0\right\}), && Y_2 \sim \mathcal{N}(\check{\mathbf{X}}_1^\intercal \boldsymbol{\theta}_1^0, 1), \\ &O_2 \sim \mathcal{N}(\check{\mathbf{H}}_{20}^\intercal \boldsymbol{\delta}^0, 2), && A_2 \sim \mathsf{Bern}\left(\sigma\left\{\mathbf{H}_{20}^\intercal \boldsymbol{\xi}_2^0 + \xi_{26}^0 O_2^2\right\}\right), \quad \text{and} && Y_3 \sim \mathcal{N}(m_3\left\{\check{\mathbf{H}}_{20}\right\}, 2), \end{aligned}$$

where $m_3\{\check{\mathbf{H}}_{20}\} = \mathbf{H}_{20}^\intercal \boldsymbol{\beta}_2^0 + A_2(\mathbf{H}_{21}^\intercal \boldsymbol{\gamma}_2^0) + \beta_{27}^0 O_2^2 Y_2 \sin\{[O_2^2(Y_2+1)]^{-1}\}$. Surrogates are generated as $W_t = \lfloor Y_{t+1} + Z_t \rfloor$, $Z_t \sim \mathcal{N}(0, \sigma_{z,t}^2)$, $t = 1, 2$ where $\lfloor x \rfloor$ corresponds to the integer part of $x \in \mathbb{R}$. Throughout, we let $\boldsymbol{\xi}_1^0 = (0.3, -0.5)^\intercal$, $\boldsymbol{\beta}_1^0 = (1,1)^\intercal$, $\boldsymbol{\gamma}_1^0 = (1,-2)^\intercal$ $\boldsymbol{\delta}^0 = (0, 0.5, -0.75, 0.25)^\intercal$, $\boldsymbol{\xi}_2^0 = (0, 0.5, 0.1, -1, -0.1)^\intercal$ $\boldsymbol{\beta}_2^0 = (.1, 3, 0, 0.1, -0.5, -0.5)^\intercal$, $\boldsymbol{\gamma}_2^0 = (1, 0.25, 0.5)^\intercal$.

We consider an additional case to mimic the structure of the EHR data set used for the real-data application. Outcomes Y_t are binary, we use a higher number of covariates for the Q functions and multivariate count surrogates $\mathbf{W}_t$ $t = 1, 2$. Data is simulated with $\mathbf{H}_{10} = (1, O_1, \ldots, O_6)^\intercal$, $\mathbf{H}_{11} = (1, O_2, \ldots, O_6)^\intercal$, $\check{\mathbf{H}}_{20} = (Y_2, 1, O_1, \ldots, O_6, A_1, Z_{21}, Z_{22})^\intercal$, and $\mathbf{H}_{21} = (1, O_1, \ldots, O_4, A_1, Z_{21}, Z_{22})^\intercal$, generated according to

$$\begin{aligned} &\mathbf{O}_1 \sim \mathcal{N}(\mathbf{0}, I_6), && A_1 \sim \mathsf{Bern}(\sigma\{\mathbf{H}_{10}^\intercal \boldsymbol{\xi}_1^0\}), && Y_2 \sim \mathsf{Bern}(\sigma\{\check{\mathbf{X}}_1^\intercal \boldsymbol{\theta}_1^0\}), \\ &\mathbf{O}_2 = \left[I\left\{Z_1 > 0\right\}, I\left\{Z_2 > 0\right\}\right]^\intercal && A_2 \sim \mathsf{Bern}\left(\tilde{m}_2\{\check{\mathbf{H}}_{20}\}\right), \quad \text{and} && Y_3 \sim \mathsf{Bern}(\tilde{m}_3\left\{\check{\mathbf{H}}_{20}\right\}), \end{aligned}$$

with $\tilde{m}_2 = \sigma\left\{\mathbf{H}_{20}^\intercal \boldsymbol{\xi}_2^0 + \tilde{\boldsymbol{\xi}}_2^\intercal \mathbf{O}_2\right\}$, $\tilde{m}_3(\check{\mathbf{H}}_{20}) = \mathbf{H}_{20}^\intercal \boldsymbol{\beta}_2^0 + A_2(\mathbf{H}_{21}^\intercal \boldsymbol{\gamma}_2^0) + \tilde{\boldsymbol{\beta}}_2^\intercal \mathbf{O}_2 Y_2 \sin\{\|\mathbf{O}_2\|_2^2/(Y_2+1)\}$ and $Z_l = O_{1l}\delta_l^0 + \epsilon_z$, $\epsilon_z \sim \mathcal{N}(0,1)$ $l = 1, 2$. The dimensions for the Q functions are 13 and 37 for the first and second stage respectively, which match with our IBD data set discussed in Section 6.2. The surrogates are generated according to $\mathbf{W}_t = \lfloor \mathbf{Z}_t \rfloor$, with $\mathbf{Z}_t \sim \mathcal{N}\left(\boldsymbol{\alpha}^\intercal(1, \mathbf{O}_t, A_t, Y_t), I\right)$. Parameters are set to $\boldsymbol{\xi}_1^0 = (-0.1, 1, -1, 0.1)^\intercal$, $\boldsymbol{\beta}_1^0 = (0.5, 0.2, -1, -1, 0.1, -0.1, 0.1)^\intercal$, $\boldsymbol{\gamma}_1^0 = (1, -2, -2, -0.1, 0.1, -1.5)^\intercal$,
$\boldsymbol{\xi}_2^0 = (0, 0.5, 0.1, -1, 1, -0.1)^\intercal$, $\boldsymbol{\beta}_2^0 = (1, \boldsymbol{\beta}_1^0, 0.25, -1, -0.5)^\intercal$,
$\boldsymbol{\gamma}_2^0 = (1, 0.1, -0.1, 0.1, -0.1, 0.25, -1, -0.5)^\intercal$, and $\boldsymbol{\alpha} = (1, \mathbf{0}, 1)^\intercal$.

Finally, we demonstrate how the method generalizes to $T > 2$ stages. We extend our continuous simulation set-up into a data generation process which depends recursively on previous time steps at any given stage. The first stage has a single covariate: $\mathbf{H}_{10} = \mathbf{H}_{11} = (1, O_1)^\intercal$, then for $t = 2, \ldots, T$ we have $\check{\mathbf{H}}_{t0} = (Y_2, \ldots, Y_t, 1, O_{t-1}, A_{t-1}, O_{t-1}A_{t-1}, O_t)^\intercal$, and $\mathbf{H}_{t1} = (1, A_{t-1}, O_t)^\intercal$. We generate the data for the first stage as

$$O_1 \sim \mathsf{Bern}(0.5), \quad A_1 \sim \mathsf{Bern}(\sigma\left\{\mathbf{H}_{10}^\intercal \boldsymbol{\xi}_1^0\right\}), \quad Y_2 \sim \mathcal{N}(\check{\mathbf{X}}_1^\intercal \boldsymbol{\theta}_1^0, 1),$$

we then proceed sequentially for $t = 2, \ldots, T+1$ generating data according to the following models:

$$\begin{aligned} &O_t \sim \mathcal{N}([1, O_{t-1}, A_{t-1}, O_{t-1}A_{t-1}]^\intercal \boldsymbol{\delta}_t^0, 2), \quad Y_{t+1} \sim \mathcal{N}(m_t\left\{\check{\mathbf{H}}_{t0}\right\}, 2), \quad \text{and} \\ &A_t \sim \mathsf{Bern}\left(\sigma\left\{[1, O_{t-1}, A_{t-1}, O_{t-1}A_{t-1}, O_t, Y_{t-1}]^\intercal \boldsymbol{\xi}_t^0 + \xi_{t*}^0 \sin O_t^2\right\}\right). \end{aligned}$$

where $m_t\{\check{\mathbf{H}}_{t0}\} = \mathbf{H}_{t0}^\intercal \boldsymbol{\beta}_t^0 + A_t(\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t^0) + \beta_{t*}^0 O_t^2 Y_t \sin\{[O_t^2(Y_t+1)]^{-1}\}$.
Throughout, we let $\boldsymbol{\xi}_1^0 = (0.3, -0.5)^\intercal$, $\boldsymbol{\beta}_1^0 = (0.1, 1)^\intercal$, $\boldsymbol{\gamma}_1^0 = (1, -0.2)^\intercal$,

$\boldsymbol{\delta}_t^0 = (0, 0.5, -0.75, 0.25, 1)^\intercal$, $\boldsymbol{\xi}_t^0 = (0, 0.5, 0.1, -1, -0.1, 1)^\intercal$, $\boldsymbol{\beta}_t^0 = (0.1, 0, 0.1, -0.5, -0.5, 0.1)^\intercal$, $\boldsymbol{\gamma}_t^0 = (1, 0.25, 0.5)^\intercal$. Surrogates are generated as $W_t = \lfloor Y_{t+1} + Z_t \rfloor$, $Z_t \sim \mathcal{N}(0, \sigma^2_{z,t})$, where $\lfloor x \rfloor$ corresponds to the integer part of $x \in \mathbb{R}$.

For two-stage settings, we fit models $Q_1(\mathbf{H}_1, A_1) = \mathbf{H}_{10}^\intercal \boldsymbol{\beta}_1^0 + A_1(\mathbf{H}_{11}^\intercal \boldsymbol{\gamma}_1^0)$, $Q_2(\check{\mathbf{H}}_2, A_2) = \check{\mathbf{H}}_{20}^\intercal \boldsymbol{\beta}_2^0 + A_2(\mathbf{H}_{21}^\intercal \boldsymbol{\gamma}_2^0)$ for the Q functions, $\pi_1(\mathbf{H}_1) = \sigma\left(\mathbf{H}_{10}^\intercal \boldsymbol{\xi}_1\right)$ and $\pi_2(\check{\mathbf{H}}_2) = \sigma\left(\check{\mathbf{H}}_{20}^\intercal \boldsymbol{\xi}_2\right)$ for the propensity scores. We fit analogous models for $T > 2$-stage settings: $Q_t(\check{\mathbf{H}}_t, A_t) = \check{\mathbf{H}}_{t0}^\intercal \boldsymbol{\beta}_t^0 + A_t(\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t^0)$, and $\pi_t(\check{\mathbf{H}}_t) = \sigma\left(\check{\mathbf{H}}_{t0}^\intercal \boldsymbol{\xi}_t\right)$ for $t = 1, \ldots, T$. To index mis-specification in the fitted Q-learning and the propensity score models, we use parameters ξ_{26}^0 and β_{27}^0, $\tilde{\boldsymbol{\xi}}_2, \tilde{\boldsymbol{\beta}}_2$, and ξ_{t*}^0, β_{t*}^0 for $T = 2$, and $T > 2$ stage settings respectively. A value of 0 for these parameters corresponds to correct specification of their respective models. For mis-specification, we set $\xi_{26}^0 = 1$, $\tilde{\boldsymbol{\xi}}_2 = \frac{1}{\|(1,\ldots,1)\|_2}(1, \ldots, 1)^\intercal$, and $\beta_{27}^0 = 1$, $\tilde{\boldsymbol{\beta}}_2 = \frac{1}{\|(1,\ldots,1)\|_2}(1, \ldots, 1)^\intercal$ for the propensity score π_2 and Q_1, Q_2 functions respectively. Similarly $\xi_{t*}^0 = 1$ and $\beta_{t*}^0 = 1$ for the general time horizon settings imply mis-specification of π_t, and Q_t respectively for $t = 1 \ldots, T$. Under mis-specification of the outcome model or propensity score model, the term omitted by the working models is highly non-linear, in which case the imputation model will be mis-specified as well. We show how our method does not need correct specification of the imputation model.

For the imputation models, we considered both random forest (RF) with 500 trees and basis expansion (BE) with piecewise-cubic splines with 2 equally spaced knots on the quantiles 33 and 67 (Hastie, 1992). We use 5-folds for our re-fitting bias correction step. For the two stage settings, we consider two choices of (n, N): $(135, 1272)$ which are similar to the sizes of our EHR study and larger sizes of $(500, 10000)$. For $T > 2$ settings, we use $(n, N) = (1000, 15000)$, and $T = 3, 5$ stages, and $(n, N) = (1500, 15000)$, and $T = 7$ stages. We report statistics of the interaction-effect coefficients from the Q functions, in particular we show mean absolute bias $\frac{1}{3}\sum_{j=1}^{2} |\boldsymbol{\gamma}_{tj}|$, and empirical standard error. For each configuration, we summarize results based on 1000 replications.

We start discussing results under correct specification of the Q functions. In Table 1, we present the results for the estimation of treatment interaction coefficients $\bar{\boldsymbol{\gamma}}_1, \bar{\boldsymbol{\gamma}}_2$, under the correct model specification, continuous outcome setting with $\beta_{27}^0 = \xi_{26}^0 = 0$. The complete tables for all $\bar{\boldsymbol{\theta}}$ parameters for the continuous and EHR-like settings can be found in Appendix C. We report bias, empirical standard error (ESE), average standard error (ASE), 95% coverage probability (CovP) and relative efficiency (RE) defined as the ratio of the supervised over the SSL estimate ESE.

Overall, compared to the supervised approach, the proposed semi-supervised Q-learning approach has substantial gains in efficiency while maintaining comparable or even lower bias. This is likely due to the refitting step which helps take care of the finite sample bias, both from the missing outcome imputation and Q function parameter estimation. Imputation with BE yields slightly better estimates than when using RF, both in terms of efficiency and bias. Coverage probabilities are close to the nominal level thanks to the strong performance of our sample-split standard error estimator shown in Section 5.2.

We next turn to Q-learning parameters under mis-specification of (1). Figure 1 shows the bias and root mean square error (RMSE) for the treatment interaction coefficients in the 2-stage Q functions. We focus on the continuous setting, where we set $\beta_{27}^0 \in \{-1, 0, 1\}$. Recall that $\beta_{27}^0 \neq 0$ implies that both Q functions are mis-specified as the fitting of Q_1 depends on formulation of Q_2 as seen in (2). Semi-supervised Q-learning is more efficient

(a) $n = 135$ and $N = 1272$

	Supervised		Semi-Supervised									
			Random Forests					Basis Expansion				
Parameter	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
γ_{11}=1.4	-0.03	0.41	0.00	0.26	0.24	0.93	1.57	0.00	0.24	0.23	0.93	1.68
γ_{12}=-2.6	0.04	0.58	-0.01	0.36	0.34	0.94	1.61	-0.02	0.35	0.31	0.90	1.69
γ_{21}=0.8	0.00	0.34	0.01	0.21	0.20	0.93	1.61	0.00	0.20	0.19	0.94	1.71
γ_{22}=0.2	-0.02	0.45	-0.01	0.28	0.28	0.95	1.60	-0.01	0.27	0.26	0.94	1.70
γ_{23}=0.5	0	0.18	0.01	0.11	0.11	0.94	1.59	0.00	0.11	0.11	0.94	1.68

(b) $n = 500$ and $N = 10,000$

	Supervised		Semi-Supervised									
			Random Forests					Basis Expansion				
Parameter	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
γ_{11}=1.4	0.01	0.22	0.01	0.12	0.11	0.92	1.76	0.01	0.12	0.11	0.92	1.80
γ_{12}=-2.6	0	0.29	0	0.17	0.16	0.93	1.73	-0.01	0.16	0.15	0.93	1.80
γ_{21}=0.8	0.00	0.17	0.00	0.10	0.09	0.93	1.80	0.00	0.09	0.09	0.93	1.86
γ_{22}=0.2	-0.01	0.23	0	0.13	0.12	0.93	1.81	0	0.13	0.12	0.94	1.83
γ_{23}=0.5	0.00	0.09	0.00	0.05	0.05	0.94	1.78	0.00	0.05	0.05	0.95	1.81

Table 1: Bias, empirical standard error (ESE) of the supervised and the SSL estimators with either random forest or basis expansion imputation strategies for $\bar{\gamma}_1, \bar{\gamma}_2$ when (a) $n = 135$ and $N = 1272$ and (b) $n = 500$ and $N = 10,000$. For the SSL estimators, we also obtain the average of the estimated standard errors (ASE) as well as the empirical coverage probabilities (CovP) of the 95% confidence intervals.

for any degree of mis-specification for both small and large finite sample settings. As the theory predicts, there is no real difference in efficiency gain of SSL across mis-specification of the Q function models. This is because asymptotic distribution of $\widehat{\bar{\gamma}}_{\mathsf{SSL}}$ shown in Theorems 2 & 3 are centered on the target parameters $\bar{\gamma}$. Thus, both SSL and SUP have negligible bias regardless of the true value of β_{27}^0.

Next we analyze performance of the doubly robust value function estimators for both continuous and EHR-like settings. Table 2 shows bias and RMSE across different sample sizes, and comparing SSL vs. SUP estimators. Results are shown for the correct specification of the Q functions and propensity scores, and when either is mis-specified. Bias across simulation settings is relatively similar between $\widehat{V}_{\mathsf{SSL_{DR}}}$ and $\widehat{V}_{\mathsf{SUP_{DR}}}$, and appears to be small relative to RMSE. The low magnitude of bias suggests both estimators are robust to model mis-specification. There is an exception on the EHR setting with small sample size, for which the bias is non-negligible. This is likely due to the fact that the Q function parameters to estimate are 13+37, and the propensity score functions have 12 parameters which add up to a large number relative to the labeled sample size: $n = 135$. The SSL bias is lower in this case which could be due to the refitting step, which helped to reduce the finite sample bias. Efficiency gains of $\widehat{V}_{\mathsf{SSL_{DR}}}$ are consistent across model specification.

Finally we discuss results for the $T > 2$ settings. Table 3 exhibits the interaction effect estimates $\boldsymbol{\gamma}_t$ of the Q functions for $t = 1, \ldots, T$, with $T = 3, 5, 7$. Results show that SSL

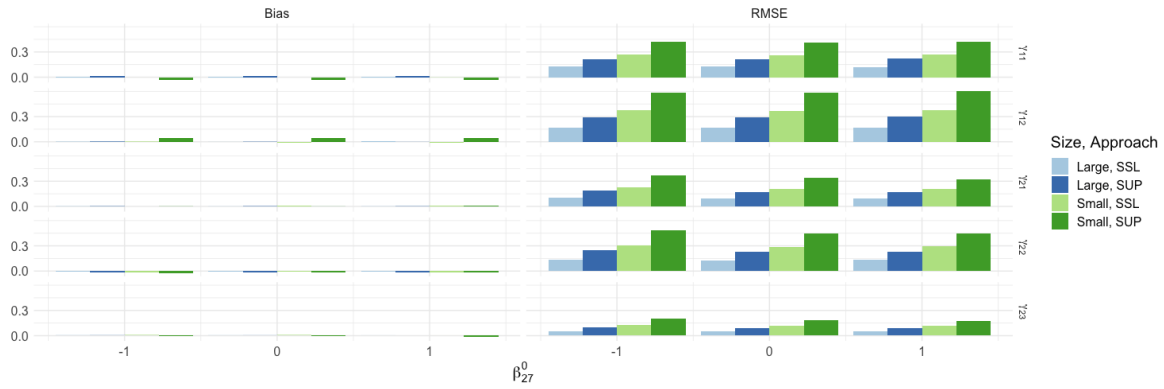

Figure 1: Monte Carlo estimates of bias and RMSE ratios for estimation of γ_{11}, γ_{12}, γ_{21}, γ_{22}, γ_{23} under mis-specification of the Q functions through β^0_{27}. Results are shown for the large ($N = 10,000$, $n = 500$) and small ($N = 1,272$, $n = 135$) data samples for the continuous setting over 1,000 simulated data sets.

estimates remain more efficient than the supervised counterpart even as the time horizon increases. Although the SSL method requires imputation of $O(T^2)$ functions of the outcome, it still has low bias, and is much more efficient than the supervised counterpart as shown by the ≈ 2 relative efficiency for comparing SSL vs. supervised learning across settings. Similarly, Table 4, which displays the bias, standard error and the efficiency of the value function estimates, shows that for all time horizons the SSL outperforms its supervised counterpart in terms of efficiency. However as expected, both estimators loose efficiency as time horizon T increases. This is due to the fundamental information-theoretic difficult nature of the estimation problem for large T, in these contexts simplifying assumptions such as MDP or others are usually made (see Uehara et al., 2022). The high variance estimates for both approaches dominate the relative efficiency gain of SSL estimation as T grows. We also note that the relative efficiency is seemingly constant across correct and mis-specified models as our theoretical results state. We next illustrate our approach using an inflammatory bowl disease (IBD) data set.

6.2 Application to an EHR Study of Inflammatory Bowel Disease

Anti–tumor necrosis factor (anti-TNF) therapy has greatly changed the management and improved the outcomes of patients with IBD (Peyrin-Biroulet, 2010). However, it remains unclear whether a specific anti-TNF agent has any advantage in efficacy over other agents, especially at the individual level. There have been few randomized clinical trials performed to directly compare anti-TNF agents for treating IBD patients (Sands et al., 2019). Retrospective studies comparing infliximab and adalimumab for treating IBD have found limited and sometimes conflicting evidence of their relative effectiveness (Inokuchi et al., 2019; Lee et al., 2019; Osterman and Lichtenstein, 2017). There is even less evidence regarding optimal DTR for choosing these treatments over time (Ananthakrishnan et al., 2016). To explore this, we performed RL using data from a cohort of IBD patients previously identified via

(a) $n = 135$ and $N = 1272$

			Supervised		Semi-Supervised									
					Random Forests					Basis Expansion				
Setting	Model	$\bar{V}$	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
Continuous	Correct	6.08	0.02	0.27	0.04	0.21	0.24	0.97	1.27	0.02	0.23	0.25	0.97	1.18
	Missp. Q	6.34	0.01	0.24	0.03	0.19	0.22	0.97	1.27	0.00	0.20	0.22	0.97	1.20
	Missp. π	6.08	0.01	0.28	0.02	0.22	0.24	0.97	1.24	0.01	0.25	0.25	0.97	1.12
EHR	Correct	1.38	0.09	0.15	0.05	0.12	0.12	0.94	1.24	0.04	0.13	0.12	0.95	1.12
	Missp. Q	1.43	0.09	0.14	0.04	0.12	0.12	0.96	1.12	0.03	0.14	0.12	0.95	1.02
	Missp. π	1.38	0.09	0.15	0.05	0.14	0.13	0.96	1.13	0.04	0.14	0.13	0.96	1.05

(b) $n = 500$ and $N = 10,000$

			Supervised		Semi-Supervised									
					Random Forests					Basis Expansion				
Setting	Model	$\bar{V}$	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
Continuous	Correct	6.08	0.02	0.15	0.03	0.11	0.12	0.96	1.32	0.02	0.13	0.13	0.95	1.16
	Missp. Q	6.34	0.01	0.13	0.03	0.10	0.10	0.96	1.31	0.01	0.11	0.11	0.96	1.16
	Missp. π	6.08	0.01	0.14	0.03	0.11	0.12	0.96	1.28	0.02	0.12	0.12	0.95	1.16
EHR	Correct	1.38	0.02	0.07	0.01	0.04	0.06	0.99	1.55	0.00	0.06	0.06	0.98	1.23
	Missp. Q	1.43	0.01	0.07	0.00	0.04	0.05	0.99	1.66	0.00	0.05	0.06	0.98	1.35
	Missp. π	1.38	0.02	0.08	0.01	0.06	0.07	0.99	1.22	0.00	0.07	0.07	0.97	1.03

Table 2: Bias, empirical standard error (ESE) of the supervised estimator $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$ and bias, ESE, average standard error (ASE) and coverage probability (CovP) for $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$ with either random forest or basis expansion imputation strategies when (a) $n = 135$ and $N = 1272$ and (b) $n = 500$ and $N = 10,000$. We show performance and relative efficiency across both simulation settings for estimation under correct models, and mis-specification of Q function or propensity score function.

machine learning algorithms from the EHR systems of two tertiary referral academic centers in the Greater Boston metropolitan area (Ananthakrishnan et al., 2012). We focused on the subset of $N = 1,272$ patients who initiated either Infliximab ($A_1 = 0$) or Adalimumab ($A_1 = 1$) and continued to be treated by either of these two therapies during the next 6 months. The observed treatment sequence distributions are shown in Table 5. The outcomes of interest are the binary indicator of treatment response at 6 months ($t = 2$) and at 12 months ($t = 3$), both of which were only available on a subset of $n = 135$ patients whose outcomes were manually annotated via chart review.

To derive the DTR, we included gender, age, Charlson co-morbidity index (Charlson et al., 1987), prior exposure to anti-TNF agents, as well as mentions of clinical terms associated with IBD such as bleeding complications extracted from the clinical notes via natural language processing (NLP). These features and confounding variables are adjusted for in the Q functions at both time points. To improve the imputation of Y_t, we use 15 relevant NLP features such as mentions of rectal or bowel resection surgery as surrogates at $t = 1, 2$. We transformed all count variables using $x \mapsto \log(1 + x)$ to decrease skewness in the distributions, and centered continuous features. We used RF with 500 trees to carry out the imputation step, and 5-fold cross-validation (CV) to estimate the value function. We only consider observations with estimated propensity scores within the [0.1,0.9] range to address the lack of overlap in the covariate distributions between treatment groups. We

	$T=3$					$T=5$					$T=7$				
	Sup.		SSL			Sup.		SSL			Sup.		SSL		
	Bias	ESE	Bias	ESE	RE	Bias	ESE	Bias	ESE	RE	Bias	ESE	Bias	ESE	RE
γ_1	0.02	0.18	0.09	0.09	2.14	0.00	0.18	0.13	0.08	2.32	0.01	0.19	0.14	0.08	2.32
γ_2	0.03	0.13	0.03	0.07	2.0	0.04	0.12	0.06	0.06	2.07	0.03	0.13	0.07	0.06	2.21
γ_3	0.01	0.11	0.1	0.06	2.08	0.01	0.12	0.06	0.06	2.09	0.05	0.12	0.06	0.06	2.09
γ_4						0.03	0.11	0.08	0.05	2.01	0.08	0.12	0.09	0.06	2.12
γ_5						0.01	0.1	0.07	0.04	1.87	0.04	0.11	0.08	0.06	1.95
γ_6											0.03	0.08	0.1	0.05	1.89
γ_7											0.02	0.03	0.08	0.04	1.97

Table 3: Mean absolute bias, and empirical standard error (ESE) of the supervised and the SSL Q function interaction effect estimates $\boldsymbol{\gamma}_t$, for $T = 3, 5, 7$ stages. Random forest imputation is used for SSL estimation.

	$T=3$					$T=5$					$T=7$				
	Sup.		SSL			Sup.		SSL			Sup.		SSL		
Model	Bias	ESE	Bias	ESE	RE	Bias	ESE	Bias	ESE	RE	Bias	ESE	Bias	ESE	RE
Correct	0.0	0.11	0.01	0.08	1.47	0.0	0.25	0.0	0.19	1.34	0.05	0.68	0.05	0.61	1.12
Missp. Q	0.0	0.11	0.01	0.07	1.52	0.02	0.32	0.01	0.25	1.29	0.02	0.61	0.02	0.54	1.09
Missp. π	0.01	0.11	0.01	0.07	1.49	0.03	0.33	0.02	0.24	1.38	0.05	0.8	0.03	0.66	1.22

Table 4: Bias, and empirical standard error (ESE) of the value function estimators $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$, $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$ for $T = 3, 5, 7$ stages. Random forests imputation is used for semi-supervised estimation. We show performance and relative efficiency for estimation under correct models, and mis-specification of Q function or propensity score functions.

use this approximation to the optimal selection of observations proposed by Crump et al. (2009). Additionally, we use ridge regularization for our natural cubic splines model of the propensity scores.

The supervised and semi-supervised estimates are shown in Table 6 for the Q-learning models and in Table 7 for the value functions associated with the estimated DTR. Similar to those observed in the simulation studies, the semi-supervised Q-learning has more power to detect significant predictors of treatment response. Relative efficiency for almost all Q function estimates is near or over 2. The supervised Q-learning does not have the power to detect predictors such as prior use of anti-TNF agents, which are clearly relevant to treatment response (Ananthakrishnan et al., 2016). Semi-supervised Q-learning is able to detect that the efficacy of Adalimumab wears off as patients get older, meaning younger patients in the first stage experienced a higher rate of treatment response to Adalimumab, a finding that cannot be detected with supervised Q-learning. Additionally, supervised Q-learning does not pick up that there is a higher rate of response to Adalimumab among patients that are male or have experienced an abscess. This translates into a far from optimal treatment rule as seen in the cross-validated value function estimates. Table 7 reflects that using our semi-supervised approach to find the regime and to estimate the value function of such treatment policy yields a more efficient estimate, as the semi-supervised value function estimate $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$ yielded a smaller standard error than that of the supervised estimate $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$. However, the standard errors are large relative to the point estimates. On the upside, they both yield estimates very close in numerical value which is reassuring: both should be unbiased as predicted by theory and simulations.

		A_1	
		0	1
A_2	0	912	327
	1	27	183

Table 5: Distribution of treatment trajectories for an observed sample of size 1407.

7. Discussion

We have proposed an efficient and robust strategy for estimating optimal DTRs and their value function in a setting where patient outcomes are scarce. In particular, we developed a two step estimation procedure amenable to non-parametric imputation of the missing outcomes. This helped us establish $\sqrt{n}$-consistency and asymptotic normality for both the Q function parameters $\widehat{\boldsymbol{\theta}}$ and the doubly robust value function estimator $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$. We additionally provided theoretical results which illustrate if and when the outcome-surrogates $\mathbf{W}$ contribute towards efficiency gain in estimation of $\widehat{\boldsymbol{\theta}}_{\mathsf{SSL}}$ and $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$. These results let us conclude that our procedure is always preferable to using the labeled data only: since estimation is robust to mis-specification of the imputation models, our approach is safe to use and will be at least as efficient as the supervised methods.

Regarding our theoretical results, we believe that no specific aspects of the proofs explicitly require the number stages to be $T = 2$. Indeed, we hypothesize that we can generalize the theory to any fixed $T > 2$ using induction, as we have already proven the results for the

Stage 1 Regression							
	Supervised			Semi-Supervised			
Parameter	Estimate	SE	P-val	Estimate	SE	P-val	RE
Intercept	**0.424**	**0.082**	**0.00**	**0.518**	**0.028**	**0.00**	2.937
Female	-0.237	0.167	0.16	**-0.184**	**0.067**	**0.007**	2.514
Age	0.155	0.088	0.081	**0.18**	**0.034**	**0.00**	2.588
Charlson Score	0.006	0.072	0.929	-0.047	0.026	0.075	2.776
Prior anti-TNF	-0.038	0.06	0.524	**-0.085**	**0.019**	**0.00**	3.177
Perianal	**0.138**	**0.06**	**0.022**	**0.179**	**0.022**	**0.00**	2.688
Bleeding	0.049	0.08	0.54	0.058	0.03	0.055	2.675
A1	0.163	0.488	0.739	0.148	0.206	0.473	2.374
Female$\times A_1$	0.168	0.696	0.81	-0.042	0.287	0.886	2.424
Age$\times A_1$	-0.177	0.264	0.503	**-0.278**	**0.109**	**0.013**	2.418
Charlson Score$\times A_1$	0.136	0.391	0.728	0.195	0.178	0.276	2.194
Perianal$\times A_1$	-0.113	0.226	0.618	-0.019	0.08	0.808	2.838
Bleeding$\times A_1$	0.262	0.364	0.474	0.127	0.161	0.431	2.267

Stage 2 Regression							
	Supervised			Semi-Supervised			
Parameter	Estimate	SE	P-val	Estimate	SE	P-val	RE
Y_1	**0.37**	**0.11**	**0.00**	**0.55**	**0.05**	**0.00**	2.08
Intercept	0.08	0.06	0.17	0.04	0.02	0.14	2.40
Female	-0.01	0.10	0.92	-0.00	0.05	0.98	2.21
Age	0.05	0.06	0.35	**0.07**	**0.02**	**0.00**	2.33
Charlson Score	0.04	0.04	0.33	**0.06**	**0.02**	**0.01**	2.06
Prior anti-TNF	-0.05	0.05	0.29	**-0.09**	**0.02**	**0.00**	2.39
Perianal	-0.01	0.04	0.80	**-0.03**	**0.02**	**0.06**	2.31
Bleeding	-0.04	0.05	0.49	-0.03	0.03	0.29	2.14
A1	0.11	0.25	0.67	0.03	0.10	0.74	2.60
Abscess_2	0.06	0.04	0.16	**0.05**	**0.01**	**0.00**	2.68
Fistula_2	0.02	0.05	0.67	0.01	0.02	0.62	2.33
Female$\times A_1$	0.13	0.38	0.74	0.17	0.16	0.30	2.37
Age$\times A_1$	-0.02	0.12	0.88	-0.09	0.06	0.17	1.94
Charlson Score$\times A_1$	-0.02	0.16	0.89	0.04	0.07	0.55	2.19
Perianal$\times A_1$	-0.14	0.09	0.15	**-0.17**	**0.04**	**0.00**	2.34
Bleeding$\times A_1$	0.13	0.20	0.51	0.03	0.09	0.76	2.17
A2	0.07	0.17	0.69	**0.22**	**0.07**	**0.00**	2.55
Female$\times A_2$	-0.39	0.28	0.16	**-0.51**	**0.11**	**0.00**	2.53
Age$\times A_2$	0.09	0.10	0.40	**0.15**	**0.04**	**0.00**	2.27
Charlson Score$\times A_2$	0.01	0.07	0.84	-0.03	0.03	0.42	2.08
Perianal$\times A_2$	**0.20**	**0.09**	**0.04**	**0.23**	**0.04**	**0.00**	2.23
Bleeding$\times A_2$	0.03	0.08	0.77	0.02	0.04	0.49	2.34
$\text{Abscess}_2 \times A_2$	-0.13	0.07	0.06	**-0.09**	**0.03**	**0.00**	2.31
$\text{Fistula}_2 \times A_2$	-0.04	0.06	0.56	-0.03	0.03	0.36	2.17

Table 6: Results for the Inflammatory Bowel Disease data set, for first and second stage regressions. Fully supervised Q-learning is shown on the left and semi-supervised is shown on the right. Last columns in the panels show relative efficiency (RE) defined as the ratio of standard errors of the semi-supervised vs. supervised method, RE greater than one favors semi-supervised. Statistically significant coefficients at the 0.05 level are in bold.

	Estimate	SE
$\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$	0.851	0.486
$\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$	0.871	0.397

Table 7: Value function estimates for the Inflammatory Bowel Disease data set. The first row shows the estimate for treatment rule learned using $\mathcal{U}$ and its respective value function, the second row shows the same for a rule estimated using $\mathcal{L}$ and its estimated value.

first couple of stages. We chose to leave this generalization for future work, as we believe no new or particularly interesting theoretical methodology is required for this extension, beyond the already cumbersome notation and book-keeping in this relatively simpler $T = 2$ setting.

Both the semi-supervised Q and value function estimation hold validity as long as the time horizon T remains finite. However, it is crucial to acknowledge that practical implementation may introduce instability to our estimators, particularly in cases involving large values of T. Instability in the presence of a large T scenario is a commonly observed phenomenon even for supervised approaches in RL problems. This issue has gained much attention in the context of policy evaluation, where it was shown that (supervised) doubly robust estimators can exhibit instability due to their reliance on products of T inverse propensity weights (cf. Jiang and Li, 2016a; Thomas and Brunskill, 2016a; Kallus and Ue-

hara, 2020b). Such terms tend to exhibit significant variability as T increases, primarily because they often involve nested products of these weights (Levine et al., 2020).

In the offline RL literature, alternative policy evaluation methods have been proposed to improve stability. For instance, methods such as the weighted doubly robust estimator, MAGIC (Thomas and Brunskill, 2016a), and IH (Liu et al., 2018), have shown promise. Detailed information can be found in Voloshin et al. (2019). However, theoretical properties of these estimators are not as well-understood. Therefore, utilizing these ideas for our problem is not feasible with currently available tools.

As the time horizon T increases in our semi-supervised method, the number of terms requiring imputation for the Q and value function estimations is naturally higher. While this leads to increased variance, our approach offers a key advantage: the complexity of the conditional means to be imputed remains constant with respect to T. Therefore, efficiency is still gained as the primary source of variability arises from higher-order propensity scores, similar to previously mentioned supervised methods.

In particular, Appendix B delves into the generalization of our SSL Q-learning algorithm and value function estimation. We provide a thorough illustration of the functions requiring imputation. The analysis reveals that these functions exclusively consist of linear or quadratic terms involving missing outcomes, analogous to the $T = 2$ case. Existing imputation techniques readily extend to $T > 2$ scenarios, but the number of terms requiring imputation grows quadratically with time horizon ($O(T^2)$). The normal equations for a three-stage setting is presented to concretely illustrate the advantages and challenges of our semi-supervised Q-learning approach. Additionally, the doubly robust SSL value function algorithm is extended for the general $T > 2$ case. As expected, the value function also requires $O(T^2)$ terms to be imputed, limited to linear or quadratic terms of missing outcomes and propensity scores, again highlighting the manageability of complexity within our framework.

Finally, we are interested in extending this framework to handle missing at random (MAR) sampling mechanisms in the future. In the EHR setting, it is feasible to sample a subset of the data completely at random in order to annotate the records. Hence, we argue that our missingness assumption, which requires that the labeled sample has the same distribution as the unlabeled sample, is satisfied by design as we choose a random sample from the unlabeled data and then label it. However, the MAR context allows us to leverage different data sources for $\mathcal{L}$ and $\mathcal{U}$. For example, we could use an annotated EHR data cohort and a large unlabeled registry data repository for our inference, ultimately making the policies and value estimation more efficient and robust. We believe this line of work has the potential to leverage massive observational cohorts, which will help to improve personalized clinical care for a wide range of diseases.

Acknowledgments

We thank the reviewers and action editor for their insightful review, which significantly improved our work through their valuable comments and suggestions. Nilanjana Laha was partially supported by the NSF-DMS grant DMS-2311098, Tianxi Cai was partially

supported by the National Institutes of Health (R01LM013614). Rajarshi Mukherjee was partially supported by NSF Grant EAGER-1941419.

References

Ashwin N Ananthakrishnan, Tianxi Cai, SC Cheng, Pj Chen, G Savova, RG Perez, Vs Gainer, Sn Murphy, P Szolovits, K Liao, Ew Karlson, S Churchill, I Kohane, and RM Plenge. Improving case definition of crohn's disease and ulcerative colitis in electronic medical records using natural language processing - a novel informatics approach. *Gastroenterology*, 142(5):S791–S791, 2012. ISSN 0016-5085.

Ashwin N Ananthakrishnan, A Cagan, Tianxi Cai, Vs Gainer, S Shaw, S Churchill, E Karlson, I Kohane, K Liao, and S Murphy. Comparative effectiveness of infliximab and adalimumab in crohn's disease and ulcerative colitis. *Gastroenterology*, 150(4):S979–S979, 2016. ISSN 0016-5085.

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89 (1):133–161, 2021.

Susan Athey, Raj Chetty, Guido W Imbens, and Hyunseung Kang. The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. *Technical report, National Bureau of Economic Research*, 2019.

Colin B. Begg and Denis H. Y. Leung. On the use of surrogate end points in randomized trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 163(1): 15–28, 2000. ISSN 09641998, 1467985X. URL `http://www.jstor.org/stable/2680505`.

Andrew Bennett and Nathan Kallus. Efficient policy learning from surrogate-loss classification reductions. In *International Conference on Machine Learning*, pages 788–798. PMLR, 2020.

G Biau, L Devroye, and G Lugosi. Consistency of random forests and other averaging classifiers. *Journal Of Machine Learning Research*, 9:2015–2033, 2008. ISSN 1532-4435.

John Blitzer and Xiaojin Zhu. Semi-supervised learning for natural language processing. In *ACL (Tutorial Abstracts)*, page 3, 2008. URL `http://www.aclweb.org/anthology/P08-5003`.

Abhishek Chakrabortty. Robust semi-parametric inference in semi-supervised settings, 2016.

Abhishek Chakrabortty, Tianxi Cai, et al. Efficient and adaptive linear regression in semi-supervised settings. *The Annals of Statistics*, 46(4):1541–1572, 2018.

Abhishek Chakrabortty, Guorong Dai, and Eric Tchetgen Tchetgen. A general framework for treatment effect estimation in semi-supervised and high dimensional settings. 2022. doi: 10.48550/ARXIV.2201.00468. URL `https://arxiv.org/abs/2201.00468`.

Bibhas Chakraborty and Erica E.M Moodie. *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. Statistics for Biology and Health. Springer New York, New York, NY, 2013 edition, 2013. ISBN 9781461474272.

Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-supervised learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 2006.

Mary E Charlson, Peter Pompei, Kathy L Ales, and C.Ronald Mackenzie. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, 40(5):373–383, 1987. ISSN 0021-9681.

David Cheng, Ashwin N Ananthakrishnan, and Tianxi Cai. Robust and efficient semi-supervised estimation of average treatment effects with application to electronic health records data. *Biometrics*, 2020.

Richard Crump, V Hotz, Guido Imbens, and Oscar Mitnik. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199, 2009. ISSN 00063444. URL `http://search.proquest.com/docview/201696479/`.

Marie Davidian, Anastasios A. Tsiatis, and Selene Leon. Semiparametric Estimation of Treatment Effect in a Pretest–Posttest Study with Missing Data. *Statistical Science*, 20(3):261 – 301, 2005. doi: 10.1214/088342305000000151. URL `https://doi.org/10.1214/088342305000000151`.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. *Proceedings of the 28th International Conference on Machine Learning*, 2011.

R.M Dudley. Balls in rk do not cut all subsets of k + 2 points. *Advances in mathematics (New York. 1965)*, 31(3):306–308, 1979. ISSN 0001-8708.

Chelsea Finn, Tianhe Yu, Justin Fu, Pieter Abbeel, and Sergey Levine. Generalizing skills with semi-supervised reinforcement learning. 2016.

T.J Hastie. *Statistical Models in S*. CRC Press, 1 edition, 1992. ISBN 041283040X.

Chuan Hong, Katherine P Liao, and Tianxi Cai. Semi-supervised validation of multiple surrogate outcomes with application to electronic medical records phenotyping. *Biometrics*, 75(1):78–89, 2019.

Toshihiro Inokuchi, Sakuma Takahashi, Sakiko Hiraoka, Tatsuya Toyokawa, Shinjiro Takagi, Koji Takemoto, Jiro Miyaike, Tsuyoshi Fujimoto, Reiji Higashi, Yuki Morito, et al. Long-term outcomes of patients with crohn's disease who received infliximab or adalimumab as the first-line biologics. *Journal of gastroenterology and hepatology*, 34(8):1329–1336, 2019.

Nan Jiang and Lihong Li. Doubly robust off-policy value evaluation for reinforcement learning. *In International Conference on Machine Learning, pages 652–661. PMLR*, 2016a.

Nan Jiang and Lihong Li. Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 652–661. PMLR, 2016b.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1):99–134, 1998. ISSN 0004-3702. doi: https://doi.org/10.1016/S0004-3702(98)00023-X. URL `https://www.sciencedirect.com/science/article/pii/S000437029800023X`.

Nathan Kallus and Xiaojie Mao. On the role of surrogates in the efficient estimation of treatment effects with limited outcome data. *arXiv preprint arXiv:2003.12408*, 2020.

Nathan Kallus and Masatoshi Uehara. Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *Journal of Machine Learning Research*, 21(167), 2020a.

Nathan Kallus and Masatoshi. Uehara. Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *Journal of Machine Learning Research, 21(167): 1–63*, 2020b.

Michael R. Kosorok and Eric B. Laber. Precision medicine. 6(1):263–286, 2019. ISSN 2326-8298.

Eric B Laber, Daniel J Lizotte, Min Qian, William E Pelham, and Susan A Murphy. Dynamic treatment regimes: technical challenges and applications. *Electronic journal of statistics*, 8(1):1225–1272, 2014. ISSN 1935-7524. URL `http://search.proquest.com/docview/1826600138/`.

Yongil Lee, Jae Hee Cheon, Yehyun Park, Soo Jung Park, Tae Il Kim, and Won Ho Kim. Comparison of long-term outcomes between infliximab and adalimumab in biologic-naive patients with ulcerative colitis. *Gut & Liver*, 13, 2019.

Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.

Qiang Liu, Lihong Li, Ziyang Tang, and Dengyong Zhou. Breaking the curse of horizon: Infinite-horizon off-policy estimation. *Advances in Neural Information Processing Systems*, 31, 2018.

S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003. doi: 10.1111/1467-9868.00389. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00389`.

SA Murphy. A generalization error for q-learning. *Journal Of Machine Learning Research*, 6:1073–1097, 2005. ISSN 1532-4435.

Susan A Murphy, Mark J van der Laan, James M Robins, and Conduct Problems Prevention Research Group. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423, 2001.

Mark T Osterman and Gary R Lichtenstein. Infliximab vs adalimumab for uc: Is there a difference? *Clinical Gastroenterology and Hepatology*, 15(8):1197–1199, 2017.

Margaret Sullivan Pepe. Inference using surrogate outcome data and a validation sample. *Biometrika*, 79(2):355–365, 1992. ISSN 00063444. URL http://www.jstor.org/stable/2336846.

L Peyrin-Biroulet. Anti-tnf therapy in inflammatory bowel diseases: a huge review. *Minerva gastroenterologica e dietologica*, 56(2):233, 2010.

Ross L. Prentice. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8(4):431–440, 1989. doi: https://doi.org/10.1002/sim.4780080407. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780080407.

Siyuan Qiao, Wei Shen, Zhishuai Zhang, Bo Wang, and Alan Yuille. Deep co-training for semi-supervised image recognition, 2018.

J. Robins. Causal inference from complex longitudinal data. *Latent Variable Modeling and Applications to Causality*, pages 69—117, 1997.

James M. Robins. *Optimal Structural Nested Models for Optimal Sequential Decisions*, pages 189–326. Springer New York, New York, NY, 2004. ISBN 978-1-4419-9076-1. doi: 10.1007/978-1-4419-9076-1_11. URL https://doi.org/10.1007/978-1-4419-9076-1_11.

Bruce E Sands, Laurent Peyrin-Biroulet, Edward V Loftus Jr, Silvio Danese, Jean-Frédéric Colombel, Murat Törüner, Laimas Jonaitis, Brihad Abhyankar, Jingjing Chen, Raquel Rogers, et al. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. *New England Journal of Medicine*, 381(13):1215–1226, 2019.

Phillip J. Schulte, Anastasios A. Tsiatis, Eric B. Laber, and Marie Davidian. **Q**- and **A**-learning methods for estimating optimal dynamic treatment regimes. *Statist. Sci.*, 29(4):640–661, 11 2014. doi: 10.1214/13-STS450. URL https://doi.org/10.1214/13-STS450.

Erwan Scornet, Gérard Biau, and Jean-Philippe Vert. Consistency of random forests. *Annals of Statistics*, 43(4):1716, 2015. ISSN 00905364. URL http://search.proquest.com/docview/1787036058/.

Aaron Sonabend-W, Nilanjana Laha, Rajarshi Mukherjee, and Tianxi Cai. Semi-supervised off policy reinforcement learning, 2020a.

Aaron Sonabend-W, Junwei Lu, Leo Anthony Celi, Tianxi Cai, and Peter Szolovits. Expert-supervised reinforcement learning for offline policy learning and evaluation. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18967–18977. Curran Associates, Inc., 2020b. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/daf642455364613e2120c636b5a1f9c7-Paper.pdf.

Aaron Sonabend W., Amelia M. Pellegrini, Stephanie Chan, Hannah E. Brown, James N. Rosenquist, Pieter J. Vuijk, Alysa E. Doyle, Roy H. Perlis, and Tianxi Cai. Integrating questionnaire measures for transdiagnostic psychiatric phenotyping using word2vec.

PLOS ONE, 15(4):1–14, 04 2020. doi: 10.1371/journal.pone.0230663. URL https://doi.org/10.1371/journal.pone.0230663.

Richard S. Sutton. *Reinforcement learning : an introduction*. Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts ; London, England, second edition. edition, 2018. ISBN 9780262039246.

Philip Thomas and Emma Brunskill. Data-efficient off-policy policy evaluation for reinforcement learning. In *ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning*, volume 48, pages 2139–2148. PMLR, 2016a.

Philip S. Thomas and Emma Brunskill. Data-efficient off-policy policy evaluation for reinforcement learning. 2016b.

Anastasios A Tsiatis. *Semiparametric Theory and Missing Data*. Springer Series in Statistics. Springer New York, New York, NY, 2006. ISBN 9780387324487.

Anastasios A Tsiatis, Marie Davidian, Shannon T Holloway, and Eric B Laber. Dynamic treatment regimes: Statistical methods for precision medicine. *CRC press.*, 2019.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, New York, NY, 2009. ISBN 978-0-387-79051-0.

Masatoshi Uehara, Chengchun Shi, and Nathan Kallus. A review of off-policy evaluation in reinforcement learning, 2022.

A. W. van der Vaart. *Asymptotic statistics*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK ; New York, NY, USA, 1998. ISBN 0521496039.

Mark J Van der Laan and Sherri. Rose. *Targeted learning in data science*. Springer, 2018.

Aad W van der Vaart and Jon A Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, New York, 1996. ISBN 9781475725476.

Aad W. Van Der Vaart and Jon A. Wellner. Empirical processes indexed by estimated functions. *Lecture Notes-Monograph Series*, 55:234–252, 2007. ISSN 07492170.

Cameron Voloshin, Hoang M Le, Nan Jiang, and Yisong Yue. Empirical study of off-policy policy evaluation for reinforcement learning. *arXiv preprint arXiv:1911.06854*, 2019.

Larry Wasserman and John D. Lafferty. Statistical analysis of semi-supervised regression. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 801–808. Curran Associates, Inc., 2008. URL http://papers.nips.cc/paper/3376-statistical-analysis-of-semi-supervised-regression.pdf.

Christopher John Cornish Hellaby Watkins. Learning from delayed rewards, 1989.

Baqun Zhang, Anastasios A Tsiatis, Eric B Laber, and Marie Davidian. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika, 100 (3):681–694*, 2013.

Yichi Zhang, Tianrun Cai, Sheng Yu, Kelly Cho, Chuan Hong, Jiehuan Sun, Jie Huang, Yuk-Lam Ho, Ashwin N Ananthakrishnan, Zongqi Xia, et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (phecap). *Nature Protocols*, 14(12):3426–3444, 2019.

Ying-Qi Zhao, Donglin Zeng, Eric B Laber, and Michael R Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015. ISSN 0162-1459. URL `http://www.tandfonline.com/doi/abs/10.1080/01621459.2014.937488`.

Ying-Qi Zhao, Eric B Laber, Yang Ning, Sumona Saha, and Bruce E Sands. Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *The Journal of Machine Learning Research*, 20(1):1821–1843, 2019.

Wang Zhixing and Chen Shaohong. Web page classification based on semi-supervised naïve bayesian em algorithm. In *2011 IEEE 3rd International Conference on Communication Software and Networks*, pages 242–245. IEEE, 2011. ISBN 9781612844855.

Doudou Zhou, Yufeng Zhang, Aaron Sonabend-W, Zhaoran Wang, Junwei Lu, and Tianxi Cai. Federated offline reinforcement learning, 2023.

Zhengyuan Zhou, Susan Athey, and Stefan Wager. Offline multi-action policy learning: Generalization and optimization. *Operations Research*, 2022a.

Zhengyuan Zhou, Susan Athey, and Stefan Wager. Offline multi-action policy learning: Generalization and optimization. *Operations Research*, 2022b.

Wensheng Zhu, Donglin Zeng, and Rui Song. Proper inference for value function in high-dimensional q-learning for dynamic treatment regimes. *Journal of the American Statistical Association*, 114(527):1404–1417, 2019. ISSN 0162-1459. URL `http://www.tandfonline.com/doi/abs/10.1080/01621459.2018.1506341`.

Xiaojin Zhu. Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison, 2008.

Appendix A. Notation

We first summarize the main notation used throughout the paper in Table 8. Note that we use a few conventions: 1) The check symbol: " ˇ " denotes the entire set of features available, including outcomes when available (for $t = 2$), this symbol is used for both the patient history $\check{\mathbf{H}}_t$, and the linear features for the $Q-$functions $\check{\mathbf{X}}_t$. 2) We use the bar symbol: " ¯ " to denote population parameters, for example $\bar{V}$ is the population value function under the optimal treatment policy $\bar{d}_t$. 3) As usual, the hat symbol "ˆ" is used to denote estimated parameters.

Notation	Definition
$\mathbf{O}_t \in \mathbb{R}^{d_t^o}$	Vector of covariates measured prior to t
$A_t \in \{0, 1\}$	A treatment indicator variable at t
$Y_{t+1} \in \mathbb{R}$	The outcome observed at $t + 1$
$\mathbf{W}_t \in \mathbb{R}^{d_t^\omega}$	A d_t^ω-dimensional vector of post-treatment covariates
$\mathbf{U}_t = (\mathbf{O}_t^\intercal, A_t, \mathbf{W}_t^\intercal)^\intercal$	All observed variables for labeled & unlabeled data at t
$\vec{\mathbf{U}} = (\mathbf{U}_1^\intercal, \mathbf{U}_2^\intercal)^\intercal$	All observed variables for $t = 1, 2$
$\boldsymbol{\phi}_{tk}(\cdot)$	Pre-specified basis functions for Q_t function regressions
$\mathbf{H}_{1k} = \boldsymbol{\phi}_{1k}(\mathbf{O}_1)$	Baseline ($k = 0$), or interaction ($k = 1$) features at $t = 1$
$\mathbf{H}_{2k} = \boldsymbol{\phi}_{2k}(\mathbf{O}_1, A_1, \mathbf{O}_2)$	Baseline ($k = 0$), or interaction ($k = 1$) features at $t = 2$
$\mathbf{H}_t \equiv [\mathbf{H}_{t0}^\intercal, \mathbf{H}_{t1}^\intercal]^\intercal$	Patient's concatenated history features at t
$\check{\mathbf{H}}_1 = \mathbf{H}_1$	Patient's history features at $t = 1$
$\check{\mathbf{H}}_2 = (Y_2, \mathbf{H}_2^\intercal)^\intercal$	Patient's history including previous outcome at $t = 2$
$\mathbf{X}_t = [\mathbf{H}_{t0}^\intercal, A_1\mathbf{H}_{t1}^\intercal]^\intercal$	Linear regression features available for all data
$\check{\mathbf{X}}_1 = \mathbf{X}_1$	Linear regression features for Q function at $t = 1$
$\check{\mathbf{X}}_2 = (Y_2, \mathbf{X}_2^\intercal)^\intercal$	Linear regression features for Q function at $t = 2$
$\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t^\intercal, \boldsymbol{\gamma}_t^\intercal)^\intercal$	Q function working model parameters
$\boldsymbol{\Sigma}_t = \mathbb{E}[\check{\mathbf{X}}_t\check{\mathbf{X}}_t^\intercal]$	Second moment for all features at stages $t = 1, 2$
$\widehat{Y}_2^* = Y_2 + \max_{a_2} Q_2(\check{\mathbf{H}}_2, a_2; \widehat{\boldsymbol{\theta}}_2)$	Labeled-data stage 1 estimated pseudo-outcome
$\bar{Y}_2^* = Y_2 + \max_{a_2} Q_2(\check{\mathbf{H}}_2, a_2; \bar{\boldsymbol{\theta}}_2)$	Stage 1 population pseudo-outcome
$\mu_{2t}(\vec{\mathbf{U}}) = \mathbb{E}(Y_2Y_t\|\vec{\mathbf{U}})$	Example of a conditional mean function
$\widehat{m}_{2t}(\vec{\mathbf{U}})$	Weakly or non-parametric estimator for $\mu_{2t}(\vec{\mathbf{U}})$
$\widehat{\eta}_{2t}$	Refitting-step linear parameter, ensures (3) is satisfied
$\widehat{\mu}_{2t}(\vec{\mathbf{U}}) = \frac{1}{K}\sum_k \widehat{m}_{2t}^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_{2t}$	Augmented cross-fitted model for $\mu_{2t}(\vec{\mathbf{U}})$
$\widehat{d}_t, \widehat{d}_t(\mathbf{H}_t), d_t(\mathbf{H}_t; \widehat{\boldsymbol{\theta}}_t)$	Estimated optimal treatment given history $\mathbf{H}_t$
$\bar{d}_t \equiv \bar{d}_t(\mathbf{H}_t) \equiv d_t(\mathbf{H}_t, \bar{\boldsymbol{\theta}}_t)$	Population-parameter optimal treatment given $\mathbf{H}_t$
$\pi_t(\check{\mathbf{H}}_t) = \mathbb{P}\{A_t = 1\|\check{\mathbf{H}}_t\}$	Treatment propensity scores at stages $t = 1, 2$
$\omega_t(\check{\mathbf{H}}_t, A_t, \boldsymbol{\Theta})$	Inverse probability weights at stages $t = 1, 2$
$\bar{V}$	Expected population outcome under optimal policy
$\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$	Supervised doubly robust estimator for $\bar{V}$
$\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$	Semi-Supervised doubly robust estimator for $\bar{V}$

Table 8: Main notation.

Appendix B. SSL Q-Learning and Off-Policy Value Estimation for $T > 2$ Stages

In this section we explore generalizing our methods to an arbitrary finite time horizon T and discuss some advantages and challenges of our method in this context. We start by showing the generalized SSL Q learning algorithm, and the functions that need imputation by writing the missing outcomes explicitly. We demonstrate that the functions of the outcome that need to be imputed consist of linear outcomes or products of at most two missing outcomes. Next, we extend the doubly robust SSL value function algorithm for $T > 2$. The value function, as expected, also has more terms to impute. However, these are all products of at most two missing outcomes or an outcome and a single propensity score function. We show that the terms in the individual products do not increase with T. Naturally, the number of conditional means needed to be imputed does increase. We end by highlighting the benefits and challenges of both algorithms applied to a time horizon larger than two stages.

SSL Q-Learning

Extending our Q function notation for a general time horizon fixed at T time steps we define

$$Q_t(\check{\mathbf{H}}_t, A_t) \equiv \mathbb{E}[Y_{t+1} + \max_a Q_{t+1}(\check{\mathbf{H}}_{t+1}, a) | \check{\mathbf{H}}_t, A_t] \text{ for } t = 1, \ldots, T,$$

and $Q_{T+1}(\check{\mathbf{H}}_{T+1}, A_{T+1}) \equiv 0$ for all $\check{\mathbf{H}}_{T+1}, A_{T+1}$.

The corresponding working linear models for the Q functions with parameters $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t^\intercal, \boldsymbol{\gamma}_t^\intercal)^\intercal$, $t = 1, \ldots, T$ are:

$$\begin{aligned}
Q_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\theta}_1^0) =& \check{\mathbf{X}}_1^\intercal \boldsymbol{\theta}_1^0 = \mathbf{H}_{10}^\intercal \boldsymbol{\beta}_1^0 + A_1(\mathbf{H}_{11}^\intercal \boldsymbol{\gamma}_1^0), \\
Q_2(\check{\mathbf{H}}_2, A_2; \boldsymbol{\theta}_2^0) =& \check{\mathbf{X}}_2^\intercal \boldsymbol{\theta}_2^0 = Y_2 \beta_{21}^0 + \mathbf{H}_{20}^\intercal \boldsymbol{\beta}_{22}^0 + A_2(\mathbf{H}_{21}^\intercal \boldsymbol{\gamma}_2^0), \\
&\vdots \\
Q_t(\check{\mathbf{H}}_t, A_t; \boldsymbol{\theta}_t^0) =& \check{\mathbf{X}}_t^\intercal \boldsymbol{\theta}_t^0 = \sum_{\ell=1}^{t-1} Y_{\ell+1} \beta_{t\ell}^0 + \mathbf{H}_{t0}^\intercal \boldsymbol{\beta}_{tt}^0 + A_t(\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t^0).
\end{aligned}$$

The pseudo-outcomes for stages 1 and $t = 2, \ldots, T$ under the linear Q function are

$$\bar{Y}_{ti}^* = Y_{ti} + \sum_{\ell=1}^{t-1} Y_{\ell+1,i} \beta_{t\ell}^0 + \mathbf{H}_{t0}^\intercal \boldsymbol{\beta}_{tt}^0 + [\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t^0]_+, t = 2, \ldots, T$$

Defining the outcome vector as $\vec{\mathbf{Y}}_{t:2} \equiv (Y_t, Y_{t-1}, \ldots, Y_2)^\intercal$, the normal equations for T, $t = 2, \ldots, T-1$ and $t = 1$ are $\mathbb{E}\left[\check{\mathbf{X}}_T(Y_{T+1} - \check{\mathbf{X}}_T^\intercal \boldsymbol{\theta}_T)\right] = \mathbf{0}$, $\mathbb{E}\left[\check{\mathbf{X}}_t(\bar{Y}_{t+1}^* - \check{\mathbf{X}}_t^\intercal \boldsymbol{\theta}_t)\right] = \mathbf{0}$, and

$\mathbb{E}\left[\check{\mathbf{X}}_1(\bar{Y}_2^* - \check{\mathbf{X}}_1^\intercal\boldsymbol{\theta}_1)\right] = \mathbf{0}$, which can be respectively written out as :

$$\mathbb{E}\begin{bmatrix} Y_T\{Y_{T+1} - (\vec{\mathbf{Y}}_{T:2}^\intercal, \mathbf{X}_T^\intercal)\boldsymbol{\theta}_T\} \\ Y_{T-1}\{Y_{T+1} - (\vec{\mathbf{Y}}_{T:2}^\intercal, \mathbf{X}_T^\intercal)\boldsymbol{\theta}_T\} \\ \vdots \\ Y_2\{Y_{T+1} - (\vec{\mathbf{Y}}_{T:2}^\intercal, \mathbf{X}_T^\intercal)\boldsymbol{\theta}_T\} \\ \mathbf{X}_T\{Y_{T+1} - (\vec{\mathbf{Y}}_{T:2}^\intercal, \mathbf{X}_T^\intercal)\boldsymbol{\theta}_T\} \end{bmatrix} = \mathbf{0},$$

$$\mathbb{E}\begin{bmatrix} Y_t\{Y_{t+1} + \sum_{\ell=1}^{t-1} Y_{\ell+1}\beta_{t+1,\ell}^0 + \mathbf{H}_{t+1,0}^\intercal\boldsymbol{\beta}_{t+1,t+1}^0 + [\mathbf{H}_{t+1,1}^\intercal\boldsymbol{\gamma}_{t+1}^0]_+ - (\vec{\mathbf{Y}}_{t:2}^\intercal, \mathbf{X}_t^\intercal)\boldsymbol{\theta}_t\} \\ Y_{t-1}\{Y_{t+1} + \sum_{\ell=1}^{t-1} Y_{\ell+1}\beta_{t+1,\ell}^0 + \mathbf{H}_{t+1,0}^\intercal\boldsymbol{\beta}_{t+1,t+1}^0 + [\mathbf{H}_{t+1,1}^\intercal\boldsymbol{\gamma}_{t+1}^0]_+ - (\vec{\mathbf{Y}}_{t:2}^\intercal, \mathbf{X}_t^\intercal)\boldsymbol{\theta}_t\} \\ \vdots \\ Y_2\{Y_{t+1} + \sum_{\ell=1}^{t-1} Y_{\ell+1}\beta_{t+1,\ell}^0 + \mathbf{H}_{t+1,0}^\intercal\boldsymbol{\beta}_{t+1,t+1}^0 + [\mathbf{H}_{t+1,1}^\intercal\boldsymbol{\gamma}_{t+1}^0]_+ - (\vec{\mathbf{Y}}_{t:2}^\intercal, \mathbf{X}_t^\intercal)\boldsymbol{\theta}_t\} \\ \mathbf{X}_t\{Y_{t+1} + \sum_{\ell=1}^{t-1} Y_{\ell+1}\beta_{t+1,\ell}^0 + \mathbf{H}_{t+1,0}^\intercal\boldsymbol{\beta}_{t+1,t+1}^0 + [\mathbf{H}_{t+1,1}^\intercal\boldsymbol{\gamma}_{t+1}^0]_+ - (\vec{\mathbf{Y}}_{t:2}^\intercal, \mathbf{X}_t^\intercal)\boldsymbol{\theta}_t\} \end{bmatrix} = \mathbf{0},$$

$$\mathbb{E}\left[\mathbf{X}_1\{Y_2(1+\beta_{21}) + \mathbf{H}_{20}^\intercal\boldsymbol{\beta}_{22} + [\mathbf{H}_{21}^\intercal\boldsymbol{\gamma}_2]_+ - \mathbf{X}_1^\intercal\boldsymbol{\theta}_1\}\right] = \mathbf{0}.$$

With the above we can generalize our robust imputation-based semi-supervised Q-learning with the same three steps: (i) imputation, (ii) refitting, and (iii) projection to the unlabeled data for T stages as follows.

Step I: Imputation

We use the usual weakly parametric or non-parametric imputation models for the conditional mean functions $\{\mu_t(\cdot), \mu_{t't}(\cdot), t', t = 2, \ldots, T+1\}$, where $\mu_t(\vec{\mathbf{U}}) = \mathbb{E}(Y_t|\vec{\mathbf{U}})$ and $\mu_{t't}(\vec{\mathbf{U}}) = \mathbb{E}(Y_{t'}Y_t|\vec{\mathbf{U}})$ consist. Notice the products consist of at most two missing outcomes. As in the case when $T = 2$, we denote the corresponding estimated mean functions as $\{\widehat{m}_t(\cdot), \widehat{m}_{t't}(\cdot), t', t = 2, \ldots, T+1\}$ under the corresponding imputation models $\{m_t(\vec{\mathbf{U}}), m_{t't}(\vec{\mathbf{U}}), t', t = 2, \ldots, T+1\}$.

Step II: Refitting

The refitting ensures the validity of the SSL estimators under potential mis-specifications of the imputation models and helps to control for overfitting bias. We update the imputation model by expanding it to include linear effects of $\{\mathbf{X}_t, t = 1, \ldots, T\}$ with cross-fitting.

In this case, the final imputation models for $\{Y_t, Y_{t'}Y_t, t = 2, \ldots, T+1\}$, denoted by $\{\bar{\mu}_t(\vec{\mathbf{U}}), \bar{\mu}_{t't}\}$, need to satisfy

$$\mathbb{E}\left[\vec{\mathbf{X}}_{t'}\{Y_t - \bar{\mu}_t(\vec{\mathbf{U}})\}\right] = \mathbf{0}, \quad t' = 1, \ldots, T, t = 2, \ldots, t'+1,$$
$$\mathbb{E}\left\{Y_{t'}Y_t - \bar{\mu}_{t't}(\vec{\mathbf{U}})\right\} = 0, \quad t' = 2, \ldots, T, t = 2, \ldots, T+1,$$

where feature vector $\mathbf{X}_t'$ has one intercept entry of 1, for $t' = 1, \ldots, T$.

Next we expand $\{m_t(\vec{\mathbf{U}}), m_{t't}(\vec{\mathbf{U}})\}$. Using the same notation as in the main paper for the K random equal sized partitions of the labeled index set $\{1, ..., n\}$, and using $\{\widehat{m}_t^{(-k)}(\vec{\mathbf{U}}), \widehat{m}_{t't}^{(-k)}(\vec{\mathbf{U}})\}$ for the counterpart of $\{\widehat{m}_t(\vec{\mathbf{U}}), \widehat{m}_{t't}(\vec{\mathbf{U}})\}$ with labeled observations in $\{1, .., n\} \setminus \mathcal{I}_k$ we obtain $\widehat{\boldsymbol{\eta}}_t$, and $\widehat{\eta}_{t't}$ respectively as the solutions to

$$\sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \vec{\mathbf{X}}_{t'i} \left\{ Y_{ti} - \widehat{m}_t^{(-k)}(\vec{\mathbf{U}}_i) - \boldsymbol{\eta}_t^{\mathsf{T}} \vec{\mathbf{X}}_i \right\} = \mathbf{0}, t' = 1, \ldots, T, t = 2, \ldots, t' + 1,$$

$$\sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \left\{ Y_{t'i} Y_{ti} - \widehat{m}_{t't}^{(-k)}(\vec{\mathbf{U}}_i) - \eta_{t't} \right\} = 0, \quad t' = 2, \ldots, T, t = 2, \ldots, T + 1.$$

Finally, we impute Y_t, and $Y_{t'}Y_t$ respectively as $\widehat{\mu}_t(\vec{\mathbf{U}}) = K^{-1} \sum_{k=1}^{K} \widehat{m}_t^{(-k)}(\vec{\mathbf{U}}) + \widehat{\boldsymbol{\eta}}_t^{\mathsf{T}} \vec{\mathbf{X}}$, and $\widehat{\mu}_{t't}(\vec{\mathbf{U}}) = K^{-1} \sum_{k=1}^{K} \widehat{m}_{t't}^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_{t't}$.

Step III: Projection

In the last step, we proceed to estimate $\widehat{\boldsymbol{\theta}}$ by replacing $\{Y_t, Y_{t'}Y_t\}$ in the Q function normal equations with their the imputed values $\{\widehat{\mu}_t(\vec{\mathbf{U}}), \widehat{\mu}_{t't}(\vec{\mathbf{U}})\}$ and project to the unlabeled data. For this we define vectors $\vec{\widehat{\mu}}_{t:2}(\vec{\mathbf{U}}) = [\widehat{\mu}_t(\vec{\mathbf{U}}), \widehat{\mu}_{t-1}(\vec{\mathbf{U}}), \ldots, \widehat{\mu}_2(\vec{\mathbf{U}})]^{\mathsf{T}}$, and $\vec{\widehat{\mu}}_{t',t:2}(\vec{\mathbf{U}}) = [\widehat{\mu}_{t't}(\vec{\mathbf{U}}), \widehat{\mu}_{t',t-1}(\vec{\mathbf{U}}), \ldots, \widehat{\mu}_{t'2}(\vec{\mathbf{U}})]^{\mathsf{T}}$.

We obtain the final SSL estimators for $\boldsymbol{\theta}_t$, $t = T, \ldots, 1$ with the following regressions (We omit the dependency of the imputed models on $\vec{\mathbf{U}}$ for notation brevity).

1. We get stage T regression parameters $\boldsymbol{\theta}_T$ from:

$$\mathbb{P}_N \begin{bmatrix} \widehat{\mu}_{T,T+1} - [\vec{\widehat{\mu}}_{T,T:2}^{\mathsf{T}}, \widehat{\mu}_T \mathbf{X}_T^{\mathsf{T}}] \widehat{\boldsymbol{\theta}}_T \\ \widehat{\mu}_{T-1,T+1} - [\vec{\widehat{\mu}}_{T-1,T:2}^{\mathsf{T}}, \widehat{\mu}_{T-1} \mathbf{X}_T^{\mathsf{T}}] \widehat{\boldsymbol{\theta}}_T \\ \vdots \\ \widehat{\mu}_{2,T+1} - [\vec{\widehat{\mu}}_{2,T:2}^{\mathsf{T}}, \widehat{\mu}_2 \mathbf{X}_T^{\mathsf{T}}] \widehat{\boldsymbol{\theta}}_T \\ \mathbf{X}_T \{ \widehat{\mu}_{T+1} - [\vec{\widehat{\mu}}_{T:2}^{\mathsf{T}}, \mathbf{X}_T^{\mathsf{T}}] \widehat{\boldsymbol{\theta}}_T \} \end{bmatrix} = \mathbf{0}$$

2. For finding stage $t = T - 1, \ldots, 2$ regression parameters, $\boldsymbol{\theta}_t$, we use:

$$\mathbb{P}_N \begin{bmatrix} \widehat{\mu}_{t,t+1} + \sum_{\ell=1}^{t-1} \widehat{\mu}_{t,\ell+1}\widehat{\beta}_{t+1,\ell} + \widehat{\mu}_t\{\mathbf{H}_{t+1,0}^\intercal \widehat{\boldsymbol{\beta}}_{t+1,t+1} + [\mathbf{H}_{t+1,1}^\intercal \widehat{\boldsymbol{\gamma}}_{t+1}]_+\} \\ \widehat{\mu}_{t-1,t+1} + \sum_{\ell=1}^{t-1} \widehat{\mu}_{t-1,\ell+1}\widehat{\beta}_{t+1,\ell} + \widehat{\mu}_{t-1}\{\mathbf{H}_{t+1,0}^\intercal \widehat{\boldsymbol{\beta}}_{t+1,t+1} + [\mathbf{H}_{t+1,1}^\intercal \widehat{\boldsymbol{\gamma}}_{t+1}]_+\} \\ \vdots \\ \widehat{\mu}_{2,t+1} + \sum_{\ell=1}^{t-1} \widehat{\mu}_{2,\ell+1}\widehat{\beta}_{t+1,\ell} + \widehat{\mu}_2\{\mathbf{H}_{t+1,0}^\intercal \widehat{\boldsymbol{\beta}}_{t+1,t+1} + [\mathbf{H}_{t+1,1}^\intercal \widehat{\boldsymbol{\gamma}}_{t+1}]_+\} \\ \mathbf{X}_t\{\widehat{\mu}_{t+1} + \sum_{\ell=1}^{t-1} \widehat{\mu}_{\ell+1}\widehat{\beta}_{t+1,\ell} + \mathbf{H}_{t+1,0}^\intercal \widehat{\boldsymbol{\beta}}_{t+1,t+1} + [\mathbf{H}_{t+1,1}^\intercal \widehat{\boldsymbol{\gamma}}_{t+1}]_+\} \end{bmatrix}$$

$$-\mathbb{P}_N \begin{bmatrix} [\vec{\widehat{\mu}}_{t,t:2}^\intercal, \widehat{\mu}_t\mathbf{X}_t^\intercal]\widehat{\boldsymbol{\theta}}_t\} \\ [\vec{\widehat{\mu}}_{t-1,t:2}^\intercal, \widehat{\mu}_{t-1}\mathbf{X}_t^\intercal]\widehat{\boldsymbol{\theta}}_t\} \\ \vdots \\ [\vec{\widehat{\mu}}_{2,t:2}^\intercal, \widehat{\mu}_2\mathbf{X}_t^\intercal]\widehat{\boldsymbol{\theta}}_t\} \\ \mathbf{X}_t[\vec{\widehat{\mu}}_{t:2}^\intercal, \mathbf{X}_t^\intercal]\widehat{\boldsymbol{\theta}}_t \end{bmatrix} = \mathbf{0}.$$

3. Finally the first stage parameters $\boldsymbol{\theta}_1$ are found using

$$\mathbb{P}_N \left[\mathbf{X}_1\{\widehat{\mu}_2(1+\widehat{\beta}_{21}) + \mathbf{H}_{20}^\intercal\widehat{\boldsymbol{\beta}}_{22} + [\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2]_+ - \mathbf{X}_1^\intercal\boldsymbol{\theta}_1\}\right] = \mathbf{0}.$$

Next we go over a simpler $T = 3$ scenario to highlight how the conditional mean models needed to be imputed are only functions of at most 2 missing outcomes. The three stage SSL Q-learning functions can be defined recursively as:

$$\begin{aligned} Q_3(\check{\mathbf{H}}_3, A_3) &\equiv \mathbb{E}[Y_4|\check{\mathbf{H}}_3, A_3] \\ Q_2(\check{\mathbf{H}}_2, A_2) &\equiv \mathbb{E}[Y_3 + \max_{a_3} Q_3(\check{\mathbf{H}}_3, a_3)|\check{\mathbf{H}}_2, A_2] \\ Q_1(\check{\mathbf{H}}_1, A_1) &\equiv \mathbb{E}[Y_2 + \max_{a_2} Q_2(\check{\mathbf{H}}_2, a_2)|\check{\mathbf{H}}_1, A_1] \end{aligned}$$

The corresponding working linear models for the Q functions with parameters $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t^\intercal, \boldsymbol{\gamma}_t^\intercal)^\intercal$, $t = 1, 2, 3$ are:

$$\begin{aligned} Q_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\theta}_1^0) =& \check{\mathbf{X}}_1^\intercal\boldsymbol{\theta}_1^0 = \mathbf{H}_{10}^\intercal\boldsymbol{\beta}_1^0 + A_1(\mathbf{H}_{11}^\intercal\boldsymbol{\gamma}_1^0), \\ Q_2(\check{\mathbf{H}}_2, A_2; \boldsymbol{\theta}_2^0) =& \check{\mathbf{X}}_2^\intercal\boldsymbol{\theta}_2^0 = Y_2\beta_{21}^0 + \mathbf{H}_{20}^\intercal\boldsymbol{\beta}_{22}^0 + A_2(\mathbf{H}_{21}^\intercal\boldsymbol{\gamma}_2^0), \\ Q_3(\check{\mathbf{H}}_3, A_3; \boldsymbol{\theta}_3^0) =& \check{\mathbf{X}}_3^\intercal\boldsymbol{\theta}_3^0 = Y_3\beta_{31}^0 + Y_2\beta_{32}^0 + \mathbf{H}_{30}^\intercal\boldsymbol{\beta}_{33}^0 + A_3(\mathbf{H}_{31}^\intercal\boldsymbol{\gamma}_3^0). \end{aligned}$$

The pseudo-outcomes for stages 1 and 2 under the linear Q function are

$$\begin{aligned} \bar{Y}_{2i}^* =& Y_{2i}(1+\beta_{21}) + \mathbf{H}_{20i}^\intercal\boldsymbol{\beta}_{22} + [\mathbf{H}_{21i}^\intercal\boldsymbol{\gamma}_2]_+, \\ \bar{Y}_{3i}^* =& Y_{3i}(1+\beta_{31}) + Y_{2i}\beta_{32} + \mathbf{H}_{30i}^\intercal\boldsymbol{\beta}_{33} + [\mathbf{H}_{31i}^\intercal\boldsymbol{\gamma}_3]_+. \end{aligned}$$

We expand the normal equations so that the outcomes Y_t, $t = 2, 3, 4$ and their products are explicit:

$$\mathbb{E}\left[\check{\mathbf{X}}_3(Y_4 - \check{\mathbf{X}}_3^\intercal\boldsymbol{\theta}_3)\right] = \mathbb{E}\begin{bmatrix} Y_3\{Y_4 - (Y_3, Y_2, \mathbf{X}_3^\intercal)\boldsymbol{\theta}_3\} \\ Y_2\{Y_4 - (Y_3, Y_2, \mathbf{X}_3^\intercal)\boldsymbol{\theta}_3\} \\ \mathbf{X}_3\{Y_4 - (Y_3, Y_2, \mathbf{X}_3^\intercal)\boldsymbol{\theta}_3\} \end{bmatrix} = \mathbf{0},$$

$$\mathbb{E}\left[\check{\mathbf{X}}_2(\bar{Y}_3^* - \check{\mathbf{X}}_2^\intercal\boldsymbol{\theta}_2)\right] = \mathbb{E}\begin{bmatrix} Y_2\{Y_3(1+\beta_{31}) + Y_2\beta_{32} + \mathbf{H}_{30}^\intercal\boldsymbol{\beta}_{33} + [\mathbf{H}_{31}^\intercal\boldsymbol{\gamma}_3]_+ - (Y_2, \mathbf{X}_2^\intercal)\boldsymbol{\theta}_2\} \\ \mathbf{X}_2\{Y_3(1+\beta_{31}) + Y_2\beta_{32} + \mathbf{H}_{30}^\intercal\boldsymbol{\beta}_{33} + [\mathbf{H}_{31}^\intercal\boldsymbol{\gamma}_3]_+ - (Y_2, \mathbf{X}_2^\intercal)\boldsymbol{\theta}_2\} \end{bmatrix} = \mathbf{0},$$

$$\mathbb{E}\left[\check{\mathbf{X}}_1(\bar{Y}_2^* - \check{\mathbf{X}}_1^\intercal\boldsymbol{\theta}_1)\right] = \mathbb{E}\left[\mathbf{X}_1\{Y_2(1+\beta_{21}) + \mathbf{H}_{20}^\intercal\boldsymbol{\beta}_{22} + [\mathbf{H}_{21}^\intercal\boldsymbol{\gamma}_2]_+ - \mathbf{X}_1^\intercal\boldsymbol{\theta}_1\}\right] = \mathbf{0}.$$

From the three-stage normal equations above, it is clear that all the imputations needed will be similar to the two-stage case: linear, pairwise products and squares of the missing outcomes. Next, we go over the generalization of our off-policy value function algorithm.

SSL Value Function Estimation

We further extend the notation of our inverse probability weights by using $\omega_0(\check{\mathbf{H}}_0, A_0; \boldsymbol{\Theta}) \equiv 1$ for all $\check{\mathbf{H}}_0, A_0$ and defining for $t = 1, \ldots, T$:

$$\omega_t(\check{\mathbf{H}}_t, A_t, \boldsymbol{\Theta}) \equiv \omega_{t-1}(\check{\mathbf{H}}_{t-1}, A_{t-1}, \boldsymbol{\Theta})\left(\frac{d_t(\mathbf{H}_t;\boldsymbol{\theta}_t)A_t}{\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)} + \frac{\{1-d_t(\mathbf{H}_t;\boldsymbol{\theta}_t)\}\{1-A_t\}}{1-\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)}\right).$$

The corresponding doubly-robust supervised value function estimator is $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}} = \mathbb{P}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}})\right\}$, where

$$\begin{aligned}\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}}) =& Q_1^o(\mathbf{H}_1; \widehat{\boldsymbol{\theta}}_1) \\ &+ \sum_{t=1}^{T} \omega_t(\check{\mathbf{H}}_t, A_t, \widehat{\boldsymbol{\Theta}})\left[Y_{t+1} - \left\{Q_t^o(\check{\mathbf{H}}_t, \widehat{\boldsymbol{\theta}}_t) - Q_{t+1}^o(\mathbf{H}_{t+1}, \widehat{\boldsymbol{\theta}}_{t+1})\right\}\right].\end{aligned}$$

As with two time steps, we define $Q_{t-}^o(\mathbf{H}_t; \boldsymbol{\theta}_t) = \mathbf{H}_{t0}^\intercal\boldsymbol{\beta}_{tt} + [\mathbf{H}_{t,1}^\intercal\boldsymbol{\gamma}_t]_+$ for $t = 1, \ldots, T$. Next, to simplify our next expression we let $\widehat{\beta}_{T+1,\ell} \equiv 0$ for $\ell = 1, \ldots, T$. Then we can re-write $\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}})$ as:

$$\begin{aligned}\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}}) =& Q_1^o(\mathbf{H}_1; \widehat{\boldsymbol{\theta}}_1) + \\ &+ \sum_{t=1}^{T} \omega_t(\check{\mathbf{H}}_t, A_t, \widehat{\boldsymbol{\Theta}})\left[(1+\widehat{\beta}_{t+1,t})Y_{t+1} + \sum_{\ell=1}^{t-1} Y_{\ell+1}(\widehat{\beta}_{t+1,\ell} - \widehat{\beta}_{t,\ell})\right] \\ &+ \sum_{t=1}^{T} \omega_t(\check{\mathbf{H}}_t, A_t, \widehat{\boldsymbol{\Theta}})\left[Q_{t+1-}^o(\mathbf{H}_{t+1}; \widehat{\boldsymbol{\theta}}_{t+1}) - Q_{t-}^o(\mathbf{H}_t; \widehat{\boldsymbol{\theta}}_t)\right].\end{aligned}$$

Note that as with the case when $t = 2$ there are only three types of functions of the outcome to impute, hence we define similar conditional mean functions:

$$\mu_2^v(\vec{\mathbf{U}}) \equiv \mathbb{E}[Y_2|\vec{\mathbf{U}}], \quad \mu_{\omega_t}^v(\vec{\mathbf{U}}) \equiv \mathbb{E}[\omega_t(\check{\mathbf{H}}_t, A_t; \bar{\boldsymbol{\Theta}})|\vec{\mathbf{U}}], \quad \mu_{t'\omega_t}^v(\vec{\mathbf{U}}) \equiv \mathbb{E}[Y_{t'}\omega_t(\check{\mathbf{H}}_t, A_t; \bar{\boldsymbol{\Theta}})|\vec{\mathbf{U}}].$$

Next we give a summary of the algorithm for the semi-supervised doubly robust value function estimation:

Step I: Imputation

As in the two-stage setting, we fit flexible, weakly parametric, or non-parametric models to the labeled data to approximate the conditional mean functions defined above. We also denote the respective imputation models as $\{m_2(\vec{\mathbf{U}}), m_{\omega_t}(\vec{\mathbf{U}}), m_{t'\omega_t}(\vec{\mathbf{U}})\}$ and their fitted values as $\{\widehat{m}_2(\vec{\mathbf{U}}), \widehat{m}_{\omega_t}(\vec{\mathbf{U}}), \widehat{m}_{t'\omega_t}(\vec{\mathbf{U}})\}$.

Step II: Refitting

Analogous to the two time step case our mis-specification and finite sample bias-corrected imputation models are $\{\bar{\mu}_2^v(\vec{\mathbf{U}}) = m_2(\vec{\mathbf{U}}) + \eta_2^v, \bar{\mu}_{\omega_t}^v(\vec{\mathbf{U}}) = m_{\omega_t}(\vec{\mathbf{U}}) + \eta_{\omega_t}^v, \bar{\mu}_{t'\omega_t}^v(\vec{\mathbf{U}}) = m_{t'\omega_t}(\vec{\mathbf{U}}) + \eta_{t'\omega_t}^v\}$. The conditional mean functions terms are of the same 3 forms as the $T = 2$ case, so we write similar constraints:

$$\begin{aligned}
\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \bar{\mathbf{\Theta}})\left\{Y_2 - \bar{\mu}_2^v(\vec{\mathbf{U}})\right\}\right] &= 0\,, \\
\mathbb{E}\left[\left\{Q_{t+1-}^o(\mathbf{H}_{t+1}; \bar{\boldsymbol{\theta}}_{t+1}) - Q_{t-}^o(\mathbf{H}_t; \bar{\boldsymbol{\theta}}_t)\right\}\left\{\omega_t(\check{\mathbf{H}}_t, A_t; \bar{\mathbf{\Theta}}) - \bar{\mu}_{\omega_t}^v(\vec{\mathbf{U}})\right\}\right] &= 0,\ t = 2, \ldots, T. \\
\mathbb{E}\left[\omega_t(\check{\mathbf{H}}_t, A_t; \bar{\mathbf{\Theta}}) Y_{t'} - \bar{\mu}_{t'\omega_t}^v(\vec{\mathbf{U}})\right] &= 0,\ t, t' = 2, \ldots, T+1.
\end{aligned}$$

To estimate η_2^v $\eta_{\omega_t}^v$, and $\eta_{t'\omega_t}^v$ under these constraints, we again employ cross-fitting using the following estimating equations

$$\begin{aligned}
\sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \omega_1(\check{\mathbf{H}}_{1i}, A_{1i}; \widehat{\mathbf{\Theta}})\left\{Y_2 - \widehat{m}_2^{(-k)}(\vec{\mathbf{U}}_i) - \widehat{\eta}_2^v\right\} &= 0, \\
\sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \left\{Q_{t+1-}^o(\mathbf{H}_{t+1}; \widehat{\boldsymbol{\theta}}_{t+1}) - Q_{t-}^o(\mathbf{H}_t; \widehat{\boldsymbol{\theta}}_t)\right\}\left\{\omega_t(\check{\mathbf{H}}_{ti}, A_{ti}; \widehat{\mathbf{\Theta}}) - \widehat{m}_{\omega_t}^{(-k)}(\vec{\mathbf{U}}_i) - \widehat{\eta}_{\omega_t}^v\right\} &= 0, \\
\sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \left\{\omega_t(\check{\mathbf{H}}_{ti}, A_{ti}; \widehat{\mathbf{\Theta}}) Y_{t'i} - \widehat{m}_{t'\omega_t}^{(-k)}(\vec{\mathbf{U}}_i) - \widehat{\eta}_{t'\omega_t}^v\right\} &= 0,
\end{aligned}$$

from the above we obtain $\widehat{\eta}_2^v$ $\widehat{\eta}_{\omega_t}^v$, and $\widehat{\eta}_{t'\omega_t}^v$and use these to construct our imputation functions:

$$\widehat{\mu}_2^v(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^{K} \widehat{m}_2^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_2^v, \quad \widehat{\mu}_{\omega_t}^v(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^{K} \widehat{m}_{\omega_t}(\vec{\mathbf{U}}) + \widehat{\eta}_{\omega_t}^v, \text{ for } t = 2, \ldots, T,$$

and

$$\widehat{\mu}_{t'\omega_t}^v(\vec{\mathbf{U}}) = K^{-1}\sum_{k=1}^{K} \widehat{m}_{t'\omega_t}^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_{t'\omega_t}^v, \text{ for } t, t' = 2, \ldots, T+1.$$

Step III: Semi-supervised augmented value function estimator.

Finally, our semi-supervised augmented estimator of the value of the policy $\bar{V}$ is

$$\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}} = \mathbb{P}_N \left\{ \mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}; \widehat{\boldsymbol{\Theta}}, \widehat{\mu}) \right\},$$

where $\widehat{\mathcal{V}}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}})$ is defined as:

$$\begin{aligned}\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}; \widehat{\boldsymbol{\Theta}}, \widehat{\mu}) =& Q_1^o(\mathbf{H}_1; \widehat{\boldsymbol{\theta}}_1) + \omega_1(\mathbf{H}_1, A_1, \widehat{\boldsymbol{\Theta}}) \left[(1 + \widehat{\beta}_{21}) \widehat{\mu}_2^v(\vec{\mathbf{U}}) + Q_{2-}^o(\mathbf{H}_2; \widehat{\boldsymbol{\theta}}_2) - Q_1^o(\mathbf{H}_1; \widehat{\boldsymbol{\theta}}_1) \right] \\ &+ \sum_{t=2}^{T} (1 + \widehat{\beta}_{t+1,t}) \widehat{\mu}_{t+1,\omega_t}^v(\vec{\mathbf{U}}) + \sum_{t=2}^{T} \sum_{\ell=1}^{t-1} \widehat{\mu}_{\ell+1,\omega_t}^v(\vec{\mathbf{U}}) (\widehat{\beta}_{t+1,\ell} - \widehat{\beta}_{t,\ell}) \\ &+ \sum_{t=2}^{T} \widehat{\mu}_{\omega_t}^v(\vec{\mathbf{U}}) \left[Q_{t+1-}^o(\mathbf{H}_{t+1}; \widehat{\boldsymbol{\theta}}_{t+1}) - Q_{t-}^o(\mathbf{H}_t; \widehat{\boldsymbol{\theta}}_t) \right].\end{aligned}$$

Next, we discuss whether the proposed method would work well for an arbitrary finite time horizon and where it might break down. Theoretically, the semi-supervised Q function and value function estimation will work as long as T is finite. However, as it is well known, we may face instability in our estimates during implementation. In the context of supervised doubly robust estimators for policy evaluation methods from (Jiang and Li, 2016b; Thomas and Brunskill, 2016a; Kallus and Uehara, 2020a) can be evaluated at the estimated optimal DTR to obtain the optimal value function. However, for large T, even these supervised estimators can be unstable as they depend on the products of the T inverse propensity weights. These become highly variable as T increases because they usually have nested products of such weights (Levine et al., 2020).

Regarding our theoretical results, as mentioned in the conclusion, there is nothing in the proofs explicitly requiring the timeline to be limited to two stages. We hypothesize that we can generalize the results to $T > 2$ using induction, as we have already proven the results for the first couple of stages. We chose to leave the theoretical results of this generalization for future work, as we believe no new or particularly interesting theoretical methodology is required for this extension, and the notation and results are already cumbersome in this more simple setting.

In the context of our semi-supervised method, as T grows, our Q- and value function estimators naturally require more terms to impute. We expect that imputing a higher number of terms will increase the variance. However, an advantage of our approach is that the conditional means to be imputed are, at most, products of two missing functions: outcomes and propensity scores, so their complexity does not increase with T. Most of the variability would come from the higher-order propensity scores, as is the case with the supervised methods previously mentioned.

Appendix C. Simulation Results for Alternative Settings

In this section we provide additional results for data generating scenarios described in Section 6. Tables C.1 and C.1 contain results for estimation of Q function parameters for the EHR simulation setting for small and large sample sizes respectively. Table C.3 contains

the complete parameter results for the continuous data generating setting for both small and large samples.

(a) $n = 135$ and $N = 1272$

	Supervised		Semi-Supervised									
			Random Forests					Basis Expansion				
Parameter	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
β_{11}=1.2	0.05	0.09	0.03	0.06	0.05	0.88	1.65	0.03	0.06	0.05	0.89	1.60
β_{12}=0	0.00	0.06	0.00	0.04	0.04	0.90	1.57	0.00	0.04	0.04	0.91	1.62
β_{13}=-0.4	0	0.07	-0.01	0.05	0.04	0.92	1.53	0	0.05	0.05	0.93	1.56
β_{14}=-0.3	0.00	0.07	-0.01	0.04	0.04	0.93	1.67	0	0.04	0.04	0.93	1.64
β_{15}=0	0.00	0.08	0.00	0.04	0.04	0.93	1.69	0.00	0.04	0.04	0.92	1.69
β_{16}=0	0	0.07	0.00	0.04	0.04	0.93	1.67	0.00	0.04	0.04	0.93	1.74
β_{17}=0	0.00	0.08	0.00	0.05	0.04	0.92	1.62	0.00	0.05	0.04	0.92	1.62
γ_{11}=0.1	-0.01	0.14	0.00	0.09	0.08	0.91	1.55	0	0.09	0.07	0.89	1.55
γ_{12}=0	-0.01	0.09	-0.01	0.06	0.05	0.92	1.53	-0.01	0.06	0.06	0.93	1.51
γ_{13}=0	0	0.08	0	0.05	0.05	0.93	1.58	0	0.05	0.05	0.94	1.58
γ_{14}=0	0	0.08	0.00	0.05	0.05	0.93	1.58	0	0.05	0.05	0.93	1.58
γ_{15}=0	0.00	0.09	0.00	0.05	0.05	0.92	1.59	0	0.05	0.05	0.95	1.65
γ_{16}=-0.1	0	0.09	0	0.06	0.05	0.92	1.52	0	0.06	0.05	0.93	1.49
β_{21}=0.1	0.00	0.10	-0.01	0.15	0.13	0.91	0.71	0	0.14	0.13	0.93	0.75
β_{22}=0.6	0	0.13	0.01	0.11	0.10	0.91	1.16	0	0.11	0.11	0.94	1.18
β_{23}=0	0.00	0.06	0.00	0.04	0.04	0.93	1.44	0.00	0.04	0.04	0.93	1.47
β_{24}=-0.2	0.00	0.06	0	0.05	0.04	0.89	1.16	0	0.05	0.05	0.93	1.20
β_{25}=-0.2	0.00	0.05	0	0.05	0.04	0.90	1.13	0	0.04	0.04	0.92	1.18
β_{26}=0	0.00	0.04	0.00	0.02	0.02	0.94	1.50	0.00	0.02	0.02	0.94	1.50
β_{27}=0	0.00	0.04	0.00	0.03	0.02	0.94	1.52	0.00	0.02	0.02	0.94	1.58
β_{28}=0	0.00	0.05	0.00	0.04	0.03	0.92	1.49	0.00	0.04	0.03	0.92	1.49
β_{29}=0	0	0.12	0.00	0.08	0.07	0.91	1.49	0.00	0.08	0.08	0.93	1.52
β_{210}=-0.2	0.00	0.11	0	0.07	0.07	0.94	1.54	0.00	0.07	0.07	0.94	1.57
β_{211}=-0.1	0.01	0.11	0.00	0.07	0.07	0.94	1.54	0.00	0.07	0.07	0.93	1.56
γ_{21}=0.1	0.01	0.16	0.01	0.11	0.10	0.92	1.47	0.01	0.11	0.10	0.94	1.51
γ_{22}=0	0.00	0.08	0.00	0.06	0.05	0.94	1.47	0.00	0.06	0.06	0.93	1.50
γ_{23}=0	0	0.08	0.00	0.06	0.05	0.94	1.45	0.00	0.05	0.05	0.94	1.48
γ_{24}=0	0	0.07	0.00	0.05	0.05	0.93	1.43	0.00	0.05	0.05	0.94	1.46
γ_{25}=0	0	0.07	0	0.05	0.05	0.94	1.48	0	0.05	0.05	0.94	1.48
γ_{26}=0	0	0.18	0	0.12	0.11	0.92	1.45	0	0.12	0.11	0.94	1.52
γ_{27}=-0.2	-0.01	0.16	-0.01	0.11	0.10	0.93	1.47	-0.01	0.11	0.10	0.94	1.48
γ_{28}=-0.1	-0.01	0.15	-0.01	0.10	0.10	0.94	1.54	-0.01	0.10	0.10	0.94	1.57

Table C.1: Bias, empirical standard error (ESE) of the supervised and the SSL estimators with either random forest imputation or basis expansion imputation strategies for $\bar{\boldsymbol{\theta}}$ when (a) $n = 135$ and $N = 1272$ under the EHR simulation setting. For the SSL estimators, we also obtain the average of the estimated standard errors (ASE) as well as the empirical coverage probabilities (CovP) of the 95% confidence intervals.

(b) $n = 500$ and $N = 10,000$

	Supervised		Semi-Supervised									
			Random Forests					Basis Expansion				
Parameter	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
β_{11}=1.2	0.01	0.05	0.00	0.02	0.02	0.91	2.09	0.00	0.02	0.02	0.92	2.00
β_{12}=0	0.00	0.03	0.00	0.01	0.01	0.91	2.07	0.00	0.01	0.01	0.92	2.07
β_{13}=-0.4	0.00	0.04	0	0.02	0.02	0.92	2.05	0	0.02	0.02	0.92	2.05
β_{14}=-0.3	0	0.04	0	0.02	0.01	0.92	2.06	0	0.02	0.02	0.92	2.06
β_{15}=0	0.00	0.04	0	0.02	0.02	0.94	2.18	0	0.02	0.02	0.94	2.06
β_{16}=0	0	0.04	0.00	0.02	0.02	0.94	2.18	0.00	0.02	0.02	0.94	2.18
β_{17}=0	0.00	0.04	0.00	0.02	0.02	0.93	2.06	0.00	0.02	0.02	0.94	2.06
γ_{11}=0.1	0	0.07	0	0.03	0.03	0.91	2.00	0	0.03	0.03	0.91	2.00
γ_{12}=0	-0.01	0.05	0	0.02	0.02	0.90	2.00	0	0.02	0.02	0.89	2.00
γ_{13}=0	0.00	0.04	0.00	0.02	0.02	0.92	2.00	0.00	0.02	0.02	0.91	1.90
γ_{14}=0	0	0.04	0.00	0.02	0.02	0.94	2.00	0.00	0.02	0.02	0.94	1.90
γ_{15}=0	0.00	0.04	0.00	0.02	0.02	0.94	2.16	0.00	0.02	0.02	0.94	2.05
γ_{16}=-0.1	0	0.04	0	0.02	0.02	0.93	2.05	0	0.02	0.02	0.92	1.95
β_{21}=0.1	0.00	0.05	0.00	0.04	0.04	0.95	1.16	0.00	0.04	0.05	0.96	1.13
β_{22}=0.6	0	0.07	0	0.04	0.04	0.95	1.74	0	0.04	0.04	0.96	1.69
β_{23}=0	0.00	0.03	0.00	0.01	0.01	0.94	1.87	0.00	0.01	0.01	0.94	1.87
β_{24}=-0.2	0.00	0.03	0.00	0.02	0.02	0.94	1.71	0.00	0.02	0.02	0.95	1.71
β_{25}=-0.2	0.00	0.02	0	0.01	0.01	0.94	1.60	0	0.01	0.01	0.95	1.60
β_{26}=0	0.00	0.02	0.00	0.01	0.01	0.92	1.90	0.00	0.01	0.01	0.93	1.90
β_{27}=0	0.00	0.02	0.00	0.01	0.01	0.94	1.89	0.00	0.01	0.01	0.94	1.89
β_{28}=0	0.00	0.03	0.00	0.01	0.01	0.94	1.92	0.00	0.01	0.01	0.94	1.92
β_{29}=0	0.00	0.06	0.00	0.03	0.03	0.92	1.94	0.00	0.03	0.03	0.93	1.88
β_{210}=-0.2	0	0.05	0	0.03	0.03	0.94	2.00	0.00	0.03	0.03	0.94	2.00
β_{211}=-0.1	0.00	0.06	0.00	0.03	0.03	0.94	2.00	0.00	0.03	0.03	0.94	2.00
γ_{21}=0.1	0	0.08	0.00	0.04	0.04	0.94	1.98	0.00	0.04	0.04	0.94	1.98
γ_{22}=0	0.00	0.04	0.00	0.02	0.02	0.93	1.95	0.00	0.02	0.02	0.93	1.86
γ_{23}=0	0	0.04	0	0.02	0.02	0.94	1.81	0	0.02	0.02	0.93	1.90
γ_{24}=0	0	0.03	0.00	0.02	0.02	0.94	1.83	0.00	0.02	0.02	0.95	1.83
γ_{25}=0	0	0.04	0	0.02	0.02	0.94	1.84	0	0.02	0.02	0.94	1.84
γ_{26}=0	-0.01	0.09	0	0.04	0.04	0.93	2.00	0	0.04	0.04	0.93	2.00
γ_{27}=-0.2	0.01	0.08	0.00	0.04	0.04	0.94	1.98	0.00	0.04	0.04	0.94	1.98
γ_{28}=-0.1	0.00	0.08	0.00	0.04	0.04	0.94	1.95	0.00	0.04	0.04	0.94	1.95

Table C.2: Bias, empirical standard error (ESE) of the supervised and the SSL estimators with either random forest imputation or basis expansion imputation strategies for $\bar{\boldsymbol{\theta}}$ when (b) $n = 500$ and $N = 10,000$ under the EHR simulation setting. For the SSL estimators, we also obtain the average of the estimated standard errors (ASE) as well as the empirical coverage probabilities (CovP) of the 95% confidence intervals.

Appendix D. Proof of Main Results

D.1 Semi-Supervised Q-Learning Asymptotics

In this section we first show the proofs for the theoretical results on the generalized semi-supervised Q-learning shown in Section 5.

(a) $n = 135$ and $N = 1272$

	Supervised		Semi-Supervised									
			Random Forests					Basis Expansion				
Parameter	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
β_{11}=4.9	0.04	0.34	0.01	0.22	0.18	0.91	1.58	0.01	0.20	0.17	0.90	1.70
β_{12}=1.1	-0.03	0.42	0.00	0.26	0.24	0.94	1.61	0.01	0.25	0.23	0.92	1.68
γ_{11}=1.4	-0.03	0.41	0.00	0.26	0.24	0.93	1.57	0.00	0.24	0.23	0.93	1.68
γ_{12}=-2.6	0.04	0.58	-0.01	0.36	0.34	0.94	1.61	-0.02	0.35	0.31	0.90	1.69
β_{21}=0.1	0.00	0.10	0.00	0.13	0.12	0.94	0.82	0.00	0.16	0.17	0.94	0.64
β_{22}=3	0.00	0.33	0.00	0.24	0.23	0.93	1.39	0	0.26	0.25	0.93	1.30
β_{23}=0	-0.01	0.34	-0.01	0.24	0.22	0.93	1.43	-0.01	0.24	0.24	0.94	1.39
β_{24}=0.1	0	0.43	0	0.29	0.28	0.94	1.49	0	0.30	0.29	0.94	1.46
β_{25}=-0.5	0.01	0.15	0	0.09	0.09	0.93	1.62	0.00	0.09	0.09	0.93	1.71
β_{26}=-0.4	0.03	0.48	0.01	0.37	0.35	0.93	1.29	0.01	0.41	0.40	0.94	1.16
γ_{21}=0.8	0.00	0.34	0.01	0.21	0.20	0.93	1.61	0.00	0.20	0.19	0.94	1.71
γ_{22}=0.2	-0.02	0.45	-0.01	0.28	0.28	0.95	1.60	-0.01	0.27	0.26	0.94	1.70
γ_{23}=0.5	0	0.18	0.01	0.11	0.11	0.94	1.59	0.00	0.11	0.11	0.94	1.68

(b) $n = 500$ and $N = 10,000$

	Supervised		Semi-Supervised									
			Random Forests					Basis Expansion				
Parameter	Bias	ESE	Bias	ESE	ASE	CovP	RE	Bias	ESE	ASE	CovP	RE
β_{11}=4.9	0.00	0.17	0	0.10	0.09	0.91	1.72	0	0.10	0.08	0.92	1.79
β_{12}=1.1	0	0.22	0.00	0.12	0.11	0.93	1.80	0.00	0.12	0.11	0.93	1.86
γ_{11}=1.4	0.01	0.22	0.01	0.12	0.11	0.92	1.76	0.01	0.12	0.11	0.92	1.80
γ_{12}=-2.6	0	0.29	0	0.17	0.16	0.93	1.73	-0.01	0.16	0.15	0.93	1.80
β_{21}=0.1	-0.01	0.05	0	0.05	0.05	0.94	1.06	0	0.07	0.08	0.95	0.74
β_{22}=3	0.00	0.17	0.00	0.11	0.10	0.93	1.60	0.00	0.12	0.11	0.94	1.45
β_{23}=0	0.00	0.17	0.00	0.10	0.10	0.95	1.66	0.00	0.11	0.11	0.95	1.54
β_{24}=0.1	0.02	0.23	0.01	0.13	0.12	0.94	1.77	0.01	0.14	0.13	0.94	1.68
β_{25}=-0.5	0.00	0.07	0.00	0.04	0.04	0.93	1.74	0.00	0.04	0.04	0.94	1.78
β_{26}=-0.4	-0.01	0.25	-0.01	0.17	0.15	0.93	1.51	-0.01	0.19	0.18	0.94	1.31
γ_{21}=0.8	0.00	0.17	0.00	0.10	0.09	0.93	1.80	0.00	0.09	0.09	0.93	1.86
γ_{22}=0.2	-0.01	0.23	0	0.13	0.12	0.93	1.81	0	0.13	0.12	0.94	1.83
γ_{23}=0.5	0.00	0.09	0.00	0.05	0.05	0.94	1.78	0.00	0.05	0.05	0.95	1.81

Table C.3: Bias, empirical standard error (ESE) of the supervised and the SSL estimators with either random forest imputation or basis expansion imputation strategies for $\bar{\boldsymbol{\theta}}$ when (a) $n = 135$ and $N = 1272$ and (b) $n = 500$ and $N = 10,000$ under the continuous outcome simulation setting. For the SSL estimators, we also obtain the average of the estimated standard errors (ASE) as well as the empirical coverage probabilities (CovP) of the 95% confidence intervals.

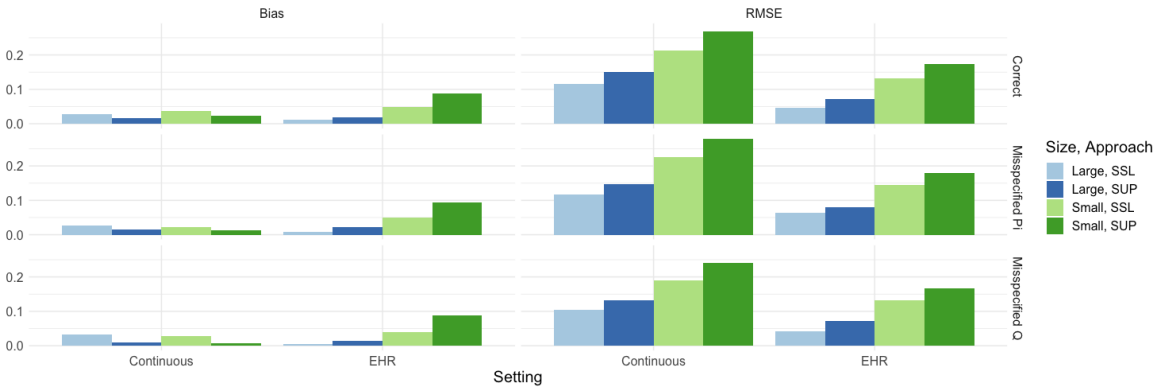

Figure C.1: Monte Carlo estimates for doubly-robust value function estimation: $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$, $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$ under continuous, and EHR settings. Columns show bias and RMSE respectively, rows show different mis-specification scenarios. Results are shown for the large ($N = 10,000$, $n = 500$) and small data samples ($N = 1,272$, $n = 135$) for the continuous setting over 1,000 simulated data sets.

D.1.1 Proofs of Theoretical Results for Q-Learning in Section 5

We first define $\boldsymbol{\theta}_{2-} \equiv (\boldsymbol{\beta}_{22}^{\intercal}, \boldsymbol{\gamma}_2^{\intercal})^{\intercal}$, and $\hat{\Delta}_s^{(-k)}(\vec{\mathbf{U}}) \equiv \widehat{m}_s^{(-k)}(\vec{\mathbf{U}}) - m_s(\vec{\mathbf{U}})$, $s \in \{2, 3, 22, 23\}$, and note that from Assumptions 1, 2 & 3 it follows that:

$$\begin{aligned}
&\sum_{k=1}^{K} \sup_{\vec{\mathbf{U}}} \left| \hat{\Delta}_{2t}^{(-k)}(\vec{\mathbf{U}}) \right| = o_{\mathbb{P}}(1) \text{ for } t = 2, 3, \\
&\sum_{k=1}^{K} \sup_{\vec{\mathbf{X}}, \vec{\mathbf{U}}} \| \vec{\mathbf{X}} \hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}) \| = o_{\mathbb{P}}(1), \\
&\sum_{k=1}^{K} \sup_{, \mathbf{X}_2 \vec{\mathbf{U}}} \| \mathbf{X}_2 \hat{\Delta}_3^{(-k)}(\vec{\mathbf{U}}) \| = o_{\mathbb{P}}(1),
\end{aligned} \tag{8}$$

Next we remind that, to ensure the validity of the SSL algorithm from the refitted imputation model, the final imputation models for $\{Y_t, Y_{2t}, t = 2, 3\}$, denoted by $\{\bar{\mu}_t(\vec{\mathbf{U}}), \bar{\mu}_{2t}, t = 2, 3\}$, need to satisfy the constraints shown in Section 3.2:

$$\begin{aligned}
\mathbb{E}\left[\vec{\mathbf{X}}\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\}\right] &= \mathbf{0}, \quad \mathbb{E}\left\{Y_2^2 - \bar{\mu}_{22}(\vec{\mathbf{U}})\right\} = 0, \\
\mathbb{E}\left[\mathbf{X}_2\{Y_3 - \bar{\mu}_3(\vec{\mathbf{U}})\}\right] &= \mathbf{0}, \quad \mathbb{E}\left\{Y_2 Y_3 - \bar{\mu}_{23}(\vec{\mathbf{U}})\right\} = 0.
\end{aligned} \tag{9}$$

where $\vec{\mathbf{X}} = (1, \mathbf{X}_1^{\intercal}, \mathbf{X}_2^{\intercal})^{\intercal}$.

Proof [Proof of Theorem 2]

Recall the estimating equation for stage 2 regression in Section 3.2 is

$$\mathbb{P}_N \begin{bmatrix} \widehat{\mu}_{23}(\vec{\mathbf{U}}) - \hat{\beta}_{21}\widehat{\mu}_{22}(\vec{\mathbf{U}}) - \widehat{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal\widehat{\boldsymbol{\theta}}_{2-} \\ \mathbf{X}_2\left\{\widehat{\mu}_3(\vec{\mathbf{U}}) - \hat{\beta}_{21}\widehat{\mu}_2(\vec{\mathbf{U}}) - \mathbf{X}_2^\intercal\widehat{\boldsymbol{\theta}}_{2-}\right\} \end{bmatrix} = \mathbf{0}.$$

Centering the above at $\bar{\boldsymbol{\theta}}_2$ we get

$$\mathbb{P}_N \begin{bmatrix} \widehat{\mu}_{22}(\vec{\mathbf{U}}), \widehat{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal \\ \mathbf{X}_2\widehat{\mu}_2(\vec{\mathbf{U}}), \mathbf{X}_2\mathbf{X}_2^\intercal \end{bmatrix} (\widehat{\boldsymbol{\theta}}_2 - \bar{\boldsymbol{\theta}}_2) = \mathbb{P}_N \begin{bmatrix} \widehat{\mu}_{23}(\vec{\mathbf{U}}) - \bar{\beta}_{21}\widehat{\mu}_{22}(\vec{\mathbf{U}}) - \widehat{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-} \\ \mathbf{X}_2\left\{\widehat{\mu}_3(\vec{\mathbf{U}}) - \bar{\beta}_{21}\widehat{\mu}_2(\vec{\mathbf{U}}) - \mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\right\} \end{bmatrix}. \tag{10}$$

Define

$$\begin{aligned}
\mathcal{R}_{\mathcal{U}} =& \mathbb{P}_N \begin{bmatrix} \bar{\mu}_{23}(\vec{\mathbf{U}}) - \bar{\beta}_{21}\bar{\mu}_{22}(\vec{\mathbf{U}}) - \bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-} \\ \mathbf{X}_2\left\{\bar{\mu}_3(\vec{\mathbf{U}}) - \bar{\beta}_{21}\bar{\mu}_2(\vec{\mathbf{U}}) - \mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\right\} \end{bmatrix}, \\
\hat{\mathcal{R}}_{\mathbb{S}}^{(K)} =& \mathbb{P}_N \begin{bmatrix} \left\{\widehat{\mu}_{23}(\vec{\mathbf{U}}) - \bar{\mu}_{23}(\vec{\mathbf{U}})\right\} - \bar{\beta}_{21}\left\{\widehat{\mu}_{22}(\vec{\mathbf{U}}) - \bar{\mu}_{22}(\vec{\mathbf{U}})\right\} - \left\{\widehat{\mu}_2(\vec{\mathbf{U}}) - \bar{\mu}_2(\vec{\mathbf{U}})\right\}\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-} \\ \mathbf{X}_2\left\{\widehat{\mu}_3(\vec{\mathbf{U}}) - \bar{\mu}_3(\vec{\mathbf{U}})\right\} - \bar{\beta}_{21}\mathbf{X}_2\left\{\widehat{\mu}_2(\vec{\mathbf{U}}) - \bar{\mu}_2(\vec{\mathbf{U}})\right\} \end{bmatrix}, \\
\Gamma_{\mathcal{U}} =& \mathbb{P}_N \begin{bmatrix} \bar{\mu}_{22}(\vec{\mathbf{U}}) & \bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal \\ \bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2 & \mathbf{X}_2\mathbf{X}_2^\intercal \end{bmatrix}, \\
\hat{\Gamma}_{\mathbb{S}}^{(K)} =& \mathbb{P}_N \begin{bmatrix} \widehat{\mu}_{22}(\vec{\mathbf{U}}) - \bar{\mu}_{22}(\vec{\mathbf{U}}) & \left\{\widehat{\mu}_2(\vec{\mathbf{U}}) - \bar{\mu}_2(\vec{\mathbf{U}})\right\}\mathbf{X}_2^\intercal \\ \left\{\widehat{\mu}_2(\vec{\mathbf{U}}) - \bar{\mu}_2(\vec{\mathbf{U}})\right\}\mathbf{X}_2 & \mathbf{0} \end{bmatrix},
\end{aligned}$$

with these we can re-write equation (10) as $\left(\Gamma_{\mathcal{U}} + \hat{\Gamma}_{\mathbb{S}}^{(K)}\right)(\widehat{\boldsymbol{\theta}}_2 - \bar{\boldsymbol{\theta}}_2) = \mathcal{R}_{\mathcal{U}} + \hat{\mathcal{R}}_{\mathbb{S}}^{(K)}$. We next deal with each term.

(I) We first consider $\hat{\mathcal{R}}_{\mathbb{S}}^{(K)}$, let

$$\begin{aligned}
\hat{\mathcal{S}}_{\mathbb{S}}^{\eta} =& \mathbb{P}_N \begin{bmatrix} (\widehat{\eta}_{23} - \eta_{23}) - \bar{\beta}_{21}(\widehat{\eta}_{22} - \eta_{22}) - (\widehat{\boldsymbol{\eta}}_2 - \boldsymbol{\eta}_2)^\intercal\mathbf{X}_2\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-} \\ \mathbf{X}_2\mathbf{X}_2^\intercal\left\{(\widehat{\boldsymbol{\eta}}_3 - \boldsymbol{\eta}_3) - \bar{\beta}_{21}(\widehat{\boldsymbol{\eta}}_2 - \boldsymbol{\eta}_2)\right\} \end{bmatrix} \\
\hat{\mathcal{S}}_{\mathbb{S}}^{(K)} =& \frac{1}{K}\sum_{k=1}^{K} \mathbb{P}_N \begin{bmatrix} \hat{\Delta}_{23}^{(-k)}(\vec{\mathbf{U}}) - \bar{\beta}_{21}\hat{\Delta}_{22}^{(-k)}(\vec{\mathbf{U}}) - \hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-} \\ \mathbf{X}_2\left\{\hat{\Delta}_3^{(-k)}(\vec{\mathbf{U}}) - \bar{\beta}_{21}\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}})\right\} \end{bmatrix} \\
\bar{\mathcal{S}}_k =& \mathbb{E}_{\mathcal{L}} \begin{bmatrix} \hat{\Delta}_{23}^{(-k)}(\vec{\mathbf{U}}) - \bar{\beta}_{21}\hat{\Delta}_{22}^{(-k)}(\vec{\mathbf{U}}) - \hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-} \\ \mathbf{X}_2\left\{\hat{\Delta}_3^{(-k)}(\vec{\mathbf{U}}) - \bar{\beta}_{21}\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}})\right\} \end{bmatrix} \quad \text{for } k \in \{1, \ldots, K\}.
\end{aligned}$$

From (3) it follows that $\hat{\mathcal{R}}_{\mathbb{S}}^{(K)} = \hat{\mathcal{S}}_{\mathbb{S}}^{\eta} + \hat{\mathcal{S}}_{\mathbb{S}}^{(K)}$. Next using (8), Assumption 2, and Lemma 15 it follows that $\hat{\mathcal{S}}_{\mathbb{S}}^{(K)} = \frac{1}{K}\sum_k \bar{\mathcal{S}}_k + O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)$, which lets us write $\hat{\mathcal{R}}_{\mathbb{S}}^{(K)} = \hat{\mathcal{S}}_{\mathbb{S}}^{\eta} + \frac{1}{K}\sum_k \bar{\mathcal{S}}_k + O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)$.

Now consider $\hat{\mathcal{S}}_{\mathbb{S}}^{\eta}$, note that by the central limit theorem (CLT) $\mathbb{P}_n\mathbf{X}_2\mathbf{X}_2 = \mathbb{E}\mathbf{X}_2\mathbf{X}_2 + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$. Thus using this, Slutsky's theorem and Assumption 1

$$(\mathbb{P}_n\mathbf{X}_2\mathbf{X}_2)^{-1}(\mathbb{P}_N\mathbf{X}_2\mathbf{X}_2) = I + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right),$$

then using (9), (3) and Assumption 2 we can write

$$
\begin{aligned}
&\mathbb{P}_N\left\{(\widehat{\boldsymbol{\eta}}_2-\boldsymbol{\eta}_2)^\intercal \mathbf{X}_2\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\right\}\\
=&\left[(\mathbb{P}_n\mathbf{X}_2\mathbf{X}_2^\intercal)^{-1}\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\mathbf{X}_{2i}\left\{Y_{2i}-\bar{\mu}_2(\vec{\mathbf{U}}_i)+m_2(\vec{\mathbf{U}}_i)-m_2^{(-k)}(\vec{\mathbf{U}}_i)\right\}\right]^\intercal\mathbb{P}_N(\mathbf{X}_2\mathbf{X}_2^\intercal)\boldsymbol{\theta}_{2-}\\
=&\left[\mathbb{P}_n\mathbf{X}_2^\intercal\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}+\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\mathbf{X}_{2i}^\intercal\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}_i)\right](\mathbb{P}_n\mathbf{X}_2\mathbf{X}_2^\intercal)^{-1}\mathbb{P}_N(\mathbf{X}_2\mathbf{X}_2^\intercal)\boldsymbol{\theta}_{2-}\\
=&\mathbb{P}_n\mathbf{X}_2\boldsymbol{\theta}_{2-}^\intercal\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}+\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\mathbf{X}_{2i}^\intercal\boldsymbol{\theta}_{2-}\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}_i)\\
+&O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\left[\mathbb{P}_n\mathbf{X}_2\boldsymbol{\theta}_{2-}^\intercal\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}+\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\mathbf{X}_{2i}^\intercal\boldsymbol{\theta}_{2-}\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}_i)\right]\\
=&\mathbb{P}_n\mathbf{X}_2\boldsymbol{\theta}_{2-}^\intercal\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}+\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\mathbf{X}_{2i}^\intercal\boldsymbol{\theta}_{2-}\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}_i)+O_{\mathbb{P}}\left(n^{-1}\right)+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)o_{\mathbb{P}}(1).
\end{aligned}
$$

Analogous derivations for all terms in $\hat{\mathcal{S}}_{\mathbb{S}}^{\eta}$ gives us

$$
\hat{\mathcal{S}}_{\mathbb{S}}^{\eta}=\mathbb{T}_{\mathcal{L}}-\mathbb{T}_{\mathcal{L}}^{(K)}+O_{\mathbb{P}}\left(n^{-1}\right)+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)o_{\mathbb{P}}(1),
$$

where

$$
\begin{aligned}
\mathbb{T}_{\mathcal{L}}=&\mathbb{P}_n\begin{bmatrix}\left\{Y_2Y_3-\bar{\mu}_{23}(\vec{\mathbf{U}})\right\}-\bar{\beta}_{21}\left\{Y_2^2-\bar{\mu}_{22}(\vec{\mathbf{U}})\right\}-\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\\ \mathbf{X}_2\left\{Y_3-\bar{\mu}_3(\vec{\mathbf{U}})\right\}-\bar{\beta}_{21}\mathbf{X}_2\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}\end{bmatrix},\\
\mathbb{T}_{\mathcal{L}}^{(K)}=&\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\begin{bmatrix}\hat{\Delta}_{23}^{(-k)}(\vec{\mathbf{U}}_i)-\bar{\beta}_{21}\hat{\Delta}_{22}^{(-k)}(\vec{\mathbf{U}}_i)-\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}_i)\mathbf{X}_{2i}^\intercal\bar{\boldsymbol{\theta}}_{2-}\\ \mathbf{X}_{2i}\left\{\hat{\Delta}_3^{(-k)}(\vec{\mathbf{U}}_i)-\bar{\beta}_{21}\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}_i)\right\}\end{bmatrix}.
\end{aligned}
$$

From the above it follows that $\hat{\mathcal{R}}_{\mathbb{S}}^{(K)}=\mathbb{T}_{\mathcal{L}}-\mathbb{T}_{\mathcal{L}}^{(K)}+\frac{1}{K}\sum_k\bar{\mathcal{S}}_k+O_{\mathbb{P}}\left(n^{-1}\right)+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)o_{\mathbb{P}}(1)$. Next by Assumption 2 and using Lemma 16 with $\hat{C}_{n,N}=1$, and setting functions $\hat{l}_n(\cdot)$, $\hat{\pi}_n(\cdot)$ to be the constant 1, and $f(\mathbf{X}_2)=\mathbf{X}_2$ to be the identity function, we have $\sqrt{n}\left(\mathbb{T}_{\mathcal{L}}^{(K)}-\frac{1}{K}\sum_k\bar{\mathcal{S}}_k\right)=O_{\mathbb{P}}\left(c_{n_K^-}\right)$. Therefore $\hat{\mathcal{R}}_{\mathbb{S}}^{(K)}=\mathbb{T}_{\mathcal{L}}+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}c_{n_K^-}\right)$.

(II) Now we consider $\mathcal{R}_{\mathcal{U}}$, from the CLT, assuming working model (1), as constraints (9) are satisfied it follows that

$$
\mathcal{R}_{\mathcal{U}}=\mathbb{E}\begin{bmatrix}\bar{\mu}_{23}(\vec{\mathbf{U}})-\bar{\beta}_{21}\bar{\mu}_{22}(\vec{\mathbf{U}})-\bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\\ \mathbf{X}_2\{\bar{\mu}_3(\vec{\mathbf{U}})-\bar{\beta}_{21}\bar{\mu}_2(\vec{\mathbf{U}})-\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\}\end{bmatrix}+O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)=\mathbf{1}O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right).
$$

(III) Next we focus on $\hat{\Gamma}_{\mathbb{S}}^{(K)}$, we use a similar expansion to (I) and define

$$\begin{aligned}
\hat{\mathcal{F}}_{\mathbb{S}}^{\eta} =& \begin{bmatrix} \widehat{\eta}_{22} - \eta_{22} & (\widehat{\eta}_2 - \eta_2)\mathbf{X}_2^\intercal \\ (\widehat{\eta}_2 - \eta_2)\,\mathbf{X}_2 & \mathbf{0} \end{bmatrix}, \\
\hat{\mathcal{F}}_{\mathbb{S}}^{(K)} =& \frac{1}{K}\sum_{k=1}^{K} \mathbb{P}_N \begin{bmatrix} \hat{\Delta}_{22}^{(-k)}(\vec{\mathbf{U}}) & \hat{\Delta}_{2}^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2^\intercal \\ \hat{\Delta}_{2}^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2 & \mathbf{0} \end{bmatrix}, \\
\bar{\mathcal{F}}_k =& \mathbb{E}_{\mathcal{L}} \begin{bmatrix} \hat{\Delta}_{22}^{(-k)}(\vec{\mathbf{U}}) & \hat{\Delta}_{2}^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2^\intercal \\ \hat{\Delta}_{2}^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2 & \mathbf{0} \end{bmatrix} \ \forall k \in \{1, \ldots, K\},
\end{aligned}$$

We argue as in (I), that from (3) it follows that $\hat{\Gamma}_{\mathbb{S}}^{(K)} = \hat{\mathcal{F}}_{\mathbb{S}}^{\eta} + \hat{\mathcal{F}}_{\mathbb{S}}^{(K)}$. Using (8), Assumptions 2 and Lemma 15 $\hat{\mathcal{F}}_{\mathbb{S}}^{(K)} - \frac{1}{K}\sum_k \bar{\mathcal{F}}_k = O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)$, therefore $\hat{\Gamma}_{\mathbb{S}}^{(K)} = \hat{\mathcal{F}}_{\mathbb{S}}^{\eta} + \frac{1}{K}\sum_k \bar{\mathcal{F}}_k + O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)$. Next we follow the same decomposition for $\hat{\mathcal{F}}_{\mathbb{S}}^{\eta}$ as we did in (I) for $\hat{\mathcal{S}}_{\mathbb{S}}^{\eta}$, it follows that

$$\begin{aligned}
\hat{\Gamma}_{\mathbb{S}}^{(K)} =& \mathbb{P}_n \begin{bmatrix} Y_2^2 - \bar{\mu}_{22}(\vec{\mathbf{U}}) & \left\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\right\}\mathbf{X}_2^\intercal \\ \left\{Y_2 - \bar{\mu}_2(\vec{\mathbf{U}})\right\}\mathbf{X}_2 & \mathbf{0} \end{bmatrix} \\
& - \frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} \begin{bmatrix} \hat{\Delta}_{22}^{(-k)}(\vec{\mathbf{U}}) & \hat{\Delta}_{2}^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2^\intercal \\ \hat{\Delta}_{2}^{(-k)}(\vec{\mathbf{U}})\mathbf{X}_2 & \mathbf{0} \end{bmatrix} + \frac{1}{K}\sum_k \bar{\mathcal{F}}_k + O_{\mathbb{P}}\left(n^{-1}\right) + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) o_{\mathbb{P}}(1).
\end{aligned}$$

The first term in the right hand side is $O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ by he CLT, the next two terms together are $O_{\mathbb{P}}\left(n^{-\frac{1}{2}} c_{n_K^-}\right)$ by Lemma 16, thus $\hat{\Gamma}_{\mathbb{S}}^{(K)} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}} c_{n_K^-}\right)$.

(IV) Finally we consider $\Gamma_{\mathcal{U}}$. By central limit theorem and (9) it follows that

$$\Gamma_{\mathcal{U}} = \mathbb{E}\begin{bmatrix} \bar{\mu}_{22}(\vec{\mathbf{U}}) & \bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal \\ \bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2 & \mathbf{X}_2\mathbf{X}_2^\intercal \end{bmatrix} + O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right) = \mathbb{E}[\check{\mathbf{X}}_2\check{\mathbf{X}}_2^\intercal] + O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right).$$

From (I)-(IV) we can write (10) as $(\widehat{\boldsymbol{\theta}}_2 - \bar{\boldsymbol{\theta}}_2) = \mathbb{E}\left[\check{\mathbf{X}}_2\check{\mathbf{X}}_2^\intercal\right]^{-1}\mathbb{T}_{\mathcal{L}} + O_{\mathbb{P}}\left(n^{-\frac{1}{2}} c_{n_K^-}\right)$, it follows that

$$\begin{aligned}
&\sqrt{n}(\widehat{\boldsymbol{\theta}}_2 - \bar{\boldsymbol{\theta}}_2) = \mathbb{E}\left[\check{\mathbf{X}}_2\check{\mathbf{X}}_2^\intercal\right]^{-1} \\
&\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \begin{bmatrix} \left\{Y_{2i}Y_{3i} - \bar{\mu}_{23}(\vec{\mathbf{U}}_i)\right\} - \bar{\beta}_{21}\left\{Y_{2i}^2 - \bar{\mu}_{22}(\vec{\mathbf{U}}_i)\right\} - \check{\mathbf{X}}_{2i}^\intercal\bar{\boldsymbol{\theta}}_{2-}\left\{Y_{2i} - \bar{\mu}_2(\vec{\mathbf{U}}_i)\right\} \\ \mathbf{X}_{2i}\left\{Y_{3i} - \bar{\mu}_3(\vec{\mathbf{U}}_i)\right\} - \bar{\beta}_{21}\mathbf{X}_{2i}\left\{Y_{2i} - \bar{\mu}_2(\vec{\mathbf{U}}_i)\right\} \end{bmatrix} \\
&+ o_{\mathbb{P}}(1).
\end{aligned}$$

■

Proof [Proof of Theorem 3] The solution to stage 1 estimating equation $\boldsymbol{\theta}_1$ in Section 3.2 satisfies

$$\mathbb{P}_N\left[\mathbf{X}_1\left\{\widehat{\mu}_2(\vec{\mathbf{U}}) + \hat{\beta}_{21}\widehat{\mu}_2(\vec{\mathbf{U}}) + \mathbf{H}_{20}^\intercal\widehat{\boldsymbol{\beta}}_{22} + [\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2]_+ - \mathbf{X}_1^\intercal\widehat{\boldsymbol{\theta}}_1\right\}\right] = \mathbf{0}.$$

We center the above at $\bar{\boldsymbol{\theta}}_1$ and get

$$\mathbb{P}_N\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right]\left(\widehat{\boldsymbol{\theta}}_1-\bar{\boldsymbol{\theta}}_1\right)=\mathbb{P}_N\left[\mathbf{X}_1\left\{\bar{\mu}_2(\vec{\mathbf{U}})+\hat{\beta}_{21}\bar{\mu}_2(\vec{\mathbf{U}})+\mathbf{H}_{20}^\intercal\widehat{\boldsymbol{\beta}}_{22}+\left[\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2\right]_+-\mathbf{X}_1^\intercal\bar{\boldsymbol{\theta}}_1\right\}\right]. \tag{11}$$

Next, with the following definitions

$$\begin{aligned}
\hat{\Sigma}_{\mathcal{U}}=&\mathbb{P}_N\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right],\quad \hat{\Sigma}_{\mathcal{L}}=\mathbb{P}_n\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right],\\
\mathcal{R}^{(1)}=&\mathbb{P}_N\left[\mathbf{X}_1\left\{\bar{\mu}_2(\vec{\mathbf{U}})+\hat{\beta}_{21}\bar{\mu}_2(\vec{\mathbf{U}})+\mathbf{H}_{20}^\intercal\widehat{\boldsymbol{\beta}}_{22}+\left[\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2\right]_+-\mathbf{X}_1^\intercal\bar{\boldsymbol{\theta}}_1\right\}\right],\\
\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}=&\mathbb{P}_N\left[\mathbf{X}_1\left\{\widehat{\mu}_2(\vec{\mathbf{U}})-\bar{\mu}_2(\vec{\mathbf{U}})\right\}\right],
\end{aligned}$$

we can write (11) as $\hat{\Sigma}_{\mathcal{U}}(\widehat{\boldsymbol{\theta}}_1-\bar{\boldsymbol{\theta}}_1)=\mathcal{R}^{(1)}+(1+\hat{\beta}_{21})\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}$. We now analyze both terms $\mathcal{R}^{(1)}$, and $(1+\hat{\beta}_{21})\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}$.

I) First we consider $(1+\hat{\beta}_{21})\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}$, define

$$\begin{aligned}
\hat{\mathcal{S}}_{\mathbb{S}}^{(1\eta)}=&\hat{\Sigma}_{\mathcal{U}}\left(\widehat{\boldsymbol{\eta}}_2-\boldsymbol{\eta}_2\right),\\
\hat{\mathcal{S}}_{\mathbb{S}}^{(1\mathbb{K})}=&\frac{1}{K}\sum_{k=1}^{K}\mathbb{P}_N\left[\mathbf{X}_1\hat{\Delta}_2^{(\text{-}k)}(\vec{\mathbf{U}})\right],\\
\bar{\mathcal{S}}_k^{(1)}=&\mathbb{E}\left[\mathbf{X}_1\hat{\Delta}_2^{(\text{-}k)}(\vec{\mathbf{U}})\right],
\end{aligned}$$

from (3) it follows that $\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}=\hat{\mathcal{S}}_{\mathbb{S}}^{(1\eta)}+\hat{\mathcal{S}}_{\mathbb{S}}^{(1\mathbb{K})}$, next from Assumptions 1, 2, we get $\sum_{k=1}^{K}\sup_{\mathbf{X}_1,\vec{\mathbf{U}}}\|\mathbf{X}_1\hat{\Delta}_2^{(\text{-}k)}(\vec{\mathbf{U}})\|=o_{\mathbb{P}}(1)$, thus by Lemma 15 $\hat{\mathcal{S}}_{\mathbb{S}}^{(1\mathbb{K})}=\bar{\mathcal{S}}_k^{(1)}+\left(N^{-\frac{1}{2}}\right)$. Using (3) again, and recalling $\bar{\mu}_2(\vec{\mathbf{U}})=m_2(\vec{\mathbf{U}})+\mathbf{X}_1^\intercal\boldsymbol{\eta}_2$ we have

$$\begin{aligned}
\hat{\mathcal{S}}_{\mathbb{S}}^{(1\eta)}=&\hat{\Sigma}_{\mathcal{U}}\hat{\Sigma}_{\mathcal{L}}^{-1}\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\mathbf{X}_{1i}\left\{Y_{2i}-\bar{\mu}_2(\vec{\mathbf{U}}_i)-\widehat{m}_2^{(\text{-}k)}(\vec{\mathbf{U}}_i)+m_2(\vec{\mathbf{U}}_i)\right\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{1i}\left\{Y_{2i}-\bar{\mu}_2(\vec{\mathbf{U}}_i)\right\}-\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\mathbf{X}_{1i}\hat{\Delta}_2^{(\text{-}k)}(\vec{\mathbf{U}}_i)+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where the last line follows by the CLT and Assumptions 1 and 2 as

$$\hat{\Sigma}_{\mathcal{U}}\hat{\Sigma}_{\mathcal{L}}^{-1}=I+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$$

Now using Lemma 15 and Assumptions 1, 2 again, it follows that

$$\hat{\mathcal{S}}_{\mathbb{S}}^{(1\mathbb{K})}=\bar{\mathcal{S}}_k^{(1)}+O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right),$$

combining the above we can write

$$\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}=\mathbb{P}_n\mathbf{X}_1\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}-\frac{1}{n}\sum_{k=1}^{K}\left\{\sum_{i\in\mathcal{I}_k}\mathbf{X}_{1i}\hat{\Delta}_2^{(\text{-}k)}(\vec{\mathbf{U}}_i)-\mathbb{E}\left[\mathbf{X}_1\hat{\Delta}_2^{(\text{-}k)}(\vec{\mathbf{U}})\right]\right\}+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).$$

Next by Assumption 2 and Lemma 16 we have

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{K}\left\{\sum_{i\in\mathcal{I}_k}\mathbf{X}_{1i}\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}}_i)-\mathbb{E}\left[\mathbf{X}_1\hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}})\right]\right\}=O_{\mathbb{P}}\left(c_{n_K^-}\right),$$

therefore $\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}=\mathbb{P}_n\mathbf{X}_1\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}c_{n_K^-}\right)$. Finally using Theorem 2 we have $\hat{\beta}_{21}-\bar{\beta}_{21}=O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, and by CLT $\mathbb{P}_n\mathbf{X}_1\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}=O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, thus we can write

$$(1+\hat{\beta}_{21})\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}=(1+\bar{\beta}_{21})\mathbb{P}_n\mathbf{X}_1\left\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\right\}+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}c_{n_K^-}\right).$$

II) Next we consider $\mathcal{R}^{(1)}$ by writing

$$\begin{aligned}\mathcal{R}^{(1)}=&\mathbb{P}_N\left[\mathbf{X}_1\left\{\bar{\mu}_2(\vec{\mathbf{U}})+\bar{\beta}_{21}\bar{\mu}_2(\vec{\mathbf{U}})+\mathbf{H}_{20}^{\intercal}\bar{\boldsymbol{\beta}}_{22}+[\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+-\mathbf{X}_1^{\intercal}\bar{\boldsymbol{\theta}}_1\right\}\right]\\&+\mathbb{P}_N\left[\mathbf{X}_1\left\{\bar{\mu}_2(\vec{\mathbf{U}})(\hat{\beta}_{21}-\bar{\beta}_{21})+\mathbf{H}_{20}^{\intercal}(\widehat{\boldsymbol{\beta}}_{22}-\bar{\boldsymbol{\beta}}_{22})+[\mathbf{H}_{21}^{\intercal}\widehat{\boldsymbol{\gamma}}_2]_+-[\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+\right\}\right],\end{aligned}$$

note that under (9) using model (1) the first term in the right hand side is mean zero, therefore from Assumption 1 and CLT

$$\mathbb{P}_N\left\{\mathbf{X}_1\left(\bar{\mu}_2(\vec{\mathbf{U}})+\bar{\boldsymbol{\beta}}_{21}\bar{\mu}_2(\vec{\mathbf{U}})+\mathbf{H}_{20}^{\intercal}\bar{\boldsymbol{\beta}}_{22}+[\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+-\mathbf{X}_1^{\intercal}\bar{\boldsymbol{\theta}}_1\right)\right\}=O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right).$$

Hence, we have

$$\begin{aligned}\sqrt{n}\mathcal{R}^{(1)}=&\sqrt{n}\mathbb{P}_N\left[\mathbf{X}_1\left\{\bar{\mu}_2(\vec{\mathbf{U}})(\hat{\beta}_{21}-\bar{\beta}_{21})+\mathbf{H}_{20}^{\intercal}(\widehat{\boldsymbol{\beta}}_{22}-\bar{\boldsymbol{\beta}}_{22})+[\mathbf{H}_{21}^{\intercal}\widehat{\boldsymbol{\gamma}}_2]_+-[\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+\right\}\right]\\&+O_{\mathbb{P}}\left(\sqrt{\frac{n}{N}}\right)\\=&\mathbb{P}_N\left[\mathbf{X}_1\left(\bar{\mu}_2(\vec{\mathbf{U}}),\mathbf{H}_{20}^{\intercal}\right)\right]\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_2-\bar{\boldsymbol{\beta}}_2\right)+\sqrt{n}\mathbb{P}_N\left[\mathbf{X}_1\left([\mathbf{H}_{21}^{\intercal}\widehat{\boldsymbol{\gamma}}_2]_+-[\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+\right)\right]\\&+O_{\mathbb{P}}\left(\sqrt{\frac{n}{N}}\right)\\=&\mathbb{E}\left[\mathbf{X}_1\left(\bar{\mu}_2(\vec{\mathbf{U}}),\mathbf{H}_{20}^{\intercal}\right)\right]n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\psi}_{2i\beta}+\sqrt{n}\mathbb{P}_N\left[\mathbf{X}_1\left([\mathbf{H}_{21}^{\intercal}\widehat{\boldsymbol{\gamma}}_2]_+-[\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+\right)\right]\\&+O_{\mathbb{P}}\left(\sqrt{\frac{n}{N}}\right),\end{aligned}$$

where the last inequality follows from the CLT, where $\boldsymbol{\psi}_{2i\beta}$ is the element corresponding to $\widehat{\boldsymbol{\beta}}_2$ of the influence function $\boldsymbol{\psi}_{2i}$ defined in Theorem 2.

Next by Theorem 2 we know that

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}}_2-\bar{\boldsymbol{\gamma}}_2)=O_{\mathbb{P}}(1),$$

using Lemma 17 (a) we have

$$\mathbb{P}\left[\sqrt{n}\mathbb{P}_N\left\{\mathbf{X}_1\left([\mathbf{H}_{21}^{\intercal}\widehat{\boldsymbol{\gamma}}_2]_+-[\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+\right)\right\}=\mathbb{P}_N\left\{\mathbf{X}_1\mathbf{H}_{21}^{\intercal}I\left(\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2>0\right)\right\}\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_2-\bar{\boldsymbol{\gamma}}_2\right)\right]\to 1.$$

Therefore, letting $\boldsymbol{\psi}_{2i\gamma}$ be the element corresponding to $\widehat{\boldsymbol{\gamma}}_2$ of the influence function $\boldsymbol{\psi}_{2i}$ defined in Theorem 2,

$$
\begin{aligned}
&\sqrt{n}\mathbb{P}_N\left\{\mathbf{X}_1\left([\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2]_+ - [\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2]_+\right)\right\}\\
&=\mathbb{P}_N\left\{\mathbf{X}_1\mathbf{H}_{21}^\intercal I(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0)I\left(\widehat{\boldsymbol{\gamma}}_2\in\mathcal{A}\right)\right\}\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_2-\bar{\boldsymbol{\gamma}}_2\right)\\
&+\sqrt{n}\mathbb{P}_N\left\{\mathbf{X}_1\left([\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2]_+ - [\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2]_+\right)\right\}I_{\{\widehat{\boldsymbol{\gamma}}_2\notin\mathcal{A}\}}\\
&=\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^\intercal|\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0,\widehat{\boldsymbol{\gamma}}_2\in\mathcal{A}\right]\mathbb{P}\left(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0\right)\mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_2\in\mathcal{A}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_{2\gamma_2 i}+O_{\mathbb{P}}\left(c_{n_K^-}\right)+o_{\mathbb{P}}(1)\\
&=\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^\intercal|\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0\right]\mathbb{P}\left(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_{2\gamma_2 i}+o_{\mathbb{P}}(1),
\end{aligned}
$$

combining all terms

$$
\begin{aligned}
\sqrt{n}\mathcal{R}^{(1)}=&\mathbb{E}\left[\mathbf{X}_1\left(\bar{\mu}_2(\vec{\mathbf{U}}),\mathbf{H}_{20}^\intercal\right)\right]n^{-\frac{1}{2}}\sum_{i=1}^n\boldsymbol{\psi}_{2i(\beta)}\\
&+\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^\intercal|\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0\right]\mathbb{P}\left(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\psi}_{2i(\gamma)}\\
&+O_{\mathbb{P}}\left(c_{n_K^-}\right).
\end{aligned}
$$

Finally, from I), II), and since $\hat{\Sigma}_{\mathcal{U}}^{-1}=\mathbb{E}\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right]^{-1}+o_{\mathbb{P}}(1)$ by the LLN, we have

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\theta}}_1-\bar{\boldsymbol{\theta}}_1)=&\mathbb{E}\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right]^{-1}\hat{\Sigma}_{\mathcal{U}}^{-1}\mathcal{R}^{(1)}+\mathbb{E}\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right]^{-1}(1+\hat{\beta}_{21})\hat{\mathcal{R}}_{\mathbb{S}}^{(1K)}+o_{\mathbb{P}}(1)\\
=&\mathbb{E}\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right]^{-1}(1+\bar{\beta}_{21})\frac{1}{\sqrt{n}}\sum_{i=1}^n\{Y_{2i}-\bar{\mu}_2(\vec{\mathbf{U}}_i)\}\\
&+\mathbb{E}\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right]^{-1}\mathbb{E}\left[\mathbf{X}_1\left(\bar{\mu}_2(\vec{\mathbf{U}}),\mathbf{H}_{20}^\intercal\right)\right]\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\psi}_{2i(\beta)}\\
&+\mathbb{E}\left[\mathbf{X}_1\mathbf{X}_1^\intercal\right]^{-1}\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^\intercal|\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0\right]\mathbb{P}\left(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2>0\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\psi}_{2i\gamma}\\
&+o_{\mathbb{P}}(1),
\end{aligned}
$$

using (9) we have $\mathbb{E}\left[\mathbf{X}_1\left(\bar{\mu}_2(\vec{\mathbf{U}}),\mathbf{H}_{20}^\intercal\right)\right]=\mathbb{E}\left[\mathbf{X}_1\left(Y_2,\mathbf{H}_{20}^\intercal\right)\right]$ which yields our required results ■

Next we discuss some results and assumptions needed for Proposition 5. First we show the asymptotic results for the supervised estimation of the Q function parameters. Recall $\widehat{\boldsymbol{\theta}}_{1\mathsf{SUP}}$, $\widehat{\boldsymbol{\theta}}_{2\mathsf{SUP}}$ are the estimators for the Q function parameters, when using the labeled data $\mathcal{L}$ only. From Laber et al. (2014) we have that the following results for $\widehat{\boldsymbol{\theta}}_{2\mathsf{SUP}}$:

$$
\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{2\mathsf{SUP}}-\bar{\boldsymbol{\theta}}_2\right)=\Sigma_2^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)\to\mathcal{N}\left(\mathbf{0},\mathbf{V}_{2\mathsf{SUP}}\left[\bar{\boldsymbol{\theta}}_2\right]\right),
$$

with

$$\begin{aligned}\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2) &= \check{\mathbf{X}}_2\{Y_{3i} - \check{\mathbf{X}}_{2i}^\intercal\bar{\boldsymbol{\theta}}_2\},\\ \mathbf{V}_{2\mathsf{SUP}}\left[\bar{\boldsymbol{\theta}}_2\right] &= \boldsymbol{\Sigma}_2^{-1}\mathbb{E}\left[\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)^\intercal\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^\intercal,\end{aligned}$$

and for $\widehat{\boldsymbol{\theta}}_{1\mathsf{SUP}}$:

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{1\mathsf{SUP}} - \bar{\boldsymbol{\theta}}_1\right) = \Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}_{1\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_1) \to \mathcal{N}\left(\mathbf{0}, \mathbf{V}_{1\mathsf{SUP}}\left[\bar{\boldsymbol{\theta}}_1\right]\right),$$

with

$$\begin{aligned}\boldsymbol{\psi}_{1\mathsf{SUP}}(\mathbf{L}_i;\bar{\boldsymbol{\theta}}_2) =& \mathbf{X}_{i1}\{Y_{2i} + Y_{2i}\bar{\beta}_{21} + \mathbf{H}_{20i}^\intercal\bar{\boldsymbol{\beta}}_{22} + [\mathbf{H}_{21i}^\intercal\bar{\boldsymbol{\gamma}}_2]_+ - \mathbf{X}_{1i}^\intercal\bar{\boldsymbol{\theta}}_1\}\\ &+\mathbb{E}\left[\mathbf{X}_1\left(Y_2, \mathbf{H}_{20}^\intercal\right)\right]\boldsymbol{\psi}_{2\mathsf{SUP},(\beta)}(\mathbf{L}_i)\\ &+\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^\intercal|\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2 > 0\right]\mathbb{P}\left(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2 > 0\right)\boldsymbol{\psi}_{2\mathsf{SUP},(\gamma)}(\mathbf{L}_i),\\ \mathbf{V}_{1\mathsf{SUP}}\left[\bar{\boldsymbol{\theta}}_1\right] =& \boldsymbol{\Sigma}_1^{-1}\mathbb{E}\left[\boldsymbol{\psi}_{1\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_1)\boldsymbol{\psi}_{1\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_1)^\intercal\right]\left(\boldsymbol{\Sigma}_1^{-1}\right)^\intercal.\end{aligned}$$

Next we discuss the assumption required for Proposition 5. We need the imputation models $\bar{\mu}_s(\vec{\mathbf{U}})$, $s \in \{2,3,22,23\}$ to satisfy several additional constraints. For example, for the stage two Q function parameters, recall $\boldsymbol{\theta}_{2-} = (\boldsymbol{\beta}_{22}^\intercal, \boldsymbol{\gamma}_2^\intercal)^\intercal$, the imputation models should satisfy:

$$\mathbb{E}\left[\vec{\mathbf{X}}^\intercal\bar{\mu}_j(\vec{\mathbf{U}})\{g_s(\mathbf{Y}) - \bar{\mu}_s(\vec{\mathbf{U}})\}\right] = \mathbf{0} \quad \mathbb{E}\left[\vec{\mathbf{X}}^\intercal\bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\{g_s(\mathbf{Y}) - \bar{\mu}_s(\vec{\mathbf{U}})\}\right] = \mathbf{0},$$

for $s, j \in \{2,3,22,23\}$, and

$$\mathbb{E}\left[\vec{\mathbf{X}}\vec{\mathbf{X}}^\intercal\bar{\mu}_j(\vec{\mathbf{U}})\{g_s(\mathbf{Y}) - \bar{\mu}_s(\vec{\mathbf{U}})\}\right] = \mathbf{0}, \quad \mathbb{E}\left[\vec{\mathbf{X}}\vec{\mathbf{X}}^\intercal\mathbf{X}_2^\intercal\bar{\boldsymbol{\theta}}_{2-}\{g_s(\mathbf{Y}) - \bar{\mu}_s(\vec{\mathbf{U}})\}\right] = \mathbf{0},$$

for $s, j \in \{2,3\}$, where $\mathbf{X} = (1, \mathbf{X}_1^\intercal, \mathbf{X}_2^\intercal)^\intercal$, $g_2(\mathbf{Y}) = Y_2$, $g_3(\mathbf{Y}) = Y_3$, $g_{22}(\mathbf{Y}) = Y_2^2$, $g_{23}(\mathbf{Y}) = Y_2Y_3$.

To summarize all the assumptions needed, we define the following functions:

$$\begin{aligned}\mathcal{E}^\theta(\vec{\mathbf{U}}) \equiv& \left\{\mathcal{E}_1(\vec{\mathbf{U}})^\intercal, \mathcal{E}_2(\vec{\mathbf{U}})^\intercal\right\}^\intercal,\\ \mathcal{E}_2(\vec{\mathbf{U}}) \equiv& \begin{bmatrix}\bar{\mu}_{23}(\vec{\mathbf{U}}) - [\bar{\mu}_{22}(\vec{\mathbf{U}}), \bar{\mu}_2(\vec{\mathbf{U}})\mathbf{X}_2^\intercal]\bar{\boldsymbol{\theta}}_2\\ \mathbf{X}_2\{\bar{\mu}_3(\vec{\mathbf{U}}) - [\bar{\mu}_2(\vec{\mathbf{U}}), \mathbf{X}_2^\intercal]\bar{\boldsymbol{\theta}}_2\}\end{bmatrix},\\ \mathcal{E}_1(\vec{\mathbf{U}}) \equiv& \mathbf{X}_1\{\bar{\mu}_2(\vec{\mathbf{U}})(1+\bar{\beta}_{21}) + Q_{2-}^o(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2) - \mathbf{X}_1^\intercal\bar{\boldsymbol{\theta}}_1\}\\ &+\mathbb{E}\left[\mathbf{X}_1\left(Y_2, \mathbf{H}_{20}^\intercal\right)\right]\mathcal{E}_{2\beta}(\vec{\mathbf{U}})\\ &+\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^\intercal|\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2 > 0\right]\mathbb{P}\left(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2 > 0\right)\mathcal{E}_{2\gamma}(\vec{\mathbf{U}}),\end{aligned} \tag{12}$$

where $\mathcal{E}_{2\beta}(\vec{\mathbf{U}})$, $\mathcal{E}_{2\gamma}(\vec{\mathbf{U}})$ are the elements corresponding to $\bar{\boldsymbol{\beta}}_2$, $\bar{\boldsymbol{\gamma}}_2$ of $\mathcal{E}_2(\vec{\mathbf{U}})$. Now we can succinctly summarize the constraints, by having $\bar{\mu}_s(\vec{\mathbf{U}})$, $s \in \{2,3,22,23\}$ satisfy

$$\mathbb{E}\left[\left\{\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2) - \mathcal{E}_2(\vec{\mathbf{U}})\right\}\mathcal{E}_2(\vec{\mathbf{U}})^\intercal\right] = \mathbf{0}.$$

This is condensed in the following assumption.

Assumption 7 *Let* $\mathcal{E}^{\theta}(\vec{\mathbf{U}})$ *be as defined in* (12)*, and*

$$\boldsymbol{\psi}_{\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}) = \left[\boldsymbol{\psi}_{1\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_1)^{\intercal}, \boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)^{\intercal}\right]^{\intercal},$$

the imputation models $\bar{\mu}_s(\vec{\mathbf{U}})$*,* $s \in \{2,3,22,23\}$ *satisfy*

$$\mathbb{E}\left[\left\{\boldsymbol{\psi}_{\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}) - \mathcal{E}^{\theta}(\vec{\mathbf{U}})\right\}\mathcal{E}^{\theta}(\vec{\mathbf{U}})^{\intercal}\right] = \mathbf{0}.$$

Proof [Proof of Proposition 5]

We first show the result is true for $\mathbf{V}_{2\mathsf{SSL}}\left[\bar{\boldsymbol{\theta}}_2\right]$. To simplify algebra, we denote the influence function from Theorem 2 as $\boldsymbol{\psi}_{2\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)$. Using the influence function of $\widehat{\boldsymbol{\theta}}_{2\mathsf{SUP}}$ and Theorem 2 we have the following relationship:

$$\boldsymbol{\psi}_{2\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2) = \boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2) - \mathcal{E}_2(\vec{\mathbf{U}}).$$

Therefore

$$\begin{aligned}
\mathbf{V}_{2\mathsf{SSL}}\left(\bar{\boldsymbol{\theta}}_2\right) =& \boldsymbol{\Sigma}_2^{-1}\mathbb{E}\left[\boldsymbol{\psi}_{2\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)\boldsymbol{\psi}_{2\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)^{\intercal}\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^{\intercal} \\
=& \boldsymbol{\Sigma}_2^{-1}\mathbb{E}\left[\left\{\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2) - \mathcal{E}_2(\vec{\mathbf{U}})\right\}\left\{\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2) - \mathcal{E}_2(\vec{\mathbf{U}})\right\}^{\intercal}\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^{\intercal} \\
=& \boldsymbol{\Sigma}_2^{-1}\mathbb{E}\left[\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)^{\intercal}\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^{\intercal} \\
&+\boldsymbol{\Sigma}_2^{-1}\mathbb{E}\left[\mathcal{E}_2(\vec{\mathbf{U}})\mathcal{E}_2(\vec{\mathbf{U}})^{\intercal}\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^{\intercal} \\
&-2\boldsymbol{\Sigma}_2^{-1}\mathbb{E}\left[\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2)\mathcal{E}_2(\vec{\mathbf{U}})^{\intercal}\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^{\intercal}
\end{aligned}$$

Now, since our imputation models satisfy Assumption 7, it follows that

$$\mathbb{E}\left[\left\{\boldsymbol{\psi}_{2\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_2) - \mathcal{E}_2(\vec{\mathbf{U}})\right\}\mathcal{E}_2(\vec{\mathbf{U}})^{\intercal}\right] = \mathbf{0}.$$

Therefore we have

$$\mathbf{V}_{2\mathsf{SSL}}\left(\bar{\boldsymbol{\theta}}_2\right) = \mathbf{V}_{2\mathsf{SUP}}\left(\bar{\boldsymbol{\theta}}_2\right) - \boldsymbol{\Sigma}_2^{-1}\mathrm{Var}\left[\mathcal{E}_2(\vec{\mathbf{U}})\right]\left(\boldsymbol{\Sigma}_2^{-1}\right)^{\intercal}.$$

To show the result is true for $\mathbf{V}_{1\mathsf{SSL}}\left[\bar{\boldsymbol{\theta}}_1\right]$, We denote by $\mathcal{E}_{2\beta}(\vec{\mathbf{U}})$ and $\mathcal{E}_{2\gamma}(\vec{\mathbf{U}})$ the vectors corresponding to $\bar{\boldsymbol{\beta}}_2$, $\bar{\boldsymbol{\gamma}}_2$ in $\mathcal{E}_2(\vec{\mathbf{U}})$ respectively, and further recall the definition of $\mathcal{E}_1(\vec{\mathbf{U}})$:

$$\begin{aligned}
\mathcal{E}_1(\vec{\mathbf{U}}) =& \mathbf{X}_1\{\bar{\mu}_2(\vec{\mathbf{U}}) + \bar{\mu}_2(\vec{\mathbf{U}})\bar{\beta}_{21} + \mathbf{H}_{20}^{\intercal}\bar{\boldsymbol{\beta}}_{22} + [\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+ - \mathbf{X}_1^{\intercal}\bar{\boldsymbol{\theta}}_1\} \\
&+\mathbb{E}\left[\mathbf{X}_1\left(Y_2, \mathbf{H}_{20}^{\intercal}\right)\right]\mathcal{E}_{2\beta}(\vec{\mathbf{U}}) \\
&+\mathbb{E}\left[\mathbf{X}_1\mathbf{H}_{21}^{\intercal}|\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2 > 0\right]\mathbb{P}\left(\mathbf{H}_{21}^{\intercal}\bar{\boldsymbol{\gamma}}_2 > 0\right)\mathcal{E}_{2\gamma}(\vec{\mathbf{U}}).
\end{aligned}$$

From the form of the influence function of $\widehat{\boldsymbol{\theta}}_{1\mathsf{SUP}}$, and Theorems 2 & 3 we have that:

$$\boldsymbol{\psi}_{1\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\theta}}_1) = \boldsymbol{\psi}_{1\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\theta}}_1) - \mathcal{E}_1(\vec{\mathbf{U}}).$$

Analogous steps for the proof of $\bar{\boldsymbol{\theta}}_2$ can then be used to show

$$\mathbf{V}_{1\mathsf{SSL}}\left(\bar{\boldsymbol{\theta}}_1\right) = \mathbf{V}_{1\mathsf{SUP}}\left(\bar{\boldsymbol{\theta}}_1\right) - \boldsymbol{\Sigma}_1^{-1}\mathrm{Var}\left[\mathcal{E}_1(\vec{\mathbf{U}})\right]\left(\boldsymbol{\Sigma}_1^{-1}\right)^{\intercal}.$$

The required result is obtained by stacking the influence functions for $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ for the supervised and semi-supervised versions, noting that

$$\boldsymbol{\psi}_{\mathsf{SSL}}(\mathbf{L}; \bar{\boldsymbol{\theta}}) = \boldsymbol{\psi}_{\mathsf{SUP}}(\mathbf{L}; \bar{\boldsymbol{\theta}}) - \mathcal{E}^{\theta}(\vec{\mathbf{U}}).$$

and repeating the steps above. ■

D.2 Value Function Results

In this section we prove the main results for our SSL value function estimator. Before the proofs we go over some useful definitions, notation and lemmas. First recall that, in order to correct for potential biases arising from finite sample estimation and model misspecifications, the final imputed models for $\{Y_2, \omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}),\, Y_t\omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}),\, t = 2, 3\}$ satisfy the following constraints:

$$\begin{aligned}
\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}})\left\{Y_2 - \bar{\mu}_2^v(\vec{\mathbf{U}})\right\}\right] &= 0, \\
\mathbb{E}\left[Q_{2-}^o(\vec{\mathbf{U}}; \boldsymbol{\theta}_2)\left\{\omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}) - \bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}})\right\}\right] &= 0, \\
\mathbb{E}\left[\omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}})Y_t - \bar{\mu}_{t\omega_2}^v(\vec{\mathbf{U}})\right] &= 0,\ t = 2, 3.
\end{aligned} \tag{13}$$

Next, define the set

$$\mathcal{S}(\delta) = \Bigg\{(\boldsymbol{\theta}, \boldsymbol{\xi})\Bigg| \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 < \delta, \|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|_2^2 < \delta, \boldsymbol{\theta}_t \in \Theta_t, \boldsymbol{\xi}_t \in \Omega_t, t = 1, 2, \\ \pi_1(\mathbf{H}_1; \boldsymbol{\xi}_1) > 0,\ \pi_2(\check{\mathbf{H}}_2; \boldsymbol{\xi}_2) > 0,\ \forall\, \mathbf{H} \in \mathcal{H}\Bigg\}.$$

We will be using the influence functions for our model parameters $\boldsymbol{\Theta}$. In this regard let $\boldsymbol{\psi}^{\theta} = (\boldsymbol{\psi}_1^{\intercal}, \boldsymbol{\psi}_2^{\intercal})^{\intercal}$. By Theorems 2 & 3 $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\psi}^{\theta}(\vec{\mathbf{U}}_i) + o_{\mathbb{P}}(1)$. Next, from Assumption 5, it can be shown that $\widehat{\boldsymbol{\xi}}$ has the following expansion: $\sqrt{n}(\widehat{\boldsymbol{\xi}} - \bar{\boldsymbol{\xi}}) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\psi}^{\xi}\left(\mathbf{L}_i; \bar{\boldsymbol{\xi}}\right) + o_{\mathbb{P}}(1)$, where

$$\boldsymbol{\psi}_t^{\xi}\left(\mathbf{L}; \bar{\boldsymbol{\xi}}\right) = \mathbb{E}\left\{\check{\mathbf{H}}_t^{\intercal}\check{\mathbf{H}}_t\sigma\left(\check{\mathbf{H}}_t^{\intercal}\bar{\boldsymbol{\xi}}_t\right)\left[1 - \sigma\left(\check{\mathbf{H}}_t^{\intercal}\bar{\boldsymbol{\xi}}_t\right)\right]\right\}^{-1}\check{\mathbf{H}}_t\left\{A_t - \sigma\left(\check{\mathbf{H}}_t^{\intercal}\bar{\boldsymbol{\xi}}_t\right)\right\},\ t = 1, 2,$$

$\boldsymbol{\psi}^{\xi}\left(\mathbf{L}; \bar{\boldsymbol{\xi}}\right) = \left[\boldsymbol{\psi}_1^{\xi}\left(\mathbf{L}; \bar{\boldsymbol{\xi}}\right), \boldsymbol{\psi}_2^{\xi}\left(\mathbf{L}; \bar{\boldsymbol{\xi}}\right)\right]$ and $\mathbb{E}[\boldsymbol{\psi}^{\xi}] = 0$, $\mathbb{E}[(\boldsymbol{\psi}^{\xi})^{\intercal}\boldsymbol{\psi}^{\xi}] < \infty$.

We now introduce a set of definitions used in this section to make the proofs easier to read. Recall from (7) we have

$\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}} = \mathbb{P}_N\left\{\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}; \widehat{\boldsymbol{\Theta}}, \widehat{\mu})\right\}$, where $\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}; \widehat{\boldsymbol{\Theta}}, \widehat{\mu})$ is the semi-supervised augmented estimator for observation $\vec{\mathbf{U}}$, we re-write $\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}; \widehat{\boldsymbol{\Theta}}, \widehat{\mu})$ as $\mathcal{V}_{\widehat{\boldsymbol{\Theta}}, \widehat{\mu}}(\vec{\mathbf{U}})$ recall its definition,

and define the following functions:

$$\begin{aligned}\mathcal{V}_{\widehat{\boldsymbol{\Theta}},\widehat{\mu}}(\vec{\mathbf{U}}) \equiv & Q_1^o(\check{\mathbf{H}}_1;\widehat{\boldsymbol{\theta}}_1)+\omega_1(\check{\mathbf{H}}_1,A_1,\widehat{\boldsymbol{\Theta}})\left[(1+\widehat{\beta}_{21})\widehat{\mu}_2^v(\vec{\mathbf{U}})-Q_1^o(\check{\mathbf{H}}_1;\widehat{\boldsymbol{\theta}}_1)+Q_{2-}^o\{\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2\}\right]\\ &+\widehat{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})-\widehat{\beta}_{21}\widehat{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})-Q_{2-}^o(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2)\widehat{\mu}_{\omega_2}^v(\vec{\mathbf{U}}),\\ \mathcal{V}_{\bar{\boldsymbol{\Theta}},\widehat{\mu}}(\vec{\mathbf{U}}) \equiv & Q_1^o(\check{\mathbf{H}}_1;\bar{\boldsymbol{\theta}}_1)+\omega_1(\check{\mathbf{H}}_1,A_1,\bar{\boldsymbol{\Theta}})\left[(1+\bar{\beta}_{21})\widehat{\mu}_2^v(\vec{\mathbf{U}})-Q_1^o(\check{\mathbf{H}}_1;\bar{\boldsymbol{\theta}}_1)+Q_{2-}^o\{\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2\}\right]\\ &+\widehat{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})-\bar{\beta}_{21}\widehat{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})-Q_{2-}^o(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\widehat{\mu}_{\omega_2}^v(\vec{\mathbf{U}}).\end{aligned} \tag{14}$$

We next replace the estimated imputation functions with their limits $\bar{\mu}_2^v$, $\bar{\mu}_{2\omega_2}^v$, $\bar{\mu}_{3\omega_2}^v$ and $\bar{\mu}_{\omega_2}^v$, and define:

$$\begin{aligned}\mathcal{V}_{\widehat{\boldsymbol{\Theta}},\bar{\mu}}(\vec{\mathbf{U}}) \equiv & Q_1^o(\check{\mathbf{H}}_1;\widehat{\boldsymbol{\theta}}_1)+\omega_1(\check{\mathbf{H}}_1,A_1,\widehat{\boldsymbol{\Theta}})\left[(1+\widehat{\beta}_{21})\bar{\mu}_2^v(\vec{\mathbf{U}})-Q_1^o(\check{\mathbf{H}}_1;\widehat{\boldsymbol{\theta}}_1)+Q_{2-}^o(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2)\right]\\ &+\bar{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})-\hat{\beta}_{21}\bar{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})-Q_{2-}^o(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2)\bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}}),\\ \mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}(\vec{\mathbf{U}}) \equiv & Q_1^o(\check{\mathbf{H}}_1;\bar{\boldsymbol{\theta}}_1)+\omega_1(\check{\mathbf{H}}_1,A_1,\bar{\boldsymbol{\Theta}})\left[(1+\bar{\beta}_{21})\bar{\mu}_2^v(\vec{\mathbf{U}})-Q_1^o(\check{\mathbf{H}}_1;\bar{\boldsymbol{\theta}}_1)+Q_{2-}^o(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\right]\\ &+\bar{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})-\bar{\beta}_{21}\bar{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})-Q_{2-}^o(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}}).\end{aligned} \tag{15}$$

Finally we define the following functions which are weighted sums of the imputation function errors:

$$\begin{aligned}\mathcal{E}_{\widehat{\boldsymbol{\Theta}}}(\vec{\mathbf{U}}) \equiv & \omega_1(\check{\mathbf{H}}_1,A_1;\widehat{\boldsymbol{\Theta}})(1+\hat{\beta}_{21})\left\{\widehat{\mu}_2^v(\vec{\mathbf{U}})-\bar{\mu}_2^v(\vec{\mathbf{U}})\right\}+\widehat{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})-\bar{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})\\ &-\hat{\beta}_{21}\left\{\widehat{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})-\bar{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})\right\}-Q_{2-}^o\left(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2\right)\left\{\widehat{\mu}_{\omega_2}^v(\vec{\mathbf{U}})-\bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}})\right\},\\ \mathcal{E}_{\bar{\boldsymbol{\Theta}}}(\vec{\mathbf{U}}) \equiv & \omega_1(\check{\mathbf{H}}_1,A_1;\bar{\boldsymbol{\Theta}})(1+\bar{\beta}_{21})\left\{\widehat{\mu}_2^v(\vec{\mathbf{U}})-\bar{\mu}_2^v(\vec{\mathbf{U}})\right\}+\widehat{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})-\bar{\mu}_{3\omega_2}^v(\vec{\mathbf{U}})\\ &-\bar{\beta}_{21}\left\{\widehat{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})-\bar{\mu}_{2\omega_2}^v(\vec{\mathbf{U}})\right\}-Q_{2-}^o\left(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2\right)\left\{\widehat{\mu}_{\omega_2}^v(\vec{\mathbf{U}})-\bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}})\right\}.\end{aligned} \tag{16}$$

These definitions will come in handy in the following proofs as we can use them to write $\mathcal{V}_{\widehat{\boldsymbol{\Theta}},\widehat{\mu}}(\vec{\mathbf{U}})=\mathcal{V}_{\widehat{\boldsymbol{\Theta}},\bar{\mu}}(\vec{\mathbf{U}})+\mathcal{E}_{\widehat{\boldsymbol{\Theta}}}(\vec{\mathbf{U}})$, $\mathcal{V}_{\bar{\boldsymbol{\Theta}},\widehat{\mu}}(\vec{\mathbf{U}})=\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}(\vec{\mathbf{U}})+\mathcal{E}_{\bar{\boldsymbol{\Theta}}}(\vec{\mathbf{U}})$. Finally, recalling that $\mathbb{P}_{\vec{\mathbf{U}}}$ is the underlying distribution of the data, we define function $g_1:\boldsymbol{\Theta}\mapsto\mathbb{R}$ as

$$g_1(\boldsymbol{\Theta})=\int \mathcal{V}_{\boldsymbol{\Theta},\bar{\mu}}(\vec{\mathbf{U}})d\mathbb{P}_{\vec{\mathbf{U}}}.$$

With the above definitions we proceed by stating three lemmas that will be used to prove Theorem 7. We defer the proofs of these lemmas for after proving the main Theorem in this section.

Lemma 11 *Under Assumptions 1-6, we have*

I) $\sqrt{n}\left\{\mathbb{P}_N\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right]-g_1\left(\bar{\boldsymbol{\Theta}}\right)\right\}=o_{\mathbb{P}}(1)$,

II) $$\sqrt{n}\left\{g_1(\widehat{\boldsymbol{\Theta}})-g_1\left(\bar{\boldsymbol{\Theta}}\right)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\left(\frac{\partial}{\partial\boldsymbol{\theta}}g_1\left(\bar{\boldsymbol{\Theta}}\right)\right)^{\mathsf{T}}\boldsymbol{\psi}^{\theta}(\vec{\mathbf{U}}_i)+\left(\frac{\partial}{\partial\boldsymbol{\xi}}g_1\left(\bar{\boldsymbol{\Theta}}\right)\right)^{\mathsf{T}}\boldsymbol{\psi}^{\xi}(\vec{\mathbf{U}}_i)\right\}+o_{\mathbb{P}}(1).$$

Lemma 12 *Under Assumptions 1-6, the following holds:*

$$\sqrt{n}\left\{\left(\mathbb{P}_N\left[\mathcal{V}_{\widehat{\mathbf{\Theta}},\bar{\mu}}\right]-g_1(\widehat{\mathbf{\Theta}})\right)-\left(\mathbb{P}_N\left[\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}\right]-g_1(\bar{\mathbf{\Theta}})\right)\right\}=o_{\mathbb{P}}\left(1\right).$$

Lemma 13 *Under Assumptions 1-6, the following assertions hold:*

$$\begin{aligned}
&I) \quad \sqrt{n}\mathbb{P}_N\left\{\mathcal{E}_{\widehat{\mathbf{\Theta}}}-\mathcal{E}_{\bar{\mathbf{\Theta}}}\right\}=o_{\mathbb{P}}(1),\\
&II) \quad \sqrt{n}\mathbb{P}_N\left[\mathcal{E}_{\bar{\mathbf{\Theta}}}\right]=\mathbb{G}_n\left\{\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\}\\
&\qquad\qquad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\theta}(\mathbf{L}_i)^{\mathsf{T}}\frac{\partial}{\partial\boldsymbol{\theta}}\int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\mathbf{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}}\\
&\qquad\qquad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\xi}(\mathbf{L}_i)^{\mathsf{T}}\frac{\partial}{\partial\boldsymbol{\xi}}\int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\mathbf{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}}\\
&\qquad\qquad +o_{\mathbb{P}}(1).
\end{aligned}$$

Proof [Proof of Theorem 7] We start by expanding the expression in (7) and using definitions (14), (15), (16):

$$
\begin{aligned}
&\sqrt{n}\left\{\mathbb{P}_N\left[\mathcal{V}_{\widehat{\bar{\boldsymbol{\Theta}}},\widehat{\bar{\mu}}}\right]-\mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right]\right\}\\
=&\sqrt{n}\left\{\underbrace{\mathbb{P}_N\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right]+\mathbb{P}_N\left[\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right]}_{(I)}-\underbrace{g_1(\bar{\boldsymbol{\Theta}})-\mathbb{E}_{\mathbb{S}}\left[\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right]}_{(II)}\right\}\\
+&\sqrt{n}\left\{\left(\mathbb{P}_N\left[\mathcal{V}_{\widehat{\bar{\boldsymbol{\Theta}}},\bar{\mu}}\right]+\mathbb{P}_N\left[\mathcal{E}_{\widehat{\bar{\boldsymbol{\Theta}}}}\right]-\underbrace{g_1(\widehat{\bar{\boldsymbol{\Theta}}})}_{(III)}\right)-\mathbb{E}_{\mathbb{S}}\left[\mathcal{E}_{\widehat{\bar{\boldsymbol{\Theta}}}}\right]\right\}\\
-&\left\{\underbrace{\mathbb{P}_N\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right]+\mathbb{P}_N\left[\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right]}_{(I)}-\underbrace{g_1(\bar{\boldsymbol{\Theta}})-\mathbb{E}_{\mathbb{S}}\left[\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right]}_{(II)}\right\}\\
+&\sqrt{n}\left\{\underbrace{g_1(\widehat{\bar{\boldsymbol{\Theta}}})+\mathbb{E}_{\mathbb{S}}\left[\mathcal{E}_{\widehat{\bar{\boldsymbol{\Theta}}}}\right]}_{(III)}-\mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right]\right\}\\
=&\sqrt{n}\left\{\mathbb{P}_N\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right]-g_1(\bar{\boldsymbol{\Theta}})\right\}\\
+&\sqrt{n}\left\{g_1(\widehat{\bar{\boldsymbol{\Theta}}})-g_1(\bar{\boldsymbol{\Theta}})\right\}\\
+&\sqrt{n}\left\{\left(\mathbb{P}_N\left[\mathcal{V}_{\widehat{\bar{\boldsymbol{\Theta}}},\bar{\mu}}\right]-g_1(\widehat{\bar{\boldsymbol{\Theta}}})\right)-\left(\mathbb{P}_N\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right]-g_1(\bar{\boldsymbol{\Theta}})\right)\right\}\\
+&\sqrt{n}\mathbb{P}_N\left[\mathcal{E}_{\widehat{\bar{\boldsymbol{\Theta}}}}-\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right]\\
+&\sqrt{n}\mathbb{P}_N\left[\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{v}_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\boldsymbol{\Theta}})+o_{\mathbb{P}}(1).
\end{aligned}
$$

which follows from Lemmas 11, 12 & 13 with the influence function $\psi^{v}_{\mathsf{SSL}_{\mathsf{DR}}}$ defined as

$$
\begin{aligned}
\psi^{v}_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})=&\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})+\boldsymbol{\psi}^{\theta}(\mathbf{L})^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}\int\left\{\mathcal{V}_{\boldsymbol{\Theta},\bar{\mu}}(\mathbf{L})+\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})\right\}d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}}\\
&+\boldsymbol{\psi}^{\xi}(\mathbf{L})^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}\int\left\{\mathcal{V}_{\boldsymbol{\Theta},\bar{\mu}}(\vec{\mathbf{U}})+\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})\right\}d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}},\\
\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})=&\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\boldsymbol{\Theta}}_1)(1+\bar{\beta}_{21})\left\{Y_2-\bar{\mu}^{v}_2(\vec{\mathbf{U}})\right\}+\omega_2(\check{\mathbf{H}}_2,A_2,\bar{\boldsymbol{\Theta}}_2)Y_3-\bar{\mu}_{3\omega_2}(\vec{\mathbf{U}})\\
&-\bar{\beta}_{21}\left\{\omega_2(\check{\mathbf{H}}_2,A_2,\bar{\boldsymbol{\Theta}}_2)Y_2-\bar{\mu}_{2\omega_2}(\vec{\mathbf{U}})\right\}-Q^{o}_{2-}(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\left\{\omega_2(\check{\mathbf{H}}_2,A_2,\bar{\boldsymbol{\Theta}}_2)-\bar{\mu}_{\omega_2}(\vec{\mathbf{U}})\right\}
\end{aligned}
$$

Next note that

$$
\int\left(\mathcal{V}_{\boldsymbol{\Theta},\bar{\mu}}(\vec{\mathbf{U}})+\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})\right)d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}}=\int\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}},
$$

where $\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})$ is defined in (5). Finally, all random variables in the expression of $\psi^v_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})$ are bounded by Assumptions 1 and 5 we have $\mathbb{E}\left[\psi^v_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})^2\right] < \infty$, the central limit theorem yields that

$$\sqrt{n}\left\{\mathbb{P}_N\left[\mathcal{V}_{\widehat{\mathbf{\Theta}},\widehat{\mu}}\right] - g_1(\bar{\mathbf{\Theta}})\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^v_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\mathbf{\Theta}}) + o_{\mathbb{P}}(1) \xrightarrow{d} N\left(0, \sigma^2_{\mathsf{SSL}_{\mathsf{DR}}}\right).$$

■

Proof [Proof of Lemma 11] I) We start with $\sqrt{n}\left\{\mathbb{P}_N\left[\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}\right] - g_1(\bar{\mathbf{\Theta}})\right\}$. Note that $\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}(\vec{\mathbf{U}})$ is a deterministic function of random variable $\vec{\mathbf{U}}$ as parameters and imputation functions are fixed. We have that $\mathbb{E}\left[\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}(\vec{\mathbf{U}})^2\right] < \infty$ holds by Assumption 1 & 5. Thus the central limit theorem yields $\mathbb{G}_N\left\{\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}\right\} \xrightarrow{d} \mathcal{N}\left(0, Var\left[\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}\right]\right)$, therefore

$$\sqrt{n}\left\{\mathbb{P}_N\left[\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}\right] - g_1(\bar{\mathbf{\Theta}})\right\} = \sqrt{\frac{n}{N}}\mathbb{G}_N\left\{\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}\right\} = O_{\mathbb{P}}\left(\frac{\sqrt{n}}{N}\right) = o_{\mathbb{P}}(1).$$

II) We next consider $\sqrt{n}\left\{g_1(\widehat{\mathbf{\Theta}}) - g_1(\bar{\mathbf{\Theta}})\right\}$. Using a Taylor series expansion

$$g_1(\widehat{\mathbf{\Theta}}) = g_1(\bar{\mathbf{\Theta}}) + (\widehat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}g_1(\bar{\mathbf{\Theta}}) + (\widehat{\boldsymbol{\xi}} - \bar{\boldsymbol{\xi}})^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}g_1(\bar{\mathbf{\Theta}}) + O_{\mathbb{P}}\left(n^{-1}\right),$$

as both $\|\widehat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}\|_2^2 = O_{\mathbb{P}}\left(n^{-1}\right)$ and $\|\widehat{\boldsymbol{\xi}} - \bar{\boldsymbol{\xi}}\|_2^2 = O_{\mathbb{P}}\left(n^{-1}\right)$ by Theorems 2, 3 and Assumption 5, therefore

$$\sqrt{n}\left\{g_1(\widehat{\mathbf{\Theta}}) - g_1(\bar{\mathbf{\Theta}})\right\} = \sqrt{n}(\widehat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}g_1(\bar{\mathbf{\Theta}}) + \sqrt{n}(\widehat{\boldsymbol{\xi}} - \bar{\boldsymbol{\xi}})^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}g_1(\bar{\mathbf{\Theta}}) + o_{\mathbb{P}}(1).$$

We can write

$$\sqrt{n}\left\{g_1(\widehat{\mathbf{\Theta}}) - g_1(\bar{\mathbf{\Theta}})\right\} = \frac{\partial}{\partial\boldsymbol{\theta}}g_1(\bar{\mathbf{\Theta}})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\theta}(\vec{\mathbf{U}}_i) + \frac{\partial}{\partial\boldsymbol{\xi}}g_1(\bar{\mathbf{\Theta}})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\xi}(\vec{\mathbf{U}}_i) + o_{\mathbb{P}}(1).$$

■

Proof [Proof of Lemma 12]

We consider $\sqrt{n}\left\{\left(\mathbb{P}_N\left[\mathcal{V}_{\widehat{\mathbf{\Theta}},\bar{\mu}}\right] - g_1(\widehat{\mathbf{\Theta}})\right) - \left(\mathbb{P}_N\left[\mathcal{V}_{\bar{\mathbf{\Theta}},\bar{\mu}}\right] - g_1(\bar{\mathbf{\Theta}})\right)\right\}$, recall that $d_t(\check{\mathbf{H}}_t, \boldsymbol{\theta}_t) = I(\mathbf{H}_{t1}^{\intercal}\boldsymbol{\gamma}_t > 0)$ $t = 1, 2$, thus the inverse probability weight functions are defined as

$$\omega_1(\check{\mathbf{H}}_1, A_1, \mathbf{\Theta}) \equiv \frac{I(\mathbf{H}_{11}^{\intercal}\boldsymbol{\gamma}_1 > 0)A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{\{1 - I(\mathbf{H}_{11}^{\intercal}\boldsymbol{\gamma}_1 > 0)\}\{1 - A_1\}}{1 - \pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}, \quad \text{and}$$

$$\omega_2(\check{\mathbf{H}}_2, A_2, \mathbf{\Theta}) \equiv \omega_1(\check{\mathbf{H}}_1, A_1, \mathbf{\Theta})\left(\frac{I(\mathbf{H}_{21}^{\intercal}\boldsymbol{\gamma}_2 > 0)A_2}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} + \frac{\{1 - I(\mathbf{H}_{21}^{\intercal}\boldsymbol{\gamma}_2 > 0)\}\{1 - A_2\}}{1 - \pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right).$$

Define the class

$$\ell_t = \{I\left(\mathbf{H}_t^{\intercal}\boldsymbol{\gamma}_t \geq 0\right) : \mathcal{H}_{t1}, \boldsymbol{\gamma} \in \mathbb{R}^{q_t}\},\ t = 1, 2$$

and the collection of half spaces $\mathcal{C}_\ell \equiv \{\mathbf{H}_t \in \mathbb{R}^{q_t} : \mathbf{H}_t^{\intercal}\boldsymbol{\gamma}_t \geq 0, \boldsymbol{\gamma} \in \mathbb{R}^{q_t}, t \in \{1, 2\}\}$. By Dudley (1979) $\mathcal{C}_\ell$ is a VC class of VC dimension $q_t + 1$. Next by van der Vaart and Wellner (1996) we have that as $\mathcal{C}_\ell$ is a VC-class ℓ_t is a class of the same index. Finally, by Theorem 2.6.7 we have that ℓ_t is a $\mathbb{P}$-Donsker class. Next define the following function

$$\begin{aligned} f_{\boldsymbol{\Theta}}(\vec{\mathbf{U}}) =& Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1) + \omega_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta}) \left[(1 + \beta_{21})\bar{\mu}_2^v(\vec{\mathbf{U}}) - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1) + Q_{2-}^o(\mathbf{H}_2; \boldsymbol{\theta}_2)\right] \\ &+ \bar{\mu}_{3\omega_2}^v(\vec{\mathbf{U}}) - \beta_{21}\bar{\mu}_{2\omega_2}^v(\vec{\mathbf{U}}) - Q_{2-}^o(\mathbf{H}_2; \boldsymbol{\theta}_2)\bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}}). \end{aligned}$$

We define the associated class of functions $\mathcal{C}_1 = \left\{ f_{\boldsymbol{\Theta}}(\vec{\mathbf{U}}) | \vec{\mathbf{U}}, \boldsymbol{\Theta} \in \mathcal{S}(\delta) \right\}$.

i) By Assumptions 3, 5 and Theorem 19.5 in Vaart (1998), ℓ_t, $\mathcal{W}_t$, $\mathcal{Q}_t, t = 1, 2$ are $\mathbb{P}$-Donsker classes. Thus it follows that $\mathcal{C}_1$ is a Donsker class.

ii) We estimate $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ for (4) with their maximum likelihood estimators, $\widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2$, solving $\mathbb{P}_n\left[S_t(\boldsymbol{\xi}_t)\right] = \mathbf{0}, t = 1, 2$. By Assumption (5) and Theorem 5.9 in Vaart (1998) $\widehat{\boldsymbol{\xi}}_t \xrightarrow{p} \bar{\boldsymbol{\xi}}_t, t = 1, 2$. Next, by Theorems 2, 3, under Assumptions 1, 2, $\widehat{\boldsymbol{\theta}}_t \xrightarrow{p} \bar{\boldsymbol{\theta}}_t, t = 1, 2$. Thus $\mathbb{P}\left(\widehat{\boldsymbol{\Theta}} \in \mathcal{S}(\delta)\right) \to 1,\ \forall \delta$.

iii) We next show $\int \left(\mathcal{V}_{\widehat{\boldsymbol{\Theta}}, \bar{\mu}} - \mathcal{V}_{\bar{\boldsymbol{\Theta}}, \bar{\mu}}\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}} \longrightarrow 0$. By Assumptions 5 (ii), 6, and bounded covariates and there exists a constant $c \in \mathbb{R}$ such that we can write

$$\begin{aligned} &\int \left(\mathcal{V}_{\widehat{\boldsymbol{\Theta}}, \bar{\mu}} - \mathcal{V}_{\bar{\boldsymbol{\Theta}}, \bar{\mu}}\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}} \\ &\leq \int \left(Q_1^o(\mathbf{H}_1; \widehat{\boldsymbol{\theta}}_1) - Q_1^o(\mathbf{H}_1; \bar{\boldsymbol{\theta}}_1)\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}} \\ &+ c\int \left(\frac{1}{1 - \pi_1(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1)} - \frac{1}{1 - \pi_1(\mathbf{H}_1; \bar{\boldsymbol{\xi}}_1)}\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}} \\ &+ c\int \left(\frac{1}{\pi_1(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1)} - \frac{1}{\pi_1(\mathbf{H}_1; \bar{\boldsymbol{\xi}}_1)}\right)^2 \\ &+ c\int \left\{Q_{2-}^o(\check{\mathbf{H}}_2; \widehat{\boldsymbol{\theta}}_2) - Q_{2-}^o(\mathbf{H}_2; \bar{\boldsymbol{\theta}}_2)\right\}^2 d\mathbb{P}_{\vec{\mathbf{U}}} \\ &+ c\int \left\{I(\mathbf{H}_{11}^{\intercal}\widehat{\boldsymbol{\gamma}}_1 > 0) - I(\mathbf{H}_{11}^{\intercal}\bar{\boldsymbol{\gamma}}_1 > 0)\right\}^2 d\mathbb{P}_{\vec{\mathbf{U}}} \\ &+ \left(\hat{\beta}_{21} - \bar{\beta}_{21}\right)^2 \\ &+ c\int \left(\mathbf{H}_{20}^{\intercal}\bar{\boldsymbol{\beta}}_{22} + [\mathbf{H}_{20}^{\intercal}\bar{\boldsymbol{\gamma}}_2]_+ - \mathbf{H}_{20}^{\intercal}\widehat{\boldsymbol{\beta}}_{22} - [\mathbf{H}_{20}^{\intercal}\widehat{\boldsymbol{\gamma}}_2]_+\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}} \end{aligned}$$

where we use $(a - b)^2, (a + b)^2 \leq 2a^2 + 2b^2\ \forall a, b \in \mathbb{R}$, $\widehat{d}_1, A_1 \leq 1$ for all $\mathbf{H} \in \mathcal{H}$, and boundedness of $\widehat{\boldsymbol{\theta}}_t, t = 1, 2$ by Assumptions 1-3. Next note that all terms outside integrals are bounded by Assumptions 1-3. Finally we consider terms within the integrals with the following example

$$\int \left(Q^o_{2-}(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2) - Q^o_{2-}(\mathbf{H}_2;\bar{\boldsymbol{\theta}})\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}}$$
$$= \int \left(\mathbf{H}_{20}^\intercal \widehat{\boldsymbol{\beta}}_{22} + [\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2]_+ - \mathbf{H}_{20}^\intercal \bar{\boldsymbol{\beta}}_{22} - [\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2]_+\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}}$$
$$= 4\|\widehat{\boldsymbol{\beta}}_{22} - \bar{\boldsymbol{\beta}}_{22}\|_2^2 \int \mathbf{H}_{20}^\intercal \mathbf{H}_{20} d\mathbb{P}_{\vec{\mathbf{U}}}$$
$$+ 4\|\widehat{\boldsymbol{\gamma}}_2 - \bar{\boldsymbol{\gamma}}_2\|_2^2 \int \mathbf{H}_{21}^\intercal \mathbf{H}_{21} d\mathbb{P}_{\vec{\mathbf{U}}} = O_{\mathbb{P}}(n^{-1}),$$

which follows from Theorem 2 and Lemma 17 (a). All similar terms can be handled accordingly. We get the convergence in probability to 0: $\int \left(\mathcal{V}_{\widehat{\boldsymbol{\Theta}},\bar{\mu}} - \mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}} \to 0$ as all other terms within expectation are $O_{\mathbb{P}}\left(n^{-1}\right)$ by the dominating convergence theorem, boundedness conditions as stated in Assumptions 2, 5, and the consistency of $\widehat{\boldsymbol{\xi}}$ and $\widehat{\boldsymbol{\theta}}$ as $\mathbb{P}\left(\widehat{\boldsymbol{\Theta}} \in \mathcal{S}(\delta)\right) \to 1$, $\forall \delta > 0$.

Finally, we have i) $\mathbb{P}\left(\widehat{\boldsymbol{\Theta}} \in \mathcal{S}(\delta)\right) \to 1$, ii) $\mathcal{C}_1$ is a Donsker class, and
iii) $\int \left(\mathcal{V}_{\widehat{\boldsymbol{\Theta}},\bar{\mu}} - \mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right)^2 d\mathbb{P}_{\vec{\mathbf{U}}} \longrightarrow 0$, then by Theorem 2.1 in Van Der Vaart and Wellner (2007),

$$\sqrt{\frac{n}{N}}\sqrt{n}\left\{\left(\mathbb{P}_N\left[\mathcal{V}_{\widehat{\boldsymbol{\Theta}},\bar{\mu}}\right] - g_1(\widehat{\boldsymbol{\Theta}})\right) - \left(\mathbb{P}_N\left[\mathcal{V}_{\bar{\boldsymbol{\Theta}},\bar{\mu}}\right] - g_1(\bar{\boldsymbol{\Theta}})\right)\right\} = \sqrt{\frac{n}{N}} o_{\mathbb{P}}(1).$$

■

Proof [Proof of Lemma 13] I) First note that from the empirical normal equations (6), we have that the solution $\widehat{\eta}_2^v$ satisfies $\widehat{\eta}_2^v - \eta_2^v = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$. Therefore

$$\begin{aligned}\sup_{\vec{\mathbf{U}}} \left|\widehat{\mu}_2^v(\vec{\mathbf{U}}) - \mu_2^v(\vec{\mathbf{U}})\right| &= \sup_{\vec{\mathbf{U}}} \left|\frac{1}{K}\widehat{m}_2^{(-k)}(\vec{\mathbf{U}}) + \widehat{\eta}_2^v - m_2(\vec{\mathbf{U}}) + \eta_2^v\right| \\ &\leq \frac{1}{K}\sup_{\vec{\mathbf{U}}} \left|\widehat{m}_2^{(-k)}(\vec{\mathbf{U}}) + m_2(\vec{\mathbf{U}})\right| + |\widehat{\eta}_2^v - \eta_2^v| \\ &= o_{\mathbb{P}}(1) + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) = o_{\mathbb{P}}(1),\end{aligned}$$

where we additionally use Assumption 6 for the difference of estimated and true imputation models $\hat{m}_2$, m_2. Similarly $\sup_{\vec{\mathbf{U}}} \left|\widehat{\mu}_{t\omega_2}^v(\vec{\mathbf{U}}) - \bar{\mu}_{t\omega_2}^v(\vec{\mathbf{U}})\right| = o_{\mathbb{P}}(1)$, $\sup_{\vec{\mathbf{U}}} \left|\widehat{\mu}_{\omega_2}^v(\vec{\mathbf{U}}) - \bar{\mu}_{\omega_2}^v(\vec{\mathbf{U}})\right| =$

$o_{\mathbb{P}}(1)$, $t = 2, 3$. Next, using the triangle and Jensen's inequalities, we have

$$
\begin{aligned}
&\mathbb{P}_N\left[\mathcal{E}_{\widehat{\boldsymbol{\Theta}}} - \mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right] \\
\leq &\mathbb{P}_N\left|\omega_1(\check{\mathbf{H}}_1, A_1; \widehat{\boldsymbol{\Theta}}_1)(1+\hat{\beta}_{21}) - \omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}}_1)(1+\bar{\beta}_{21})\right| \sup_{\vec{\mathbf{U}}}\left|\widehat{\mu}_2^v(\vec{\mathbf{U}}) - \bar{\mu}_2^v(\vec{\mathbf{U}})\right| \\
+&\left|\hat{\beta}_{21} - \bar{\beta}_{21}\right| \sup_{\vec{\mathbf{U}}}\left|\widehat{\mu}_{2\omega_2}(\vec{\mathbf{U}}) - \mu_{2\omega_2}(\vec{\mathbf{U}})\right| \\
+&\mathbb{P}_N\left|Q_{2-}^o(\mathbf{H}_2; \widehat{\boldsymbol{\theta}}_2) - Q_{2-}^o(\mathbf{H}_2; \bar{\boldsymbol{\theta}}_2)\right| \sup_{\vec{\mathbf{U}}}\left|\widehat{\mu}_{\omega_2}(\vec{\mathbf{U}}) - \bar{\mu}_{\omega_2}(\vec{\mathbf{U}})\right| \\
\leq &\mathbb{P}_N\left|\omega_1(\check{\mathbf{H}}_1, A_1; \widehat{\boldsymbol{\Theta}}_1) - \omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}}_1)\right| o_{\mathbb{P}}(1) \\
+&\mathbb{P}_N\left|\omega_1(\check{\mathbf{H}}_1, A_1; \widehat{\boldsymbol{\Theta}}_1)\hat{\beta}_{21} - \omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}}_1)\bar{\beta}_{21}\right| o_{\mathbb{P}}(1) \\
+&\left|\hat{\beta}_{21} - \bar{\beta}_{21}\right| o_{\mathbb{P}}(1) + \mathbb{P}_N\left|\left(\widehat{\boldsymbol{\beta}}_{22} - \widehat{\boldsymbol{\beta}}_{22}\right)^{\mathsf{T}} \mathbf{H}_{20} + [\widehat{\boldsymbol{\gamma}}_2^{\mathsf{T}}\mathbf{H}_{21}]_+ - [\bar{\boldsymbol{\gamma}}_2^{\mathsf{T}}\mathbf{H}_{21}]_+\right| o_{\mathbb{P}}(1).
\end{aligned}
$$

By Theorem 2 we have $\widehat{\boldsymbol{\theta}}_2 - \bar{\boldsymbol{\theta}}_2 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, also from Lemma 17 (a) it follows that $\mathbb{P}_N\left([\mathbf{H}_{21}^{\mathsf{T}}\widehat{\boldsymbol{\gamma}}_2]_+ - [\mathbf{H}_{21}^{\mathsf{T}}\bar{\boldsymbol{\gamma}}_2]_+\right) = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, hence as covariates are bounded we have

$$
\begin{aligned}
&\left|\hat{\beta}_{21} - \bar{\beta}_{21}\right| o_{\mathbb{P}}(1) + \mathbb{P}_N\left|\left(\widehat{\boldsymbol{\beta}}_{22} - \widehat{\boldsymbol{\beta}}_{22}\right)^{\mathsf{T}} \mathbf{H}_{20} + [\widehat{\boldsymbol{\gamma}}_2^{\mathsf{T}}\mathbf{H}_{21}]_+ - [\bar{\boldsymbol{\gamma}}_2^{\mathsf{T}}\mathbf{H}_{21}]_+\right| \\
&\leq \left\{o_{\mathbb{P}}(1) + \sup_{\mathbf{H}_{20}}\|\mathbf{H}_{20}\|_2\|\widehat{\boldsymbol{\beta}}_{22} - \bar{\boldsymbol{\beta}}_{22}\|_2 + \sup_{\mathbf{H}_{21}}\|\mathbf{H}_{21}\|_2\right\} O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).
\end{aligned}
$$

Next, we can write

$$
\omega_1(\mathbf{H}_1, A_1; \widehat{\boldsymbol{\Theta}}_1) = I\left\{A_1 = d_1\left(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1\right)\right\}\left\{\frac{A_1}{\pi_1\left(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1\right)} + \frac{1 - A_1}{1 - \pi_1\left(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1\right)}\right\}.
$$

By Lemma 17 (b) it follows that

$$
\begin{aligned}
\mathbb{P}_N\left[I\left\{A_1 = d_1\left(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1\right)\right\} - I\left\{A_1 = d_1\left(\mathbf{H}_1; \bar{\boldsymbol{\xi}}_1\right)\right\}\right] &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right), \\
\mathbb{P}_N\left[\frac{A_1}{\pi_1(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1)} - \frac{A_1}{\pi_1\left(\mathbf{H}_1; \bar{\boldsymbol{\xi}}_1\right)}\right] &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right), \\
\mathbb{P}_N\left[\frac{1 - A_1}{1 - \pi_1(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1)} - \frac{1 - A_1}{1 - \pi_1\left(\mathbf{H}_1; \bar{\boldsymbol{\xi}}_1\right)}\right] &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).
\end{aligned}
$$

Using the above and Lemma 14 we get

$$
\begin{aligned}
\hat{\beta}_{21}\mathbb{P}_N\left\{\omega_1(\check{\mathbf{H}}_1, A_1; \widehat{\boldsymbol{\Theta}}_1)\right\} - \bar{\beta}_{21}\mathbb{P}_N\left\{\omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}}_1)\right\} &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right), \\
\mathbb{P}_N\left\{\omega_1(\check{\mathbf{H}}_1, A_1; \widehat{\boldsymbol{\Theta}}_1)\right\} - \mathbb{P}_N\left\{\omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}}_1)\right\} &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).
\end{aligned}
$$

From the above we get

$$\mathbb{P}_N\left\{\mathcal{E}_{\widehat{\mathbf{\Theta}}} - \mathcal{E}_{\bar{\mathbf{\Theta}}}\right\} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) o_{\mathbb{P}}(1).$$

II) To show the relevant result, we first recall the definition of $\nu_{\mathsf{SSL}_{\mathsf{DR}}}$ from Theorem 7 and show that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\widehat{\mathbf{\Theta}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\mathbf{\Theta}}) \\ + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left[\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\mathbf{\Theta}})\right]\right)^{\intercal}\boldsymbol{\psi}^{\theta}(\mathbf{L}_i) \\ + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\boldsymbol{\xi}}\mathbb{E}\left[\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\mathbf{\Theta}})\right]\right)^{\intercal}\boldsymbol{\psi}^{\xi}(\mathbf{L}_i) + o_{\mathbb{P}}(1).\end{aligned} \tag{17}$$

We start expanding $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\widehat{\mathbf{\Theta}})$ as

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\widehat{\mathbf{\Theta}}) \\ &= \mathbb{G}_n\left\{\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\} + \mathbb{G}_n\left\{\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}) - \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\} + \sqrt{n}\int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}},\end{aligned}$$

we next consider the limit of each term above.

1) Using a Taylor series expansion on $\int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}}$ we get

$$\int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}} = \int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}} + \left(\widehat{\mathbf{\Theta}} - \bar{\mathbf{\Theta}}\right)^{\intercal}\frac{\partial}{\partial\mathbf{\Theta}}\int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} + O_{\mathbb{P}}\left(n^{-1}\right),$$

where the remaining terms are of order $O\left\{\left(\widehat{\mathbf{\Theta}} - \bar{\mathbf{\Theta}}\right)^2\right\}$ which by Theorems 2 & 3 are $O_{\mathbb{P}}\left(n^{-1}\right)$. Next note that from (13) it follows that $\int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}} = 0$, and thus letting $g_2(\mathbf{\Theta}) = \int\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})d\mathbb{P}_{\mathbf{L}}$ we have

$$\sqrt{n}g_2(\widehat{\mathbf{\Theta}}) = \sqrt{n}(\widehat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^{\intercal}\frac{\partial}{\partial\mathbf{\Theta}}g_2(\mathbf{\Theta})\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} + \sqrt{n}(\widehat{\boldsymbol{\xi}} - \bar{\boldsymbol{\xi}})^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}g_2(\mathbf{\Theta})\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} + o_{\mathbb{P}}(1).$$

We can write

$$\sqrt{n}g_2(\widehat{\mathbf{\Theta}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\theta}(\mathbf{L}_i)^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}g_2(\mathbf{\Theta})\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\xi}(\mathbf{L}_i)^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}g_2(\mathbf{\Theta})\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}} + o_{\mathbb{P}}(1).$$

2) We next show

$$\mathbb{G}_n\left\{\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}) - \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\} = o_{\mathbb{P}}(1),$$

define the class

$$\ell_t = \{I(\mathbf{H}^{\intercal}\boldsymbol{\gamma}_t \geq 0) : \mathcal{H}_{t1}, \boldsymbol{\gamma} \in \mathbb{R}^{q_t}\},\ t = 1, 2$$

and the collection of half spaces $\mathcal{C}_\ell \equiv \{\mathbf{H}_t \in \mathbb{R}^{q_t} : \mathbf{H}^\intercal \boldsymbol{\gamma}_t \geq 0, \boldsymbol{\gamma} \in \mathbb{R}^{q_t}, t \in \{1,2\}\}$, by Dudley (1979) $\mathcal{C}_\ell$ is a VC class of VC dimension $q_t + 1$, next by van der Vaart and Wellner (1996) we have that as $\mathcal{C}_\ell$ is a VC-class ℓ_t is a class of the same index. Finally, by Theorem 2.6.7 we have that ℓ_t is a Donsker class.

$$
\begin{aligned}
f_{\boldsymbol{\Theta}}(\mathbf{L}_i) =& \omega_1(\check{\mathbf{H}}_{1i}, A_{1i}; \boldsymbol{\Theta}_1)(1+\beta_{21}) \left\{ Y_{2i} - \bar{\mu}_2^v(\vec{\mathbf{U}}_i) \right\} + \omega_2(\check{\mathbf{H}}_{2i}, A_{2i}; \boldsymbol{\Theta}_2) Y_{3i} - \bar{\mu}_{3\omega_2}(\vec{\mathbf{U}}_i) \\
& - \beta_{21} \left\{ \omega_2(\check{\mathbf{H}}_{2i}, A_{2i}; \boldsymbol{\Theta}_2) Y_{2i} - \bar{\mu}_{2\omega_2}(\vec{\mathbf{U}}_i) \right\} - Q_{2-}^o(\mathbf{H}_{2i}; \boldsymbol{\theta}_2) \left\{ \omega_2(\check{\mathbf{H}}_{2i}, A_{2i}; \boldsymbol{\Theta}_2) - \bar{\mu}_{\omega_2}(\vec{\mathbf{U}}_i) \right\},
\end{aligned}
$$

we define the class of functions $\mathcal{C}_2 = \{ f_{\boldsymbol{\Theta}}(\mathbf{L}) | \boldsymbol{\Theta} \in \mathcal{S}(\delta) \}$.

i) By Assumptions 3, 5 and Theorem 19.5 in Vaart (1998), $\mathcal{W}_t$, $\mathcal{Q}_t, t = 1,2$ are a $\mathbb{P}$-Donsker class. Additionally, the terms in the $\omega_t(\mathbf{H}_t, A_t; \boldsymbol{\Theta}_t)$ functions of the form $\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t I(\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t > 0)$ constitute a $\mathbb{P}$-Donsker class, as $\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t$ is linear in $\boldsymbol{\gamma}_t$ and $I(\mathbf{H}_{t1}^\intercal \boldsymbol{\gamma}_t > 0)$ is $\mathbb{P}$-Donsker. Thus it follows that $\mathcal{C}_2$ is a $\mathbb{P}$-Donsker class.

ii) We estimate $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ for (4) with their maximum likelihood estimators, $\widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2$, solving $\mathbb{P}_n [S_t(\boldsymbol{\xi}_t)] = \mathbf{0}, t = 1,2$, by Assumption 5 and Theorem 5.9 in Vaart (1998) $\widehat{\boldsymbol{\xi}}_t \xrightarrow{p} \bar{\boldsymbol{\xi}}_t, t = 1,2$. Next, by Theorems 2, 3, under Assumptions 1, 2, $\widehat{\boldsymbol{\theta}}_t \xrightarrow{p} \bar{\boldsymbol{\theta}}_t, t = 1,2$. Thus $\mathbb{P}\left(\widehat{\boldsymbol{\Theta}} \in \mathcal{S}(\delta) \right) \xrightarrow{p} 1$, $\forall \delta$. Therefore, we have $\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}}) \in \mathcal{C}_2$ with high probability.

iii) We then show $\int \left\{ \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}}) - \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \bar{\boldsymbol{\Theta}}) \right\}^2 d\mathbb{P}_{\mathbf{L}} \longrightarrow 0$. Using simple algebra for a large enough constant c we have

$$
\begin{aligned}
& \int \left\{ \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}}) - \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \bar{\boldsymbol{\Theta}}) \right\}^2 d\mathbb{P}_{\mathbf{L}} \\
\leq & c \sup_{Y_2, \vec{\mathbf{U}}} \left\{ Y_2 - \bar{\mu}_2^v(\vec{\mathbf{U}}) \right\}^2 \\
& \times \sup_{\check{\mathbf{H}}_1, A_1} \left\{ (1+\hat{\beta}_{21}) \omega_1(\check{\mathbf{H}}_1, A_1; \widehat{\boldsymbol{\Theta}}_1) - (1+\bar{\beta}_{21}) \omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}}_1) \right\}^2 \\
& + c \sup_{Y_3} Y_3^2 \sup_{\check{\mathbf{H}}_2, A_2} \left\{ \omega_2(\check{\mathbf{H}}_2, A_2; \widehat{\boldsymbol{\Theta}}_2) - \omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}_2) \right\}^2 \\
& + c \sup_{Y_2} Y_2^2 \sup_{\check{\mathbf{H}}_2, A_2} \left\{ \hat{\beta}_{21} \omega_2(\check{\mathbf{H}}_2, A_2; \widehat{\boldsymbol{\Theta}}_2) - \bar{\beta}_{21} \omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}_2) \right\}^2 \\
& + c \sup_{\vec{\mathbf{U}}} \bar{\mu}_{2\omega_3}(\vec{\mathbf{U}})^2 \left(\hat{\beta}_{21} - \bar{\beta}_{21} \right)^2 \\
& + c \sup_{\check{\mathbf{H}}_2, A_2} \left\{ Q_{2-}^o(\mathbf{H}_2; \widehat{\boldsymbol{\theta}}_2) \omega_2(\check{\mathbf{H}}_2, A_2; \widehat{\boldsymbol{\Theta}}_2) - Q_{2-}^o(\mathbf{H}_2; \bar{\boldsymbol{\theta}}_2) \omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}_2) \right\}^2 \\
& + c \sup_{\vec{\mathbf{U}}} \bar{\mu}_{2\omega_2}(\vec{\mathbf{U}})^2 \sup_{\mathbf{H}_2} \left\{ Q_{2-}^o(\mathbf{H}_2; \widehat{\boldsymbol{\theta}}_2) - Q_{2-}^o(\mathbf{H}_2; \bar{\boldsymbol{\theta}}_2) \right\}^2 \\
& \xrightarrow{p} 0
\end{aligned}
$$

where we use $(a-b)^2, (a+b)^2 \leq 2a^2 + 2b^2$ $\forall a, b \in \mathbb{R}$, boundedness of $\bar{\boldsymbol{\Theta}}$ and covariates by Assumptions 1, 2 to bound all supremum quantities.

By Theorems 2 and 3 we have $\widehat{\boldsymbol{\theta}}_2 - \bar{\boldsymbol{\theta}}_2 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, $\widehat{\boldsymbol{\theta}}_1 - \bar{\boldsymbol{\theta}}_1 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, also from Lemma 17 (a) it follows that

$$\begin{aligned}
&\sup_{\mathbf{H}_2}\left\{Q^o_{2-}(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2) - Q^o_{2-}(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\right\}^2 \\
&\leq 2\sup_{\mathbf{H}_{20}}\|\mathbf{H}_{20}\|_2^2\|\widehat{\boldsymbol{\beta}}_{22} - \bar{\boldsymbol{\beta}}_{22}\|_2^2 + 2\sup_{\mathbf{H}_{21}}\|\mathbf{H}_{21}\|_2^2\|\widehat{\boldsymbol{\gamma}}_{22} - \bar{\boldsymbol{\gamma}}_{22}\|_2 \\
&= O_{\mathbb{P}}\left(n^{-1}\right).
\end{aligned}$$

Next, we can write

$$\begin{aligned}
\omega_1(\mathbf{H}_1, A_1; \widehat{\boldsymbol{\Theta}}_1) =& I\left\{A_1 = d_1\left(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1\right)\right\}\left\{\frac{A_1}{\pi_1\left(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1\right)} + \frac{1-A_1}{1-\pi_1\left(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1\right)}\right\} \\
\omega_2(\check{\mathbf{H}}_2, A_2; \widehat{\boldsymbol{\Theta}}_1) =& \omega_1(\mathbf{H}_1, A_1; \widehat{\boldsymbol{\Theta}}_1) I\left\{A_2 = d_2\left(\mathbf{H}_2; \widehat{\boldsymbol{\xi}}_2\right)\right\}\left\{\frac{A_2}{\pi_2\left(\check{\mathbf{H}}_2; \widehat{\boldsymbol{\xi}}_2\right)} + \frac{1-A_2}{2-\pi_2\left(\check{\mathbf{H}}_2; \widehat{\boldsymbol{\xi}}_2\right)}\right\}.
\end{aligned}$$

By Lemma 17 (b) it follows that

$$\begin{aligned}
\sup_{\mathbf{H}_1, \mathbf{a}_1}\left|I(\widehat{d}_1 = A_1) - I(\bar{d}_1 = A_1)\right| =& o_{\mathbb{P}}(1), \\
\sup_{\mathbf{H}_2, \mathbf{a}_2}\left|I(\widehat{d}_1 = A_1)I(A_2 = \widehat{d}_2) - I(\bar{d}_1 = A_1)I(\bar{d}_2 = A_2)\right| =& o_{\mathbb{P}}(1), \\
\sup_{\mathbf{H}_1}\left|\frac{1}{\pi_1(\mathbf{H}_1; \widehat{\boldsymbol{\xi}}_1)} - \frac{1}{\pi_1(\mathbf{H}_1; \bar{\boldsymbol{\xi}}_1)}\right| =& O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).
\end{aligned}$$

Using the above and Lemma 14 we get

$$\begin{aligned}
\sup_{\check{\mathbf{H}}_1, A_1}\left\{(1+\hat{\beta}_{21})\omega_1(\check{\mathbf{H}}_1, A_1; \widehat{\boldsymbol{\Theta}}_1) - (1+\bar{\beta}_{21})\omega_1(\check{\mathbf{H}}_1, A_1; \bar{\boldsymbol{\Theta}}_1)\right\}^2 &= o_{\mathbb{P}}(1), \\
\sup_{\check{\mathbf{H}}_2, A_2}\left\{(1+\hat{\beta}_{21})\omega_2(\check{\mathbf{H}}_2, A_2; \widehat{\boldsymbol{\Theta}}_2) - (1+\bar{\beta}_{21})\omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}_2)\right\}^2 &= o_{\mathbb{P}}(1), \\
\sup_{\check{\mathbf{H}}_2, A_2}\left\{Q^o_{2-}(\mathbf{H}_2; \widehat{\boldsymbol{\theta}}_2)\omega_2(\check{\mathbf{H}}_2, A_2; \widehat{\boldsymbol{\Theta}}_2) - Q^o_{2-}(\mathbf{H}_2; \bar{\boldsymbol{\theta}}_2)\omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}_2)\right\}^2 &= o_{\mathbb{P}}(1), \\
\sup_{\check{\mathbf{H}}_2, A_2}\left\{\hat{\beta}_{21}\omega_2(\check{\mathbf{H}}_2, A_2; \widehat{\boldsymbol{\Theta}}_2) - \bar{\beta}_{21}\omega_2(\check{\mathbf{H}}_2, A_2; \bar{\boldsymbol{\Theta}}_2)\right\}^2 &= o_{\mathbb{P}}(1).
\end{aligned}$$

which gives us $\int\left\{\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}}) - \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \bar{\boldsymbol{\Theta}})\right\}^2 d\mathbb{P}_{\mathbf{L}} \xrightarrow{p} 0$.

Therefore we have i) $\mathbb{P}\left(\widehat{\boldsymbol{\Theta}} \in \mathcal{S}(\delta)\right) \to 1$, $\forall\delta$, ii) $\mathcal{C}_2$ is a $\mathbb{P}$-Donsker class, and iii) $\int\left(\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}}) - \nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \bar{\boldsymbol{\Theta}})\right)^2 d\mathbb{P}_{\mathbf{L}} \to 0$. By Theorem 2.1 in Van Der Vaart and Wellner (2007)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\left(\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i; \widehat{\boldsymbol{\Theta}}) - \mathbb{E}_{\mathbb{S}}[\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \widehat{\boldsymbol{\Theta}})]\right) - \left(\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i; \bar{\boldsymbol{\Theta}}) - \mathbb{E}_{\mathbb{S}}[\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}; \bar{\boldsymbol{\Theta}})]\right)\right\} = o_{\mathbb{P}}(1).$$

by 1), 2) and noting that $\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\boldsymbol{\Theta}})$ has mean zero we obtain the result in (17).

We next re-write $\sqrt{n}\mathbb{P}_N\left[\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right]$ by expressing the estimated imputation functions in $\mathcal{E}_{\bar{\boldsymbol{\Theta}}}$ in terms of the labeled sample $\mathcal{L}$. Letting

$$\hat{C}^{(1)}_{n,N}=\frac{(1+\bar{\beta}_{21})\mathbb{P}_N\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\boldsymbol{\Theta}})\right\}}{(1+\hat{\beta}_{21})\mathbb{P}_n\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\widehat{\boldsymbol{\Theta}})\right\}},\ \hat{C}^{(2)}_{n,N}=\frac{\mathbb{P}_N\left\{Q^o_{2-}(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\right\}}{\mathbb{P}_n\left\{Q^o_{2-}(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2)\right\}},$$

we can write:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\omega_1(\check{\mathbf{H}}_{1j},A_{1j},\bar{\boldsymbol{\Theta}})(1+\bar{\beta}_{21})\left\{\widehat{\mu}^v_2(\vec{\mathbf{U}}_j)-\bar{\mu}^v_2(\vec{\mathbf{U}}_j)\right\}\\
=&\frac{1}{N}\sum_{j=1}^{N}\omega_1(\check{\mathbf{H}}_{1j},A_{1j},\bar{\boldsymbol{\Theta}})(1+\bar{\beta}_{21})\left\{\frac{1}{K}\sum_{k=1}^{K}\widehat{m}^{(\text{-}k)}_2(\vec{\mathbf{U}}_j)+\widehat{\eta}^v_2-m_2(\vec{\mathbf{U}}_j)-\eta^v_2\right\}\\
=&(1+\bar{\beta}_{21})\frac{1}{KN}\sum_{j=1}^{N}\sum_{k=1}^{K}\omega_1(\check{\mathbf{H}}_{1j},A_{1j},\bar{\boldsymbol{\Theta}})\hat{\Delta}^{(\text{-}k)}_2(\vec{\mathbf{U}}_j)+(\widehat{\eta}^v_2-\eta^v_2)\frac{1}{N}\sum_{j=1}^{N}\omega_1(\check{\mathbf{H}}_{1j},A_{1j},\bar{\boldsymbol{\Theta}})(1+\bar{\beta}_{21}),
\end{aligned}$$

where the first step follows from constrains shown in (6) and we simply regroup terms in the second step.

Next note that we can use Lemma 15 to replace

$$\mathbb{P}_N\left[(1+\bar{\beta}_{21})\omega_1(\check{\mathbf{H}}_1,A_1,\bar{\boldsymbol{\Theta}})\hat{\Delta}^{(\text{-}k)}_2(\vec{\mathbf{U}}_j)\right]$$

by

$$\mathbb{E}_{\mathcal{L}}\left[(1+\bar{\beta}_{21})\omega_1(\check{\mathbf{H}}_1,A_1,\bar{\boldsymbol{\Theta}})\hat{\Delta}^{(\text{-}k)}_2(\vec{\mathbf{U}}_j)\right]+O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right),$$

using $\mathbb{E}_{\mathcal{L}}[\cdot]$ to denote expectation with respect to $\mathcal{L}$. Additionally, using (6) and the definition of $\bar{\mu}^v_2(\vec{\mathbf{U}})$ for the second term we get:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\omega_1(\check{\mathbf{H}}_{1j},A_{1j};\bar{\boldsymbol{\Theta}})(1+\bar{\beta}_{21})\left\{\widehat{\mu}^v_2(\vec{\mathbf{U}}_j)-\bar{\mu}^v_2(\vec{\mathbf{U}}_j)\right\}\\
=&\mathbb{E}_{\mathcal{L}}\left[\frac{1}{K}\sum_{k=1}^{K}(1+\bar{\beta}_{21})\omega_1(\check{\mathbf{H}}_1,A_1,\bar{\boldsymbol{\Theta}})\hat{\Delta}^{(\text{-}k)}_2(\vec{\mathbf{U}})\right]+O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)\\
-&\hat{C}^{(1)}_{n,N}\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}(1+\hat{\beta}_{21})\omega_1(\check{\mathbf{H}}_{1i},A_{1i};\widehat{\boldsymbol{\Theta}})\hat{\Delta}^{(\text{-}k)}_2(\vec{\mathbf{U}}_i)\\
+&\hat{C}^{(1)}_{n,N}(1+\hat{\beta}_{21})\frac{1}{n}\sum_{i=1}^{n}\omega_1(\check{\mathbf{H}}_{1i},A_{1i};\widehat{\boldsymbol{\Theta}})\left\{Y_{2i}-\bar{\mu}^v_2(\vec{\mathbf{U}}_i)\right\}\\
=&\left\{1+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\right\}(1+\hat{\beta}_{21})\frac{1}{n}\sum_{i=1}^{n}\omega_1(\check{\mathbf{H}}_{1i},A_{1i};\widehat{\boldsymbol{\Theta}})\left\{Y_{2i}-\bar{\mu}^v_2(\vec{\mathbf{U}}_i)\right\}+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}c_{n^-_K}\right),
\end{aligned}$$

where the last step follows from Assumption 6 and Lemma 16 choosing f to be the constant function 1, setting $\hat{\Delta}_k(\vec{\mathbf{U}}) = \hat{\Delta}_2^{(-k)}(\vec{\mathbf{U}})$, $\hat{l}(\check{\mathbf{H}}_1) = A_1 I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0)$, and $\hat{\pi}(\check{\mathbf{H}}_1) = \pi_1(\check{\mathbf{H}}_1; \widehat{\boldsymbol{\xi}}_1)$ and with $\hat{C}_{n,N} = \hat{C}_{n,N}^{(1)}$ -which satisfies $\hat{C}_{n,N}^{(1)} = 1 + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ by Lemma 17 (c).

Using similar arguments we have

$$
\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N} Q_{2-}^{o}(\mathbf{H}_{2j};\bar{\boldsymbol{\theta}}_2)\{\widehat{\mu}_{\omega_2}(\vec{\mathbf{U}}_j) - \bar{\mu}_{\omega_2}(\vec{\mathbf{U}}_j)\} \\
&= \frac{1}{N}\sum_{j=1}^{N} Q_{2-}^{o}(\mathbf{H}_{2j};\bar{\boldsymbol{\theta}}_2)\left\{\frac{1}{K}\sum_{k=1}^{K}\widehat{m}_{\omega_2}^{(-k)}(\vec{\mathbf{U}}_j) + \widehat{\eta}_{\omega_2}^{v} - m_{\omega_2}(\vec{\mathbf{U}}_j) - \eta_{\omega_2}^{v}\right\} \\
&= \frac{1}{KN}\sum_{j=1}^{N}\sum_{k=1}^{K} Q_{2-}^{o}(\mathbf{H}_{2j};\bar{\boldsymbol{\theta}}_2)\hat{\Delta}_{\omega_2 k}(\vec{\mathbf{U}}_j) + \left(\widehat{\eta}_{\omega_2}^{v} - \eta_{\omega_2}^{v}\right)\frac{1}{N}\sum_{j=1}^{N} Q_{2-}^{o}(\mathbf{H}_{2j};\bar{\boldsymbol{\theta}}_2) \\
&= \mathbb{E}_{\mathcal{L}}\left[\frac{1}{K}\sum_{k=1}^{K} Q_{2-}^{o}(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_2)\hat{\Delta}_{\omega_2 k}(\vec{\mathbf{U}})\right] - \hat{C}_{n,N}^{(2)}\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} Q_{2-}^{o}(\mathbf{H}_{2i};\bar{\boldsymbol{\theta}}_2)\hat{\Delta}_{\omega_2 k}(\vec{\mathbf{U}}_i) \\
&+ \hat{C}_{n,N}^{(2)}\frac{1}{n}\sum_{i=1}^{n} Q_{2-}^{o}(\mathbf{H}_{2i};\widehat{\boldsymbol{\theta}}_2)\{\omega_2(\check{\mathbf{H}}_{2i}, A_{2i};\widehat{\boldsymbol{\Theta}}) - \bar{\mu}_{\omega_2}(\vec{\mathbf{U}}_i)\} + O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right) \\
&= \left\{1 + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\right\}\frac{1}{n}\sum_{i=1}^{n} Q_{2-}^{o}(\mathbf{H}_{2i};\widehat{\boldsymbol{\theta}}_2)\{\omega_2(\check{\mathbf{H}}_{2i}, A_{2i};\widehat{\boldsymbol{\Theta}}) - \bar{\mu}_{\omega_2}(\vec{\mathbf{U}}_i)\} + O_{\mathbb{P}}\left(n^{-\frac{1}{2}} c_{n_K^-}\right),
\end{aligned}
$$

and for $t = 2, 3$

$$
\begin{aligned}
\frac{1}{N}\sum_{j=1}^{N}\left\{\widehat{\mu}_{t\omega_2}^{v}(\vec{\mathbf{U}}_j) - \bar{\mu}_{t\omega_2}^{v}(\vec{\mathbf{U}}_j)\right\} =& \frac{1}{KN}\sum_{j=1}^{N}\left\{\sum_{k=1}^{K}\widehat{m}_{t\omega_2}^{(-k)}(\vec{\mathbf{U}}_j) + \widehat{\eta}_{t\omega_2}^{v} - m_{t\omega_2}(\vec{\mathbf{U}}_j) - \eta_{t\omega_2}^{v}\right\} \\
=& \frac{1}{KN}\sum_{j=1}^{N}\sum_{k=1}^{K}\hat{\Delta}_{3\omega_2 k}(\vec{\mathbf{U}}_j) + (\widehat{\eta}_{t\omega_2}^{v} - \eta_{t\omega_2}^{v}) \\
=& \mathbb{E}_{\mathcal{L}}\left[\frac{1}{K}\sum_{k=1}^{K}\hat{\Delta}_{3\omega_2 k}(\vec{\mathbf{U}})\right] - \frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\hat{\Delta}_{3\omega_2 k}(\vec{\mathbf{U}}_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\omega_2(\check{\mathbf{H}}_{2i}, A_{2i};\widehat{\boldsymbol{\Theta}})Y_{ti} - \bar{\mu}_{t\omega_2}^{v}(\vec{\mathbf{U}}_i) + O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right) \\
=& \frac{1}{n}\sum_{i=1}^{n}\omega_2(\check{\mathbf{H}}_{2i}, A_{2i};\widehat{\boldsymbol{\Theta}})Y_{ti} - \bar{\mu}_{t\omega_2}^{v}(\vec{\mathbf{U}}_i) + O_{\mathbb{P}}\left(n^{-\frac{1}{2}} c_{n_K^-}\right),
\end{aligned}
$$

finally by Theorem 2, $\hat{\beta}_{21} - \bar{\beta}_{21} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.

Therefore, recalling the definition of $\nu_{\mathsf{SSL}_{\mathsf{DR}}}$ from Theorem 7, using the derivations above, we can write

$$\begin{aligned}\frac{\sqrt{n}}{N}\sum_{j=1}^{N}\mathcal{E}_{\bar{\boldsymbol{\Theta}}}(\vec{\mathbf{U}}_j) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}_i;\widehat{\boldsymbol{\Theta}}) \\ + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_1(\check{\mathbf{H}}_1,A_1;\widehat{\boldsymbol{\Theta}}_1)(1+\hat{\beta}_{21})\left\{Y_2-\bar{\mu}_2^v(\vec{\mathbf{U}})\right\} \\ - O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\beta}_{21}\left\{\omega_2(\check{\mathbf{H}}_2,A_2,\widehat{\boldsymbol{\Theta}}_2)Y_2-\bar{\mu}_{2\omega_2}(\vec{\mathbf{U}})\right\} \\ - O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q^o_{2-}(\mathbf{H}_2;\widehat{\boldsymbol{\theta}}_2)\left\{\omega_2(\check{\mathbf{H}}_2,A_2,\bar{\boldsymbol{\Theta}}_2)-\bar{\mu}_{\omega_2}(\vec{\mathbf{U}})\right\} \\ + O_{\mathbb{P}}\left(c_{n_K^-}\right).\end{aligned} \tag{18}$$

Using result (17) we know $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}}_i;\widehat{\boldsymbol{\Theta}}) = O_{\mathbb{P}}(1)$, therefore the second, third and fourth terms in (18) are $o_{\mathbb{P}}(1)$. Using (17) again for the first term in (18) we get our required result:

$$\begin{aligned}\sqrt{n}\mathbb{P}_N\left[\mathcal{E}_{\bar{\boldsymbol{\Theta}}}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\boldsymbol{\Theta}}) \\ + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left[\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\boldsymbol{\Theta}})\right]\right)^{\mathsf{T}}\boldsymbol{\psi}^{\theta}(\mathbf{L}_i) \\ + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\boldsymbol{\xi}}\mathbb{E}\left[\nu_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L}_i;\bar{\boldsymbol{\Theta}})\right]\right)^{\mathsf{T}}\boldsymbol{\psi}^{\xi}(\mathbf{L}_i) + o_{\mathbb{P}}(1).\end{aligned}$$

■

Proof [Proof of Proposition 8]
Recall the definition of $\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})$ in (5), using (13) we have $\mathbb{E}\left[\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\boldsymbol{\Theta}},\bar{\mu})\right] = \mathbb{E}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})\right]$, therefore

$$\text{Bias}\left\{\bar{V},\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}\left(\mathbf{L};\bar{\boldsymbol{\Theta}}\right)\right\} = \text{Bias}\left\{\bar{V},\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\boldsymbol{\Theta}},\bar{\mu})\right\}.$$

Therefore, by Lemma 18 we have

$$\begin{aligned}&\text{Bias}\left\{\bar{V},\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\boldsymbol{\Theta}},\bar{\mu})\right\} \\ \leq&\sqrt{\sup_{\check{\mathbf{H}}_1}|\{1-\pi_1(\check{\mathbf{H}}_1;\bar{\boldsymbol{\xi}}_1)\}^{-1}|}\sqrt{\|\pi_1(\check{\mathbf{H}}_1;\bar{\boldsymbol{\xi}}_1)-\pi_1(\check{\mathbf{H}}_1)\|_{L_2(\mathbb{P})}}\sqrt{\|Q_1^o(\check{\mathbf{H}}_1;\bar{\boldsymbol{\theta}}_1)-Q_1^o(\check{\mathbf{H}}_1)\|_{L_2(\mathbb{P})}} \\ +&\sqrt{\sup_{\check{\mathbf{H}}_2}\left|\left\{\frac{A_1}{\pi_1(\check{\mathbf{H}}_1;\bar{\boldsymbol{\xi}}_1)}+\frac{1-A_1}{1-\pi_1(\check{\mathbf{H}}_1;\bar{\boldsymbol{\xi}}_1)}\right\}\{1-\pi_1(\check{\mathbf{H}}_1;\bar{\boldsymbol{\xi}}_1)\}^{-1}\{1-\pi_2(\check{\mathbf{H}}_2;\bar{\boldsymbol{\xi}}_2)\}^{-1}\right|} \\ \times&\sqrt{\|\pi_2(\check{\mathbf{H}}_2;\bar{\boldsymbol{\xi}}_2)-\pi_2(\check{\mathbf{H}}_2)\|_{L_2(\mathbb{P})}}\sqrt{\|Q_2^o(\check{\mathbf{H}}_2;\bar{\boldsymbol{\theta}}_2)-Q_2^o(\check{\mathbf{H}}_2)\|_{L_2(\mathbb{P})}}.\end{aligned}$$

Next using Theorem 7

$$\sqrt{n}\left\{\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}-\bar{V}\right\}+\sqrt{n}\text{Bias}\left\{\bar{V},\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\mathbf{\Theta}},\bar{\mu})\right\}\xrightarrow{d}N\left(0,\sigma^2_{\mathsf{SSL}_{\mathsf{DR}}}\right), \tag{19}$$

if either (1) or (4) are correct then Bias $\left\{\bar{V},\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\mathbf{\Theta}},\bar{\mu})\right\}=o_{\mathbb{P}}(1)$, multiplying (19) by $n^{-\frac{1}{2}}$ we have

$$\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}-\bar{V}\xrightarrow{\mathbb{P}}0,$$

which is the required result for Proposition 8 (a).

Next, if $\sqrt{\|\pi_t(\check{\mathbf{H}}_t;\widehat{\boldsymbol{\xi}}_t)-\pi_t(\check{\mathbf{H}}_t)\|_{L_2(\mathbb{P})}}\sqrt{\|Q^o_t(\check{\mathbf{H}}_t;\widehat{\boldsymbol{\theta}}_t)-Q^o_t(\check{\mathbf{H}}_t)\|_{L_2(\mathbb{P})}}=O_{\mathbb{P}}\left(n^{-1}\right)$ for $t=1,2$ then Bias $\left\{\bar{V},\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\mathbf{\Theta}},\bar{\mu})\right\}=O_{\mathbb{P}}\left(n^{-1}\right)$ and from (19) we get

$$\sqrt{n}\left\{\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}-\bar{V}\right\}\xrightarrow{d}N\left(0,\sigma^2_{\mathsf{SSL}_{\mathsf{DR}}}\right),$$

which is the required result for Proposition 8 (b). ■

Before proving Proposition 9, we introduce a useful definition and state the necessary assumption to prove the result. Let $\psi^{\xi}_{\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\xi}})$ and $\psi^{\xi}_{\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\xi}})$ be the supervised and SSL influence functions respectively for $\boldsymbol{\xi}$, then we define

$$\begin{aligned}\mathcal{E}^v(\vec{\mathbf{U}})=&\mathcal{V}_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\mathbf{\Theta}},\bar{\mu})-\mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right]+\mathcal{E}^{\theta}(\vec{\mathbf{U}})^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}\int\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}}\\&+\mathcal{E}^{\xi}(\vec{\mathbf{U}})^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}\int\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}},\\\mathcal{E}^{\xi}(\vec{\mathbf{U}})=&\psi^{\xi}_{\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\xi}})-\psi^{\xi}_{\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\xi}}).\end{aligned}$$

We need to ensure that the imputation models $\bar{\mu}^v_2(\vec{\mathbf{U}})$, $\bar{\mu}^v_{\omega_2}(\vec{\mathbf{U}})$, $\bar{\mu}^v_{t\omega_2}(\vec{\mathbf{U}})$, $t=2,3$ used in the SSL value function estimator $V_{\mathsf{SSL}_{\mathsf{DR}}}$ are unbiased when multiplied by several functions. For example, we need additional constraints of the type:

$$\begin{aligned}\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)Q^o_{2-}(\mathbf{H}_2;\bar{\boldsymbol{\theta}}_1)\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\}\right]=&\mathbf{0},\\\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)^2\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\}\right]=&\mathbf{0},\\\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)^2\{Y_2-\bar{\mu}_2(\vec{\mathbf{U}})\}\right]=&\mathbf{0},\end{aligned}$$

so the imputation models are unbiased in expectation when multiplied by every term and cross-product of terms in $\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})$, $\mathcal{E}^v(\vec{\mathbf{U}})$. These constraints can be summarized in the following Assumption.

Assumption 8 *Imputation models* $\bar{\mu}^v_2(\vec{\mathbf{U}})$, $\bar{\mu}^v_{\omega_2}(\vec{\mathbf{U}})$, $\bar{\mu}^v_{t\omega_2}(\vec{\mathbf{U}})$, $t=2,3$ *satisfy*

$$\mathbb{E}\left[\left\{\mathcal{E}^v(\vec{\mathbf{U}})-\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\}\mathcal{E}^v(\vec{\mathbf{U}})\right]=0.$$

Proof [Proof of Proposition 9] From Theorem 19 in Appendix F.1 we have that the influence function for the fully-supervised value function estimator (5) is:

$$\begin{aligned}\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) =&\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) - \mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})\right] + \boldsymbol{\psi}^{\theta}_{\mathsf{SUP}}(\mathbf{L})^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}}\\ &+\boldsymbol{\psi}^{\xi}_{\mathsf{SUP}}(\mathbf{L})^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})d\mathbb{P}_{\mathbf{L}}\Big|_{\boldsymbol{\Theta}=\bar{\boldsymbol{\Theta}}}.\end{aligned}$$

Next, as we estimate $\boldsymbol{\xi}$ with a semi-supervised approach such that $\boldsymbol{\psi}^{\xi}_{\mathsf{SSL}}(\mathbf{L};\bar{\boldsymbol{\xi}}) = \boldsymbol{\psi}^{\xi}_{\mathsf{SUP}}(\mathbf{L};\bar{\boldsymbol{\xi}}) - \mathcal{E}^{\xi}(\vec{\mathbf{U}})$, simple algebra can be used to show that

$$\psi^v_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) = \psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) - \mathcal{E}^v(\vec{\mathbf{U}}).$$

Using the above we can write

$$\begin{aligned}\sigma^2_{\mathsf{SSL}_{\mathsf{DR}}} = \mathbb{E}\left[\psi^v_{\mathsf{SSL}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})^2\right] =&\mathbb{E}\left[\left\{\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}}) - \mathcal{E}^v(\vec{\mathbf{U}})\right\}^2\right]\\ =&\mathbb{E}\left[\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})^2\right] + \mathbb{E}\left[\mathcal{E}^v(\vec{\mathbf{U}})^2\right]\\ &-2\mathbb{E}\left[\psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})\mathcal{E}^v(\vec{\mathbf{U}})\right].\end{aligned}$$

By Assumption 8, we have $\mathbb{E}\left[\left\{\mathcal{E}^v(\vec{\mathbf{U}}) - \psi^v_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})\right\}\mathcal{E}^v(\vec{\mathbf{U}})\right] = 0$, hence

$$\sigma^2_{\mathsf{SSL}_{\mathsf{DR}}} = \sigma^2_{\mathsf{SUP}_{\mathsf{DR}}} - \mathrm{Var}\left[\mathcal{E}^v(\vec{\mathbf{U}})\right].$$

■

D.2.1 Variance Estimation for $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$

As discussed in Remark 10, to estimate standard errors for $V_{\mathsf{SSL}_{\mathsf{DR}}}(\vec{\mathbf{U}};\bar{\boldsymbol{\Theta}})$, we will approximate the derivatives of the expectation terms $\frac{\partial}{\partial\boldsymbol{\Theta}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\boldsymbol{\Theta}})d\mathbb{P}_{\mathbf{L}}$ using kernel smoothing to replace the indicator functions. In particular, let $\mathbb{K}_h(x) = \frac{1}{h}\sigma(x/h)$, with σ defined as in (4), we approximate $d_t(\mathbf{H}_t,\boldsymbol{\theta}_t) = I(\mathbf{H}_{t1}^{\intercal}\boldsymbol{\gamma}_t > 0)$ with $\mathbb{K}_h(\mathbf{H}_{t1}^{\intercal}\boldsymbol{\gamma}_t)$ $t = 1, 2$, and define the smoothed propensity score weights as

$$\begin{aligned}\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta}) &\equiv \frac{A_1\mathbb{K}_h(\mathbf{H}_{11}^{\intercal}\boldsymbol{\gamma}_1)}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{\{1-A_1\}\{1-\mathbb{K}_h(\mathbf{H}_{11}^{\intercal}\boldsymbol{\gamma}_1)\}}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}, \quad \text{and}\\ \tilde{\omega}_2(\check{\mathbf{H}}_2, A_2, \boldsymbol{\Theta}) &\equiv \tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta})\left[\frac{A_2\mathbb{K}_h(\mathbf{H}_{21}^{\intercal}\boldsymbol{\gamma}_2)}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} + \frac{\{1-A_2\}\{1-\mathbb{K}_h(\mathbf{H}_{21}^{\intercal}\boldsymbol{\gamma}_2)\}}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right].\end{aligned}$$

For simplicity we'll set $h = 1$, the derivatives are as follows:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}}\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}Q_1^o(\mathbf{H}_1;\boldsymbol{\theta}_1) &+ \left\{\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta})\right\}\left[Y_2 - \left\{Q_1^o(\mathbf{H}_1,\boldsymbol{\theta}_1) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right] \\ &+\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta})\left[-\frac{\partial}{\partial\boldsymbol{\theta}}Q_1^o(\mathbf{H}_1,\boldsymbol{\theta}_1) + \frac{\partial}{\partial\boldsymbol{\theta}}Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right] \\ &+\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\omega}_2(\check{\mathbf{H}}_2, A_2, \boldsymbol{\Theta})\right\}\left[Y_3 - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right] \\ &-\tilde{\omega}_2(\check{\mathbf{H}}_2, A_2, \boldsymbol{\Theta})\frac{\partial}{\partial\boldsymbol{\theta}}Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2),\end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}}Q_1^o(\mathbf{H}_1;\boldsymbol{\theta}_1) &= [\mathbf{H}_{10}^\intercal, \mathbf{H}_{11}^\intercal I\left(\mathbf{H}_{11}^\intercal\boldsymbol{\gamma}_1 > 0\right), \mathbf{0}^\intercal]^\intercal, \\ \frac{\partial}{\partial\boldsymbol{\theta}}Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2) &= [\mathbf{0}^\intercal, \mathbf{H}_{20}^\intercal, \mathbf{H}_{21}^\intercal I\left(\mathbf{H}_{21}^\intercal\boldsymbol{\gamma}_2 > 0\right)]^\intercal, \\ \frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta}) &= \left\{\frac{A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} - \frac{1-A_1}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\} \\ &\quad\times [\mathbf{0}^\intercal, \mathbf{H}_{11}^\intercal\mathbb{K}_h(\mathbf{H}_{11}^\intercal\boldsymbol{\gamma}_1)\{1-\mathbb{K}_h(\mathbf{H}_{11}^\intercal\boldsymbol{\gamma}_1)\}, \mathbf{0}^\intercal]^\intercal \\ \frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\omega}_2(\check{\mathbf{H}}_2, A_2, \boldsymbol{\Theta}) &= \frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta})\left\{\frac{A_2 d_2(\mathbf{H}_2;\boldsymbol{\theta}_2)}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} + \frac{\{1-A_2\}\{1-d_2(\mathbf{H}_2;\boldsymbol{\theta}_2)\}}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\} \\ +\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta})&\left[\mathbf{0}^\intercal, \mathbf{H}_{21}^\intercal\mathbb{K}_h(\mathbf{H}_{21}^\intercal\boldsymbol{\gamma}_2)(1-\mathbb{K}_h(\mathbf{H}_{21}^\intercal\boldsymbol{\gamma}_2))\left\{\frac{A_2}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} - \frac{1-A_2}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\}\right]^\intercal.\end{aligned}$$

Next we have

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\xi}}\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\boldsymbol{\Theta}) = &\left\{\frac{\partial}{\partial\boldsymbol{\xi}}\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta})\right\}\left[Y_2 - \left\{Q_1^o(\mathbf{H}_1,\boldsymbol{\theta}_1) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right] \\ &+\left\{\frac{\partial}{\partial\boldsymbol{\xi}}\tilde{\omega}_2(\check{\mathbf{H}}_2, A_2, \boldsymbol{\Theta})\right\}\left[Y_3 - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right],\end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\xi}}\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta}) &= \left[\varpi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)^\intercal, \mathbf{0}^\intercal\right]^\intercal, \\ \frac{\partial}{\partial\boldsymbol{\xi}}\tilde{\omega}_2(\check{\mathbf{H}}_2, A_2, \boldsymbol{\Theta}) &= \left[\tilde{\omega}_1(\check{\mathbf{H}}_1, A_1, \boldsymbol{\Theta})\varpi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)^\intercal, \mathbf{0}^\intercal\right]^\intercal, \\ \varpi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t) &\equiv \check{\mathbf{H}}_{t1}\{1-d_t(\mathbf{H}_t,\boldsymbol{\theta}_t)\}\{1-A_t\}\frac{\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)}{1-\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)} \\ &\quad - \check{\mathbf{H}}_{t1}d_t(\check{\mathbf{H}}_t,\boldsymbol{\theta}_t)A_t\frac{1-\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)}{\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)}.\end{aligned}$$

Appendix E. Technical Lemmas

We start with a simple Lemma that will save us some algebra:

Lemma 14 *For a fixed ℓ, let $\mathbf{X} \in \mathbb{R}^\ell$ be a random bounded vector and functions $g_1(\mathbf{X}), g_2(\mathbf{X})$ be measurable functions of $\mathbf{X}$. Let $\mathbb{S}_n = \{\mathbf{X}\}_{i=1}^n$ be an i.i.d. sample, and $\hat{g}_1(\cdot)$, $\hat{g}_2(\cdot)$ be the estimators for functions g_1, $g_2 \in \mathbb{R}$ respectively with $\sup_{\mathbf{X}} |g_1(\mathbf{X})|$, $\sup_{\mathbf{X}} |g_2(\mathbf{X})|$, $\sup_{\mathbf{X}} |\hat{g}_1(\mathbf{X})|$, $\sup_{\mathbf{X}} |\hat{g}_2(\mathbf{X})| < \kappa$ for fixed $\kappa \in \mathbb{R}$. If $\mathbb{P}_n\{\hat{g}_k - g_k\} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, for $k = 1, 2$, then $\mathbb{P}_n\{\hat{g}_1\hat{g}_2 - g_1g_2\} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.*

Proof [Proof of Lemma 14] By definition, $\mathbb{P}_n\{\hat{g}_1\hat{g}_2 - g_1g_2\} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ if and only if for a given any $\epsilon > 0$, $\exists M_\epsilon > 0$ such that
$\mathbb{P}\left(|\mathbb{P}_n\{\hat{g}_1\hat{g}_2 - g_1g_2\}| > M_\epsilon n^{-\frac{1}{2}}\right) \leq \epsilon \ \forall n$. Let $M_\epsilon > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(|\mathbb{P}_n\{g_1g_2 - g_1g_2\}| > M_\epsilon n^{-\frac{1}{2}}\right) \\
=&\mathbb{P}\left(|\mathbb{P}_n\{\hat{g}_1\hat{g}_2 - \hat{g}_1g_2 + \hat{g}_1g_2 - g_1g_2\}| > M_\epsilon n^{-\frac{1}{2}}\right) \\
&\leq \mathbb{P}\left(|\mathbb{P}_n\{\hat{g}_1(\hat{g}_2 - g_2)\}| + |\mathbb{P}_n\{g_2(\hat{g}_1 - g_1)\}| > M_\epsilon n^{-\frac{1}{2}}\right) \\
&\leq \mathbb{P}\left(\sup_{\mathbf{X}} |\hat{g}_1(\mathbf{X})||\mathbb{P}_n\{\hat{g}_2 - g_2\}| + \sup_{\mathbf{X}} |g_2(\mathbf{X})||\mathbb{P}_n\{\hat{g}_1 - g_1\}| > M_\epsilon n^{-\frac{1}{2}}\right)
\end{aligned}
$$

which follows from bounded functions, the union bound, now since $\mathbb{P}_n\{\hat{g}_k(\mathbf{X}) - g_k(\mathbf{X})\} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, $k = 1, 2$, there exists $M_\epsilon > 0$ such that

$$
\mathbb{P}\left(|\mathbb{P}_n\{\hat{g}_2 - g_2\}| > M_\epsilon n^{-\frac{1}{2}}\frac{1}{\kappa}\right) + \mathbb{P}\left(|\mathbb{P}_n\{\hat{g}_1 - g_1\}| > M_\epsilon n^{-\frac{1}{2}}\frac{1}{\kappa}\right) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
$$

■

Lemma 15 *(Lemma (A.1) (a) in Chakrabortty et al. (2018))*
Let $\mathbf{X} \in \mathbb{R}^\ell$ be any random vector and $g(\mathbf{X}) \in \mathbb{R}^\ell$ be any measurable function of $\mathbf{X}$, with ℓ and d fixed. Let $\mathbb{S}_n = \{\mathbf{X}\}_{i=1}^n, \mathbb{S}_N = \{\mathbf{X}\}_{j=1}^N$ be two random samples of n and N i.i.d observations of $\mathbf{X}$ respectively, such that $\mathbb{S}_n \perp\!\!\!\perp \mathbb{S}_N$. Let $\hat{g}_n(\cdot)$ be any estimator of $g(\cdot)$ estimated with $\mathbb{S}_n$ such that the random sequence: $\hat{T}_n = \sup_{x \in \mathcal{X}} \|\hat{g}_n(\cdot)\| = O_{\mathbb{P}}(1)$, where $\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^\ell$. Further define the following random sequences: $\hat{\boldsymbol{G}}_{n,N} \equiv \frac{1}{N}\sum_{j=1}^N \hat{g}_n(\mathbf{X}_j)$, and $\bar{\boldsymbol{G}}_n \equiv \mathbb{E}_{\mathbb{S}_N}\left[\hat{G}_{n,N}\right] = \mathbb{E}_{\mathbf{X}}\left[\hat{g}_n(\mathbf{X})\right]$, where $\mathbb{E}_{\mathbf{X}}$ is the expectation with respect to $\mathbf{X} \in \mathbb{S}_N$. We assume all expectations involved are finite almost surely (a.s.) $\mathbb{S}_n \ \forall n$. Then $G_{n,N} - \bar{G}_n = O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)$.

Proof [Proof of lemma 15]

The following proof follows similar arguments to Chakrabortty et al. (2018). Let $\mathcal{G}_{n,N}$, $\bar{\mathcal{G}}_n$ be the j^{th} element of $\hat{\boldsymbol{G}}_{n,N}$ and $\bar{\boldsymbol{G}}_n$ respectively, with $j \in \{1, \ldots, \ell\}$. We show that $\mathcal{G}_{n,N} - \bar{\mathcal{G}}_n = O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right)$, which implies Lemma 15 for any ℓ dimensional $\hat{\boldsymbol{G}}_{n,N}, \bar{\boldsymbol{G}}_n$. Denote by

$\mathbb{P}_{\mathbb{S}_n}$, $\mathbb{P}_{\mathbb{S}_n,\mathbb{S}_N}$ denote the joint probability distributions of samples $\mathbb{S}_n$ and $\mathbb{S}_n, \mathbb{S}_N$ respectively. Further let $\mathbb{E}_{\mathbb{S}_n}[\cdot]$ denote the expectation with respect to $\mathbb{S}_n$. Since $\mathbb{S}_n \perp\!\!\!\perp \mathbb{S}_N$ using Hoeffding's inequality

$$\mathbb{P}_{\mathbb{S}_N}\left(\left|\hat{\mathcal{G}}_{n,N} - \hat{\mathcal{G}}_n\right| > N^{-\frac{1}{2}}t \middle| \mathbb{S}_n\right) \leq 2\exp\left(-\frac{2N^2t^2}{4N^2\hat{T}_n^2}\right) \text{ a.s. } \mathbb{P}_{\mathbb{S}_n}.$$

Also, as $\mathbb{S}_n \perp\!\!\!\perp \mathbb{S}_N$ we have

$$\mathbb{P}_{\mathbb{S}_n,\mathbb{S}_N}\left[\left|\hat{\mathcal{G}}_{n,N} - \hat{\mathcal{G}}_n\right| > N^{-\frac{1}{2}}t\right] = \mathbb{E}_{\mathbb{S}_n}\left[\mathbb{P}_{\mathbb{S}_N}\left\{\left|\hat{\mathcal{G}}_{n,N} - \hat{\mathcal{G}}_n\right| > N^{-\frac{1}{2}}t \middle| \mathbb{S}_n\right\}\right].$$

Next, we have that $\hat{T}_n = \sup_{x\in\mathcal{X}} \|\hat{g}_n(\cdot)\| = O_{\mathbb{P}}(1)$ and is non-negative, thus $\forall \epsilon > 0$ $\exists\, \delta(\epsilon) > 0$ such that
$\mathbb{P}_{\mathbb{S}_n}\left(\hat{T}_n > \delta(\epsilon)\right) < \epsilon/4$, using the above we have that $\forall\ n, N$:

$$\begin{aligned}
&\mathbb{P}_{\mathbb{S}_n,\mathbb{S}_N}\left(\left|\hat{\mathcal{G}}_{n,N} - \hat{\mathcal{G}}_n\right| > N^{-\frac{1}{2}}t\right) \leq \mathbb{E}_{\mathbb{S}_n}\left[2\exp\left(-\frac{2N^2t^2}{4N^2\hat{T}_n^2}\right)\right] \\
=&\mathbb{E}_{\mathbb{S}_n}\left[2\exp\left(-\frac{t^2}{2\hat{T}_n^2}\right)\right] = \mathbb{E}_{\mathbb{S}_n}\left[2\exp\left(-\frac{t^2}{2\hat{T}_n^2}\right)\left(I\{\hat{T}_n > \delta(\epsilon)\} + I\{\hat{T}_n \leq \delta(\epsilon)\}\right)\right] \\
&\leq 2\mathbb{P}_{\mathbb{S}_n}\left(\hat{T}_n < \delta(\epsilon)\right) + 2\exp\left(-\frac{t^2}{2\delta^2(\epsilon)}\right)\mathbb{P}_{\mathbb{S}_n}\left(\hat{T}_n > \delta(\epsilon)\right) \leq 2\exp\left(-\frac{t^2}{2\delta^2(\epsilon)}\right) + \frac{\epsilon}{2} \leq \frac{2\epsilon}{2} = \epsilon,
\end{aligned}$$

where the last step follows from choosing t large enough such that $\exp\left(-\frac{t^2}{2\delta^2(\epsilon)}\right) \leq \epsilon/4$. ■

For Assumption 9 and Lemma 16 we first define some notation and set up the problem. Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) \in \mathbb{R}^{\ell_1+\ell_2}$ be any random vector and $g(\mathbf{X}_1) \in \mathbb{R}$ be any measurable function of $\mathbf{X}_1 \in \mathbb{R}^{\ell_1}$ with ℓ_1, ℓ_2 fixed. Suppose we're interested in estimating $m(\mathbf{X}_2) = \mathbb{E}[g(\mathbf{X}_1)|\mathbf{X}_2]$. Let $\mathbb{S}_n = \{\mathbf{X}\}_{i=1}^n$ be a random sample of n i.i.d. observations of $\mathbf{X}$, and $\mathbb{S}_{k=1}^K$ denote a random partition of $\mathbb{S}_n$ into K disjoint subsets of size $n_K = \frac{n}{K}$ with index sets $\{\mathcal{I}_k\}_{k=1}^K$. We will use cross-validation to estimate $\widehat{m}(\mathbf{X}_2)$, that is, we use subset $\mathcal{I}_k$ to train estimator $\widehat{m}_k$ and we estimate $m(\mathbf{X}_2)$ with: $\widehat{m}(\mathbf{X}_2) = K^{-1}\sum_{k=1}^K \sum_{i\in\mathcal{I}_k} \widehat{m}_k(\mathbf{X}_2)$, $K \geq 2$. Denote by $\hat{C}_{n,N} \in \mathbb{R}$ an estimator which depends on both samples $\mathbb{S}_n, \mathbb{S}_N$. Additionally, let function $\hat{\pi}_n(\cdot) : \mathbb{R}^{\ell_2} \to (0,1)$ be a random function with limit $\pi(\cdot)$, $\hat{l}_n(\mathbf{X}_2) : \mathbb{R}^{\ell_2} \to \{0,1\}$, be a random function with limit $l(\mathbf{X}_2)$, and finally function $f : \mathbb{R}^{\ell_2} \to \mathbb{R}^d$, $d \leq \ell_2$ be any deterministic function of $\mathbf{X}_2$.

Assumption 9 *Let $\mathcal{X} \subset \mathbb{R}^p$ for an arbitrary $p \in \mathbb{N}$ i) function $w : \mathcal{X} \mapsto \mathbb{R}$ and estimator $\hat{\pi}_n$ are such that $\sup_{\mathbf{X}_2}\left|\hat{\pi}_n(\mathbf{X}_2)^{-1} - \pi(\mathbf{X}_2)^{-1}\right| = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, ii) function $l : \mathcal{X} \mapsto \{0,1\}$ and estimator $\hat{l}_n$ are such that $\sup_{\mathbf{X}_2}\left|\hat{l}_n(\mathbf{X}_2) - l(\mathbf{X}_2)\right| = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, and iii) function $f : \mathbb{R}^{\ell_2} \to \mathbb{R}^d$, $d \leq \ell_2$ is such that $\sup_{\mathbf{X}_2}\|f(\mathbf{X}_2)\| < \infty$.*

Lemma 16 *Define* $\hat{\boldsymbol{G}}_k^n(\mathbf{X}_2) = \hat{C}_{n,N}\frac{\hat{l}_n(\mathbf{X}_2)}{\hat{\pi}_n(\mathbf{X}_2)}f(\mathbf{X}_2)\hat{\Delta}_k(\mathbf{X}_2) - \mathbb{E}\left[\frac{l(\mathbf{X}_2)}{\pi(\mathbf{X}_2)}f(\boldsymbol{x}_2)\hat{\Delta}_k(\mathbf{X}_2)\right]$ *for* $\hat{\Delta}_k(\mathbf{X}_2) = \widehat{m}_k(\mathbf{X}_2) - m(\mathbf{X}_2)$, *and* $\hat{C}_{n,N} \in \mathbb{R}$ *which satisfies* $\hat{C} = 1 + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$. *Under Assumptions 6 and 9, there is* $c_{n_K^-} = o(1)$ *such that* $\mathbb{G}_{n,K} = n^{-\frac{1}{2}}\sum_{k=1}^K \sum_{i\in\mathcal{I}_k} \hat{\boldsymbol{G}}_k^n(\mathbf{X}_2) = O_{\mathbb{P}}\left(c_{n_K^-}\right)$,

Proof [Proof of Lemma 16] First we define

$$\mathcal{G}_k^{(n)} = n^{-\frac{1}{2}}\sum_{i\in\mathcal{I}_k}\frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})}f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i}) - \mathbb{E}\left[\frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})}f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i})\right],$$

for any sample subset $\mathbb{S}_K \subseteq \mathcal{L}$, let $\mathbb{P}_{\mathbb{S}_K}$ denote the joint probability distribution of $\mathbb{S}_K$, and let $\mathbb{E}_{\mathbb{S}_K}[\cdot]$ denote expectation with respect to $\mathbb{P}_{\mathbb{S}_K}$, and $\mathbb{G}_{n,K} = K^{-\frac{1}{2}}\sum_{k=1}^K \mathcal{G}_k^{(n)}$, Next by Assumption 6 we have $\widehat{d}_k \equiv \sup_{\mathbf{X}_2}\hat{\Delta}(\mathbf{X}_2) = o_{\mathbb{P}}(1)$. Finally let $B_1 = \sup_{\mathbf{X}_2}\|f(\mathbf{X}_2)\|_2 < \infty$, $B_2 < \infty$ be the upperbound to $\sup_{\mathbf{X}_2}|\pi(\mathbf{X}_2)^{-1}|, \sup_{\mathbf{X}_2}|\hat{l}_n(\mathbf{X}_2)|\sup_{\mathbf{X}_2}|\frac{\hat{l}_n(\mathbf{X}_2)}{\hat{\pi}_n(\mathbf{X}_2)}|$.

First note that

$$\begin{aligned}
&\|\mathbb{G}_{n,K}\|_2 \\
&= \left\| n^{-\frac{1}{2}}\sum_{k=1}^K\sum_{i\in\mathcal{I}_k}\hat{C}_{n,N}\frac{\hat{l}_n(\mathbf{X}_{2i})}{\hat{\pi}_n(\mathbf{X}_{2i})}f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i}) - \mathbb{E}\left[\frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})}f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i})\right]\right\|_2 \\
&\le \left\| \left(\hat{C}_{n,N} - 1\right) n^{-\frac{1}{2}}\sum_{k=1}^K\sum_{i\in\mathcal{I}_k} f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i})\frac{\hat{l}_n(\mathbf{X}_{2i})}{\hat{\pi}_n(\mathbf{X}_{2i})}\right\|_2 \\
&+ \left\| n^{-\frac{1}{2}}\sum_{k=1}^K\sum_{i\in\mathcal{I}_k} f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i})\hat{l}_n(\mathbf{X}_{2i})\left(\frac{1}{\hat{\pi}_n(\mathbf{X}_{2i})} - \frac{1}{\pi(\mathbf{X}_{2i})}\right)\right\|_2 \\
&+ \left\| n^{-\frac{1}{2}}\sum_{k=1}^K\sum_{i\in\mathcal{I}_k} f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i})\frac{1}{\pi(\mathbf{X}_{2i})}\left(\hat{l}_n(\mathbf{X}_{2i}) - l(\mathbf{X}_{2i})\right)\right\|_2 \\
&+ \left\| n^{-\frac{1}{2}}\sum_{k=1}^K\sum_{i\in\mathcal{I}_k}\frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})}f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i}) - \mathbb{E}\left[\frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})}f(\mathbf{X}_{2i})\hat{\Delta}_k(\mathbf{X}_{2i})\right]\right\|_2,
\end{aligned}$$

which follows from the triangle inequality, next as $f(\cdot), \hat{\pi}_n(\cdot)^{-1}, \pi(\cdot)^{-1}, \hat{l}_n(\cdot)$ are bounded $\forall \mathbf{X}_2 \in \mathcal{X}$, and using uniform bounds of $O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ for the difference terms we have

$$\begin{aligned}
\|\mathbb{G}_{n,K}\|_2 \leq & O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) n^{\frac{1}{2}} B_1 B_2 \left|\sum_{k=1}^{K} \widehat{d}_k\right| + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) n^{\frac{1}{2}} B_1 B_2 \left|\sum_{k=1}^{K} \widehat{d}_k\right| \\
& + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) n^{\frac{1}{2}} B_1 B_2 \left|\sum_{k=1}^{K} \widehat{d}_k\right| + \left\|\frac{1}{K}\sum_{k=1}^{K} \mathcal{G}_k^{(n)}\right\|_2, \\
\leq & \left\| n^{-\frac{1}{2}} \frac{1}{K} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})} f(\mathbf{X}_{2i}) \hat{\Delta}_k(\mathbf{X}_{2i}) - \mathbb{E}\left[\frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})} f(\mathbf{X}_{2i}) \hat{\Delta}_k(\mathbf{X}_{2i})\right] \right\|_2 + o_{\mathbb{P}}(1).
\end{aligned}$$

where the last step follows from $\widehat{d}_k = o_{\mathbb{P}}(1)$. Next we want to bound the first term above by $c_{n_K^-}$ in probability, note that $\forall \epsilon \; \exists M > 0$ such that

$$\begin{aligned}
& \mathbb{P}\left(\left\|\sum_{k=1}^{K} \mathcal{G}_k^{(n)}\right\|_2 > M c_{n_K^-}\right) \leq \mathbb{P}\left(K^{-\frac{1}{2}} \left\|\sum_{k=1}^{K} \mathcal{G}_k^{(n)}\right\|_2 > M c_{n_K^-}\right) \\
& \leq \sum_{k=1}^{K} \mathbb{P}\left(\left\|\mathcal{G}_k^{(n)}\right\|_2 > \frac{M c_{n_K^-}}{K^{\frac{1}{2}}}\right) \leq \sum_{k=1}^{K}\sum_{j=1}^{d} \mathbb{P}\left(\left|\mathcal{G}_{k[j]}^{(n)}\right| > \frac{M c_{n_K^-}}{(Kd)^{\frac{1}{2}}}\right) \\
& \leq \sum_{k=1}^{K}\sum_{j=1}^{d} \mathbb{E}_{\mathcal{L}_k^-}\left[\mathbb{P}_{\mathcal{L}_k}\left(\left|\mathcal{G}_{k[j]}^{(n)}\right| > \frac{M c_{n_K^-}}{(Kd)^{\frac{1}{2}}} \middle| \mathcal{L}_k^-\right)\right],
\end{aligned}$$

where the first 3 steps follow from applying Boole's inequality and the triangle inequality, the fourth step follows from iterated expectations for the the event $\left\{\left|\mathcal{G}_{k[j]}^{(n)}\right| > \frac{M c_{n_K^-}}{(Kd)^{\frac{1}{2}}}\right\}$.

Next, we have $\mathcal{L}_k^- \perp\!\!\!\perp \mathcal{L}_k$, $\forall\, k \in \{1, \ldots, K\}$, thus conditional on $\mathcal{L}_k^-$, $n^{\frac{1}{2}}\mathcal{G}_k^{(n)}$ is a sum of iid centered random vectors $\left\{\frac{l(\mathbf{X}_{2i})}{\pi(\mathbf{X}_{2i})} f(\mathbf{X}_{2i}) \hat{\Delta}_k(\mathbf{X}_{2i})\right\}_{i \in \mathcal{I}_k}$ which are bounded a.s. $\mathbb{P}_{\mathcal{L}}^-$, $\forall k, n$. Thus we can apply Hoeffding's inequality to $\mathcal{G}_{k[j]}^{(n)} \forall j$:

$$\mathbb{P}_{\mathcal{L}_k}\left(\left|\mathcal{G}_{k[j]}^{(n)}\right| > \frac{M c_{n_K^-}}{(Kd)^{\frac{1}{2}}} \middle| \mathcal{L}_k^-\right) \leq 2\exp\left\{-\frac{M^2 c_{n_K^-}^2}{2KdB^2\widehat{d}_k^2}\right\} \tag{20}$$

a.s. $\mathbb{P}_{\mathcal{L}_k^-} \forall n$; and for each $k \in \{1, \ldots, K\}, j \in \{1 \ldots, d\}$. Note that $\frac{c_{n_K^-}}{D_k} \geq 0$ is stochastically bounded away from zero as $\widehat{d}_k = o_{\mathbb{P}}(1)$, therefore $\forall k$ and given $\epsilon > 0$, $\exists \delta(\epsilon, k) > 0$ such that $\mathbb{P}_{\mathcal{L}_k^-}\left(\frac{c_{n_K^-}}{D_k} \leq \delta(\epsilon, k)\right) \leq \frac{\epsilon}{4Kd}$, let $\delta^*(\epsilon, k) = \min_k\{\delta(\epsilon, k)\}$, we have that $\mathbb{P}_{\mathcal{L}_k^-}\left(\frac{c_{n_K^-}}{D_k} \leq \delta^*(\epsilon, k)\right) \leq \frac{\epsilon}{4Kd}$.

Therefore using the bound in (20) and event $\left\{\frac{c_{n_K^-}}{D_k} \leq \delta^*(\epsilon,k)\right\}$:

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{k=1}^{K}\mathcal{G}_k^{(n)}\right\|_2 > Mc_{n_K^-}\right)\\
&\leq \sum_{k=1}^{K}\sum_{j=1}^{d}\mathbb{E}_{\mathcal{L}_k^-}\left[\mathbb{P}_{\mathcal{L}_k}\left(\left|\mathcal{G}_{k[j]}^{(n)}\right| > \frac{Mc_{n_K^-}}{(Kd)^{\frac{1}{2}}}\middle|\mathcal{L}_k^-\right)\right]\\
&\leq \sum_{k=1}^{K}\sum_{j=1}^{d}\mathbb{E}_{\mathcal{L}_k^-}\left[2\exp\left\{-\frac{M^2c_{n_K^-}^2}{2KdB^2\widehat{d}_k^2}\right\}\left(I\left\{\frac{c_{n_K^-}}{D_k}\leq\delta^*(\epsilon,k)\right\}+I\left\{\frac{c_{n_K^-}}{D_k}\delta^*(\epsilon,k)\right\}\right)\right]\\
&\leq 2Kd\exp\left\{-\frac{M^2\delta^*(\epsilon,k)^2}{2KdB^2}\right\}\mathbb{P}_{\mathcal{L}_k^-}\left(\frac{c_{n_K^-}}{D_k}\leq\delta^*(\epsilon,k)\right)+2Kd\mathbb{P}_{\mathcal{L}_k^-}\left(\frac{c_{n_K^-}}{D_k}>\delta^*(\epsilon,k)\right)\\
&\leq 2Kd\frac{\epsilon}{4Kd}+2Kd\exp\left\{-\frac{M^2\delta^*(\epsilon,k)^2}{2KdB^2}\right\}\mathbb{P}_{\mathcal{L}_k^-}\left(\frac{c_{n_K^-}}{D_k}>\delta^*(\epsilon,k)\right),
\end{aligned}$$

next note that choosing a large enough M such that $\exp\left\{-\frac{M^2\delta^*(\epsilon,k)^2}{2KdB^2}\right\} < \frac{\epsilon}{4Kd}$, since $\mathbb{P}_{\mathcal{L}_k^-}\left(\frac{c_{n_K^-}}{D_k}>\delta^*(\epsilon,k)\leq 1\right)$ we get $\mathbb{P}\left(\left\|\sum_{k=1}^{K}\mathcal{G}_k^{(n)}\right\|_2 > Mc_{n_K^-}\right)\leq\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon$.
Finally we have

$$\mathbb{G}_{n,K} = O_{\mathbb{P}}\left(c_{n_K^-}\right)+o_{\mathbb{P}}(1) = O_{\mathbb{P}}\left(c_{n_K^-}\right).$$

■

Lemma 17 *Let $\widehat{\boldsymbol{\gamma}}\in\mathbb{R}^d$ be a random variable such that $\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}-\bar{\boldsymbol{\gamma}}\right)=O_{\mathbb{P}}(1)$, then for any fixed vector $\mathbf{a}\in\mathbb{R}^d$ we have that (a) $\sqrt{n}\left([\mathbf{a}^\intercal\widehat{\boldsymbol{\gamma}}]_+-[\mathbf{a}^\intercal\bar{\boldsymbol{\gamma}}]_+\right)=\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}-\bar{\boldsymbol{\gamma}}\right)I(\mathbf{a}^\intercal\bar{\boldsymbol{\gamma}}>0)+o_{\mathbb{P}}(1)$, (b) Functions $\widehat{d}_t$ $t=1,2$, defined in Section 4 and propensity scores π_1 in (4) satisfy*

$$\begin{aligned}
\sup_{\mathbf{H}_1,\mathbf{a}_1}\left|I(\widehat{d}_1=A_1)-I(\bar{d}_1=A_1)\right| &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right),\\
\sup_{\mathbf{H}_2,\mathbf{a}_2}\left|I(\widehat{d}_1=A_1)I(A_2=\widehat{d}_2)-I(\bar{d}_1=A_1)I(\bar{d}_2=A_2)\right| &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right),\\
\sup_{\mathbf{H}_1}\left|\frac{1}{\pi_1(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1)}-\frac{1}{\pi_1(\mathbf{H}_1;\bar{\boldsymbol{\xi}}_1)}\right| &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).
\end{aligned}$$

(c) For $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\xi}}$ estimated via our semi-supervised approach, and limits $\bar{\boldsymbol{\theta}}$, $\bar{\boldsymbol{\xi}}$ defined in Assumptions 3 and 5 respectively

$$\hat{C}_{n,N}^{(1)}=\frac{(1+\hat{\beta}_{21})\mathbb{P}_N\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\boldsymbol{\Theta}}_1)\right\}}{(1+\hat{\beta}_{21})\mathbb{P}_n\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\widehat{\boldsymbol{\Theta}}_1)\right\}},\ \hat{C}_{n,N}^{(2)}=\frac{\mathbb{P}_N\left\{Q_{2-}^o(\mathbf{H}_2,A_2;\bar{\boldsymbol{\theta}}_2)\right\}}{\mathbb{P}_n\left\{Q_{2-}^o(\mathbf{H}_2,A_2;\widehat{\boldsymbol{\theta}}_2)\right\}},$$

satisfy $\hat{C}_{n,N}^{(1)}=1+O_{\mathbb{P}}(n^{-\frac{1}{2}})$, $\hat{C}_{n,N}^{(2)}=1+O_{\mathbb{P}}(n^{-\frac{1}{2}})$.

Proof [Proof of Lemma 17]

Define set $\mathcal{A}_q$ for any q dimensional vector $\hat{\boldsymbol{\gamma}}$ as

$$\mathcal{A}_q = \left\{ \widehat{\boldsymbol{\gamma}} \in \mathbb{R}^q \;\middle|\; \frac{1}{2}\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}} < \mathbf{a}^\intercal \widehat{\boldsymbol{\gamma}} < 2\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}},\ \forall \mathbf{a} \in \mathbb{R}^q \right\}.$$

Now consider $\widehat{\boldsymbol{\gamma}} \in \mathcal{A}_q$:

- if $\text{sign}(\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}}) = 1$, then $0 < \frac{1}{2}\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}} < \mathbf{a}^\intercal \widehat{\boldsymbol{\gamma}} \implies \text{sign}(\mathbf{a}^\intercal \widehat{\boldsymbol{\gamma}}) = 1$,
- if $\text{sign}(\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}}) = -1$, then $\mathbf{a}^\intercal \widehat{\boldsymbol{\gamma}} < 2\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}} < 0 \implies \text{sign}(\mathbf{a}^\intercal \widehat{\boldsymbol{\gamma}}) = -1$.

Assuming $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}) = O_{\mathbb{P}}(1)$, $\mathcal{A}_q$ exists and in fact it is such that $\mathbb{P}\left(\widehat{\boldsymbol{\gamma}} \in \mathcal{A}_q\right) \xrightarrow{p} 1$.
(a) Using the above:

$$\begin{aligned}
\sqrt{n}\left([\mathbf{a}^\intercal \widehat{\boldsymbol{\gamma}}]_+ - [\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}}]_+\right) =& \sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}\right) I(\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}} > 0) I\left(\widehat{\boldsymbol{\gamma}} \in \mathcal{A}_q\right) + \sqrt{n}\left([\mathbf{a}^\intercal \widehat{\boldsymbol{\gamma}}]_+ - [\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}}]_+\right) I\left(\widehat{\boldsymbol{\gamma}} \notin \mathcal{A}_q\right) \\
=& \sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}\right) I(\mathbf{a}^\intercal \bar{\boldsymbol{\gamma}} > 0) + o_{\mathbb{P}}(1).
\end{aligned}$$

(b) As $A_{ti} \in \{0, 1\}$, $t = 1, 2$, we can write

$$\begin{aligned}
I(\widehat{d}_1 = A_1) I(\widehat{d}_2 = A_2) =& I\left\{A_1 = I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0)\right\} I\left\{A_2 = I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0)\right\} \\
=& I\{A_1 = I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0)\} I\{A_2 = I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0)\} \\
=& A_1 A_2 I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0) \\
&+ (1 - A_1)(1 - A_2) I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 < 0) \\
&+ A_1 (1 - A_2) I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 < 0) \\
&+ (1 - A_1) A_2 I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0),
\end{aligned}$$

therefore

$$\begin{aligned}
&\left| I(\widehat{d}_1 = A_1) I(\widehat{d}_2 = A_2) - I(\bar{d}_1 = A_1) I(\bar{d}_2 = A_2) \right| \\
=& \Big| A_1 A_2 \left\{ I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 > 0) \right\} \\
&+ (1 - A_1)(1 - A_2) \left\{ I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 < 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 < 0) \right\} \\
&+ A_1 (1 - A_2) \left\{ I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 < 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 < 0) \right\} \\
&+ (1 - A_1) A_2 \left\{ I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 > 0) \right\} \Big| \\
\leq& A_1 A_2 \left| I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 > 0) \right| \\
&+ (1 - A_1)(1 - A_2) \left| I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 < 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 < 0) \right| \\
&+ A_1 (1 - A_2) \left| I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 < 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 > 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 < 0) \right| \\
&+ (1 - A_1) A_2 \left| I(\mathbf{H}_{11}^\intercal \widehat{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \widehat{\boldsymbol{\gamma}}_2 > 0) - I(\mathbf{H}_{11}^\intercal \bar{\boldsymbol{\gamma}}_1 < 0) I(\mathbf{H}_{21}^\intercal \bar{\boldsymbol{\gamma}}_2 > 0) \right|
\end{aligned}$$

where the first step follows from above, the second step from the triangle inequality, now as $\widehat{\boldsymbol{\gamma}}_1$, $\widehat{\boldsymbol{\gamma}}_2$ have dimensions q_{12}, q_{22} respectively, we use sets $\mathcal{A}_{q_{12}}$, $\mathcal{A}_{q_{22}}$ and have

$$\begin{aligned}
&\left|I(\widehat{d}_1 = A_1)I(\widehat{d}_2 = A_2) - I(\bar{d}_1 = A_1)I(\bar{d}_2 = A_2)\right| \\
\leq& A_1A_2I(\widehat{\boldsymbol{\gamma}}_1 \notin \mathcal{A}_{q_{12}})I(\widehat{\boldsymbol{\gamma}}_2 \notin \mathcal{A}_{q_{22}}) + (1-A_1)(1-A_2)I(\widehat{\boldsymbol{\gamma}}_1 \notin \mathcal{A}_{q_{12}})I(\widehat{\boldsymbol{\gamma}}_2 \notin \mathcal{A}_{q_{22}}) \\
&+A_1(1-A_2)I(\widehat{\boldsymbol{\gamma}}_1 \notin \mathcal{A}_{q_{12}})I(\widehat{\boldsymbol{\gamma}}_2 \notin \mathcal{A}_{q_{22}}) + (1-A_1)A_2I(\widehat{\boldsymbol{\gamma}}_1 \notin \mathcal{A}_{q_{12}})I(\widehat{\boldsymbol{\gamma}}_2 \notin \mathcal{A}_{q_{22}}) \\
=& I(\widehat{\boldsymbol{\gamma}}_1 \notin \mathcal{A}_{q_{12}})I(\widehat{\boldsymbol{\gamma}}_2 \notin \mathcal{A}_{q_{22}})
\end{aligned}$$

which follows from the fact that for any term within absolute value:

$$\left|I(\mathbf{H}_{11}^\intercal\widehat{\boldsymbol{\gamma}}_1 < 0)I(\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2 > 0) - I(\mathbf{H}_{11}^\intercal\bar{\boldsymbol{\gamma}}_1 < 0)I(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2 > 0)\right| = I(\widehat{\boldsymbol{\gamma}}_1 \notin \mathcal{A}_{q_{12}})I(\widehat{\boldsymbol{\gamma}}_2 \notin \mathcal{A}_{q_{22}})$$

since for $I(\mathbf{H}_{11}^\intercal\bar{\boldsymbol{\gamma}}_1 < 0)I(\mathbf{H}_{21}^\intercal\bar{\boldsymbol{\gamma}}_2 > 0) \neq I(\mathbf{H}_{11}^\intercal\widehat{\boldsymbol{\gamma}}_1 < 0)I(\mathbf{H}_{21}^\intercal\widehat{\boldsymbol{\gamma}}_2 > 0)$ both $\widehat{\boldsymbol{\gamma}}_1, \widehat{\boldsymbol{\gamma}}_2$ have to be outside sets $\mathcal{A}_{q_{12}}, \mathcal{A}_{q_{22}}$ respectively. Thus $\left|I(\widehat{d}_1 = A_1)I(\widehat{d}_2 = A_2) - I(\bar{d}_1 = A_1)I(\bar{d}_2 = A_2)\right| = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, we can analogous show that $\left|I(\widehat{d}_1 = A_1) - I(\bar{d}_1 = A_1)\right| = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\ \forall i$. Next to see $\sup_{\mathbf{H}_1}\left|\frac{1}{\pi_1(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1)} - \frac{1}{\pi_1(\mathbf{H}_1;\bar{\boldsymbol{\xi}}_1)}\right| = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, note that as $\mathcal{H}_1$, Ω_1 are bounded sets we have

$$\begin{aligned}
&\sup_{\mathbf{H}_1}\left|\frac{1}{\pi_1(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1)} - \frac{1}{\pi_1(\mathbf{H}_1;\bar{\boldsymbol{\xi}}_1)}\right| = \sup_{\mathbf{H}_1\in\mathcal{H}_1}\left|e^{-\mathbf{H}_1^\intercal\hat{\boldsymbol{\xi}}_1} - e^{-\mathbf{H}_1^\intercal\bar{\boldsymbol{\xi}}_1}\right| \\
\leq& \sup_{\mathbf{H}_1\in\mathcal{H}_1,\boldsymbol{\xi}_1\in\Omega_1}\left|\frac{d}{dx}e^{-x}\big|_{x=\mathbf{H}_1^\intercal\boldsymbol{\xi}_1}\right| \sup_{\mathbf{H}_1\in\mathcal{H}_1}\left|\mathbf{H}_1^\intercal\hat{\boldsymbol{\xi}}_1 - \mathbf{H}_1^\intercal\bar{\boldsymbol{\xi}}\right| \\
\leq& \sup_{\mathbf{H}_1\in\mathcal{H}_1,\boldsymbol{\xi}_1\in\Omega_1}\left|\frac{d}{dx}e^{-x}\big|_{x=\mathbf{H}_1^\intercal\boldsymbol{\xi}_1}\right| \sup_{\mathbf{H}_1\in\mathcal{H}_1}\|\mathbf{H}_1\|\left\|\hat{\boldsymbol{\xi}}_1 - \bar{\boldsymbol{\xi}}_1\right\|_2 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where we use the definition of π_1 in (4), Lipschitz and $\left\|\hat{\boldsymbol{\xi}}_1 - \bar{\boldsymbol{\xi}}_1\right\|_2 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ from Assumptions (5) and Theorem 5.21 in Vaart (1998) as we are using Z-estimation for $\boldsymbol{\xi}_1$.
(c) By Theorem 2 we have $\hat{\beta}_{21} - \bar{\beta}_{21} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$. Next, we can write

$$\omega_1(\mathbf{H}_1, A_1; \widehat{\boldsymbol{\Theta}}_1) = I\left\{A_1 = d_1\left(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1\right)\right\}\left\{\frac{A_1}{\pi_1\left(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1\right)} + \frac{1-A_1}{1-\pi_1\left(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1\right)}\right\}.$$

By Lemma 17 (b) it follows that

$$\begin{aligned}
\mathbb{P}_n\left[I\left\{A_1 = d_1\left(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1\right)\right\} - I\left\{A_1 = d_1\left(\mathbf{H}_1;\bar{\boldsymbol{\xi}}_1\right)\right\}\right] &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right), \\
\mathbb{P}_n\left[\frac{A_1}{\pi_1(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1)} - \frac{A_1}{\pi_1\left(\mathbf{H}_1;\bar{\boldsymbol{\xi}}_1\right)}\right] &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right), \\
\mathbb{P}_n\left[\frac{1-A_1}{1-\pi_1(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_1)} - \frac{1-A_1}{1-\pi_1\left(\mathbf{H}_1;\bar{\boldsymbol{\xi}}_1\right)}\right] &= O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).
\end{aligned}$$

Using the above and Lemma 14 we get

$$(1+\hat{\beta}_{21})\mathbb{P}_n\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\widehat{\bar{\mathbf{\Theta}}}_1)\right\}=(1+\bar{\beta}_{21})\mathbb{P}_n\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)\right\}+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$$

Also by CLT we have

$$\begin{aligned}(1+\bar{\beta}_{21})\mathbb{P}_n\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)\right\}&=(1+\bar{\beta}_{21})\mathbb{E}\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)\right\}+O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right),\\(1+\bar{\beta}_{21})\mathbb{P}_N\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)\right\}&=(1+\bar{\beta}_{21})\mathbb{E}\left\{\omega_1(\check{\mathbf{H}}_1,A_1;\bar{\mathbf{\Theta}}_1)\right\}+O_{\mathbb{P}}\left(N^{-\frac{1}{2}}\right),\end{aligned}$$

finally by Slutsky's theorem $\hat{C}^{(1)}_{n,N}-1=O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$. With similar arguments, and using Lemma 17 (a) to see $\mathbb{P}_n\left([\mathbf{H}_{21}^{\intercal}\widehat{\gamma}_2]_+-[\mathbf{H}_{21}^{\intercal}\bar{\gamma}_2]_+\right)=O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, we can show $\hat{C}^{(2)}_{n,N}-1=O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.

■

Lemma 18 *Let $Q_t(\check{\mathbf{H}}_t;\boldsymbol{\theta}_t)$, $\pi_t(\check{\mathbf{H}}_t;\boldsymbol{\xi}_t)$ $t=1,2$ be estimator functions of (1) & (4) respectively and define the bias as Bias $\left(\bar{V},\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})\right)\equiv\bar{V}-\mathbb{E}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})\right]$, then*

$$\begin{aligned}&Bias\left(\bar{V},\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})\right)\\=&\mathbb{E}\left[\left\{1-\frac{\pi_1(\check{\mathbf{H}}_1)}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_1^o(\check{\mathbf{H}}_1)-Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right\}\right]\\+&\mathbb{E}\left[\left\{1-\frac{1-\pi_1(\check{\mathbf{H}}_1)}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_1^o(\check{\mathbf{H}}_1)-Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right\}\right]\\+&\mathbb{E}\left[\left\{\frac{A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}+\frac{1-A_1}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{1-\frac{\pi_2(\check{\mathbf{H}}_2)}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\}\left\{Q_2^o(\check{\mathbf{H}}_2)-Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right]\\+&\mathbb{E}\left[\left\{\frac{A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}+\frac{1-A_1}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{1-\frac{1-\pi_2(\check{\mathbf{H}}_2)}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\}\left\{Q_2^o(\check{\mathbf{H}}_2)-Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right].\end{aligned}$$

where $\bar{V}=\mathbb{E}[\mathbb{E}[Y_2+\mathbb{E}[Y_3|\mathbf{H}_2,Y_2,A_2=\bar{d}_2(\check{\mathbf{H}}_2)]|\mathbf{H}_1,A_1=\bar{d}_1(\check{\mathbf{H}}_1)]]$ is the mean population value under the optimal treatment rule.

Proof [Proof of Lemma 18]

$$\begin{aligned}\text{Bias}\left(\bar{V},\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})\right)=&\mathbb{E}[\mathbb{E}[Y_2+\mathbb{E}[Y_3|\mathbf{H}_2,Y_2,A_2=\bar{d}_2]|\mathbf{H}_1,A_1=\bar{d}_1]]-\mathbb{E}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})\right]\\=&\mathbb{E}\left[Q_1^o(\mathbf{H}_1)-Q_1^o(\mathbf{H}_1;\boldsymbol{\theta}_1)\right]\\-&\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1,A_1;\mathbf{\Theta}_1)\left\{Y_2-Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right\}\right]\\-&\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1,A_1;\mathbf{\Theta}_1)Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right]\\-&\mathbb{E}\left[\omega_2(\check{\mathbf{H}}_2,A_2;\mathbf{\Theta}_2)\left\{Y_3-Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right].\end{aligned}$$

Adding and subtracting $\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1) Q_2^o(\check{\mathbf{H}}_2)\right] = \mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1)\mathbb{E}[Y_3|\mathbf{H}_2, \bar{d}_2(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2), Y_2]\right]$,

$$
\begin{aligned}
&\text{Bias}\left(\bar{V}, \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \boldsymbol{\Theta})\right) \\
&= \mathbb{E}\left[Q_1^o(\mathbf{H}_1) - Q_1^o(\mathbf{H}_1; \boldsymbol{\theta}_1)\right] \\
&- \mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1)\left\{Y_2 + \mathbb{E}[Y_3|\mathbf{H}_2\bar{d}_2(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2), Y_2] - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1)\right\}\right] \\
&- \mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1)\left\{Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2) - Q_2^o(\check{\mathbf{H}}_2)\right\}\right] \\
&- \mathbb{E}\left[\omega_2(\check{\mathbf{H}}_2, A_2; \boldsymbol{\Theta}_2)\left\{Y_3 - Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2)\right\}\right],
\end{aligned}
$$

using iterated expectations in the second and fourth terms:

$$
\begin{aligned}
&\text{Bias}\left(\bar{V}, \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \boldsymbol{\Theta})\right) \\
&= \mathbb{E}\left[Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1)\right] \\
&- \mathbb{E}\left[\mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1)\left\{Y_2 + \mathbb{E}[Y_3|\check{\mathbf{H}}_2, \bar{d}_2(\check{\mathbf{H}}_2), Y_2] - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1)\right\}\middle|\check{\mathbf{H}}_1, A_1\right]\right] \\
&- \mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1)\left\{Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2) - Q_2^o(\check{\mathbf{H}}_2)\right\}\right] \\
&- \mathbb{E}\left[\mathbb{E}\left[\omega_2(\check{\mathbf{H}}_2, A_2; \boldsymbol{\Theta}_2)\left\{Y_3 - Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2)\right\}\middle|\check{\mathbf{H}}_2, A_2, Y_2\right]\right] \\
&= \mathbb{E}\left[Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1)\right] \\
&- \mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1)\left\{\mathbb{E}\left[Y_2 + \mathbb{E}[Y_3|\check{\mathbf{H}}_2, \bar{d}_2(\check{\mathbf{H}}_2), Y_2]\middle|\check{\mathbf{H}}_1, A_1\right] - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1)\right\}\right] \\
&- \mathbb{E}\left[\omega_1(\check{\mathbf{H}}_1, A_1; \boldsymbol{\Theta}_1)\left\{Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2) - Q_2^o(\check{\mathbf{H}}_2)\right\}\right] \\
&- \mathbb{E}\left[\omega_2(\check{\mathbf{H}}_2, A_2; \boldsymbol{\Theta}_2)\left\{\mathbb{E}\left[Y_3|\check{\mathbf{H}}_2, A_2, Y_2\right] - Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2)\right\}\right].
\end{aligned}
$$

using definitions of $\omega_t(\check{\mathbf{H}}_t, A_t; \boldsymbol{\Theta}_t)$ $t = 1, 2$ we can write:

$$
\begin{aligned}
&\text{Bias}\left(\bar{V}, \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \boldsymbol{\Theta})\right) \\
=&\mathbb{E}\left[Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1)\right] \\
-&\mathbb{E}\left[\left\{\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1; \boldsymbol{\xi}_1)} + \frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1; \boldsymbol{\xi}_1)}\right\}\left\{Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1; \boldsymbol{\theta}_1)\right\}\right] \\
-&\mathbb{E}\left[\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1; \boldsymbol{\xi}_1)}\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2)\right\}\right] - \mathbb{E}\left[\frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1; \boldsymbol{\xi}_1)}\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2)\right\}\right] \\
-&\mathbb{E}\Bigg[\left\{\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1; \boldsymbol{\xi}_1)} + \frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1; \boldsymbol{\xi}_1)}\right\}\left\{\frac{\bar{d}_2 A_2}{\pi_2(\check{\mathbf{H}}_2; \boldsymbol{\xi}_2)} + \frac{(1-\bar{d}_2)(1-A_2)}{1-\pi_2(\check{\mathbf{H}}_2; \boldsymbol{\xi}_2)}\right\} \\
&\times\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2; \boldsymbol{\theta}_2)\right\}\Bigg]
\end{aligned}
$$

assuming $A_1 \perp A_2 | \mathbf{H}_2, Y_2$, we use iterated expectations:

$$
\begin{aligned}
&\text{Bias}\left(\bar{V}, \mathcal{V}_{\mathsf{SUP_{DR}}}(\mathbf{L};\boldsymbol{\Theta})\right)\\
=&\mathbb{E}\left[Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right]\\
-&\mathbb{E}\left[\mathbb{E}\left[\left\{\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right\}\middle|\check{\mathbf{H}}_1\right]\right]\\
-&\mathbb{E}\left[\left\{\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right]\\
-&\mathbb{E}\Bigg[\mathbb{E}\Bigg[\left\{\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{\frac{\bar{d}_2 A_2}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} + \frac{(1-\bar{d}_2)(1-A_2)}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\}\\
&\times\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\Bigg|\check{\mathbf{H}}_2\Bigg]\Bigg]\\
=&\mathbb{E}\left[Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right]\\
-&\mathbb{E}\left[\left\{\frac{\bar{d}_1 \pi_1(\check{\mathbf{H}}_1)}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{\{1-\bar{d}_1\}\{1-\pi_1(\check{\mathbf{H}}_1)\}}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right\}\right]\\
-&\mathbb{E}\left[\left\{\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right]\\
-&\mathbb{E}\Bigg[\left\{\frac{\bar{d}_1 A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{(1-\bar{d}_1)(1-A_1)}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{\frac{\bar{d}_2 \pi_2(\check{\mathbf{H}}_2)}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)} + \frac{\{1-\bar{d}_2\}\{1-\pi_2(\check{\mathbf{H}}_2)\}}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\}\\
&\times\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\Bigg]
\end{aligned}
$$

finally, factorizing common terms:

$$
\begin{aligned}
&\text{Bias}\left(\bar{V}, \mathcal{V}_{\mathsf{SUP_{DR}}}(\mathbf{L};\boldsymbol{\Theta})\right)\\
=&\mathbb{E}\left[\bar{d}_1\left\{1 - \frac{\pi_1(\check{\mathbf{H}}_1)}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right\}\right]\\
+&\mathbb{E}\left[\{1-\bar{d}_1\}\left\{1 - \frac{1-\pi_1(\check{\mathbf{H}}_1)}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{Q_1^o(\check{\mathbf{H}}_1) - Q_1^o(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)\right\}\right]\\
+&\mathbb{E}\left[\bar{d}_2\left\{\frac{A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{1-A_1}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{1 - \frac{\pi_2(\check{\mathbf{H}}_2)}{\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\}\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right]\\
+&\mathbb{E}\left[\{1-\bar{d}_2\}\left\{\frac{A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{1-A_1}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\left\{1 - \frac{1-\pi_2(\check{\mathbf{H}}_2)}{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)}\right\}\left\{Q_2^o(\check{\mathbf{H}}_2) - Q_2^o(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\right]\\
\leq&\sqrt{\sup_{\check{\mathbf{H}}_1}|\{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)\}^{-1}|}\sqrt{\|\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1) - \pi_1(\check{\mathbf{H}}_1)\|_{L_2(\mathbb{P})}}\sqrt{\|Q_1^o(\check{\mathbf{H}}_1;\widehat{\boldsymbol{\theta}}_1) - Q_1^o(\check{\mathbf{H}}_1)\|_{L_2(\mathbb{P})}}\\
+&\sqrt{\sup_{\check{\mathbf{H}}_2}\left|\left\{\frac{A_1}{\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)} + \frac{1-A_1}{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)}\right\}\{1-\pi_1(\check{\mathbf{H}}_1;\boldsymbol{\xi}_1)\}^{-1}\{1-\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2)\}^{-1}\right|}\\
\times&\sqrt{\|\pi_2(\check{\mathbf{H}}_2;\boldsymbol{\xi}_2) - \pi_2(\check{\mathbf{H}}_2)\|_{L_2(\mathbb{P})}}\sqrt{\|Q_2^o(\check{\mathbf{H}}_2;\widehat{\boldsymbol{\theta}}_2) - Q_2^o(\check{\mathbf{H}}_2)\|_{L_2(\mathbb{P})}},
\end{aligned}
$$

which follows by Cauchy–Schwarz Inequality. ■

Appendix F. Additional Theoretical Results

F.1 Augmented Value Function Estimation

We first re-write Assumption 5 to account for only using sample $\mathcal{L}$ in estimation of the Q functions and propensity scores.

Assumption 10 *Define the following class of functions:*

$$\begin{aligned}
\mathcal{Q}_1 &\equiv \{Q_1(\mathbf{H}_1, A_1; \boldsymbol{\theta}_1) | \boldsymbol{\theta}_1 \in \Theta_1 \subset \mathbb{R}^{q_1}\}, \\
\mathcal{Q}_2 &\equiv \{Q_2(\mathbf{H}_2, A_2, Y_2; \boldsymbol{\theta}_2) | \boldsymbol{\theta}_2 \in \Theta_2 \subset \mathbb{R}^{q_2}\}, \\
\mathcal{W}_1 &\equiv \{\pi_1(\mathbf{H}_1; \boldsymbol{\xi}_1) | \boldsymbol{\xi}_1 \in \Omega_1 \subset \mathbb{R}^{p_1}\}, \\
\mathcal{W}_2 &\equiv \left\{\pi_2(\check{\mathbf{H}}_2; \boldsymbol{\xi}_2) | \boldsymbol{\xi}_2 \in \Omega_2 \subset \mathbb{R}^{p_2}\right\},
\end{aligned} \tag{21}$$

with p_1, p_2, q_1, q_2 fixed under model definitions (1) & (4). Let the population equations $\mathbb{E}\left[S_t^{\xi}(\boldsymbol{\xi}_t)\right] = \mathbf{0}, t = 1, 2$ have solutions $\bar{\boldsymbol{\xi}}_1, \bar{\boldsymbol{\xi}}_2$, where

$$\begin{aligned}
S_1^{\xi}(\boldsymbol{\xi}_1) =& \frac{\partial}{\partial \boldsymbol{\xi}_1} \log\left[\pi_1(\mathbf{H}_1; \boldsymbol{\xi}_1)^{A_1}(1 - \pi_1(\mathbf{H}_1; \boldsymbol{\xi}_1))^{(1-A_1)}\right], \\
S_2^{\xi}(\boldsymbol{\xi}_2) =& \frac{\partial}{\partial \boldsymbol{\xi}_2} \log\left[\pi_2(\check{\mathbf{H}}_2; \boldsymbol{\xi}_2)^{A_2}(1 - \pi_2(\check{\mathbf{H}}_2; \boldsymbol{\xi}_2))^{(1-A_2)}\right],
\end{aligned}$$

and the population equations for the Q functions $\mathbb{E}[S_t^{\theta}(\boldsymbol{\theta}_t)] = \mathbf{0}, t = 1, 2$ have solutions $\bar{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2$, where

$$\begin{aligned}
S_2^{\theta}(\boldsymbol{\theta}_2) =& \frac{\partial}{\partial \boldsymbol{\theta}_2^{\intercal}} \|Y_3 - Q_2(\check{\mathbf{H}}_2, A_2; \boldsymbol{\theta}_2)\|_2^2, \\
S_1^{\theta}(\boldsymbol{\theta}_1) =& \frac{\partial}{\partial \boldsymbol{\theta}_1^{\intercal}} \|Y_2 + \bar{Q}_2(\check{\mathbf{H}}_2; \bar{\boldsymbol{\theta}}_2) - Q_1(\mathbf{H}_1, A_1; \boldsymbol{\theta}_1)\|_2^2,
\end{aligned}$$

(i) ξ_1, ξ_2 are bounded sets. (ii) Θ_1, Θ_2 are open bounded sets and for some $r > 0$ and $g_t(\cdot)$

$$\left|Q_t(\cdot; \boldsymbol{\theta}_t) - Q_t(\cdot; \boldsymbol{\theta}_t')\right| \leq g_t(\cdot)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_t'\| \ \forall \boldsymbol{\theta}_t, \boldsymbol{\theta}_t' \in \Theta_t, \ \mathbb{E}\left[|g_t(\cdot)|^r\right] < \infty, \ t = 1, 2. \tag{22}$$

(iii) The population minimizers satisfy $\bar{\boldsymbol{\theta}}_t \in \Theta_t, \bar{\boldsymbol{\xi}}_t \in \Omega_t, t = 1, 2$. (iv) For $\bar{\boldsymbol{\xi}}_t, t = 1, 2$, $\bar{\pi}_1(\mathbf{H}_1; \bar{\boldsymbol{\xi}}_1) > 0, \bar{\pi}_2(\check{\mathbf{H}}_2; \bar{\boldsymbol{\xi}}_2) > 0 \ \forall \mathbf{H} \in \mathcal{H}$.

Existence of solutions $\bar{\boldsymbol{\theta}}_t \in \Theta_t$, $t = 1, 2$ is clear as Θ_1, Θ_2 are open and bounded.

Theorem 19 (Asymptotic Normality for $\widehat{V}_{\mathsf{SSL}_{\mathsf{DR}}}$) *Under Assumptions 1, 4, and 10, $\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}}$ as defined in (5) is such that*

$$\sqrt{n}\left\{\widehat{V}_{\mathsf{SUP}_{\mathsf{DR}}} - \mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L}; \bar{\boldsymbol{\Theta}})\right]\right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\mathsf{SUP}_{\mathsf{DR}}}^{v}(\mathbf{L}_i; \bar{\boldsymbol{\Theta}}) + o_{\mathbb{P}}(1) \xrightarrow{d} N\left(0, \sigma_{\mathsf{SUP}_{\mathsf{DR}}}^2\right).$$

where

$$\begin{aligned}\psi^{v}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}}) =&\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}}) - \mathbb{E}_{\mathbb{S}}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right] + \boldsymbol{\psi}^{\theta}_{\mathsf{SUP}}(\mathbf{L})^{\intercal}\frac{\partial}{\partial\boldsymbol{\theta}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}}\\ &+\boldsymbol{\psi}^{\xi}_{\mathsf{SUP}}(\mathbf{L})^{\intercal}\frac{\partial}{\partial\boldsymbol{\xi}}\int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})d\mathbb{P}_{\mathbf{L}}\Big|_{\mathbf{\Theta}=\bar{\mathbf{\Theta}}},\\ \sigma^{2}_{\mathsf{SUP}_{\mathsf{DR}}} =&\mathbb{E}\left[\psi^{v}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})^{2}\right].\end{aligned}$$

Proof [proof of theorem 19]

Letting $g(\mathbf{\Theta}) = \int \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\mathbf{\Theta})d\mathbb{P}_{\mathbf{L}}$, we start by centering (5) and scaling by $\sqrt{n}$:

$$\begin{aligned}&\sqrt{n}\left\{\mathbb{P}_n\left(\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}})\right) - \mathbb{E}\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right]\right\}\\ =&\mathbb{G}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\} + \mathbb{G}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\} + \sqrt{n}\left\{g(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - g(\bar{\mathbf{\Theta}})\right\}\end{aligned}$$

I) Empirical Process Term

We first show that under Assumption 10, $\mathbb{G}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\} = o_{\mathbb{P}}(1)$, let

$$\begin{aligned}f_{\mathbf{\Theta}}(\vec{\mathbf{U}}) = Q^{o}_{1}(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1)+&\omega_1(\check{\mathbf{H}}_1, A_1, \mathbf{\Theta})\left\{Y_2 - Q^{o}_{1}(\check{\mathbf{H}}_1;\boldsymbol{\theta}_1) + Q^{o}_{2}(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\}\\ &+\omega_2(\check{\mathbf{H}}_1, A_1;\mathbf{\Theta})\left\{Y_3 - Q^{o}_{2}(\check{\mathbf{H}}_2;\boldsymbol{\theta}_2)\right\},\end{aligned}$$

we define the class of functions $\mathcal{C}_3 = \left\{f_{\mathbf{\Theta}}(\vec{\mathbf{U}})|\vec{\mathbf{U}}, \mathbf{\Theta} \in \mathcal{S}(\delta)\right\}$, and

$$\ell = \{l : \{0,1\}^2 \mapsto \{0,1\}\}.$$

i) By Assumptions 10 and Theorem 19.5 in Vaart (1998), ℓ, $\mathcal{W}_t$, $\mathcal{Q}_t, t = 1, 2$ are a $\mathbb{P}$-Donsker class, thus it follows that $\mathcal{C}_3$ is a Donsker class.

ii) We estimate $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ from (21) with their maximum likelihood estimator $\widehat{\boldsymbol{\xi}}_{1\mathsf{SUP}}, \widehat{\boldsymbol{\xi}}_{2\mathsf{SUP}}$, solving $\mathbb{P}_n\left[S_t(\boldsymbol{\xi}_t)\right] = \mathbf{0}, t = 1, 2$ and estimate functions $\pi_1(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_{1\mathsf{SUP}})$, $\pi_2(\check{\mathbf{H}}_2;\widehat{\boldsymbol{\xi}}_{2_{\mathsf{SUP}}})$ with $\widehat{\boldsymbol{\xi}}_{1\mathsf{SUP}}, \widehat{\boldsymbol{\xi}}_{2\mathsf{SUP}}$. By Assumption 10 and weak law of large numbers $\widehat{\boldsymbol{\xi}}_{t\mathsf{SUP}} \xrightarrow{p} \bar{\boldsymbol{\xi}}_t, t = 1, 2$.

Analogous, under regularity conditions, (2) have unique solutions $\widehat{\boldsymbol{\theta}}_{t\mathsf{SUP}}$ for which $\widehat{\boldsymbol{\theta}}_{t\mathsf{SUP}} \xrightarrow{p} \bar{\boldsymbol{\theta}}_t, t = 1, 2$ by Assumption 10 and weak law of large numbers. Both regardless of whether models (1) & (4) are correct. Thus $\mathbb{P}\left(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}} \in \mathcal{S}(\delta)\right) \to 1$, $\forall\delta$.

iii) We next show $\int \left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\}^2 d\mathbb{P}_{\mathbf{L}} \longrightarrow 0$. Using (7), for a large enough constant c we can write

$$\begin{aligned}
&\int \left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\}^2 d\mathbb{P}_{\mathbf{L}} \\
\leq & c \sup_{\mathbf{H}_1} \left(\mathbf{H}_{10}^{\mathsf{T}}\bar{\boldsymbol{\beta}}_1 + [\mathbf{H}_{11}^{\mathsf{T}}\bar{\gamma}_1]_+ - \mathbf{H}_{10}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_{1\mathsf{SUP}} - [\mathbf{H}_{11}^{\mathsf{T}}\widehat{\gamma}_{1\mathsf{SUP}}]_+\right)^2 \\
+ & c \sup_{\check{\mathbf{H}}_2} \left(\check{\mathbf{H}}_{20}^{\mathsf{T}}\bar{\boldsymbol{\beta}}_2 + [\mathbf{H}_{21}^{\mathsf{T}}\bar{\gamma}_2]_+ - \check{\mathbf{H}}_{20}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_{2\mathsf{SUP}} - [\mathbf{H}_{21}^{\mathsf{T}}\widehat{\gamma}_{2\mathsf{SUP}}]_+\right)^2 \\
+ & c \sup_{\mathbf{H}_1, A_1} \left\{\omega_1(\mathbf{H}_1, A_1;\widehat{\mathbf{\Theta}}_{1\mathsf{SUP}}) - \omega_1(\mathbf{H}_1, A_1;\bar{\mathbf{\Theta}}_1)\right\}^2 + \left(\hat{\beta}_{21\mathsf{SUP}} - \bar{\beta}_{21}\right)^2 \\
& \longrightarrow 0
\end{aligned}$$

where we use $(a-b)^2, (a+b)^2 \leq 2a^2 + 2b^2 \; \forall a, b \in \mathbb{R}$, boundedness of $\bar{\mathbf{\Theta}}$ and covariates by Assumptions 1, 2, and 10. Next,

from assumption (7) it can be shown that $\widehat{\boldsymbol{\theta}}_{2\mathsf{SUP}} - \bar{\boldsymbol{\theta}}_2 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, $\widehat{\boldsymbol{\theta}}_{1\mathsf{SUP}} - \bar{\boldsymbol{\theta}}_1 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$, also from Lemma 17 (a) it follows that for $t = 1, 2$

$$\begin{aligned}
&\sup_{\check{\mathbf{H}}_t} \left(\mathbf{H}_{t0}^{\mathsf{T}}\bar{\boldsymbol{\beta}}_t + [\mathbf{H}_{t1}^{\mathsf{T}}\bar{\gamma}_t]_+ - \mathbf{H}_{t0}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_{t\mathsf{SUP}} - [\mathbf{H}_{t1}^{\mathsf{T}}\widehat{\gamma}_{t\mathsf{SUP}}]_+\right)^2 \\
&\leq 2 \sup_{\check{\mathbf{H}}_{t0}} \|\check{\mathbf{H}}_{t0}\|_2^2 \|\widehat{\boldsymbol{\beta}}_t - \bar{\boldsymbol{\beta}}_t\|_2^2 + 2 \sup_{\mathbf{H}_{t1}} \|\mathbf{H}_{t1}\|_2^2 \|\widehat{\gamma}_t - \bar{\gamma}_t\|_2 \\
&= O_{\mathbb{P}}\left(n^{-1}\right).
\end{aligned}$$

Next, we can write

$$\omega_1(\mathbf{H}_1, A_1;\widehat{\mathbf{\Theta}}_{1\mathsf{SUP}}) = I\left\{A_1 = d_1\left(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_{1\mathsf{SUP}}\right)\right\}\left\{\frac{A_1}{\pi_1\left(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_{1\mathsf{SUP}}\right)} + \frac{1-A_1}{1-\pi_1\left(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_{1\mathsf{SUP}}\right)}\right\}.$$

By Lemma 17 (b) it follows that

$$\sup_{\mathbf{H}_1}\left|\frac{1}{\pi_1(\mathbf{H}_1;\widehat{\boldsymbol{\xi}}_{1\mathsf{SUP}})} - \frac{1}{\pi_1(\mathbf{H}_1;\bar{\boldsymbol{\xi}}_1)}\right| = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right).$$

Using the above and Lemma 14 we get

$$\sup_{\check{\mathbf{H}}_1, A_1}\left\{\omega_1(\check{\mathbf{H}}_1, A_1;\widehat{\mathbf{\Theta}}_{1\mathsf{SUP}}) - \omega_1(\check{\mathbf{H}}_1, A_1;\bar{\mathbf{\Theta}}_1)\right\}^2 = o_{\mathbb{P}}(1),$$

which gives us $\int \left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\}^2 d\mathbb{P}_{\mathbf{L}} \longrightarrow 0$.

Hence, we have i) $\mathbb{P}\left(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}} \in \mathcal{S}(\delta)\right) \to 1$, $\forall \delta$, ii) $\mathcal{C}_1$ is a Donsker class, and iii) $\int \left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - \mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\}^2 d\mathbb{P}_{\mathbf{L}} \longrightarrow 0$, then by Theorem 2.1 in Van Der Vaart and Wellner (2007)

$$\sqrt{n}\left[\mathbb{P}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - g(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}})\right\} - \mathbb{P}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}}) - g(\bar{\mathbf{\Theta}})\right\}\right] = o_{\mathbb{P}}(1).$$

Centered Sample Average

Next we consider $\mathbb{G}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\}$. Note that $\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})$ is a deterministic function of random variable $\mathbf{L}$ as parameters are fixed. We have that $\mathbb{E}\left[\left(\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}}\right)^2\right] < \infty$ holds by Assumption 1 & 10. Thus the central limit theorem yields

$$\mathbb{G}_n\left\{\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right\} \xrightarrow{d} \mathcal{N}\left(0, Var\left[\mathcal{V}_{\mathsf{SUP}_{\mathsf{DR}}}(\mathbf{L};\bar{\mathbf{\Theta}})\right]\right).$$

Bias Term

We finally analyze the bias: $\sqrt{n}\left\{g(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - g(\bar{\mathbf{\Theta}})\right\}$. Using a Taylor series expansion

$$g(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) = g(\bar{\mathbf{\Theta}}) + (\widehat{\boldsymbol{\theta}}_{\mathsf{SUP}} - \bar{\boldsymbol{\theta}})^{\intercal}\frac{\partial}{\partial \boldsymbol{\theta}_{\mathsf{SUP}}}g(\bar{\mathbf{\Theta}}) + (\widehat{\boldsymbol{\xi}}_{\mathsf{SUP}} - \bar{\boldsymbol{\xi}})^{\intercal}\frac{\partial}{\partial \boldsymbol{\xi}_{\mathsf{SUP}}}g(\bar{\mathbf{\Theta}}) + O_{\mathbb{P}}\left(n^{-1}\right),$$

therefore

$$\sqrt{n}\left\{g(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - g(\bar{\mathbf{\Theta}})\right\} = \sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathsf{SUP}} - \bar{\boldsymbol{\theta}})^{\intercal}\frac{\partial}{\partial \boldsymbol{\theta}_{\mathsf{SUP}}}g(\bar{\mathbf{\Theta}}) + \sqrt{n}(\widehat{\boldsymbol{\xi}}_{\mathsf{SUP}} - \bar{\boldsymbol{\xi}})^{\intercal}\frac{\partial}{\partial \boldsymbol{\xi}_{\mathsf{SUP}}}g(\bar{\mathbf{\Theta}}) + o_{\mathbb{P}}(1).$$

Using the Q function and propensity score function influence functions we can write

$$\sqrt{n}\left\{g(\widehat{\mathbf{\Theta}}_{\mathsf{SUP}}) - g(\bar{\mathbf{\Theta}})\right\} = \frac{\partial}{\partial \boldsymbol{\theta}_{\mathsf{SUP}}}g(\bar{\mathbf{\Theta}})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\theta}_{\mathsf{SUP}}(\mathbf{L}_i) + \frac{\partial}{\partial \boldsymbol{\xi}}g(\bar{\mathbf{\Theta}})\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{\xi}_{\mathsf{SUP}}(\mathbf{L}_i) + o_{\mathbb{P}}(1)$$

■