3-2024

Bootstrap inference for quantile treatment effects in randomized experiments with matched pairs

Liang JIANG

Xiaobin LIU

Peter C B Phillips
Singapore Management University, peterphillips@smu.edu.sg

Yichong ZHANG
Singapore Management University, yczhang@smu.edu.sg

Bootstrap Inference for Quantile Treatment Effects in Randomized Experiments with Matched Pairs*

Liang Jiang† Xiaobin Liu ‡ Peter C.B. Phillips § Yichong Zhang¶

August 4, 2020

Abstract

This paper examines methods of inference concerning quantile treatment effects (QTEs) in randomized experiments with matched-pairs designs (MPDs). We derive the limit distribution of the QTE estimator under MPDs, highlighting the difficulties that arise in analytical inference due to parameter tuning. We show that the naive weighted bootstrap fails to approximate the limit distribution of the QTE estimator under MPDs because it ignores the dependence structure within the matched pairs. To address this difficulty we propose two bootstrap methods that can consistently approximate the limit distribution: the gradient bootstrap and the weighted bootstrap of the inverse propensity score weighted (IPW) estimator. The gradient bootstrap is free of tuning parameters but requires knowledge of the pair identities. The weighted bootstrap of the IPW estimator does not require such knowledge but involves one tuning parameter. Both methods are straightforward to implement and able to provide pointwise confidence intervals and uniform confidence bands that achieve exact limiting coverage rates. We demonstrate their finite sample performance using simulations and provide an empirical application to a well-known dataset in microfinance.

Keywords: Bootstrap inference, matched pairs, quantile treatment effect, randomized control trials

JEL codes: C14, C21

*We thank Kengo Kato and Yu-Chin Hsu for their valuable comments and suggestions. We are grateful to Esther Duflo and Cynthia Kinnan for providing the data used in the empirical application. Yichong Zhang acknowledges the financial support from Singapore Ministry of Education Tier 2 grant under grant MOE2018-T2-2-169 and the Lee Kong Chian fellowship. Any and all errors are our own.

†Fudan University. E-mail address: jiangliang@fudan.edu.cn.

‡The corresponding author. School of Economics, Academy of Financial Research, and Institute for Fiscal Big-Data & Policy of Zhejiang University. E-mail address: liuxiaobin@zju.edu.cn.

§Yale University, University of Auckland, University of Southampton, and Singapore Management University. E-mail address: peter.phillips@yale.edu

¶Singapore Management University. E-mail address: yczhang@smu.edu.sg.

1 Introduction

Matched-pairs designs (MPDs) have recently seen widespread and increasing use in various randomized experiments conducted by economists. By MPD we mean a randomization scheme that first pairs units based on the closeness of their baseline covariates and then randomly assigns one unit in the pair to be treated. In development economics, researchers routinely pair villages, neighborhoods, micro-enterprises, or townships in their experiments (Banerjee, Duflo, Glennerster, and Kinnan, 2015; Crepon, Devoto, Duflo, and Pariente, 2015; Glewwe, Park, and Zhao, 2016; Groh and Mckenzie, 2016). In labor economics, especially in the field of education, researchers pair schools or students to evaluate the effects of various education interventions (Angrist and Lavy, 2009; Beuermann, Cristia, Cueto, Malam 2015; Fryer, 2017; Fryer, Devi, and Holden, 2017; Bold, Kimenyi, Mwabu, Nganga, and Sandefur, 2018; Fryer, 2018). Bruhn and McKenzie (2009) surveyed leading experts in development field experiments and reported that 56% of them explicitly match pairs of observations on baseline characteristics.

Researchers often use randomized experiments to estimate quantile treatment effects (QTEs) as well as average treatment effects (ATEs). Quantile effects can capture heterogeneity in both the sign and magnitude of treatment effects, which may vary according to position within the distribution of outcomes. A common practice in conducting inference on QTEs is to use bootstrap rather than analytical methods because the latter usually require tuning parameters in implementation. However, the treatment assignment in MPDs introduces negative *dependence* because exactly half of the units are treated. Standard bootstrap inference procedures that rely on cross-sectional *independence* are therefore conservative and lack power. This difficulty raises the question of how to conduct bootstrap inference for QTEs in MPDs in a manner that mitigates these shortcomings.

The present paper addresses this question by showing that both the gradient bootstrap and the weighted bootstrap of the inverse propensity score weighted (IPW) estimator can consistently approximate the limit distribution of the original QTE estimator under MPDs, thereby eliminating asymptotic size distortion in inference. In particular, for testing null hypotheses that the QTEs equal some pre-specified values involving single or multiple quantile indexes (or some pre-specified function over a compact set of quantile indexes), the usual pointwise confidence interval or uniform confidence band constructed by using the corresponding bootstrap standard errors achieves a limiting rejection probability under the null equal to the nominal level.

Our starting point is to derive the limit distribution of the two-sample-difference type QTE estimator in MPDs uniformly over a compact set of quantile indexes. Analytic computation of the variance of the QTE estimator using this limit theory requires estimation of two infinite dimensional nuisance parameters. By implication two tuning parameters are needed for every quantile index of interest. This procedure is inevitably cumbersome and provides the motivation to develop bootstrap methods of inference that reduce the need for tuning parameters.

As noted above, observations under MPDs are generally dependent within the pairs, whereas the usual bootstrap counterparts are asymptotically independent conditional on the data. In accord with this contrasting property of the bootstrap we show that the naive weighted bootstrap fails to approximate the limit distribution of the QTE estimator. Consequently, usual bootstrap tests of the null hypothesis that the QTE equals a pre-specified value are conservative and lack power.

To tackle this shortcoming we propose a gradient bootstrap method and show that it can consistently approximate the limit distribution of the QTE estimator under MPDs uniformly over a compact set of quantile indexes. Hagemann (2017) proposed using the gradient bootstrap for the cluster-robust inference in linear quantile regression models. Like Hagemann (2017), we rely on the gradient bootstrap to avoid estimating the Hessian matrix that involves the infinite-dimensional nuisance parameters. The gradient bootstrap procedure is therefore free of tuning parameters. On the other hand and differing from Hagemann (2017), we construct a specific perturbation of the score based on pair and adjacent pairs of observations, which can capture the dependence structure in the original data.

To implement our gradient bootstrap method, researchers need to know the identities of pairs. Such information may not be available when they are using an experiment that was run by someone else in the past and the randomization procedure may not have been fully described. To address this issue, we propose a weighted bootstrap of the IPW QTE estimator, which can be implemented without such knowledge. We show that such a bootstrap can consistently estimate the asymptotic distribution of the QTE estimator under MPDs. There is a cost to not using information about pair identities as the method requires one tuning parameter for the nonparametric estimation of the propensity score. In spite of this additional cost, this weighted bootstrap method still has an advantage over direct analytic inference because practical implementation of the latter requires more than one tuning parameter.

The contributions in the present paper relate to other recent research. Bai, Shaikh, and Romano (2019) first pointed out that in MPDs the two-sample t-test for the null hypothesis that the ATE equals a pre-specified value is conservative. They then proposed adjusting the standard error of the estimator and studied the validity of the permutation test. This paper complements those results by considering the QTEs and by developing new methods of bootstrap inference. Unlike the permutation test, our methods of bootstrap inference do not require studentization, which is cumbersome in the QTE context. In addition, our weighted bootstrap method complements their results by providing a way to perform inference relating to both ATEs and QTEs when pair identities are unknown. In other work, Bai (2019) investigated the optimality of MPDs in randomized experiments. Zhang and Zheng (2020) considered bootstrap inference under covariate-adaptive randomization. A key difference in our contribution is that in MPDs the number of strata is proportional to the sample size, whereas in covariate-adaptive randomization that number is fixed. In consequence, the present work uses fundamentally different asymptotic arguments and bootstrap methods from those employed by Zhang and Zheng (2020). The present paper also fits within a growing literature that studies inference in randomized experiments (e.g., Hahn, Hirano, and Karlan (2011), Athey and Imbens (2017), Abadie, Chingos, and West (2018), Bugni, Canay, and Shaikh (2018), Tabord-Meehan (2018), and Bugni, Canay, and Shaikh (2019), among others).

The remainder of the paper is organized as follows. Section 2 describes the model setup and notation. Section 3 develops the asymptotic properties of our QTE estimator. In Section 4 we study the naive weighted bootstrap, the gradient bootstrap, and the weighted bootstrap of the IPW estimator. Section 5 provides computational details and recommendations for practitioners. Section 6 reports simulation results. Section 7 gives an empirical applicaiton of our methods of bootstrap inference to the data in Banerjee et al. (2015), examining both the ATEs and QTEs of microfinance on the take-up rates of microcredit. Section 8 concludes. Proofs of all results are in the appendix.

2 Setup and Notation

Denote the potential outcomes for treated and control groups as $Y(1)$ and $Y(0)$, respectively. The treatment status is written as A, where $A = 1$ means treated and $A = 0$ means untreated. The researcher can only observe $\{Y_i, X_i, A_i\}_{i=1}^{2n}$ where $Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$, and $X_i \in \Re^{d_x}$ is a collection of baseline covariates, where d_x is the dimension of X. The parameter of interest is the τth QTE, denoted as

$$q(\tau) = q_1(\tau) - q_0(\tau),$$

where $q_1(\tau)$ and $q_0(\tau)$ are the τth quantiles of $Y(1)$ and $Y(0)$, respectively. The testing problems of interest involve single, multiple, or even a continuum of quantile indexes, as in the following null hypotheses

$$\begin{aligned}
&\mathcal{H}_0 : q(\tau) = \underline{q} \quad \text{versus} \quad q(\tau) \neq \underline{q}, \\
&\mathcal{H}_0 : q(\tau_1) - q(\tau_2) = \underline{q} \quad \text{versus} \quad q(\tau_1) - q(\tau_2) \neq \underline{q}, \text{ and} \\
&\mathcal{H}_0 : q(\tau) = \underline{q}(\tau) \ \forall \tau \in \Upsilon \quad \text{versus} \quad q(\tau) \neq \underline{q}(\tau) \text{ for some } \tau \in \Upsilon,
\end{aligned}$$

for some pre-specified value $\underline{q}$ or function $\underline{q}(\tau)$, where Υ is some compact subset of $(0, 1)$.

The units are grouped into pairs based on the closeness of their baseline covariates, which will be made clear next. We denote the pairs of units as

$$(\pi(2j - 1), \pi(2j)) \text{ for } j \in [n],$$

where $[n] = \{1, \cdots, n\}$ and π is a permutation of $2n$ units based on $\{X_i\}_{i=1}^{2n}$ as specified in Assumption 1(iv) below. Within the pair, one unit is randomly assigned to treatment and the other to control. Specifically, we make the following assumption on the data generating process (DGP) and the treatment assignment rule.

Assumption 1. *(i)* $\{Y_i(1), Y_i(0), X_i\}_{i=1}^{2n}$ *is i.i.d.*

(ii) $\{Y_i(1), Y_i(0)\}_{i=1}^{2n} \perp\!\!\!\perp \{A_i\}_{i=1}^{2n} | \{X_i\}_{i=1}^{2n}$.

(iii) Conditionally on $\{X_i\}_{i=1}^{2n}$*,* $(\pi(2j-1), \pi(2j))$ *for* $j \in [n]$*, are i.i.d. and each uniformly distributed over the values in* $\{(1,0),(0,1)\}$.

(iv) $\frac{1}{n}\sum_{j=1}^{n} \left\|X_{\pi(2j)} - X_{\pi(2j-1)}\right\|_2^r \xrightarrow{p} 0$ *for* $r = 1, 2$.

Assumption 1 is used in Bai et al. (2019) to which we refer readers for more discussion. In Assumption 1(iv), $||\cdot||_2$ denotes Euclidean distance. However, all our results hold if $||\cdot||_2$ is replaced by any distance that is equivalent to it, such as L_∞ distance, L_1 distance, and the Mahalanobis distance when all the eigenvalues of the covariance matrix are bounded and bounded away from zero. Later in Section 4 and following Assumption 4 we provide two cases for which Assumption 1(iv) holds.

3 Estimation

Let $\hat{q}_1(\tau)$ and $\hat{q}_0(\tau)$ be the τth percentiles of outcomes in the treated and control groups, respectively. Then, the τth QTE estimator we consider is just

$$\hat{q}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau).$$

To facilitate further analysis and motivate our bootstrap procedure, we note that $\hat{q}(\tau)$ can be equivalently computed by direct quantile regression. Let

$$(\hat{\beta}_0(\tau), \hat{\beta}_1(\tau)) = \arg\min_b \sum_{i=1}^{2n} \rho_\tau(Y_i - \dot{A}'b),$$

where $\dot{A}_i = (1, A_i)^T$ and $\rho_\tau(u) = u(\tau - 1\{u \leq 0\})$. Then, $\hat{q}(\tau) = \hat{\beta}_1(\tau)$ and $\hat{q}_0(\tau) = \hat{\beta}_0(\tau)$.

Assumption 2. *For* $a = 0, 1$*, denote* $F_a(\cdot)$*,* $F_a(\cdot|x)$*,* $f_a(\cdot)$*, and* $f_a(\cdot|x)$ *as the CDF of* $Y_i(a)$*, the conditional CDF of* $Y_i(a)$ *given* $X_i = x$*, the PDF of* $Y_i(a)$*, and the conditional PDF of* $Y_i(a)$ *given* $X_i = x$*, respectively.*

(i) $f_a(q_a(\tau))$ *is bounded and bounded away from zero uniformly over* $\tau \in \Upsilon$*, and* $f_a(q_a(\tau)|x)$ *is uniformly bounded for* $(x, \tau) \in Supp(X) \times \Upsilon$.

(ii) There exists a function $C(x)$ such that

$$\sup_{\tau\in\Upsilon}|f_a(q_a(\tau)+v|x)-f_a(q_a(\tau)|x)|\leq C(x)|v| \quad \textit{and} \quad \mathbb{E}C(X_i)<\infty.$$

(iii) Let $\mathcal{N}_0$ be a neighborhood of 0. Then, there exists a constant C such that for any $x,x'\in$ Supp(X)

$$\sup_{\tau\in\Upsilon,v\in\mathcal{N}_0}|f_a(q_a(\tau)+v|x')-f_a(q_a(\tau)+v|x)|\leq C||x'-x||_2$$

and

$$\sup_{\tau\in\Upsilon,v\in\mathcal{N}_0}|F_a(q_a(\tau)+v|x)-F_a(q_a(\tau)+v|x')|\leq C||x-x'||_2.$$

Assumption 2(i) is the standard regularity condition widely assumed in quantile estimation. The Lipschitz conditions in Assumptions 2(ii) and 2(iii) are similar in spirit to those assumed in Bai et al. (2019, Assumption 2.1) and ensure that units that are "close" in terms of their baseline covariates are suitably comparable. For $a=0,1$, let $m_{a,\tau}(x,q)=\mathbb{E}(\tau-1\{Y(a)\leq q\}|X=x)$ and $m_{a,\tau}(x)=m_{a,\tau}(x,q_a(\tau))$.

Theorem 3.1. *Suppose Assumptions 1 and 2 hold. Then, uniformly over $\tau\in\Upsilon$,*

$$\sqrt{n}(\hat{q}(\tau)-q(\tau))\rightsquigarrow\mathcal{B}(\tau),$$

where $\mathcal{B}(\tau)$ is a Gaussian process with covariance kernel $\Sigma(\cdot,\cdot)$ such that

$$\begin{aligned}\Sigma(\tau,\tau')=&\frac{\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X)}{f_1(q_1(\tau))f_1(q_1(\tau'))}+\frac{\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{0,\tau}(X)m_{0,\tau'}(X)}{f_0(q_0(\tau))f_0(q_0(\tau'))}\\&+\frac{1}{2}\mathbb{E}\left(\frac{m_{1,\tau}(X)}{f_1(q_1(\tau))}-\frac{m_{0,\tau}(X)}{f_0(q_0(\tau))}\right)\left(\frac{m_{1,\tau'}(X)}{f_1(q_1(\tau'))}-\frac{m_{0,\tau'}(X)}{f_0(q_0(\tau'))}\right).\end{aligned}$$

Several remarks are in order. First, the asymptotic variance of $\hat{q}(\tau)$ under MPDs is

$$\Sigma(\tau,\tau)=\frac{\tau-\tau^2-\mathbb{E}m^2_{1,\tau}(X)}{f_1^2(q_1(\tau))}+\frac{\tau-\tau^2-\mathbb{E}m^2_{0,\tau}(X)}{f_0^2(q_0(\tau))}+\frac{1}{2}\mathbb{E}\left(\frac{m_{1,\tau}(X)}{f_1(q_1(\tau))}-\frac{m_{0,\tau}(X)}{f_0(q_0(\tau))}\right)^2.$$

Further note that the asymptotic variance of $\hat{q}(\tau)$ under simple random sampling is

$$\Sigma^\dagger(\tau,\tau)=\frac{\tau-\tau^2}{f_1^2(q_1(\tau))}+\frac{\tau-\tau^2}{f_0^2(q_0(\tau))}.$$

It is clear that

$$\Sigma^{\dagger}(\tau,\tau)-\Sigma(\tau,\tau)=\frac{1}{2}\mathbb{E}\left(\frac{m_{1,\tau}(X)}{f_1(q_1(\tau))}+\frac{m_{0,\tau}(X)}{f_0(q_0(\tau))}\right)^2\geq 0. \tag{3.1}$$

Equality in the last expression holds when both $m_{1,\tau}(X)$ and $m_{0,\tau}(X)$ are zero, which implies that X is irrelevant to the τth quantiles of $Y(0)$ and $Y(1)$.

Second, the asymptotic variance $\Sigma(\tau,\tau)$ coincides with the semiparametric efficiency bound of the QTE estimator established in Firpo (2007) and Donald and Hsu (2014) for observational data under unconfoundedness.[1] Hahn (1998) points out that, even in the case of simple random sampling, to achieve the semiparametric efficiency bound one needs to use the IPW estimator with a nonparametrically estimated propensity score. We view the MPD as an alternative to achieving such efficiency without nonparametric estimation.[2]

Third, to provide an analytic estimate of the asymptotic variance $\Sigma(\tau,\tau)$ it is necessary at least to estimate the infinite dimensional nuisance parameters $f_1(q_1(\tau))$ and $f_0(q_0(\tau))$, which requires two tuning parameters. Hence, if a researcher is interested in testing a null hypothesis that involves G quantile indexes, $2G$ tuning parameters are needed to estimate $2G$ densities, a cumbersome task in practical work; and to construct a uniform confidence band for the QTE analytically, two tuning parameters are needed at each grid point of the quantile indexes. Moreover, if pair identities are unknown, analytic methods of inference potentially require nonparametric estimation of the quantities $m_{a,\tau}(\cdot)$ for $a=0,1$ as well. There are other practical difficulties. Nonparametric estimation is sometimes sensitive to the choice of tuning parameters and rule-of-thumb tuning parameter selection may not be appropriate for every data generating process (DGP) or every quantile. Use of cross-validation in selecting the tuning parameters is possible in principle but in practice time-consuming. These practical difficulties of analytic methods of inference provide a strong motivation to investigate bootstrap inference procedures are much less reliant on tuning parameters.

4 Bootstrap Inference

This section examines three bootstrap inference procedures for the QTEs in MPDs. We first show that a naive weighted bootstrap method fails to approximate the limit distribution of the QTE estimator derived in Section 3. We then propose two bootstrap methods that can consistently estimate the asymptotic distribution of the QTE estimator.

[1] The propensity score is just a constant of 1/2.

[2] Whether the efficiency bound remains the same under MPDs is still an open question and is an interesting topic for future research.

4.1 Naive Weighted Bootstrap Inference

We first consider the naive weighted bootstrap estimators of $\hat{\beta}_0(\tau)$ and $\hat{\beta}_1(\tau)$. Let

$$(\hat{\beta}_0^w(\tau), \hat{\beta}_1^w(\tau)) = \arg\min_b \sum_{i=1}^{2n} \xi_i \rho_\tau(Y_i - \dot{A}'b),$$

where ξ_i is the bootstrap weight defined in the next assumption.

Assumption 3. *Suppose $\{\xi_i\}_{i=1}^{2n}$ is a sequence of nonnegative i.i.d. random variables with unit expectation and variance and a sub-exponential upper tail.*

Denote $\hat{q}^w(\tau) = \hat{\beta}_1^w(\tau)$ and recall that $\hat{q}(\tau) = \hat{\beta}_1(\tau)$.

Theorem 4.1. *If Assumptions 1–3 hold, then conditional on the data and uniformly over $\tau \in \Upsilon$,*

$$\sqrt{n}(\hat{q}^w(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}^w(\tau),$$

where $\mathcal{B}^w(\tau)$ is a Gaussian process with covariance kernel $\Sigma^\dagger(\cdot,\cdot)$ such that

$$\Sigma^\dagger(\tau,\tau') = \frac{\min(\tau,\tau') - \tau\tau'}{f_1(q_1(\tau))f_1(q_1(\tau'))} + \frac{\min(\tau,\tau') - \tau\tau'}{f_0(q_0(\tau))f_0(q_0(\tau'))}.$$

Three remarks are in order. First, $\Sigma^\dagger(\tau,\tau')$ is just the covariance kernel of the QTE estimator when simple random sampling (instead of the MPD) is used as the treatment assignment rule. It follows that the naive weighted bootstrap fails to approximate the limit distribution of $\hat{q}(\tau)$ ($\hat{\beta}_1(\tau)$). The intuition is straightforward. Given the data, the bootstrap weights are i.i.d. and thus unable to mimic the cross-sectional dependence in the original sample.

Second, it is possible to consider the conventional nonparametric bootstrap in which the bootstrap sample is generated from the empirical distribution of the data. If the observations are i.i.d., van der Vaart and Wellner (1996, Section 3.6) showed that the conventional bootstrap is first-order equivalent to a weighted bootstrap with Poisson(1) weights. However, in the current setting, $\{A_i\}_{i\in[2n]}$ are dependent. It is technically challenging to show rigorously that the above equivalence still holds and this is left for future research.

Third, an alternative procedure is to bootstrap the pairs of observations, i.e., to use the same bootstrap weights for observations indexed by $\pi(2j-1)$ and $\pi(2j)$. But such a bootstrap alone is unable to mimic the dependence structure in the original sample. In fact, the gradient bootstrap procedure proposed below follows this idea and uses the same weight for the observations in the same pair to construct the score $S^*_{n,1}$ defined in (4.5). But in order to construct a final score that can mimic the dependence in the data we need an extra score $S^*_{n,2}$, which is defined in (4.6).

4.2 Gradient Bootstrap Inference

We now approximate the asymptotic distribution of the QTE estimator via the gradient bootstrap. Let $u = \sqrt{n}(b - \beta(\tau))$ be a localizing estimation error parameter. From the derivations in Theorem 3.1, we see that

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = \arg\min_u \sum_{i=1}^{2n} \rho_\tau \left(Y_i - \dot{A}^T \beta(\tau) - \frac{\dot{A}^T u}{\sqrt{n}} \right),$$

where

$$\sum_{i=1}^{2n} \left[\rho_\tau(Y_i - \dot{A}^T \beta(\tau) - \frac{\dot{A}^T u}{\sqrt{n}}) - \rho_\tau(Y_i - \dot{A}^T \beta(\tau)) \right] \approx -u' \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} S_n(\tau) + \frac{u^T Q(\tau) u}{2}, \qquad (4.1)$$

$$S_n(\tau) = \begin{pmatrix} \sum_{i=1}^{2n} \frac{A_i}{\sqrt{n}} (\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\ \sum_{i=1}^{2n} \frac{(1-A_i)}{\sqrt{n}} (\tau - 1\{Y_i(0) \le q_0(\tau)\}) \end{pmatrix},$$

and

$$Q(\tau) = \begin{pmatrix} f_1(q_1(\tau)) + f_0(q_0(\tau)) & f_1(q_1(\tau)) \\ f_1(q_1(\tau)) & f_1(q_1(\tau)) \end{pmatrix}.$$

Minimizing the right side of (4.1) gives

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \approx Q^{-1}(\tau) \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} S_n(\tau). \qquad (4.2)$$

The gradient bootstrap proposes to perturb the objective function by some random error $S_n^*(\tau)$, which will be specified later. This error in turn perturbs the score function $S_n(\tau)$. The corresponding bootstrap estimator $\hat{\beta}^*(\tau)$ solves the following optimization problem

$$\hat{\beta}^*(\tau) = \arg\min_b \sum_{i=1}^{2n} \rho_\tau(Y_i - \dot{A}' b) - \sqrt{n} b^T \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} S_n^*(\tau). \qquad (4.3)$$

By a change of variable and (4.1) we have

$$\sqrt{n}(\hat{\beta}^*(\tau) - \beta(\tau)) \approx \arg\min_u -u' \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} [S_n(\tau) + S_n^*(\tau)] + \frac{u^T Q(\tau) u}{2},$$

which implies

$$\sqrt{n}(\hat{\beta}^*(\tau) - \beta(\tau)) \approx Q^{-1}(\tau) \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} [S_n(\tau) + S_n^*(\tau)]. \tag{4.4}$$

Taking the difference between (4.2) and (4.4), we have

$$\sqrt{n}(\hat{\beta}^*(\tau) - \hat{\beta}(\tau)) \approx Q^{-1}(\tau) \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} S_n^*(\tau).$$

The second element of $\hat{\beta}^*(\tau)$ in (4.3) is the bootstrap version of the QTE estimator, which is denoted $\hat{q}^*(\tau)$. By solving (4.3) we avoid estimating the Hessian $Q(\tau)$, which involves the infinite-dimensional nuisance parameters. Then, for the gradient bootstrap to consistently approximate the limit distribution of the original estimator $\hat{\beta}(\tau)$, we need only construct $S_n^*(\tau)$ in such a way that its weak limit given the data coincides with that of the original score $S_n(\tau)$.

Accordingly, we now show how to specify $S_n^*(\tau)$. Let $\{\eta_j\}_{j=1}^n$ and $\{\hat{\eta}_k\}_{k=1}^{\lfloor n/2 \rfloor}$ be two mutually independent i.i.d. sequences of standard normal random variables. Use the indexes $(j,1),(j,0)$ to denote the indexes in $(\pi(2j-1), \pi(2j))$ with $A=1$ and $A=0$, respectively. For example, if $A_{\pi(2j)} = 1$ and $A_{\pi(2j-1)} = 0$, then $(j,1) = \pi(2j)$ and $(j,0) = \pi(2j-1)$. Similarly, use indexes $(k,1), \cdots, (k,4)$ to denote the first index in $(\pi(4k-3), \cdots, \pi(4k))$ with $A=1$, the first index with $A=0$, the second index with $A=1$, and the second index with $A=0$, respectively. Now let

$$S_n^*(\tau) = \frac{S_{n,1}^*(\tau) + S_{n,2}^*(\tau)}{\sqrt{2}},$$

where

$$S_{n,1}^*(\tau) = \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^n \eta_j \left(\tau - 1\{Y_{(j,1)} \leq \hat{q}_1(\tau)\}\right) \\ \sum_{j=1}^n \eta_j \left(\tau - 1\{Y_{(j,0)} \leq \hat{q}_0(\tau)\}\right) \end{pmatrix} \tag{4.5}$$

and

$$S_{n,2}^*(\tau) = \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{k=1}^{\lfloor n/2 \rfloor} \hat{\eta}_k \left[\left(\tau - 1\{Y_{(k,1)} \leq \hat{q}_1(\tau)\}\right) - \left(\tau - 1\{Y_{(k,3)} \leq \hat{q}_1(\tau)\}\right)\right] \\ \sum_{k=1}^{\lfloor n/2 \rfloor} \hat{\eta}_k \left[\left(\tau - 1\{Y_{(k,2)} \leq \hat{q}_0(\tau)\}\right) - \left(\tau - 1\{Y_{(k,4)} \leq \hat{q}_0(\tau)\}\right)\right] \end{pmatrix}. \tag{4.6}$$

In Section 5 we show how to compute the bootstrap estimator $\hat{\beta}^*(\tau)$ directly from the sub-gradient condition of (4.3). This method avoids the optimization inherent in (4.3) and computation is fast. The following assumption imposes the condition that baseline covariates in adjacent pairs are also 'close'.

Assumption 4. *Suppose that* $\frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \left\| X_{(k,l)} - X_{(k,l')} \right\|_2^r \xrightarrow{p} 0$ *for* $r = 1, 2$ *and* $l, l' \in [4]$.

Assumption 4 and Assumption 1(iv) are jointly equivalent to Bai et al. (2019, Assumption 2.4). We refer readers to Bai et al. (2019) for further discussion of this assumption. In particular, Bai et al. (2019, Theorems 4.1 and 4.2) established two cases under which both Assumption 4 and Assumption 1(iv) hold. We repeat their results below for completeness.

Case (1). Suppose X is a scalar and $\mathbb{E}X^2 < \infty$. Let π be any permutation of $2n$ elements such that $X_{\pi(1)} \leq \cdots \leq X_{\pi(2n)}$. Then, both Assumption 4 and Assumption 1(iv) hold.

Case (2). Suppose $\text{Supp}(X) \subset [0,1]^{d_x}$. Let $\breve{\pi}$ be any permutation of $2n$ elements minimizing $\frac{1}{n}\sum_{j=1}^{n} ||X_{\breve{\pi}(2j-1)} - X_{\breve{\pi}(2j)}||_2$, let $\overline{X}_j = \frac{1}{2}\left(X_{\breve{\pi}(2j-1)} + X_{\breve{\pi}(2j)}\right)$, and let $\overline{\pi}$ be any permutation of n elements minimizing $\frac{1}{n}\sum_{j=1}^{n} ||\overline{X}_{\overline{\pi}(j)} - \overline{X}_{\overline{\pi}(j-1)}||_2$. Then, the permutation π with $\pi(2j) = \breve{\pi}(2\overline{\pi}(j))$ and $\pi(2j-1) = \breve{\pi}(2\overline{\pi}(j)-1)$ satisfies Assumption 4 and Assumption 1(iv).

Denote $\hat{q}^*(\tau) = \hat{\beta}_1^*(\tau)$ and recall that $\hat{q}(\tau) = \hat{\beta}_1(\tau)$. We now have the following result.

Theorem 4.2. *Suppose Assumptions 1, 2, and 4 hold. Then, conditional on the data and uniformly over* $\tau \in \Upsilon$, $\sqrt{n}(\hat{q}^*(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}(\tau)$, *where* $\mathcal{B}(\tau)$ *is the same Gaussian process defined in Theorem 3.1.*

Three remarks on Theorem 4.2 are in order. First, the bootstrap estimator $\hat{q}^*(\tau)$ has the following objectives: (i) to avoid estimating densities; and (ii) to mimick the distribution of the original estimator $\hat{\beta}(\tau)$ under MPDs. Objective (i) relates to the Hessian (Q) and (ii) to the score (S_n) of the quantile regression. The gradient bootstrap provide a flexible approach to achieve both goals.

Second, Bai et al. (2019) showed that adjacent pairs can be used to construct a valid standard error for the ATE estimator under MPDs. Our approach follows their lead and bootstraps pairs and adjacent pairs of units. Theorem 4.2 shows that the limit distribution of the resulting bootstrapped perturbation $S_n^*(\tau)$ given that the data can consistently approximate that of the original score $S_n(\tau)$ uniformly over $\tau \in \Upsilon$. For inference concerning the ATE, it is not necessary to use the gradient bootstrap as the Hessian does not contain any infinite-dimensional nuisance parameters. In fact, the way we compute the perturbation $S_n^*(\tau)$ leads directly to a variance estimator $\hat{\nu}_n^2$ for the ATE estimator $\hat{\Delta} = \frac{1}{n}\sum_{j=1}^{n}(Y_{(j,1)} - Y_{(j,0)})$, where

$$\hat{\nu}_n^2 = \frac{1}{2n}\sum_{j=1}^{n}(Y_{(j,1)} - Y_{(j,0)} - \hat{\Delta})^2 + \frac{1}{2n}\sum_{k=1}^{\lfloor n/2 \rfloor}\left[(Y_{(k,1)} - Y_{(k,3)}) - (Y_{(k,2)} - Y_{(k,4)})\right]^2.$$

By some manipulation, one can show that $\hat{\nu}_n^2$ is numerically the same as the estimate used in the adjusted t-test of Bai et al. (2019, Section 3.3).

Third, to implement the gradient bootstrap, researchers need to know pair identities. That information may not be available when the base experiment was run by others and the randomization procedure not fully detailed. In such cases, we propose bootstrapping the IPW estimator of the QTE, whose validity is established in the next section.

4.3 Weighted Bootstrap of Inverse Propensity Score Weighted Estimator

As indicated in Section 3, the QTE estimator under MPDs achieves the semiparametric efficiency bound established for independent observational data. If we use independent bootstrap weights and seek to maintain efficiency, we need to bootstrap an estimator that can achieve the semiparametric efficiency bound under independent data. As pointed out by Hahn (1998) and Firpo (2007), the IPW estimator with a nonparametrically estimated propensity score satisfies this requirement. Accordingly, we now propose a weighted bootstrap version of the IPW estimator to approximate the limit distribution of the QTE estimator in MPDs.

The sieve method is used to estimate the propensity score. Let $b(X)$ be the K-dimensional sieve basis on X and and $\hat{A}_i$ the estimated propensity score for the ith individual. Then,

$$\hat{A}_i = b(X_i)'\hat{\theta}, \tag{4.7}$$

where ξ_i is the bootstrap weight defined in Assumption 3 and $\hat{\theta} = \arg\min_\theta \sum_{i=1}^{2n} \xi_i (A_i - b(X_i)'\theta)^2$.

Because the true propensity score is 1/2, by setting the first component of $b(X)$ to unity, we have $1/2 = b'(X)\theta_0$ where $\theta_0 = (0.5, 0, \cdots, 0)^T$. The linear probability model for the propensity score is correctly specified. It is possible to use sieve logistic regression to compute the propensity score, as done by Hirano, Imbens, and Ridder (2003), Firpo (2007), and Donald and Hsu (2014). The main benefit of using logistic regression is to guarantee that the estimated propensity score lies between zero and one. For simplicity, we use a linear sieve regression here.

The weighted bootstrap IPW estimator can be computed as

$$\hat{q}^w_{ipw}(\tau) = \hat{q}^w_{ipw,1}(\tau) - \hat{q}^w_{ipw,0}(\tau),$$

where

$$\hat{q}^w_{ipw,1}(\tau) = \arg\min_q \sum_{i=1}^{2n} \frac{\xi_i A_i}{\hat{A}_i} \rho_\tau(Y_i - q) \quad \text{and} \quad \hat{q}^w_{ipw,0}(\tau) = \arg\min_q \sum_{i=1}^{2n} \frac{\xi_i (1-A_i)}{1-\hat{A}_i} \rho_\tau(Y_i - q). \tag{4.8}$$

Assumption 5. *(i) The support of X is compact. The first component of $b(X)$ is 1.*

(ii) $\max_{k\in[K]} \mathbb{E} b_k^2(X_i) \le \overline{C} < \infty$ *for some constant* $\overline{C} > 0$. $\sup_{x \in Supp(X)} ||b(x)||_2 = \zeta(K)$.

(iii) $K^2 \zeta(k)^2 \log(n) = o(n)$.

(iv) With probability approaching one, there exist constants $\underline{C}$ and $\overline{C}$ such that

$$0 < \underline{C} \le \lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{2n} \xi_i b(X_i) b(X_i)'\right) \le \lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{2n} \xi_i b(X_i) b(X_i)'\right) \le \overline{C} < \infty,$$

where $\lambda_{\min}(\mathcal{M})$ *and* $\lambda_{\max}(\mathcal{M})$ *denote the minimum and maximum eigenvalues of matrix* $\mathcal{M}$*.*

(v) There exist $\gamma_1(\tau) \in \Re^K$ *and* $\gamma_0(\tau) \in \Re^K$ *such that*

$$B_{a,\tau}(x) = m_{a,\tau}(x) - b'(x)\gamma_a(\tau), \quad a = 0, 1,$$

and $\sup_{a=0,1,\tau\in\Upsilon,x\in Supp(X)} |B_{a,\tau}(x)| = o(1/\sqrt{n})$*.*

Two remarks are in order. First, requiring X to have a compact support is common in nonparametric sieve estimation. Second, the quantity $\zeta(K)$ depends on the choice of basis functions. For example, $\zeta(K) = O(K^{1/2})$ for B-splines and $\zeta(K) = O(K)$ for power series[3]. Taking B-splines as an example, Assumption 5(iii) requires $K = o(n^{1/3})$. Assumption 5(iv) is standard because $K \ll n$. Assumption 5(v) requires that the approximation error of $m_{a,\tau}(x)$ via a linear sieve function is sufficiently small. For instance, suppose $m_{a,\tau}(x)$ is s-times continuously differentiable in x with all derivatives uniformly bounded by some constant $\overline{C}$, then $\sup_{a=0,1,\tau\in\Upsilon,x\in\text{Supp}(X)} |B_{a,\tau}(x)| = O(K^{-s/d_x})$. Assumptions 5(iii) and 5(v) imply that $K = n^h$ for some $h \in (d_x/(2s), 1/3)$, which implicitly requires $s > 3d_x/2$. The choice of K reflects the usual bias-variance trade-off and is the only tuning parameter that researchers need to specify when implementing this bootstrap method.

Theorem 4.3. *Suppose Assumptions 1–3 and 5 hold, then conditionally on the data and uniformly over* $\gamma \in \Upsilon$*,* $\sqrt{n}(\hat{q}^w_{ipw}(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}(\tau)$*, where* $\mathcal{B}(\tau)$ *is the same Gaussian process as defined in Theorem 3.1.*

The benefit of the weighted bootstrap of the IPW estimator is that it does not require knowledge of the pair identities. The cost is that we have to nonparametrically estimate the propensity score, which requires one tuning parameter and is subject to the usual curse of dimensionality. Nonetheless, we still prefer this bootstrap method of inference to the analytic approach. Analytic estimation of the standard error of the QTE estimator without the knowledge of pair identities requires nonparametric estimation of $\{m_{a,\tau}(X), f_a(q_a(\tau))\}_{a=0,1}$, which involves four tuning parameters. The number of tuning parameters further increases with the number of quantile indexes involved in the null hypothesis and uniform confidence bands for QTE over τ requires $4G$ tuning parameters for grid size G. By contrast, implementation of the weighted bootstrap for the IPW estimator requires estimation of the propensity score only once, requiring use of a single tuning parameter.

Inference concerning the *ATE* in MPDs can also be accomplished via the weighted bootstrap of the IPW *ATE* estimator. A similar argument shows that such a bootstrap can consistently approximate the asymptotic distribution of the ATE estimator under MPDs. This result complements that established by Bai et al. (2019) because it provides a way to make inferences about the ATE in MPDs when information on pair identities is unavailable. That pair identity information is required by Bai et al. (2019) in computing standard errors for their adjusted t-test.

[3]See Chen (2007) for a full discussion of the sieve method.

5 Computation and Guidance for Practitioners

5.1 Computation of the Gradient Bootstrap

In practice, the order of pairs in the dataset is usually arbitrary and does not satisfy Assumption 4. To apply the gradient bootstrap, researchers first need to re-order the pairs. For the jth pair with units indexed by $(j,1)$ and $(j,0)$ in the treatment and control groups, let $\overline{X}_j = \frac{1}{2}\{X_{(j,1)} + X_{(j,0)}\}$. Then, let $\overline{\pi}$ be any permutation of n elements that minimizes

$$\frac{1}{n}\sum_{j=1}^{n} ||\overline{X}_{\overline{\pi}(j)} - \overline{X}_{\overline{\pi}(j-1)}||_2.$$

The pairs are re-ordered by indexes $\overline{\pi}(1), \cdots, \overline{\pi}(n)$. With an abuse of notation, we still index the pairs after re-ordering by $1, \cdots, n$. Note that the original QTE estimator $\hat{q}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau)$ is invariant to the re-ordering.

For the bootstrap sample, we directly compute $\hat{\beta}^*(\tau)$ from the sub-gradient condition of (4.3). Specifically, we compute $\hat{\beta}_0^*(\tau)$ as $Y^0_{(h_0)}$ and $\hat{q}^*(\tau) \equiv \hat{\beta}_1^*(\tau)$ as $Y^1_{(h_1)} - Y^0_{(h_0)}$, where $Y^0_{(h_0)}$ and $Y^1_{(h_1)}$ are the h_0th and h_1th order statistics of outcomes in the treatment and control groups, respectively,[4] and h_0 and h_1 are two integers satisfying

$$n\tau + T^*_{n,a}(\tau) + 1 \geq h_a \geq n\tau + T^*_{n,a}(\tau), \; a = 0, 1, \tag{5.1}$$

with

$$\begin{pmatrix} T^*_{n,1}(\tau) \\ T^*_{n,0}(\tau) \end{pmatrix} = \sqrt{n}S^*_n(\tau) = \frac{1}{\sqrt{2}}\Bigg[\begin{pmatrix} \sum_{j=1}^n \eta_j \left(\tau - 1\{Y_{(j,1)} \leq \hat{q}_1(\tau)\}\right) \\ \sum_{j=1}^n \eta_j \left(\tau - 1\{Y_{(j,0)} \leq \hat{q}_0(\tau)\}\right) \end{pmatrix}$$
$$+ \begin{pmatrix} \sum_{k=1}^{\lfloor n/2 \rfloor} \hat{\eta}_k \left[\left(\tau - 1\{Y_{(k,1)} \leq \hat{q}_1(\tau)\}\right) - \left(\tau - 1\{Y_{(k,3)} \leq \hat{q}_1(\tau)\}\right)\right] \\ \sum_{k=1}^{\lfloor n/2 \rfloor} \hat{\eta}_k \left[\left(\tau - 1\{Y_{(k,2)} \leq \hat{q}_0(\tau)\}\right) - \left(\tau - 1\{Y_{(k,4)} \leq \hat{q}_0(\tau)\}\right)\right] \end{pmatrix}\Bigg].$$

As the probability of $n\tau + T^*_{n,a}(\tau)$ being an integer is zero, h_a is uniquely defined with probability one.

We summarize the steps in the bootstrap procedure as follows.

1. Re-order the pairs.

2. Compute the original estimator $\hat{q}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau)$.

3. Let B be the number of bootstrap replications. Let $\mathcal{G}$ be a grid of quantile indexes. For $b \in [B]$, generate $\{\eta_j\}_{j \in [n]}$ and $\{\hat{\eta}_k\}_{k \in \lfloor n/2 \rfloor}$. Compute $\hat{q}^{*b}(\tau) = Y^1_{(h_1)} - Y^0_{(h_0)}$ for $\tau \in \mathcal{G}$, where h_0 and h_1 are computed in (5.1). Obtain $\{\hat{q}^{*b}(\tau)\}_{\tau \in \mathcal{G}}$.

[4] We assume $Y^a_{(1)} \leq \cdots \leq Y^a_{(n)}$ for $a = 0, 1$.

4. Repeat the above step for $b \in [B]$ and obtain B bootstrap estimators of the QTE, denoted as $\{\hat{q}^{*b}(\tau)\}_{b\in[B],\tau\in\mathcal{G}}$.

5.2 Computation of the Weighted Bootstrap of the IPW estimator

We first provide more details on the sieve basis. Let $b(x) \equiv (b_1(x), \cdots, b_K(x))'$, where $\{b_k(\cdot)\}_{k=1}^K$ are K basis functions of a linear sieve space $\mathcal{B}$. Given that all d_x elements of X are continuously distributed, the sieve space $\mathcal{B}$ can be constructed as follows.

1. For each element $X^{(l)}$ of X, $l = 1, \cdots, d_x$, let $\mathcal{B}_l$ be the univariate sieve space of dimension J_n. One example of $\mathcal{B}_l$ is the linear span of the J_n dimensional polynomials given by

$$\mathcal{B}_l = \left\{ \sum_{k=0}^{J_n} \alpha_k x^k, x \in \text{Supp}(X^{(l)}), \alpha_k \in \Re \right\};$$

   Another is the linear span of r-order splines with J_n nodes given by

$$\mathcal{B}_l = \left\{ \sum_{k=0}^{r-1} \alpha_k x^k + \sum_{j=1}^{J_n} b_j[\max(x - t_j, 0)]^{r-1}, x \in \text{Supp}(X^{(l)}), \alpha_k, b_j \in \Re \right\},$$

   where the grid $-\infty = t_0 \leq t_1 \leq \cdots \leq t_{J_n} \leq t_{J_n+1} = \infty$ partitions $\text{Supp}(X^{(l)})$ into $J_n + 1$ subsets $I_j = [t_j, t_{j+1}) \cap \text{Supp}(X^{(l)})$, $j = 1, \cdots, J_n - 1$, $I_0 = (t_0, t_1) \cap \text{Supp}(X^{(l)})$, and $I_{J_n} = (t_{J_n}, t_{J_n+1}) \cap \text{Supp}(X^{(l)})$.

2. Let $\mathcal{B}$ be the tensor product of $\{\mathcal{B}_l\}_{l=1}^{d_x}$, which is defined as a linear space spanned by the functions $\prod_{l=1}^{d_x} g_l$, where $g_l \in \mathcal{B}_l$. The dimension of $\mathcal{B}$ is then $K \equiv d_x J_n$.

Given the sieve basis, we can estimate the propensity score following (4.7). We then obtain $\hat{q}^w_{ipw,1}(\tau)$ and $\hat{q}^w_{ipw,0}(\tau)$ by solving the sub-gradient conditions for the two optimizations in (4.8). Specifically, we have $\hat{q}^w_{ipw,1}(\tau) = Y_{h'_1}$ and $\hat{q}^w_{ipw,0}(\tau) = Y_{h'_0}$, where the indexes h'_0 and h'_1 satisfy $A_{h'_a} = a$, $a = 0, 1$,

$$\tau \left(\sum_{i=1}^{2n} \frac{\xi_i A_i}{\hat{A}_i} \right) - \frac{\xi_{h'_1}}{\hat{A}_{h'_1}} \leq \sum_{i=1}^{2n} \frac{\xi_i A_i}{\hat{A}_i} 1\{Y_i < Y_{h'_1}\} \leq \tau \left(\sum_{i=1}^{2n} \frac{\xi_i A_i}{\hat{A}_i} \right), \tag{5.2}$$

and

$$\tau \left(\sum_{i=1}^{2n} \frac{\xi_i (1 - A_i)}{1 - \hat{A}_i} \right) - \frac{\xi_{h'_0}}{1 - \hat{A}_{h'_0}} \leq \sum_{i=1}^{2n} \frac{\xi_i (1 - A_i)}{1 - \hat{A}_i} 1\{Y_i < Y_{h'_0}\} \leq \tau \left(\sum_{i=1}^{2n} \frac{\xi_i (1 - A_i)}{1 - \hat{A}_i} \right). \tag{5.3}$$

In the implementation, we set $\{\xi_i\}_{i\in[2n]}$ as i.i.d. standard exponential random variables. In this case, all the equalities in (5.2) and (5.3) hold with probability zero. Thus, h'_1 and h'_0 are uniquely defined with probability one.

We summarize the bootstrap procedure as follows.

1. Compute the original estimator $\hat{q}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau)$.

2. Let B be the number of bootstrap replications. Let $\mathcal{G}$ be a grid of quantile indexes. For $b \in [B]$, generate $\{\xi_i\}_{i\in[2n]}$ as a sequence of i.i.d. exponential random variables. Estimate the propensity score following (4.7). Compute $\hat{q}^{w,b}_{ipw}(\tau) = Y_{h'_1} - Y_{h'_0}$ for $\tau \in \mathcal{G}$, where h'_0 and h'_1 are computed in (5.2) and (5.3), respectively. Obtain $\{\hat{q}^{w,b}_{ipw}(\tau)\}_{\tau\in\mathcal{G}}$.

3. Repeat the above step for $b \in [B]$ and obtain B bootstrap estimators of the QTE, denoted as $\{\hat{q}^{w,b}_{ipw}(\tau)\}_{b\in[B],\tau\in\mathcal{G}}$.

For comparison, we also consider the naive weighted bootstrap in our simulations. Its computation follows a procedure similar to the above with only one difference: the nonparametric estimate $\hat{A}_i$ of the propensity score is replaced by the truth, that is, 1/2.

5.3 Bootstrap Confidence Intervals

Given the bootstrap estimates, we discuss how to conduct bootstrap inference for the null hypotheses with single, multiple, and a continuum of quantile indexes. We take the gradient bootstrap as an example. If the IPW bootstrap is used, one can just replace $\{\hat{q}^{*b}(\tau)\}_{b\in[B],\tau\in\mathcal{G}}$ by $\{\hat{q}^{w,b}_{ipw}(\tau)\}_{b\in[B],\tau\in\mathcal{G}}$ in the following cases.

Case (1). We aim to test the single null hypothesis that $\mathcal{H}_0 : q(\tau) = \underline{q}$ vs. $q(\tau) \neq \underline{q}$. Let $\mathcal{G} = \{\tau\}$ in the procedures described above. Further denote $\mathcal{Q}(\nu)$ as the νth empirical quantile of the sequence $\{\hat{q}^{*b}(\tau)\}_{b\in[B]}$. Let $\alpha \in (0,1)$ be the significance level. We suggest using the bootstrap estimator to construct the standard error of $\hat{q}(\tau)$ as $\hat{\sigma} = \frac{\mathcal{Q}(0.975)-\mathcal{Q}(0.025)}{C_{0.975}-C_{0.025}}$, where C_μ is the μth standard normal critical value. Then the valid confidence interval and Wald test using this standard error are

$$CI_1(\alpha) = (\hat{q}(\tau) - C_{1-\alpha/2}\hat{\sigma}, \hat{q}(\tau) + C_{\alpha/2}\hat{\sigma}),$$

and $1\{\left|\frac{\hat{q}(\tau)-\underline{q}}{\hat{\sigma}}\right| \geq C_{1-\alpha/2}\}$, respectively.

Further denote the standard and percentile bootstrap confidence intervals as CI_2 and CI_3, respectively, where

$$CI_2(\alpha) = (2\hat{q}(\tau) - \mathcal{Q}(1-\alpha/2), 2\hat{q}(\tau) - \mathcal{Q}(\alpha/2))$$

and

$$CI_3(\alpha) = (\mathcal{Q}(\alpha/2), \mathcal{Q}(1-\alpha/2)).$$

Theoretically, CI_1, CI_2, and CI_3 are all valid. When $\alpha = 0.05$, CI_1, CI_2, and CI_3 are centered at $\hat{q}(\tau)$, $2\hat{q}(\tau) - \frac{1}{2}\{\mathcal{Q}(0.975) + \mathcal{Q}(0.025)\}$, and $\frac{1}{2}\{\mathcal{Q}(0.975) + \mathcal{Q}(0.025)\}$, respectively, but share the same length $\mathcal{Q}(0.975) - \mathcal{Q}(0.025)$. In (unreported) simulations, we found that in small samples, CI_1 usually has the best size control while CI_2 over-rejects and CI_3 under-rejects.

Case (2). We aim to test the null hypothesis that $\mathcal{H}_0 : q(\tau_1) - q(\tau_2) = \underline{q}$ vs. $q(\tau_1) - q(\tau_2) \neq \underline{q}$. In this case, let $\mathcal{G} = \{\tau_1, \tau_2\}$. Further, let $\mathcal{Q}(\nu)$ denote the νth empirical quantile of the sequence $\{\hat{q}^{*b}(\tau_1) - \hat{q}^{*b}(\tau_2)\}_{b\in[B]}$, and let $\alpha \in (0, 1)$ be the significance level. For the same reason discussed in case (1), we suggest using the bootstrap standard error to construct the valid confidence interval and Wald test as

$$CI_1(\alpha) = (\hat{q}(\tau_1) - \hat{q}(\tau_2) - C_{1-\alpha/2}\hat{\sigma}, \hat{q}(\tau_1) - \hat{q}(\tau_2) + C_{\alpha/2}\hat{\sigma}),$$

and $1\{\left|\frac{\hat{q}(\tau_1)-\hat{q}(\tau_2)-\underline{q}}{\hat{\sigma}}\right| \geq C_{1-\alpha/2}\}$, respectively, where $\hat{\sigma} = \frac{\mathcal{Q}(0.975)-\mathcal{Q}(0.025)}{C_{0.975}-C_{0.025}}$.

Case (3). We aim to test the null hypothesis that

$$\mathcal{H}_0 : q(\tau) = \underline{q}(\tau) \ \forall \tau \in \Upsilon \text{ vs. } q(\tau) \neq \underline{q}(\tau) \ \exists \tau \in \Upsilon.$$

In theory, we should let $\mathcal{G} = \Upsilon$. In practice, we let $\mathcal{G} = \{\tau_1, \cdots, \tau_G\}$ be a fine grid of Υ where G should be as large as computationally possible. Further, let $\mathcal{Q}_\tau(\nu)$ denote the νth empirical quantile of the sequence $\{\hat{q}^{*b}(\tau)\}_{b\in[B]}$ for $\tau \in \mathcal{G}$. Compute the standard error of $\hat{q}(\tau)$ as

$$\hat{\sigma}_\tau = \frac{\mathcal{Q}_\tau(0.975) - \mathcal{Q}_\tau(0.025)}{C_{0.975} - C_{0.025}}.$$

The uniform confidence band with an α significance level is constructed as

$$CB(\alpha) = \{\hat{q}(\tau) - \mathcal{C}_\alpha\hat{\sigma}_\tau, \hat{q}(\tau) + \mathcal{C}_\alpha\hat{\sigma}_\tau : \tau \in \mathcal{G}\},$$

where the critical value $\mathcal{C}_\alpha$ is computed as

$$\mathcal{C}_\alpha = \inf\left\{z : \frac{1}{B}\sum_{b=1}^{B} 1\left\{\sup_{\tau\in\mathcal{G}}\left|\frac{\hat{q}^{*b}(\tau) - \tilde{q}(\tau)}{\hat{\sigma}_\tau}\right| \leq z\right\} \geq 1 - \alpha\right\}$$

and $\tilde{q}(\tau)$ is first-order equivalent to $\hat{q}(\tau)$ in the sense that $\sup_{\tau\in\Upsilon} |\tilde{q}(\tau) - \hat{q}(\tau)| = o_p(1/\sqrt{n})$. We suggest choosing $\tilde{q}(\tau) = \frac{1}{2}\{\mathcal{Q}_\tau(0.975) + \mathcal{Q}_\tau(0.025)\}$ over other choices such as $\tilde{q}(\tau) = \mathcal{Q}_\tau(0.5)$ and $\tilde{q}(\tau) = \hat{q}(\tau)$ due to its better finite-sample performance. We reject $\mathcal{H}_0$ at an α significance level if $\underline{q}(\cdot) \notin CB(\alpha)$.

5.4 Practical Recommendations

Our practical recommendations are straightforward. If pair identities are known, we suggest using the gradient bootstrap for inference. If pair identities are unknown, we suggest using the weighted bootstrap of the IPW estimator with a nonparametrically estimated propensity score for inference.

6 Simulation

In this section, we assess the finite-sample performance of the methods discussed in Section 4 with a Monte Carlo simulation study. In all cases, potential outcomes for $a \in \{0,1\}$ and $1 \leq i \leq 2n$ are generated as

$$Y_i(a) = \mu_a + m_a(X_i) + \sigma_a(X_i)\varepsilon_{a,i},\ a = 0,1, \tag{6.1}$$

where $\mu_a, m_a(X_i), \sigma_a(X_i)$, and $\varepsilon_{a,i}$ are specified as follows. In each of the specifications below, $n \in \{50, 100\}$ and $(X_i, \varepsilon_{0,i}, \varepsilon_{1,i})$ are i.i.d. The number of replications is 10,000. For bootstrap replications we set $B = 5,000$.

Model 1 $X_i \sim \text{Unif}[0,1]$; $m_0(X_i) = 0$; $m_1(X_i) = 10\left(X_i^2 - \frac{1}{3}\right)$; $\varepsilon_{a,i} \sim N(0,1)$ for $a = 0,1$; $\sigma_0(X_i) = \sigma_0 = 1$ and $\sigma_1(X_i) = \sigma_1$.

Model 2 As in Model 1, but $\sigma_0(X_i) = \left(1 + X_i^2\right)$ and $\sigma_1(X_i) = \left(1 + X_i^2\right)\sigma_1$.

Model 3 $X_i = (\Phi(V_{i1}), \Phi(V_{i2}))'$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and

$$V_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

$m_0(X_i) = \gamma' X_i - 1$; $m_1(X_i) = m_0(X_i) + 10\left(\Phi^{-1}(X_{i1})\Phi^{-1}(X_{i2}) - \rho\right)$; $\varepsilon_{a,i} \sim N(0,1)$ for $a = 0,1$; $\sigma_0(X_i) = \sigma_0 = 1$ and $\sigma_1(X_i) = \sigma_1$. We set $\gamma = (1,1)'$, $\sigma_1 = 1$, $\rho = 0.2$.

Model 4 As in Model 3, but with $\gamma = (1,4)'$, $\sigma_1 = 2$, $\rho = 0.7$.

Pairs are determined similarly to those in Bai et al. (2019). Specifically, if X_i is a scalar, then pairs are determined by sorting $\{X_i\}_{i\in[2n]}$ as described in Case (1) in Section 4.2. If X_i is multi-dimensional, then the pairs are determined by the permutation π described in Case (2) in Section 4.2, which can be obtained by using the *R* package *nbpMatching*. After forming the pairs, we assign treatment status within each pair through a random draw from the uniform distribution over $\{(0,1),(1,0)\}$.

We examine the performance of various tests for ATEs and QTEs at the nominal level $\alpha = 5\%$. For the ATE, we consider the hypothesis that

$$\mathbb{E}(Y(1) - Y(0)) = \text{truth} + \Delta \quad vs. \quad \mathbb{E}(Y(1) - Y(0)) \neq \text{truth} + \Delta.$$

For the QTE, we consider the hypotheses that

$$q(\tau) = \text{truth} + \Delta \quad vs. \quad q(\tau) \neq \text{truth} + \Delta,$$

for $\tau = 0.25$, 0.5, and 0.75,

$$q(0.25) - q(0.75) = \text{truth} + \Delta \quad vs. \quad q(0.25) - q(0.75) \neq \text{truth} + \Delta, \tag{6.2}$$

and

$$q(\tau) = \text{truth} + \Delta \quad \forall \tau \in [0.25, 0.75] \quad vs. \quad q(\tau) \neq \text{truth} + \Delta \quad \exists \tau \in [0.25, 0.75]. \tag{6.3}$$

To illustrate size and power of the tests, we set $\mathcal{H}_0 : \Delta = 0$ and $\mathcal{H}_1 : \Delta = 1/2$. The true value for the ATE is 0, whereas the true values for the QTEs are simulated with a $10,000$ sample size and replications. The computational procedures described in Section 5 are followed to perform the bootstrap and calculate the test statistics. To test the single null hypothesis involving one or two quantile indexes, we use the Wald tests specified in Section 5.3. To test the null hypothesis involving a continuum of quantile indexes, we use the uniform confidence band $CB(\alpha)$ defined in Case (3) in the same section.

The results for the ATEs appear in Table 1. Each row presents a different model and each column reports the rejection probabilities for the various methods. The column 'Naive' refers to the two-sample t-test and 'Adj' refers to the adjusted t-test in Bai et al. (2019); the column 'IPW' corresponds to the t-test with standard errors generated by the weighted bootstrap of the IPW ATE estimator. In all cases, we find that (i) the two-sample t-test has rejection probability under $\mathcal{H}_0$ far below the nominal level and is the least powerful test among the three, and (ii) the adjusted t-test has rejection probability under $\mathcal{H}_0$ close to the nominal level and is not conservative. These results are consistent with those in Bai et al. (2019). The IPW t-test proposed in this paper has performance similar to the adjusted t-test.[5] Under $\mathcal{H}_0$, the test has rejection probability close to 5%; under $\mathcal{H}_1$, it is more powerful than the naive method and has power similar to the adjusted t-test. These findings indicate that the IPW t-test provides an alternative to the adjusted t-test when pair identities are unknown.

The results for QTEs are summarized in Tables 2 and 3. Each table has four panels (Models 1-4). Each row in the panel displays the rejection probabilities for the tests using the standard errors estimated by various bootstrap methods. Specifically, the rows 'Naive weight', 'Gradient',

[5] Throughout this section, we use B-splines to nonparametrically estimate the propensity score in the weighted bootstrap of the IPW estimator. If $\dim(X_i)$=1, we choose the bases $\{1, X, [\max(X - qx_0, X - qx_{0.5})]^2\}$ where qx_0 and $qx_{0.5}$ are quantiles of X at 0 and 50%, respectively; if $\dim(X_i)$=2, we choose the bases $\{1, \max(X_1 - qx_{1,0}, X_1 - x_{1,0.5}), \max(X_2 - qx_{2,0}, X_2 - qx_{2,0.5}), X_1 X_2\}$. The choices of the sieve basis functions and K are adhoc. It is possible to use data-driven methods to select them but a rigorous analysis of the validity of various data-driven methods is beyond the scope of this paper.

Table 1: The Empirical Size and Power of Tests for ATEs

Model	$\mathcal{H}_0$: $\Delta = 0$						$\mathcal{H}_1$: $\Delta = 1/2$					
	$n = 50$			$n = 100$			$n = 50$			$n = 100$		
	Naive	Adj	IPW	Naive	Adj	IPW	Naive	Adj	IPW	Naive	Adj	IPW
1	1.32	5.47	5.44	1.22	5.75	6.00	11.80	29.10	29.44	27.67	49.79	50.46
2	1.85	5.35	5.59	1.64	5.63	5.89	10.43	23.26	24.24	23.72	40.42	41.68
3	1.20	4.76	4.92	0.77	4.68	5.16	1.31	5.66	5.91	1.92	8.13	8.74
4	2.32	6.47	6.01	1.25	5.33	4.74	1.08	5.16	4.35	0.93	5.65	4.89

Notes: The table presents the rejection probabilities for tests of ATEs. The columns 'Naive' and 'Adj' correspond to the two-sample t-test and the adjusted t-test in Bai et al. (2019), respectively; the column 'IPW' corresponds to the t-test using the standard errors estimated by the weighted bootstrap of the IPW ATE estimator.

and 'IPW' respectively correspond to the results of the naive weighted bootstrap, the gradient bootstrap, and the weighted bootstrap of the IPW QTE estimator.

Table 2 reports empirical size and power of the tests with a single null hypothesis involving one or two quantile indexes. Columns '0.25', '0.50', and '0.75' correspond to tests with quantiles at 25%, 50%, and 75%. Column 'Dif' corresponds to the test with null hypothesis (6.2). As expected given Theorem 4.1, the test with standard errors estimated by the naive method performs poorly in all cases. It is conservative under $\mathcal{H}_0$ and lacks power under $\mathcal{H}_1$. In contrast, the test using the standard errors estimated by either the gradient bootstrap or the IPW method has a rejection probability under $\mathcal{H}_0$ that is close to the nominal level in almost all specifications. When the number of pairs is 50, the tests in the 'Dif' column constructed based on either the gradient or the IPW method are slightly conservative. Sizes approach the nominal level when n increases to 100.

Table 3 reports empirical size and power of the uniform confidence bands for the hypothesis specified in (6.3) with a grid $\mathcal{G} = \{0.25, 0.27, \cdots, 0.47, 0.49, 0.5, 0.51, 0.53, \cdots, 0.73, 0.75\}$. The test using standard errors estimated by the naive method has rejection probabilities under $\mathcal{H}_0$ far below the nominal level in all specifications. In Models 1-2, the test using standard errors estimated by either the gradient bootstrap or the IPW bootstrap yields a rejection probability under $\mathcal{H}_0$ that is very close to the nominal level even when the number of pairs is as small as 50. Nonetheless, in Models 3-4, the tests constructed based on both methods are conservative when the number of pairs equals 50. When the number of pairs increases to 100, both tests perform much better and have a rejection probability under $\mathcal{H}_0$ that is close to the nominal level. Under $\mathcal{H}_1$, the tests based on both the gradient and IPW methods are more powerful than those based on the naive method.

In summary, the simulation results in Tables 2 and 3 are consistent with the results in Theorems 4.2 and 4.3: both the gradient bootstrap and the IPW bootstrap provide valid pointwise and uniform inference for QTEs under MPDs. The findings also show that when the information on pair identities is unavailable the IPW method continues to provide a sound basis for inference.

Table 2: The Empirical Size and Power of Tests for QTEs

	$\mathcal{H}_0$: $\Delta = 0$								$\mathcal{H}_1$: $\Delta = 1/2$							
	$n = 50$				$n = 100$				$n = 50$				$n = 100$			
	0.25	0.50	0.75	Dif	0.25	0.50	0.75	Dif	0.25	0.50	0.75	Dif	0.25	0.50	0.75	Dif
Model 1																
Naive weight	3.00	2.00	2.22	1.98	3.12	2.06	1.93	1.73	16.67	6.05	5.56	3.96	34.93	11.56	8.11	7.35
Gradient	5.13	4.82	4.92	3.66	5.07	5.62	5.30	4.04	23.76	13.03	11.27	8.18	42.92	22.91	17.30	14.57
IPW	5.47	5.31	6.17	4.24	5.26	5.83	5.65	3.95	24.81	13.48	12.12	8.40	43.93	23.33	17.21	13.91
Model 2																
Naive weight	3.08	2.32	2.55	1.96	3.64	2.53	2.08	1.87	14.82	6.54	4.71	3.68	30.29	11.50	7.46	6.88
Gradient	4.57	4.63	4.39	3.44	5.00	5.42	5.28	3.68	19.51	12.25	8.76	6.57	35.38	20.86	14.79	12.25
IPW	4.93	5.12	5.78	4.45	5.17	5.73	5.88	4.00	20.29	12.90	10.40	7.35	36.38	21.53	15.14	12.53
Model 3																
Naive weight	2.11	1.03	2.10	0.92	1.56	1.37	1.58	0.86	4.98	2.85	1.92	0.98	6.57	7.14	1.73	1.43
Gradient	5.24	3.06	3.14	1.76	4.83	4.20	4.27	3.01	9.71	7.43	3.22	2.39	13.80	16.72	5.67	4.40
IPW	4.76	3.19	5.61	2.60	4.77	3.71	4.95	3.02	8.75	7.81	5.35	3.09	13.04	15.42	6.06	4.21
Model 4																
Naive weight	2.59	1.71	1.98	1.65	2.65	1.66	1.55	1.23	6.09	1.94	1.76	1.28	9.85	2.98	1.19	1.18
Gradient	4.75	4.00	3.33	2.82	4.70	4.74	5.06	3.88	9.37	5.76	3.35	2.87	14.67	8.88	5.27	4.25
IPW	3.97	3.97	4.91	3.68	4.23	4.51	5.01	3.48	8.08	5.37	4.79	3.26	13.50	8.33	5.17	3.51

Note: The table presents the rejection probabilities for tests of QTEs involving a continuum of quantile indexes. The columns '0.25', '0.50', and '0.75' correspond to tests with quantiles at 25%, 50%, and 75%, respectively; the column 'Dif' corresponds to the test with the null hypothesis specified in (6.2). The rows 'Naive weight', 'Gradient', and 'IPW' correspond to the results of the naive weighted bootstrap, the gradient bootstrap, and the weighted bootstrap of the IPW QTE estimator, respectively.

Table 3: The Empirical Size and Power of Uniform Inferences for QTEs

	$\mathcal{H}_0$: $\Delta = 0$		$\mathcal{H}_1$: $\Delta = 1/2$	
	$n = 50$	$n = 100$	$n = 50$	$n = 100$
Model 1				
Naive weight	1.07	1.52	7.50	18.12
Gradient	4.08	4.64	17.88	33.30
IPW	4.49	4.94	16.30	32.40
Model 2				
Naive weight	1.37	1.85	6.73	16.50
Gradient	3.66	4.57	14.30	27.64
IPW	4.25	4.91	14.27	27.47
Model 3				
Naive weight	0.63	0.63	1.43	3.50
Gradient	1.90	3.07	5.19	13.33
IPW	2.19	2.99	4.25	11.34
Model 4				
Naive weight	0.99	1.00	1.40	3.05
Gradient	2.87	3.72	4.47	8.57
IPW	2.78	3.36	3.18	6.98

Notes: The table presents the rejection probabilities for tests of QTEs. The rows 'Naive weight', 'Gradient' and 'IPW' correspond respectively to the results of the naive weighted bootstrap, the gradient bootstrap, and the weighted bootstrap of the IPW QTE estimator.

7 Empirical Application

Questions surrounding the effectiveness of microfinance as a development tool has sparked a great deal of interest from both policymakers and economists. To answer such questions a growing number of studies have implemented randomized experiments in different settings (see Banerjee, Karlan, and Zinman, 2015, and the references therein). In particular, Banerjee et al. (2015) adopted MPD in their randomization. In this section, we apply the bootstrap methods of inference developed in this paper to their data to examine both the ATEs and QTEs on the take-up rates of microcredit to assess the effectiveness of microfinance.[6]

The sample consists of 104 areas in the city of Hyderabad in India. Based on average per capita consumption and per-household outstanding debt, the areas were grouped into pairs of similar neighborhoods. This segmentation gives 52 pairs in the sample; one area in each pair was

[6]The public-use data provided by the authors does not contain information on pair assignment. We thank Esther Duflo and Cynthia Kinnan for providing us with this information.

Table 4: Summary Statistics

	Total	Treatment group	Control group
Loan take-up rate			
Spandana	0.128(0.140)	0.193(0.131)	0.062(0.117)
Any MFI	0.224(0.152)	0.265(0.151)	0.182(0.143)
Matching variable			
Consumption	1026.4(184.4)	1047.8(195.7)	1005.0(171.5)
Debt	36184.7(36036.5)	32694.1(17755.5)	39675.3(47776.8)
Observations	104	52	52

Notes: Unit of observation: area. The table presents the means and standard deviations (in parentheses) of two outcome variables: the take-up rate of loans from Spandana and the take-up rate of loans from any MFI, and two pair-matching variables: average per capita consumption and per-household debt.

Table 5: ATEs of Micofinance on Take-up Rates of Microcredit

	Naive	Adj	IPW
Spandana	0.131(0.024)	0.131(0.022)	0.131(0.022)
Any MFI	0.083(0.029)	0.083(0.024)	0.083(0.027)

Notes: The table presents the ATE estimates of the effect of microfinance on the takeup rates of microcredit. Standard errors are in parentheses. The columns "Naive" and "Adj" correspond to the two-sample t-test and the adjusted t-test in Bai et al. (2019), respectively. The column "IPW" corresponds to the t-test using the standard errors estimated by the weighted bootstrap of the IPW ATE estimator.

randomly assigned to the treatment group and the other to the control group. In the treatment areas, a group-lending microcredit program was implemented. Banerjee et al. (2015) then examined the impacts of expanding access to microfinance on various outcome variables at two endlines.

Here we focus on the impacts of microfinance on two area-level outcome variables at the first endline. One is the area's take-up rate of loans from Spandana, a microfinance organization that implemented the group-lending microcredit program. The other is the area's take-up rate of loans from any microfinance institutions (MFIs). Table 4 gives descriptive statistics (means and standard deviations) for these two outcome variables as well as the matching variables used by Banerjee et al. (2015) to form the pairs in their experiments.

Table 5 reports the results on the ATE estimates of the effect of microfinance on the take-up rates of microcredit with the standard errors (in parentheses) calculated by three methods. Specifically, the columns 'Naive' and 'Adj' correspond to the two-sample t-test and the adjusted t-test in Bai et al. (2019), respectively; the column 'IPW' corresponds to the t-test using standard

Table 6: QTEs of Micofinance on Take-up Rates of Microcredit

	Naive weight	Gradient	IPW
Panel A. Spandana			
25%	0.082(0.021)	0.082(0.026)	0.082(0.020)
50%	0.182(0.024)	0.182(0.021)	0.182(0.023)
75%	0.229(0.047)	0.229(0.046)	0.229(0.047)
Panel B. Any MFI			
25%	0.056(0.045)	0.056(0.043)	0.056(0.042)
50%	0.082(0.040)	0.082(0.034)	0.082(0.040)
75%	0.141(0.054)	0.141(0.054)	0.141(0.049)

Notes: The table presents the QTE estimates of the effect of microfinance on the take-up rates of microcredit at quantiles 25%, 50%, and 75%. Standard errors are in parentheses. The columns "Naive weight," "Gradient," and "IPW" correspond to the results of the naive weighted bootstrap, the gradient bootstrap, and the weighted bootstrap of the IPW QTE estimator, respectively.

errors estimated by the weighted bootstrap of the IPW ATE estimator.[7] The results lead to the following observations. First, consistent with the findings in Banerjee et al. (2015), the ATE estimates show that expanding access to microfinance has highly significant average effects on the take-up rates of microcredit from both Spandana and any MFIs. Second, the standard errors in the adjusted t-test are lower than those in the naive t-test. This result is consistent with the finding in Bai et al. (2019). More importantly, the standard errors estimated by the IPW weighted bootstrap are also lower than those in the naive t-test and similar to those for the adjusted t-test. For example, in the test of the ATE on the take-up rate of microcredit from Spandana, the IPW method reduces the standard error by 8% compared with the naive one. The magnitude of the reduction is the same as that in the adjusted t-test. These results corroborate our earlier finding that the IPW method is an alternative to the approach adopted in Bai et al. (2019), especially when the information on pair identities is unavailable.

Next, we estimate the QTEs of microfinance on the take-up rates of microcredit and estimate their standard errors by the three methods discussed in Section 4. Table 6 presents the results on the QTE estimates at quantile indexes 0.25, 0.5, and 0.75 with the standard errors (in parentheses) estimated by three different methods. Specifically, the columns 'Naive weight', 'Gradient', and 'IPW' correspond to the results of the naive weighted bootstrap, the gradient bootstrap,[8] and the weighted bootstrap of the IPW QTE estimator, respectively. These results lead to the following

[7]Throughout this section, to nonparametrically estimate the propensity score in the IPW weighted bootstrap, we first standardize the data to have mean zero and variance one and then fit the standardized data via the sieve estimation based on the B-splines with the same basis as used in Section 6.

[8]Using the original pair identities and matching variables in Banerjee et al. (2015), we can re-order the pairs according to the procedure described in Section 5.1. We follow Banerjee et al. (2015) in using Euclidean distance to measure the distance between the covariates in distinctive pairs.

two observations.

First, consistent with the theory in Section 4, the standard errors estimated by the gradient bootstrap or the IPW weighted bootstrap are mostly lower than those estimated by the naive weighted bootstrap. For example in Panel A, at the median, compared with the naive weighted bootstrap, the gradient bootstrap reduces the standard errors by 12.5% and the IPW weighted bootstrap reduces the standard errors by over 4%. In Panel B, all the standard errors computed using methods Gradient and IPW are smaller than those computed using the naive method.

Second, there seem to be considerable heterogeneity in the effects of microfinance. Specifically, the treatment effects of microfinance on the take-up rates of microcredit increase as the quantile indexes increase and the increases are economically substantial. For example, in Panel A, the treatment effect increases by about 122% from the 25th percentile to the median and by about 26% from the median to the 75th percentile. In Panel B, the treatment effect at the 25th percentile is positive but not statistically significantly different from zero. The treatment effect increases by over 46% from the 25th percentile to the median and by about 72% from the median to the 75th percentile. These findings may imply that expanding access to microfinance has small, if not negligible, effects on the take-up rates of microcredit for areas in the lower tail of the distribution but that these effects become stronger for upper-ranked areas, thereby exhibiting the so-called Matthew effect.

The second observation in Table 6 indicates that the heterogeneous effects of microfinance on the take-up rates of microcredit are economically substantial. Are they statistically significant too? In Table 7, we provide statistical tests for the heterogeneity of the QTEs. Specifically, we test the null hypotheses that $q(0.50) - q(0.25) = 0$ and $q(0.75) - q(0.50) = 0$. We find that only the difference between the 25th and median QTEs in Panel A is statistically significant. This finding implies that the statistical evidence of heterogeneous treatment effects of microfinance is strong only for the areas in the lower tail of the distribution and when the loans are from Spandana.

8 Conclusion

This paper has studied estimation and inference of QTEs under MPDs and developed new bootstrap methods to improve statistical performance. Derivation of the limit distribution of QTE estimators under MPDs reveals that analytic methods of inference based on asymptotic theory requires estimation of two infinite-dimensional nuisance parameters for every quantile index of interest. A further limitation is that the naive weighted bootstrap fails to approximate the limit distribution of the QTE estimator as it does not preserve the dependence structure in the original sample. Instead, we propose a gradient bootstrap approach that can consistently approximate the limit distribution of the original estimator and is free of tuning parameters. Implementation of the gradient bootstrap requires knowledge of pair identities. So when such information is unavailable we propose a weighted bootstrap procedure based on the IPW estimator of the QTE and show that

Table 7: Tests for the Difference between Two QTEs of Micofinance

	Naive weight	Gradient	IPW
Panel A. Spandana			
$q(0.50) - q(0.25)$	0.099(0.023)	0.099(0.024)	0.099(0.022)
$q(0.75) - q(0.50)$	0.047(0.046)	0.047(0.046)	0.047(0.045)
Panel B. Any MFI			
$q(0.50) - q(0.25)$	0.026(0.043)	0.026(0.044)	0.026(0.044)
$q(0.75) - q(0.50)$	0.059(0.049)	0.059(0.046)	0.059(0.046)

Notes: The table presents tests for the difference between two QTEs of microfinance on the take-up rates of microcredit. Standard errors are in parentheses. The columns 'Naive weight', 'Gradient', and 'IPW' correspond to the results of the naive weighted bootstrap, the gradient bootstrap, and the weighted bootstrap of the IPW QTE estimator, respectively.

it can consistently approximate the limit distribution of the original QTE estimator. Simulations provide finite-sample evidence of these procedures that support the asymptotic findings. In our empirical application of these bootstrap methods to the real dataset in Banerjee et al. (2015) we find considerable evidence of heterogeneity in the effects of microfinance on the take-up rates of microcredit. In both the simulations and the empirical application, the two recommended bootstrap methods of inference perform well in the sense that they usually provide smaller standard errors and greater inferential accuracy than those obtained by naive bootstrap methods.

A Proof of Theorem 3.1

Let $u = (u_0, u_1)' \in \Re^2$ and

$$L_n(u,\tau) = \sum_{i=1}^{2n} \left[\rho_\tau(Y_i - \dot{A}_i'\beta(\tau) - \dot{A}_i'u/\sqrt{n}) - \rho_\tau(Y_i - \dot{A}_i'\beta(\tau)) \right].$$

Then, by change of variables we have

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = \arg\min_u L_n(u,\tau).$$

Note that $L_n(u,\tau)$ is convex in u for each τ and bounded in τ for each u. We divide the proof into three steps. In Step (1), we show that there exists

$$g_n(u,\tau) = -u'W_n(\tau) + \frac{u'Q(\tau)u}{2}$$

such that for each u,

$$\sup_{\tau\in\Upsilon} |L_n(u,\tau) - g_n(u,\tau)| \xrightarrow{p} 0;$$

and the maximum eigenvalue of $Q(\tau)$ is bounded from above and the minimum eigenvalue of $Q(\tau)$ is bounded away from 0, uniformly over $\tau \in \Upsilon$. In Step (2), we show $W_n(\tau)$ as a stochastic process over $\tau \in \Upsilon$ is tight. Then by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = [Q(\tau)]^{-1}W_n(\tau) + r_n(\tau),$$

where $\sup_{\tau\in\Upsilon} ||r_n(\tau)|| = o_p(1)$. Last, in Step (3), we establish weak convergence of $[Q(\tau)]^{-1}W_n(\tau)$ uniformly over $\tau \in \Upsilon$. The second element of the limiting process is $\mathcal{B}(\tau)$ stated in Theorem 3.1.

Step (1). By Knight's identity (Knight, 1998), we have

$$\begin{aligned}
&L_n(u,\tau) \\
=& -\sum_{i=1}^{2n} \frac{u'}{\sqrt{n}}\dot{A}_i \left(\tau - 1\{Y_i \le \dot{A}_i'\beta(\tau)\}\right) + \sum_{i=1}^{2n} \int_0^{\frac{\dot{A}_i'u}{\sqrt{n}}} \left(1\{Y_i - \dot{A}_i'\beta(\tau) \le v\} - 1\{Y_i - \dot{A}_i'\beta(\tau) \le 0\}\right) dv \\
\equiv& -u'W_n(\tau) + Q_n(u,\tau),
\end{aligned}$$

where

$$W_n(\tau) = \sum_{i=1}^{2n} \frac{1}{\sqrt{n}}\dot{A}_i \left(\tau - 1\{Y_i \le \dot{A}_i'\beta(\tau)\}\right)$$

and

$$
\begin{aligned}
Q_n(u,\tau) =& \sum_{i=1}^{2n} \int_0^{\frac{\dot{A}_i' u}{\sqrt{n}}} \left(1\{Y_i - \dot{A}_i'\beta(\tau) \le v\} - 1\{Y_i - \dot{A}_i'\beta(\tau) \le 0\}\right) dv \\
=& \sum_{i=1}^{2n} A_i \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv \\
&+ \sum_{i=1}^{2n} (1-A_i) \int_0^{\frac{u_0}{\sqrt{n}}} \left(1\{Y_i(0) - q_0(\tau) \le v\} - 1\{Y_i(0) - q_0(\tau) \le 0\}\right) dv \\
\equiv& Q_{n,1}(u,\tau) + Q_{n,0}(u,\tau). \qquad \text{(A.1)}
\end{aligned}
$$

We first consider $Q_{n,1}(u,\tau)$. Let

$$
H_n(X_i,\tau) = \mathbb{E}\left(\int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv | X_i \right). \qquad \text{(A.2)}
$$

Then,

$$
\begin{aligned}
Q_{n,1}(u,\tau) =& \sum_{i=1}^{2n} \frac{H_n(X_i,\tau)}{2} + \sum_{i=1}^{2n} \left(A_i - \frac{1}{2}\right) H_n(X_i,\tau) \\
&+ \sum_{i=1}^{2n} A_i \left[\int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv - H_n(X_i,\tau) \right]. \qquad \text{(A.3)}
\end{aligned}
$$

For the first term on the RHS of (A.3), we have, uniformly over $\tau \in \Upsilon$,

$$
\sum_{i=1}^{2n} \frac{H_n(X_i,\tau)}{2} = \frac{1}{4n} \sum_{i=1}^{2n} f_1(q_1(\tau) + \tilde{v}|X_i)(u_0+u_1)^2 \xrightarrow{p} \frac{f_1(q_1(\tau))(u_0+u_1)^2}{2}, \qquad \text{(A.4)}
$$

where $\tilde{v}$ is between 0 and $|u_0+u_1|/\sqrt{n}$ and we use the fact that, due to Assumption 2,

$$
\sup_{\tau\in\Upsilon} \frac{1}{2n} \sum_{i=1}^{2n} |f_1(q_1(\tau) + \tilde{v}|X_i) - f_1(q_1(\tau)|X_i)| \le \left(\frac{1}{2n} \sum_{i=1}^{2n} C(X_i) \right) \frac{|u_0+u_1|}{\sqrt{n}} \xrightarrow{p} 0.
$$

Lemma E.2 shows

$$
\sup_{\tau\in\Upsilon} \left| \sum_{i=1}^{2n} \left(A_i - \frac{1}{2}\right) H_n(X_i,\tau) \right| = o_p(1) \qquad \text{(A.5)}
$$

and

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n} A_i\left[\int_0^{\frac{u_0+u_1}{\sqrt{n}}}\left(1\{Y_i(1)-q_1(\tau)\le v\}-1\{Y_i(1)-q_1(\tau)\le 0\}\right)dv - H_n(X_i,\tau)\right]\right| = o_p(1). \tag{A.6}$$

Combining (A.3)–(A.6), we have

$$\sup_{\tau\in\Upsilon}\left|Q_{n,1}(u,\tau)-\frac{f_1(q_1(\tau))(u_0+u_1)^2}{2}\right| = o_p(1). \tag{A.7}$$

By a similar argument, we can show that

$$\sup_{\tau\in\Upsilon}\left|Q_{n,0}(u,\tau)-\frac{f_0(q_0(\tau))u_0^2}{2}\right| = o_p(1). \tag{A.8}$$

Combining (A.7) and (A.8), we have

$$Q_n(u,\tau)\xrightarrow{p}\frac{u'Q(\tau)u}{2},$$

where

$$Q(\tau)=\begin{pmatrix} f_1(q_1(\tau))+f_0(q_0(\tau)) & f_1(q_1(\tau)) \\ f_1(q_1(\tau)) & f_1(q_1(\tau))\end{pmatrix}. \tag{A.9}$$

Then,

$$\sup_{\tau\in\Upsilon}|L_n(u,\tau)-g_n(u,\tau)| = \sup_{\tau\in\Upsilon}\left|Q_n(u,\tau)-\frac{u'Q(\tau)u}{2}\right| = o_p(1).$$

Last, because $f_a(q_a(\tau))$ for $a=0,1$ is bounded and bounded away from zero uniformly over $\tau\in\Upsilon$, so are the eigenvalues of $Q(\tau)$ uniformly over $\tau\in\Upsilon$.

Step (2). Let $e_1=(1,1)^T$, $e_0=(1,0)^T$. Then,

$$\begin{aligned} W_n(\tau) =& \sum_{i=1}^{2n}\frac{e_1}{\sqrt{n}}A_i\left(\tau-1\{Y_i(1)\le q_1(\tau)\}\right)+\sum_{i=1}^{2n}\frac{e_0}{\sqrt{n}}(1-A_i)\left(\tau-1\{Y_i(0)\le q_0(\tau)\}\right) \\ \equiv& e_1W_{n,1}(\tau)+e_0W_{n,0}(\tau). \end{aligned} \tag{A.10}$$

Recall $m_{1,\tau}(X_i)=\mathbb{E}(\tau-1\{Y_i(1)\le q_1(\tau)\}|X_i)$. Denote

$$\eta_{i,1}(\tau)=\tau-1\{Y_i(1)\le q_1(\tau)\}-m_{1,\tau}(X_i).$$

For $W_{n,1}(\tau)$, we have

$$W_{n,1}(\tau) = \sum_{i=1}^{2n} \frac{A_i}{\sqrt{n}} \eta_{i,1}(\tau) + \sum_{i=1}^{2n} \frac{1}{2\sqrt{n}} m_{1,\tau}(X_i) + R_1(\tau) \tag{A.11}$$

where

$$R_1(\tau) = \sum_{i=1}^{2n} \frac{(A_i - 1/2)}{\sqrt{n}} m_{1,\tau}(X_i).$$

By Lemma E.3, we have

$$\sup_{\tau \in \Upsilon} |R_1(\tau)| = o_p(1).$$

Next, we focus on the first two terms on the RHS of (A.11). Note $\{Y_i(1)\}_{i=1}^{2n}$ given $\{X_i\}_{i=1}^{2n}$ is an independent sequence that is also independent of $\{A_i\}_{i=1}^{2n}$. Let $\tilde{Y}_j(1)|\tilde{X}_j$ be distributed according to $Y_{i_j}(1)|X_{i_j}$ where i_j is the j-th smallest index in the set $\{i \in [2n] : A_i = 1\}$ and $\tilde{X}_j = X_{i_j}$. Then, by noticing that $\sum_{i=1}^{2n} A_i = n$, we have

$$\sum_{i=1}^{2n} \frac{A_i}{\sqrt{n}} \eta_{i,1}(\tau) | \{A_i, X_i\}_{i=1}^{2n} \stackrel{d}{=} \sum_{j=1}^{n} \frac{\tilde{\eta}_{j,1}(\tau)}{\sqrt{n}} \bigg| \{\tilde{X}_j\}_{j=1}^{n}, \tag{A.12}$$

where $\tilde{\eta}_{j,1}(\tau) = \tau - 1\{\tilde{Y}_j(1) \leq q_1(\tau)\} - m_{1,\tau}(\tilde{X}_j)$, and given $\{\tilde{X}_j\}_{j=1}^{n}$, $\{\tilde{\eta}_{j,1}(\tau)\}_{j=1}^{n}$ is a sequence of independent random variables. Further denote the conditional distribution of $\tilde{Y}_j(1)$ given $\tilde{X}_j$ as $\mathbb{P}^{(j)}$ and $\Lambda_\tau(x) = F_1(q_1(\tau)|x)(1 - F_1(q_1(\tau)|x))$. Then,

$$\begin{aligned}
\frac{1}{n} \sum_{j=1}^{n} \mathbb{P}^{(j)} (\tilde{\eta}_{j,1}(\tau))^2 &= \frac{1}{n} \sum_{j=1}^{n} \Lambda_\tau(\tilde{X}_j) \\
&= \frac{1}{n} \sum_{i=1}^{2n} A_i \Lambda_\tau(X_i) \\
&= \frac{1}{2n} \sum_{i=1}^{2n} \Lambda_\tau(X_i) + \frac{1}{2n} \sum_{j=1}^{n} (A_{\pi(2j-1)} - A_{\pi(2j)}) \left[\Lambda_\tau(X_{\pi(2j-1)}) - \Lambda_\tau(X_{\pi(2j)})\right] \\
&\xrightarrow{p} \mathbb{E}\Lambda_\tau(X_i),
\end{aligned}$$

where the last convergence holds because

$$\frac{1}{2n} \sum_{i=1}^{2n} \Lambda_\tau(X_i) \xrightarrow{p} \mathbb{E}\Lambda_\tau(X_i),$$

and

$$\left|\frac{1}{2n}\sum_{j=1}^{n}(A_{\pi(2j-1)}-A_{\pi(2j)})\left[\Lambda_\tau(X_{\pi(2j-1)})-\Lambda_\tau(X_{\pi(2j)})\right]\right| \lesssim \frac{1}{2n}\sum_{j=1}^{n}||X_{\pi(2j-1)}-X_{\pi(2j)}||_2 \xrightarrow{p} 0.$$

In addition, because $\tilde{\eta}_{j,1}(\tau)$ is bounded, the Lyapounov condition holds, i.e.,

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{P}^{(j)}|\tilde{\eta}_{j,1}(\tau)|^3 \xrightarrow{p} 0.$$

Therefore, by the triangular array CLT, for fixed τ, we have

$$\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)|\{A_i,X_i\}_{i=1}^{2n} \stackrel{d}{=} \sum_{j=1}^{n}\frac{\tilde{\eta}_{j,1}(\tau)}{\sqrt{n}}\bigg|\{\tilde{X}_j\}_{j=1}^{n} \rightsquigarrow \mathcal{N}(0,\mathbb{E}\Lambda_\tau(X_i)).$$

It is straightforward to extend the results to finite-dimensional convergence by the Cramér-Wold device. In particular, the covariance between $\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)$ and $\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau')$ conditionally on $\{X_i\}_{i=1}^{2n}$ converges to

$$\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X).$$

Next, we show that the process $\{\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau):\tau\in\Upsilon\}$ is stochastically equicontinuous. Denote $\overline{\mathbb{P}}f=\frac{1}{n}\sum_{j=1}^{n}\mathbb{P}^{(j)}f$ for a generic function f. Let

$$\mathcal{F}_1=\{[\tau-1\{Y\le q_1(\tau)\}]-\left[\tau'-1\{Y\le q_1(\tau')\}\right]:\tau,\tau'\in\Upsilon,|\tau-\tau'|\le\varepsilon\}$$

which is a VC-class with a fixed VC-index and has an envelope $F_i=2$. In addition,

$$\sigma_n^2=\sup_{f\in\mathcal{F}_1}\overline{\mathbb{P}}f^2\lesssim\sup_{\tilde{\tau}\in\Upsilon}\frac{1}{n}\sum_{i=1}^{n}\left[\varepsilon^2+\frac{f_1(q_1(\tilde{\tau})|\tilde{X}_j)\varepsilon}{f_1(q_1(\tilde{\tau}))}\right]\lesssim\varepsilon\ a.s.$$

Then, by Lemma E.1,

$$\begin{aligned}\mathbb{E}\left[\sup_{\tau,\tau'\in\Upsilon,|\tau-\tau'|\le\varepsilon}\left|\sum_{j=1}^{n}\frac{\tilde{\eta}_{j,1}(\tau)-\tilde{\eta}_{j,1}(\tau')}{\sqrt{n}}\right|\bigg|\{\tilde{X}_j\}_{j=1}^{n}\right]=&\mathbb{E}\left[\|\mathbb{P}_n-\overline{\mathbb{P}}\|_{\mathcal{F}_1}\bigg|\{\tilde{X}_j\}_{j=1}^{n}\right]\\ \lesssim&\sqrt{\varepsilon\log(1/\varepsilon)}+\frac{\log(1/\varepsilon)}{\sqrt{n}}\ a.s.\end{aligned}$$

For any $\delta, \eta > 0$, we can find an $\varepsilon > 0$ such that

$$\begin{aligned}
&\limsup_n \mathbb{P}\left(\sup_{\tau,\tau'\in\Upsilon,|\tau-\tau'|\leq\varepsilon}\left|\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\left(\eta_{i,1}(\tau)-\eta_{i,1}(\tau')\right)\right|\geq\delta\right)\\
=&\limsup_n \mathbb{E}\mathbb{P}\left(\sup_{\tau,\tau'\in\Upsilon,|\tau-\tau'|\leq\varepsilon}\left|\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\left(\eta_{i,1}(\tau)-\eta_{i,1}(\tau')\right)\right|\geq\delta\middle|\{A_i,X_i\}_{i=1}^{2n}\right)\\
\leq&\limsup_n \mathbb{E}\frac{\mathbb{E}\left[\sup_{\tau,\tau'\in\Upsilon,|\tau-\tau'|\leq\varepsilon}\left|\sum_{j=1}^{n}\frac{\tilde{\eta}_{j,1}(\tau)-\tilde{\eta}_{j,1}(\tau')}{\sqrt{n}}\right|\middle|\{\tilde{X}_j\}_{j=1}^{n}\right]}{\delta}\\
\lesssim&\limsup_n \frac{\sqrt{\varepsilon\log(1/\varepsilon)}+\frac{\log(1/\varepsilon)}{\sqrt{n}}}{\delta}\leq\eta,
\end{aligned}$$

where the last inequality holds because $\varepsilon\log(1/\varepsilon)\to 0$ as $\varepsilon\to 0$. This implies $\{\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau) : \tau\in\Upsilon\}$ is stochastically equicontinuous, and hence tight.

In addition, note $\{X_i\}_{i=1}^{2n}$ are i.i.d. and $\{m_{1,\tau}(x) : \tau\in\Upsilon\}$ is Donsker, then $\{\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau}(X_i) : \tau\in\Upsilon\}$ is tight. This leads to the desired result that $\{W_{n,1}(\tau) : \tau\in\Upsilon\}$ is tight. In the same manner, we can show that $\{W_{n,0}(\tau) : \tau\in\Upsilon\}$ is tight, which leads to tightness of $\{W_n(\tau) : \tau\in\Upsilon\}$.

Step (3). Recall $m_{0,\tau}(X_i) = \mathbb{E}(\tau - 1\{Y_i(0)\leq q_0(\tau)\}|X_i)$ and let $\eta_{i,0}(\tau) = \tau - 1\{Y_i(0)\leq q_0(\tau)\} - m_{0,\tau}(X_i)$. Then, based on the previous two steps, we have

$$\sqrt{n}(\hat{\beta}(\tau)-\beta(\tau)) = Q^{-1}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)+\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau}(X_i)\\ \sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}\eta_{i,0}(\tau)+\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{0,\tau}(X_i)\end{pmatrix}+R(\tau) \tag{A.13}$$

where $\sup_{\tau\in\Upsilon}|R(\tau)| = o_p(1)$. In addition, we have already established the stochastic equicontinuity and finite-dimensional convergence of

$$\begin{pmatrix}\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)+\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau}(X_i)\\ \sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}\eta_{i,0}(\tau)+\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{0,\tau}(X_i)\end{pmatrix}.$$

Thus, in order to derive the weak limit of $\sqrt{n}(\hat{\beta}(\tau)-\beta(\tau))$ uniformly over $\tau\in\Upsilon$, it suffices to consider its covariance kernel. First, note that, by construction, $\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau) \perp\!\!\!\perp \sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}\eta_{i,0}(\tau')$ for any $(\tau,\tau')\in\Upsilon$. Second, note that $\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)$ is asymptotically independent of $\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau'}(X_i)$. To see this, let $(s,t)\in\Re^2$, then

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)\leq t, \sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau'}(X_i)\leq s\right)\\
=&\mathbb{E}\left\{\mathbb{P}\left(\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)\leq t\middle|\{A_i,X_i\}_{i=1}^{2n}\right)1\left\{\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau'}(X_i)\leq s\right\}\right\}
\end{aligned}$$

$$
\begin{aligned}
=&\Phi(t/\sqrt{\mathbb{E}\Lambda_\tau(X_i)})\mathbb{P}\left(\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau'}(X_i)\le s\right)\\
&+\mathbb{E}\left\{\left[\mathbb{P}\left(\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)\le t\middle|\{A_i,X_i\}_{i=1}^{2n}\right)-\Phi(t/\sqrt{\mathbb{E}\Lambda_\tau(X_i)})\right]1\left\{\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau'}(X_i)\le s\right\}\right\}\\
\to&\Phi(t/\sqrt{\mathbb{E}\Lambda_\tau(X_i)})\Phi(s/\sqrt{\mathbb{E}m_{1,\tau}^2(X_i)/2}),
\end{aligned}
$$

where the last convergence holds due to the fact that

$$
\mathbb{P}\left(\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau)\le t\middle|\{A_i,X_i\}_{i=1}^{2n}\right)-\Phi(t/\sqrt{\mathbb{E}\Lambda_\tau(X_i)})\xrightarrow{p}0.
$$

We can extend the independence result to multiple τ and τ', implying that the two stochastic processes

$$
\left\{\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau):\tau\in\Upsilon\right\}\quad\text{and}\quad\left\{\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau}(X_i):\tau\in\Upsilon\right\}
$$

are asymptotically independent. For the same reason, we can show

$$
\left\{\left(\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau),\sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}\eta_{i,0}(\tau)\right):\tau\in\Upsilon\right\}\quad\text{and}\quad\left\{\left(\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau}(X_i),\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{0,\tau}(X_i)\right):\tau\in\Upsilon\right\}
$$

are asymptotically independent. Last, it is tedious but straightforward to show that, uniformly over $\tau\in\Upsilon$,

$$
\left(\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta_{i,1}(\tau),\sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}\eta_{i,0}(\tau)\right)\rightsquigarrow\tilde{\mathcal{B}}_1(\tau),
$$

and

$$
\left(\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau}(X_i),\sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{0,\tau}(X_i)\right)\rightsquigarrow\tilde{\mathcal{B}}_2(\tau),
$$

where $\tilde{\mathcal{B}}_1(\tau)$ and $\tilde{\mathcal{B}}_2(\tau)$ are two Gaussian processes with covariance kernels

$$
\tilde{\Sigma}_1(\tau,\tau')=\begin{pmatrix}\mathbb{E}\left[\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X)\right] & 0\\ 0 & \mathbb{E}\left[\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{0,\tau}(X)m_{0,\tau'}(X)\right]\end{pmatrix}
\tag{A.14}
$$

and

$$\tilde{\Sigma}_2(\tau,\tau') = \frac{1}{2}\begin{pmatrix} \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X) & \mathbb{E}m_{1,\tau}(X)m_{0,\tau'}(X) \\ \mathbb{E}m_{1,\tau'}(X)m_{0,\tau}(X) & \mathbb{E}m_{0,\tau}(X)m_{0,\tau'}(X) \end{pmatrix}, \quad \text{respectively.} \tag{A.15}$$

This implies $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \rightsquigarrow \tilde{\mathcal{B}}(\tau)$, where $\tilde{\mathcal{B}}(\tau)$ is a Gaussian process with covariance kernel

$$\tilde{\Sigma}(\tau,\tau') = Q^{-1}(\tau)\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\left(\tilde{\Sigma}_1(\tau,\tau') + \tilde{\Sigma}_2(\tau,\tau')\right)\left[\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} Q^{-1}(\tau')\right]^T.$$

Focusing on the second element of $\hat{\beta}(\tau)$, we have

$$\sqrt{n}(\hat{q}(\tau) - q(\tau)) \rightsquigarrow \mathcal{B}(\tau),$$

where $\mathcal{B}(\tau)$ is a Gaussian process with covariance kernel

$$\begin{aligned}\Sigma(\tau,\tau') =& \frac{\min(\tau,\tau') - \tau\tau' - \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X)}{f_1(q_1(\tau))f_1(q_1(\tau'))} + \frac{\min(\tau,\tau') - \tau\tau' - \mathbb{E}m_{0,\tau}(X)m_{0,\tau'}(X)}{f_0(q_0(\tau))f_0(q_0(\tau'))} \\ &+ \frac{1}{2}\mathbb{E}\left(\frac{m_{1,\tau}(X)}{f_1(q_1(\tau))} - \frac{m_{0,\tau}(X)}{f_0(q_0(\tau))}\right)\left(\frac{m_{1,\tau'}(X)}{f_1(q_1(\tau'))} - \frac{m_{0,\tau'}(X)}{f_0(q_0(\tau'))}\right).\end{aligned}$$

B Proof of Theorem 4.1

Let $u = (u_0, u_1)' \in \Re^2$ and

$$L_n^w(u,\tau) = \sum_{i=1}^{2n} \xi_i\left[\rho_\tau(Y_i - \dot{A}_i'\beta(\tau) - \dot{A}_i'u/\sqrt{n}) - \rho_\tau(Y_i - \dot{A}_i'\beta(\tau))\right].$$

Then, by change of variables we have

$$\sqrt{n}(\hat{\beta}^w(\tau) - \beta(\tau)) = \arg\min_u L_n^w(u,\tau).$$

Notice that $L_n^w(u,\tau)$ is convex in u for each τ and bounded in τ for each u. In the following, we divide the proof into three steps. In Step (1), we show that there exists

$$g_n^w(u,\tau) = -u'W_n^w(\tau) + \frac{u'Q(\tau)u}{2}$$

such that for each u,

$$\sup_{\tau\in\Upsilon} |L_n^w(u,\tau) - g_n^w(u,\tau)| \xrightarrow{p} 0$$

and $Q(\tau)$ is defined in the proof of Theorem 3.1. In Step (2), we show $W_n^w(\tau)$ as a stochastic process over $\tau \in \Upsilon$ is tight. Then by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^w(\tau) - \beta(\tau)) = [Q(\tau)]^{-1} W_n^w(\tau) + r_n(\tau),$$

where $\sup_{\tau \in \Upsilon} ||r_n(\tau)||_2 = o_p(1)$. Last, in Step (3), we establish the weak convergence of

$$\sqrt{n}(\hat{\beta}^w(\tau) - \hat{\beta}(\tau))$$

conditionally on data.

Step (1). Similar to Step (1) in the previous section, we have

$$L_n^w(u,\tau) = -u' W_n^w(\tau) + Q_n^w(u,\tau),$$

where

$$W_n^w(\tau) = \sum_{i=1}^{2n} \frac{\xi_i}{\sqrt{n}} \dot{A}_i \left(\tau - 1\{Y_i \leq \dot{A}_i' \beta(\tau)\}\right)$$

and

$$\begin{aligned}
Q_n^w(u,\tau) =& \sum_{i=1}^{2n} \xi_i \int_0^{\frac{\dot{A}_i' u}{\sqrt{n}}} \left(1\{Y_i - \dot{A}_i'\beta(\tau) \leq v\} - 1\{Y_i - \dot{A}_i'\beta(\tau) \leq 0\}\right) dv \\
=& \sum_{i=1}^{2n} \xi_i A_i \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \leq v\} - 1\{Y_i(1) - q_1(\tau) \leq 0\}\right) dv \\
&+ \sum_{i=1}^{2n} \xi_i (1-A_i) \int_0^{\frac{u_0}{\sqrt{n}}} \left(1\{Y_i(0) - q_0(\tau) \leq v\} - 1\{Y_i(0) - q_0(\tau) \leq 0\}\right) dv \\
\equiv& Q_{n,1}^w(u,\tau) + Q_{n,0}^w(u,\tau).
\end{aligned} \tag{B.1}$$

We first consider $Q_{n,1}^w(u,\tau)$. Note

$$\begin{aligned}
H_n(X_i,\tau) =& \mathbb{E}\xi_i \left(\int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \leq v\} - 1\{Y_i(1) - q_1(\tau) \leq 0\}\right) dv | X_i\right) \\
=& \mathbb{E}\left(\int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \leq v\} - 1\{Y_i(1) - q_1(\tau) \leq 0\}\right) dv | X_i\right).
\end{aligned} \tag{B.2}$$

Then,

$$Q_{n,1}^w(u,\tau) = \sum_{i=1}^{2n} \frac{H_n(X_i,\tau)}{2} + \sum_{i=1}^{2n}\left(A_i - \frac{1}{2}\right) H_n(X_i,\tau) + \sum_{i=1}^{2n} A_i \left[\xi_i \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv - H_n(X_i,\tau)\right]. \tag{B.3}$$

By (A.4), we have, uniformly over $\tau \in \Upsilon$,

$$\sum_{i=1}^{2n} \frac{H_n(X_i,\tau)}{2} \xrightarrow{p} \frac{f_1(q_1(\tau))(u_0+u_1)^2}{2}.$$

In addition, (A.5) implies

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\left(A_i - \frac{1}{2}\right) H_n(X_i,\tau)\right| = o_p(1).$$

Last, Lemma E.2 implies

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n} A_i \left[\xi_i \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv - H_n(X_i,\tau)\right]\right| = o_p(1).$$

Combining the above results, we have

$$\sup_{\tau\in\Upsilon}\left|Q_{n,1}^w(u,\tau) - \frac{f_1(q_1(\tau))(u_0+u_1)^2}{2}\right| = o_p(1). \tag{B.4}$$

By a similar argument, we can show that

$$\sup_{\tau\in\Upsilon}\left|Q_{n,0}^w(u,\tau) - \frac{f_0(q_0(\tau))u_0^2}{2}\right| = o_p(1). \tag{B.5}$$

Combining (B.4) and (B.5), we have

$$Q_n^w(u,\tau) \xrightarrow{p} \frac{u'Q(\tau)u}{2},$$

where $Q(\tau)$ is defined in (A.9). Then,

$$\sup_{\tau\in\Upsilon}|L_n^w(u,\tau) - g_n^w(u,\tau)| = \sup_{\tau\in\Upsilon}\left|Q_n^w(u,\tau) - \frac{u'Q(\tau)u}{2}\right| = o_p(1).$$

Step (2). We have

$$
\begin{aligned}
W_n^w(\tau) =& \sum_{i=1}^{2n} \frac{e_1}{\sqrt{n}} \xi_i A_i \left(\tau - 1\{Y_i(1) \leq q_1(\tau)\}\right) + \sum_{i=1}^{2n} \frac{e_0}{\sqrt{n}} (1-A_i)\xi_i \left(\tau - 1\{Y_i(0) \leq q_0(\tau)\}\right) \\
\equiv & e_1 W_{n,1}^w(\tau) + e_0 W_{n,0}^w(\tau).
\end{aligned} \tag{B.6}
$$

Recall $m_{1,\tau}(X_i) = \mathbb{E}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}|X_i)$, $e_1 = (1,1)^T$, and $e_0 = (1,0)^T$, and denote

$$
\eta_{i,1}^w(\tau) = \xi_i(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) - m_{1,\tau}(X_i).
$$

Then, for $W_{n,1}^w(\tau)$, we have

$$
W_{n,1}^w(\tau) = \sum_{i=1}^{2n} \frac{A_i}{\sqrt{n}} \eta_{i,1}^w(\tau) + \sum_{i=1}^{2n} \frac{1}{2\sqrt{n}} m_{1,\tau}(X_i) + R_1(\tau), \tag{B.7}
$$

where by Lemma E.3,

$$
\sup_{\tau \in \Upsilon} |R_1(\tau)| = \sup_{\tau \in \Upsilon} \left| \sum_{i=1}^{2n} \frac{(A_i - 1/2)}{\sqrt{n}} m_{1,\tau}(X_i) \right| = o_p(1).
$$

The second term on the RHS of (B.7) is stochastically equicontinuous and tight. Next, we focus on the first term. Similar to the argument in Step (2) in the previous section, we have

$$
\sum_{i=1}^{2n} \frac{A_i}{\sqrt{n}} \eta_{i,1}^w(\tau) | \{A_i, X_i\}_{i=1}^{2n} \stackrel{d}{=} \sum_{j=1}^{n} \frac{\tilde{\eta}_{j,1}^w(\tau)}{\sqrt{n}} \Bigg| \{\tilde{X}_j\}_{j=1}^n, \tag{B.8}
$$

where $\tilde{\eta}_{j,1}^w(\tau) = \tilde{\xi}_j(\tau - 1\{\tilde{Y}_j(1) \leq q_1(\tau)\}) - m_{1,\tau}(\tilde{X}_j)$, $(\tilde{Y}_j(1), \tilde{X}_j)$ are as defined before, $\tilde{\xi}_j = \xi_{i_j}$, i_j is the j-th smallest index in the set $\{i \in [2n] : A_i = 1\}$, and given $\{\tilde{X}_j\}_{j=1}^n$, $\{\tilde{\eta}_{j,1}^w(\tau)\}_{j=1}^n$ is a sequence of independent random variables. Further, denote the conditional distribution of $(\tilde{\xi}_j, \tilde{Y}_j(1))$ given $\tilde{X}_j$ as $\mathbb{P}^{(j)}$. Then,

$$
\frac{1}{n} \sum_{j=1}^{n} \mathbb{P}^{(j)} (\tilde{\eta}_{j,1}^w(\tau))^2 = \frac{1}{n} \sum_{j=1}^{n} \left\{ \mathbb{E}\left[(\tilde{\xi}_j^w)^2 (\tau - 1\{\tilde{Y}_j(1) \leq q_1(\tau)\})^2 | \tilde{X}_j \right] - m_{1,\tau}^2(\tilde{X}_j) \right\} \leq \overline{C} < \infty,
$$

for some constant $\overline{C} > 0$. This implies that pointwise in $\tau \in \Upsilon$,

$$
\sum_{i=1}^{2n} \frac{A_i}{\sqrt{n}} \eta_{i,1}^w(\tau) | \{A_i, X_i\}_{i=1}^{2n} \stackrel{d}{=} \sum_{j=1}^{n} \frac{\tilde{\eta}_{j,1}^w(\tau)}{\sqrt{n}} \Bigg| \{\tilde{X}_j\}_{j=1}^n = O_p(1).
$$

In addition, let

$$\mathcal{F}_2 = \{\xi\left[\tau - 1\{Y \le q_1(\tau)\}\right] - \xi\left[\tau' - 1\{Y \le q_1(\tau')\}\right] : \tau, \tau' \in \Upsilon, |\tau - \tau'| \le \varepsilon\}$$

which is a VC-class with a fixed VC-index and has an envelope $F_i = 2\xi_i$. In addition, $||\max_{i\in[n]} F_i||_{\mathbb{P},2} \le C\log(n)$ and

$$\sigma_n^2 = \sup_{f\in\mathcal{F}_2} \overline{\mathbb{P}} f^2 \lesssim \sup_{\tilde{\tau}\in\Upsilon} \frac{1}{n}\sum_{i=1}^n \left[\varepsilon^2 + \frac{f_1(q_1(\tilde{\tau})|\tilde{X}_j)\varepsilon}{f_1(q_1(\tilde{\tau}))}\right] \lesssim \varepsilon \ a.s.$$

Then, by Lemma E.1,

$$\begin{aligned}\mathbb{E}\left[\sup_{\tau,\tau'\in\Upsilon,|\tau-\tau'|\le\varepsilon}\left|\sum_{j=1}^n \frac{\tilde{\eta}^w_{j,1}(\tau) - \tilde{\eta}^w_{j,1}(\tau')}{\sqrt{n}}\right|\Bigg|\{\tilde{X}_j\}_{j=1}^n\right] =& \mathbb{E}\left[||\mathbb{P}_n - \overline{\mathbb{P}}||_{\mathcal{F}_2}\Big|\{\tilde{X}_j\}_{j=1}^n\right] \\ \lesssim& \sqrt{\varepsilon\log(1/\varepsilon)} + \frac{\log(1/\varepsilon)\log(n)}{\sqrt{n}} \ a.s.\end{aligned}$$

The RHS of the above display vanishes as $n \to \infty$ followed by $\varepsilon \to 0$, which implies

$$\sum_{i=1}^{2n} \frac{A_i}{\sqrt{n}}\eta^w_{i,1}(\tau)|\{A_i, X_i\}_{i=1}^{2n} \overset{d}{=} \sum_{j=1}^n \frac{\tilde{\eta}^w_{j,1}(\tau)}{\sqrt{n}}\Bigg|\{\tilde{X}_j\}_{j=1}^n \tag{B.9}$$

is stochastically equicontinuous. Therefore, $\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta^w_{i,1}(\tau)|\{A_i,X_i\}_{i=1}^{2n}$, and hence $W^w_{n,1}(\tau)$ is tight. Similarly, we can show $W^w_{n,0}(\tau)$ is tight.

Step (3). Based on the previous two steps, we have

$$\sqrt{n}(\hat{\beta}^w(\tau) - \beta(\tau)) = Q^{-1}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}\eta^w_{i,1}(\tau) + \sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{1,\tau}(X_i)\\ \sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}\eta^w_{i,0}(\tau) + \sum_{i=1}^{2n}\frac{1}{2\sqrt{n}}m_{0,\tau}(X_i)\end{pmatrix} + R^w(\tau) \tag{B.10}$$

where $\sup_{\tau\in\Upsilon}||R^w(\tau)||_2 = o_p(1)$ and $\sqrt{n}(\hat{\beta}^w(\tau) - \beta(\tau))$ is stochastically equicontinuous. Taking the difference between (A.13) and (B.10), we have

$$\sqrt{n}(\hat{\beta}^w(\tau) - \hat{\beta}(\tau)) = Q^{-1}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}(\xi_i - 1)(\tau - 1\{Y_i(1) \le q_1(\tau)\})\\ \sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}(\xi_i - 1)(\tau - 1\{Y_i(0) \le q_0(\tau)\})\end{pmatrix} + R^*(\tau), \tag{B.11}$$

where $\sup_{\tau\in\Upsilon}|R^*(\tau)| = o_p(1)$. In addition, because both $\sqrt{n}(\hat{\beta}^w(\tau) - \beta(\tau))$ and $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$ are stochastically equicontinuous, so be $\sqrt{n}(\hat{\beta}^w(\tau) - \hat{\beta}(\tau))$. Then by Markov inequality, $\sqrt{n}(\hat{\beta}^w(\tau) - \hat{\beta}(\tau))$ is stochastically equicontinuous conditionally on data as well. In order to derive the limiting distribution of $\sqrt{n}(\hat{\beta}^w(\tau) - \hat{\beta}(\tau))$ conditionally on data, we only need to compute the covariance kernel. Note that

$$\mathbb{E}\left[\begin{pmatrix}\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}(\xi_i-1)(\tau-1\{Y_i(1)\le q_1(\tau)\})\\ \sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}(\xi_i-1)(\tau-1\{Y_i(0)\le q_0(\tau)\})\end{pmatrix}\begin{pmatrix}\sum_{i=1}^{2n}\frac{A_i}{\sqrt{n}}(\xi_i-1)(\tau'-1\{Y_i(1)\le q_1(\tau')\})\\ \sum_{i=1}^{2n}\frac{1-A_i}{\sqrt{n}}(\xi_i-1)(\tau'-1\{Y_i(0)\le q_0(\tau')\})\end{pmatrix}^T\middle|Data\right]$$

$$=\frac{1}{n}\sum_{i=1}^{2n}\begin{pmatrix}A_i(\tau-1\{Y_i(1)\le q_1(\tau)\})(\tau'-1\{Y_i(1)\le q_1(\tau')\}) & 0\\ 0 & (1-A_i)(\tau-1\{Y_i(0)\le q_0(\tau)\})(\tau'-1\{Y_i(0)\le q_0(\tau')\})\end{pmatrix}.$$

For the $(1,1)$ entry, we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{2n}A_i(\tau-1\{Y_i(1)\le q_1(\tau)\})(\tau'-1\{Y_i(1)\le q_1(\tau')\})\\ &=\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau)\eta_{1,i}(\tau')+\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau)m_{1,\tau'}(X_i)+\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau')m_{1,\tau}(X_i)+\frac{1}{n}\sum_{i=1}^{2n}A_im_{1,\tau}(X_i)m_{1,\tau'}(X_i).\end{aligned}$$

Note that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau)\eta_{1,i}(\tau') &\stackrel{d}{=}\frac{1}{n}\sum_{j=1}^{n}\tilde{\eta}_{1,j}(\tau)\tilde{\eta}_{1,j}(\tau')\\ &\xrightarrow{p}\lim_n\frac{1}{n}\sum_{j=1}^{n}(F_1(q_1(\min(\tau,\tau'))|\tilde{X}_j)-F_1(q_1(\tau)|\tilde{X}_j)F_1(q_1(\tau')|\tilde{X}_j))\\ &=\min(\tau,\tau')-\mathbb{E}F_1(q_1(\tau)|X_i)F_1(q_1(\tau')|X_i). \qquad \text{(B.12)}\end{aligned}$$

Lemma E.4 shows

$$\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau)m_{1,\tau'}(X_i)\xrightarrow{p}0$$

and

$$\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau')m_{1,\tau}(X_i)\xrightarrow{p}0.$$

Lemma E.6 implies

$$\frac{1}{n}\sum_{i=1}^{2n}A_im_{1,\tau}(X_i)m_{1,\tau'}(X_i)\xrightarrow{p}\mathbb{E}m_{1,\tau}(X_i)m_{1,\tau'}(X_i).$$

This means

$$\frac{1}{n}\sum_{i=1}^{2n}A_i(\tau-1\{Y_i(1)\le q_1(\tau)\})(\tau'-1\{Y_i(1)\le q_1(\tau')\})\xrightarrow{p}\min(\tau,\tau')-\tau\tau'.$$

For the same reason,

$$\frac{1}{n}\sum_{i=1}^{2n}(1-A_i)(\tau - 1\{Y_i(0)\leq q_0(\tau)\})(\tau' - 1\{Y_i(0)\leq q_0(\tau')\} \xrightarrow{p} \min(\tau,\tau') - \tau\tau'.$$

Then, for the second element $\hat{\beta}_1^w(\tau)$ of $\hat{\beta}^w(\tau)$, conditional on the data, we have

$$\sqrt{n}(\hat{\beta}_1^w(\tau) - \hat{\beta}_1(\tau)) \rightsquigarrow \mathcal{B}^w(\tau),$$

where $\mathcal{B}^w(\tau)$ is a Gaussian process with covariance kernel

$$\Sigma^\dagger(\tau,\tau') = \frac{\min(\tau,\tau') - \tau\tau'}{f_1(q_1(\tau))f_1(q_1(\tau'))} + \frac{\min(\tau,\tau') - \tau\tau'}{f_0(q_0(\tau))f_0(q_0(\tau'))}.$$

C Proof of Theorem 4.2

Let $u \in \Re^2$ and

$$L_n^*(u,\tau) = \sum_{i=1}^{2n}\left[\rho_\tau(Y_i - \dot{A}_i'\beta(\tau) - \dot{A}_i'u/\sqrt{n}) - \rho_\tau(Y_i - \dot{A}_i'\beta(\tau))\right] - u^T\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}S_n^*(\tau).$$

Then,

$$\sqrt{n}\left(\hat{\beta}^*(\tau) - \beta(\tau)\right) = \arg\min_u L_n^*(u,\tau).$$

By the same argument as in the proof of Theorem 3.1, we have

$$L_n^*(u,\tau) = -u^TW_n(\tau) + Q_n(u,\tau) - u^T\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}S_n^*(\tau) = -u^T\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}(S_n(\tau) + S_n^*(\tau)) + Q_n(u,\tau).$$

Further note that $S_n^*(\tau) = \frac{1}{\sqrt{2}}\left(S_{n,1}^*(\tau) + S_{n,2}^*(\tau)\right)$. In the following, we divide the proof into three steps. In Step (1), we derive the weak limit of $S_{n,1}^*(\tau)$ given data. In Step (2), we derive the weak limit of $S_{n,2}^*(\tau)$. In Step (3), we derive the desired result of this theorem.

Step (1). Given the data, $S_{n,1}^*(\tau)$ is a Gaussian process with covariance kernel

$$\tilde{\Sigma}_1^*(\tau,\tau') = \begin{pmatrix}\tilde{\Sigma}_{1,1,1}^*(\tau,\tau') & \tilde{\Sigma}_{1,1,2}^*(\tau,\tau')\\ \tilde{\Sigma}_{1,2,1}^*(\tau,\tau') & \tilde{\Sigma}_{1,2,2}^*(\tau,\tau')\end{pmatrix}$$

where

$$\tilde{\Sigma}^*_{1,1,1}(\tau,\tau') = \frac{1}{n}\sum_{j=1}^{n}\left(\tau - 1\{Y_{(j,1)} \le \hat{q}_1(\tau)\}\right)\left(\tau' - 1\{Y_{(j,1)} \le \hat{q}_1(\tau')\}\right),$$

$$\tilde{\Sigma}^*_{1,1,2}(\tau,\tau') = \frac{1}{n}\sum_{j=1}^{n}\left(\tau - 1\{Y_{(j,1)} \le \hat{q}_1(\tau)\}\right)\left(\tau' - 1\{Y_{(j,0)} \le \hat{q}_0(\tau')\}\right),$$

$$\tilde{\Sigma}^*_{1,2,1}(\tau,\tau') = \frac{1}{n}\sum_{j=1}^{n}\left(\tau - 1\{Y_{(j,0)} \le \hat{q}_0(\tau)\}\right)\left(\tau' - 1\{Y_{(j,1)} \le \hat{q}_1(\tau')\}\right),$$

and

$$\tilde{\Sigma}^*_{1,2,2}(\tau,\tau') = \frac{1}{n}\sum_{j=1}^{n}\left(\tau - 1\{Y_{(j,0)} \le \hat{q}_0(\tau)\}\right)\left(\tau' - 1\{Y_{(j,0)} \le \hat{q}_0(\tau')\}\right).$$

Next, we derive the limit of $\tilde{\Sigma}^*_1(\tau,\tau')$ uniformly over $\tau,\tau' \in \Upsilon$. Recall $m_{1,\tau}(X_i,q) = \mathbb{E}\left(\tau - 1\{Y_i(1) \le q\}|X_i\right)$ and define $\eta_{1,i}(q,\tau) = (\tau - 1\{Y_i(1) \le q\}) - m_{1,\tau}(X_i,q)$. Then

$$\begin{aligned}\tilde{\Sigma}_{1,1,1}(\tau,\tau') =& \frac{1}{n}\sum_{j=1}^{n}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau') + \frac{1}{n}\sum_{j=1}^{n}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)m_{1,\tau'}(X_{(j,1)},\hat{q}_1(\tau'))\\ &+ \frac{1}{n}\sum_{j=1}^{n}\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau')m_{1,\tau}(X_{(j,1)},\hat{q}_1(\tau)) + \frac{1}{n}\sum_{j=1}^{n}m_{1,\tau}(X_{(j,1)},\hat{q}_1(\tau))m_{1,\tau'}(X_{(j,1)},\hat{q}_1(\tau'))\\ =& I(\tau,\tau') + II(\tau,\tau') + III(\tau,\tau') + IV(\tau,\tau'),\end{aligned} \tag{C.1}$$

where we use the fact that $Y_{(j,1)} = Y_{(j,1)}(1)$ and $Y_{(j,0)} = Y_{(j,0)}(0)$. Given $\{A_i, X_i\}_{i=1}^{2n}$, $\{Y_{(j,1)}(1)\}_{j=1}^{n}$ is a sequence of independent random variables with probability measure $\Pi_{j=1}^{n}\mathbb{P}^{(j)}$, where $\mathbb{P}^{(j)}$ is the conditional probability of $Y(1)$ given X evaluated at $X = X_{(j,1)}$. Therefore,

$$I(\tau,\tau') = \overline{\mathbb{P}}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau') + \left(\mathbb{P}_n - \overline{\mathbb{P}}\right)\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau'), \tag{C.2}$$

where $\overline{\mathbb{P}}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau')$ is interpreted as $\overline{\mathbb{P}}\eta_{1,(j,1)}(q,\tau)\eta_{1,(j,1)}(q',\tau')|_{q=\hat{q}_1(\tau),q'=\hat{q}(\tau')}$. In addition, by Theorem 3.1, for any $\varepsilon > 0$, it is possible to find a sufficiently large constant L such that

$$\mathbb{P}(\sup_{\tau\in\Upsilon}|\hat{q}(\tau) - q(\tau)| \le L/\sqrt{n}) \ge 1 - \varepsilon. \tag{C.3}$$

Therefore, we have,

$$
\begin{aligned}
&\overline{\mathbb{P}}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau') \\
=&\frac{1}{n}\sum_{j=1}^{n}\left[F_1(\min(\hat{q}_1(\tau),\hat{q}_1(\tau'))|X_{(j,1)}) - F_1(\hat{q}_1(\tau)|X_{(j,1)})F_1(\hat{q}_1(\tau')|X_{(j,1)})\right] \\
=&\frac{1}{n}\sum_{j=1}^{n}\left[F_1(\min(q_1(\tau),q_1(\tau'))|X_{(j,1)}) - F_1(q_1(\tau)|X_{(j,1)})F_1(q_1(\tau')|X_{(j,1)})\right] + R_I(\tau,\tau') \\
=&\frac{1}{n}\sum_{i=1}^{2n}A_i\left[F_1(\min(q_1(\tau),q_1(\tau'))|X_i) - F_1(q_1(\tau)|X_i)F_1(q_1(\tau')|X_i)\right] + R_I(\tau,\tau') \\
=&\frac{1}{2n}\sum_{i=1}^{2n}\left[F_1(\min(q_1(\tau),q_1(\tau'))|X_i) - F_1(q_1(\tau)|X_i)F_1(q_1(\tau')|X_i)\right] \\
+&\frac{1}{n}\sum_{i=1}^{2n}\left(A_i-\frac{1}{2}\right)\left[F_1(\min(q_1(\tau),q_1(\tau'))|X_i) - F_1(q_1(\tau)|X_i)F_1(q_1(\tau')|X_i)\right] + R_I(\tau,\tau'), \qquad \text{(C.4)}
\end{aligned}
$$

where $\sup_{\tau,\tau'\in\Upsilon}|R_I(\tau,\tau')| \xrightarrow{p} 0$ due to (C.3) and Lipschitz continuity of $F_1(\cdot|X)$.

By the standard uniform convergence theorem (van der Vaart and Wellner (1996, Theorem 2.4.1)), uniformly over $\tau,\tau'\in\Upsilon$,

$$\frac{1}{2n}\sum_{i=1}^{2n}\left[F_1(\min(q_1(\tau),q_1(\tau'))|X_i) - F_1(q_1(\tau)|X_i)F_1(q_1(\tau')|X_i)\right] \xrightarrow{p} \min(\tau,\tau') - \tau\tau' - \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X).$$

By the same argument in Lemma E.3,

$$\sup_{\tau,\tau'\in\Upsilon}\left|\frac{1}{n}\sum_{i=1}^{2n}\left(A_i-\frac{1}{2}\right)\left[F_1(\min(q_1(\tau),q_1(\tau'))|X_i) - F_1(q_1(\tau)|X_i)F_1(q_1(\tau')|X_i)\right]\right| \xrightarrow{p} 0$$

Therefore, uniformly over $\tau,\tau'\in\Upsilon$,

$$\overline{\mathbb{P}}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau') \xrightarrow{p} \min(\tau,\tau') - \tau\tau' - \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X).$$

To deal with the second term in (C.2), first denote

$$\mathcal{F}_3 = \{(\tau - 1\{Y\le q_1(\tau)+v\})\left(\tau' - 1\{Y\le q_1(\tau')+v'\}\right) : \tau,\tau'\in\Upsilon, |v|,|v'|\le L/\sqrt{n}\}.$$

Note $\mathcal{F}_3$ has an envelope $F=1$ and is nested by a VC-class of functions with a fixed VC-index. Then, by Lemma E.1,

$$\mathbb{E}\|\mathbb{P}_n - \overline{\mathbb{P}}\|_{\mathcal{F}_3} \lesssim 1/\sqrt{n}.$$

This implies, with probability greater than $1-\varepsilon$,

$$\sup_{\tau,\tau'\in\Upsilon} | \left(\mathbb{P}_n - \overline{\mathbb{P}}\right) \eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau')| \xrightarrow{p} 0. \tag{C.5}$$

Since ε in (C.3) is arbitrary, we have, uniformly over $\tau,\tau'\in\Upsilon$,

$$I(\tau,\tau') \xrightarrow{p} \min(\tau,\tau') - \tau\tau' - \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X). \tag{C.6}$$

By Lemma E.5, we have

$$\sup_{\tau,\tau'\in\Upsilon} |II(\tau,\tau')| = o_p(1) \quad \text{and} \quad \sup_{\tau,\tau'\in\Upsilon} |III(\tau,\tau')| = o_p(1).$$

For $IV(\tau,\tau')$, we note that

$$\begin{aligned} IV(\tau,\tau') =& \frac{1}{n}\sum_{j=1}^{n} m_{1,\tau}(X_{(j,1)})m_{1,\tau'}(X_{(j,1)}) + R_{IV}(\tau,\tau') \\ =& \frac{1}{n}\sum_{i=1}^{2n} A_i m_{1,\tau}(X_i)m_{1,\tau'}(X_i) + R_{IV}(\tau,\tau') \\ =& \frac{1}{2n}\sum_{i=1}^{2n} m_{1,\tau}(X_i)m_{1,\tau'}(X_i) + \frac{1}{n}\sum_{i=1}^{2n}\left(A_i - \frac{1}{2}\right) m_{1,\tau}(X_i)m_{1,\tau'}(X_i) + R_{IV}(\tau,\tau'). \end{aligned} \tag{C.7}$$

By the standard uniform convergence theorem (van der Vaart and Wellner (1996, Theorem 2.4.1)), uniformly over $\tau,\tau'\in\Upsilon$,

$$\frac{1}{2n}\sum_{i=1}^{2n} m_{1,\tau}(X_i)m_{1,\tau'}(X_i) \xrightarrow{p} \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X).$$

Lemma E.6 further shows

$$\sup_{\tau,\tau'\in\Upsilon} |R_{IV}(\tau,\tau')| = o_p(1) \quad \text{and} \quad \sup_{\tau,\tau'\in\Upsilon} \left| \frac{1}{n}\sum_{i=1}^{2n}\left(A_i - \frac{1}{2}\right) m_{1,\tau}(X_i)m_{1,\tau'}(X_i) \right| = o_p(1).$$

Combining the above results, we have, uniformly over $\tau,\tau'\in\Upsilon$,

$$\tilde{\Sigma}^*_{1,1,1}(\tau,\tau') \xrightarrow{p} \min(\tau,\tau') - \tau\tau'.$$

Now we turn to $\tilde{\Sigma}^*_{1,1,2}(\tau,\tau')$. Recall $m_{0,\tau}(X_i,q) = \mathbb{E}\left(\tau - 1\{Y_i(0)\le q\}|X_i\right)$ and define $\eta_{0,i}(q,\tau) =$

$(\tau - 1\{Y_i(0) \le q\}) - m_{0,\tau}(X_i, q)$. Then,

$$
\begin{aligned}
\tilde{\Sigma}^*_{1,1,2}(\tau,\tau') =& \frac{1}{n}\sum_{j=1}^n \eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{0,(j,0)}(\hat{q}_0(\tau'),\tau') + \frac{1}{n}\sum_{j=1}^n \eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)m_{0,\tau'}(X_{(j,0)},\hat{q}_0(\tau')) \\
&+ \frac{1}{n}\sum_{j=1}^n \eta_{0,(j,0)}(\hat{q}_0(\tau'),\tau')m_{1,\tau}(X_{(j,1)},\hat{q}_1(\tau)) + \frac{1}{n}\sum_{j=1}^n m_{1,\tau}(X_{(j,1)},\hat{q}_1(\tau))m_{0,\tau'}(X_{(j,0)},\hat{q}_0(\tau')) \\
=& \widetilde{I}(\tau,\tau') + \widetilde{II}(\tau,\tau') + \widetilde{III}(\tau,\tau') + \widetilde{IV}(\tau,\tau').
\end{aligned}
$$

We derive the uniform limit for each term on the RHS of the above display. First, note that

$$
\widetilde{I}(\tau,\tau') = \overline{\mathbb{P}}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{0,(j,0)}(\hat{q}_0(\tau'),\tau') + (\mathbb{P}_n - \overline{\mathbb{P}})\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{0,(j,0)}(\hat{q}_0(\tau'),\tau'). \tag{C.8}
$$

Similar to (C.4), we have

$$
\sup_{\tau,\tau'\in\Upsilon} \left|\overline{\mathbb{P}}\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{0,(j,0)}(\hat{q}_0(\tau'),\tau') - \overline{\mathbb{P}}\eta_{1,(j,1)}(q_1(\tau),\tau)\eta_{0,(j,0)}(q_0(\tau'),\tau')\right| \xrightarrow{p} 0.
$$

Furthermore, because $(j,1) \neq (j,0)$, conditionally on $\{A_i, X_i\}_{i=1}^{2n}$, $\eta_{1,(j,1)}(q_1(\tau),\tau) \perp\!\!\!\perp \eta_{1,(j,0)}(q_0(\tau),\tau)$,

$$
\overline{\mathbb{P}}\eta_{1,(j,1)}(q_1(\tau),\tau)\eta_{0,(j,0)}(q_0(\tau'),\tau') = 0.
$$

Similar to (C.5), we have

$$
\sup_{\tau,\tau'\in\Upsilon} \left|(\mathbb{P}_n - \overline{\mathbb{P}})\eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{0,(j,0)}(\hat{q}_0(\tau'),\tau')\right| \xrightarrow{p} 0.
$$

This implies that, uniformly over $\tau,\tau' \in \Upsilon$, $\widetilde{I}(\tau,\tau') \xrightarrow{p} 0$. By the same argument as in the proof of Lemma E.5, we can show that

$$
\sup_{\tau,\tau'\in\Upsilon}\left|\widetilde{II}(\tau,\tau')\right| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\tau,\tau'\in\Upsilon}\left|\widetilde{III}(\tau,\tau')\right| \xrightarrow{p} 0.
$$

Last, by the same argument in the proof of Lemma E.6, we can show that, uniformly over $\tau,\tau' \in \Upsilon$,

$$
\begin{aligned}
\widetilde{IV}(\tau,\tau') &= \frac{1}{n}\sum_{j=1}^n m_{1,\tau}(X_{(j,1)})m_{0,\tau'}(X_{(j,0)}) + o_p(1) \\
&= \frac{1}{n}\sum_{j=1}^n m_{1,\tau}(X_{(j,1)})m_{0,\tau'}(X_{(j,1)}) + \frac{1}{n}\sum_{j=1}^n m_{1,\tau}(X_{(j,1)})[m_{0,\tau'}(X_{(j,0)}) - m_{0,\tau'}(X_{(j,1)})] + o_p(1) \\
&\xrightarrow{p} \mathbb{E}m_{1,\tau}(X)m_{0,\tau'}(X),
\end{aligned}
$$

where the $o_p(1)$ holds uniformly over $\tau, \tau' \in \Upsilon$, and the last line holds because $m_{1,\tau}(x)$ is bounded and $m_{0,\tau}(x)$ is Lipschitz.

Combining the above results, we have uniformly over $\tau, \tau' \in \Upsilon$,

$$\tilde{\Sigma}^*_{1,1,2}(\tau, \tau') \xrightarrow{p} \mathbb{E} m_{1,\tau}(X) m_{0,\tau'}(X).$$

The limits of $\tilde{\Sigma}^*_{1,2,1}$ and $\tilde{\Sigma}^*_{1,2,2}$ can be derived similarly. To sum up, we have established that, uniformly over $\tau, \tau' \in \Upsilon$,

$$\tilde{\Sigma}^*_1(\tau, \tau') \xrightarrow{p} \begin{pmatrix} \min(\tau, \tau') - \tau\tau' & \mathbb{E} m_{1,\tau}(X_i) m_{0,\tau'}(X_i) \\ \mathbb{E} m_{0,\tau}(X_i) m_{1,\tau'}(X_i) & \min(\tau, \tau') - \tau\tau' \end{pmatrix}.$$

Lemma E.7 shows $S^*_{n,1}(\tau)$ is stochastically equicontinuous and tight. This concludes the proof of this step.

Step (2). Given the data, $S^*_{n,2}(\tau)$ is a Gaussian process with covariance kernel

$$\tilde{\Sigma}^*_2(\tau, \tau') = \begin{pmatrix} \tilde{\Sigma}^*_{2,1,1}(\tau, \tau') & \tilde{\Sigma}^*_{2,1,2}(\tau, \tau') \\ \tilde{\Sigma}^*_{2,2,1}(\tau, \tau') & \tilde{\Sigma}^*_{2,2,2}(\tau, \tau') \end{pmatrix}$$

where

$$\begin{aligned} \tilde{\Sigma}^*_{2,1,1}(\tau, \tau') = & \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\left(\tau - 1\{Y_{(k,1)} \leq \hat{q}_1(\tau)\}\right) - \left(\tau - 1\{Y_{(k,3)} \leq \hat{q}_1(\tau)\}\right) \right] \\ & \times \left[\left(\tau' - 1\{Y_{(k,1)} \leq \hat{q}_1(\tau')\}\right) - \left(\tau' - 1\{Y_{(k,3)} \leq \hat{q}_1(\tau')\}\right) \right], \end{aligned}$$

$$\begin{aligned} \tilde{\Sigma}^*_{2,1,2}(\tau, \tau') = & \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\left(\tau - 1\{Y_{(k,1)} \leq \hat{q}_1(\tau)\}\right) - \left(\tau - 1\{Y_{(k,3)} \leq \hat{q}_1(\tau)\}\right) \right] \\ & \times \left[\left(\tau' - 1\{Y_{(k,2)} \leq \hat{q}_0(\tau')\}\right) - \left(\tau' - 1\{Y_{(k,4)} \leq \hat{q}_0(\tau')\}\right) \right], \end{aligned}$$

$$\begin{aligned} \tilde{\Sigma}^*_{2,2,1}(\tau, \tau') = & \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\left(\tau - 1\{Y_{(k,2)} \leq \hat{q}_0(\tau)\}\right) - \left(\tau - 1\{Y_{(k,4)} \leq \hat{q}_0(\tau)\}\right) \right] \\ & \times \left[\left(\tau' - 1\{Y_{(k,1)} \leq \hat{q}_1(\tau')\}\right) - \left(\tau' - 1\{Y_{(k,3)} \leq \hat{q}_1(\tau')\}\right) \right], \end{aligned}$$

and

$$\tilde{\Sigma}^*_{2,2,2}(\tau, \tau') = \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\left(\tau - 1\{Y_{(k,2)} \leq \hat{q}_0(\tau)\}\right) - \left(\tau - 1\{Y_{(k,4)} \leq \hat{q}_0(\tau)\}\right) \right]$$

$$\times \left[\left(\tau' - 1\{Y_{(k,2)} \leq \hat{q}_0(\tau')\} \right) - \left(\tau' - 1\{Y_{(k,4)} \leq \hat{q}_0(\tau')\} \right) \right].$$

In the following, we derive the limit of $\tilde{\Sigma}_2^*(\tau,\tau')$. For $\tilde{\Sigma}_{2,1,1}^*(\tau,\tau')$, we have

$$\begin{aligned}
&\tilde{\Sigma}_{2,1,1}^*(\tau,\tau') \\
&= \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \left[\eta_{1,(k,1)}(\hat{q}_1(\tau),\tau) - \eta_{1,(k,3)}(\hat{q}_1(\tau),\tau)\right]\left[\eta_{1,(k,1)}(\hat{q}_1(\tau'),\tau') - \eta_{1,(k,3)}(\hat{q}_1(\tau'),\tau')\right] \\
&\quad + \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \left[\eta_{1,(k,1)}(\hat{q}_1(\tau),\tau) - \eta_{1,(k,3)}(\hat{q}_1(\tau),\tau)\right]\left[m_{1,\tau'}(X_{(k,1)},\hat{q}_1(\tau')) - m_{1,\tau'}(X_{(k,3)},\hat{q}_1(\tau'))\right] \\
&\quad + \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \left[m_{1,\tau}(X_{(k,1)},\hat{q}_1(\tau)) - m_{1,\tau}(X_{(k,3)},\hat{q}_1(\tau))\right]\left[\eta_{1,(k,1)}(\hat{q}_1(\tau'),\tau') - \eta_{1,(k,3)}(\hat{q}_1(\tau'),\tau')\right] \\
&\quad + \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \left[m_{1,\tau}(X_{(k,1)},\hat{q}_1(\tau)) - m_{1,\tau}(X_{(k,3)},\hat{q}_1(\tau))\right]\left[m_{1,\tau'}(X_{(k,1)},\hat{q}_1(\tau')) - m_{1,\tau'}(X_{(k,3)},\hat{q}_1(\tau'))\right] \\
&\equiv \widehat{I}(\tau,\tau') + \widehat{II}(\tau,\tau') + \widehat{III}(\tau,\tau') + \widehat{IV}(\tau,\tau').
\end{aligned}$$

Also note that

$$\begin{aligned}
&\widehat{I}(\tau,\tau') \\
&= \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \left[\eta_{1,(k,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(k,1)}(\hat{q}_1(\tau'),\tau') + \eta_{1,(k,3)}(\hat{q}_1(\tau),\tau)\eta_{1,(k,3)}(\hat{q}_1(\tau'),\tau')\right] \\
&\quad - \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \eta_{1,(k,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(k,3)}(\hat{q}_1(\tau'),\tau') - \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \eta_{1,(k,1)}(\hat{q}_1(\tau'),\tau')\eta_{1,(k,3)}(\hat{q}_1(\tau),\tau) \\
&= \frac{1}{n}\sum_{j=1}^{n} \eta_{1,(j,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(j,1)}(\hat{q}_1(\tau'),\tau') \\
&\quad - \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \eta_{1,(k,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(k,3)}(\hat{q}_1(\tau'),\tau') - \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \eta_{1,(k,1)}(\hat{q}_1(\tau'),\tau')\eta_{1,(k,3)}(\hat{q}_1(\tau),\tau).
\end{aligned}$$

The first term on the RHS of the above display is just $I(\tau,\tau')$ defined in Step (1), whose limit is established in (C.6). For the second and third terms, we note that $(k,1) \neq (k,3)$, which implies, given $\{X_i, A_i\}_{i=1}^{2n}$, $(\eta_{1,(k,1)}(\hat{q}_1(\tau),\tau), \eta_{1,(k,1)}(\hat{q}_1(\tau'),\tau')) \perp\!\!\!\perp (\eta_{1,(k,3)}(\hat{q}_1(\tau),\tau), \eta_{1,(k,3)}(\hat{q}_1(\tau'),\tau'))$. Then, by the same argument in (C.8) and the discussion below, we have

$$\sup_{\tau,\tau' \in \Upsilon} \left| \frac{1}{n}\sum_{k=1}^{\lfloor n/2 \rfloor} \eta_{1,(k,1)}(\hat{q}_1(\tau),\tau)\eta_{1,(k,3)}(\hat{q}_1(\tau'),\tau') \right| \xrightarrow{p} 0$$

and

$$\sup_{\tau,\tau'\in\Upsilon}\left|\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}\eta_{1,(k,1)}(\hat{q}_1(\tau'),\tau')\eta_{1,(k,3)}(\hat{q}_1(\tau),\tau)\right| \xrightarrow{p} 0.$$

This implies that, uniformly over $\tau,\tau'\in\Upsilon$,

$$\widehat{I}(\tau,\tau') \xrightarrow{p} \min(\tau,\tau') - \tau\tau' - \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X).$$

By the same argument in the proof of Lemma E.5, we have

$$\sup_{\tau,\tau'\in\Upsilon}\left|\widehat{II}(\tau,\tau')\right| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\tau,\tau'\in\Upsilon}\left|\widehat{III}(\tau,\tau')\right| \xrightarrow{p} 0.$$

For $\widehat{IV}(\tau,\tau')$, we note $m_{1,\tau}(x,q)$ is Lipschitz in x by Assumption 2. Therefore, by Assumption 4, we have

$$\sup_{\tau,\tau'\in\Upsilon}\left|\widehat{IV}(\tau,\tau')\right| \lesssim \frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}||X_{(k,1)}-X_{(k,3)}||_2^2 \xrightarrow{p} 0.$$

Combining the above results, we show that, uniformly over $\tau,\tau'\in\Upsilon$,

$$\tilde{\Sigma}^*_{2,1,1}(\tau,\tau') \xrightarrow{p} \min(\tau,\tau') - \tau\tau' - \mathbb{E}m_{1,\tau}(X)m_{1,\tau'}(X).$$

For $\tilde{\Sigma}^*_{2,1,2}(\tau,\tau')$, we have

$$\begin{aligned}
&\tilde{\Sigma}^*_{2,1,1}(\tau,\tau')\\
=&\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}\left[\eta_{1,(k,1)}(\hat{q}_1(\tau),\tau)-\eta_{1,(k,3)}(\hat{q}_1(\tau),\tau)\right]\left[\eta_{0,(k,2)}(\hat{q}_0(\tau'),\tau')-\eta_{0,(k,4)}(\hat{q}_0(\tau'),\tau')\right]\\
&+\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}\left[\eta_{1,(k,1)}(\hat{q}_1(\tau),\tau)-\eta_{1,(k,3)}(\hat{q}_1(\tau),\tau)\right]\left[m_{0,\tau'}(X_{(k,2)},\hat{q}_0(\tau'))-m_{0,\tau'}(X_{(k,4)},\hat{q}_0(\tau'))\right]\\
&+\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}\left[m_{1,\tau}(X_{(k,1)},\hat{q}_1(\tau))-m_{1,\tau}(X_{(k,3)},\hat{q}_1(\tau))\right]\left[\eta_{0,(k,2)}(\hat{q}_0(\tau'),\tau')-\eta_{0,(k,4)}(\hat{q}_0(\tau'),\tau')\right]\\
&+\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}\left[m_{1,\tau}(X_{(k,1)},\hat{q}_1(\tau))-m_{1,\tau}(X_{(k,3)},\hat{q}_1(\tau))\right]\left[m_{0,\tau'}(X_{(k,2)},\hat{q}_0(\tau'))-m_{0,\tau'}(X_{(k,4)},\hat{q}_0(\tau'))\right]\\
\equiv&\overline{I}(\tau,\tau')+\overline{II}(\tau,\tau')+\overline{III}(\tau,\tau')+\overline{IV}(\tau,\tau').
\end{aligned}$$

Because $(k,1),\cdots,(k,4)$ are distinctive,

$$\big(\eta_{1,(k,1)}(q,\tau),\eta_{1,(k,3)}(q,\tau),\eta_{0,(k,2)}(q',\tau),\eta_{0,(k,4)}(q',\tau)\big)$$

are mutually independent conditionally on $\{X_i,A_i\}_{i=1}^{2n}$. Then, by the same arguments as in (C.4) and (C.5), we have

$$\sup_{\tau,\tau'\in\Upsilon}|\overline{I}(\tau,\tau')| \xrightarrow{p} 0.$$

By the same argument as in the proof of Lemma E.5, we also have

$$\sup_{\tau,\tau'\in\Upsilon}|\overline{II}(\tau,\tau')| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\tau,\tau'\in\Upsilon}|\overline{III}(\tau,\tau')| \xrightarrow{p} 0.$$

Last, by Assumption 4, we have

$$\begin{aligned}\sup_{\tau,\tau'\in\Upsilon}|\overline{IV}(\tau,\tau')| \lesssim& \frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}||X_{(k,1)}-X_{(k,3)}||_2||X_{(k,2)}-X_{(k,4)}||_2\\ \lesssim& \frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}||X_{(k,1)}-X_{(k,3)}||_2^2+\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}||X_{(k,2)}-X_{(k,4)}||_2^2 \xrightarrow{p} 0.\end{aligned}$$

Combining the above results, we have

$$\sup_{\tau,\tau'\in\Upsilon}|\tilde{\Sigma}^*_{2,1,2}(\tau,\tau')| \xrightarrow{p} 0.$$

We can derive the limits of $\tilde{\Sigma}^*_{2,2,1}(\tau,\tau')$ and $\tilde{\Sigma}^*_{2,2,2}(\tau,\tau')$ in the same manner. To sum up, uniformly over $\tau,\tau'\in\Upsilon$, we have

$$\tilde{\Sigma}^*_2 \xrightarrow{p} \begin{pmatrix}\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{1,\tau}(X_i)m_{1,\tau'}(X_i) & 0\\ 0 & \min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{0,\tau}(X_i)m_{0,\tau'}(X_i)\end{pmatrix}$$

The stochastic equicontinuity and tightness of $S^*_{n,2}(\tau)$ can be established similarly to $S^*_{n,1}(\tau)$.

Step (3). Because both $S_n(\tau)$ and $S^*_n(\tau)$ are stochastically equicontinuous and tight, we can apply Kato (2009, Theorem 2) and have

$$\sqrt{n}(\hat{\beta}^*(\tau)-\beta(\tau)) = Q^{-1}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}(S_n(\tau)+S^*_n(\tau))+R^*(\tau), \tag{C.9}$$

where $\sup_{\tau\in\Upsilon}||R^*(\tau)||_2 = o_p(1)$. Taking the difference between (C.9) and (A.13), we have

$$\sqrt{n}(\hat{\beta}^*(\tau) - \hat{\beta}(\tau)) = Q^{-1}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix} S_n^*(\tau) + \tilde{R}^*(\tau),$$

where $\sup_{\tau\in\Upsilon}||\tilde{R}^*(\tau)||_2 = o_p(1)$. In addition, given the data, $S_{n,1}^*(\tau)$ and $S_{n,2}^*(\tau)$ are independent. Steps (1) and (2) show that uniformly over $\tau \in \Upsilon$ and conditionally on data, $S_n^*(\tau) = \frac{S_{n,1}^*(\tau)+S_{n,2}^*(\tau)}{\sqrt{2}}$ converges to a Gaussian process with covariance kernel

$$\frac{1}{2}\left[\tilde{\Sigma}_1(\tau,\tau') + \tilde{\Sigma}_2(\tau,\tau')\right],$$

where $\tilde{\Sigma}_1(\tau,\tau')$ and $\tilde{\Sigma}_2(\tau,\tau')$ are defined in (A.14) and (A.15), respectively. The weak limit of $S_n^*(\tau)$ given data coincides with the weak limit of $S_n(\tau)$. This implies, given the data, that

$$\sqrt{n}(\hat{q}^*(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}(\tau),$$

where $\mathcal{B}(\tau)$ is the Gaussian process defined in Theorem 3.1. This concludes the proof.

D Proof of Theorem 4.3

We first focus on $\hat{q}^w_{ipw,1}(\tau)$. Let $u \in \Re$ and

$$\tilde{L}_n^w(u,\tau) = \sum_{i=1}^{2n} \frac{\xi_i A_i}{2\hat{A}_i}\left[\rho_\tau(Y_i - q_1(\tau) - u/\sqrt{n}) - \rho_\tau(Y_i - q_1(\tau))\right].$$

Then, by change of variables, we have

$$\sqrt{n}(\hat{q}^w_{ipw,1}(\tau) - q_1(\tau)) = \arg\min_u \tilde{L}_n^w(u,\tau).$$

Notice that $\tilde{L}_n^w(u,\tau)$ is convex in u for each τ and bounded in τ for each u. In the following, we divide the proof into three steps. In Step (1), we show that there exists

$$\tilde{g}_n^w(u,\tau) = -u'\widetilde{W}^w_{n,1}(\tau) + \frac{f_1(q_1(\tau))u^2}{2}$$

such that for each u,

$$\sup_{\tau\in\Upsilon}|\tilde{L}_n^w(u,\tau) - \tilde{g}_n^w(u,\tau)| \xrightarrow{p} 0.$$

In Step (2), we show $\widetilde{W}^w_{n,1}(\tau)$ as a stochastic process over $\tau \in \Upsilon$ is tight. Then by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{q}^w_{ipw,1}(\tau) - q_1(\tau)) = [f_1(q_1(\tau))]^{-1}\widetilde{W}^w_{n,1}(\tau) + \tilde{r}_{n,1}(\tau),$$

where $\sup_{\tau\in\Upsilon} |\tilde{r}_{n,1}(\tau)| = o_p(1)$. For the same reason, we can show

$$\sqrt{n}(\hat{q}^w_{ipw,0}(\tau) - q_0(\tau)) = [f_0(q_0(\tau))]^{-1}\widetilde{W}^w_{n,0}(\tau) + \tilde{r}_{n,0}(\tau),$$

for some $\widetilde{W}^w_{n,0}(\tau)$ to be specified later and $\sup_{\tau\in\Upsilon} |\tilde{r}_{n,0}(\tau)| = o_p(1)$. Last, in Step (3), we establish the weak convergence of

$$\sqrt{n}(\hat{q}^w_{ipw}(\tau) - \hat{q}(\tau))$$

conditionally on data.

Step (1). Similar to Step (1) in the previous section, we have

$$\tilde{L}^w_n(u,\tau) = -\widetilde{W}^w_{n,1}(\tau)u + \tilde{Q}^w_n(u,\tau),$$

where

$$\widetilde{W}^w_{n,1}(\tau) = \sum_{i=1}^{2n} \frac{\xi_i A_i}{2\sqrt{n}\hat{A}_i}\left(\tau - 1\{Y_i(1) \le q_1(\tau)\}\right),$$

and

$$\begin{aligned}
\tilde{Q}^w_n(u,\tau) =& \sum_{i=1}^{2n} \frac{\xi_i A_i}{2\hat{A}_i}\int_0^{\frac{u}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv \\
=& \sum_{i=1}^{2n} \xi_i A_i \int_0^{\frac{u}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv \\
&+ \sum_{i=1}^{2n} \frac{\xi_i A_i(1/2 - \hat{A}_i)}{\hat{A}_i}\int_0^{\frac{u}{\sqrt{n}}} \left(1\{Y_i(1) - q_1(\tau) \le v\} - 1\{Y_i(1) - q_1(\tau) \le 0\}\right) dv \\
\equiv& \tilde{Q}^w_{n,1}(u,\tau) + \tilde{Q}^w_{n,2}(u,\tau).
\end{aligned} \tag{D.1}$$

Exactly the same as $Q^w_{n,1}(u,\tau)$ in Section B, we have

$$\sup_{\tau\in\Upsilon}\left|\tilde{Q}^w_{n,1}(u,\tau) - \frac{f_1(q_1(\tau))u^2}{2}\right| = o_p(1). \tag{D.2}$$

For $\tilde{Q}^w_{n,2}(u,\tau)$, we have, with probability approaching one,

$$
\begin{aligned}
|\tilde{Q}^w_{n,2}(u,\tau)| &\leq \max_{i\in[2n]}|\hat{A}_i - 1/2|\sum_{i=1}^{2n}\frac{\xi_i}{1/2-\max_{i\in[2n]}|\hat{A}_i-1/2|}1\{|Y_i(1)-q_1(\tau)|\leq u/\sqrt{n}\}\frac{|u|}{\sqrt{n}} \\
&\leq \max_{i\in[2n]}|\hat{A}_i - 1/2|\sum_{i=1}^{2n}4\xi_i 1\{|Y_i(1)-q_1(\tau)|\leq u/\sqrt{n}\}\frac{|u|}{\sqrt{n}}, \qquad \text{(D.3)}
\end{aligned}
$$

where the second inequality follows the fact that, w.p.a.1, $|\hat{A}_i - 1/2| \leq 1/4$ as proved in Lemma E.8. Because $\{\xi_i, Y_i(1)\}_{i\in[2n]}$ are i.i.d., by the usual maximal inequality, we can show that

$$
\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}4\xi_i 1\{|Y_i(1)-q_1(\tau)|\leq u/\sqrt{n}\}\frac{|u|}{\sqrt{n}} - \mathbb{E}\sum_{i=1}^{2n}4\xi_i 1\{|Y_i(1)-q_1(\tau)|\leq u/\sqrt{n}\}\frac{|u|}{\sqrt{n}}\right| = o_p(1). \qquad \text{(D.4)}
$$

In addition,

$$
\mathbb{E}\sum_{i=1}^{2n}4\xi_i 1\{|Y_i(1)-q_1(\tau)|\leq u/\sqrt{n}\}\frac{|u|}{\sqrt{n}} \lesssim \sqrt{n}u\left(F_1(q_1(\tau)+\frac{|u|}{\sqrt{n}}) - F_1(q_1(\tau)-\frac{|u|}{\sqrt{n}})\right) \lesssim u^2. \quad \text{(D.5)}
$$

Combining (D.3)–(D.5) with the fact that $\max_{i\in[2n]}|\hat{A}_i - 1/2| = o_p(1)$ as proved in Lemma E.8, we have

$$
\sup_{\tau\in\Upsilon}|\tilde{Q}^w_{n,2}(u,\tau)| = o_p(1).
$$

This concludes the proof of Step (1).

Step (2). We have

$$
\begin{aligned}
\widetilde{W}^w_{n,1}(\tau) =& \sum_{i=1}^{2n}\frac{\xi_i A_i}{\sqrt{n}}(\tau - 1\{Y_i(1)\leq q_1(\tau)\}) - \sum_{i=1}^{2n}\frac{2\xi_i A_i(\hat{A}_i - 1/2)}{\sqrt{n}}(\tau - 1\{Y_i(1)\leq q_1(\tau)\}) \\
&+ \sum_{i=1}^{2n}\frac{2\xi_i A_i(1/2-\hat{A}_i)^2}{\sqrt{n}\hat{A}_i}(\tau - 1\{Y_i(1)\leq q_1(\tau)\}) \\
\equiv& \widetilde{W}^w_{n,1,1}(\tau) - \widetilde{W}^w_{n,1,2}(\tau) + \widetilde{W}^w_{n,1,3}(\tau). \qquad \text{(D.6)}
\end{aligned}
$$

First, $\widetilde{W}^w_{n,1,1}(\tau)$ is tight following the exact same argument as in Step (2) of Section B. Second, we have

$$
\widetilde{W}^w_{n,1,2}(\tau) = \sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)(\hat{A}_i - 1/2)}{\sqrt{n}} + \sum_{i=1}^{2n}\frac{2\xi_i(A_i-1/2)m_{1,\tau}(X_i)(\hat{A}_i-1/2)}{\sqrt{n}} + \sum_{i=1}^{2n}\frac{2\xi_i A_i\eta_{1,i}(\tau)(\hat{A}_i-1/2)}{\sqrt{n}}
$$

$$\equiv I(\tau) + II(\tau) + III(\tau).$$

Lemma E.9 shows

$$\sup_{\tau\in\Upsilon}\left|I(\tau) - \sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)(A_i-1/2)}{\sqrt{n}}\right| = o_p(1),$$

$$\sup_{\tau\in\Upsilon}|II(\tau)| = o_p(1), \quad \text{and} \quad \sup_{\tau\in\Upsilon}|III(\tau)| = o_p(1).$$

Combining the above results, we have

$$\sup_{\tau\in\Upsilon}\left|\widetilde{W}^w_{n,1,2}(\tau) - \sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)(A_i-1/2)}{\sqrt{n}}\right| = o_p(1). \tag{D.7}$$

Last, we have, w.p.a.1,

$$\begin{aligned}\sup_{\tau\in\Upsilon}|\widetilde{W}^w_{n,1,3}(\tau)| \leq & \sum_{i=1}^{2n}\frac{2\xi_i}{\sqrt{n}(1/2-\max_{i\in[2n]}|1/2-\hat{A}_i|)}(1/2-\hat{A}_i)^2 \\ \lesssim & \frac{4}{\sqrt{n}}\sum_{i=1}\xi_i(1/2-\hat{A}_i)^2 = o_p(1),\end{aligned} \tag{D.8}$$

where the first inequality holds because $\sup_{\tau\in\Upsilon}|\tau - 1\{Y_i(1)\leq q_1(\tau)\}|\leq 1$, the second inequality holds because $\max_i|1/2-\hat{A}_i|\leq 1/4$ w.p.a.1 as proved in Lemma E.8, and the last inequality holds due to Lemma E.8.

Combining (D.6)–(D.8), we have

$$\widetilde{W}^w_{n,1}(\tau) = \sum_{i=1}^{2n}\frac{\xi_i A_i \eta_{1,i}(\tau)}{\sqrt{n}} + \sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)}{2\sqrt{n}} + o_p(1),$$

where the $o_p(1)$ term holds uniformly over $\tau\in\Upsilon$. By (B.9) and the argument above, we can show $\sum_{i=1}^{2n}\frac{\xi_i A_i\eta_{1,i}(\tau)}{\sqrt{n}}$ as a stochastic process over $\tau\in\Upsilon$ is stochastically equicontinuous and tight. Furthermore, $\{\xi_i, X_i\}_{i\in[2n]}$ is a sequence of i.i.d. random variables. Then, by the usual maximal inequality, we can show $\sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)}{2\sqrt{n}}$ as a stochastic process over $\tau\in\Upsilon$ is stochastically equicontinuous and tight. This implies, $\widetilde{W}^w_{n,1}(\tau)$ as a stochastic process over $\tau\in\Upsilon$ is stochastically equicontinuous and tight, and thus, is stochastically equicontinuous conditionally on data by the Markov inequality. Therefore, we have

$$\sqrt{n}(\hat{q}^w_{ipw,1}(\tau) - q_1(\tau)) = \frac{1}{f_1(q_1(\tau))}\left(\sum_{i=1}^{2n}\frac{\xi_i A_i\eta_{1,i}(\tau)}{\sqrt{n}} + \sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)}{2\sqrt{n}}\right) + \tilde{r}_{n,1}(\tau),$$

where $\sup_{\tau\in\Upsilon}|\tilde{r}_{n,1}(\tau)| = o_p(1)$. Similarly, we can show that

$$\sqrt{n}(\hat{q}^w_{ipw,0}(\tau) - q_0(\tau)) = \frac{1}{f_0(q_0(\tau))}\left(\sum_{i=1}^{2n}\frac{\xi_i(1-A_i)\eta_{0,i}(\tau)}{\sqrt{n}} + \sum_{i=1}^{2n}\frac{\xi_i m_{0,\tau}(X_i)}{2\sqrt{n}}\right) + \tilde{r}_{n,0}(\tau),$$

where $\sup_{\tau\in\Upsilon}|\tilde{r}_{n,1}(\tau)| = o_p(1)$.

Step (3). In the proof of Theorem 3.1, we established that

$$\begin{aligned}
&\sqrt{n}(\hat{q}(\tau) - q(\tau))\\
=&\frac{1}{f_1(q_1(\tau))}\left(\sum_{i=1}^{2n}\frac{\xi_i A_i\eta_{1,i}(\tau)}{\sqrt{n}} + \sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)}{2\sqrt{n}}\right)\\
&- \frac{1}{f_0(q_0(\tau))}\left(\sum_{i=1}^{2n}\frac{\xi_i(1-A_i)\eta_{0,i}(\tau)}{\sqrt{n}} + \sum_{i=1}^{2n}\frac{\xi_i m_{0,\tau}(X_i)}{2\sqrt{n}}\right) + r_b(\tau),
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|r_b(\tau)| = o_p(1)$. Then, we have

$$\begin{aligned}
\sqrt{n}(\hat{q}^w_{ipw}(\tau) - \hat{q}(\tau)) =&\frac{1}{f_1(q_1(\tau))}\left(\sum_{i=1}^{2n}\frac{(\xi_i-1)A_i\eta_{1,i}(\tau)}{\sqrt{n}}\right) - \frac{1}{f_0(q_0(\tau))}\left(\sum_{i=1}^{2n}\frac{(\xi_i-1)(1-A_i)\eta_{0,i}(\tau)}{\sqrt{n}}\right)\\
&+\sum_{i=1}^{2n}\frac{(\xi_i-1)}{2\sqrt{n}}\left(\frac{m_{1,\tau}(X_i)}{f_1(q_1(\tau))} - \frac{m_{0,\tau}(X_i)}{f_0(q_0(\tau))}\right) + \tilde{r}_b(r),
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|\tilde{r}_b(\tau)| = o_p(1)$. The conditional stochastic equicontinuity of the first three terms on the RHS of the above display has been established in Step (2). Here, we only need to determine the covariance kernel of $\sqrt{n}(\hat{q}^w_{ipw}(\tau) - \hat{q}(\tau))$ given data. Specifically, the covariance kernel is the limit of the display below:

$$\begin{aligned}
&\frac{1}{f_1(q_1(\tau))f_1(q_1(\tau'))}\sum_{i=1}^{2n}\frac{A_i\eta_{1,i}(\tau)\eta_{1,i}(\tau')}{n} + \frac{1}{f_0(q_0(\tau))f_0(q_0(\tau'))}\sum_{i=1}^{2n}\frac{(1-A_i)\eta_{0,i}(\tau)\eta_{0,i}(\tau')}{n}\\
&+\sum_{i=1}^{2n}\frac{1}{4n}\left(\frac{m_{1,\tau}(X_i)}{f_1(q_1(\tau))} - \frac{m_{0,\tau}(X_i)}{f_0(q_0(\tau))}\right)\left(\frac{m_{1,\tau'}(X_i)}{f_1(q_1(\tau'))} - \frac{m_{0,\tau'}(X_i)}{f_0(q_0(\tau'))}\right)\\
&+\frac{1}{2n}\sum_{i=1}^{2n}\frac{(1-A_i)\eta_{0,i}(\tau)}{f_0(q_0(\tau))}\left(\frac{m_{1,\tau'}(X_i)}{f_1(q_1(\tau'))} - \frac{m_{0,\tau'}(X_i)}{f_0(q_0(\tau'))}\right) + \frac{1}{2n}\sum_{i=1}^{2n}\frac{A_i\eta_{1,i}(\tau)}{f_1(q_1(\tau))}\left(\frac{m_{1,\tau'}(X_i)}{f_1(q_1(\tau'))} - \frac{m_{0,\tau'}(X_i)}{f_0(q_0(\tau'))}\right)\\
&+\frac{1}{2n}\sum_{i=1}^{2n}\frac{(1-A_i)\eta_{0,i}(\tau')}{f_0(q_0(\tau'))}\left(\frac{m_{1,\tau}(X_i)}{f_1(q_1(\tau))} - \frac{m_{0,\tau}(X_i)}{f_0(q_0(\tau))}\right) + \frac{1}{2n}\sum_{i=1}^{2n}\frac{A_i\eta_{1,i}(\tau')}{f_1(q_1(\tau'))}\left(\frac{m_{1,\tau}(X_i)}{f_1(q_1(\tau))} - \frac{m_{0,\tau}(X_i)}{f_0(q_0(\tau))}\right).
\end{aligned} \tag{D.9}$$

Note that (B.12) implies

$$\frac{1}{f_1(q_1(\tau))f_1(q_1(\tau'))}\sum_{i=1}^{2n}\frac{A_i\eta_{1,i}(\tau)\eta_{1,i}(\tau')}{n}\xrightarrow{p}\frac{\min(\tau,\tau')-\mathbb{E}F_1(q_1(\tau)|X_i)F_1(q_1(\tau')|X_i)}{f_1(q_1(\tau))f_1(q_1(\tau'))}$$
$$=\frac{\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{1,\tau}(X_i)m_{1,\tau'}(X_i)}{f_1(q_1(\tau))f_1(q_1(\tau'))}.$$

Similarly,

$$\frac{1}{f_0(q_0(\tau))f_0(q_0(\tau'))}\sum_{i=1}^{2n}\frac{(1-A_i)\eta_{0,i}(\tau)\eta_{0,i}(\tau')}{n}\xrightarrow{p}\frac{\min(\tau,\tau')-\tau\tau'-\mathbb{E}m_{0,\tau}(X_i)m_{0,\tau'}(X_i)}{f_0(q_0(\tau))f_0(q_0(\tau'))}.$$

By the law of large numbers,

$$\sum_{i=1}^{2n}\frac{1}{4n}\left(\frac{m_{1,\tau}(X_i)}{f_1(q_1(\tau))}-\frac{m_{0,\tau}(X_i)}{f_0(q_0(\tau))}\right)\left(\frac{m_{1,\tau'}(X_i)}{f_1(q_1(\tau'))}-\frac{m_{0,\tau'}(X_i)}{f_0(q_0(\tau'))}\right)$$
$$\xrightarrow{p}\frac{1}{2}\mathbb{E}\left(\frac{m_{1,\tau}(X_i)}{f_1(q_1(\tau))}-\frac{m_{0,\tau}(X_i)}{f_0(q_0(\tau))}\right)\left(\frac{m_{1,\tau'}(X_i)}{f_1(q_1(\tau'))}-\frac{m_{0,\tau'}(X_i)}{f_0(q_0(\tau'))}\right).$$

Last, by Lemma E.4, the last four terms on the RHS of (D.9) will vanish. Hence,

$$(\text{D.9})\xrightarrow{p}\Sigma(\tau,\tau'),$$

where $\Sigma(\tau,\tau')$ is defined in Theorem 3.1. This concludes the proof.

E Technical Lemmas

E.1 A Maximal Inequality with i.n.i.d. Random Variables

Although Chernozhukov, Chetverikov, and Kato (2014) derived their Corollary 5.1 for i.i.d. data, the result is still valid when the data are independent but not identically distributed (i.n.i.d.). In this section, we restate their corollary for i.n.i.d. data and provide a brief justification. The proof is due to Chernozhukov et al. (2014). We include this section purely for clarification purpose. Let $\{W_i\}_{i=1}^n$ be a sequence of i.n.i.d. random variables taking values in a measurable space $(S,\mathcal{S})$ with distributions $\Pi_{i=1}^n\mathbb{P}^{(i)}$. Let $\mathcal{F}$ be a generic class of measurable functions $S\mapsto\Re$ with envelope F. Further denote $\overline{\mathbb{P}}f=\frac{1}{n}\sum_{i=1}^n\mathbb{P}^{(i)}f$, $||f||_{\overline{\mathbb{P}},2}=\sqrt{\overline{\mathbb{P}}f^2}$ and $\mathbb{P}_nf$ is the usual empirical process $\mathbb{P}_nf=\frac{1}{n}\sum_{i=1}^n f(W_i)$, $\sigma^2=\sup_{f\in\mathcal{F}}\overline{\mathbb{P}}f^2\leq\overline{\mathbb{P}}F^2$, and $M=\max_{i\in[n]}F(W_i)$.

Lemma E.1. *Suppose $\overline{\mathbb{P}}F^2<\infty$ and there exist constants $a\geq e$ and $v\geq 1$ such that*

$$\sup_Q N(\mathcal{F},e_Q,\varepsilon||F||_{Q,2})\leq\left(\frac{a}{\varepsilon}\right)^v,\quad\forall\varepsilon\in(0,1],\tag{E.1}$$

where $e_Q(f,g) = ||f-g||_{Q,2}$ and the supremum is taken over all finitely discrete probability measures on $(S,\mathcal{S})$. Then,

$$\mathbb{E}||\sqrt{n}(\mathbb{P}_n - \overline{\mathbb{P}})||_{\mathcal{F}} \lesssim \sqrt{v\sigma^2 \log\left(\frac{a||F||_{\overline{\mathbb{P}},2}}{\sigma}\right)} + \frac{v||M||_2}{\sqrt{n}} \log\left(\frac{a||F||_{\overline{\mathbb{P}},2}}{\sigma}\right).$$

The proof of Lemma E.1 is exactly the same as that for Chernozhukov et al. (2014, Corollary 5.1) with $\mathbb{P}$ replaced by $\overline{\mathbb{P}}$. For brevity, we just highlight some key steps below.

Proof. Let $\{\varepsilon_i\}_{i=1}^n$ be a sequence of Rademacher random variables that is independent of $\{W_i\}_{i=1}^n$, $\sigma_n^2 = \sup_{f\in\mathcal{F}} \mathbb{P}_n f^2$, and $Z = \mathbb{E}\left[\left|\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(W_i)\right|\right|_{\mathcal{F}}\right]$. Then, by van der Vaart and Wellner (1996, Lemma 2.3.1) or Ledoux and Talagrand (2013, Lemma 6.3),

$$\mathbb{E}||\sqrt{n}(\mathbb{P}_n - \overline{\mathbb{P}})||_{\mathcal{F}} \leq 2Z.$$

Note Ledoux and Talagrand (2013, Lemma 6.3) only requires $\{W_i\}_{i=1}^n$ to be independent. In addition, let the uniform entropy integral be

$$J(\delta) \equiv J(\delta, \mathcal{F}, F) = \int_0^\delta \sup_Q \sqrt{1 + \log N(\mathcal{F}, e_Q, \varepsilon||F||_{Q,2})}d\varepsilon \tag{E.2}$$

where $e_Q(f,g) = ||f-g||_{Q,2}$ and the supremum is taken over all finitely discrete probability measures on $(S,\mathcal{S})$. Then, we have

$$\begin{aligned} Z =& \mathbb{E}\mathbb{E}\left[\left|\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(W_i)\right|\right|_{\mathcal{F}} | W_1, \cdots, W_n\right] \\ \lesssim& \mathbb{E}\left[||F||_{\mathbb{P}_n,2} J(\sigma_n/||F||_{\mathbb{P}_n,2})\right] \\ \lesssim& ||F||_{\overline{\mathbb{P}},2} J(\sqrt{\mathbb{E}\sigma_n^2}/||F||_{\overline{\mathbb{P}},2}), \end{aligned} \tag{E.3}$$

where the second inequality is due to the Jensen's inequality and the fact that $J(\sqrt{x/y})\sqrt{y}$ is concave in (x,y) as shown by Chernozhukov et al. (2014). To see the first inequality, note that by the Hoeffding's inequality,

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(W_i)\right| \geq t \middle| \{W_i\}_{i=1}^n\right) \lesssim \exp\left(-\frac{t^2/2}{\frac{1}{n}\sum_{i=1}^n f(W_i)^2}\right),$$

which implies the stochastic process $\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(W_i)$ indexed by f is sub-Gaussian conditionally on $\{W_i\}_{i=1}^n$. Then, the first inequality in (E.3) follows van der Vaart and Wellner (1996, Corollary 2.2.8), where we let $\delta = \sigma_n/||F||_{\mathbb{P}_n,2}$ and σ_n can be viewed as the diameter of the class of functions

$\mathcal{F}$. We also note that this is a conditional argument, which is still valid even when $\{W_i\}_{i=1}^n$ is i.n.i.d.

Next, we aim to bound $\mathbb{E}\sigma_n^2$. Recall $\sigma^2 = \sup_{f\in\mathcal{F}} \overline{\mathbb{P}} f^2$. We have, for i.n.i.d. $\{W_i\}_{i=1}^n$,

$$\begin{aligned}
\mathbb{E}\sigma_n^2 &\leq \sigma^2 + \mathbb{E}(||(\mathbb{P}_n - \overline{\mathbb{P}})f^2||_{\mathcal{F}}) \\
&\leq \sigma^2 + 2\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f^2(W_i)\right\|_{\mathcal{F}}\right] \\
&\leq \sigma^2 + 8\mathbb{E}\left[M\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(W_i)\right\|_{\mathcal{F}}\right] \\
&\leq \sigma^2 + 8||M||_{\mathbb{P},2}\{\mathbb{E}[||\mathbb{P}_n\varepsilon_i f(W_i)||_{\mathcal{F}}^2]\}^{1/2} \\
&\leq \sigma^2 + C||M||_{\mathbb{P},2}\{\mathbb{E}[||\mathbb{P}_n\varepsilon_i f(W_i)||_{\mathcal{F}}] + n^{-1}||M||_{\mathbb{P},2}\} \\
&= \sigma^2 + Cn^{-1/2}||M||_{\mathbb{P},2}Z + Cn^{-1}||M||_{\mathbb{P},2}^2,
\end{aligned} \tag{E.4}$$

where the first inequality is due to the triangle inequality, the second inequality is due to Ledoux and Talagrand (2013, Lemma 6.3), the third inequality is due to Ledoux and Talagrand (2013, Theorem 4.12), the fourth inequality is due to the Cauchy-Schwarz inequality, and the fifth inequality is due to Ledoux and Talagrand (2013, Lemma 6.8) with $q = 2$.

Given (E.4), Chernozhukov et al. (2014) then proved the results that, for $\delta = \sigma/||F||_{\overline{\mathbb{P}},2}$,

$$\mathbb{E}[\sqrt{n}||\mathbb{P}_n - \overline{\mathbb{P}}||_{\mathcal{F}}] \lesssim J(\delta, \mathcal{F}, F)||F||_{\overline{\mathbb{P}},2} + \frac{||M||_{\mathbb{P}_s,2}J^2(\delta,\mathcal{F},F)}{\delta^2\sqrt{n}}. \tag{E.5}$$

In this step, they relied on the facts that $J(\delta) = J(\delta, \mathcal{F}, F)$ is concave in δ and $\delta \mapsto J(\delta)/\delta$ is nonincreasing. The desired result is a quick corollary of (E.5) by noticing that, under (E.1),

$$J(\delta) \leq \int_0^\delta \sqrt{1 + \nu\log\left(\frac{a}{\varepsilon}\right)}d\varepsilon \leq 2\sqrt{2\nu}\delta\sqrt{\log\left(\frac{a}{\delta}\right)}. \tag{E.6}$$

□

E.2 Technical Lemmas Used in the Proof of Theorem 3.1

Lemma E.2. *Recall $H_n(X_i, \tau)$ defined in* (A.2)*. Under the assumptions in Theorem 3.1,*

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\left(A_i - \frac{1}{2}\right)H_n(X_i,\tau)\right| = o_p(1),$$

and

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n} A_i\left[\xi_i^*\int_0^{\frac{u_0+u_1}{\sqrt{n}}}\left(1\{Y_i(1)-q_1(\tau)\le v\}-1\{Y_i(1)-q_1(\tau)\le 0\}\right)dv - H_n(X_i,\tau)\right]\right| = o_p(1),$$

where either $\xi_i^*=1$ *or* $\xi_i^*=\xi_i$ *which satisfies Assumption 3.*

Proof. For the first result, we have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\left(A_i-\frac{1}{2}\right)H_n(X_i,\tau)\right| \\
&\le \frac{1}{2}\sum_{j=1}^{n}\sup_{\tau\in\Upsilon}|H_n(X_{\pi(2j-1),\tau})-H_n(X_{\pi(2j),\tau})| \\
&\le \sum_{j=1}^{n}\frac{1}{2}\int_0^{\frac{|u_0+u_1|}{\sqrt{n}}}\sup_{\tau\in\Upsilon}|f_1(q_1(\tau)+\tilde{v}_j|X_{\pi(2j-1)})-f_1(q_1(\tau)+\tilde{v}_j|X_{\pi(2j)})|vdv \\
&\lesssim \sum_{j=1}^{n}\int_0^{\frac{|u_0+u_1|}{\sqrt{n}}}||X_{\pi(2j-1)}-X_{\pi(2j)}||_2 vdv \\
&\lesssim \frac{(u_0+u_1)^2}{n}\sum_{j=1}^{n}||X_{\pi(2j-1)}-X_{\pi(2j)}||_2 \xrightarrow{p} 0,
\end{aligned}$$

where the first inequality is due to the fact that for the j-th pair, $(A_{\pi(2j-1)}-1/2, A_{\pi(2j)}-1/2)$ is either $(1/2,-1/2)$ or $(-1/2,1/2)$, the second inequality is by standard Taylor expansion to the first order where $|\tilde{v}_j|\le(|u_0+u_1|)/\sqrt{n}$, the third inequality is due to Assumption 2, and the last convergence is due to Assumption 1.

Let $(\tilde{\xi}_j^*,\tilde{Y}_j(1),\tilde{X}_j)=(\xi_{i_j}^*,Y_{i_j}(1),X_{i_j})$ where i_j is the j-th smallest index in the set $\{i\in[2n]: A_i=1\}$. Then, similar to (B.8), we have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n} A_i\left[\xi_i^*\int_0^{\frac{u_0+u_1}{\sqrt{n}}}\left(1\{Y_i(1)-q_1(\tau)\le v\}-1\{Y_i(1)-q_1(\tau)\le 0\}\right)dv - H_n(X_i,\tau)\right]\right|\Big|\{A_i,X_i\}_{i=1}^{2n} \\
&\stackrel{d}{=} ||\mathbb{P}_n-\overline{\mathbb{P}}||_{\mathcal{F}_4}|\{\tilde{X}_j\}_{j=1}^n,
\end{aligned}$$

where $\mathcal{F}_4=\{\tilde{\xi}^*\int_0^{(u_0+u_1)/\sqrt{n}}\left(1\{\tilde{Y}(1)\le q_1(\tau)+v\}-1\{\tilde{Y}(1)\le q_1(\tau)\}\right)dv : \tau\in\Upsilon\}$, $\mathbb{P}_n f$ is the usual empirical process, $\overline{\mathbb{P}}f=\frac{1}{n}\sum_{j=1}^n\mathbb{P}^{(j)}f$, and $\mathbb{P}^{(j)}$ denotes the probability measure of $(\tilde{\xi}_j^*,\tilde{Y}_j(1))$ given $\tilde{X}_j$. Note $\mathcal{F}_4$ is a VC-class with a fixed VC index, has an envelope $F_j=(|u_0+u_1|\tilde{\xi}_j^*)/\sqrt{n}$, $M=\max_{j\in[n]}F_j=(|u_0+u_1|\log(n))/\sqrt{n}$, and

$$\sigma^2=\sup_{f\in\mathcal{F}_4}\overline{\mathbb{P}}f^2\le\sup_{\tau\in\Upsilon}\frac{1}{n}\sum_{j=1}^n\left[F_1\left(q_1(\tau)+\frac{|u_0+u_1|}{\sqrt{n}}\middle|\tilde{X}_j\right)-F_1\left(q_1(\tau)-\frac{|u_0+u_1|}{\sqrt{n}}\middle|\tilde{X}_j\right)\right]\frac{u^2}{n}$$

$$\le \frac{1}{n}\sum_{j=1}^{n} C(\tilde{X}_j)\frac{(u_0+u_1)^2}{n^{3/2}}$$
$$= \frac{1}{n}\sum_{i=1}^{2n} A_i C(X_i)\frac{(u_0+u_1)^2}{n^{3/2}}$$
$$\le \left(\frac{1}{n}\sum_{i=1}^{2n} C(X_i)\right)\frac{(u_0+u_1)^2}{n^{3/2}}.$$

As $\left(\frac{1}{n}\sum_{i=1}^{2n} C(X_i)\right) \stackrel{a.s.}{\to} \mathbb{E}2C(X_i)$, we have $\left(\frac{1}{n}\sum_{i=1}^{2n} C(X_i)\right) \le 3\mathbb{E}C(X_i)$ a.s. Given such a sequence $\{X_i\}_{i\ge 1}$, Lemma E.1 implies

$$\mathbb{E}\left[||\mathbb{P}_n - \overline{\mathbb{P}}||_{\mathcal{F}_4}|\{\tilde{X}_j\}_{i=1}^{n}\right] \lesssim \sqrt{\frac{3\mathbb{E}C(X_i)\log(n)}{n^{3/2}}} + \frac{\log^2(n)}{n} = o_{a.s.}(1).$$

This implies

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n} A_i\left[\xi_i^*\int_0^{\frac{u_0+u_1}{\sqrt{n}}}\left(1\{Y_i(1)-q_1(\tau)\le v\} - 1\{Y_i(1)-q_1(\tau)\le 0\}\right)dv - H_n(X_i,\tau)\right]\right| = o_p(1).$$

□

Lemma E.3. *Under the assumptions in Theorem 3.1,*

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{(A_i-1/2)}{\sqrt{n}}m_{1,\tau}(X_i)\right| = o_p(1).$$

Proof. We have

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{(A_i-1/2)}{\sqrt{n}}m_{1,\tau}(X_i)\right| = \sup_{\tau\in\Upsilon}\left|\sum_{j=1}^{n}\frac{1}{2\sqrt{n}}(A_{\pi(2j-1)}-A_{\pi(2j)})(F_1(q_1(\tau)|X_{\pi(2j-1)}) - F_1(q_1(\tau)|X_{\pi(2j)}))\right|.$$

Note that

$$\mathcal{F}_5 = \{F_1(q_1(\tau)|X) - F_1(q_1(\tau)|X') : \tau\in\Upsilon\}$$

is a VC-class with a fixed VC-index and has an envelope $F = 2$. This implies (E.1) holds with some constants $a \ge e$ and $v \ge 1$. Then, as discussed in the (E.6), the uniform entropy integral $J(\delta)$ of $\mathcal{F}_5$ satisfies

$$J(\delta) \le \int_0^\delta \sqrt{1+\nu\log\left(\frac{a}{\varepsilon}\right)}d\varepsilon \le 2\sqrt{2\nu}\delta\sqrt{\log\left(\frac{a}{\delta}\right)}.$$

In addition,

$$\sigma_n^2 = \sup_{\tau\in\Upsilon}\frac{1}{n}\sum_{j=1}^{n}(F_1(q_1(\tau)|X_{\pi(2j-1)}) - F_1(q_1(\tau)|X_{\pi(2j)}))^2 \lesssim \frac{1}{n}\sum_{j=1}^{n}\left\|X_{\pi(2j-1)} - X_{\pi(2j)}\right\|_2^2 \xrightarrow{p} 0.$$

We focus on the set $\mathcal{A}_n = \{\sigma_n^2 \leq \varepsilon\}$ for some arbitrary $\varepsilon > 0$ so that $\mathbb{P}(\mathcal{A}_n) \geq 1-\varepsilon$ for n sufficiently large. Note that $\mathcal{A}_n$ belongs to the sigma field generated by $\{X_i\}_{i=1}^{2n}$. In addition, note that conditional on $\{X_i\}_{i=1}^{2n}$, $\{A_{\pi(2j-1)} - A_{\pi(2j)}\}_{j=1}^{n}$ is a sequence of i.i.d. Rademacher random variables. Then, following the same argument in (E.3)

$$\begin{aligned}
&\mathbb{E}\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{(A_i-1/2)}{\sqrt{n}}m_{1,\tau}(X_i)\right|1\{\mathcal{A}_n\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\left\|\frac{1}{2\sqrt{n}}\sum_{j=1}^{n}(A_{\pi(2j-1)} - A_{\pi(2j)})f(X_{\pi(2j-1)}, X_{\pi(2j)})\right\|_{\mathcal{F}_5}\middle|\{X_i\}_{i=1}^{2n}\right]1\{\mathcal{A}_n\}\right\} \\
&\lesssim \mathbb{E}J(\sigma_n/2)1\{\mathcal{A}_n\} \\
&\lesssim J(\varepsilon/2) \lesssim \sqrt{2\nu}\varepsilon\sqrt{\log\left(\frac{2a}{\varepsilon}\right)},
\end{aligned}$$

where the first inequality is due to van der Vaart and Wellner (1996, Corollary 2.2.8) and the fact that, by the Hoeffding's inequality, for any $f \in \mathcal{F}_5$,

$$\mathbb{P}\left(\left|\sum_{j=1}^{n}(A_{\pi(2j-1)} - A_{\pi(2j)})f(X_{\pi(2j-1)}, X_{\pi(2j)})\right| \geq x\middle|\{X_i\}_{i=1}^{2n}\right) \leq 2\exp\left(-\frac{1}{2}\frac{x^2}{\sum_{j=1}^{n}f^2(X_{\pi(2j-1)}, X_{\pi(2j)})}\right).$$

As $\sqrt{2\nu}\varepsilon\sqrt{\log\left(\frac{2a}{\varepsilon}\right)} \to 0$ as $\varepsilon \to 0$, we derive the desired result by letting $n \to \infty$ followed by $\varepsilon \to 0$.

□

E.3 Technical Lemmas Used in the Proof of Theorem 4.1

Lemma E.4. *Suppose the assumptions in Theorem 4.1 hold, then*

$$\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau)m_{j,\tau'}(X_i) \xrightarrow{p} 0,$$

$$\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau)m_{0,\tau'}(X_i) \xrightarrow{p} 0,$$

$$\frac{1}{n}\sum_{i=1}^{2n}(1-A_i)\eta_{0,i}(\tau)m_{0,\tau'}(X_i) \xrightarrow{p} 0,$$

and

$$\frac{1}{n}\sum_{i=1}^{2n}(1-A_i)\eta_{0,i}(\tau)m_{1,\tau'}(X_i) \xrightarrow{p} 0.$$

Proof. We focus on the first statement. The rest can be proved in the same manner. Based on the notation in Section 4.2, we have

$$\frac{1}{n}\sum_{i=1}^{2n}A_i\eta_{1,i}(\tau)m_{1,\tau'}(X_i) = \frac{1}{n}\sum_{j=1}^{n}\eta_{1,(j,1)}(q_1(\tau),\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')).$$

where $\eta_{1,i}(q,\tau) = (\tau - 1\{Y_i(1) \le q\}) - m_{1,\tau}(X_i,q)$. Then, (E.7) implies the desired result.

□

E.4 Technical Lemmas Used in the Proof of Theorem 4.2

Lemma E.5. *Recall $II(\tau,\tau')$ and $III(\tau,\tau')$ defined in* (C.1)*. Suppose the assumptions in Theorem 3.1 hold, then*

$$\sup_{\tau,\tau'\in\Upsilon}|II(\tau,\tau')| \xrightarrow{p} 0 \quad \textit{and} \quad \sup_{\tau,\tau'\in\Upsilon}|III(\tau,\tau')| \xrightarrow{p} 0.$$

Proof. We focus on bounding $II(\tau,\tau')$. The bound for $III(\tau,\tau')$ can be established similarly. By (C.3), we have, with probability greater than $1-\varepsilon$,

$$|II(\tau,\tau')| \le \sup_{\tau,\tau'\in\Upsilon,|v|,|v'|\le L/\sqrt{n}}\left|\frac{1}{n}\sum_{j=1}^{n}\eta_{1,(j,1)}(q_1(\tau)+v,\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')+v')\right|. \tag{E.7}$$

We aim to bound the RHS. Let $\{\varepsilon_j\}_{j=1}^n$ denote a sequence of i.i.d. Rademacher random variables that is independent of the data. Further denote the class of functions

$$\mathcal{F}_6 = \{\eta_{1,(j,1)}(q_1(\tau)+v,\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')+v') : \tau,\tau'\in\Upsilon, |v|,|v'|\le L/\sqrt{n}\}.$$

Note $\mathcal{F}_6$ has an envelope $F=1$ and is nested by a VC-class of functions with a fixed VC-index. Then,

$$\mathbb{E}\left[\sup_{\tau,\tau'\in\Upsilon,|v|,|v'|\le L/\sqrt{n}}\left|\frac{1}{n}\sum_{j=1}^{n}\eta_{1,(j,1)}(q_1(\tau)+v,\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')+v')\right|\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left\{\mathbb{E}\left[\sup_{\tau,\tau'\in\Upsilon,|v|,|v'|\leq L/\sqrt{n}}\left|\frac{1}{n}\sum_{j=1}^{n}\eta_{1,(j,1)}(q_1(\tau)+v,\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')+v')\right| |\{X_i,A_i\}_{i=1}^{2n}\right]\right\} \\
&\lesssim \mathbb{E}\left\{\mathbb{E}\left[\sup_{\tau,\tau'\in\Upsilon,|v|,|v'|\leq L/\sqrt{n}}\left|\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\eta_{1,(j,1)}(q_1(\tau)+v,\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')+v')\right| |\{X_i,A_i\}_{i=1}^{2n}\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\sup_{\tau,\tau'\in\Upsilon,|v|,|v'|\leq L/\sqrt{n}}\left|\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\eta_{1,(j,1)}(q_1(\tau)+v,\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')+v')\right| |\{X_i,A_i,Y_i(1)\}_{i=1}^{2n}\right]\right\} \\
&\leq \frac{||F||_{\mathbb{P},2}J(\sqrt{\mathbb{E}\sigma_n^2}/||F||_{\mathbb{P},2})}{\sqrt{n}} \lesssim \frac{1}{\sqrt{n}}, \qquad \text{(E.8)}
\end{aligned}
$$

where the first equality is due to the law of iterated expectation, the first inequality is due to Ledoux and Talagrand (2013, Lemma 6.3) and the fact that $\{\eta_{1,(j,1)}(q_1(\tau)+v,\tau)\}_{j=1}^n$ is a sequence of independent and centered random variables given $\{X_i,A_i\}_{i=1}^{2n}$, the second inequality follows the same argument in (E.3) with $F=2$,

$$
\sigma_n^2 = \sup_{\tau,\tau'\in\Upsilon,|v|,|v'|\leq L/\sqrt{n}} \frac{1}{n}\sum_{j=1}^{n}\left[\eta_{1,(j,1)}(q_1(\tau)+v,\tau)m_{1,\tau'}(X_{(j,1)},q_1(\tau')+v')\right]^2 \leq 4,
$$

and $J(\cdot)$ being the uniform entropy integral for the class of functions $\mathcal{F}_6$ defined in (E.2), and the last inequality holds because when $\mathcal{F}_6$ is nested by a VC-class, ε_i is bounded, and thus, has a sub-Gaussian tail, and $\delta = \sqrt{\mathbb{E}\sigma_n^2}/||F||_{\mathbb{P},2} \leq 1$, we have

$$
J(\delta) \lesssim \delta \max(\sqrt{\log(1/\delta)}, 1) \lesssim 1,
$$

as shown in (E.6). This implies, uniformly over $\tau,\tau'\in\Upsilon$,

$$
II(\tau,\tau') \xrightarrow{p} 0.
$$

□

Lemma E.6. *Recall* $R_{IV}(\tau,\tau')$ *defined in* (C.7). *Suppose assumptions in Theorem 3.1 hold, then*

$$
\sup_{\tau,\tau'\in\Upsilon}|R_{IV}(\tau,\tau')| = o_p(1) \quad \textit{and} \quad \sup_{\tau,\tau'\in\Upsilon}\left|\frac{1}{n}\sum_{i=1}^{2n}\left(A_i-\frac{1}{2}\right)m_{1,\tau}(X_i)m_{1,\tau'}(X_i)\right| = o_p(1).
$$

Proof. Note

$$
R_{IV}(\tau,\tau') = \frac{1}{n}\sum_{j=1}^{n}\left[m_{1,\tau}(X_{(j,1)})m_{1,\tau'}(X_{(j,1)}) - m_{1,\tau}(X_{(j,1)},\hat{q}_1(\tau))m_{1,\tau'}(X_{(j,1)},\hat{q}_1(\tau'))\right].
$$

By (C.3) and the fact that $F_1(\cdot|X)$ is Lipschitz continuous, we have

$$\begin{aligned}
&\sup_{\tau,\tau'\in\Upsilon}|R_{IV}(\tau,\tau')| \\
&\leq \sup_{\tau,\tau'\in\Upsilon}\frac{1}{n}\sum_{j=1}^{n}\left|m_{1,\tau}(X_{(j,1)})m_{1,\tau'}(X_{(j,1)}) - m_{1,\tau}(X_{(j,1)},\hat{q}_1(\tau))m_{1,\tau'}(X_{(j,1)},\hat{q}_1(\tau'))\right| \xrightarrow{p} 0.
\end{aligned}$$

By the same argument as in the proof of Lemma E.3, we have

$$\sup_{\tau,\tau'\in\Upsilon}\left|\frac{1}{n}\sum_{i=1}^{2n}\left(A_i - \frac{1}{2}\right)m_{1,\tau}(X_i)m_{1,\tau'}(X_i)\right| \xrightarrow{p} 0.$$

□

Lemma E.7. *Recall $S^*_{n,1}(\tau)$ defined in (4.5). Suppose assumptions in Theorem 3.1 hold. Then, $\{S^*_{n,1}(\tau) : \tau \in \Upsilon\}$ is stochastically equicontinuous and tight.*

Proof. It suffices to show the two marginals of $S^*_{n,1}(\tau)$ are stochastically equicontinuous and tight. We focus on the first marginal

$$\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\eta_j(\tau - 1\{Y_{(j,1)} \leq \hat{q}_1(\tau)\}) : \tau \in \Upsilon\right\}.$$

By (C.3), it suffices to establish the stochastic equicontinuity and tightness of

$$\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\eta_j(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\}) : \tau \in \Upsilon, |v| \leq L\right\}$$

for any fixed L. Let

$$\mathcal{F}_7 = \left\{\begin{matrix}(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\}) - (\tau' - 1\{Y_{(j,1)} \leq q_1(\tau') + v'/\sqrt{n}\}) : \\ \tau,\tau' \in \Upsilon, |v|, |v'| \leq L, |\tau - \tau'| \leq \varepsilon, |v - v'| \leq \varepsilon\end{matrix}\right\},$$

which is nested by a VC-class with envelope 2. Then, by (E.2) and (E.6), the uniform entropy integral $J(\delta)$ of $\mathcal{F}_7$ satisfies

$$J(\delta) \lesssim \delta\max(1, \sqrt{\log(1/\delta)}).$$

By the calculation of $\tilde{\Sigma}^*_{1,1,1}(\tau,\tau')$ (with $\hat{q}_1(\tau)$ replaced by $q_1(\tau) + \frac{v}{\sqrt{n}}$) in Section C, we have, uniformly over $\tau, \tau' \in \Upsilon$, $v, v' \in [-L, L]$,

$$\begin{aligned}\sigma_n^2(\tau,\tau',v,v') =& \frac{1}{n}\sum_{j=1}^n \left[(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\}) - (\tau' - 1\{Y_{(j,1)} \leq q_1(\tau') + v'/\sqrt{n}\})\right]^2 \\ \xrightarrow{p}& \tau(1-\tau) + \tau'(1-\tau') - 2(\min(\tau,\tau') - \tau\tau') = |\tau - \tau'| - (\tau-\tau')^2. \qquad \text{(E.9)}\end{aligned}$$

Let $\mathcal{A}_n(\varepsilon) = 1\{\sup_{\tau,\tau' \in \Upsilon, v, v' \in [-L,L]} |\sigma_n^2(\tau,\tau',v,v') - \left(|\tau-\tau'| - (\tau-\tau')^2\right)| \leq \varepsilon\}$, which will occur with probability approaching one. Also by construction, conditionally on data, $\frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_j(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\})$ is a sub-Gaussian process. Then,

$$\begin{aligned}&\mathbb{E}\left[\sup \frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_j(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\}) - (\tau' - 1\{Y_{(j,1)} \leq q_1(\tau') + v'/\sqrt{n}\})\middle| Data\right] 1\{\mathcal{A}_n(\varepsilon)\} \\ &\lesssim J\left(\frac{\sup \sigma_n(\tau,\tau',v,v')}{2}\right) 1\{\mathcal{A}_n(\varepsilon)\} \\ &\lesssim J(\sqrt{\varepsilon}) \lesssim \sqrt{\varepsilon}\max(1, \sqrt{\log(1/\varepsilon)}),\end{aligned}$$

where the supremum is taken over $\tau, \tau' \in \Upsilon, |v|, |v'| \leq L, |\tau - \tau'| \leq \varepsilon, |v - v'| \leq \varepsilon$, the first inequality is due to (van der Vaart and Wellner, 1996, Corollary 2.2.8), and the second inequality is due to (E.9) and the definition of $\mathcal{A}_n$. Then, for any $t > 0$

$$\begin{aligned}&\mathbb{P}\left(\sup \frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_j\left[(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\}) - (\tau' - 1\{Y \leq q_1(\tau') + v'/\sqrt{n}\})\right] \geq t\right) \\ &\leq \mathbb{P}(\mathcal{A}_n^c(\varepsilon)) + \mathbb{P}\left(\sup \frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_j\left[(\tau - 1\{Y \leq q_1(\tau) + v/\sqrt{n}\}) - (\tau' - 1\{Y \leq q_1(\tau') + v'/\sqrt{n}\})\right] \geq t, \mathcal{A}_n(\varepsilon)\right) \\ &\leq \mathbb{E}\left\{\frac{\mathbb{E}\left[\sup \frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_j(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\}) - (\tau' - 1\{Y_{(j,1)} \leq q_1(\tau') + v'/\sqrt{n}\})\middle| Data\right] 1\{\mathcal{A}_n(\varepsilon)\}}{t}\right\} \\ &+ \mathbb{P}(\mathcal{A}_n^c(\varepsilon)) \\ &\lesssim \mathbb{P}(\mathcal{A}_n^c(\varepsilon)) + \frac{\sqrt{\varepsilon}\max(1, \sqrt{\log(1/\varepsilon)})}{t},\end{aligned}$$

where the supremum is taken over $\tau, \tau' \in \Upsilon, |v|, |v'| \leq L, |\tau - \tau'| \leq \varepsilon, |v - v'| \leq \varepsilon$. Let $n \to \infty$ followed by $\varepsilon \to 0$, we have

$$\lim_{\varepsilon \to 0}\limsup_n \mathbb{P}\left(\sup \frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_j\left[(\tau - 1\{Y_{(j,1)} \leq q_1(\tau) + v/\sqrt{n}\}) - (\tau' - 1\{Y_{(j,1)} \leq q_1(\tau') + v'/\sqrt{n}\})\right] \geq t\right) = 0,$$

which implies $\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^n \eta_j(\tau - 1\{Y_{(j,1)} \leq \hat{q}_1(\tau)\}) : \tau \in \Upsilon\right\}$ is stochastically equicontinuous. In addition, for any fixed τ,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\eta_j(\tau - 1\{Y_{(j,1)} \le \hat{q}_1(\tau)\}) = O_p(1).$$

This implies it is also tight over $\tau \in \Upsilon$. □

E.5 Technical Lemmas Used in the Proof of Theorem 4.3

Lemma E.8. *Suppose the assumptions in Theorem 4.3 hold, then*

$$\max_{i\in[2n]}|\hat{A}_i - 1/2| = o_p(1)$$

and

$$\frac{1}{n}\sum_{i=1}^{2n}\xi_i(\hat{A}_i - 1/2)^2 = o_p(n^{-1/2}).$$

Proof. Let $\theta_0 = (0.5, 0, \cdots, 0)^T$ be a $K \times 1$ vector. Then,

$$\begin{aligned}
||\hat{\theta} - \theta_0||_2 &= \left\| \left[\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)b(X_i)^T\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)(A_i - \frac{1}{2})\right]\right\|_2 \\
&\lesssim \left\|\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)(A_i - \frac{1}{2})\right\|_2 \\
&\lesssim \sqrt{K}\left\|\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)(A_i - \frac{1}{2})\right\|_\infty .
\end{aligned}$$

Next, we aim to bound $\left\|\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)(A_i - \frac{1}{2})\right\|_\infty$. Let $b_k(X)$ be the kth component of $b(X)$. Then,

$$\begin{aligned}
&\max_{k\in[K]}\frac{1}{n}\sum_{j=1}^{n}(\xi_{\pi(2j-1)}b_k(X_{\pi(2j-1)}) - \xi_{\pi(2j)}b_k(X_{\pi(2j)}))^2 \\
&\lesssim \max_{k\in[K]}\frac{1}{n}\sum_{i=1}^{2n}\xi_i^2 b_k^2(X_i) \\
&\lesssim \max_{k\in[K]}\mathbb{E}\xi_i^2 b_k^2(X_i) + \max_{k\in[K]}\left|\frac{1}{n}\sum_{i=1}^{2n}\left[\xi_i^2 b_k^2(X_i) - \mathbb{E}\xi_i^2 b_k^2(X_i)\right]\right|.
\end{aligned}$$

The first term on the RHS of the above display is bounded by $\overline{C}$ based on Assumption 5. Let

$\{\varepsilon_i\}_{i\in[2n]}$ be a sequence of i.i.d. Rademacher random variables. Then,

$$\mathbb{E}\max_{k\in[K]}\left|\frac{1}{n}\sum_{i=1}^{2n}\left[\xi_i^2 b_k^2(X_i)-\mathbb{E}\xi_i^2 b_k^2(X_i)\right]\right| \le 2\mathbb{E}\max_{k\in[K]}\left|\frac{1}{n}\sum_{i=1}^{2n}\varepsilon_i\left[\xi_i^2 b_k^2(X_i)-\mathbb{E}\xi_i^2 b_k^2(X_i)\right]\right|.$$

By Hoeffding's inequality,

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{2n}}\sum_{i=1}^{2n}\varepsilon_i\left[\xi_i^2 b_k^2(X_i)-\mathbb{E}\xi_i^2 b_k^2(X_i)\right]\right|\ge t\,\middle|\,\{\xi_i,X_i\}_{i\in[2n]}\right)\le 2\exp(-\frac{t^2}{2\sigma_k^2}),$$

where $\sigma_k^2=\frac{1}{2n}\sum_{i=1}^{2n}\left[\xi_i^2 b_k^2(X_i)-\mathbb{E}\xi_i^2 b_k^2(X_i)\right]^2$. Then, by van der Vaart and Wellner (1996, Lemmas 2.2.1 and 2.2.2),

$$\mathbb{E}\left[\max_{k\in[K]}\left|\frac{1}{n}\sum_{i=1}^{2n}\varepsilon_i\xi_i^2 b_k^2(X_i)\right|\,\middle|\,\{\xi_i,X_i\}_{i\in[2n]}\right]\lesssim\sqrt{\frac{\log(K)}{n}}\sqrt{\max_{k\in[K]}\sigma_k^2}.$$

Applying expectation on both sides and noticing that the square root function is concave, we have

$$\begin{aligned}\mathbb{E}\max_{k\in[K]}\left|\frac{1}{n}\sum_{i=1}^{2n}\varepsilon_i\xi_i^2 b_k^2(X_i)\right| &\lesssim\sqrt{\frac{\log(K)}{n}}\sqrt{\mathbb{E}\max_{k\in[K]}\sigma_k^2}\\ &\lesssim\sqrt{\frac{\log(K)}{n}}\sqrt{\sum_{k\in[K]}\mathbb{E}\sigma_k^2}\\ &\lesssim\sqrt{\frac{\log(K)}{n}}\zeta(K)\sqrt{K}=o(1).\end{aligned}$$

Therefore,

$$\max_{k\in[K]}\left|\frac{1}{n}\sum_{i=1}^{2n}\left[\xi_i^2 b_k^2(X_i)-\mathbb{E}\xi_i^2 b_k^2(X_i)\right]\right|=o_p(1)$$

and with probability approaching one,

$$\max_{k\in[K]}\frac{1}{n}\sum_{j=1}^{n}(\xi_{\pi(2j-1)}b_k(X_{\pi(2j-1)})-\xi_{\pi(2j)}b_k(X_{\pi(2j)}))^2\le 2\overline{C}.$$

Let $I_n'=\{\max_{k\in[K]}\frac{1}{n}\sum_{j=1}^{n}(\xi_{\pi(2j-1)}b_k(X_{\pi(2j-1)})-\xi_{\pi(2j)}b_k(X_{\pi(2j)}))^2\le 2\overline{C}\}$. For $t=\sqrt{\log(n)\overline{C}}$, we have

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)(A_i-\frac{1}{2})\right\|_\infty\ge t/\sqrt{n},I_n'\right)$$

$$
\begin{aligned}
=&\mathbb{E}\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)(A_i-\frac{1}{2})\right\|_\infty \geq t/\sqrt{n}\middle|\{X_i,\xi_i\}_{i\in[2n]}\right)1\{I_n'\}\\
=&\mathbb{E}\mathbb{P}\left(\left\|\sum_{j=1}^{n}(A_{\pi(2j-1)}-A_{\pi(2j)})(\xi_{\pi(2j-1)}b(X_{\pi(2j-1)})-\xi_{\pi(2j)}b(X_{\pi(2j)}))\right\|_\infty \geq 2t\sqrt{n}\middle|\{X_i,\xi_i\}_{i\in[2n]}\right)1\{I_n'\}\\
\leq&\sum_{k=1}^{K}\mathbb{E}\mathbb{P}\left(\left|\sum_{j=1}^{n}(A_{\pi(2j-1)}-A_{\pi(2j)})(\xi_{\pi(2j-1)}b_k(X_{\pi(2j-1)})-\xi_{\pi(2j)}b_k(X_{\pi(2j)}))\right| \geq 2t\sqrt{n}\middle|\{X_i,\xi_i\}_{i\in[2n]}\right)1\{I_n'\}\\
\leq&\sum_{k=1}^{K}2\mathbb{E}\exp\left(\frac{-2t^2n}{\sum_{j=1}^{n}(\xi_{\pi(2j-1)}b_k(X_{\pi(2j-1)})-\xi_{\pi(2j)}b_k(X_{\pi(2j)}))^2}\right)1\{I_n'\}\\
\leq&2\exp\left(\log(K)-\frac{t^2}{\overline{C}}\right)\to 0,
\end{aligned}
$$

where the second last inequality is due to the Hoeffding's inequality and the fact that given $\{X_i,\xi_i\}_{i\in[2n]}$, $\{A_{\pi(2j-1)}-A_{\pi(2j)}\}_{j\in[n]}$ is i.i.d. sequence of Rademacher random variables.

This implies,

$$
||\hat{\theta}-\theta_0||_2 = O_p(\sqrt{\frac{K\log(n)}{n}}),
$$

and thus

$$
\max_{i\in[2n]}|\hat{A}_i-1/2| = \max_i |b(X_i)'(\hat{\theta}-\theta_0)| = O_p\left(\zeta(K)\sqrt{\frac{K\log(n)}{n}}\right) = o_p(1).
$$

For the second result, we have

$$
\frac{1}{n}\sum_{i=1}^{2n}\xi_i(\hat{A}_i-1/2)^2 \leq \lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^{2n}\xi_i b(X_i)b(X_i)'\right)||\hat{\theta}-\theta_0||_2^2 = O_p\left(\frac{K\log(n)}{n}\right) = o_p(n^{-1/2}),
$$

as $K^2\log^2(n) = o(n)$.

□

Lemma E.9. *Suppose assumptions in Theorem 4.3 hold, then*

$$
\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)(\hat{A}_i-1/2)}{\sqrt{n}} - \sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)(A_i-1/2)}{\sqrt{n}}\right| = o_p(1),
$$

$$
\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{2\xi_i(A_i-1/2)m_{1,\tau}(X_i)(\hat{A}_i-1/2)}{\sqrt{n}}\right| = o_p(1),
$$

and

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{2\xi_i A_i\eta_{1,i}(\tau)(\hat{A}_i-1/2)}{\sqrt{n}}\right| = o_p(1).$$

Proof. For the first result, note $m_{1,\tau}(X_i) = b(X_i)'\gamma_1(\tau)+B_\tau(X_i)$ such that $\sup_{x\in \text{Supp}(X),\tau\in\Upsilon}|B_\tau(x)| = o(1/\sqrt{n})$. Then,

$$\begin{aligned}
&\sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)(\hat{A}_i-1/2)}{\sqrt{n}}\\
&=\sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)b(X_i)'(\hat\theta-\theta_0)}{\sqrt{n}}\\
&=\gamma_1'(\tau)\left[\sum_{i=1}^{2n}\frac{\xi_i b(X_i)b(X_i)'}{\sqrt{n}}\right](\hat\theta-\theta_0)+\sum_{i=1}^{2n}\frac{\xi_i B_\tau(X_i)b(X_i)'(\hat\theta-\theta_0)}{\sqrt{n}}\\
&=\sum_{i=1}^{2n}\frac{\xi_i \gamma_1(\tau)'b(X_i)(A_i-1/2)}{\sqrt{n}}+\sum_{i=1}^{2n}\frac{\xi_i B_\tau(X_i)b(X_i)'(\hat\theta-\theta_0)}{\sqrt{n}}\\
&=\sum_{i=1}^{2n}\frac{\xi_i m_{1,\tau}(X_i)(A_i-1/2)}{\sqrt{n}}-\sum_{i=1}^{2n}\frac{\xi_i B_\tau(X_i)(A_i-1/2)}{\sqrt{n}}+\sum_{i=1}^{2n}\frac{\xi_i B_\tau(X_i)b(X_i)'(\hat\theta-\theta_0)}{\sqrt{n}},
\end{aligned}$$

where the third equality holds because

$$\hat\theta-\theta_0=\left[\sum_{i=1}^{2n}\frac{\xi_i b(X_i)b(X_i)'}{n}\right]^{-1}\left[\sum_{i=1}^{2n}\frac{\xi_i b(X_i)(A_i-1/2)}{n}\right].$$

Furthermore,

$$\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{\xi_i B_\tau(X_i)(A_i-1/2)}{\sqrt{n}}\right|\leq o_p(1)\left(\frac{1}{2n}\sum_{i=1}^{2n}\xi_i\right)=o_p(1)$$

and

$$\begin{aligned}
\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{\xi_i B_\tau(X_i)b(X_i)'(\hat\theta-\theta_0)}{\sqrt{n}}\right| &\leq \sum_{i=1}^{2n}\frac{\xi_i\zeta(K)||\hat\theta-\theta_0||_2}{\sqrt{n}}o_p(1/\sqrt{n})\\
&=\left(\sum_{i=1}^{2n}\frac{\xi_i}{n}\right)o_p\left(\sqrt{\frac{K\zeta^2(K)\log(n)}{n}}\right)=o_p(1).
\end{aligned}$$

This leads to the first result.

For the second result, we have

$$\left|\sum_{i=1}^{2n} \frac{2\xi_i(A_i - 1/2)m_{1,\tau}(X_i)(\hat{A}_i - 1/2)}{\sqrt{n}}\right| \leq \left|\left|\sum_{i=1}^{2n} \frac{2\xi_i(A_i - 1/2)m_{1,\tau}(X_i)b(X_i)}{\sqrt{n}}\right|\right|_2 ||\hat{\theta} - \theta_0||_2.$$

In addition,

$$\begin{aligned}
&\sup_{\tau\in\Upsilon}\left|\left|\sum_{i=1}^{2n} \frac{2\xi_i(A_i - 1/2)m_{1,\tau}(X_i)b(X_i)}{\sqrt{n}}\right|\right|_2 \\
=& \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1} \sum_{i=1}^{2n} \frac{2\xi_i(A_i - 1/2)m_{1,\tau}(X_i)b'(X_i)\rho}{\sqrt{n}} \\
=& \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1} \sum_{j=1}^{n} \frac{(A_{\pi(2j-1)} - A_{\pi(2j)})(\xi_{\pi(2j-1)}m_{1,\tau}(X_{\pi(2j-1)})b'(X_{\pi(2j-1)}) - \xi_{\pi(2j)}m_{1,\tau}(X_{\pi(2j)})b'(X_{\pi(2j)}))\rho}{\sqrt{n}}.
\end{aligned}$$

Conditional on $\{X_i, \xi_i\}_{i\in[2n]}$, $\{(A_{\pi(2j-1)} - A_{\pi(2j)})\}_{j=1}^n$ is a sequence of i.i.d. Rademacher random variables. In addition, let

$$\mathcal{F}_8 = \{(\xi_{\pi(2j-1)}m_{1,\tau}(X_{\pi(2j-1)})b'(X_{\pi(2j-1)}) - \xi_{\pi(2j)}m_{1,\tau}(X_{\pi(2j)})b'(X_{\pi(2j)}))\rho : \tau \in \Upsilon, \rho \in R^K, ||\rho||_2 = 1\}$$

with envelope $F_j = (\xi_{\pi(2j-1)}\zeta(K) + \xi_{\pi(2j)}\zeta(K))$. Then, w.p.a.1,

$$\mathbb{E}\frac{1}{n}\sum_{j=1}^{n} F_j^2 \leq \frac{1}{n}\sum_{i=1}^{2n} \mathbb{E}\xi_i^2\zeta^2(K) \leq \overline{C}\zeta^2(K).$$

In addition, for some constant $c > 0$,

$$\sup_Q N(\mathcal{F}_8, e_Q, \varepsilon||F||_{Q,2}) \leq \left(\frac{a}{\varepsilon}\right)^{cK}, \quad \forall\varepsilon \in (0, 1].$$

Let $\sigma_n^2 = \sup_{f\in\mathcal{F}_8} \mathbb{P}_n f^2$ and $\delta^2 = \frac{\sigma_n^2}{\frac{1}{n}\sum_{j=1}^n F_j^2} \leq 1$. Then, by van der Vaart and Wellner (1996, Corollary 2.2.8), (E.2) and (E.6), we have, w.p.a.1,

$$\begin{aligned}
&\mathbb{E}\mathbb{E}\left[\sup_{\tau\in\Upsilon}\left|\left|\sum_{i=1}^{2n} \frac{2\xi_i(A_i - 1/2)m_{1,\tau}(X_i)b(X_i)}{\sqrt{n}}\right|\right|_2 \middle| \{X_i, \xi_i\}_{i\in[2n]}\right] \\
\lesssim& \mathbb{E}\int_0^{\sigma_n} \sqrt{1 + \log(N(\mathcal{F}_8, e_{\mathbb{P}_n}, \varepsilon))}d\varepsilon \\
\lesssim& \mathbb{E}\sqrt{\frac{1}{n}\sum_{j=1}^{n} F_j^2}\int_0^{\delta} \sqrt{1 + \log\sup_Q N(\mathcal{F}_8, e_Q, \varepsilon||F||_{Q,2})}d\varepsilon
\end{aligned}$$

$$\leq \left(\mathbb{E}\sqrt{\frac{1}{n}\sum_{j=1}^{n}F_j^2}\right)\sqrt{K}J(1)$$

$$\leq \left(\sqrt{\mathbb{E}\frac{1}{n}\sum_{j=1}^{n}F_j^2}\right)\sqrt{K}J(1)$$

$$\lesssim \sqrt{K}\zeta(K).$$

This implies

$$\sup_{\tau\in\Upsilon}\left\|\sum_{i=1}^{2n}\frac{2\xi_i(A_i-1/2)m_{1,\tau}(X_i)b(X_i)}{\sqrt{n}}\right\|_2 = O_p(\sqrt{K}\zeta(K))$$

and

$$\sup_{\tau\in\Upsilon}\left\|\sum_{i=1}^{2n}\frac{2\xi_i(A_i-1/2)m_{1,\tau}(X_i)(\hat{A}_i-1/2)}{\sqrt{n}}\right\|_2 = O_p\left(\sqrt{\frac{K^2\zeta^2(K)\log(n)}{n}}\right) = o_p(1).$$

Last, for the third result, we have

$$\begin{aligned}\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{2\xi_iA_i\eta_{1,i}(\tau)(\hat{A}_i-1/2)}{\sqrt{n}}\right| &\leq \sup_{\tau\in\Upsilon}\left\|\sum_{i=1}^{2n}\frac{2\xi_iA_i\eta_{1,i}(\tau)b(X_i)}{\sqrt{n}}\right\|_2 ||\hat{\theta}-\theta_0||_2 \\ &\leq \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{i=1}^{2n}\frac{2\xi_iA_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\right]||\hat{\theta}-\theta_0||_2. \quad \text{(E.10)}\end{aligned}$$

Let $\{\tilde{\varepsilon}_j\}_{j\in[n]}$ and $\{\varepsilon_i\}_{i\in[2n]}$ be two sequences of i.i.d. Rademacher random variables that are independent of the data. By (A.12), we have

$$\sum_{i=1}^{2n}\frac{2\xi_iA_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\Bigg|\{A_i,X_i\}_{i\in[2n]} \stackrel{d}{=} \sum_{j=1}^{n}\frac{2\tilde{\xi}_j\tilde{\eta}_{1,j}(\tau)b'(\tilde{X}_j)\rho}{\sqrt{n}}\Bigg|\{\tilde{X}_j\}_{j\in[n]},$$

and

$$\sum_{i=1}^{2n}\frac{2\varepsilon_i\xi_iA_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\Bigg|\{A_i,X_i\}_{i\in[2n]} \stackrel{d}{=} \sum_{j=1}^{n}\frac{2\tilde{\varepsilon}_j\tilde{\xi}_j\tilde{\eta}_{1,j}(\tau)b'(\tilde{X}_j)\rho}{\sqrt{n}}\Bigg|\{\tilde{X}_j\}_{j\in[n]},$$

where conditionally on $\{\tilde{X}_j\}_{j\in[n]}$, $\{\tilde{\xi}_j\tilde{\eta}_{1,j}(\tau)\}_{j\in[n]}$ is a sequence of independent random variables. Then, by the same argument as in (E.8), we have

$$\mathbb{E}\sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{i=1}^{2n}\frac{2\xi_iA_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left\{\mathbb{E} \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{i=1}^{2n}\frac{2\xi_i A_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\middle|\{X_i,A_i\}_{i\in[2n]}\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E} \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{j=1}^{n}\frac{2\tilde{\xi}_j\tilde{\eta}_{1,j}(\tau)b'(\tilde{X}_j)\rho}{\sqrt{n}}\middle|\{\tilde{X}_j\}_{j\in[n]}\right]\right\} \\
&\lesssim \mathbb{E}\left\{\mathbb{E} \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{j=1}^{n}\frac{2\tilde{\varepsilon}_j\tilde{\xi}_j\tilde{\eta}_{1,j}(\tau)b'(\tilde{X}_j)\rho}{\sqrt{n}}\middle|\{\tilde{X}_j\}_{j\in[n]}\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E} \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{i=1}^{2n}\frac{2\varepsilon_i\xi_i A_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\middle|\{X_i,A_i\}_{i\in[2n]}\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E} \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{i=1}^{2n}\frac{2\varepsilon_i\xi_i A_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\middle|\{X_i,A_i,Y_i(1)\}_{i\in[2n]}\right]\right\}.
\end{aligned}
$$

Let

$$
\mathcal{F}_9 = \{2\xi_i A_i\eta_{1,i}(\tau)b'(X_i)\rho : \tau\in\Upsilon, \rho\in\Re^K, ||\rho||_2=1\},
$$

with envelope $F_i = 2\xi_i\zeta(K)$. In addition, for some constant $c > 0$,

$$
\sup_Q N(\mathcal{F}_9, e_Q, \varepsilon||F||_{Q,2}) \le \left(\frac{a}{\varepsilon}\right)^{cK}, \quad \forall\varepsilon\in(0,1].
$$

Then, following (E.3) and (E.6), we have

$$
\mathbb{E}\left\{\mathbb{E} \sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\left[\sum_{i=1}^{2n}\frac{2\varepsilon_i\xi_i A_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}}\middle|\{X_i,A_i,Y_i(1)\}_{i\in[2n]}\right]\right\} \lesssim ||F||_{\mathbb{P},2}\sqrt{K}J(1) \lesssim \sqrt{K}\zeta(K).
$$

This implies

$$
\sup_{\tau\in\Upsilon,\rho\in\Re^K,||\rho||_2=1}\sum_{i=1}^{2n}\frac{2\xi_i A_i\eta_{1,i}(\tau)b'(X_i)\rho}{\sqrt{n}} = O_p(\sqrt{K}\zeta(K)).
$$

Then, by (E.10) and Lemma E.8, we have

$$
\sup_{\tau\in\Upsilon}\left|\sum_{i=1}^{2n}\frac{2\xi_i A_i\eta_{1,i}(\tau)(\hat{A}_i-1/2)}{\sqrt{n}}\right| = O_p\left(\sqrt{\frac{K^2\zeta^2(K)\log(n)}{n}}\right) = o_p(1).
$$

□

References

Abadie, A., M. M. Chingos, and M. R. West (2018). Endogenous stratification in randomized experiments. *The Review of Economics and Statistics 100*(4), 567–580.

Angrist, J. D. and V. Lavy (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *The American Economic Review 99*(4), 1384–1414.

Athey, S. and G. Imbens (2017). Chapter 3 - the econometrics of randomized experimentsa. In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Field Experiments*, Volume 1 of *Handbook of Economic Field Experiments*, pp. 73 – 140. North-Holland.

Bai, Y. (2019). Optimality of matched-pair designs in randomized controlled trials. *Available at SSRN 3483834*.

Bai, Y., A. Shaikh, and J. P. Romano (2019). Inference in experiments with matched pairs. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-63).

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015). The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Economics 7*(1), 22–53.

Banerjee, A., D. Karlan, and J. Zinman (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics 7*(1), 1–21.

Beuermann, D. W., J. Cristia, S. Cueto, O. Malamud, and Y. Cruzaguayo (2015). One laptop per child at home: Short-term impacts from a randomized experiment in peru. *American Economic Journal: Applied Economics 7*(2), 53–80.

Bold, T., M. S. Kimenyi, G. Mwabu, A. Nganga, and J. Sandefur (2018). Experimental evidence on scaling up education reforms in kenya. *Journal of Public Economics 168*(12), 1–20.

Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association 113*(524), 1741–1768.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics 10*(4), 1747–1785.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics 6*, 5549–5632.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*(4), 1564–1597.

Crepon, B., F. Devoto, E. Duflo, and W. Pariente (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in morocco. *American Economic Journal: Applied Economics 7*(1), 123–150.

Donald, S. G. and Y.-C. Hsu (2014). Estimation and inference for distribution functions and quantile functions in treatment effect models. *Journal of Econometrics 178*, 383–397.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica 75*(1), 259–276.

Fryer, R. G. (2017). Management and student achievement: Evidence from a randomized field experiment. *National Bureau of Economic Research*.

Fryer, R. G. (2018). The 'pupil' factory: Specialization and the production of human capital in schools. *The American Economic Review 108*(3), 616–656.

Fryer, R. G., T. Devi, and R. Holden (2017). Vertical versus horizontal incentives in education: Evidence from randomized trials. *National Bureau of Economic Research*.

Glewwe, P., A. F. Park, and M. Zhao (2016). A better vision for development: Eyeglasses and academic performance in rural primary schools in china. *Journal of Development Economics 122*(9), 170–182.

Groh, M. and D. J. Mckenzie (2016). Macroinsurance for microenterprises: A randomized experiment in post-revolution egypt. *Journal of Development Economics 118*(1), 1–38.

Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association 112*(517), 446–456.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*(2), 315–331.

Hahn, J., K. Hirano, and D. Karlan (2011). Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics 29*(1), 96–108.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis 100*(8), 1816–1829.

Knight, K. (1998). Limiting distributions for l1 regression estimators under general conditions. *The Annals of Statistics 26*(2), 755–770.

Ledoux, M. and M. Talagrand (2013). *Probability in Banach Spaces: Isoperimetry and Processes.* Springer Science & Business Media.

Tabord-Meehan, M. (2018). Stratification trees for adaptive randomization in randomized controlled trials. *arXiv preprint arXiv:1806.05127*.

van der Vaart, A. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Zhang, Y. and X. Zheng (2020). Quantile treatment effects and bootstrap inference under covariate-adaptive randomization. *Quantitative Economics, forthcoming 11*(3), 957–982.