EFFICIENT ESTIMATION OF COUNTERFACTUAL DISTRIBUTIONS AND TESTING DISTRIBUTIONAL TREATMENT EFFECTS

A Thesis

by

LUKANG HUANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Raymond Ka Wai Wong
Committee Members,	Yonghong An
	Xianyang Zhang
Head of Department,	Jianhua Huang

August 2019

Major Subject: Statistics

ABSTRACT

This article considers efficient estimation of and inference on counterfactual distributions in a discrete (binary or multivalued) treatment. The counterfactual distributions are estimated by weighted sample averages and the distributional effects are tested by the Mann-Whitney statistics. The difference between this study and other studies in the literature is the way to estimate the weighting functings. While other studies estimate the weighting functions either parametrically or semiparametrically or nonparametrically without incoporating the restrictions on the weighting functions, we estimate the weighting functions by imposing those restrictions. As a result, our estimated counterfactual distribution functions and the Mann-Whitney statistics are efficient, attaining the semiparametric efficiency bounds which are also derived in the paper. A small scale simulation study and an application to the job training program illustrate the practical value of the proposed approach.

DEDICATION

To my mother, my father, my sister.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Wong Professor Zhang from the department of statistics and Professor An from the department of economics.

All work for the dissertation was completed independently by myself.

Funding Sources

I completed all work independently without outside financial support.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND LITERATURE RIVIEW

1.1 Introduction

Evaluating the effect of the exposure of a set of units to a program, or treatment, on some outcome, is the central problem studied in the literature on the causal effect of programs. The object of interest of this literature is a comparison of the two outcomes for the same unit when exposed and when not exposed to the treatment, and a key feature is the accommodation of general heterogeneity in the outcomes. Most of this literature focus on the comparison of the two average outcomes over the whole population (e.g., the so-called average treatment effect (ATE)) or the treated subpopulation (e.g., the so-called average treatment effect on the treated (ATT)). Example studies of this sort include [1], [2], [3], [4], [5], [6], [7], and [8] among others. Few studies focus on the comparison of the quantiles of the outcomes (e.g., the so-called quantile treatment effect (QTE) or quantile treatment effect on the treated (QTT)). See [9]) for an example, [10] for a general methodology, [11] and [12] for an extensive literature survey.

Despite simplicity of the average and quantile treatment effects, these parameters do not tell the whole story. It is possible that, on average, there does not exist the treatment effect when some units experience positive treament effects while other units experience negative treatment effects. Thus, focusing on the average treatment effect may miss out the treatment effects on subgroups. It is also possible that some quantile treatment effects are positive while other quantile treatment effects are negative. Again, focusing on some quantiles may miss out the treatment effects on other quantiles. Ideally, one would like to compare the outcome distributions, not just some moments. For example, one may be interested in the comparison of the two outcome distributions for the whole population when exposed and when not exposed to the treatment. These distribution could reveal not just the magnitude of the effects (if any) but also the proportion of the population benefitted from (or harmed by) the treatment. One may also be interested in the comparison of the two outcome distributions for a subpopulation (e.g., treated or untreated population) when exposed and when not exposed to the treatment. These distributions may help decision makers developing targeted programs or treaments. In light of the importance of the distribution comparison, the literature call the distribution difference as the distributional effect.

Clearly, to detect and analyze the distributional effects, one must have consistent distribution estimates available. Since we only observe each unit either exposed or not exposed to the treatment, participation selection, common in observational data, could complicate the estimation. Under the condition of selection on observables, [13] and [14] propose some consistent estimates of the marginal distributions of outcomes and establish their asymptotic distributions. [14] also suggest some stochastic dominance statistics for detecting the distributional effects. [15] allows for selection on unobservables but requires a special instrument to identify the marginal distributions of outcomes for the complier subpopulation. He also proposes a stochastic dominance statistic for detecting the distributional effects. In a slightly different setting, [16], [17] and [18] propose Mann-Whitney statistics to detect the distributional effects. Their focus, however, is not on estimation of the outcome distributions. Moreover, their estimation of Mann-Whitney type parameters are based on parametric models, hence they suffer from the model misspecification problems. In a missing data setting where two samples are generated independently from two distributions, [19], [20], [21] and [22] propose Mann-Whitney statistics to test the difference of the two distributions. Again, estimation of the distribution functions is not their focus. [22] also derive the semiparametric efficiency bound of the Mann-Whitney statistics for the two-samples problem and show that their Mann-Whitney statistics attain the semiparametric efficiency bound under some additional index restrictions. We note that the treatment effect problem is not the same as the two-samples problem. For example, condition 5 of [22] is not satisfied by the treatment effect model. Even if we can create two artificial samples from the treatment effect data using the approach suggested in [21] and [22], the two samples are not independent. Therefore, the semiparametric efficiency bound derived in [22] for the two-samples problem may not be the semiparametric efficiency bound for the treatment model.

To summarize, efficient estimation of the outcome distributions for the whole or target population and efficient inference on the distributional effects have not received any attention from the treatment effects literature. The main objective of this paper is to fill in this gap. Specifically, we shall derive the semiparametric efficiency bounds of all counterfactual distributions as well as the Mann-Whitney indicators of distributional effects for the treatment effect models under the condition of selection on observables. We shall then propose estimation of the counterfactual distributions and Mann-Whitney statistics, derive their asymptotic distributions and show that they attain the semiparametric efficiency bounds.

The remainder of the paper is organized as follows: Section 2.1 describes the counterfactuals and the Mann-Whitney indicators in the discrete treatment model, Section 2.2 derives the efficiency bounds for the counterfactual distributions and the the Mann-Whitney indicators in Section 2.1, Section 2.3 presents an estimation of the counterfactual distributions and the Mann-Whitney indicators in Section 2.1, Section 2.4 derives large sample properties of the proposed estimators in Section 2.3, Section 2.5 provides a data-driven approach to select the smoothing parameter and some consistent covariance matrices for the proposed estimators, Section 3.1 reports on a small scale simulation study, Section 3.2 reports on an empirical application, followed by some concluding remarks in Section 4.1. Detailed proofs of the main results are contained in Appendix A.

2. MODEL AND THEOREM

2.1 Model Setup

Consider a multivalued treatment (J-valued treatment with $J \geq 1$). Let $(Y(0), Y(1), ..., Y(J))$ denote the potential outcomes for the same unit when not exposed and when exposed to various treatment levels respectively and let $T \in \{0, 1, ..., J\}$ denote the treatment assignment variable. For any $y \in \mathbb{R}$ and $0 \leq s \leq J$, let $F_s(y) = \mathbb{P}(Y(s) \leq y)$ denote the marginal distribution. A comparison of $F_s(y)$ and $F_t(y)$ can reveal the distributional effect of treatment t relative to treatment s. Similarly, let $F_{st}(y) = \mathbb{P}(Y(s) \leq y | T = t)$ for $s, t \in \{0, 1, ..., J\}$ denote the counterfactual distribution of treatment s for those receiving treatment t. Then the comparison of $F_{st}(y)$ and $F_{tt}(y)$ can reveal the distributional effect of those receiving treatment t if they instead receive treatment s. Our primary interest is the efficient estimation of the distributions $(F_s(y), F_{st}(y); s, t \in \{0, 1\})$ and efficient inference on the distributional effects.

To detect the difference of any two distributions, say $F(y)$ and $G(y)$, we shall use the Mann-Whitney indicator $\theta = \int_{-\infty}^{\infty} F(y)dG(y)$. It is easy to show that $\theta = 1/2$ if and only if $F(y) = G(y)$ for all $y \in \mathbb{R}$, $\theta > \frac{1}{2}$ if $F(y) > G(y)$ for all $y \in \mathbb{R}$ and $\theta < \frac{1}{2}$ if $F(y) < G(y)$ for all $y \in \mathbb{R}$. Thus, the deviation of θ from $1/2$ can be used to test the null hypothesis of $H_0 : F(y) = G(y)$ for all $y \in \mathbb{R}$ against the alternative $H_1 : F(y) \neq G(y)$ for some $y \in \mathbb{R}$ or against the alternative hypothesis $H_1 : F(y) > G(y)$ for all $y \in \mathbb{R}$ or against the alternative $H_1 : F(y) < G(y)$ for all $y \in \mathbb{R}$. Applying this indicator to our distribution functions, we obtain:

$$\theta_{st} = \int_{-\infty}^{\infty} F_s(y)dF_t(y) \text{ and } \theta_{stt} = \int_{-\infty}^{\infty} F_{st}(y)dF_{tt}(y) \text{ for any } 0 \leq s, t \leq J.$$

We shall estimate the indicators

$$\theta_0 = ((\theta_{st}; \text{ for all } 0 \leq s \neq t \leq J), (\theta_{stt}; \text{ for all } 0 \leq s \neq t \leq J))^\top$$

by plugging in the estimates of the distributions

$$\mathbb{F}(\cdot) = (F_s(\cdot), F_{st}(\cdot); \text{for all } 0 \leq s, t \leq J)^\top.$$

Notice that $F_{tt}(\cdot)$ can be estimated directly by the empirical distribution using observations from treatment t. However, since there are no observations drawn from the counterfactual distributions, the counteractual distributions are not identified unless some restrictions are imposed. To identify the counterfactual distributions, we follow the lead of the literature by imposing “selection on observables" (e.g., [1, 2, 23, 6, 7]). Specifically, let X denote a vector of observed covariates. The following condition shall be maintained throughout the paper.

Assumption 2.1.1. *Given X, T is independent of $(Y(s); s \in \{0,1\})$, namely $T \perp (Y(s); 0 \leq s \leq J)|X$.*

Denote the observed outcome by $Y = \sum_{s=0}^{J} \mathbb{1}\left(T = s\right) Y(s)$. Assumption 2.1.1 implies that, for any $0 \leq s \leq J$, the counterfactual distribution $F_s(y)$ is identified as:

$$F_s(y) = \mathbb{E}\left[\mathbb{1}\{T = s\}\pi_s(X)\mathbb{1}(Y \leq y)\right],$$

where $\pi_s(X) = 1/\mathbb{P}(T = s|X)$ is the inverse weighting probability, and that, for any $s \neq t$, the counterfactual distribution $F_{st}(y)$ is identified as

$$F_{st}(y) = \mathbb{E}\left[\pi_{st}(X)\mathbb{1}(Y \leq y)|T = s\right],$$

where $\pi_{st}(X) = f_{X|T}(X|T = t)/f_{X|T}(X|T = s)$ is the ratio of conditional densities.

2.2 Efficiency Bound

Since testing the significance of the distributional effects is the key task of the treatment effect literature, it is important to use the most powerful tests. The power of the tests depends on the accuracy of the estimated distributions and the Mann-Whitney indicators. In order to evaluate the accuracy of these estimates, we derive the semiparametric efficiency bounds for the distribution functions and the indicators. Specifically, with $D_s = \mathbb{1}(T = s)$ and $p_s = \mathbb{E}(D_s)$, we denote:

$$S_{F_s} = \pi_s(X)\left[\mathbb{1}(Y \le y)D_s - F_{Y(s)|X}(y|X)D_s\right] + F_{Y(s)|X}(y|X) - F_s(y);$$
$$S_{F_{st}} = \frac{1(Y \le y)D_s\pi_{st}(X)}{p_s} - \left\{\frac{D_s\pi_{st}(X)}{p_s} - \frac{D_t}{p_t}\right\}F_{Y(s)|X}(y|X) - \frac{D_tF_{st}(y)}{p_t};$$

and the efficient scores of θ_0 by

$$\begin{aligned}
S_{\theta_{st}} = {} & D_s\pi_s(X)\left\{1 - F_t(Y) - \int F_{Y(s)|X}(y|X)dF_t(y)\right\} \\
& + D_t\pi_t(X)\left\{F_s(Y) - 1 + \int F_{Y(t)|X}(y|X)dF_s(y)\right\} \\
& + \left(1 - \int F_{Y(t)|X}(y|X)dF_s(y)\right) - \theta_{st} + \int F_{Y(s)|X}(y|X)dF_t(y) - \theta_{st}; \\
S_{\theta_{stt}} = {} & \frac{D_s}{p_s}\left\{(1 - F_{tt}(Y))\pi_{st}(X) - \pi_{st}(X)\int F_{Y(s)|X}(y|X)dF_{tt}(y)\right\} \\
& + \frac{D_t}{p_t}\left\{F_{st}(Y) - 2\theta_{stt} + \int F_{Y(s)|X}(y|X)dF_{tt}(y)\right\}.
\end{aligned}$$

The following results are derived in the supplemental material.

Theorem 2.2.1. *Under Assumption 2.1.1, we obatin that:*

1. *for a fixed $y \in R$, the efficient score of $\mathbb{F}(y)$ is*

$$\mathbb{S}_F(T, X, Y, y) = (S_{F_s}, S_{F_{st}};\ \textit{for all}\ 0 \le s, t \le J)^\top$$

and hence the semiparametric efficiency bound is $V_F(y) = \mathbb{E}[\mathbb{S}_F(T, X, Y, y)\mathbb{S}_F(T, X, Y, y)^\top]$;

2. *the efficient score of θ_0 is*

$$\mathbb{S}_\theta(T, X, Y) = ((S_{\theta_{st}};\ \textit{for all}\ 0 \le s \ne t \le J), (S_{\theta_{stt}};\textit{for all}\ 0 \le s \ne t \le J))^\top$$

and hence the semiparametric efficiency bound is $V_\theta = \mathbb{E}[\mathbb{S}_\theta(T, X, Y)\mathbb{S}_\theta(T, X, Y)^\top]$.

It is worth noting that, if we estimate $F_{tt}(y)$ by the sample average of $\mathbb{1}(T = t \,\&\, Y \leq y)/p_t$, the resulting estimator attains the semiparametric efficiency bound. However, if we estimate $F_{st}(y)$ for $s \neq t$ by the sample average of $\mathbb{1}(T = s \,\& Y \ \leq y)\pi_{st}(X)/p_s$ with known $\pi_{st}(X)$, the resulting estimtor has variance $\mathbb{E}[\widetilde{S}^2_{F_{st}}]$ with

$$\widetilde{S}_{F_{st}} = S_{F_{st}} + \left\{ \frac{\mathbb{1}(T = s)}{p_s} \pi_{st}(X) - \frac{\mathbb{1}(T = t)}{p_t} \right\} F_{Y(s)|X}(y|X).$$

It can be shown that

$$\mathbb{E}\left[\widetilde{S}^2_{F_{st}}\right] > \mathbb{E}\left[S^2_{F_{st}}\right].$$

This seems counterintuitive since knowing $\pi_{st}(X)$ is additional information that should improve, at least not worsen, the efficiency of $F_{st}(y)$. The reason for this seemingly counterintuitive result is that the information contained in the restriction on $\pi_{st}(X)$,

$$\mathbb{E}\left[\pi_{st}(X)u(X)|T = s\right] = \mathbb{E}\left[u(X)|T = t\right] \text{ holds for any integrable function } u(X),$$

is ignored in the estimation of $F_{st}(y)$. When this restriction is exploited, the efficiency of $F_{st}(y)$ can actually be improved. Similar insight is also noted by [7] in the efficiency of ATE. Here we show that the same insight holds true for the counterfactual distributions. The insight suggests that obtaining the most accurate estimates of the weighting functions is not nearly as important as incorporating the restrictions: for any integrable function $u(X)$,

$$\mathbb{E}\left[\mathbb{1}(T = t)\pi_t(X)u(X)\right] = \mathbb{E}[u(X)] = \mathbb{E}\left[\mathbb{1}(T = s)\pi_s(X)u(X)\right]; \tag{2.1}$$

$$\mathbb{E}\left[\pi_{st}(X)u(X)|T = s\right] = \mathbb{E}\left[u(X)|T = t\right]. \tag{2.2}$$

We notice that [22] derive the semiparametric efficiency bound of the Mann-Whitney indicator for the two-samples problem with missing data under the condition that the two samples are independent of each other. Although the binary treatment model (e.g., $J = 1$) can be formulated as a two-samples problem with missing data, the two samples consist of the same units and hence are not independent of each other. Failing to recognize the dependence of the two artifical samples, their semiparametric efficiency bound may not be the semiparametric efficiency bound for the binary treament model.

2.3 Efficient Estimation

The discussion in the preceding section suggests that, in order to estimate $\mathbb{F}(\cdot)$ and θ_0 efficiently, it is vitally important to incorporate the restrictions on the weighting functions into estimation. The difficulty, however, is that these restrictions hold on an infinite dimmensional functional space and we only have a finite sample of observations. To overcome this difficulty, we propose to impose the restrictions on a finite dimensional sieve space that grows with sample sizes. Specifically, let $u_K(X) = (u_{K,1}(X), \ldots, u_{K,K}(X))^\top$ denote known basis functions that can approximate any suitable function $u(X)$ arbitrarily well (see [24]). For normalization purpose, we always include constant 1 in $u_K(X)$. Restrictions (2.1)-(2.2) imply:

$$\mathbb{E}\left[\mathbb{1}(T=t)\pi_t(X)u_K(X)\right] = \mathbb{E}[u_K(X)] = \mathbb{E}\left[\mathbb{1}(T=s)\pi_s(X)u_K(X)\right], \tag{2.3}$$

$$\mathbb{E}\left[\pi_{st}(X)u_K(X)|T=s\right] = \mathbb{E}\left[u_K(X)|T=t\right]. \tag{2.4}$$

Here we use the same smoothing parameter K in the restrictions on all weighting functions to simplify notation. In principle, one could use different smoothing parameters for different weighting functions. As long as our sufficient conditions below are satisfied by all smoothing parameters, our main results still hold.

To estimate the weighting functions, let $\rho(v)$ denote a strictly increasing and concave function, and let $\rho'(v)$ denote the first derivative. For a sample of independent and identically distributed observations $\{(Y_i, T_i, X_i), i = 1, 2, ..., N\}$, denote

$$\hat{\pi}_s(X_i) = \rho'(\hat{\lambda}_s^\top u_K(X_i)) \text{ and } \hat{\pi}_{st}(X_i) = \rho'(\hat{\lambda}_{st}^\top u_K(X_i)),$$

where $\hat{\lambda}_s$,$\hat{\lambda}_{st} \in R^K$ maximize respectively the following concave objective functions:

$$\hat{Q}_s(\lambda) = \frac{1}{N}\sum_{i=1}^{N}\left[\mathbb{1}(T_i = s)\rho\left(\lambda^\top u_K(X_i)\right) - \lambda^\top u_K(X_i)\right],$$
$$\hat{Q}_{st}(\lambda) = \frac{\sum_{i=1}^{N}\mathbb{1}(T_i = s)\rho(\lambda^\top u_K(X_i))}{\sum_{i=1}^{N}\mathbb{1}(T_i = s)} - \frac{\sum_{i=1}^{N}\mathbb{1}(T_i = t)\lambda^\top u_K(X_i)}{\sum_{i=1}^{N}\mathbb{1}(T_i = t)}.$$

It is straightforward to show that the estimates $\hat{\pi}_s(X)$ and $\hat{\pi}_{st}(X)$ satisfy the sample analogue of (2.3)-(2.4):

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}(T_i = s)\hat{\pi}_s(X_i)u_K(X_i) = \frac{1}{N}\sum_{i=1}^{N}u_K(X_i),$$
$$\frac{\sum_{i=1}^{N}\mathbb{1}(T_i = s)\hat{\pi}_{st}(X_i)u_K(X_i)}{\sum_{i=1}^{N}\mathbb{1}(T_i = s)} = \frac{\sum_{i=1}^{N}\mathbb{1}(T_i = t)u_K(X_i)}{\sum_{i=1}^{N}\mathbb{1}(T_i = t)}.$$

Moreover, these estimates have empirical likelihood interpretations. For example, $\rho(v) = -\exp(-v)$ corresponds to the exponential tilting [25, 26], $\rho(v) = \log(1+v)$ corresponds to the empirical likelihood [27, 28], $\rho(v) = -(1-v)^2/2$ corresponds to the continuous updating of the generalized method of moments [29, 30] and $\rho(v) = v - \exp(-v)$ corresponds to the inverse logistic. Regardless of the choice of $\rho(\cdot)$, $\hat{Q}_s(\lambda)$ and $\hat{Q}_{st}(\lambda)$ are all globally concave and hence their maximands are easy to compute.

With the estimated weighting functions in hand, we now estimate the distribution functions by

$$\hat{F}_s(y) = \frac{1}{N}\sum_{i=1}^{N}\hat{\pi}_s(X_i)\mathbb{1}(T_i = s \;\&\; Y_i \le y),$$
$$\hat{F}_{st}(y) = \frac{\sum_{i=1}^{N}\hat{\pi}_{st}(X_i)\mathbb{1}(T_i = s \;\&\; Y_i \le y)}{\sum_{i=1}^{N}\mathbb{1}(T_i = s)},$$
$$\hat{F}_{tt}(y) = \frac{\sum_{i=1}^{N}\mathbb{1}(T_i = t \;\&\; Y_i \le y)}{\sum_{i=1}^{N}\mathbb{1}(T_i = t)}.$$

Notice that, since the basis functions $u_K(X)$ include the constant 1, the estimated weighting functions satisfy:

$$N^{-1}\sum_{i=1}^{N}\mathbb{1}(T_i=s)\hat{\pi}_s(X_i)=1,$$
$$\sum_{i=1}^{N}\mathbb{1}(T_i=s)\hat{\pi}_{st}(X_i)=\sum_{i=1}^{N}\mathbb{1}(T_i=s).$$

Consequently, the estimated distributions are garanteed to be in $[0,1]$.

It is worth noting that, in the binary treatment case, [14] propose similar estimates of the distribution functions (i.e., $\widehat{F}_s(y)$). The difference is that they estimate the weighting functions by inverse of the nonparametric estimate of the propensity score $P(T=s|X)$. A disadvantage of their approach is that their estimated weighting functions are quite sensitive to the extreme small values of the estimated propensity score (see [31] for further commets). In contrast, we estimate the inverse directly and avoid the small propensity score problem. Moreover, we estimate other distributions as well as the Mann-Whitney indicators.

By pluging in the estimated distributions, we estimate θ_0 by the following Mann-Whitney statistics:

$$\widehat{\theta}_{st}=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{1}(T_i=s\ \&\ T_j=t)\hat{\pi}_s(X_i)\hat{\pi}_t(X_j)\mathbb{1}(Y_i\le Y_j), \tag{2.5}$$
$$\hat{\theta}_{stt}=\frac{\sum_{j=1}^{N}\sum_{i=1}^{N}\mathbb{1}(T_i=s\ \&\ T_j=t)\hat{\pi}_{st}(X_i)\mathbb{1}(Y_i\le Y_j)}{\sum_{j=1}^{N}\sum_{i=1}^{N}\mathbb{1}(T_i=s\ \&\ T_j=t)}. \tag{2.6}$$

The Mann-Whitney statistic $\widehat{\theta}_{st}$ is similar to the one constructed in [20], [21] and [22] for the two-samples missing data problem. [20] requires a parametric modeling of the weighting functions and his approach is not robust to the risk of misspecification. Moreover, he does not incorporate the restrictions on the weighting functions and hence his statistic is unlikely efficient. [21] first construct a conditional version (conditional on covariates or the propensity score) of the Mann-Whitney statistic and then average it over the population. The conditional version is estimated either nonparametrically or semiparametrically if the propensity score function is specified. Again, since restrictions on the weighting functions are not imposed, their statistics are unlikely efficient. [22] is essentially an improved version of [21]. Instead of assuming parametric propensity score, [21] assume the propensity score function is an index function and estimate the index from the data. Although they show that their statistic attains their semiparametric efficiency bound for the two-samples problem, provided an index assumption is satisfied, we argue that the treatment effect problem is not the same as the two-samples missing data problem. Therefore their statistic is unlikely to attain our bound.

2.4 Asymptotic Properties

To establish the large sample properties of the proposed estimators

$$\widehat{\mathbb{F}}(\cdot) = (\widehat{F}_s(\cdot), \widehat{F}_{st}(\cdot);\ \text{for all } 0 \leq s, t \leq J)^\top$$

and

$$\widehat{\theta} = ((\widehat{\theta}_{st};\ \text{for all } 0 \leq s \neq t \leq J), (\widehat{\theta}_{stt};\ \text{for all } 0 \leq s \neq t \leq J))^\top,$$

we impose the following sufficient conditions.

Assumption 2.4.1. *(i) The support of r-dimensional covariates X, denoted by $\mathcal{X}$, is Cartesian product of r compact intervals; (ii) For all $x \in \mathcal{X}$, $\pi_0(x), \pi_1(x), \pi_{01}(x)$ and $\pi_{10}(x)$ are bounded and bounded away from 0.*

Assumption 2.4.2. *For all $0 \leq s, t \leq 1$, there exist λ_{sK} and λ_{stK} and $\alpha > 0$ such that*

$$\begin{aligned} \sup_{x\in\mathcal{X}} \left|(\rho'^{-1}(\pi_s(x)) - \lambda_{sK}^\top u_K(x)\right| &= O(K^{-\alpha}) \text{ holds uniformly in } K; \\ \sup_{x\in\mathcal{X}} \left|(\rho'^{-1}(\pi_{st}(x)) - \lambda_{stK}^\top u_K(x)\right| &= O(K^{-\alpha}) \text{ holds uniformly in } K. \end{aligned}$$

Assumption 2.4.3. *For any $t \in \{0,1\}$, the conditional distribution $F_{Y(t)|X}(y|x) := P(Y_i(t) \leq y|X_i = x)$ is continuously differentiable in x and is Lipschitz continuous in y.*

Assumption 2.4.4. *(i) For every K, the smallest eigenvalue of $\mathbb{E}\left[u_K(X)u_K(X)^\top\right]$ is bounded* *away from zero uniformly in K. (ii) There exists a sequence of constants $\zeta(K)$ satisfying $\sup_{x\in\mathcal{X}}\|u_K(x)\| \leq \zeta(K)$ such that $\zeta(K)K^{-\alpha} \to 0$ and $\zeta(K)\sqrt{K/N} \to 0$ as $N \to \infty$.*

Assumption 2.4.5. *For all i, $\rho(\cdot)$ is a strictly concave function defined on R and $\rho'(\cdot)$ is bounded.*

Assumption 2.4.1 (i) restricts the covariates to be bounded. This condition is restrictive but familiar in the nonparametric regression literature. It can be relaxed, however, if the tail distribution of X is restricted. Assumption 2.4.1 (ii) requires that each type of individuals (denoted by X) must present in both groups. This restriction allows us to match individuals in one group with statistical counterparts in the other group. Assumption 2.4.2 requires the sieve approximation error of $\rho'^{-1}(\pi_s(x))$ and $\rho'^{-1}(\pi_{st}(x))$ to shrink at a polynomial rate. This condition is satisfied for a variety of sieve basis functions. For example, if X is discrete, then the approximation error is zero for sufficient large K and in this case Assumption 2.4.2 is satisfied with $\alpha = +\infty$. If some components of X are continuous, the polynomial rate depends positively on the smoothness of $\rho'^{-1}(\pi_s(x))$ and $\rho'^{-1}(\pi_{st}(x))$ and negatively on the number of the continuous covariates. Assumption 2.4.3 is needed for ensuring the stochastic equicontinuity that is used for weak convergence. Assumption 2.4.4 (i) rules out the near multicollinearity of the basis functions. This condition is commonly imposed in the sieve regression literature (see [32, 33]). Assumption 2.4.4 (ii) restricts the growth rate of the smoothing parameter K. K must grow not too fast and not too slow. Assumption 2.4.5 is a mild restriction on ρ and it is satisfied by all important special cases that have been considered in the literature.

Under these conditions, we first establish the consistency of the estimated weighting functions under the L_2-norm.

Theorem 2.4.6. *Suppose that Assumptions 2.4.1-2.4.5 hold. Then, for any $0 \leq s \neq t \leq J$, we* *have:*

$$\int_{\mathcal{X}} |\widehat{\pi}_s(x) - \pi_s(x)|^2 dF_X(x) = O_p\left(\max\left\{K^{-2\alpha}, \frac{K}{N}\right\}\right),$$
$$\int_{\mathcal{X}} |\widehat{\pi}_{st}(x) - \pi_{st}(x)|^2 dF_X(x) = O_p\left(\max\left\{K^{-2\alpha}, \frac{K}{N}\right\}\right),$$
$$\frac{1}{N}\sum_{i=1}^{N} |\widehat{\pi}_s(X_i) - \pi_s(X_i)|^2 = O_p\left(\max\left\{K^{-2\alpha}, \frac{K}{N}\right\}\right),$$
$$\frac{1}{N}\sum_{i=1}^{N} |\widehat{\pi}_{st}(X_i) - \pi_{st}(X_i)|^2 = O_p\left(\max\left\{K^{-2\alpha}, \frac{K}{N}\right\}\right).$$

The convergence rates in Theorem 2.4.6 are quite similar to the convergence rates of regression functions in the nonparametric literature (e.g., [33]), and these rates will be used to regulate the growing rate of the smoothing parameter K so that the estimated Mann-Whitney statistics and distribution functions are consistent and asymptotically normally distributed. Under the following under-smoothing condition, we now establish the asymptotic normality of the Mann-Whitney statistics.

Assumption 2.4.7. $\sqrt{N}K^{-\alpha} \to 0$ *as* $N \to \infty$.

Theorem 2.4.8. *Under Assumption 2.1.1 and Assumptions 2.4.1-2.4.7, we have*

$$\sqrt{N}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_\theta)\ ,$$

where V_θ is the semiparametric efficiency bound of θ_0.

The asymptotic normality of the Mann-Whitney statistics can be used for detecting evidence of distributional effects. For instance, with $v_{11\theta}$ denoting the first main diagonal element of V_θ, the above theorem shows that $\sqrt{N}(\widehat{\theta}_{01} - \theta_{01})/\sqrt{v_{11\theta}}$ asymptotically has standard normal distribution. Under the null hypothesis of no distributional effect $H_0 : F_0(\cdot) = F_1(\cdot)$, $\sqrt{N}(\widehat{\theta}_{01} - 1/2)/\sqrt{v_{11\theta}}$ is asymptotically standard normal. At significance level $\alpha = 0.05$, we conclude no evidence of distributional effects if $\left|\sqrt{N}(\widehat{\theta}_{01} - 1/2)/\sqrt{v_{11\theta}}\right| < 1.96$, strong evidence of positive distributional effects if $\sqrt{N}(\widehat{\theta}_{01} - 1/2)/\sqrt{v_{11\theta}} > 1.96$ and strong evidence of negative effects if $\sqrt{N}(\widehat{\theta}_{01} - 1/2)/\sqrt{v_{11\theta}} < -1.96$. The fact that the Mann-Whitney statistics attain the semiparametric efficiency bound implies that these inferences are also efficient. If strong evidences of distributional effects are detected in the whole population or target populations, we might be interested in which outcome levels are benefited from (or harmed by) the treatment. But to do so, we need the asymptotic distribution of the estimated distribution functions, which are established in the following theorem.

Theorem 2.4.9. *Under Assumption 2.1.1 and Assumptions 2.4.1-2.4.7, we have, for any fixed* $y_0 \in \mathbb{R}$,

$$\sqrt{N}\{\widehat{\mathbb{F}}(y_0) - \mathbb{F}(y_0)\} \xrightarrow{d} \mathcal{N}(0, V_F(y_0)),$$

where $V_F(y_0)$ *is the semiparametric efficiency bound of* $\mathbb{F}(y_0)$*. Furthermore, we have*

$$\sqrt{N}\{\widehat{\mathbb{F}}(\cdot) - \mathbb{F}(\cdot)\} \Rightarrow G(\cdot, \cdot),$$

where " $\Rightarrow$ *" denotes the weak convergence and* $G(\cdot, \cdot)$ *is a Gaussian process with kernel function*

$$\Omega(y_1, y_2) = \mathbb{E}\left[\mathbb{S}_F(T, X, Y; y_1)\mathbb{S}_F(T, X, Y; y_2)\right].$$

The asymptotic results above allow us to construct the confidence band for the distributional differences. For example, with $v_{F_sF_t}(y_0)$ as the variance of $\widehat{F}_s(y_0) - \widehat{F}_t(y_0)$, the 5% confidence band is given by

$$\left[\widehat{F}_s(\cdot) - \widehat{F}_t(\cdot) - 1.96\frac{v_{F_sF_t}(\cdot)}{\sqrt{N}}, \widehat{F}_s(\cdot) - \widehat{F}_t(\cdot) + 1.96\frac{v_{F_sF_t}(\cdot)}{\sqrt{N}}\right],$$

which can be graphed and show which segments of population (defined by outcome levels) are benefited from (or harmed by) the treatment. Similar procedure can be applied to the treated or untreated population. The fact that these estimated distributions attain the semiparametric efficiency bound implies that these inferences are efficient.

2.5 Smoothing Parameter and Consistent Covariance

In order to carry out the statistical inferences discussed in the previous section, we must provide a data-driven smoothing parameter K so that the estimated distributions and the Mann-Whitney statistics can be computed, and provide some consistent covariance matrices so that the confidence bands can be constructed. In theory, the smoothing parameter K is required to satisfy Assumption 2.4.4 and Assumption 2.4.7, which unfortunately do not determine a unique value of K. This presents a dilemma for applied researchers who only have a finite sample of observations.

2.5.1 Selection of K

We shall select K to minimize the penalized loss function. Notice that the weighting functions satisfy the following moment conditions:

$$\mathbb{E}\left[D_s\pi_s(X)\exp(X)\right] = \mathbb{E}[\exp(X)] \text{ and } \mathbb{E}\left[\pi_{st}(X)\exp(X)|T=s\right] = \mathbb{E}[\exp(X)|T=t].$$

We propose to select K to minimize the following penalized loss function:

$$\hat{K}_s = \arg\min \left\| \left\{ \frac{1}{N}\sum_{i=1}^N D_{is}\hat{\pi}_s(X_i)\exp(X_i) - \frac{1}{N}\sum_{i=1}^N \exp(X_i) \right\}^2 \right\|_1 /(1-K/N)^2,$$
$$\hat{K}_{st} = \arg\min \left\| \left\{ \frac{\sum_{i=1}^N D_{is}\hat{\pi}_{st}(X_i)\exp(X_i)}{\sum_{j=1}^N D_{js}} - \frac{\sum_{i=1}^N D_{it}\exp(X_i)}{\sum_{j=1}^N D_{jt}} \right\}^2 \right\|_1 /(1-K/N)^2.$$

Evidently, we allow different smoothing parameter values to be used in estimation of different weighting functions, though a single smoothing parameter is used in theory. As long as these smoothing parameters satisfy the sufficient conditions on the smoothing parameter, all large sample properties still hold. Allowing for different smoohting parameter values, however, may improve finite sample performance.

2.5.2 Covariance

In this subsection, we estimate the asymptotic covariances by estimating the efficient scores. Notice that the estimated counterfactual distributions and the Mann-Whitney statistics depend on the estimated weighting functions. These estimated weighting functions contribute to the efficient scores through the following term:

$$-\frac{\partial \tilde{\pi}_s(X)}{\partial \lambda^\top}\left(\frac{\partial^2 \hat{Q}_s(\tilde{\lambda}_s)}{\partial \lambda \partial \lambda^\top}\right)^{-1}\frac{\partial \hat{Q}_s(\lambda_s^*)}{\partial \lambda}, \tag{2.7}$$

$$-\frac{\partial \tilde{\pi}_{st}(X)}{\partial \lambda^\top}\left(\frac{\partial^2 \hat{Q}_{st}(\tilde{\lambda}_{st})}{\partial \lambda \partial \lambda^\top}\right)^{-1}\frac{\partial \hat{Q}_{st}(\lambda_{st}^*)}{\partial \lambda}, \tag{2.8}$$

where $\tilde{\lambda}_s$ (resp. $\tilde{\lambda}_{st}$) lies between $\hat{\lambda}_s$ (resp. $\hat{\lambda}_{st}$) and λ_s^* (resp. λ_{st}^*), and $\partial \tilde{\pi}_s(X)/\partial \lambda^\top = \rho''(\tilde{\lambda}_s^\top u_K(X))u_K^\top(X)$ and $\partial \tilde{\pi}_{st}(X)/\partial \lambda^\top = \rho''(\tilde{\lambda}_{st}^\top u_K(X))u_K^\top(X)$, see Appendix A for more detailed explanation.

Denote

$$\begin{aligned}
\widehat{\Delta}_s &= -\frac{1}{N}\sum_{i=1}^N \mathbb{1}(T_i = s \,\&\, Y_i \le y)\cdot \rho''(\hat{\lambda}_s^\top u_K(X_i))\cdot u_K(X_i)^\top \left(\frac{\partial^2 \hat{Q}_s(\widehat{\lambda}_s)}{\partial \lambda \partial \lambda^\top}\right)^{-1}, \\
\widehat{\Delta}_{st} &= -\frac{\sum_{i=1}^N \mathbb{1}(T_i = s \,\&\, Y_i \le y)\cdot \rho''(\hat{\lambda}_{st}^\top u_K(X_i))\cdot u_K(X_i)^\top \left(\frac{\partial^2 \hat{Q}_{st}(\widehat{\lambda}_{st})}{\partial \lambda \partial \lambda^\top}\right)^{-1}}{\sum_{i=1}^N \mathbb{1}(T_i = s)}.
\end{aligned}$$

The efficient scores of the counterfactual distributions are estimated by

$$\begin{aligned}
\widehat{S}_{iF_s}(y) &= \widehat{\pi}_s(X_i)\mathbb{1}(T_i = s \,\&\, Y_i \le y) - \hat{F}_s(y) + \widehat{\Delta}_s u_K(X_i)\left[\mathbb{1}(T_i = s)\hat{\pi}_s(X_i) - 1\right], \\
\widehat{S}_{iF_{st}}(y) &= \frac{\hat{\pi}_{st}(X_i)\mathbb{1}(T_i = s \,\&\, Y_i \le y)}{N^{-1}\sum_{i=1}^N \mathbb{1}(T_i = s)} - \hat{F}_{st}(y) \\
&\quad + \widehat{\Delta}_{st} u_K(X_i)\left[\frac{\mathbb{1}(T_i = s)\hat{\pi}_{st}(X_i)}{N^{-1}\sum_{i=1}^N \mathbb{1}(T_i = s)} - \frac{\mathbb{1}(T_i = t)}{N^{-1}\sum_{i=1}^N \mathbb{1}(T_i = t)}\right], \\
\widehat{S}_{iF_{tt}}(y) &= \frac{\mathbb{1}(T_i = t \,\&\, Y_i \le y)}{N^{-1}\sum_{i=1}^N \mathbb{1}(T_i = t)} - \frac{\mathbb{1}(T_i = t)}{N^{-1}\sum_{i=1}^N \mathbb{1}(T_i = t)}\hat{F}_{tt}(y).
\end{aligned}$$

With

$$\widehat{\mathbb{S}}_{iF} = (\widehat{S}_{iF_s}, \widehat{S}_{iF_{st}};\ \text{for all } 0 \le s, t \le J)^\top,$$

the asymptotic covariance of the estimated distributions is estimated by

$$\widehat{V}_F(y) = \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathbb{S}}_{iF}\widehat{\mathbb{S}}_{iF}^\top.$$

To estimate the asymptotic covariance of the Mann-Whitney statistics, we denote

$$\widehat{\delta}_{st1} = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{1}(T_i = s \,\&\, T_j = t \,\&\, Y_i \le Y_j)\rho''(\hat{\lambda}_s^\top u_K(X_i))\hat{\pi}_t(X_j)u_K(X_i)^\top \left(\frac{\partial^2 \hat{Q}_s(\widehat{\lambda}_s)}{\partial\lambda\partial\lambda^\top}\right)^{-1},$$

$$\widehat{\delta}_{st2} = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{1}(T_i = s \,\&\, T_j = t \,\&\, Y_i \le Y_j)\hat{\pi}_s(X_i)\rho''(\hat{\lambda}_t^\top u_K(X_j))u_K(X_j)^\top \left(\frac{\partial^2 \hat{Q}_t(\widehat{\lambda}_t)}{\partial\lambda\partial\lambda^\top}\right)^{-1},$$

$$\widehat{\delta}_{stt} = -\frac{\sum_{j=1}^{N}\sum_{i=1}^{N} \mathbb{1}(T_i = s \,\&\, T_j = t \,\&\, Y_i \le Y_j)\rho''(\hat{\lambda}_{st}^\top u_K(X_i))u_K(X_i)^\top}{\sum_{j=1}^{N}\sum_{i=1}^{N} \mathbb{1}(T_i = s \,\&\, T_j = t)} \left(\frac{\partial^2 \hat{Q}_{st}(\widehat{\lambda}_{st})}{\partial\lambda\partial\lambda^\top}\right)^{-1}.$$

The efficient score of the Mann-Whitney statistics is estimated by

$$\begin{aligned}
\widehat{S}_{i\theta_{st}} &= \frac{1}{N}\sum_{j=1}^{N} \hat{\pi}_s(X_i)\hat{\pi}_t(X_j)\mathbb{1}(T_i = s \,\&\, T_j = t \,\&\, Y_i \le Y_j) - \hat{\theta}_{st} \\
&\quad + \frac{1}{N}\sum_{j=1}^{N} \hat{\pi}_s(X_j)\hat{\pi}_t(X_i)\mathbb{1}(T_j = s \,\&\, T_i = t \,\&\, Y_j \le Y_i) - \hat{\theta}_{st} \\
&\quad + \widehat{\delta}_{st1}u_K(X_i)\left[\mathbb{1}(T_i = s)\hat{\pi}_s(X_i) - 1\right] + \widehat{\delta}_{st2}u_K(X_i)\left[\mathbb{1}(T_i = t)\hat{\pi}_t(X_i) - 1\right],
\end{aligned}$$

and

$$\begin{aligned}
\widehat{S}_{i\theta_{stt}} &= \frac{\sum_{j=1}^{N} \hat{\pi}_{st}(X_i)\mathbb{1}(T_i = s\&T_j = t\&Y_i \le Y_j)}{N^{-1}\sum_{j=1}^{N}\sum_{i=1}^{N} \mathbb{1}(T_i = s\&T_j = t)} - \frac{\mathbb{1}(T_i = s)}{N^{-1}\sum_{j=1}^{N} \mathbb{1}(T_i = s)}\hat{\theta}_{stt} \\
&\quad + \frac{\sum_{j=1}^{N} \hat{\pi}_{st}(X_j)\mathbb{1}(T_j = s\&T_i = t\&Y_j \le Y_i)}{N^{-1}\sum_{j=1}^{N}\sum_{i=1}^{N} \mathbb{1}(T_i = s\&T_j = t)} - \frac{\mathbb{1}(T_i = t)}{N^{-1}\sum_{j=1}^{N} \mathbb{1}(T_i = t)}\hat{\theta}_{stt} \\
&\quad + \widehat{\delta}_{stt}u_K(X_i)\left[\frac{\mathbb{1}(T_i = s)\hat{\pi}_{st}(X_i)}{N^{-1}\sum_{i=1}^{N} \mathbb{1}(T_i = s)} - \frac{\mathbb{1}(T_i = t)}{N^{-1}\sum_{i=1}^{N} \mathbb{1}(T_i = t)}\right].
\end{aligned}$$

With

$$\widehat{\mathbb{S}}_{i\theta} = ((\widehat{S}_{i\theta_{st}}; 0 \leq s < t \leq J), (\widehat{S}_{i\theta_{stt}}; 0 \leq s \neq t \leq J))^\top,$$

we estimate the asymptotic covariance of the Mann-Whitney statistics by

$$\widehat{V}_\theta = \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbb{S}}_{i\theta} \widehat{\mathbb{S}}_{i\theta}^\top.$$

The following theorem shows that the estimated covariances are consistent.

Theorem 2.5.1. *Under Assumption 2.1.1 and Assumptions 2.4.1-2.4.5, $\hat{V}_F(y)$ is a consistent estimator of $V_F(y)$ for any fixed y and $\hat{V}_\theta$ is a consistent estimator of V_θ.*

3. SIMULATION AND EMPIRICAL APPLICATIONS

3.1 Simulation Studies

To evaluate the finite sample performance of the Mann-Whitney statistic $\widehat{\theta}_{011}$ and the distributional difference $\widehat{F}_{01} - \widehat{F}_{11}$, we conduct a small scale simulation study on a binary treatment. We consider three scenarios for DGPs. In all scenarios, we set $X = (X_1, X_2)$, $X_1 = \exp(Z_1/2)$ and $X_2 = Z_2/(1+\exp(Z_1))$, where Z_1, Z_2 are independently drawn from the uniform distribution over $[-2, 2]$.

- Scenario I: $\mathbb{P}(T = 1 \mid X) = 1/(1 + \exp(Z_1 - 0.05Z_2)$, $Y(0) = 200 - b(Z)$ and $Y(1) = 200 - b(Z)$, where $b(Z) = 10Z_1 + 7.5Z_2$. The true value is $\theta_{011} = 0.500$.
- Scenario II: $\mathbb{P}(T = 1 \mid X) = 1/(1 + \exp(0.6Z_1 - 0.5Z_2)$, $Y(0) = 200 - 0.5b(Z) + N$ and $Y(1) = 212 + b(Z) + N$, where $b(Z) = 10Z_1 + 7.5Z_2$ and N is the standard normal random variable. The true value is $\theta_{011} = 0.719$.
- Scenario III: $\mathbb{P}(T = 1 \mid X) = 1/(1 + \exp(0.5Z_1 - 0.5Z_2)$, $Y(0) = 210 + b(Z) + N$ and $Y(1) = 200 - 0.6b(Z) + N$, where $b(Z) = 10.5Z_1 + 9.5Z_2$ and N is a standard normal random variable. The true value is $\theta_{011} = 0.318$.

In each Monte Carlo run, we generate a sample of data from DGPs of two sample sizes: 600 and 800. The Monte Carlo runs are repeated for 500 times. We compare our covariate balancing estimator (CBE), the inverse probability weighting (IPW) estimator of [14], and the kernel estimator of [21]. Details of calculations are given below:

1. our covariate balancing estimator (CBE) is computed using the proposed approach with $\rho(v) = -\exp(-v)$. The data-driven K is computed from the following basis functions:

$$u_1(X) = 1,\ u_2(X) = [1, X_1, X_2], u_3(X) = [\text{first and second order polynomials of } X],$$
$$u_4(X) = [\text{first, second and third order polynomials of } X].$$

2. the inverse probability weighting (IPW) estimator of [14] is constructed with Logit propensity score function proposed by [7].

3. the kernel estimator of [21] is constructed with the Gaussian kernel and bandwith $h = 0.5$.

Table 3.1 reports the bias, standard deviation (stdev), and the root mean square error (rmse) of $\widehat{\theta}_{011}$. Glancing at Table 3.1, we have the following observations:

Table 3.1: Results of Estimation of θ_{011}

Sample Size	Scenario	CBE	IPW	Kernel Method
		bias, stdev, rmse	bias, stdev, rmse	bias, stdev, rmse
N=600	I	0.005, 0.005, 0.007	0.069, 0.015, 0.070	0.100, 0.013, 0.101
	II	0.002, 0.027, 0.027	-0.088, 0.029, 0.092	0.012, 0.024, 0.027
	III	-0.007, 0.028, 0.029	-0.035, 0.028, 0.044	-0.017, 0.025, 0.030
N=800	I	0.005, 0.005, 0.007	0.069, 0.013, 0.070	0.079, 0.035, 0.086
	II	0.004, 0.023, 0.024	-0.087, 0.024, 0.091	-0.030, 0.023, 0.037
	III	-0.007, 0.026, 0.027	-0.034, 0.025, 0.042	0.013, 0.022, 0.025

Table 3.2: Rejection Rates

Sample Size	Scenario	Rejection Rate
N=600	I	0
	II	0.680
	III	0.988
N=800	I	0
	II	0.972
	III	0.996

1. CBE is unbiased in all scenarios and has the smallest standard deviation.

2. IPW estimator and kernel estimator have large bias in all scenarios. The bias does not shrink to zero as the sample size increases.

Table 3.2 reports the rejection rates (power) of the Mann-Whitney test for $H_0 : F_{01} = F_{11}$. Figure 3.1 plots $F_{01}(y) - F_{11}(y)$, $\widehat{F}_{01}(y) - \widehat{F}_{11}(y)$ and the confidence band. Glancing at Table 3.2 and Figure 3.1, we have the following observations:

1. In all scenarios, the confidence band contains the true curve $F_{01}(y) - F_{11}(y)$.

2. In scenario I, $F_{01}(y) = F_{11}(y)$ for all y. The null hypothesis is rejected with probability less than 0.05. In scenario II, the null hypothesis is false and there exists a positive distributional effect. Our test rejects the null with probability tending to one quickly as sample size increases. Moreover, the graph shows significant positive distributional effect. In scenario III, the null hypothesis is false and there exists a negative distributional effect. Our test rejects the null with pprobability tending to one as sample size increases, and the graph shows negative distributional effect.

Overall, the proposed estimators and tests perform better relative to the other alternatives.

3.2 Empirical Applications

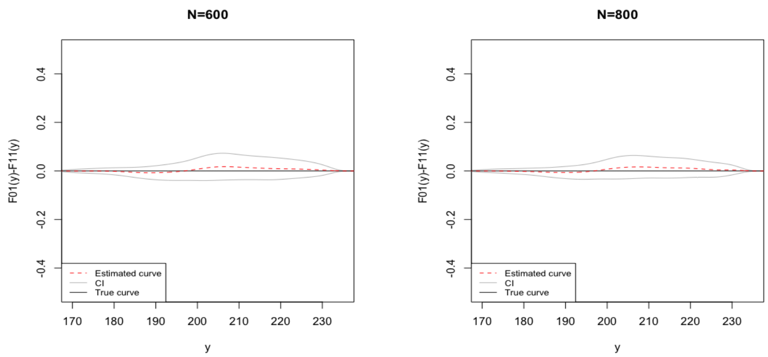
To illustrate the applicability of the proposed approach, we analyze the evidence of racial discrimination. Let T be the race indicator, i.e $T = 1$ if the individual is black and $T = 0$ if the individual is white. Let $Y(1)$ denote the wage of black man and $Y(0)$ denote the wage of white man. Let X denote the individual characteristics. Recall that $F_{st}(y)$ denote the conditional distribution of $Y(s)$ conditional on $T = t$. Notice that

$$F_{11}(y) - F_{00}(y) = \{F_{11}(y) - F_{01}(y)\} + \{F_{01}(y) - F_{00}(y)\},$$

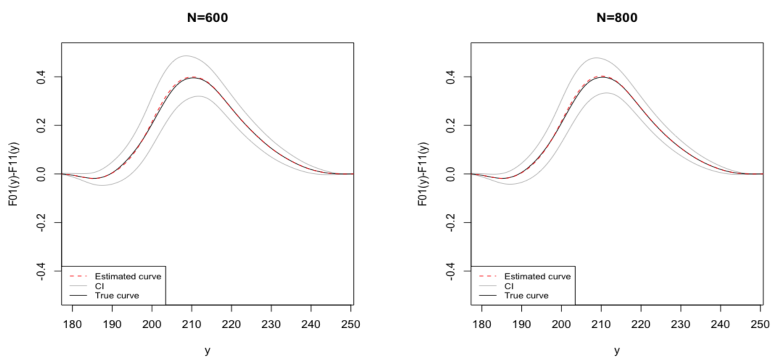
where the first term on the right hand side is a measure of race discrimination. If there is no race discrimination, then $F_{11}(y) = F_{01}(y)$ and $\theta_{011} = \int F_{01}(y)dF_{11}(y)$=0.5.

Figure 3.1: $F_{01}(y) - F_{11}(y)$

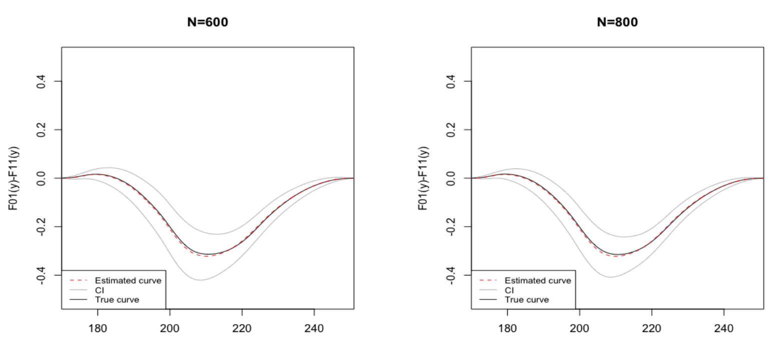
(a) Scenario I



(b) Scenario II



(c) Scenario III

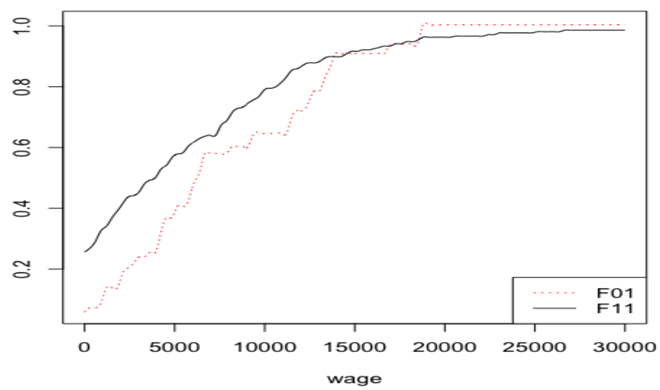


The red dotted line is our estimated curve; The black solid line is the true curve; The grey solid lines denote the confidence band.

Our sample consists of all males who participated in a training programe from the National Supported Work Demonstration (see [34]). The data include observations on age, education, Hispanic, marital status, a binary high school degree dummy, and real earning in 1975. The outcome variable is real earnings in 1978. For detailed description of the dataset, see [35].

We apply the proposed procedure to compute $\widehat{F}_{01}(y)$ and $\widehat{F}_{11}(y)$ and $\widehat{\theta}_{011}$. We find that $\widehat{\theta}_{011} = 0.358$ with p-value less than 0.05. Thus the null hypothesis of no racial discrimination is rejected at 5% level. Indeed Figure 3.2 shows that White males earned higher wages than Black males when both earn less than 15000. But for high wage earners, there is no evidence of racial discrimination.

Figure 3.2: Plot of $F_{01}(y)$ and $F_{11}(y)$



4. SUMMARY AND CONCLUSIONS

4.1 Concluding Remarks

We study the distributional treatment effect which is measured by the Mann-Whitney statistics. The difficulty of our method lies in the estimation for the counterfactual distribution which can not be observed. We propose covariate balancing estimators for counterfactual distributions and provide the asymptotic normality of the constructed Mann-Whitney statistics. We also obtained a consistent estimator of the asymptotic variance. Our procedure assumes all confounders are observed. The simulation results show that our method performs quite well. Extension of the proposed methodology to the situation with unmeasured confounders is certainly of great interest. We consider this an important problem for our future research.

REFERENCES

[1] P. R. Rosenbaum and D. B. Rubin, “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.

[2] P. R. Rosenbaum and D. B. Rubin, “Reducing bias in observational studies using subclassification on the propensity score,” *J. Am. Statist. Ass.*, vol. 79, no. 387, pp. 516–524, 1984.

[3] P. R. Rosenbaum, “Model-based direct adjustment,” *J. Am. Statist. Ass.*, vol. 82, no. 398, pp. 387–394, 1987.

[4] J. J. Heckman, H. Ichimura, and P. E. Todd, “Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme,” *The review of economic studies*, vol. 64, no. 4, pp. 605–654, 1997.

[5] J. J. Heckman, H. Ichimura, and P. Todd, “Matching as an econometric evaluation estimator,” *The review of economic studies*, vol. 65, no. 2, pp. 261–294, 1998.

[6] J. Hahn, “On the role of the propensity score in efficient semiparametric estimation of average treatment effects,” *Econometrica*, vol. 66, no. 2, pp. 315–331, 1998.

[7] K. Hirano, G. Imbens, and G. Ridder, “Efficient estimation of average treatment effects using the estimated propensity score,” *Econometrica*, vol. 71, pp. 1161–1189, July 2003.

[8] K. C. G. Chan, S. C. P. Yam, and Z. Zhang, “Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 78, no. 3, pp. 673–700, 2016.

[9] S. Firpo, “Efficient semiparametric estimation of quantile treatment effects,” *Econometrica*, vol. 75, no. 1, pp. 259–276, 2007.

[10] C. Ai, O. Linton, K. Motegi, and Z. Zhang, “A unified framework for efficient estimation of general treatment models,” *arXiv preprint arXiv:1808.04936*, 2018.

[11] G. W. Imbens and J. M. Wooldridge, "Recent developments in the econometrics of program evaluation," *Journal of economic literature*, vol. 47, no. 1, pp. 5–86, 2009.

[12] A. Abadie and M. D. Cattaneo, "Econometric methods for program evaluation," *Annual Review of Economics*, vol. 10, pp. 465–503, 2018.

[13] C. Victor, F. Iván, and M. Blaise, "Inference on counterfactual distributions," *Econometrica*, vol. 81, no. 6, pp. 2205–2268, 2013.

[14] S. G. Donald and Y. C. Hsu, "Estimation and inference for distribution functions and quantile functions in treatment effect models," *Journal of Econometrics*, vol. 178, no. 1, pp. 383–397, 2014.

[15] A. Abadie, "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 284–292, 2002.

[16] P. Wu, Y. Han, T. Chen, and X. Tu, "Causal inference for mann–whitney–wilcoxon rank sum and other nonparametric statistics," *Statistics in medicine*, vol. 33, no. 8, pp. 1261–1271, 2014.

[17] K. Vermeulen, O. Thas, and S. Vansteelandt, "Increasing the power of the mann-whitney test in randomized experiments through flexible covariate adjustment," *Statistics in medicine*, vol. 34, no. 6, pp. 1012–1030, 2015.

[18] Z. Zhang, S. Ma, C. Shen, and C. Liu, "Estimating mann–whitney-type causal effects," *Working Paper*, 2019.

[19] E. Schisterman and A. Rotnitzky, "Estimation of the mean of a k-sample u-statistic with missing outcomes and auxiliaries," *Biometrika*, vol. 88, no. 3, pp. 713–725, 2001.

[20] Y. K. Cheung, "Exact two-sample inference with missing data," *Biometrics*, vol. 61, no. 2, pp. 524–531, 2005.

[21] S. X. Chen, J. Qin, and C. Y. Tang, "Mann–whitney test with adjustments to pretreatment variables for missing values and observational study," *J. R. Statist. Soc. B (Statisticial Methodology)*, vol. 75, no. 1, pp. 81–102, 2013.

[22] H. Lin, F. Zhou, Q. Wang, L. Zhou, and J. Qin, "Robust and efficient estimation for the treatment effect in causal inference and missing data problems," *Journal of Econometrics*, 2018.

[23] P. R. Rosenbaum and D. B. Rubin, "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *The American Statistician*, vol. 39, no. 1, pp. 33–38, 1985.

[24] X. Chen, "Large sample sieve estimation of semi-nonparametric models," *Handbook of econometrics*, vol. 6, pp. 5549–5632, 2007.

[25] Y. Kitamura and M. Stutzer, "An information-theoretic alternative to generalized method of moments estimation," *Econometrica*, vol. 65, no. 4, pp. 861–874, 1997.

[26] G. Imbens, R. Spady, and P. Johnson, "Information theoretic approaches to inference in moment condition models," *Econometrica*, vol. 66, no. 2, pp. 333–357, 1998.

[27] A. Owen, "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.

[28] J. Qin and J. Lawless, "Empirical likelihood and general estimating equations," *Ann. Statist.*, vol. 22, pp. 300–325, 1994.

[29] L. Hansen, "Large sample properties of generalized method of moments estimators," *Econometrica*, vol. 50, pp. 1029–1054, 1982.

[30] L. Hansen, J. Heaton, and A. Yaron, "Finite-sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, vol. 14, no. 3, pp. 262–280, 1996.

[31] J. Kang and J. Schafer, "Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data," *Statistical science*, vol. 22, no. 4, pp. 523–539, 2007.

[32] W. K. Newey, "The asymptotic variance of semiparametric estimators," *Econometrica*, vol. 62, pp. 1349–1382, November 1994.

[33] W. K. Newey, "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, vol. 79, pp. 147–168, July 1997.

[34] R. Lalonde, "Evaluating the econometric evaluations of training programs," *American Economic Review*, vol. 76, pp. 604–620, 1986.

[35] R. H. Dehejia and S. Wahba, "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American statistical Association*, vol. 94, no. 448, pp. 1053–1062, 1999.

[36] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*. The Johns Hopkins University Press, 1993.

[37] W. Hoeffding, "A class of statistics with asymptotically normal distribution," in *Breakthroughs in Statistics*, pp. 308–334, Springer, 1992.

[38] R. J. Serfling, *Approximation theorems of mathematical statistics*, vol. 162. John Wiley & Sons, 2009.

[39] V. S. Korolyuk and Y. V. Borovskich, *Theory of U-statistics*, vol. 273. Springer Science & Business Media, 2013.

[40] D. W. Andrews, "Empirical process methods in econometrics," *Handbook of econometrics*, vol. 4, pp. 2247–2294, 1994.

A. APPENDIX

A.1 Proof of Theorem 2.2.1

This section utilizes the method of [36] to show the efficient influence functions of $F_{01}(y_0)$ and θ_{011}, which are denoted by $S_{F_{01}}$ and $S_{\theta_{011}}$ respectively, are

$$S_{F_{01}} = \frac{1-T}{p_0}\pi_{01}(X)\mathbb{1}(Y \le y_0) - \left\{\frac{1-T}{p_0}\pi_{01}(X) - \frac{T}{p_1}\right\} F_{Y(0)|X}(y_0|X) - \frac{T}{p_1}F_{01}(y_0); \quad \text{(A.1)}$$

$$\begin{aligned} S_{\theta_{011}} = &\frac{1-T}{p_0}\left\{\bar{F}_{11}(Y)\pi_{01}(X) - \pi_{01}(X)\int F_{Y(0)|X}(y \mid X) f_{Y(1)|T=1}(y)dy\right\} \\ &+ \frac{T}{p_1}\left\{F_{01}(Y) - \theta_{011} + \int F_{Y(0)|X}(y \mid X) f_{Y(1)|T=1}(y)dy - \theta_{011}\right\}. \end{aligned} \quad \text{(A.2)}$$

By applying the similar argument, it is easy to show that the efficient influence functions of $F_1(y_0), F_0(y_0), F_{10}(y_0), \theta_{01}$ and θ_{10t} are S_{F_0}, S_{F_0}, $S_{F_{10}}$, and $S_{\theta_{01}}$ defined in Theorem 1.

Consider a regular parametric submodel $\{f_{Y,X,T}(y,x,t;\tau) : \tau \in \mathbb{R}\}$ with $f_{Y,X,T}(y,x,t;\tau)|_{\tau=\tau_0} = f_{Y,X,T}(y,x,t)$. The likelihood function of $\{Y, X, T\}$ is

$$\begin{aligned} f_{Y,X,T}(y,x,t;\tau) = &\left[f_{Y|X,T}(y|x,T=1;\tau) f_{X|T}(x|T=1;\tau) P(T=1;\tau)\right]^t \\ &\times \left[f_{Y|T,X}(y|x,T=0;\tau) f_{X|T}(x|T=0;\tau) P(T=0;\tau)\right]^{(1-t)}. \end{aligned}$$

The score function of $f_{Y,X,T}(y,x,t;\tau)$ is

$$\begin{aligned} s(y,t,x;\tau) :=& \frac{d}{d\tau}\log f_{Y,X,T}(y,x,t;\tau) \\ =& t \cdot s_1(y|x,T=1;\tau) \\ &+ (1-t)\cdot s_0(y|x,T=0;\tau) + \frac{t - \mathbb{P}(T=1;\tau)}{\mathbb{P}(T=1;\tau)(1-\mathbb{P}(T=1;\tau))}\cdot \mathbb{P}'(T=1;\tau) \\ &+ t\cdot h_1(x|T=1;\tau) + (1-t)\cdot h_0(x|T=0;\tau), \end{aligned} \quad \text{(A.3)}$$

where

$$
\begin{aligned}
s_1(y|x, T=1;\tau) &= \frac{d}{d\tau}\log f_{Y|X,T}(y|x, T=1;\tau),\\
s_0(y|x, T=0;\tau) &= \frac{d}{d\tau}\log f_{Y|X,T}(y|x, T=0;\tau),\\
\mathbb{P}'(T=1;\tau) &= \frac{d}{d\tau}\mathbb{P}(T=1;\tau),\\
h_1(x|T=1;\tau) &= \frac{d}{d\tau}\log f_{X|T}(x|T=1;\tau),\\
h_0(x|T=0;\tau) &= \frac{d}{d\tau}\log f_{X|T}(x|T=0;\tau).
\end{aligned}
$$

From (A.3), we obtain the tangent space of this model:

$$
\Lambda := \{t\cdot s_1(y|x) + (1-t)\cdot s_0(y|x) + a\cdot(t - \mathbb{P}(T=1)) + t\cdot h_1(x) + (1-t)\cdot h_0(x)\},
$$

where $\int s_j(y|x) f_{Y|X,T}(y|x, T=j)dy = 0$, $\int h_j(x) f_{Y|X,T}(x|T=j)dx = 0$, $j\in\{0,1\}$, and $a\in\mathbb{R}$.

The parameters $\{F_{01}(y_0), \theta_{011}\}$ under the model $f_{Y,X,T}(y,x,t;\tau)$ are represented by

$$
\begin{aligned}
F_{01}(y_0;\tau) &= \int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x;\tau) f_{X|T=1}(x;\tau)dx,\\
\theta_{011}(\tau) &= \int_{y\in\mathcal{Y}} \mathbb{E}\left[\mathbb{1}(Y(0)\le y)|T=1\right] f_{Y|T=1}(y;\tau)dy\\
&= \int_{y\in\mathcal{Y}} \left\{\int_{x\in\mathcal{X}} F_{Y|T=0,X}(y|x;\tau) f_{X|T=1}(x;\tau)dx\right\} f_{Y|T=1}(y;\tau)dy\\
&= \int_{y\in\mathcal{Y}} \left\{\int_{x\in\mathcal{X}} F_{Y|T=0,X}(y|x;\tau) f_{X|T=1}(x;\tau)dx\right\} \left\{\int_{\mathcal{X}} f_{Y|X,T=1}(y|x;\tau) f_{X|T=1}(x;\tau)dx\right\} dy.
\end{aligned}
$$

Then

$$
\begin{aligned}
\frac{\partial F_{01}(y_0;\tau_0)}{\partial\tau} &= \int_{x\in\mathcal{X}} \frac{\partial F_{Y(0)|X}(y_0|x;\tau_0)}{\partial\tau} f_{X|T=1}(x;\tau_0)dx\\
&\quad + \int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x;\tau_0) h_1(x|T=1;\tau_0) f_{X|T=1}(x;\tau_0)dx,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \theta_{011}(\tau_0)}{\partial \tau} &= \int_{y\in\mathcal{Y}} \left\{ \int_{x\in\mathcal{X}} \frac{\partial F_{Y|T=0,X}(y|x;\tau_0)}{\partial \tau} f_{X|T=1}(x;\tau_0)dx \right\} f_{Y|T=1}(y;\tau_0)dy \\
&+ \int_{y\in\mathcal{Y}} \left\{ \int_{x\in\mathcal{X}} F_{Y|T=0,X}(y|x;\tau_0) h_1(x|T=1;\tau_0) f_{X|T=1}(x;\tau_0)dx \right\} f_{Y|T=1}(y;\tau_0)dy \\
&+ \int_{y\in\mathcal{Y}} \left[\int_{x\in\mathcal{X}} F_{Y|T=0,X}(y|x;\tau_0) f_{X|T=1}(x;\tau_0)dx \right] \\
&\quad \left[\int_{x\in\mathcal{X}} \{ s_1(y|x,T=1;\tau_0) f_{Y|X,T=1}(y|x;\tau_0) f_{X|T=1}(x;\tau_0) \right. \\
&\quad \left. + f_{Y|X,T=1}(y|x;\tau_0) h_1(x|T=1;\tau_0) f_{X|T=1}(x;\tau_0) \} dx \right] dy.
\end{aligned}
$$

It is obvious to see that $S_{F_{01}}(y_0)$ and $S_{\theta_{011}}$ are in the tangent space Λ. Hence, to prove $S_{F_{01}}(y_0)$ and $S_{\theta_{011}}$ are the efficient influence functions of $F_{01}(y_0)$ and θ_{011}, it is sufficient to verify the following equations hold:

$$
\frac{\partial F_{01}(y_0;\tau_0)}{\partial \tau} = \mathbb{E}\left[S_{F_{01}}(Y,T,X) \cdot s(Y,T,X;\tau_0) \right], \tag{A.4}
$$

$$
\frac{\partial \theta_{011}(\tau_0)}{\partial \tau} = \mathbb{E}\left[S_{\theta_{011}}(Y,T,X) \cdot s(Y,T,X;\tau_0) \right], \tag{A.5}
$$

where $s(Y,T,X;\tau_0)$ is defined in (A.3). To prove (A.4), we compute the expectation on the right hand side:

$$
\begin{aligned}
&\mathbb{E}\left[S_{F_{01}}(Y,T,X)\cdot s(Y,T,X;\tau_0)\right]\\
=&\mathbb{E}\Bigg\{\left[\frac{1-T}{p_0}\pi_{01}(X)\mathbb{1}(Y\le y_0)-\left\{\frac{1-T}{p_0}\pi_{01}(X)-\frac{T}{p_1}\right\}F_{Y(0)|X}(y_0|X)-\frac{T}{p_1}F_{01}(y_0)\right]\\
&\quad\times\Bigg[T\cdot s_1(Y|X,T=1;\tau_0)+(1-T)\cdot s_0(Y|X,T=0;\tau_0)\\
&\quad+\frac{T-\mathbb{P}(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\cdot\mathbb{P}'(T=1;\tau_0)\\
&\quad+T\cdot h_1(X|T=1;\tau_0)+(1-T)\cdot h_0(X|T=0;\tau_0)\Bigg]\Bigg\}\\
=&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)\mathbb{1}(Y\le y_0)s_0(Y|X,T=0;\tau_0)\right\}\\
&+\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)\mathbb{1}(Y\le y_0)h_0(X|T=0;\tau_0)\right\}\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)\mathbb{1}(Y\le y_0)(T-p_1)\right\}\\
&-\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y(0)|X}(y_0|X)s_0(Y|X,T=0;\tau_0)\right\}\\
&-\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y(0)|X}(y_0|X)h_0(X|T=0;\tau_0)\right\}\\
&-\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y(0)|X}(y_0|X)(T-p_1)\right\}\\
&+\mathbb{E}\left\{\frac{T}{p_1}F_{Y(0)|X}(y_0|X)s_1(Y|X,T=1;\tau_0)\right\}\\
&+\mathbb{E}\left\{\frac{T}{p_1}F_{Y(0)|X}(y_0|X)h_1(X|T=1;\tau_0)\right\}\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{T}{p_1}F_{Y(0)|X}(y_0|X)(T-p_1)\right\}\\
&-\mathbb{E}\left\{\frac{T}{p_1}F_{01}(y_0)s_1(Y|X,T=1;\tau_0)\right\}\\
&-\mathbb{E}\left\{\frac{T}{p_1}F_{01}(y_0)h_1(X|T=1;\tau_0)\right\}\\
&-\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{T}{p_1}F_{01}(y_0)(T-p_1)\right\}.
\end{aligned}
$$

Computing every term in $\mathbb{E}\left[S_{F_{01}}(Y,T,X)\cdot s(Y,T,X;\tau_0)\right]$, we have

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)\mathbb{1}(Y\le y_0)s_0(Y|X,T=0;\tau_0)\right\}\\
=&\int_{x\in\mathcal{X}}\left\{\int_{y\in\mathcal{Y}}\pi_{01}(x)\mathbb{1}(y\le y_0)s_0(y|x,T=0;\tau_0)f_{Y|X,T=0}(y|x;\tau_0)dy\right\}f_{X|T=0}(x;\tau_0)dx\\
=&\int_{x\in\mathcal{X}}\frac{\partial F_{Y(0)|X}(y_0|x;\tau_0)}{\partial\tau}f_{X|T=1}(x;\tau_0)dx,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)\mathbb{1}(Y\le y_0)h_0(X|T=0;\tau_0)\right\}\\
=&\int_{x\in\mathcal{X}}\left\{\int_{y\in\mathcal{Y}}\pi_{01}(x)\mathbb{1}(y\le y_0)h_0(x|T=0;\tau_0)f_{Y|X,T=0}(y|x;\tau_0)dy\right\}f_{X|T=0}(x;\tau_0)dx\\
=&\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0;\tau_0)\pi_{01}(x)\frac{\partial f_{X|T=0}(x;\tau_0)}{\partial\tau}dx,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)\mathbb{1}(Y\le y_0)(T-p_1)\right\} = -p_1\int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x;\tau_0)f_{X|T=1}(x;\tau_0)dx,\\
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y(0)|X}(y_0|X)s_0(Y|X,T=0;\tau_0)\right\} = 0,\\
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y(0)|X}(y_0|X)h_0(X|T=0;\tau_0)\right\}\\
&=\int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x;\tau_0)\pi_{01}(x)\frac{\partial f_{X|T=0}(x;\tau_0)}{\partial\tau}dx,\\
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y(0)|X}(y_0|X)(T-p_1)\right\} = -p_1\int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x;\tau_0)f_{X|T=1}(x;\tau_0)dx,\\
&\mathbb{E}\left\{\frac{T}{p_1}F_{Y(0)|X}(y_0|X)s_1(Y|X,T=1;\tau_0)\right\} = 0,\\
&\mathbb{E}\left\{\frac{T}{p_1}F_{Y(0)|X}(y_0|X)h_1(X|T=1;\tau_0)\right\} = \int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x;\tau_0)\frac{\partial f_{X|T=1}(x;\tau_0)}{\partial\tau}dx,\\
&\mathbb{E}\left\{\frac{T}{p_1}F_{Y(0)|X}(y_0|X)(T-p_1)\right\} = (1-p_1)\int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x)f_{X|T=1}(x;\tau_0)dx,\\
&\mathbb{E}\left\{\frac{T}{p_1}F_{01}(y_0)s_1(Y|X,T=1;\tau_0)\right\} = 0,\\
&\mathbb{E}\left\{\frac{T}{p_1}F_{01}(y_0)h_1(X|T=1;\tau_0)\right\} = 0,\\
&\mathbb{E}\left\{\frac{T}{p_1}F_{01}(y_0)(T-p_1)\right\} = (1-p_1)\int_{x\in\mathcal{X}} F_{Y(0)|X}(y_0|x)f_{X|T=1}(x;\tau_0)dx.
\end{aligned}
$$

Then we get

$$
\begin{aligned}
&\mathbb{E}\left[S_{F_{01}}(Y,T,X)\cdot s(Y,T,X;\tau_0)\right]\\
=&\int_{x\in\mathcal{X}}\frac{\partial F_{Y(0)|X}(y_0|x;\tau_0)}{\partial\tau}f_{X|T=1}(x;\tau_0)dx+\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0;\tau_0)\pi_{01}(x)\frac{\partial f_{X|T=0}(x;\tau_0)}{\partial\tau}dx\\
&-\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\times p_1\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0|x;\tau_0)f_{X|T=1}(x;\tau_0)dx\\
&-\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0|x;\tau_0)\pi_{01}(x)\frac{\partial f_{X|T=0}(x;\tau_0)}{\partial\tau}dx\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\times p_1\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0|x;\tau_0)f_{X|T=1}(x;\tau_0)dx\\
&+\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0|x;\tau_0)\frac{\partial f_{X|T=1}(x;\tau_0)}{\partial\tau}dx\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\times(1-p_1)\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0|x)f_{X|T=1}(x;\tau_0)dx\\
&-\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\times(1-p_1)\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0|x)f_{X|T=1}(x;\tau_0)dx\\
=&\int_{x\in\mathcal{X}}\frac{\partial F_{Y(0)|X}(y_0|x;\tau_0)}{\partial\tau}f_{X|T=1}(x;\tau_0)dx\\
&+\int_{x\in\mathcal{X}}F_{Y(0)|X}(y_0|x;\tau_0)h_1(x|T=1;\tau_0)f_{X|T=1}(x;\tau_0)dx\\
=&\frac{\partial F_{01}(y_0;\tau_0)}{\partial\tau}\ .
\end{aligned}
$$

We next verify Equation (A.5) holds:

$$
\begin{aligned}
&\mathbb{E}\left[S_{\theta_{011}}(Y,T,X)\cdot s(Y,T,X;\tau_0)\right] \\
=&\mathbb{E}\Bigg\{\Bigg[\frac{1-T}{p_0}\left\{\pi_{01}(X)-F_{11}(Y)\pi_{01}(X)-\pi_{01}(X)\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dy\right\} \\
&\qquad+\frac{T}{p_1}\left\{F_{01}(Y)-\theta_{011}+\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dy-\theta_{011}\right\}\Bigg] \\
&\times\Bigg[T\cdot s_1(Y|X,T=1;\tau_0)+(1-T)\cdot s_0(Y|X,T=0;\tau_0) \\
&+\frac{T-\mathbb{P}(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\cdot\mathbb{P}'(T=1;\tau_0)+T\cdot h_1(X|T=1;\tau_0)+(1-T)\cdot h_0(X|T=0;\tau_0)\Bigg]\Bigg\} \\
=&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)s_0(Y|X,T=0;\tau_0)\right\} \\
&+\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)h_0(X|T=0;\tau_0)\right\} \\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)(T-p_1)\right\} \\
&-\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y|T=1}(Y)s_0(Y|X,T=0;\tau_0)\right\} \\
&-\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y|T=1}(Y)h_0(X|T=0;\tau_0)\right\} \\
&-\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y|T=1}(Y)(T-p_1)\right\} \\
&-\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X))\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dys_0(Y|X,T=0;\tau_0)\right\} \\
&-\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X))\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dyh_0(X|T=0;\tau_0)\right\} \\
&-\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X))\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dy(T-p_1)\right\} \\
&+\mathbb{E}\left\{\frac{T}{p_1}\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dys_1(Y|X,T=0;\tau_0)\right\} \\
&+\mathbb{E}\left\{\frac{T}{p_1}\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dyh_1(X|T=0)\right\} \\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{T}{p_1}\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dy(T-p_1)\right\} \\
&+\mathbb{E}\left\{\frac{T}{p_1}F_{01}(Y)s_1(Y|X,T=0;\tau_0)\right\} \\
&+\mathbb{E}\left\{\frac{T}{p_1}F_{01}(Y)h_1(X|T=0;\tau_0)\right\} \\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{T}{p_1}F_{01}(Y)(T-p_1)\right\} \\
&-2\times\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}\mathbb{E}\left\{\frac{T}{p_1}\theta_{011}(T-p_1)\right\}.
\end{aligned}
$$

We compute the expectation terms of above expression, and obtain the following results:

$$\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)s_0(Y|X,T=0;\tau_0)\right\} = 0 \ ,$$
$$\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)h_0(X|T=0;\tau_0)\right\} = \int_{x\in\mathcal{X}} h_0(x|T=0;\tau_0)f_{X|T=1}(x;\tau_0)dx \ ,$$
$$\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)(T-p_1)\right\} = -p_1 \ ,$$

and

$$\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y|T=1}(Y)s_0(Y|X,T=0;\tau_0)\right\} \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} F_{Y|T=1}(y)\frac{\partial f_{Y|X,T=0}(y|x;\tau_0)}{\partial\tau} f_{X|T=1}(x;\tau_0)dxdy \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} \left[\int_{-\infty}^{y} f_{Y|T=1}(y')dy'\right]\frac{\partial f_{Y|X,T=0}(y|x;\tau_0)}{\partial\tau} f_{X|T=1}(x;\tau_0)dxdy \\
&= \int_{x\in\mathcal{X}}\int_{y'\in\mathcal{Y}} \left[\int_{y'}^{+\infty} \frac{\partial f_{Y|X,T=0}(y|x;\tau_0)}{\partial\tau}dy\right] f_{Y|T=1}(y')f_{X|T=1}(x;\tau_0)dxdy' \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} \left[\frac{\partial(1-F_{Y|T=0,X}(y|x;\tau_0))}{\partial\tau}\right] f_{Y|T=1}(y)f_{X|T=1}(x;\tau_0)dxdy \\
&= -\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} \frac{\partial F_{Y|T=0,X}(y|x;\tau_0)}{\partial\tau} f_{X|T=1}(x;\tau_0)f_{Y|T=1}(y)dydx \ ,
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y|T=1}(Y)h_0(X|T=0;\tau_0)\right\} \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} F_{Y|T=1}(y)h_0(x|T=0;\tau_0)f_{Y|X,T=0}(y|x;\tau_0)f_{X|T=1}(x;\tau_0)dxdy,
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X)F_{Y|T=1}(Y)(T-p_1)\right\}\\
&=-p_1\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}}F_{Y|T=1}(y)f_{Y|X,T=0}(y|x;\tau_0)f_{X|T=1}(x;\tau_0)dxdy\\
&=-p_1(1-\theta_{011}),\\
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X))\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dys_0(Y|X,T=0;\tau_0)\right\}=0\ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X))\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dyh_0(X|T=0;\tau_0)\right\}\\
=&\int_{x\in\mathcal{X}}\left\{\int_{y\in\mathcal{Y}}F_{Y|T=0,X}(y|x)f_{Y|T=1}(y)dy\right\}h_0(x|T=0;\tau_0)f_{X|T=1;\tau_0}(x)dx\\
=&\int_{x\in\mathcal{X}}\left\{1-\int_{y\in\mathcal{Y}}F_{Y|T=1}(y)f_{Y|T=0,X}(y|x)dy\right\}h_0(x|T=0;\tau_0)f_{X|T=1;\tau_0}(x)dx\\
=&\int_{x\in\mathcal{X}}h_0(x|T=0;\tau_0)f_{X|T=1}(x;\tau_0)dx\\
&-\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}}F_{Y|T=1}(y)h_0(x|T=0;\tau_0)f_{Y|X,T=0}(y|x;\tau_0)f_{X|T=1}(x;\tau_0)dxdy\ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{1}{p_0}(1-T)\pi_{01}(X))\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dy(T-p_1)\right\}\\
=&-p_1\int_{x\in\mathcal{X}}\left\{\int_{y\in\mathcal{Y}}F_{Y|T=0,X}(y|x;\tau_0)f_{Y|T=1}(y)dy\right\}f_{X|T=1}(x;\tau_0)dx\\
=&-p_1\theta_{011}\ ,
\end{aligned}
$$

and

$$
\mathbb{E}\left\{\frac{T}{p_1}\int F_{Y(0)|X}(y\mid X)f_{Y(1)|T=1}(y)dys_1(Y|X,T=1;\tau_0)\right\}=0\ ,
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{T}{p_1}\int F_{Y(0)|X}(y \mid X)f_{Y(1)|T=1}(y)dyh_1(X|T=1;\tau_0)\right\} \\
=&\int_{x\in\mathcal{X},y\in\mathcal{Y}} F_{Y(0)|X}(y|x;\tau_0)f_{Y|T=1}(y)h_1(x|T=1;\tau_0)f_{X|T=1}(x;\tau_0)dydx \ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{T}{p_1}\int F_{Y(0)|X}(y \mid X)f_{Y(1)|T=1}(y)dy(T-p_1)\right\} \\
=&(1-p_1)\int_{\mathcal{X}}\left\{\int_{\mathcal{Y}} F_{Y|X,T=0}(y'|x;\tau_0)f_{Y|T=1}(y')dy'\right\}f_{X|T=1}(x;\tau_0)dx=(1-p_1)\theta_{011} \ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{T}{p_1}F_{01}(Y)s_1(Y|X,T=0;\tau_0)\right\} \\
=&\int_{x\in\mathcal{X}}\left\{\int_{y\in\mathcal{Y}}\left[\int_{x'\in\mathcal{X}} F_{Y(0)|X}(y|x')f_{X|T=1}(x';\tau_0)dx'\right]\cdot s_1(y|x,T=0;\tau_0)f_{Y|X,T=1}(y|x;\tau_0)dy\right\} \\
&f_{X|T=1}(x;\tau_0)dx,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{T}{p_1}F_{01}(Y)h_1(X|T=0;\tau_0)\right\} \\
=&\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}}\left[\int_{x'\in\mathcal{X}} F_{Y(0)|X}(y|x';\tau_0)f_{X|T=1}(x';\tau_0)dx'\right]f_{Y|X,T=1}(y|x;\tau_0)h_1(x|T=0;\tau_0) \\
&f_{X|T=1}(x;\tau_0)dxdy \ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left\{\frac{T}{p_1}F_{01}(Y)(T-p_1)\right\}=(1-p_1)\cdot\mathbb{E}[F_{01}(Y)|T=1]=(1-p_1)\theta_{011} \ , \\
&\mathbb{E}\left\{\frac{T}{p_1}\theta_{011}(T-p_1)\right\}=(1-p_1)\theta_{011}.
\end{aligned}
$$

Then we get

$$
\begin{aligned}
&\mathbb{E}\left[S_{\theta_{011}}(Y,T,X)\cdot s(Y,T,X;\tau_0)\right]\\
=&\int_{x\in\mathcal{X}} h_0(x|T=0;\tau_0)f_{X|T=1}(x;\tau_0)dx-\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}p_1\\
&+\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}}\frac{\partial F_{Y|T=0,X}(y|x;\tau_0)}{\partial\tau}f_{X|T=1}(x;\tau_0)f_{Y|T=1}(y)dydx\\
&-\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}}F_{Y|T=1}(y)h_0(x|T=0;\tau_0)f_{Y|X,T=0}(y|x;\tau_0)f_{X|T=1}(x;\tau_0)dxdy\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}p_1(1-\theta_{011})\\
&-\int_{x\in\mathcal{X}} h_0(x|T=0;\tau_0)f_{X|T=1}(x;\tau_0)dx\\
&+\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}}F_{Y|T=1}(y)h_0(x|T=0;\tau_0)f_{Y|X,T=0}(y|x;\tau_0)f_{X|T=1}(x;\tau_0)dxdy\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}p_1\theta_{011}\\
&+\int_{x\in\mathcal{X},y\in\mathcal{Y}}F_{Y(0)|X}(y|x)f_{Y|T=1}(y|x;\tau_0)h_1(x|T=1;\tau_0)f_{X|T=1}(x;\tau_0)dydx\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}(1-p_1)\theta_{011}\\
&+\int_{x\in\mathcal{X}}\left\{\int_{y\in\mathcal{Y}}\left[\int_{x'\in\mathcal{X}}F_{Y(0)|X}(y|x';\tau_0)f_{X|T=1}(x';\tau_0)dx'\right]\cdot s_1(y|x,T=0;\tau_0)f_{Y|X,T=1}(y|x;\tau_0)dy\right\}\\
&\quad f_{X|T=1}(x;\tau_0)dx\\
&+\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}}\left[\int_{x'\in\mathcal{X}}F_{Y(0)|X}(y|x';\tau_0)f_{X|T=1}(x';\tau_0)dx'\right]f_{Y|X,T=1}(y|x;\tau_0)h_1(x|T=0;\tau_0)f_{X|T=1}(x;\tau_0)dxdy\\
&+\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}(1-p_1)\theta_{011}\\
&-2\times\frac{\mathbb{P}'(T=1;\tau_0)}{\mathbb{P}(T=1;\tau_0)(1-\mathbb{P}(T=1;\tau_0))}(1-p_1)\theta_{011}\\
=&\int_{y\in\mathcal{Y}}\int_{x\in\mathcal{X}}\frac{\partial F_{Y|T=0,X}(y|x;\tau_0)}{\partial\tau}f_{X|T=1}(x;\tau_0)dxf_{Y|T=1}(y;\tau_0)dy\\
&+\int_{y\in\mathcal{Y}}\int_{x\in\mathcal{X}}F_{Y|T=0,X}(y|x;\tau_0)h_1(x|T=1;\tau_0)f_{X|T=1}(x;\tau_0)dxf_{Y|T=1}(y;\tau_0)dy\\
&+\int_{y\in\mathcal{Y}}\left[\int_{x\in\mathcal{X}}F_{Y|T=0,X}(y|x;\tau_0)f_{X|T=1}(x;\tau_0)dx\right]\left[\int_{x\in\mathcal{X}}\left\{s_1(y|x,T=1;\tau_0)f_{Y|X,T=1}(y|x;\tau_0)f_{X|T=1}(x;\tau_0)\right.\right.\\
&\left.\left.+f_{Y|X,T=1}(y|x;\tau_0)h_1(x|T=1;\tau_0)f_{X|T=1}(x;\tau_0)\right\}dx\right]dy\\
=&\frac{\partial\theta_{011}(\tau_0)}{\partial\tau},
\end{aligned}
$$

which justifies Equation (A.5).

A.2 Proof of Theorem 2.4.6

This section establishes the convergence rates of $\hat{\pi}_0 \to \pi_0$, $\hat{\pi}_1 \to \pi_1$, $\hat{\pi}_{01} \to \pi_{01}$ and $\hat{\pi}_{10} \to \pi_{10}$. Because of Assumption 2.4.4 (i), we can assume the sieve basis $u_K(X)$ are orthonormalized, namely

$$\mathbb{E}\left[u_K(X)u_K^\top(X)\right] = I_K. \tag{A.6}$$

We introduce some notation which will be used later. Let $\{Q_0^*(\lambda), Q_1^*(\lambda), Q_{01}^*(\lambda), Q_{10}^*(\lambda)\}$ be the theoretical counterparts of $\{\hat{Q}_0(\lambda), \hat{Q}_1(\lambda), \hat{Q}_{01}(\lambda), \hat{Q}_{10}(\lambda)\}$ respectively:

$$\begin{aligned}
Q_0^*(\lambda) &:= \mathbb{E}\left[(1-T)\rho(\lambda^\top u_K(X)) - \lambda^\top u_K(X)\right], \\
Q_1^*(\lambda) &:= \mathbb{E}\left[T\rho(\lambda^\top u_K(X)) - \lambda^\top u_K(X)\right], \\
Q_{01}^*(\lambda) &:= \mathbb{E}\left[\rho(\lambda^\top u_K(X))|T=0\right] - \mathbb{E}\left[\lambda^\top u_K(X)|T=1\right], \\
Q_{10}^*(\lambda) &:= \mathbb{E}\left[\rho(\lambda^\top u_K(X))|T=1\right] - \mathbb{E}\left[\lambda^\top u_K(X)|T=0\right].
\end{aligned}$$

Note that $Q_0^*(\lambda), Q_1^*(\lambda), Q_{01}^*(\lambda)$ and $Q_{10}^*(\lambda)$ are strictly concave functions, and we denote λ_1^*, λ_0^*, λ_{01}^* and λ_{10}^* to be the unique maximizer of $Q_0^*(\lambda), Q_1^*(\lambda), Q_{01}^*(\lambda)$ and $Q_{10}^*(\lambda)$ respectively. Moreover, we define

$$\begin{aligned}
\pi_0^*(x) &:= \rho'\left((\lambda_0^*)^\top u_K(x)\right), \\
\pi_1^*(x) &:= \rho'\left((\lambda_1^*)^\top u_K(x)\right), \\
\pi_{01}^*(x) &:= \rho'\left((\lambda_{01}^*)^\top u_K(x)\right), \\
\pi_{10}^*(x) &:= \rho'\left((\lambda_{10}^*)^\top u_K(x)\right).
\end{aligned}$$

Theorem 2.4.6 is a consequence of combining Lemmas A.2.1 and A.2.2 whose proof will be given in the following two subsections.

A.2.1 Lemma $A.2.1$

Lemma A.2.1. *Under Assumptions 2.4.1-2.4.5, we have*

$$\int_{\mathcal{X}} |\pi_1(x) - \pi_1^*(x)|^2 \, dF_X(x) = O\left(K^{-2\alpha}\right) \ ,$$
$$\int_{\mathcal{X}} |\pi_0(x) - \pi_0^*(x)|^2 \, dF_X(x) = O\left(K^{-2\alpha}\right) \ ,$$
$$\int_{\mathcal{X}} |\pi_{10}(x) - \pi_{10}^*(x)|^2 \, dF_X(x) = O\left(K^{-2\alpha}\right) \ ,$$
$$\int_{\mathcal{X}} |\pi_{01}(x) - \pi_{01}^*(x)|^2 \, dF_X(x) = O\left(K^{-2\alpha}\right) \ ,$$

and

$$\frac{1}{N}\sum_{i=1}^{N} |\pi_1(X_i) - \pi_1^*(X_i)|^2 = O_P\left(K^{-2\alpha}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^{N} |\pi_0(X_i) - \pi_0^*(X_i)|^2 = O_P\left(K^{-2\alpha}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^{N} |\pi_{10}(X_i) - \pi_{10}^*(X_i)|^2 = O_P\left(K^{-2\alpha}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^{N} |\pi_{01}(X_i) - \pi_{01}^*(X_i)|^2 = O_P\left(K^{-2\alpha}\right) \ .$$

Proof. We prove

$$\int_{\mathcal{X}} |\pi_{01}(x) - \pi_{01}^*(x)|^2 \, dF_X(x) = O\left(K^{-2\alpha}\right),$$
$$\frac{1}{N}\sum_{i=1}^{N} |\pi_{01}(X_i) - \pi_{01}^*(X_i)|^2 = O_P\left(K^{-2\alpha}\right),$$

and other results can be established by using a similar argument. By Assumption 2.4.1, there exist two positive constants η_1 and η_2 such that $\pi_{01}(x) \in [\eta_1, \eta_2], \ \forall x \in \mathcal{X}$. By Assumption 2.4.5, $(\rho')^{-1}$

is strictly decreasing, and we define

$$\overline{\gamma} := \sup_{x\in\mathcal{X}}(\rho')^{-1}(\pi_{01}(x)) \le (\rho')^{-1}(\eta_1) \;\text{ and }\; \underline{\gamma} := \inf_{x\in\mathcal{X}}(\rho')^{-1}(\pi_{01}(x)) \ge (\rho')^{-1}(\eta_2)\ ,$$

which are finite constants. By Assumption 2.4.2, there exist $C_1 > 0$ and $\lambda_{01K} \in \mathbb{R}^K$ such that

$$\sup_{x\in\mathcal{X}}\left|(\rho')^{-1}(\pi_{01}(x)) - \lambda_{01K}^\top u_K(x)\right| < C_1 K^{-\alpha}, \tag{A.7}$$

which implies that for all $x \in \mathcal{X}$,

$$\lambda_{01K}^\top u_K(x) \in \big((\rho')^{-1}(\pi_{01}(x)) - C_1K^{-\alpha}, (\rho')^{-1}(\pi_{01}(x)) + C_1K^{-\alpha}\big) \subset [\underline{\gamma} - C_1K^{-\alpha}, \overline{\gamma} + C_1K^{-\alpha}], \tag{A.8}$$

and

$$\begin{aligned}&\rho'(\lambda_{01K}^\top u_K(x) + C_1K^{-\alpha}) - \rho'(\lambda_{01K}^\top u_K(x)) < \pi_{01}(x) - \rho'(\lambda_{01K}^\top u_K(x))\\ &< \rho'(\lambda_{01K}^\top u_K(x) - C_1K^{-\alpha}) - \rho'(\lambda_{01K}^\top u_K(x)),\ \ \forall x \in \mathcal{X}.\end{aligned}$$

By Mean Value Theorem, for large enough K, there exist

$$\begin{aligned}&\xi_1(x) \in (\lambda_{01K}^\top u_K(x), \lambda_{01K}^\top u_K(x) + C_1K^{-\alpha}) \subset [\underline{\gamma} - C_1K^{-\alpha}, \overline{\gamma} + 2C_1K^{-\alpha}] \subset \Gamma_1\\ &\xi_2(x) \in (\lambda_{01K}^\top u_K(x) - C_1K^{-\alpha}, \lambda_{01K}^\top u_K(x)) \subset [\underline{\gamma} - 2C_1K^{-\alpha}, \overline{\gamma} + C_1K^{-\alpha}] \subset \Gamma_1\ ,\end{aligned}$$

where

$$\Gamma_1 := [\underline{\gamma} - 1, \overline{\gamma} + 1]\ ,$$

such that for all $x \in \mathcal{X}$

$$\rho'(\lambda_{01K}^\top u_K(x) + C_1 K^{-\alpha}) - \rho'(\lambda_{01K}^\top u_K(x)) = \rho''(\xi_1(x)) C_1 K^{-\alpha} \geq -a_1 C_1 K^{-\alpha},$$
$$\rho'(\lambda_{01K}^\top u_K(x) - C_1 K^{-\alpha}) - \rho'(\lambda_{01K}^\top u_K(x)) = -\rho''(\xi_2(x)) C_1 K^{-\alpha} \leq a_2 C_1 K^{-\alpha}\ ,$$

where $-a_1 := \inf_{\gamma \in \Gamma_1} \rho''(\gamma)$, $a_2 := \sup_{\gamma \in \Gamma_1} (-\rho''(\gamma))$. Note that $a := \max\{a_1, a_2\}$ is a finite constant because Γ_1 is compact and independent of x. Therefore, we have

$$\sup_{x \in \mathcal{X}} \left| \pi_{01}(x) - \rho'(\lambda_{01K}^\top u_K(x)) \right| < a C_1 K^{-\alpha}. \tag{A.9}$$

For some fixed $C_2 > 0$ (to be chosen later), define

$$\Lambda_K := \left\{ \lambda \in \mathbb{R}^K : \|\lambda - \lambda_{01K}\| \leq C_2 K^{-\alpha} \right\}.$$

For sufficiently large K, by (A.8), Assumption 2.4.4, we have $\forall \lambda \in \Lambda_K$, $\forall x \in \mathcal{X}$,

$$|\lambda^\top u_K(x) - \lambda_{01K}^\top u_K(x)| = |(\lambda - \lambda_{01K})^\top u_K(x)| \leq \|\lambda - \lambda_{01K}\| \|u_K(x)\| \leq C_2 K^{-\alpha} \zeta(K)$$
$$\Rightarrow \lambda^\top u_K(x) \in (\lambda_{01K}^\top u_K(x) - C_2 K^{-\alpha}\zeta(K), \lambda_{01K}^\top u_K(x) + C_2 K^{-\alpha}\zeta(K))$$
$$\subset \left[\underline{\gamma} - C_1 K^{-\alpha} - C_2 K^{-\alpha}\zeta(K), \overline{\gamma} + C_1 K^{-\alpha} + C_2 K^{-\alpha}\zeta(K) \right] \subset \Gamma_1\ . \tag{A.10}$$

By (A.9), (A.10), and $\mathbb{E}[u_K(X)u_K(X)^\top] = I_K$, we can deduce that

$$
\begin{aligned}
&\|(Q_{01}^*)'(\lambda_{01K})\| = \left\|\mathbb{E}\left[\rho'\left(\lambda_{01K}^\top u_K(X)\right)u_K(X)|T=0\right] - \mathbb{E}\left[u_K(X)|T=1\right]\right\| \\
=&\left\|\mathbb{E}\left[\rho'\left(\lambda_{01K}^\top u_K(X)\right)u_K(X) - \pi_{01}(X)u_K(X)|T=0\right]\right\| \\
=&\frac{1}{p_0}\left\|\mathbb{E}\left[(1-T)\left\{\rho'\left(\lambda_{01K}^\top u_K(X)\right) - \pi_{01}(X)\right\}u_K(X)\right]\right\| \\
=&\frac{1}{p_0}\sqrt{\mathbb{E}\left[\left\{\mathbb{E}\left[(1-T)\left\{\rho'\left(\lambda_{01K}^\top u_K(X)\right) - \pi_{01}(X)\right\}u_K(X)^\top\right]\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right]^{-1}u_K(X)\right\}^2\right]} \\
=&\frac{1}{p_0}\left\|\mathbb{E}\left[\left\{(1-T)\rho'\left(\lambda_{01K}^\top u_K(X)\right) - \pi_{01}(X)\right\}u_K(X)^\top\right]\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right]^{-1}u_K(X)\right\|_{L^2} \\
\leq&\frac{1}{p_0}\left\|(1-T)\left\{\rho'\left(\lambda_{01K}^\top u_K(X)\right) - \pi_{01}(X)\right\}\right\|_{L^2} \leq \frac{a}{p_0}C_1K^{-\alpha}\,, \qquad \text{(A.11)}
\end{aligned}
$$

where the third equality follows from the definition of Frobenius norm and the fact $\mathbb{E}[u_K(X)u_K(X)^\top] = I_K$; the first inequality follows from the fact that

$$
\mathbb{E}\left[(1-T)\left\{\rho'\left(\lambda_{01K}^\top u_K(X)\right) - \pi_{01}(X)\right\}u_K(X)^\top\right]\cdot\mathbb{E}\left[u_K(X)u_K(X)^\top\right]^{-1}u_K(X)
$$

is the $L^2(dF_X)$-projection of $(1-T)\left\{\rho'\left(\lambda_{01K}^\top u_K(X)\right) - \pi_{01}(X)\right\}$ on the space spanned by $u_K(X)$.

Then for any $\lambda \in \partial\Lambda_K$, i.e. $\|\lambda - \lambda_{01K}\| = C_2K^{-\alpha}$, we can deduce that

$$
\begin{aligned}
&Q_{01}^*(\lambda) - Q_{01}^*(\lambda_{01K}) \\
&= (\lambda-\lambda_{01K})^\top(Q_{01}^*)'(\lambda_{01K}) + \frac{1}{2}(\lambda-\lambda_{01K})^\top(Q_{01}^*)''(\bar{\lambda}_{01})(\lambda-\lambda_{01K}) \\
&\leq \|\lambda-\lambda_{01K}\|\|(Q_{01}^*)'(\lambda_{01K})\| + \frac{1}{2}(\lambda-\lambda_{01K})^\top\mathbb{E}\left[\rho''\left(\bar{\lambda}_{01}^\top u_K(X)\right)u_K(X)u_K^\top(X)|T=0\right](\lambda-\lambda_{01K}) \\
&\leq \|\lambda-\lambda_{01K}\|\|(Q_{01}^*)'(\lambda_{01K})\| - \frac{a_1}{2p_0}(\lambda-\lambda_{01K})^\top\mathbb{E}[(1-T)u_K(X)u_K^\top(X)](\lambda-\lambda_{01K}) \\
&= \|\lambda-\lambda_{01K}\|\|(Q_{01}^*)'(\lambda_{01K})\| - \frac{a_1}{2p_0}(\lambda-\lambda_{01K})^\top\mathbb{E}[P(T=0|X)u_K(X)u_K^\top(X)](\lambda-\lambda_{01K}) \\
&\leq \|\lambda-\lambda_{01K}\|\|(Q_{01}^*)'(\lambda_{01K})\| - \frac{a_1}{2p_0}\|\lambda-\lambda_{01K}\|^2\cdot\inf_{x\in\mathcal{X}}\{P(T=0|X=x)\}\cdot\lambda_{\min}(\mathbb{E}[u_K(X)u_K^\top(X)]) \\
&= \|\lambda-\lambda_{01K}\|\left(\|(Q_{01}^*)'(\lambda_{01K})\| - \frac{a_1}{2p_0}\|\lambda-\lambda_{01K}\|\cdot\inf_{x\in\mathcal{X}}\{P(T=0|X=x)\}\right) \\
&\leq \|\lambda-\lambda_{01K}\|\left(\frac{a}{p_0}C_1K^{-\alpha} - \frac{a_1}{2p_0}\cdot C_2K^{-\alpha}\cdot\inf_{x\in\mathcal{X}}\{P(T=0|X=x)\}\right), \qquad \text{(A.12)}
\end{aligned}
$$

where $a_1 = \inf_{y\in\Gamma_1}\{-\rho''(y)\} > 0$ is a finite positive constant, and the last inequality follows from

(A.11). By choosing

$$C_2 > \frac{2aC_1}{a_1 \cdot \inf_{x\in\mathcal{X}} \{P(T=0|X=x)\}} ,$$

we can obtain that

$$Q_{01}^*(\lambda) < Q_{01}^*(\lambda_{01K}) , \quad \lambda \in \partial\Lambda_K .$$

In light of the continuity of Q_{01}^*, there is a local maximum of Q_{01}^* in the interior of Λ_K. On the other hand, Q_{01}^* is a strictly concave function with a unique global maximum point λ_{01}^*, therefore we can claim

$$\lambda_{01}^* \in \Lambda_K^\circ , \ \ i.e. \ \|\lambda_{01}^* - \lambda_{01K}\| \le C_2 K^{-\alpha} . \tag{A.13}$$

By Mean Value Theorem, for large enough K, there exists $\xi^*(x)$ in between $(\lambda_{01}^*)^\top u_K(x)$ and $\lambda_{01K}^\top u_K(x)$, i.e. $\xi^*(x) \in \Gamma_1$ such that for any $x \in \mathcal{X}$,

$$|\rho'(\lambda_{01K}^\top u_K(x)) - \rho'((\lambda_{01}^*)^\top u_K(x))| = |\rho''(\xi^*(x))||\lambda_{01K}^\top u_K(x) - (\lambda_{01}^*)^\top u_K(x)|. \tag{A.14}$$

By (A.9), (A.13), (A.14) and the fact $\mathbb{E}[u_K(X)u_K(X)^\top] = I_K$, we can deduce that

$$\begin{aligned}
&\int_{\mathcal{X}} |\pi_{01}(x) - \pi_{01}^*(x)|^2 \, dF_X(x) \\
\le& 2\int_{\mathcal{X}} \left|\pi_{01}(x) - \rho'\left(\lambda_{01K}^\top u_K(x)\right)\right|^2 dF_X(x) + 2\int_{\mathcal{X}} \left|\rho'\left(\lambda_{01K}^\top u_K(x)\right) - \rho'\left((\lambda_{01}^*)^\top u_K(x)\right)\right|^2 dF_X(x) \\
\le& 2\sup_{x\in\mathcal{X}} \left|\pi_{01}(x) - \rho'\left(\lambda_{01K}^\top u_K(x)\right)\right|^2 + 2\sup_{x\in\mathcal{X}} |\rho''(\xi^*(x))|^2 \cdot (\lambda_{01K} - \lambda_{01}^*)^\top \cdot \mathbb{E}[u_K(x)u_K(x)^\top] \cdot (\lambda_{01K} - \lambda_{01}^*) \\
=& O(K^{-2\alpha}) + O(1) \cdot O(K^{-2\alpha}) = O(K^{-2\alpha}).
\end{aligned}$$

We can also obtain that

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left|\pi_{01}(X_i)-\pi_{01}^{*}(X_i)\right|^2\\
\leq&\frac{2}{N}\sum_{i=1}^{N}\left|\pi_{01}(X_i)-\rho'\left(\lambda_{01K}^{\top}u_K(X_i)\right)\right|^2+\frac{2}{N}\sum_{i=1}^{N}\left|\rho'\left(\lambda_{01K}^{\top}u_K(X_i)\right)-\rho'\left((\lambda_{2K}^{*})^{\top}u_K(X_i)\right)\right|^2\\
\leq&2\sup_{x\in\mathcal{X}}\left|\pi_{01}(x)-\rho'\left(\lambda_{01K}^{\top}u_K(x)\right)\right|^2\\
&+2\sup_{x\in\mathcal{X}}|\rho''(\xi^*(x))|^2\cdot(\lambda_{01K}-\lambda_{01}^{*})^{\top}\cdot\left\{\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^{\top}\right\}(\lambda_{01K}-\lambda_{01}^{*})\\
\leq&O(K^{-2\alpha})+2\sup_{x\in\mathcal{X}}|\rho''(\xi^*(x))|^2\cdot\|\lambda_{01K}-\lambda_{01}^{*}\|^2\cdot\lambda\text{max}\left\{\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^{\top}\right\}\\
=&O(K^{-2\alpha})+O(1)\cdot O(K^{-2\alpha})\cdot O_P(1)=O(K^{-2\alpha})\ ,
\end{aligned}
$$

where $\lambda_{\max}(A)$ denotes the largest eigenvalue of a matrix A; the second equality follows from Chebyshev's inequality and the following facts

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}u_K(X_i)u_K(X_i)^{\top}-\mathbb{E}[u_K(X)u_K(X)^{\top}]\right\|^2\right] \qquad\qquad \text{(A.15)}\\
=&\frac{1}{N}\mathbb{E}\left[\left\|u_K(X)u_K(X)^{\top}-\mathbb{E}[u_K(X)u_K(X)^{\top}]\right\|^2\right]\\
=&\frac{1}{N}\mathbb{E}\left[\text{tr}\left\{u_K(X)u_K(X)^{\top}u_K(X)u_K(X)^{\top}\right\}\right]-\frac{1}{N}\text{tr}\left\{\mathbb{E}[u_K(X)u_K(X)^{\top}]\cdot\mathbb{E}[u_K(X)u_K(X)^{\top}]\right\}\\
\leq&\frac{1}{N}\cdot\zeta(K)^2\mathbb{E}\left[\|u_K(X)\|^2\right]=\zeta(K)^2\frac{K}{N}\to 0\ ,
\end{aligned}
$$

and $\mathbb{E}\left[u_K(X)u_K(X)^{\top}\right]=I_K$. □

A.2.2 Lemma $A.2.2$

Lemma A.2.2. *Under Assumptions 2.4.1-2.4.5, we have*

$$\|\hat{\lambda}_0 - \lambda_0^*\| = O_P\left(\sqrt{\frac{K}{N}}\right),$$

$$\|\hat{\lambda}_1 - \lambda_1^*\| = O_P\left(\sqrt{\frac{K}{N}}\right),$$

$$\|\hat{\lambda}_{01} - \lambda_{01}^*\| = O_P\left(\sqrt{\frac{K}{N}}\right),$$

$$\|\hat{\lambda}_{10} - \lambda_{10}^*\| = O_P\left(\sqrt{\frac{K}{N}}\right),$$

and

$$\int_{\mathcal{X}} |\hat{\pi}_0(x) - \pi_0^*(x)|^2 dF_X(x) = O_P\left(\frac{K}{N}\right) \ ,$$

$$\int_{\mathcal{X}} |\hat{\pi}_1(x) - \pi_1^*(x)|^2 dF_X(x) = O_P\left(\frac{K}{N}\right) \ ,$$

$$\int_{\mathcal{X}} |\hat{\pi}_{01}(x) - \pi_{01}^*(x)|^2 dF_X(x) = O_P\left(\frac{K}{N}\right) \ ,$$

$$\int_{\mathcal{X}} |\hat{\pi}_{10}(x) - \pi_{10}^*(x)|^2 dF_X(x) = O_P\left(\frac{K}{N}\right) \ ,$$

and

$$\frac{1}{N}\sum_{i=1}^{N}|\hat{\pi}_0(X_i)-\pi_0^*(X_i)|^2 = O_P\left(\frac{K}{N}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^{N}|\hat{\pi}_1(X_i)-\pi_1^*(X_i)|^2 = O_P\left(\frac{K}{N}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^{N}|\hat{\pi}_{01}(X_i)-\pi_{01}^*(X_i)|^2 = O_P\left(\frac{K}{N}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^{N}|\hat{\pi}_{10}(X_i)-\pi_{10}^*(X_i)|^2 = O_P\left(\frac{K}{N}\right) \ ,$$

Proof. We prove the results:

$$\|\hat{\lambda}_{01}-\lambda_{01}^*\| = O_P\left(\sqrt{\frac{K}{N}}\right),$$
$$\int_{\mathcal{X}}|\hat{\pi}_{01}(x)-\pi_{01}^*(x)|^2 dF_X(x) = O_P\left(\frac{K}{N}\right) \ ,$$
$$\frac{1}{N}\sum_{i=1}^{N}|\hat{\pi}_{01}(X_i)-\pi_{01}^*(X_i)|^2 = O_P\left(\frac{K}{N}\right),$$

and other results can be established by using the similar argument. Define

$$\hat{S}_N := \frac{1}{N}\sum_{i=1}^{N}(1-T_i)u_K(X_i)u_K(X_i)^\top.$$

Obviously, $\hat{S}_N$ is a symmetric matrix and $\mathbb{E}[\hat{S}_N] = \mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right]$. We have

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\hat{S}_N - \mathbb{E}\left[(1-T)u_K(X)u_K^\top(X)\right]\right\|^2\right] \\
=&\operatorname{tr}\Bigg(\mathbb{E}[\hat{S}_N\hat{S}_N] - 2\mathbb{E}[\hat{S}_N]\cdot\mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right] \\
&\quad + \mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right]\cdot\mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right]\Bigg) \\
=&\operatorname{tr}\Bigg(\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N(1-T_i)^2u_K(X_i)u_K^\top(X_i)u_K(X_i)u_K^\top(X_i)\right] \qquad \text{(A.16)}\\
&\quad + \mathbb{E}\left[\frac{1}{N^2}\sum_{i,j=1,i\neq j}^N(1-T_i)(1-T_j)u_K(X_i)u_K^\top(X_i)u_K(X_j)u_K^\top(X_j)\right]. \\
&\quad - \mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right]\mathbb{E}[P(T=0|X)u_K(X)u_K^\top(X)]\Bigg) \\
=&\frac{1}{N}\mathbb{E}\left[P(T=0|X)u_K(X)^\top u_K(X)u_K(X)^\top u_K(X)\right] \\
&\quad - \frac{1}{N}\operatorname{tr}\left(\mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right]\cdot\mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right]\right) \\
\leq&\zeta(K)^2\frac{K}{N}, \qquad \text{(A.17)}
\end{aligned}
$$

where the last inequality follows from the facts $\sup_{x\in\mathcal{X}}\|u_K(x)\| \leq \zeta(K)$, $0 < P(T=0|X) < 1$, and $\mathbb{E}[u_K^\top(X)u_K(X)] = K$.

Consider the event set

$$
E_N := \left\{(\lambda-\lambda_{01}^*)^\top\hat{S}_N(\lambda-\lambda_{01}^*) > (\lambda-\lambda_{01}^*)^\top\left(\mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right] - \frac{\eta_1'}{2}\cdot I_K\right)(\lambda-\lambda_{01}^*), \lambda\neq\lambda_{01}^*\right\},
$$

where $\eta_1' := \inf_{x\in\mathcal{X}} P(T=0|X=x)$. By Chebyshev's inequality, (A.16), and Assumption 2.4.4 we have

$$
\begin{aligned}
&\mathbb{P}\left(\left|(\lambda-\lambda_{01}^*)^\top\hat{S}_N(\lambda-\lambda_{01}^*) - (\lambda-\lambda_{01}^*)^\top\mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right](\lambda-\lambda_{01}^*)\right| \geq \frac{\eta_1'}{2}\|\lambda-\lambda_{01}^*\|^2, \lambda\neq\lambda_{01}^*\right) \\
&\leq \frac{4\cdot\|\lambda-\lambda_{01}^*\|^4\mathbb{E}\left[\left\|\hat{S}_N - \mathbb{E}\left[(1-T)u_K(X)u_K^\top(X)\right]\right\|^2\right]}{{\eta_1'}^2\cdot\|\lambda-\lambda_{01}^*\|^4} \leq \frac{4\cdot\zeta(K)^2K}{{\eta_1'}^2N} \to 0.
\end{aligned}
$$

Hence, for any $\epsilon > 0$, there exists $N_0(\epsilon)\in\mathbb{N}$ such that $N > N_0(\epsilon)$ large enough

$$
\mathbb{P}\left((E_N)^c\right) < \frac{\epsilon}{2}\,. \qquad \text{(A.18)}
$$

Note that λ_{01}^* is the unique maximizer of $Q_{01}^*(\lambda)$, which implies

$$(Q_{01}^*)'(\lambda_{01}^*) = \mathbb{E}[\rho'((\lambda_{01}^*)^\top u_K(X))u_K(X)|T=0] - \mathbb{E}[u_K(X)|T=1] = 0.$$

Define

$$\bar{Q}_{01}(\lambda) := \frac{1}{Np_0}\sum_{i=1}^N (1-T_i)\rho'(\lambda^\top u_K(X_i)) - \frac{1}{Np_1}\sum_{j=1}^N T_j \lambda^\top u_K(X_j),$$

where $p_1 = \mathbb{P}(T=1)$ and $p_0 = \mathbb{P}(T=0)$. Note that $\bar{Q}_{01}(\lambda)$ is a concave function, and denote the unique maximizer of $\bar{Q}_{01}(\lambda)$ by $\bar{\lambda}_{01}$.

We first compute the order of $\|\bar{\lambda}_{01} - \lambda_{01}^*\|$. Note that

$$\bar{Q}_{01}'(\lambda_{01}^*) = \frac{1}{Np_0}\sum_{i=1}^N (1-T_i)\rho'((\lambda_{01}^*)^\top u_K(X_i))u_K(X_i) - \frac{1}{Np_1}\sum_{j=1}^N T_j u_K(X_j),$$

then for K sufficiently large

$$\begin{aligned}
&\mathbb{E}[\|\bar{Q}_{01}'(\lambda_{01}^*)\|^2] = \mathbb{E}[\bar{Q}_{01}'(\lambda_{01}^*)^\top \bar{Q}_{01}(\lambda_{01}^*)] \\
=&\frac{1}{N}\mathbb{E}\left[\left\|\frac{(1-T)}{p_0}\rho'((\lambda_{01}^*)^\top u_K(X))u_K(X) - \frac{T}{p_1}u_K(X)\right\|^2\right] \\
=&\frac{1}{Np_0}\mathbb{E}\left[\left(\rho'\left((\lambda_{01}^*)^\top u_K(X)\right)\right)^2 u_K^\top(X)u_K(X)|T=0\right] + \frac{1}{Np_1}\mathbb{E}\left[u_K^\top(X)u_K(X)|T=1\right] \\
=&\frac{1}{Np_0}\mathbb{E}\left[\left(\rho'\left((\lambda_{01}^*)^\top u_K(X)\right)\right)^2 u_K^\top(X)u_K(X)|T=0\right] + \frac{1}{Np_1}\mathbb{E}\left[\pi_{01}(X)u_K^\top(X)u_K(X)|T=0\right] \\
\leq&\frac{1}{Np_0 \wedge Np_1}\left(\eta_2 + \sup_{\gamma\in\Gamma_1}\left(\rho'(\gamma)^2\right)\right)\cdot \mathbb{E}[u_K^\top(X)u_K(X)|T=0] \\
=&\frac{1}{Np_0 \wedge Np_1}\left(\eta_2 + \sup_{\gamma\in\Gamma_1}\left(\rho'(\gamma)^2\right)\right)\cdot \frac{1}{p_0}\cdot \mathbb{E}[P(T=0|X)u_K^\top(X)u_K(X)] \\
\leq&C_4^2\cdot\frac{K}{N}\ ,
\end{aligned} \tag{A.19}$$

where $C_4^2 := \left(\eta_2 + \sup_{\gamma\in\Gamma_1}\left(\rho'(\gamma)^2\right)\right)\cdot(p_0\wedge p_1)^{-1}\cdot p_0^{-1} < +\infty$.

Let $\epsilon > 0$, and fix some $C_5(\epsilon) > 0$ (to be chosen later). Define

$$\hat{\Lambda}_K(\epsilon) := \left\{ \lambda \in \mathbb{R}^K : \|\lambda - \lambda_{01}^*\| \leq C_5(\epsilon) C_4 \sqrt{\frac{K}{N}} \right\} . \tag{A.20}$$

For $\forall \lambda \in \hat{\Lambda}_K(\epsilon),\ \forall x \in \mathcal{X}$, when N sufficiently large, by Assumption 2.4.4 and (A.8), we have

$$\begin{aligned}
&|\lambda^\top u_K(x) - (\lambda_{01}^*)^\top u_K(x)| \leq \|\lambda - \lambda_{01}^*\| \|u_K(x)\| \leq C_5(\epsilon) C_4 \sqrt{\frac{K}{N}} \zeta(K) \\
&\Rightarrow \lambda^\top u_K(x) \in \left[(\lambda_{01}^*)^\top u_K(x) - C_5(\epsilon) C_4 \zeta(K) \sqrt{\frac{K}{N}},\ (\lambda_{01}^*)^\top u_K(x) + C_5(\epsilon) C_4 \zeta(K) \sqrt{\frac{K}{N}} \right] \\
&\qquad \subset \left[\underline{\gamma} - C_1 K^{-\alpha} - C_2 K^{-\alpha} \zeta(K) - C_5(\epsilon) C_4 \zeta(K) \sqrt{\frac{K}{N}}, \right. \\
&\qquad\qquad \left. \overline{\gamma} + C_1 K^{-\alpha} + C_2 K^{-\alpha} \zeta(K) + C_5(\epsilon) C_4 \zeta(K) \sqrt{\frac{K}{N}} \right] \subset \Gamma_2(\epsilon) ,
\end{aligned} \tag{A.21}$$

where

$$\Gamma_2(\epsilon) := [\underline{\gamma} - 1 - C_5(\epsilon), \overline{\gamma} + 1 + C_5(\epsilon)] ,$$

is a compact set and independent of x.

By Taylor's Theorem, for any $\lambda \in \partial \hat{\Lambda}_K(\epsilon)$, there exists λ_ϵ lying on the line joining λ and λ_{01}^* such that

$$\bar{Q}_{01}(\lambda) = \bar{Q}_{01}(\lambda_{01}^*) + (\lambda - \lambda_{01}^*)^\top \bar{G}_K'(\lambda_{01}^*) + \frac{1}{2} (\lambda - \lambda_{01}^*)^\top \bar{Q}_{01}''(\lambda_\epsilon)(\lambda - \lambda_{01}^*).$$

For the second order term of above expression, when N is large enough, we have

$$\begin{aligned}
(\lambda - \lambda_{01}^*)^\top \bar{Q}_{01}''(\lambda_\epsilon)(\lambda - \lambda_{01}^*) &= \frac{1}{N p_0} \sum_{i=1}^N (1 - T_i) \rho''(\lambda_\epsilon^\top u_K(X_i))(\lambda - \lambda_{01}^*)^\top u_K(X_i) u_K(X_i)^\top (\lambda - \lambda_{01}^*) \\
&\leq -\bar{b}(\epsilon) \frac{1}{N p_0} \sum_{i=1}^N (1 - T_i)(\lambda - \lambda_{01}^*)^\top u_K(X_i) u_K(X_i)^\top (\lambda - \lambda_{01}^*) \\
&= -\frac{\bar{b}(\epsilon)}{p_0} \cdot (\lambda - \lambda_{01}^*)^\top \hat{S}_N (\lambda - \lambda_{01}^*) ,
\end{aligned} \tag{A.22}$$

where $-\bar{b}(\epsilon) := \sup_{\gamma \in \Gamma_2(\epsilon)} \rho''(\gamma) < \infty$ because $\Gamma_2(\epsilon)$ is compact and ρ'' is a continuous function.

Then on the event E_N with large enough N, we can have for any $\lambda \in \partial\hat{\Lambda}_K(\epsilon)$,

$$\begin{aligned}
&\bar{Q}_{01}(\lambda) - \bar{Q}_{01}(\lambda_{01}^*) \\
&= (\lambda - \lambda_{01}^*)^\top \bar{Q}'_{01}(\lambda_{01}^*) + \frac{1}{2}(\lambda - \lambda_{01}^*)^\top \bar{Q}''_{01}(\lambda_\epsilon)(\lambda - \lambda_{01}^*) \\
&\le \|\lambda - \lambda_{01}^*\| \|\bar{Q}'_{01}(\lambda_{01}^*)\| - \frac{\bar{b}(\epsilon)}{2p_0}(\lambda - \lambda_{01}^*)^\top \hat{S}_N(\lambda - \lambda_{01}^*) \\
&\le \|\lambda - \lambda_{01}^*\| \|\bar{Q}'_{01}(\lambda_{01}^*)\| - \frac{\bar{b}(\epsilon)}{2p_0}(\lambda - \lambda_{01}^*)^\top \left(\mathbb{E}\left[P(T=0|X)u_K(X)u_K^\top(X)\right] - \frac{\eta_1'}{2} I_K \right)(\lambda - \lambda_{01}^*) \\
&\le \|\lambda - \lambda_{01}^*\| \|\bar{Q}'_{01}(\lambda_{01}^*)\| - \frac{\bar{b}(\epsilon)}{2p_0}(\lambda - \lambda_{01}^*)^\top \left(\eta_1' I_K - \frac{\eta_1'}{2} I_K \right)(\lambda - \lambda_{01}^*) \\
&< \|\lambda - \lambda_{01}^*\| \left(\|\bar{Q}'_{01}(\lambda_{01}^*)\| - \frac{\eta_1' \bar{b}(\epsilon)}{4p_0} \|\lambda - \lambda_{01}^*\| \right), \qquad (A.23)
\end{aligned}$$

where the first inequality follows from (A.22) and $\eta_1' = \inf_{x \in \mathcal{X}} P(T = 0|X = x)$. By Chebyshev's inequality and Inequality (A.19), for sufficiently large N, we have

$$\mathbb{P}\left\{ \|\bar{Q}'_{01}(\lambda_{01}^*)\| \ge \frac{\eta_1' \bar{b}(\epsilon)}{4p_0} \|\lambda - \lambda_{01}^*\| \right\} \le \frac{16p_0^2}{\bar{b}(\epsilon)^2 C_5^2(\epsilon) {\eta_1'}^2} \le \frac{\epsilon}{2}, \qquad (A.24)$$

where the last inequality holds by choosing $C_5(\epsilon) \ge \frac{4p_0}{\eta_1' \bar{b}(\epsilon)} \sqrt{\frac{2}{\epsilon}}$. Therefore, for sufficiently large N, by (A.18) and (A.24) we can derive

$$\begin{aligned}
&\mathbb{P}\left((E_N)^c \text{ or } \|\bar{Q}'_{01}(\lambda_{01}^*)\| \ge \frac{\eta_1' \bar{b}(\epsilon)}{4p_0} \|\lambda - \lambda_{01}^*\| \right) \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \\
\Rightarrow &\mathbb{P}\left(E_N \text{ and } \|\bar{Q}'_{01}(\lambda_{01}^*)\| < \frac{\eta_1' \bar{b}(\epsilon)}{4p_0} \|\lambda - \lambda_{01}^*\| \right) > 1 - \epsilon . \qquad (A.25)
\end{aligned}$$

Then by (A.23) and (A.25), we get

$$\mathbb{P}\{\bar{Q}_{01}(\lambda) - \bar{Q}_{01}(\lambda_{01}^*) < 0 \ \ \forall \lambda \in \partial\hat{\Lambda}_K\} \ge 1 - \epsilon ,$$

for sufficiently large N. Note that the event $\{\bar{Q}_{01}(\lambda_{01}^*) > \bar{Q}_{01}(\lambda), \ \forall \lambda \in \partial\hat{\Lambda}_K(\epsilon)\}$ implies that there exists a local maximum point in the interior of $\hat{\Lambda}_K(\epsilon)$. On the other hand, $\bar{Q}_{01}$ is strictly concave function and $\bar{\lambda}_K$ is the unique global maximum point of $\bar{Q}_{01}$, then we can conclude that

$$\mathbb{P}\left(\bar{\lambda}_{01} \in \hat{\Lambda}_K(\epsilon)\right) > 1 - \epsilon \,, \tag{A.26}$$

i.e.

$$\|\bar{\lambda}_{01} - \lambda_{01}^*\| = O_P\left(\sqrt{\frac{K}{N}}\right).$$

Next, we compute the order of $\|\bar{\lambda}_{01} - \hat{\lambda}_{01}\|$. Note that $\bar{\lambda}_{01}$ and $\hat{\lambda}_{01}$ are the minimizer of $\bar{Q}_{01}(\lambda)$ and $\hat{Q}_{01}(\lambda)$ respectively, then the first order conditions yield:

$$\frac{1}{N}\sum_{i=1}^{N}(1-T_i)\rho'(\hat{\lambda}_{01}u_K(X_i)u_K(X_i) = \frac{1}{N}\frac{\sum_{j=1}^{N}(1-T_j)}{\sum_{i=1}^{N}T_i}\sum_{j=1}^{N}T_j u_K(X_j), \tag{A.27}$$

$$\frac{1}{N}\sum_{i=1}^{N}(1-T_i)\rho'(\bar{\lambda}_{01}u_K(X_i)u_K(X_i) = \frac{1}{N}\frac{p_0}{p_1}\sum_{j=1}^{N}T_j u_K(X_j). \tag{A.28}$$

By substracting Equation (A.27) from Equation (A.28) and using Mean Value Theorem, we can obtain

$$(\hat{\lambda}_{01} - \bar{\lambda}_{01})^\top \times \frac{1}{N}\sum_{i=1}^{N}(1-T_i)\rho''(\lambda_\zeta^\top u_K(X_i))u_K(X_i)u_K^\top(X_i) = \left(\frac{\sum_{j=1}^{N}(1-T_j)/N}{\sum_{i=1}^{N}T_i/N} - \frac{p_0}{p_1}\right)\frac{1}{N}\sum_{j=1}^{N}T_j u_K^\top(X_j),$$

where λ_ζ lies on the line joining $\bar{\lambda}_{01}$ and $\hat{\lambda}_{01}$. Then we have

$$
\begin{aligned}
&\|\hat{\lambda}_{01}-\bar{\lambda}_{01}\| \\
\leq &\left\|\left(\frac{\sum_{j=1}^N(1-T_j)/N}{\sum_{i=1}^N T_i/N}-\frac{p_0}{p_1}\right)\right\|\left\|\left[\frac{1}{N}\sum_{i=1}^N(1-T_i)\rho''(\lambda_\zeta^\top u_K(X_i))u_K(X_i)u_K^\top(X_i)\right]^{-1}\frac{1}{N}\sum_{j=1}^N T_j u_K(X_j)\right\| \\
=&O_P\left(\frac{1}{\sqrt{N}}\right)\times\sqrt{\left[\frac{1}{N}\sum_{j=1}^N T_j u_K(X_j)\right]^\top\left\{\frac{1}{N}\sum_{i=1}^N(1-T_i)\rho''(\lambda_\zeta^\top u_K(X_i))u_K(X_i)u_K^\top(X_i)\right\}^{-2}\frac{1}{N}\sum_{j=1}^N T_j u_K(X_j)} \\
\leq &O_P\left(\frac{1}{\sqrt{N}}\right)\times\frac{1}{\inf_{x\in\mathcal{X}}|\rho''(\lambda_\zeta^\top u_K(x))|} \\
&\quad\times\sqrt{\left[\frac{1}{N}\sum_{j=1}^N T_j u_K(X_j)\right]^\top\left\{\frac{1}{N}\sum_{i=1}^N(1-T_i)u_K(X_i)u_K^\top(X_i)\right\}^{-2}\frac{1}{N}\sum_{j=1}^N T_j u_K(X_j)} \\
\leq &O_P\left(\frac{1}{\sqrt{N}}\right)\times O_p(1)\times\lambda_{\min}^{-1}\left(\frac{1}{N}\sum_{i=1}^N(1-T_i)u_K(X_i)u_K^\top(X_i)\right)\times\sqrt{\left[\frac{1}{N}\sum_{j=1}^N T_j u_K(X_j)\right]^\top\left[\frac{1}{N}\sum_{j=1}^N T_j u_K(X_j)\right]} \\
=&O_P\left(\frac{1}{\sqrt{N}}\right)\times O_P(\sqrt{K})=O_P\left(\sqrt{\frac{K}{N}}\right)
\end{aligned}
$$

Therefore,

$$
\|\hat{\lambda}_{01}-\lambda_{01}^*\|\leq\|\hat{\lambda}_{01}-\bar{\lambda}_{01}\|+\|\lambda_{01}^*-\bar{\lambda}_{01}\|=O_P\left(\sqrt{\frac{K}{N}}\right)+O_P\left(\sqrt{\frac{K}{N}}\right)=O_P\left(\sqrt{\frac{K}{N}}\right). \tag{A.29}
$$

We next show that $\int_{\mathcal{X}}|\hat{\pi}_{01}(x)-\pi_{01}^*(x)|^2dF_X(x)=O_P\left(K/N\right)$. By Mean Value Theorem, we can have

$$
\hat{\pi}_{01}(x)-\pi_{01}^*(x)=\rho'\left(\hat{\lambda}_{01}^\top u_K(x)\right)-\rho'\left((\lambda_{01}^*)^\top u_K(x)\right)=\rho''\left(\tilde{\lambda}_{01}^\top u_K(x)\right)(\hat{\lambda}_{01}-\lambda_{01}^*)^\top u_K(x)\,,
$$

where $\tilde{\lambda}_{01}$ lies on the line joining $\hat{\lambda}_{01}$ and λ_{01}^*. From (A.21) and (A.29), we can have

$$
\sup_{x\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_{01}^\top u_K(x)\right)\right|=O_P(1). \tag{A.30}
$$

By Mean Value Theorem, (A.30), and the fact $\mathbb{E}[\|u_K(X)\|^2] = K$, we can have

$$\begin{aligned}
&\int_{\mathcal{X}} |\hat{\pi}_{01}(x) - \pi_{01}^*(x)|^2 dF_X(x) \\
\leq &\sup_{x\in\mathcal{X}} \left|\rho''\left(\tilde{\lambda}_{01}^\top u_K(x)\right)\right|^2 \cdot \|\hat{\lambda}_{01} - \lambda_{01}^*\|^2 \int_{\mathcal{X}} u_K(x)u_K(x)^\top dF_X(x) \\
= &O_P(1) \cdot O_P\left(\frac{K}{N}\right) = O_P\left(\frac{K}{N}\right) \ . \qquad \text{(A.31)}
\end{aligned}$$

Similarly, we can obtain that

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N |\hat{\pi}_{01}(X_i) - \pi_{01}^*(X_i)|^2 \\
\leq &\sup_{x\in\mathcal{X}} \left|\rho''\left(\tilde{\lambda}_{01}^\top u_K(x)\right)\right|^2 \cdot (\hat{\lambda}_{01} - \lambda_{01}^*)^\top \cdot \left\{\frac{1}{N}\sum_{i=1}^N u_K(X_i)u_K(X_i)^\top\right\}(\hat{\lambda}_{01} - \lambda_{01}^*) \\
= &\sup_{x\in\mathcal{X}} \left|\rho''\left(\tilde{\lambda}_{01}^\top u_K(x)\right)\right|^2 \cdot \|\hat{\lambda}_{01} - \lambda_{01}^*\|^2 \cdot \lambda\text{max}\left\{\sum_{i=1}^N u_K(X_i)u_K(X_i)^\top\right\} \\
\leq &O_P(1) \cdot O_P\left(\frac{K}{N}\right) \cdot O_P(1) = O_P\left(\frac{K}{N}\right) \ . \qquad \text{(A.32)}
\end{aligned}$$

□

A.3 Proof of Theorem 2.4.8

We provide detailed argument of proving $\sqrt{N}\{\hat{\theta}_{011} - \theta_{011}\} = N^{-1/2}\sum_{i=1}^N S_{\theta_{011}}(T_i, X_i, Y_i) + o_P(1)$. A similar argument can be applied to establish other results.

Note that

$$
\begin{aligned}
\sqrt{N}\left(\hat{\theta}_{011}-\theta_{011}\right) =&\sqrt{N}\left(\int \hat{F}_{01}(y)d\hat{F}_{11}(y)-\int F_{01}(y)dF_{11}(y)\right)\\
=&\sqrt{N}\int\{\hat{F}_{01}(y)-F_{01}(y)\}d\hat{F}_{11}(y)+\sqrt{N}\int F_{01}(y)d\{\hat{F}_{11}(y)-F_{11}(y)\}\\
=&\sqrt{N}\int\{\hat{F}_{01}(y)-F_{01}(y)\}d\hat{F}_{11}(y)-\sqrt{N}\int\{\hat{F}_{11}(y)-F_{11}(y)\}dF_{01}(y)
\end{aligned}
$$

By Theorem 2.4.9, $\hat{F}_{01}(\cdot)$ and $\hat{F}_{11}(\cdot)$ have the following representation: for $\forall y \in \mathbb{R}$:

$$
\begin{aligned}
&\sqrt{N}\{\hat{F}_{01}(y)-F_{01}(y)\}\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\pi_{01}(X_i)\mathbb{1}(Y_i\leq y)-\left(\frac{1}{p_0}(1-T_i)\pi_{01}(X_i)-\frac{T_i}{p_1}\right)F_{Y(0)|X}(y|X_i)-\frac{T_i}{p_1}F_{01}(y)\right\}\\
&+o_P(1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\sqrt{N}\{\hat{F}_{11}(y)-F_{11}(y)\}\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}F_{11}(y)\left(1-\frac{T_i}{p_1}\right)+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{T_i}{p_1}\mathbb{1}(Y_i\leq y)-F_{11}(y)\right\}\\
&+o_P(1).
\end{aligned}
$$

Then we can have

$$
\sqrt{N}\int\{\hat{F}_{11}(y)-F_{11}(y)\}dF_{01}(y)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_i}{p_1}(1-F_{01}(Y_i))-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_i}{p_1}(1-\theta_{011})+o_P(1).
$$

and

$$
\begin{aligned}
&\sqrt{N}\int\{\hat{F}_{01}(y)-F_{01}(y)\}d\hat{F}_{11}(y)\\
=&\frac{1}{\sqrt{N}}\frac{1}{\sum_{i=1}^{N}T_i}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\frac{(1-T_i)T_j}{p_0}\pi_{01}(X_i)\mathbb{1}(Y_i\le Y_j)-\left(\frac{1}{p_0}(1-T_i)T_j\pi_{01}(X_i)-\frac{T_iT_j}{p_1}\right)F_{Y(0)|X}(Y_j|X_i)\right.\\
&\qquad\qquad\left.-\frac{T_iT_j}{p_1}F_{01}(Y_j)\right\}+o_P(1)\\
=&\left(\frac{1}{\sum_{i=1}^{N}T_i/N}-\frac{1}{p_1}\right)\frac{\sqrt{N}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\frac{(1-T_i)T_j}{p_0}\pi_{01}(X_i)\mathbb{1}(Y_i\le Y_j)\right.\\
&\qquad\left.-\left(\frac{1}{p_0}(1-T_i)T_j\pi_{01}(X_i)-\frac{T_iT_j}{p_1}\right)F_{Y(0)|X}(Y_j|X_i)-\frac{T_iT_j}{p_1}F_{01}(Y_j)\right\}+o_P(1)\\
&+\frac{1}{p_1}\frac{\sqrt{N}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\frac{(1-T_i)T_j}{p_0}\pi_{01}(X_i)\mathbb{1}(Y_i\le Y_j)-\left(\frac{1}{p_0}(1-T_i)T_j\pi_{01}(X_i)-\frac{T_iT_j}{p_1}\right)F_{Y(0)|X}(Y_j|X_i)\right.\\
&\qquad\qquad\left.-\frac{T_iT_j}{p_1}F_{01}(Y_j)\right\}+o_P(1)\\
=&\frac{\sqrt{N}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\frac{(1-T_i)T_j}{p_1p_0}\pi_{01}(X_i)\mathbb{1}(Y_i\le Y_j)-\frac{(1-T_i)T_j}{p_1p_0}\pi_{01}(X_i)F_{Y(0)|X}(Y_j|X_i)\right.\\
&\qquad\left.+\frac{T_iT_j}{p_1^2}F_{Y(0)|X}(Y_j|X_i)-\frac{T_iT_j}{p_1^2}F_{01}(Y_j)\right\}+o_P(1)\\
=&W_{1N}-W_{2N}+W_{3N}-W_{4N}+o_P(1), \qquad\qquad (A.33)
\end{aligned}
$$

where

$$
\begin{aligned}
W_{1N}&=\frac{\sqrt{N}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(1-T_i)T_j}{p_1p_0}\pi_{01}(X_i)\mathbb{1}(Y_i\le Y_j)\ ,\\
W_{2N}&=\frac{\sqrt{N}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(1-T_i)T_j}{p_1p_0}\pi_{01}(X_i)F_{Y(0)|X}(Y_j|X_i)\ ,\\
W_{3N}&=\frac{\sqrt{N}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{T_iT_j}{p_1^2}F_{Y(0)|X}(Y_j|X_i)\ ,\\
W_{4N}&=\frac{\sqrt{N}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{T_iT_j}{p_1^2}F_{01}(Y_j)\ .
\end{aligned}
$$

Now we apply the theory of U-statistics to represent (A.33) as a sum of $i.i.d.$ random variables. In the following, we use (X,Y,T) denoting an independent copy of $\{X_i,Y_i,T_i\}_{i=1}^N$.

For W_{1N}, define the projection statistics

$$\begin{aligned}
\tilde{W}_{1N} =& \mathbb{E}[W_{1N}] + \sum_{i=1}^{N}(1-T_i)\{\mathbb{E}[W_{1N}|X_i,Y_i,T_i=0] - \mathbb{E}[W_{1N}]\} + \sum_{j=1}^{N} T_j\{\mathbb{E}[W_{1N}|X_j,Y_j,T_j=1] - \mathbb{E}[W_{1N}]\} \\
=& \frac{\sqrt{N}N(N-1)}{N^2}\theta_{011} + \sum_{i=1}^{N}\sqrt{N}(1-T_i)\left\{\frac{N-1}{N^2 p_0}\pi_{01}(X_i)\mathbb{E}\left[\mathbb{1}(Y_i \leq Y)|Y_i, T=1\right] - \frac{2(N-1)}{N^2}\theta_{011}\right\} \\
&+ \sum_{j=1}^{N}\sqrt{N}T_j\left\{\frac{N-1}{N^2 p_1}\mathbb{E}\left[\pi_{01}(X)\mathbb{1}(Y \leq Y_j)|T=0, Y_j\right] - \frac{2(N-1)}{N^2}\theta_{011}\right\} \\
=& \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\pi_{01}(X_i)\left(1 - F_{Y(1)|T=1}(Y_i)\right) + \frac{T_i}{p_1}\mathbb{E}\left[\pi_{01}(X)\mathbb{1}(Y \leq Y_i)|T=0, Y_i\right] - \theta_{011}\right\} \\
&+ O_P\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

For W_{2N}, define the projection statistics

$$\begin{aligned}
\tilde{W}_{2N} =& \mathbb{E}[W_{2N}] + \sum_{i=1}^{N}(1-T_i)\{\mathbb{E}[W_{2N}|X_i,Y_i,T_i=0] - \mathbb{E}[W_{2N}]\} + \sum_{j=1}^{N} T_j\{\mathbb{E}[W_{2N}|X_j,Y_j,T_j=1] - \mathbb{E}[W_{2N}]\} \\
=& \frac{\sqrt{N}N(N-1)}{N^2}\theta_{011} + \sum_{i=1}^{N}\sqrt{N}(1-T_i)\left\{\frac{N-1}{N^2 p_0}\mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right] - \frac{2(N-1)}{N^2}\theta_{011}\right\} \\
&+ \sum_{j=1}^{N}\sqrt{N}T_j\left\{\frac{N-1}{N^2 p_1}\mathbb{E}\left[\pi_{01}(X)F_{Y(0)|X}(Y_j|X)|T=0, Y_j\right] - \frac{2(N-1)}{N^2}\theta_{011}\right\} \\
=& \frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\pi_{01}(X_i)\mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right] + \frac{T_i}{p_1}\mathbb{E}\left[\pi_{01}(X)F_{Y(0)|X}(Y_i|X)|T=0, Y_i\right] - \theta_{011}\right\} \\
&+ O_P\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

For W_{3N}, define the projection statistics

$$\begin{aligned}
\tilde{W}_{3N} =&\mathbb{E}[W_{3N}] + \sum_{i=1}^{N}(1-T_i)\{\mathbb{E}[W_{3N}|X_i,Y_i,T_i=0]-\mathbb{E}[W_{3N}]\} + \sum_{j=1}^{N}T_j\{\mathbb{E}[W_{3N}|X_j,Y_j,T_j=1]-\mathbb{E}[W_{3N}]\}\\
=&\frac{\sqrt{N}N(N-1)}{N^2}\theta_{011} + \frac{\sqrt{N}N}{N^2p_1}\mathbb{E}\left[F_{Y(0)|X}(Y|X)|T=1\right]\\
&+\sum_{i=1}^{N}\sqrt{N}(1-T_i)\left\{-\frac{2(N-2)}{N^2}\theta_{011} - \frac{1}{N^2}\cdot\frac{1}{p_1}\mathbb{E}\left[F_{Y(0)|X}(Y|X)|T=1\right]\right\}\\
&+\sum_{j=1}^{N}\sqrt{N}T_j\left\{\frac{1}{N^2p_1^2}F_{Y(0)|X}(Y_j|X_j) + \frac{N-1}{N^2p_1}\mathbb{E}\left[F_{Y(0)|X}(Y|X_j)|X_j,T=1\right]\right.\\
&\left.+\frac{N-1}{N^2p_1}\mathbb{E}\left[F_{Y(0)|X}(Y_j|X)|Y_j,T=1\right] - \frac{2(N-2)}{N^2}\theta_{011} - \frac{1}{N^2p_1}\mathbb{E}\left[F_{Y(0)|X}(Y|X)|T=1\right]\right\}\\
=&\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{T_i}{p_1}\mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i,T=1\right] + \frac{T_i}{p_1}\mathbb{E}\left[F_{Y(0)|X}(Y_i|X)|Y_i,T=1\right] - \theta_{011}\right\} + O_P\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

For W_{4N}, define the projection statistics

$$\begin{aligned}
\tilde{W}_{4N} =&\mathbb{E}[W_{4N}] + \sum_{i=1}^{N}(1-T_i)\{\mathbb{E}[W_{4N}|X_i,Y_i,T_i=0]-\mathbb{E}[W_{4N}]\} + \sum_{j=1}^{N}T_j\{\mathbb{E}[W_{4N}|X_j,Y_j,T_j=1]-\mathbb{E}[W_{4N}]\}\\
=&\frac{\sqrt{N}N(N-1)}{N^2}\theta_{011} + \frac{\sqrt{N}N}{N^2p_1}\mathbb{E}\left[F_{01}(Y)|T=1\right]\\
&+\sum_{i=1}^{N}\sqrt{N}(1-T_i)\left\{-\frac{2(N-2)}{N^2}\theta_{011} - \frac{1}{N^2p_1}\mathbb{E}\left[F_{01}(Y)|T=1\right]\right\}\\
&+\sum_{j=1}^{N}\sqrt{N}T_j\left\{\frac{1}{N^2p_1^2}F_{01}(Y_j) + \frac{N-1}{N^2p_1}\mathbb{E}\left[F_{01}(Y)|T=1\right]\right.\\
&\left.+\frac{N-1}{N^2p_1}F_{01}(Y_j) - \frac{2(N-2)}{N^2}\theta_{011} - \frac{1}{N^2p_1}\mathbb{E}\left[F_{01}(Y)|T=1\right]\right\}\\
=&\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{T_i}{p_1}\mathbb{E}\left[F_{01}(Y)|T=1\right] + \frac{T_i}{p_1}F_{01}(Y_i) - \theta_{011}\right\} + O_P\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

By the theory of U-statistics [37, 38, 39],

$$
\begin{aligned}
W_{1N} - \mathbb{E}[W_{1N}] &= \left\{\tilde{W}_{1N} - \mathbb{E}[\tilde{W}_{1N}]\right\}(1 + o_P(1)),\\
W_{2N} - \mathbb{E}[W_{2N}] &= \left\{\tilde{W}_{2N} - \mathbb{E}[\tilde{W}_{2N}]\right\}(1 + o_P(1)),\\
W_{3N} - \mathbb{E}[W_{3N}] &= \left\{\tilde{W}_{3N} - \mathbb{E}[\tilde{W}_{3N}]\right\}(1 + o_P(1)),\\
W_{4N} - \mathbb{E}[W_{4N}] &= \left\{\tilde{W}_{4N} - \mathbb{E}[\tilde{W}_{4N}]\right\}(1 + o_P(1)).
\end{aligned}
$$

Then we can obtain that

$$
\begin{aligned}
&\sqrt{N}\left(\hat{\theta}_{011} - \theta_{011}\right)\\
=& -\sqrt{N}\int\{\hat{F}_{11}(y) - F_{11}(y)\}dF_{01}(y) + \left\{\tilde{W}_{1N} - \mathbb{E}[\tilde{W}_{1N}] + \mathbb{E}[W_{1N}]\right\}\\
&-\left\{\tilde{W}_{1N} - \mathbb{E}[\tilde{W}_{1N}] + \mathbb{E}[W_{2N}]\right\} + \left\{\tilde{W}_{3N} - \mathbb{E}[\tilde{W}_{3N}] + \mathbb{E}[W_{3N}]\right\} - \left\{\tilde{W}_{4N} - \mathbb{E}[\tilde{W}_{4N}] + \mathbb{E}[W_{4N}]\right\} + o_P(1)\\
=& -\sqrt{N}\int\{\hat{F}_{11}(y) - F_{11}(y)\}dF_{01}(y) + \tilde{W}_{1N} - \tilde{W}_{2N} + \tilde{W}_{3N} - \tilde{W}_{4N} + o_P(1)\\
=&\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\pi_{01}(X_i)\left(1 - F_{Y(1)|T=1}(Y_i)\right) + \frac{T_i}{p_1}\mathbb{E}\left[\pi_{01}(X)\mathbb{1}(Y \le Y_i)|T=0, Y_i\right] - \theta_{011}\right\}\\
&-\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\pi_{01}(X_i)\mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right] + \frac{T_i}{p_1}\mathbb{E}\left[\pi_{01}(X)F_{Y(0)|X}(Y_i|X)|T=0, Y_i\right] - \theta_{011}\right\}\\
&+\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{T_i}{p_1}\mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right] + \frac{T_i}{p_1}\mathbb{E}\left[F_{Y(0)|X}(Y_i|X)|Y_i, T=1\right] - \theta_{011}\right\}\\
&-\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left\{\frac{T_i}{p_1}\mathbb{E}\left[F_{01}(Y)|T=1\right] + \frac{T_i}{p_1}F_{01}(Y_i) - \theta_{011}\right\}\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_i}{p_1}(1 - F_{01}(Y_i)) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_i}{p_1}(1 - \theta_{011}) + o_P(1)\\
&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1-T_i}{p_0}\left\{(1 - F_{Y(1)|T=1}(Y_i))\pi_{01}(X_i) - \pi_{01}(X_i)\mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right]\right\}\right.\\
&\left.+\frac{T_i}{p_1}\left\{\mathbb{E}\left[F_{Y(0)|X}(Y_i|X)|Y_i, T=1\right] - \mathbb{E}\left[F_{01}(Y)|T=1\right] + \mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right] - \theta_{011}\right\}\right] + o_P(1)\\
&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1-T_i}{p_0}\left\{(1 - F_{11}(Y_i))\pi_{01}(X_i) - \pi_{01}(X_i)\mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right]\right\}\right.\\
&\left.+\frac{T_i}{p_1}\left\{F_{01}(Y_i) - \theta_{011} + \mathbb{E}\left[F_{Y(0)|X}(Y|X_i)|X_i, T=1\right] - \theta_{011}\right\}\right] + o_P(1)\\
&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}S_{\theta_{011}}(T_i, X_i, Y_i) + o_P(1)\,.
\end{aligned}
$$

A.4 Proof of Theorem 2.4.9

We provide detailed argument of proving $\sqrt{N}\{\hat{F}_{01}(y) - F_{01}(y)\} = N^{-1/2}\sum_{i=1}^{N} S_{F_{01}}(T_i, X_i, Y_i; y) + o_P(1)$. A similar argument can be applied to establish other results.

Note that

$$
\begin{aligned}
&\sqrt{N}\{\hat{F}_{01}(y) - F_{01}(y)\} \\
=&\sqrt{N}\left\{\frac{\sum_{i=1}^{N}(1-T_i)\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i \le y)}{\sum_{i=1}^{N}(1-T_i)} - F_{01}(y)\right\} \\
=&\left\{\frac{1}{N^{-1}\sum_{i=1}^{N}(1-T_i)} - \frac{1}{p_0}\right\}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{(1-T_i)\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i \le y) - \mathbb{E}[(1-T)\pi_{01}(X)\mathbb{1}(Y \le y)]\right\} \\
&-\mathbb{E}[(1-T)\pi_{01}(X)\mathbb{1}(Y \le y)]\cdot\frac{1}{N^{-1}\sum_{i=1}^{N}(1-T_i)}\cdot\frac{1}{p_0}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{(1-T_i)-p_0\} \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\cdot\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i \le y) - F_{01}(y)\right\}.
\end{aligned}
$$

We shall show that for all $y \in \mathbb{R}$,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\cdot\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i \le y) - F_{01}(y)\right\} \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\pi_{01}(X_i)I(Y_i \le y) - \frac{1}{p_0}(1-T_i)\pi_{01}(X_i)F_{Y(0)|X}(y|X_i) + \frac{T_i}{p_1}F_{Y(0)|X}(y|X_i)\right. \\
&\left.\qquad + \frac{1-T_i}{p_0}F_{01}(y) - \frac{T_i}{p_1}F_{01}(y) - F_{01}(y)\right\} + o_P(1). \qquad \text{(A.34)}
\end{aligned}
$$

By using Law of Large Number, (A.34) and Central Limit Theorem, we can claim that

$$
\begin{aligned}
&\left\{\frac{1}{N^{-1}\sum_{i=1}^{N}(1-T_i)} - \frac{1}{p_0}\right\}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{(1-T_i)\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i \le y) - \mathbb{E}[(1-T)\pi_{01}(X)\mathbb{1}(Y \le y)]\right\} \\
=&o_P(1)\cdot O_P(1) = o_P(1).
\end{aligned}
$$

Therefore, we can obtain:

$$
\begin{aligned}
&\sqrt{N}\{\hat{F}_{01}(y) - F_{01}(y)\} \\
=&o_P(1) - \frac{F_{01}(y)}{p_0} \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{(1-T_i) - p_0\} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \frac{1-T_i}{p_0} \pi_{01}(X_i) \mathbb{1}(Y_i \le y) \right. \\
&\left. - \frac{1}{p_0}(1-T_i)\pi_{01}(X_i) F_{Y(0)|X}(y|X_i) + \frac{T_i}{p_1} F_{Y(0)|X}(y|X_i) + \frac{1-T_i}{p_0} F_{01}(y) - \frac{T_i}{p_1} F_{01}(y) - F_{01}(y) \right\} \\
=&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \frac{1-T_i}{p_0} \pi_{01}(X_i) \mathbb{1}(Y_i \le y) - \left\{ \frac{1-T_i}{p_0} \pi_{01}(X_i) - \frac{T_i}{p_1} \right\} F_{Y(0)|X}(y|X_i) - \frac{T}{p_1} F_{01}(y) \right\} + o_P(1) \\
=&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} S_{F_{01}}(T_i, X_i, Y_i; y) + o_P(1).
\end{aligned}
$$

By Assumption 2.4.3, it is easy to verify that $\{S_{F_{01}}(T_i, X_i, Y_i; y) : y \in \mathbb{R}\}$ satisfies the L^2-Lipschitz condition (Condition (5.3) of [40]), i.e., there exists some constant $C > 0$ such that

$$
\mathbb{E}\left[\sup_{|y_1 - y_2| \le \delta} |S_{F_{01}}(T, X, Y; y_1) - S_{F_{01}}(T, X, Y; y_2)|^2 \right] \le C \cdot \delta^2.
$$

Then by Theorem 4 and 5 of [40], the function class $\{S_{F_{01}}(T_i, X_i, Y_i; y) : y \in \mathbb{R}\}$ is *stochastically equicontinuous*. Therefore, $\sqrt{N}\{\hat{F}_{01}(\cdot) - F_{01}(\cdot)\}$ weakly converges to a Gaussian process with covariance function $\{\mathbb{E}[S_{F_{01}}(T_i, X_i, Y_i; y_1) S_{F_{01}}(T_i, X_i, Y_i; y_2)] : y_1, y_2 \in \mathbb{R}\}$, which is our desired result.

Now it remains to prove (A.34). Before starting the formal proof, we introduce some notation that will be used later:

$$
\begin{aligned}
\Omega_K &:= (Q_{01}^*)''(\lambda_{01}^*) = \mathbb{E}\left[\rho''((\lambda_{01}^*)^\top u_K(X)) u_K(X) u_K(X)^\top | T = 0 \right], \\
\Phi_K(y) &:= -\mathbb{E}\left[F_{Y(0)|X}(y|X) \rho''((\lambda_{01}^*)^\top u_K(X)) | T = 0 \right].
\end{aligned}
$$

Note that $\Psi_K(y)^\top \Sigma_K^{-1} u_K(x)$ can be regarded as a weighted L^2-projection (with respect to the weighting function $-\rho''((\lambda_1^*)^\top u_K(x)) dF_{X|T=0}(x)$) of $F_{Y(0)|X}(y|x)$ on the space spanned by $\{u_K(x)\}$.

We define the empirical version of Ω_K and $\Phi_K(y)$ by

$$\tilde{\Omega}_K := \widehat{Q}''_{01}(\tilde{\lambda}_{01}) = \frac{\sum_{i=1}^N (1-T_i)\rho''\left(\tilde{\lambda}_{01}^\top u_K(X_i)\right) u_K(X_i) u_K^\top(X_i)}{\sum_{i=1}^N (1-T_i)},$$
$$\tilde{\Phi}_K(y) := -\int_{\mathcal{X}} F_{Y(0)|X}(y|x)\rho''\left(\tilde{\lambda}_{01}^\top u_K(x)\right) u_K(x) dF_{X|T=0}(x),$$

where $\tilde{\lambda}_{01}$ lies between $\hat{\lambda}_{01}$ and λ_{01}^* satisfying the Mean Value Theorem:

$$0 = \widehat{Q}'_{01}(\hat{\lambda}_{01}) = \widehat{Q}'_{01}(\lambda_{01}^*) + \widehat{Q}''_{01}(\tilde{\lambda}_{01})(\hat{\lambda}_{01} - \lambda_{01}^*),$$

which is also equivalent to

$$\begin{aligned} &\hat{\lambda}_{01} - \lambda_{01}^* = -\widehat{Q}''_{01}(\tilde{\lambda}_{01})^{-1}\widehat{Q}'_{01}(\lambda_{01}^*) = -\tilde{\Omega}_K^{-1}\widehat{Q}'_{01}(\lambda_{01}^*) \\ =& -\tilde{\Omega}_K^{-1} \cdot \left\{ \frac{\sum_{i=1}^N (1-T_i)\rho'((\lambda_{01}^*)^\top u_K(X_i)) u_K(X_i)}{\sum_{j=1}^N (1-T_j)} - \frac{\sum_{i=1}^N T_i u_K(X_i)}{\sum_{j=1}^N T_j} \right\}. \end{aligned} \tag{A.35}$$

We can decompose $N^{-1/2}\sum_{i=1}^{N}\{(1-T_i)\cdot\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i\le y)/p_0 - F_{01}(y)\}$ as follows:

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\cdot\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i\le y)-F_{01}(y)\right\}\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1-T_i}{p_0}\cdot\{\hat{\pi}_{01}(X_i)-\pi^*_{01}(X_i)\}\mathbb{1}(Y_i\le y)-\int_{\mathcal{X}}\{\hat{\pi}_{01}(x)-\pi^*_{01}(x)\}F_{Y(0)|X}(y|x)dF_{X|T=0}(x)\right] \quad \text{(A.36)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{1-T_i}{p_0}\cdot\{\pi^*_{01}(X_i)-\pi_{01}(X_i)\}\mathbb{1}(Y_i\le y)-\frac{1}{p_0}\mathbb{E}[(1-T)\{\pi^*_{01}(X)-\pi_{01}(X)\}\mathbb{1}(Y\le y)]\right] \quad \text{(A.37)}\\
&+\sqrt{N}\cdot\frac{1}{p_0}\cdot\mathbb{E}[(1-T)\{\pi^*_{01}(X)-\pi_{01}(X)\}\mathbb{1}(Y\le y)] \quad \text{(A.38)}\\
&+\sqrt{N}\left\{\int_{\mathcal{X}}\{\hat{\pi}_{01}(x)-\pi^*_{01}(x)\}F_{Y(0)|X}(y|x)dF_{X|T=0}(x)-\tilde{\Phi}_K(y)\tilde{\Omega}_K^{-1}\hat{Q}'_{01}(\lambda^*_{01})\right\} \quad \text{(A.39)}\\
&+\sqrt{N}\cdot\left[\tilde{\Phi}_K^{\top}(y)\tilde{\Omega}_K^{-1}-\Phi_K^{\top}(y)\Omega_K^{-1}\right]\hat{Q}'_{01}(\lambda^*_{01}) \quad \text{(A.40)}\\
&+\sqrt{N}\Bigg\{\Phi_K^{\top}(y)\Omega_K^{-1}\hat{Q}'_{01}(\lambda^*_{01})+\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1-T_i}{p_0}\pi_{01}(X_i)-\frac{T_i}{p_1}\right)F_{Y(0)|X}(y|X_i)\\
&\qquad -\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1-T_i}{p_0}F_{01}(y)-\frac{T_i}{p_1}F_{01}(y)\right)\Bigg\} \quad \text{(A.41)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg\{\frac{1-T_i}{p_0}\pi_{01}(X_i)\mathbb{1}(Y_i\le y)-\frac{1}{p_0}(1-T_i)\pi_{01}(X_i)F_{Y(0)|X}(y|X_i)+\frac{T_i}{p_1}F_{Y(0)|X}(y|X_i)\\
&\qquad +\frac{1-T_i}{p_0}F_{01}(y)-\frac{T_i}{p_1}F_{01}(y)-F_{01}(y)\Bigg\} \quad \text{(A.42)}
\end{aligned}
$$

We shall show the terms (A.36)-(A.41) are of $o_p(1)$ uniformly in $y\in\mathbb{R}$.

For the term (A.36)**:** Note that

$$
\begin{aligned}
\text{(A.36)}=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{1-T_i}{p_0}\left(\hat{\pi}_{01}(X_i)-\pi^*_{01}(X_i)\right)\mathbb{1}(Y_i\le y)-\frac{P(T_i=0|X_i)}{p_0}\left(\hat{\pi}_{01}(X_i)-\pi^*_{01}(X_i)\right)F_{Y(0)|X}(y|X_i)\right\} \quad \text{(A.43)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg\{\frac{P(T_i=0|X_i)}{p_0}\left(\hat{\pi}_{01}(X_i)-\pi^*_{01}(X_i)\right)F_{Y(0)|X}(y|X_i)\\
&-\int_{\mathcal{X}}\frac{P(T_i=0|X_i=x)}{p_0}F_{Y(0)|X}(y|x)(\hat{\pi}_{01}(x)-\pi^*_{01}(x))dF_X(x)\Bigg\}. \quad \text{(A.44)}
\end{aligned}
$$

Consider the term (A.43). Given the σ-algebra $\sigma(X_i, i\ge 1)$, $\hat{\pi}_{01}(x)$ is a deterministic function of x, and the summands $\{(\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i))(1-T_i)\mathbb{1}(Y_i \le y) - (\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i)) P(T_i = 0|X_i) F_{Y(0)|X}(y|X_i)\}_{i=1}^N$ are conditionally $i.i.d.$ with conditional mean zero. For any fixed $y \in \mathbb{R}$, by computing the conditional second moment of (A.43) and using Lemma A.2.2, we can obtain

$$
\begin{aligned}
&\mathbb{E}\left[|(\text{A.43})|^2|\sigma(X_t, t \ge 1)\right] \\
=&\frac{1}{p_0^2} \cdot \frac{1}{N}\sum_{i=1}^N \mathbb{E}\bigg[|(\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i))(1-T_i)\mathbb{1}(Y_i \le y) \\
&\qquad\qquad - (\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i)) P(T_i = 0|X_i) F_{Y(0)|X}(y|X_i)|^2|\sigma(X_t, t \ge 1)\bigg] \\
\le&\frac{1}{p_0^2} \cdot \frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[|(\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i))(1-T_i)\mathbb{1}(Y_i \le y)|^2|\sigma(X_t, t \ge 1)\right] \\
\le&\frac{1}{p_0^2} \cdot \frac{1}{N}\sum_{i=1}^N (\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i))^2 = O_p(K/N) = o_p(1),
\end{aligned}
$$

then by Chebyshev's inequality the term (A.43) is of $o_p(1)$ for fixed $y \in \mathbb{R}$. We next show that (A.43) is of $o_p(1)$ uniformly in $y \in \mathbb{R}$. Denote

$$
\begin{aligned}
&f(X_i, T_i, Y_i; y) \\
:=&(\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i))(1-T_i)\mathbb{1}(Y_i \le y) - (\hat{\pi}_{01}(X_i) - \pi^*_{01}(X_i)) P(T_i = 0|X_i) F_{Y(0)|X}(y|X_i).
\end{aligned}
$$

Given the σ-algebra $\sigma(X_t, t \ge 1)$, for large enough N and fixed $\delta > 0$, the following L^2-continuity (conditional version) holds:

$$
\begin{aligned}
&\left\{\mathbb{E}\left[\sup_{|y_1-y_2|\le\delta}|f(X_i,T_i,Y_i;y_1)-f(X_i,T_i,Y_i;y_2)|^2\middle|\sigma(X_t,t\ge 1)\right]\right\}^{1/2}\\
=&\Bigg\{(\hat{\pi}_{01}(X_i)-\pi^*_{01}(X_i))^2\\
\times&\mathbb{E}\Bigg[\Bigg\{(1-T_i)(\mathbb{1}(Y_i\le y_1)-\mathbb{1}(Y_i\le y_2))\\
&-P(T_i=0|X_i)(F_{Y(0)|X}(y_1|X_i)-F_{Y(0)|X}(y_2|X_i))\Bigg\}^2\Bigg|\sigma(X_t,t\ge 1)\Bigg]\Bigg\}^{1/2}\\
=&|\hat{\pi}_{01}(X_i)-\pi^*_{01}(X_i)|\\
\times&\Bigg\{\Bigg(P(T_i=0|X_i)\cdot\left|F_{Y(0)|X}(y_1|X_i)-F_{Y(0)|X}(y_2|X_i)\right|\\
&-P(T_i=0|X_i)^2\cdot\left|F_{Y(0)|X}(y_1|X_i)-F_{Y(0)|X}(y_2|X_i)\right|^2\Bigg)\Bigg\}^{1/2}\\
\le&|\hat{\pi}_{01}(X_i)-\pi^*_{01}(X_i)|\times\left\{P(T_i=0|X_i)\cdot\left|F_{Y(0)|X}(y_1|X_i)-F_{Y(0)|X}(y_2|X_i)\right|\right\}^{1/2}\\
\le&\sup_{x\in\mathcal{X}}|\hat{\pi}_{01}(x)-\pi^*_{01}(x)|\cdot\sqrt{L}\cdot|y_1-y_2|^{1/2}\le|y_1-y_2|^{1/2},
\end{aligned}
$$

where L denotes the Lipschitz's coefficient of $F_{Y(0)|X}(y|x)$ by Assumption 2.4.3, and the last inequality follows from Lemma A.2.2. By the law of iterated expectation, the following L^2-continuity (unconditional version) holds for large enough N:

$$
\left\{\mathbb{E}\left[\sup_{|y_1-y_2|\le\delta}|f(X_i,T_i,Y_i;y_1)-f(X_i,T_i,Y_i;y_2)|^2\right]\right\}^{1/2}\le|y_1-y_2|^{1/2}\text{ for fixed }\delta>0,
$$

therefore, Condition (5.3) of [40] is satisfied. By Theorems 4 and 5 of [40], the function class $\left\{N^{-1/2}\sum_{i=1}^N f(X_i,T_i,Y_i;y);y\in\mathbb{R}\right\}$ is *stochastically equicontinuous*. Therefore,

$$
\text{(A.43)}=o_p(1)\text{ holds uniformly in }y\in\mathbb{R}.
$$

Next we consider (A.44). Using Mean Value Theorem twice, we can decompose (A.44) as follows:

$$
\begin{aligned}
\text{(A.44)} =& \frac{1}{p_0}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg[P(T_i=0|X_i)F_{Y(0)|X}(y|X_i)\rho''(\tilde{\lambda}_{01}^{\top}u_K(X_i))u_K(X_i)^{\top} \\
&\qquad - \int_{\mathcal{X}} F_{Y(0)|X}(y|x)P(T_i=0|X_i=x)\rho''(\tilde{\lambda}_{01}^{\top}u_K(x))u_K(x)^{\top}dF_X(x)\Bigg](\hat{\lambda}_{01}-\lambda_{01}^{*}) \\
=& \frac{1}{p_0}\Bigg\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg[P(T_i=0|X_i)F_{Y(0)|X}(y|X_i)\rho''((\lambda_{01}^{*})^{\top}u_K(X_i))u_K(X_i)^{\top} \\
&\qquad - \int_{\mathcal{X}} F_{Y(0)|X}(y|x)P(T_i=0|X_i=x)\rho''((\lambda_{01}^{*})^{\top}u_K(x))u_K(x)^{\top}dF_X(x)\Bigg] \\
&+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}P(T_i=0|X_i)F_{Y(0)|X}(y|X_i)\rho'''(\xi_3(X_i))(\tilde{\lambda}_{01}-\lambda_{01}^{*})^{\top}u_K(X_i)u_K(X_i)^{\top} \\
&- \sqrt{N}\int_{\mathcal{X}} F_{Y(0)|X}(y|x)P(T_i=0|X_i=x)\rho'''(\xi_3(x))(\tilde{\lambda}_{01}-\lambda_{01}^{*})^{\top}u_K(x)u_K(x)^{\top}dF_X(x)\Bigg\}(\hat{\lambda}_1-\lambda_1^{*}) \\
=& \frac{1}{p_0}\cdot\left(W_K^{(1)}(y)+W_K^{(2)}(y)+W_K^{(3)}(y)\right)^{\top}(\hat{\lambda}_{01}-\lambda_{01}^{*}),
\end{aligned}
$$

where

$$
\begin{aligned}
W_K^{(1)}(y) :=& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg[P(T_i=0|X_i)F_{Y(0)|X}(y|X_i)\rho''((\lambda_{01}^{*})^{\top}u_K(X_i))u_K(X_i) \\
&\qquad - \int_{\mathcal{X}} F_{Y(0)|X}(y|x)P(T_i=0|X_i=x)\rho''((\lambda_{01}^{*})^{\top}u_K(x))u_K(x)dF_X(x)\Bigg], \\
W_K^{(2)}(y) :=& \frac{1}{\sqrt{N}}\left[\sum_{i=1}^{N}P(T_i=0|X_i)F_{Y(0)|X}(y|X_i)\rho'''(\xi_3(X_i))u_K(X_i)u_K(X_i)^{\top}\right](\tilde{\lambda}_{01}-\lambda_{01}^{*}), \\
W_K^{(3)}(y) :=& -\sqrt{N}\int_{\mathcal{X}} P(T=0|X=x)F_{Y(0)|X}(y|x)\rho'''(\xi_3(x))u_K(x)u_K(x)^{\top}dF_X(x)\cdot(\tilde{\lambda}_{01}-\lambda_{01}^{*}),
\end{aligned}
$$

and $\tilde{\lambda}_{01}$ lies on the line joining λ_{01}^{*} and $\hat{\lambda}_{01}$, $\xi_3(x)$ lies between $\tilde{\lambda}_{01}^{\top}u_K(x)$ and $(\lambda_{01}^{*})^{\top}u_K(x)$.

Consider $W_K^{(1)}(y)$. Noting $\mathbb{E}[W_K^{(1)}(y)]=0$ and computing the second moment of $W_K^{(1)}(y)$, we have:

$$\begin{aligned}\mathbb{E}\left[|W_K^{(1)}(y)|^2\right] =&\mathbb{E}[P(T_i=0|X_i)^2 F_{Y(0)|X}(y|X_i)^2\rho''((\lambda_{01}^*)^\top u_K(X_i))^2 u_K(X_i)^\top u_K(X_i)]\\ &-\mathbb{E}[F_{Y(0)|X}(y|X_i)P(T_i=0|X_i)\rho''((\lambda_{01}^*)^\top u_K(X_i))u_K^\top(X_i)]\\ &\quad\times\mathbb{E}[F_{Y(0)|X}(y|X_i)P(T_i=0|X_i)\rho''((\lambda_{01}^*)^\top u_K(X_i))u_K^\top(X_i)]\\ \le&\mathbb{E}[P(T_i=0|X_i)^2 F_{Y(0)|X}(y|X_i)^2\rho''((\lambda_{01}^*)^\top u_K(X_i))^2 u_K(X_i)^\top u_K(X_i)]\\ \le&\, O(1)\cdot\mathbb{E}[\|u_K(X)\|^2]=O(K).\end{aligned}$$

Then by Chebyshev's inequality, we have that for each fixed $y\in\mathbb{R}$,

$$\|W_K^{(1)}(y)\| = O_p(\sqrt{K}). \tag{A.45}$$

We next consider $\sup_{y\in\mathbb{R}}\|W_{jK}^{(3)}(y)\|$. By Lemma A.2.2, we have that

$$\begin{aligned}&\sup_{y\in\mathbb{R}}\|W_K^{(3)}(y)\|^2\\ =&N\cdot\sup_{y\in\mathbb{R}}\left\{(\tilde{\lambda}_{01}-\lambda_{01}^*)^\top\int_{\mathcal{X}}P(T=0|X=x)F_{Y(0)|X}(y|x)\rho'''(\xi_3(x))u_K(x)u_K^\top(x)dF_X(x)\right.\\ &\left.\qquad\times\int_{\mathcal{X}}P(T=0|X=x)F_{Y(0)|X}(y|x)\rho'''(\xi_3(x))u_K(x)u_K(x)^\top dF_X(x)\times(\tilde{\lambda}_{01}-\lambda_{01}^*)\right\}\\ \le&N\cdot\left(\sup_{x\in\mathcal{X}}|\rho'''(\xi_3(x))|\right)^2\cdot(\tilde{\lambda}_{01}-\lambda_{01}^*)^\top\left\{\int_{\mathcal{X}}u_K(x)u_K(x)^\top dF_X(x)\right\}^2(\tilde{\lambda}_{01}-\lambda_{01}^*)\\ \le&N\cdot O_p(1)\cdot\|\tilde{\lambda}_{01}-\lambda_{01}^*\|^2=O_p(K)\,,\end{aligned}$$

where $\xi_3(x)$ lies between $\tilde{\lambda}_{01}^\top u_K(x)$ and $(\lambda_{01}^*)^\top u_K(x)$. From the proof of Lemma A.2.2, we know that $\sup_{x\in\mathcal{X}}|\rho'''(\xi_3(x))|=O_p(1)$. Note $\tilde{\lambda}_{01}$ lies on the line joining $\hat{\lambda}_{01}$ and λ_{01}^*, by Lemma A.2.2, we have $\|\tilde{\lambda}_{01}-\lambda_{01}^*\|=O_p\left(\sqrt{K/N}\right)$. Therefore,

$$\sup_{y\in\mathbb{R}}\|W_K^{(3)}(y)\| = O(\sqrt{K}). \tag{A.46}$$

Finally, we compute the probability order of $\sup_{y\in\mathbb{R}} \|W_K^{(2)}(y)\|$.

$$\begin{aligned}\sup_{y\in\mathbb{R}} \|W_K^{(2)}(y)\|^2 =& N \cdot \sup_{y\in\mathbb{R}} \left\{ (\tilde{\lambda}_{01} - \lambda_{01}^*)^\top \left[\frac{1}{N} \sum_{i=1}^N P(T_i = 0|X_i) F_{Y(0)|X}(y|X_i) \rho'''(\xi_3(X_i)) u_K(X_i) u_K(X_i)^\top \right] \right. \\ & \left. \times \left[\frac{1}{N} \sum_{i=1}^N P(T_i = 0|X_i) F_{Y(0)|X}(y|X_i) \rho'''(\xi_3(X_i)) u_K(X_i) u_K X_i)^\top \right] (\tilde{\lambda}_{01} - \lambda_{01}^*) \right\} \\ \leq& N \cdot \sup_{x\in\mathcal{X}} |\rho'''(\xi_3(x))|^2 \cdot \left\{ (\tilde{\lambda}_{01} - \lambda_{01}^*)^\top \left[\frac{1}{N} \sum_{i=1}^N u_K(X_i) u_K(X_i)^\top \right]^2 (\tilde{\lambda}_{01} - \lambda_{01}^*) \right\} \\ \leq& N \cdot \sup_{x\in\mathcal{X}} |\rho'''(\xi_3(x))|^2 \cdot \|\tilde{\lambda}_{01} - \lambda_{01}^*\|^2 \cdot \lambda_{\max}^2 \left(\frac{1}{N} \sum_{i=1}^N u_K(X_i) u_K(X_i)^\top \right) \\ \leq& N \cdot O_p(1) \cdot O_p(K/N) \cdot O_p(1) = O_p(K),\end{aligned}$$

where $\lambda_{\max} \left(N^{-1} \sum_{i=1}^N u_K(X_i) u_K(X_i)^\top \right)$ denotes the largest eigenvalue of $N^{-1} \sum_{i=1}^N u_K(X_i) u_K(X_i)^\top$, and it converges to $\lambda_{\max}(\mathbb{E}[u_K(X) u_K(X)^\top]) < \infty$. Therefore,

$$\sup_{y\in\mathbb{R}} \|W_K^{(2)}(y)\| = O_p(\sqrt{K}). \tag{A.47}$$

Then in light of Lemma A.2.2 and Assumption 2.4.4, we have that for fixed $y \in \mathbb{R}$,

$$\begin{aligned}\text{(A.44)} &\leq \left(\|W_K^{(1)}(y)\| + \|W_K^{(2)}(y)\| + \|W_K^{(3)}(y)\| \right) \cdot \|\hat{\lambda}_{01} - \lambda_{01}^*\| \\ &\leq \left(O_p(\sqrt{K}) + O_p(\sqrt{K}) + O_p(\sqrt{K}) \right) \cdot O_p\left(\sqrt{\frac{K}{N}} \right) = O_p\left(\sqrt{\frac{K^2}{N}} \right) = o_p(1).\end{aligned}$$

Using a similar argument of showing (A.43) is of $o_p(1)$ uniformly in $y \in \mathbb{R}$, we can obtain

$$\text{(A.44)} = o_p(1) \text{ holds uniformly in } y \in \mathbb{R}.$$

Hence,

$$\text{(A.36)} = o_p(1) \text{ holds uniformly in } y \in \mathbb{R}.$$

***For term* (A.37):** For fixed $y \in \mathbb{R}$, by computing the second moment of (A.37) and using Lemma

A.2.1, we obtain

$$
\begin{aligned}
&\mathbb{E}\left[|(\text{A.37})|^2\right] \\
=&\frac{1}{p_0^2}\mathbb{E}\left[\left\{(\pi_{01}^*(X_i)-\pi_{01}(X_i))\,(1-T_i)\mathbb{1}(Y_i\le y)-\mathbb{E}\left[(\pi_{01}^*(X_i)-\pi_{01}(X_i))\,F_{Y(0)|X}(y|X_i)P(T_i=0|X_i)\right]\right\}^2\right] \\
\le&\frac{1}{p_0^2}\mathbb{E}\left[(\pi_{01}^*(X_i)-\pi_{01}(X_i))^2\cdot(1-T_i)^2\cdot\mathbb{1}(Y_i\le y)^2\right] \\
\le&\frac{1}{p_0^2}\mathbb{E}\left[(\pi_{01}^*(X_i)-\pi_{01}(X_i))^2\right]=O(K^{-2\alpha})=o(1),
\end{aligned}
$$

which implies that for fixed $y\in\mathbb{R}$, (A.37) is of $o_p(1)$. Using a similar argument of showing (A.43) is of $o_p(1)$ uniformly in $y\in\mathbb{R}$, we can obtain that (A.37) is of $o_p(1)$ uniformly in $y\in\mathbb{R}$.

For term (A.38)**:** By Lemma A.2.1, Assumption 2.4.4, we can derive

$$
\begin{aligned}
&\sup_{y\in\mathbb{R}}\left|\sqrt{N}\mathbb{E}\left[F_{Y(0)|X}(y|X)P(T=0|X=x)\,(\pi_{01}^*(X)-\pi_{01}(X))\right]\right| \\
\le&\sqrt{N}\mathbb{E}[|\pi_{01}^*(X)-\pi_{01}(X)|^2]^{\frac{1}{2}}=\sqrt{N}\cdot O\left(K^{-\alpha}\right)=o(1).
\end{aligned}
$$

For term (A.39)**:** Using Mean Value Theorem, we have

$$
\begin{aligned}
&\sqrt{N}\int_{\mathcal{X}}F_{Y(0)|X}(y|x)(\widehat{\pi}_{01}(x)-\pi_{01}^*(x))dF_{X|T=0}(x) \\
=&\sqrt{N}\int_{\mathcal{X}}F_{Y(0)|X}(y|x)\rho''(\tilde{\lambda}_{01}^\top u_K(x))u_K(x)^\top dF_{X|T=0}(x)(\hat{\lambda}_{01}-\lambda_{01}^*) \\
=&-\sqrt{N}\tilde{\Phi}_K^\top(y)(\hat{\lambda}_{01}-\lambda_{01}^*) \\
=&-\sqrt{N}\tilde{\Phi}_K^\top(y)\tilde{\Omega}_K^{-1}\widehat{Q}_{01}'(\lambda_{01}^*).\ \text{(using (A.35))}
\end{aligned}
$$

Therefore, the term (A.39) is exactly zero.

For term (A.40)**:** We can decompose (A.40) as follows:

$$\begin{aligned}
\text{(A.40)} &= \left\{\tilde{\Phi}_K^\top(y)\tilde{\Omega}_K^{-1} - \Phi_K^\top(y)\Omega_K^{-1}\right\}\sqrt{N}\hat{Q}_{01}'(\lambda_{01}^*) \\
&= \left\{\tilde{\Phi}_K(y) - \Phi_K(y)\right\}^\top \tilde{\Omega}_K^{-1}\sqrt{N}\hat{Q}_{01}'(\lambda_{01}^*) + \Phi_K^\top(y)\left\{\tilde{\Omega}_K^{-1} - \Omega_K^{-1}\right\}\sqrt{N}\hat{Q}_{01}'(\lambda_{01}^*). \quad \text{(A.48)}
\end{aligned}$$

Consider the first term in (A.48). Similar to prove the result $\|\bar{Q}_{01}(\lambda_{01}^*)\| = O_p(\sqrt{K/N})$ in (A.19) , we proved that

$$\|\widehat{Q}_{01}(\lambda_{01}^*)\| = O_p\left(\sqrt{\frac{K}{N}}\right). \quad \text{(A.49)}$$

By Mean Value Theorem, we have

$$\begin{aligned}
&\tilde{\Phi}_K(y) - \Phi_K(y) = -\int_{\mathcal{X}} F_{Y(0)|X}(y|x)\left[\rho''(\tilde{\lambda}_{01}^\top u_K(x)) - \rho''((\lambda_{01}^*)^\top u_K(x))\right]u_K(x)dF_{X|T=0}(x) \\
&= -\frac{1}{p_0}\int_{\mathcal{X}} P(T=0|X=x)F_{Y(0)|X}(y|x)\left[\rho''(\tilde{\lambda}_{01}^\top u_K(x)) - \rho''((\lambda_{01}^*)^\top u_K(x))\right]u_K(x)dF_X(x) \\
&= \frac{1}{p_0}\cdot\frac{W_K^{(3)}(y)}{\sqrt{N}}.
\end{aligned}$$

Note that the matrix $\tilde{\Omega}_K$ is negative definite with probability approaching one, $\lambda_{\min}(\tilde{\Omega}_K^{-1}) = \lambda_{\max}(\tilde{\Omega}_K)^{-1} < 0$ and $\lambda_{\max}(\tilde{\Omega}_K)$ is bounded away from zero with probability approaching one, then $|\lambda_{\min}(\tilde{\Omega}_K^{-1})| = O_p(1)$. Using (A.46) and (A.49), we have

$$\begin{aligned}
&\sup_{y\in\mathbb{R}}|(\tilde{\Phi}_K(y) - \Phi_K(y))^\top\tilde{\Omega}_K^{-1}\sqrt{N}\hat{Q}_{01}'(\lambda_{01}^*)| \quad \text{(A.50)} \\
&\leq \frac{1}{p_0}\cdot\sup_{y\in\mathbb{R}}\|(W_K^{(3)}(y))^\top\tilde{\Omega}_K^{-1}\|\|\hat{Q}_{01}'(\lambda_{01}^*)\| \\
&= \frac{1}{p_0}\cdot\sup_{y\in\mathbb{R}}\sqrt{(W_K^{(3)}(y))^\top\left(\tilde{\Omega}_K^{-1}\right)^2 W_K^{(3)}(y)}\cdot\|\hat{Q}_{01}'(\lambda_{01}^*)\| \\
&\leq \frac{1}{p_0}\cdot\sup_{y\in\mathbb{R}}\sqrt{\lambda_{\min}^2(\tilde{\Omega}_K^{-1})W_K^{(3)}(y)^\top\cdot I_K\cdot W_K^{(3)}(y)}\cdot\|\hat{Q}_{01}'(\lambda_{01}^*)\| \\
&\leq O_p(1)O_p(K^{\frac{1}{2}})O_p\left(\sqrt{\frac{K}{N}}\right) = O_p\left(\sqrt{\frac{K^2}{N}}\right).
\end{aligned}$$

For the second term in (A.48),

$$
\begin{aligned}
&\sup_{y\in\mathbb{R}}|\Phi_K^\top(y)(\tilde{\Omega}_K^{-1}-\Omega_K^{-1})\sqrt{N}\hat{Q}_{01}'(\lambda_{01}^*)| \qquad\qquad (A.51)\\
=&\sup_{y\in\mathbb{R}}\sqrt{N}|\hat{Q}_{01}'(\lambda_{01}^*)^\top\tilde{\Omega}_K^{-1}(\Omega_K-\tilde{\Omega}_K)\Omega_K^{-1}\Phi_K(y)|\\
\leq&\sqrt{N}\|\hat{Q}_{01}'(\lambda_{01}^*)\|\cdot\sup_{y\in\mathbb{R}}\|\Phi_K(y)\|\cdot\|\tilde{\Omega}_K^{-1}(\Omega_K-\tilde{\Omega}_K)\Omega_K^{-1}\|\\
=&\sqrt{N}\|\hat{Q}_{01}'(\lambda_{01}^*)\|\cdot\sup_{y\in\mathbb{R}}\|\Phi_K(y)\|\cdot\text{tr}\left(\tilde{\Omega}_K^{-1}(\Omega_K-\tilde{\Omega}_K)\Omega_K^{-1}\Omega_K^{-1}(\Omega_K-\tilde{\Omega}_K)\tilde{\Omega}_K^{-1}\right)^{\frac{1}{2}}\\
=&\sqrt{N}\|\hat{Q}_{01}'(\lambda_{01}^*)\|\cdot\sup_{y\in\mathbb{R}}\|\Phi_K(y)\|\cdot\text{tr}\left(\Omega_K^{-1}\Omega_K^{-1}(\Omega_K-\tilde{\Omega}_{jK})\tilde{\Omega}_K^{-1}\tilde{\Omega}_K^{-1}(\Omega_K-\tilde{\Omega}_K)\right)^{\frac{1}{2}}\\
\leq&\sqrt{N}\|\hat{Q}_{01}'(\lambda_{01}^*)\|\cdot\sup_{y\in\mathbb{R}}\|\Phi_K(y)\|\cdot\left|\lambda_{\max}(\Omega_K^{-1}\Omega_K^{-1})\right|^{\frac{1}{2}}\cdot\text{tr}\left((\Omega_K-\tilde{\Omega}_K)\tilde{\Omega}_K^{-1}\tilde{\Omega}_K^{-1}(\Omega_K-\tilde{\Omega}_K)\right)^{\frac{1}{2}}\\
=&\sqrt{N}\|\hat{Q}_{01}'(\lambda_{01}^*)\|\cdot\sup_{y\in\mathbb{R}}\|\Phi_K(y)\|\cdot\left|\lambda_{\max}(\Omega_K^{-1}\Omega_K^{-1})\right|^{\frac{1}{2}}\cdot\text{tr}\left(\tilde{\Omega}_K^{-1}\tilde{\Omega}_K^{-1}(\Sigma_K-\tilde{\Omega}_K)(\Omega_K-\tilde{\Omega}_K)\right)^{\frac{1}{2}}\\
\leq&\sqrt{N}\|\hat{Q}_{01}'(\lambda_{01}^*)\|\cdot\sup_{y\in\mathbb{R}}\|\Phi_K(y)\|\cdot\left|\lambda_{\max}(\Omega_K^{-1}\Omega_K^{-1})\right|^{\frac{1}{2}}\cdot\left|\lambda_{\max}(\tilde{\Omega}_K^{-1}\tilde{\Omega}_K^{-1})\right|^{\frac{1}{2}}\cdot\text{tr}\left((\Omega_K-\tilde{\Omega}_K)(\Omega_K-\tilde{\Omega}_K)\right)^{\frac{1}{2}}\\
=&\sqrt{N}\|\hat{Q}_{01}'(\lambda_{01}^*)\|\cdot\sup_{y\in\mathbb{R}}\|\Phi_K(y)\|\cdot\left|\lambda_{\min}(\Omega_K^{-1})\right|\cdot\left|\lambda_{\min}(\tilde{\Omega}_K^{-1})\right|\cdot\left\|\Omega_K-\tilde{\Omega}_K\right\|,
\end{aligned}
$$

where the second and third inequalities follow from the fact that $\text{tr}(AB)\leq\lambda_{\max}(B)\text{tr}(A)$ for any symmetric B and positive semidefinite matrix A.

We now estimate the probability order of $\left\|\Omega_K-\tilde{\Omega}_K\right\|$. Using Mean Value Theorem, triangle inequality, and Lemma A.2.2, we can deduce that

$$
\begin{aligned}
&\left\|\Omega_K-\tilde{\Omega}_K\right\| \qquad\qquad (A.52)\\
=&\left\|\frac{1}{p_0}\cdot\mathbb{E}\left[(1-T)\rho''((\lambda_{01}^*)^\top u_K(X))u_K(X)u_K(X)^\top\right]-\frac{\sum_{i=1}^N(1-T_i)\rho''((\tilde{\lambda}_{01})^\top u_K(X_i))u_K(X_i)u_K(X_i)^\top}{\sum_{j=1}^N(1-T_j)}\right\|\\
\leq&\left\|\left\{\frac{1}{p_0}-\frac{1}{N^{-1}\sum_{j=1}^N(1-T_j)}\right\}\cdot\mathbb{E}\left[(1-T)\rho''((\lambda_{01}^*)^\top u_K(X))u_K(X)u_K(X)^\top\right]\right\|\\
&+\frac{1}{N^{-1}\sum_{j=1}^N(1-T_j)}\\
&+\times\left\|\mathbb{E}\left[(1-T)\rho''((\lambda_{01}^*)^\top u_K(X))u_K(X)u_K(X)^\top\right]-\frac{1}{N}\sum_{i=1}^N(1-T_i)\rho''((\lambda_{01}^*)^\top u_K(X_i))u_K(X_i)u_K(X_i)^\top\right\|\\
&+\frac{1}{N^{-1}\sum_{j=1}^N(1-T_j)}\times\left\|\frac{1}{N}\sum_{i=1}^N(1-T_i)\rho'''(\xi_3(X_i))u_K(X_i)u_K(X_i)^\top\cdot(\tilde{\lambda}_K-\lambda_K^*)^\top u_K(X_i)\right\|\\
\leq&O_p(N^{-1/2})+O_p(\zeta(K)\sqrt{K/N})+O_p(\zeta(K)\sqrt{K^2/N})=O_p(\zeta(K)\sqrt{K^2/N}),
\end{aligned}
$$

where $\xi_3(x)$ lies between $(\lambda_{01}^*)^\top u_K(x)$ and $\tilde{\lambda}_{01}^\top u_K(x)$. Note that

$$\sup_{y\in\mathbb{R}} \|\Phi_K(y)\| \leq \mathbb{E}[|\rho''((\lambda_{01}^*)^\top u_K(X))| \cdot \|u_K(X)\||T=0] \leq O(1)\cdot \mathbb{E}[\|u_K(X))\|^2|T=0]^{\frac{1}{2}} = O(\sqrt{K}). \quad \text{(A.53)}$$

Combining (A.51)-(A.53), we can obtain that

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} \|\Phi_{jK}^\top(y)(\tilde{\Omega}_K^{-1} - \Omega_K^{-1})\sqrt{N}\hat{Q}_{01}'(\lambda_{01}^*)\| \\
=&\sqrt{N} O_p\left(\sqrt{\frac{K}{N}}\right) O(\sqrt{K}) O_p(1) O_p(1) O_p\left(\zeta(K)\sqrt{\frac{K^2}{N}}\right) = \ O_p\left(\zeta(K)\sqrt{\frac{K^4}{N}}\right),
\end{aligned}$$

then with (A.48), (A.50) and Assumption 2.4.4, we have

$$\text{(A.40)} = O_p\left(\sqrt{\frac{K^2}{N}}\right) + O_p\left(\zeta(K)\sqrt{\frac{K^4}{N}}\right) = o_p(1) \text{ holds uniformly in } y\in\mathbb{R}.$$

For term (A.41)**:** Note that

$$\begin{aligned}
&\sqrt{N}\cdot\Phi_K^\top(y)\Omega_K^{-1}\hat{Q}_{01}'(\lambda_{01}^*) \\
=&\sqrt{N}\cdot\Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{\sum_{i=1}^N(1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)u_K(X_i)}{\sum_{j=1}^N(1-T_j)} - \frac{\sum_{i=1}^N T_i u_K(X_i)}{\sum_{j=1}^N T_j}\right\} \\
=&\sqrt{N}\cdot\Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{1}{Np_0}\sum_{i=1}^N(1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)u_K(X_i)\right\}\left\{\frac{p_0}{N^{-1}\sum_{j=1}^N(1-T_j)} - 1\right\} \\
&-\sqrt{N}\cdot\Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{N^{-1}\sum_{i=1}^N T_i u_K(X_i)}{p_1}\right\}\left\{\frac{p_1}{N^{-1}\sum_{j=1}^N T_j} - 1\right\} \\
&+\sqrt{N}\cdot\Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{1}{Np_0}\sum_{i=1}^N(1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)u_K(X_i)\right\} \\
&-\sqrt{N}\cdot\Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{N^{-1}\sum_{i=1}^N T_i u_K(X_i)}{p_1}\right\}.
\end{aligned}$$

We claim the following results hold:

$$\sqrt{N} \cdot \Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{1}{Np_0}\sum_{i=1}^N (1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)u_K(X_i)\right\}\left\{\frac{p_0}{N^{-1}\sum_{j=1}^N(1-T_j)}-1\right\}$$
$$=F_{01}(y)\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\frac{1-T_i}{p_0}-1\right\}+o_P(1) \tag{A.54}$$

and

$$\sqrt{N} \cdot \Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{N^{-1}\sum_{i=1}^N T_i u_K(X_i)}{p_1}\right\}\left\{\frac{p_1}{N^{-1}\sum_{j=1}^N T_j}-1\right\}$$
$$=F_{01}(y)\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\frac{T_i}{p_1}-1\right\}+o_P(1) \tag{A.55}$$

and

$$\sqrt{N} \cdot \Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{1}{Np_0}\sum_{i=1}^N (1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)u_K(X_i)\right\}$$
$$=\frac{1}{\sqrt{N}}\sum_{i=1}^N\frac{(1-T_i)\pi_{01}(X_i)F_{Y(0)|X}(y|X_i)}{p_0}+o_P(1) \tag{A.56}$$

and

$$\sqrt{N} \cdot \Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{N^{-1}\sum_{i=1}^N T_i u_K(X_i)}{p_1}\right\}=\frac{1}{\sqrt{N}}\sum_{i=1}^N\frac{T_i\pi_{01}(X_i)F_{Y(0)|X}(y|X_i)}{p_1}+o_P(1). \tag{A.57}$$

Combining (A.54)-(A.57), we can have that (A.41) is of $o_P(1)$.

It remains to prove (A.54)-(A.57). For (A.54), by Lemma A.2.1 we can deduce that

$$
\begin{aligned}
&\sqrt{N}\cdot\Phi_K^\top(y)\Omega_K^{-1}\left\{\frac{1}{Np_0}\sum_{i=1}^N(1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)u_K(X_i)\right\}\left\{\frac{p_0}{N^{-1}\sum_{j=1}^N(1-T_j)}-1\right\}\\
=&\left\{\frac{1}{Np_0}\sum_{i=1}^N(1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)\cdot\left[-\Phi_K^\top(y)\Omega_K^{-1}u_K(X_i)\right]\right\}\frac{1}{\sqrt{N}}\sum_{j=1}^N\frac{(1-T_j)-p_0}{N^{-1}\sum_{i=1}^N(1-T_i)}\\
=&\left\{\frac{1}{Np_0}\sum_{i=1}^N(1-T_i)\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)F_{Y(0)|X}(y|X_i)\right\}\frac{1}{\sqrt{N}}\sum_{j=1}^N\{(1-T_j)-p_0\}\cdot\frac{1}{N^{-1}\sum_{i=1}^N(1-T_i)}\\
&+o_P(1)\\
=&\frac{\mathbb{E}[(1-T)\pi_{01}(X)F_{Y(0)|X}(y|X)]}{\mathbb{E}[1-T]}\cdot\frac{1}{\sqrt{N}}\sum_{j=1}^N\left\{\frac{1-T_j}{p_0}-1\right\}\\
=&F_{01}(y)\cdot\frac{1}{\sqrt{N}}\sum_{j=1}^N\left\{\frac{1-T_j}{p_0}-1\right\}+o_P(1),
\end{aligned}
$$

where the second equality holds by noting the fact that $-\Phi_K^\top(y)\Omega_K^{-1}u_K(x)$ is a weighted least square projection (with the weighting function $-\rho''((\lambda_{01}^*)^\top u_K(x))dF_{X|T=0}(x)$) of $F_{Y(0)|X}(y|x)$ on the space spanned by $\{u_K(x)\}$; the third equality holds by Lemma A.2.1 and Law of Large Number. Using the similar projection argument and Lemma A.2.1, it is easy to show (A.55)-(A.57) also hold.

Finally, by combining the results that (A.36)-(A.41) are of $o_P(1)$, we can claim that (A.34) holds.

A.5 Proof of Theorem 2.5.1

A.5.1 Variance Estimators

By using Taylor's expansion and the first order conditions $\partial\hat{Q}_s(\hat{\lambda}_s)/\partial\lambda=0$ and $\partial\hat{Q}_{st}(\hat{\lambda}_{st})/\partial\lambda=0$, we have

$$
\begin{aligned}
\hat{\lambda}_s-\lambda_s^*&=\left(-\frac{\partial^2\hat{Q}_s(\tilde{\lambda}_s)}{\partial\lambda\partial\lambda^\top}\right)^{-1}\frac{\partial\hat{Q}_s(\lambda_s^*)}{\partial\lambda},\\
\hat{\lambda}_{st}-\lambda_{st}^*&=\left(-\frac{\partial^2\hat{Q}_{st}(\tilde{\lambda}_{st})}{\partial\lambda\partial\lambda^\top}\right)^{-1}\frac{\partial\hat{Q}_{st}(\lambda_{st}^*)}{\partial\lambda}.
\end{aligned}
$$

By Taylor's expansion, Theorems 3 and 4, we have

$$
\begin{aligned}
\hat{F}_s(y) - F_s(y) =& \frac{1}{N}\sum_{i=1}^{N} D_{is}\pi_s^*(X_i)\mathbb{1}(Y_i \le y) - F_s(y) \\
&+ \left[\frac{1}{N}\sum_{i=1}^{N} D_{is}\frac{\partial}{\partial\lambda}\tilde{\pi}_s(X_i)\mathbb{1}(Y_i \le y)u_K^\top(X_i)\right](\hat{\lambda}_s - \lambda_s^*), \\
=& \frac{1}{N}\sum_{i=1}^{N} S_{F_s}(T_i, X_i, Y_i; y) + o_P(N^{-1/2}), \\
\hat{F}_{st}(y) - F_{st}(y) =& \frac{1}{\sum_{i=1}^N D_{is}}\sum_{i=1}^{N} D_{is}\pi_{st}^*(X_i)\mathbb{1}(Y_i \le y) - F_{st}(y) \\
&+ \left[\frac{1}{\sum_{i=1}^N D_{is}}\sum_{i=1}^{N} D_{is}\frac{\partial}{\partial\lambda}\tilde{\pi}_{st}(X_i)\mathbb{1}(Y_i \le y)u_K^\top(X_i)\right](\hat{\lambda}_{st} - \lambda_{st}^*) \\
=& \frac{1}{N}\sum_{i=1}^{N} S_{F_{st}}(T_i, X_i, Y_i; y) + o_P(N^{-1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\theta}_{st} - \theta_{st} =& \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}\pi_s^*(X_i)\pi_t^*(X_j)\mathbb{1}(Y_i \le Y_j) - \theta_{st} \\
&+ \left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}\left(\frac{\partial}{\partial\lambda}\tilde{\pi}_s(X_i)\right)\pi_t^*(X_j)\mathbb{1}(Y_i \le Y_j)u_K^\top(X_i)\right](\hat{\lambda}_s - \lambda_s^*) \\
&+ \left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}\left(\frac{\partial}{\partial\lambda}\tilde{\pi}_t(X_j)\right)\pi_s^*(X_i)\mathbb{1}(Y_i \le Y_j)u_K^\top(X_j)\right](\hat{\lambda}_t - \lambda_t^*) \\
=& \frac{1}{N}\sum_{i=1}^{N} S_{\theta_{st}}(T_i, X_i, Y_i) + o_P(N^{-1/2}) \\
\hat{\theta}_{stt} - \theta_{stt} =& \frac{1}{\sum_{i=1}^N\sum_{j=1}^N D_{is}D_{jt}}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}\pi_{st}^*(X_i)\mathbb{1}(Y_i \le Y_j) - \theta_{stt} \\
&+ \left[\frac{1}{\sum_{i=1}^N\sum_{j=1}^N D_{is}D_{jt}}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}\left(\frac{\partial}{\partial\lambda}\tilde{\pi}_{st}(X_i)\right)\mathbb{1}(Y_i \le Y_j)u_K^\top(X_i)\right] \\
&\quad \times (\hat{\lambda}_{st} - \lambda_{st}^*) \\
=& \frac{1}{N}\sum_{i=1}^{N} S_{\theta_{stt}}(T_i, X_i, Y_i) + o_P(N^{-1/2}).
\end{aligned}
$$

For the double summation terms of $\hat{\theta}_{st} - \theta_{st}$ and $\hat{\theta}_{stt} - \theta_{stt}$, by applying the theory of U-statistics, we can represent them as sum of $i.i.d.$ random variables:

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}\pi_s^*(X_i)\pi_t^*(X_j)\mathbb{1}(Y_i \le Y_j) - \theta_{st} \\
=&\frac{1}{N}\sum_{i=1}^{N}\Bigg\{D_{it}\pi_t^*(X_i)\mathbb{E}\left[D_s\pi_s^*(X)\mathbb{1}(Y \le Y_i)|Y_i\right] \\
&\qquad + D_{is}\pi_s^*(X_i)\mathbb{E}\left[D_t\pi_t^*(X)\mathbb{1}(Y_i \le Y)|Y_i\right]\Bigg\} - 2\theta_{st} + o_P\left(\frac{1}{\sqrt{N}}\right), \qquad \text{(A.58)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{is}D_{jt}\pi_{st}^*(X_i)\mathbb{1}(Y_i \le y) - \theta_{stt} \\
=&\frac{1}{N}\sum_{i=1}^{N}\Bigg\{\frac{D_{it}}{p_tp_s}\mathbb{E}\left[D_s\pi_{st}^*(X)\mathbb{1}(Y \le Y_i)|Y_i\right] + \frac{D_{is}}{p_tp_s}\pi_{st}^*(X_i)\mathbb{E}\left[D_t\mathbb{1}(Y_i \le Y)|Y_i\right] \\
&\qquad - \frac{D_{is}}{p_s}\theta_{stt} - \frac{D_{it}}{p_t}\theta_{stt}\Bigg\} + o_P\left(\frac{1}{\sqrt{N}}\right). \qquad \text{(A.59)}
\end{aligned}
$$

With $\pi_s^*(X), \pi_{st}^*(X)$ replaced by their estimators $\hat{\pi}_s(X)$ $\hat{\pi}_{st}(X)$, and $\frac{\partial^2\hat{Q}_s(\tilde{\lambda}_s)}{\partial\lambda\partial\lambda^\top}$, $\frac{\partial^2\hat{Q}_{st}(\tilde{\lambda}_{st})}{\partial\lambda\partial\lambda^\top}$are replaced by $\frac{\partial^2\hat{Q}_s(\hat{\lambda}_s)}{\partial\lambda\partial\lambda^\top}$, $\frac{\partial^2\hat{Q}_s(\hat{\lambda}_{st})}{\partial\lambda\partial\lambda^\top}$, and $\frac{\partial\hat{Q}_s(\lambda_s^*)}{\partial\lambda}$, $\frac{\partial\hat{Q}_{st}(\lambda_{st}^*)}{\partial\lambda}$ are replaced by $\frac{\partial\hat{Q}_s(\hat{\lambda}_s)}{\partial\lambda}$, $\frac{\partial\hat{Q}_s(\hat{\lambda}_{st})}{\partial\lambda}$, and the expectation terms are replaced by their sample version, we get the estimators for all of the efficient influence $(S_{F_s}, S_{F_{st}}, S_{\theta_{st}}, S_{\theta_{stt}})$.

A.5.2 Consistency

Without loss of generality, we prove $\hat{V}_{\theta_{011}} \xrightarrow{p} V_{\theta_{011}}$, other results can be established in a similar way. Based on the analysis in Section A.5.1, we have

$$\begin{aligned}
&\sqrt{N}\{\hat{\theta}_{011} - \theta_{011}\} \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{T_i}{p_1p_0}\mathbb{E}\left[(1-T)\pi_{01}^*(X)\mathbb{1}(Y \leq Y_i)|Y_i\right] + \frac{1-T_i}{p_1p_0}\pi_{01}^*(X_i)\mathbb{E}\left[T\mathbb{1}(Y_i \leq Y)|Y_i\right] - \frac{T_i}{p_1}\theta_{011} - \frac{1-T_i}{p_0}\theta_{011}\right\} \\
&-\left[\frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N}(1-T_j)T_j}\sum_{i=1}^{N}\sum_{j=1}^{N}(1-T_i)T_j\left(\frac{\partial}{\partial\lambda}\tilde{\pi}_{01}(X_i)\right)\mathbb{1}(Y_i \leq Y_j)u_K^\top(X_i)\right]\frac{\partial^2\hat{Q}_{01}(\tilde{\lambda}_{01})}{\partial\lambda\partial\lambda^\top}^{-1} \\
&\quad\times\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{N(1-T_i)\pi_{01}^*(X_i)u_K(X_i)}{\sum_{j=1}^{N}(1-T_j)} - \frac{NT_iu_K(X_i)}{\sum_{j=1}^{N}T_j}\right\} + o_P(1) \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{S}_{\theta_{011}}(T_i, X_i, Y_i) + o_P(1),
\end{aligned}$$

where

$$\begin{aligned}
&\tilde{S}_{\theta_{011}}(T_i, X_i, Y_i) \\
=&\frac{T_i}{p_1p_0}\mathbb{E}\left[(1-T)\pi_{01}^*(X)\mathbb{1}(Y \leq Y_i)|Y_i\right] + \frac{1-T_i}{p_1p_0}\pi_{01}^*(X_i)\mathbb{E}\left[T\mathbb{1}(Y_i \leq Y)|Y_i\right] - \frac{T_i}{p_1}\theta_{011} - \frac{1-T_i}{p_0}\theta_{011} \\
&-\left[\frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N}(1-T_j)T_j}\sum_{i=1}^{N}\sum_{j=1}^{N}(1-T_i)T_j\left(\frac{\partial}{\partial\lambda}\tilde{\pi}_{01}(X_i)\right)\mathbb{1}(Y_i \leq Y_j)u_K^\top(X_i)\right]\frac{\partial^2\hat{Q}_{01}(\tilde{\lambda}_{01})}{\partial\lambda\partial\lambda^\top}^{-1} \\
&\times\left\{\frac{N(1-T_i)\pi_{01}^*(X_i)u_K(X_i)}{\sum_{j=1}^{N}(1-T_j)} - \frac{NT_iu_K(X_i)}{\sum_{j=1}^{N}T_j}\right\}.
\end{aligned}$$

The estimated influence function is defined by

$$
\begin{aligned}
&\hat{S}_{\theta_{011}}(T_i, X_i, Y_i) \\
=&\frac{N}{\sum_{i=1}^N \sum_{j=1}^N (1-T_i)T_j} \sum_{j=1}^N \{(1-T_j)T_i\hat{\pi}_{01}(X_j)\mathbb{1}(Y_j \le Y_i) + (1-T_i)T_j\hat{\pi}_{01}(X_i)\mathbb{1}(Y_i \le Y_j)\} \\
&-\left(\frac{T_i}{\sum_{i=1}^N T_i/N} - \frac{1-T_i}{\sum_{i=1}^N (1-T_i)/N}\right)\hat{\theta}_{011} \\
&-\left[\frac{1}{\sum_{i=1}^N \sum_{j=1}^N (1-T_j)T_j} \sum_{i=1}^N \sum_{j=1}^N (1-T_i)T_j \left(\frac{\partial}{\partial\lambda}\hat{\pi}_{01}(X_i)\right)\mathbb{1}(Y_i \le Y_j) u_K^\top(X_i)\right] \frac{\partial^2 \hat{Q}_{01}(\hat{\lambda}_{01})}{\partial\lambda\partial\lambda^\top}^{-1} \\
&\quad \times \left\{\frac{N(1-T_i)\hat{\pi}_{01}(X_i)u_K(X_i)}{\sum_{j=1}^N (1-T_j)} - \frac{NT_i u_K(X_i)}{\sum_{j=1}^N T_j}\right\}.
\end{aligned}
$$

The estimator of efficient variance $V_{\theta_{011}}$ is $\hat{V}_{\theta_{011}} = N^{-1}\sum_{i=1}^N \hat{S}_{\theta_{011}}(T_i, X_i, Y_i)^2$. By Theorem 4, we have $N^{-1}\sum_{i=1}^N \tilde{S}_{\theta_{011}}(T_i, X_i, Y_i)^2 \xrightarrow{p} V_{\theta_{011}}$. Therefore, to prove $\hat{V}_{\theta_{011}} \xrightarrow{p} V_{\theta_{011}}$, it is sufficient to prove $N^{-1}\sum_{i=1}^N \hat{S}_{\theta_{011}}(T_i, X_i, Y_i)^2 - N^{-1}\sum_{i=1}^N \tilde{S}_{\theta_{011}}(T_i, X_i, Y_i)^2 \to 0$.

Note that $\tilde{S}_{\theta_{011}}(T_i, X_i, Y_i)$ can be written as follows:

$$
\tilde{S}_{\theta_{011}}(T_i, X_i, Y_i) = \left(-\tilde{L}_K, 1\right)\tilde{b}_{Ki},
$$

where

$$
\begin{aligned}
\tilde{L}_K :=& \left(\frac{1}{\sum_{i=1}^N (1-T_i)\sum_{j=1}^N T_j} \sum_{i=1}^N \sum_{j=1}^N (1-T_i)T_j \rho''\left(\tilde{\lambda}_{01}^\top u_K(X_i)\right)\mathbb{1}(Y_i \le Y_j) u_K^\top(X_i)\right) \\
&\times \left(\frac{1}{\sum_{i=1}^N (1-T_i)} \sum_{i=1}^N (1-T_i)\rho''\left(\tilde{\lambda}_{01}^\top u_K(X_i)\right) u_K(X_i) u_K^\top(X_i)\right)^{-1},
\end{aligned}
$$

and

$$
\tilde{b}_{Ki} := \begin{pmatrix} \frac{N(1-T_i)}{\sum_{i=1}^N (1-T_i)}\rho'\left((\lambda_{01}^*)^\top u_K(X_i)\right)u_K(X_i) - \frac{NT_i}{\sum_{j=1}^N T_j} u_K(X_i) \\ \\ \frac{T_i}{p_1 p_0}\mathbb{E}\left[(1-T)\pi_{01}^*(X)\mathbb{1}(Y \le Y_i)|Y_i\right] + \frac{1-T_i}{p_1 p_0}\pi_{01}^*(X_i)\mathbb{E}\left[T\mathbb{1}(Y_i \le Y)|Y_i\right] - \frac{T_i}{p_1}\theta_{011} - \frac{1-T_i}{p_0}\theta_{011} \end{pmatrix}.
$$

Then

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{S}_{\theta_{011}}(T_i,X_i,Y_i)^2=\left(-\tilde{L}_K,1\right)\tilde{P}_K\left(-\tilde{L}_K,1\right)^\top,$$

where

$$\tilde{P}_K:=\frac{1}{N}\sum_{i=1}^{N}\tilde{b}_{Ki}b_{Ki}^\top,$$

Similarly, $\hat{S}_{\theta_{011}}(T_i,X_i,Y_i)$ can be written as follows:

$$\hat{S}_{\theta_{011}}(T_i,X_i,Y_i)=\left(-\hat{L}_K,1\right)\hat{b}_{Ki},$$

where

$$\begin{aligned}\hat{L}_K:=&\left(\frac{1}{\sum_{i=1}^N(1-T_i)\sum_{j=1}^N T_j}\sum_{i=1}^N\sum_{j=1}^N(1-T_i)T_j\rho''\left(\hat{\lambda}_{01}^\top u_K(X_i)\right)\mathbb{1}(Y_i\le Y_j)u_K^\top(X_i)\right)\\&\times\left(\frac{1}{\sum_{i=1}^N(1-T_i)}\sum_{i=1}^N(1-T_i)\rho''\left(\hat{\lambda}_{01}^\top u_K(X_i)\right)u_K(X_i)u_K^\top(X_i)\right)^{-1},\end{aligned}$$

and

$$\hat{b}_{Ki}=\begin{pmatrix}\frac{N(1-T_i)}{\sum_{i=1}^N(1-T_i)}\rho'\left(\hat{\lambda}_{01}^\top u_K(X_i)\right)u_K(X_i)-\frac{NT_i}{\sum_{j=1}^N T_j}u_K(X_i)\\ \\ \frac{NT_i}{\sum_{i=1}^N\sum_{j=1}^N(1-T_i)T_j}\sum_{j=1}^N(1-T_j)\rho'\left(\hat{\lambda}_{01}^\top u_K(X_j)\right)\mathbb{1}(Y_j\le Y_i)\\ +\frac{N(1-T_i)}{\sum_{i=1}^N\sum_{j=1}^N(1-T_i)T_j}\rho'\left(\hat{\lambda}_{01}^\top u_K(X_i)\right)\sum_{j=1}^N T_j\mathbb{1}(Y_i\le Y_j)-\left(\frac{1-T_i}{\sum_{i=1}^N(1-T_i)/N}+\frac{T_i}{\sum_{j=1}^N T_j/N}\right)\hat{\theta}_{011}\end{pmatrix}.$$

Then

$$\frac{1}{N}\sum_{i=1}^{N}\hat{S}_{\theta_{011}}(T_i,X_i,Y_i)^2=\left(-\hat{L}_K,1\right)\hat{P}_K\left(-\hat{L}_K,1\right)^\top,$$

where

$$\hat{P}_K := \frac{1}{N}\sum_{i=1}^{N}\hat{b}_{Ki}\hat{b}_{Ki}^{\top},$$

Therefore, to prove $N^{-1}\sum_{i=1}^{N}\hat{S}_{\theta_{011}}(T_i,X_i,Y_i)^2 - N^{-1}\sum_{i=1}^{N}\tilde{S}_{\theta_{011}}(T_i,X_i,Y_i)^2 \to 0$, it suffices to prove $\|\tilde{L}_K - \hat{L}_K\| = o_P(1)$ and $\|\tilde{P}_K - \hat{P}_K\| = o_P(1)$, which is also equavalent to prove

$$\left\|\frac{1}{\sum_{i=1}^{N}(1-T_i)}\sum_{i=1}^{N}(1-T_i)\left(\rho''(\tilde{\lambda}_{01}u_K(X_i)) - \rho''(\hat{\lambda}_{01}u_K(X_i))\right)u_K(X_i)u_K^{\top}(X_i)\right\| = o_P(1), \tag{A.60}$$

$$\left\|\frac{1}{\sum_{j=1}^{N}\sum_{i=1}^{N}(1-T_i)T_j}\sum_{i=1}^{N}(1-T_i)\left(\rho''(\tilde{\lambda}_{01}u_K(X_i)) - \rho''(\hat{\lambda}_{01}u_K(X_i))\right)\mathbb{1}(Y_i\le Y_j)u_K^{\top}(X_i)\right\| = o_P(1), \tag{A.61}$$

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}_{Ki}\|^2 - \frac{1}{N}\sum_{i=1}^{N}\|\tilde{b}_{Ki}\|^2\right\| = o_P(1). \tag{A.62}$$

For (A.60), we can deduce that

$$\begin{aligned}
&\left\|\frac{1}{\sum_{i=1}^{N}(1-T_i)}\sum_{i=1}^{N}(1-T_i)\left(\rho''(\tilde{\lambda}_{01}u_K(X_i)) - \rho''(\hat{\lambda}_{01}u_K(X_i))\right)u_K(X_i)u_K^{\top}(X_i)\right\| \\
\le&\left\|\frac{1}{\sum_{i=1}^{N}(1-T_i)}\sum_{i=1}^{N}(1-T_i)\rho'''\left(\xi_5(X_i)\right)u_K(X_i)u_K^{\top}(X_i)\cdot(\tilde{\lambda}_{01}-\hat{\lambda}_{01})^{\top}u_K(X_i)\right\| \\
\le&\frac{1}{\sum_{i=1}^{N}(1-T_i)}\sum_{i=1}^{N}|1-T_i||\rho'''(\xi_5(X_i))|\cdot\|u_K(X_i)\|^3\cdot\|\tilde{\lambda}_{01}-\hat{\lambda}_{01}\| \\
\le&\sup_{x\in\chi}|\rho'''(\xi_5(x))|\cdot\|\tilde{\lambda}_{01}-\hat{\lambda}_{01}\|^2\cdot\frac{Np_0}{\sum_{i=1}^{N}(1-T_i)}\cdot\frac{1}{Np_0}\sum_{i=1}^{N}\|u_K(X_i)\|^3 \\
=&\sup_{x\in\chi}|\rho'''(\xi_5(x))|\cdot\|\tilde{\lambda}_{01}-\hat{\lambda}_{01}\|^2\cdot\frac{Np_0}{\sum_{i=1}^{N}(1-T_i)}\cdot\frac{1}{p_0}\left\{\mathbb{E}\left[\|u_K(X)\|^3\right] + O_P\left(\zeta(K)^2\sqrt{\frac{K}{N}}\right)\right\} \\
\le&\sup_{x\in\chi}|\rho'''(\xi_5(x))|\cdot\|\tilde{\lambda}_{01}-\hat{\lambda}_{01}\|^2\cdot\frac{Np_0}{\sum_{i=1}^{N}(1-T_i)}\cdot\frac{1}{p_0}\left\{\zeta(K)\cdot\mathbb{E}\left[\|u_K(X)\|^2\right] + O_P\left(\zeta(K)^2\sqrt{\frac{K}{N}}\right)\right\} \\
\le&O_P(1)\cdot\frac{K}{N}\cdot\zeta(K)\cdot K = O_P\left(\zeta(K)\frac{K^2}{N}\right) = o_P(1)
\end{aligned}$$

where $\xi_5(X_i)$ lies between $\tilde{\lambda}_{01}^{\top}u_K(X_i)$ and $\hat{\lambda}_{01}^{\top}u_K(X_i)$. By using the similar argument, we can easily obtain (A.61) and (A.62).