12-2018

Quantile treatment effects and bootstrap inference under covariate-adaptive randomization

Xin ZHENG
Singapore Management University, xin.zheng.2015@phdecons.smu.edu.sg

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

Quantile Treatment Effects and Bootstrap Inference under Covariate-Adaptive Randomization*

Xin Zheng† Yichong Zhang‡

Abstract

This paper studies the estimation and inference of the quantile treatment effect under covariate-adaptive randomization. We propose three estimation methods: (1) the simple quantile regression (QR), (2) the QR with strata fixed effects, and (3) the inverse propensity score weighted QR. For the three estimators, we derive their asymptotic distributions uniformly over a set of quantile indexes and show that the estimator obtained from inverse propensity score weighted QR weakly dominates the other two in terms of efficiency, for a wide range of randomization schemes. For inference, we show that the weighted bootstrap tends to be conservative for methods (1) and (2) while has asymptotically exact type I error for method (3). We also show that the covariate-adaptive bootstrap inference has the exact asymptotic size for all three methods. We illustrate the finite sample performance of the new estimation and inference methods using both simulated and real datasets.

Keywords: Bootstrap inference, quantile treatment effect

JEL codes: C12, C14

1 Introduction

The randomized control trial (RCT), as pointed out in Angrist and Pischke (2008), is one of the five most common methods (along with instrumental variable regressions, matching estimations, differences-in-differences, and regression continuity designs) for causal inference. Researchers can use it to estimate not only average treatment effects (ATEs) but also quantile treatment effects (QTEs), which capture the heterogeneity of sign and magnitude of treatment effects varying depending on their place in the overall distribution of outcomes. RCTs have been routinely implemented with covariate-adaption, so that individuals are first stratified based on some baseline covariates,

*We are grateful to Jia Li for his valuable comments. Zhang acknowledges the financial support from Singapore Ministry of Education Tier 2 grant under grant no. MOE2018-T2-2-169. Any and all errors are our own.

†Singapore Management University. E-mail address: xin.zheng.2015@phdecons.smu.edu.sg.

‡The corresponding author. Singapore Management University. E-mail address: yczhang@smu.edu.sg.

and then, within each strata, the treatment status is assigned (independent of covariates) to achieve some balance between the sizes of treatment and control groups. See, for example, Imbens and Rubin (2015, Chapter 9) for a textbook treatment of the topic and Duflo, Glennerster, and Kremer (2007) and Bruhn and McKenzie (2009) for two excellent surveys focused on implementing RCTs in development economics. To achieve such balance, treatment status for each individual is assigned sequentially and dependently, which introduces (negative) cross-sectional *dependence*. The standard inference procedures that rely on cross-sectional *independence* are usually conservative and lack power. How to estimate QTEs under covariate-adaptive randomization? What are the asymptotic distributions for the QTE estimators, and how to make proper inference? These questions have yet to be addressed.

We propose three ways to estimate QTEs: (1) the simple QR, (2) the QR with strata fixed effects, (3) the inverse propensity score weighted QR. We establish the weak limits for the three estimators uniformly over a set of quantile indexes and show that the estimator obtained from method (3) weakly dominates the other two in terms of efficiency, for a wide range of randomization schemes. In particular, when strong balance[1] is achieved, the three estimators are asymptotically first-order equivalent. For inference, we show that the weighted bootstrap inference is conservative for methods (1) and (2), but has asymptotically exact size for method (3). In addition, we also study the covariate-adaptive bootstrap which respects the cross-sectional dependence when generating the bootstrap sample. We show that the estimator based on the covariate-adaptive bootstrap sample can mimic that of the original sample in terms of standard error. Thus, the covariate-adaptive bootstrap inference can produce asymptotically exact size for all three estimators.

As originally proposed by Doksum (1974) and Firpo (2007), the QTE, for a fixed percentile, corresponds to the horizontal difference between the marginal distributions of the potential outcomes for treatment and control groups. Our estimators (1) and (3) directly follow those in Doksum (1974) and Firpo (2007), respectively. When estimating ATEs, Bruhn and McKenzie (2009) recommend running a linear regression of outcomes on treatment assignment and indicators for each of the strata. We modify such a regression and incorporate strata fixed effects when estimating QTEs, which leads to our second method.

Under covariate-adaption, Shao, Yu, and Zhong (2010) first point out that the usual two-sample t-test for the ATE is conservative. Then, they propose a covariate-adaptive bootstrap which can produce the correct standard error. Shao and Yu (2013) extend the results to generalized linear models. However, both papers parametrize the (transformed) conditional mean equation by a specific linear model and focus on a specific randomization scheme (covariate-adaptive biased coin method). Ma, Qin, Li, and Hu (2018) derive the theoretical properties of ATE estimators based on general covariate-adjusted randomization under the linear model framework. Bugni, Canay, and Shaikh (2018a) substantially generalize the previous results to a fully nonparametric setting

[1]We will define "strong balance" in Section 2.

with a general class of randomization schemes. However, they mainly focus on the ATE and show that the standard two-sample t-test and the t-test based on the linear regression with strata fixed effects are conservative. Then, they obtain analytical estimators for the correct standard errors and study the validity of permutation tests. More recently, Bugni, Canay, and Shaikh (2018b) study the estimation of ATE with multiple treatments and propose a fully saturated estimator.

Our paper complements the above papers in four aspects. First, we consider the estimation and inference for QTEs, which are functions of quantile index τ. We rely on the empirical processes theories in van der Vaart and Wellner (1996) and Chernozhukov, Chetverikov, and Kato (2014) to obtain uniformly weak convergence of our estimators over a compact set of τ. Based on the uniform convergence, we can construct not only point-wise but also uniform confidence band. Second, we study the asymptotic properties of the inverse propensity score weighted estimator under covariate-adaptive randomization. We show it is weakly more efficient than the other two estimators considered in the paper. Analogously, for estimating the ATE, we show that the inverse propensity score weighted estimate is weakly more efficient than the two estimators considered in Bugni et al. (2018a), and is asymptotically first-order equivalent to the fully saturated estimator proposed by Bugni et al. (2018b). Third, we show that the weighted bootstrap inference ignores the (negative) cross-sectional dependence generated due to the covariate-adaptive randomization and is conservative for estimators (1) and (2). However, estimator (3) is robust to such dependence because the randomization scheme does not affect its influence function. Therefore, the weighted bootstrap inference is valid for estimator (3) under a wide range of randomization schemes. Fourth, we establish that the covariate-adaptive bootstrap has the exact asymptotic size for all three estimation methods paired with a wide range of randomization schemes. Shao et al. (2010) first propose the covariate-adaptive bootstrap and establish its validity for the ATE in a linear regression model under the null hypothesis that the treatment effect is not only zero but also homogeneous.[2] We modify the covariate-adaptive bootstrap and establish its validity for the QTE in the nonparametric setting proposed by Bugni et al. (2018a). In addition, our results rely on neither the null hypothesis nor homogeneity of the treatment effect. Compared with the analytical inference, the two bootstrap inferences for QTEs studied in this paper avoid estimating the infinite-dimensional nuisance parameters such as the densities of the potential outcomes, and thus, the choices of tuning parameters. In addition, unlike the permutation tests studied in Bugni et al. (2018a), the validity of bootstrap inferences does not require either the strong balance condition or studentization.

The rest of the paper is organized as follows. Section 2 describes the model set-up and notation. Sections 3.1, 3.2, and 3.3 discuss the asymptotic properties for estimators (1), (2), and (3), respectively. Sections 4 and 5 investigate the validity of the weighted bootstrap and covariate-adaptive bootstrap inferences, respectively. Section 6 examines the finite-sample performance of the estimation and inference methods. Section 7 applies the new methods to estimate and infer the average

[2]We say the average treatment effect is homogeneous if the conditional average treatment effect given covariates is the same as the unconditional one.

and quantile treatment effects of iron efficiency on educational attainment. Section 8 concludes. An appendix provides proofs for all results.

2 Setup and Notation

First, denote the potential outcomes for treated and control groups as $Y(1)$ and $Y(0)$, respectively. The treatment status is denoted as A, where $A = 1$ means treated and $A = 0$ means untreated. The researcher can only observe $\{Y_i, Z_i, A_i\}_{i=1}^n$ where $Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$, and Z_i is a collection of baseline covariates. Stratum are constructed from Z using a function $S : \text{Supp}(Z) \mapsto \mathcal{S}$, where $\mathcal{S}$ is a finite set. For $1 \leq i \leq n$, Let $S_i = S(Z_i)$ and $p(s) = \mathbb{P}(S_i = s)$. We make the following assumption on the data generating process (DGP) and the treatment assignment rule:

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

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

3. $\left\{\left\{\frac{D_n(s)}{\sqrt{n}}\right\}_{s\in\mathcal{S}} \middle| \{S_i\}_{i=1}^n\right\} \rightsquigarrow N(0, \Sigma_D)$ a.s., where

$$D_n(s) = \sum_{i=1}^n (A_i - \pi)1\{S_i = s\} \quad \text{and} \quad \Sigma_D = \text{diag}\{p(s)\gamma(s) : s \in \mathcal{S}\}$$

with $0 \leq \gamma(s) \leq \pi(1 - \pi)$.

Several remarks are in order. First, Assumptions 1.2 and 1.3 are exactly the same as Bugni et al. (2018a, Assumption 2.2). Assumption 1.1 is also maintained in Bugni et al. (2018a) implicitly. We refer interested readers to Bugni et al. (2018a) for more discussion of these assumptions. Second, note that, in Assumption 1.3, the parameter π is the target proportion of treatment for each strata and $D_n(s)$ measures the imbalance. Bugni et al. (2018b) study the more general case that π can take distinct values for different stratum. Extending the results in this paper to this general set-up is left as an interesting topic for future research. Third, we follow the terminology in Bugni et al. (2018a), which follows Efron (1971) and Hu and Hu (2012), saying a treatment assignment rule achieves strong balance if $\gamma(s) = 0$. Fourth, we do not require that the treatment status is assigned independently. Instead, we only require Assumption 1.3, which is satisfied by several treatment assignment rules such as simple random sampling (SRS), biased-coin design (BCD), adaptive biased-coin design (WEI), and stratified block randomization (SBR). Bugni et al. (2018a, Section 3) provides an excellent summary of these four examples. For completeness, we briefly repeat their descriptions below. Note that both BCD and SBR assignment rules achieve strong balance.

Example 1 (SRS). *Let $\{A_i\}_{i=1}^n$ be drawn independently and independent of $\{S_i\}$ as Bernoulli random variables with success rate π, i.e.,*

$$\mathbb{P}\left(A_k = 1 \middle| \{S_i\}_{i=1}^n, \{A_j\}_{j=1}^k\right) = \mathbb{P}(A_k = 1) = \pi.$$

Then, Assumption 1.3 holds with $\gamma(s) = \pi(1-\pi)$.

Example 2 (WEI). *The design is first proposed by Wei (1978). Let $D_{k-1}(s) = \sum_{i=1}^{k-1}\left(A_i - \frac{1}{2}\right)1\{S_i = s\}$, $n_{k-1}(S_k) = \sum_{i=1}^{k-1} 1\{S_i = S_k\}$, and*

$$\mathbb{P}\left(A_k = 1 \middle| \{S_i\}_{i=1}^k, \{A_i\}_{i=1}^{k-1}\right) = \phi\left(\frac{D_{k-1}(S_k)}{n_{k-1}(S_k)}\right),$$

where $\phi(\cdot) : [-1,1] \mapsto [0,1]$ is a pre-specified non-increasing function satisfying $\phi(-x) = 1 - \phi(x)$. Here, $\frac{D_0(S_1)}{0}$ is understood to be zero. Then, Bugni et al. (2018a) show that Assumption 1.3 holds with $\pi = \frac{1}{2}$ and $\gamma(s) = \frac{1}{4}(1 - 4\phi'(0))^{-1}$.

Example 3 (BCD). *The treatment status is determined sequentially for $1 \leq k \leq n$ as*

$$\mathbb{P}\left(A_k = 1 | \{S_i\}_{i=1}^k, \{A_i\}_{i=1}^{k-1}\right) = \begin{cases} \frac{1}{2} & \text{if } D_{k-1}(S_k) = 0 \\ \lambda & \text{if } D_{k-1}(S_k) < 0 \\ 1-\lambda & \text{if } D_{k-1}(S_k) > 0, \end{cases}$$

where $D_{k-1}(s)$ is defined as above and $\frac{1}{2} < \lambda \leq 1$. Then, Bugni et al. (2018a) show that Assumption 1.3 holds with $\pi = \frac{1}{2}$ and $\gamma(s) = 0$.

Example 4 (SBR). *For each strata, $\lfloor \pi n(s) \rfloor$ units are assigned to treatment and the rest is assigned to control. Then, Bugni et al. (2018a) show that Assumption 1.3 holds with $\gamma(s) = 0$.*

Our parameter of interest is the τ-th QTE defined as

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

where $\tau \in (0,1)$ is a quantile index and $q_j(\tau)$ is the τ-th quantile of random variable $Y(j)$ for $j = 0, 1$. The following regularity conditions are common in the literature of quantile estimations.

Assumption 2. *For $j = 0, 1$, denote $f_j(\cdot)$ and $f_j(\cdot|s)$ as the PDFs of $Y_i(j)$ and $Y_i(j)|S_i = s$, respectively. Then (1) $f_j(q_j(\tau))$ and $f_j(q_j(\tau)|s)$ are bounded and bounded away from zero uniformly over $\tau \in \Upsilon$ and $s \in \mathcal{S}$, where Υ is a compact subset of $(0,1)$; (2) $f_j(\cdot)$ and $f_j(\cdot|s)$ are Lipschitz over $\{q_j(\tau) : \tau \in \Upsilon\}$.*

3 Estimation

3.1 Simple Quantile Regression

In this section, we propose to estimate $q(\tau)$ by a QR of Y_i on A_i. Denote $\beta(\tau) = (\beta_0(\tau), \beta_1(\tau))'$, $\beta_0(\tau) = q_0(\tau)$, and $\beta_1(\tau) = q(\tau)$. We estimate $\beta(\tau)$ by $\hat{\beta}$, where

$$\hat{\beta}(\tau) = \operatorname*{arg\,min}_{b=(b_0,b_1)'\in\Re^2} \sum_{i=1}^{n} \rho_\tau\left(Y_i - \dot{A}_i' b\right),$$

$\dot{A}_i = (1, A_i)'$, and $\rho_\tau(u) = u(\tau - 1\{u \leq 0\})$ is the standard check function.

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

$$\sqrt{n}\left(\hat{\beta}_1(\tau) - q(\tau)\right) \rightsquigarrow \mathcal{B}_{sqr}(\tau),$$

where $\mathcal{B}_{sqr}(\cdot)$ is a Gaussian process with covariance kernel $\Sigma_{sqr}(\cdot,\cdot)$. The expression for $\Sigma_{sqr}(\cdot,\cdot)$ can be found in the Appendix.

In particular, the asymptotic variance for $\sqrt{n}\left(\hat{\beta}_1(\tau) - \beta_1(\tau)\right)$ is $\xi_Y^2(\pi,\tau) + \xi_A^2(\pi,\tau) + \xi_S^2(\tau)$, where

$$\zeta_Y^2(\pi,\tau) = \frac{\tau(1-\tau) - \mathbb{E}m_1^2(S,\tau)}{\pi f_1^2(q_1(\tau))} + \frac{\tau(1-\tau) - \mathbb{E}m_0^2(S,\tau)}{(1-\pi) f_0^2(q_0(\tau))},$$

$$\zeta_A^2(\pi,\tau) = \mathbb{E}\gamma(S)\left(\frac{m_1(S,\tau)}{\pi f_1(q_1(\tau))} + \frac{m_0(S,\tau)}{(1-\pi) f_0(q_0(\tau))}\right)^2,$$

$$\zeta_S^2(\tau) = \mathbb{E}\left(\frac{m_1(S,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S,\tau)}{f_0(q_0(\tau))}\right)^2,$$

and $m_j(s,\tau) = \mathbb{E}(\tau - 1\{Y(j) \leq q_j(\tau)\}|S = s)$. Note that, when the treatment assignment rule achieves strong balance, $\zeta_A^2(\pi,\tau) = 0$.

3.2 Quantile Regression with Strata Fixed Effects

The strata fixed effects estimator for the ATE is obtained by a linear regression of outcome Y_i on the treatment status A_i, controlling for strata dummies $\{1\{S_i = s\}_{s\in\mathcal{S}}\}$. Bugni et al. (2018a) point out that, due to the Frisch–Waugh–Lovell theorem, this estimator is equal to the linear coefficient in the regression of Y_i on $\tilde{A}_i$, in which $\tilde{A}_i$ is the projection of A_i on the strata dummies. Unlike the expectation, the quantile operator is nonlinear. Therefore, we cannot consistently estimate QTEs by a linear QR of Y_i on A_i and strata dummies. Instead, based on the equivalence relationship, we propose to run the QR of Y_i on $\tilde{A}_i$. Formally, let $\tilde{A}_i = A_i - \hat{\pi}(S_i)$ and $\dot{\tilde{A}}_i = (1, \tilde{A}_i)'$, where $\hat{\pi}(s) = n_1(s)/n(s)$, $n_1(s) = \sum_{i=1}^n A_i 1\{S_i = s\}$, and $n(s) = \sum_{i=1}^n 1\{S_i = s\}$. Then, the strata fixed effects estimator for the QTE is $\hat{\beta}_{sfe,1}(\tau)$, where

$$\hat{\beta}_{sfe}(\tau) \equiv \left(\hat{\beta}_{sfe,0}(\tau), \hat{\beta}_{sfe,1}(\tau)\right)' = \underset{b=(b_0,b_1)'\in\Re^2}{\arg\min} \sum_{i=1}^n \rho_\tau\left(Y_i - \dot{\tilde{A}}_i' b\right).$$

Theorem 3.2. *If Assumptions 1 and 2 hold, then, uniformly over $\tau \in \Upsilon$,*

$$\sqrt{n}\left(\hat{\beta}_{sfe,1}(\tau) - q(\tau)\right) \rightsquigarrow \mathcal{B}_{sfe}(\tau),$$

where $\mathcal{B}_{sfe}(\cdot)$ is a Gaussian process with covariance kernel $\Sigma_{sfe}(\cdot,\cdot)$. The expression for $\Sigma_{sfe}(\cdot,\cdot)$ can be found in the Appendix.

In particular, the asymptotic variance for $\hat{\beta}_{sfe,1}(\tau)$ is

$$\zeta_Y^2(\pi,\tau) + \zeta_A'^2(\pi,\tau) + \zeta_S^2(\tau),$$

where $\zeta_Y^2(\pi,\tau)$ and $\zeta_S^2(\tau)$ are the same as those defined below Theorem 3.1,

$$\begin{aligned}\zeta_A'^2(\pi,\tau) =& \mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau) - m_0(S,\tau))\left(\frac{1-\pi}{\pi f_1(q_1(\tau))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right) \\ &+ q(\tau)\left(\frac{f_1(q_1(\tau)|S)}{f_1(q_1(\tau))} - \frac{f_0(q_0(\tau)|S)}{f_0(q_0(\tau))}\right)\Bigg]^2,\end{aligned}$$

and $f_j(\cdot|s)$ is the conditional density of $Y(j)$ given $S = s$.

Three remarks are in order. First, if the treatment assignment rule achieve strong balance, then $\zeta_A'^2 = 0$ and the asymptotic variances for $\hat{\beta}_1(\tau)$ and $\hat{\beta}_{sfe,1}(\tau)$ are the same. Second, if the treatment assignment rule does not achieve strong balance, then it is difficult to compare the asymptotic variances of $\hat{\beta}_1(\tau)$ and $\hat{\beta}_{sfe,1}(\tau)$. Based on our simulation results in Section 6, the QR estimator with strata fixed effects usually has a smaller standard error. Third, in order to analytically compute the asymptotic variance $\hat{\beta}_{sfe,1}(\tau)$, one needs to nonparametrically estimate not only the unconditional densities $f_j(\cdot)$ but also the conditional densities $f_j(\cdot|s)$ for $j = 0, 1$ and $s \in \mathcal{S}$. However, such difficulty can be avoided by the bootstrap inference considered in Section 5.

3.3 Inverse Propensity Score weighted Quantile Regression

In the simple random sampling, $\hat{\pi}(S_i)$ defined in the previous section is also an estimator of the propensity score, i.e., π. In addition, Assumption 1.2 implies that the unconfoundedness condition holds. Therefore, following the lead of Firpo (2007), we can estimate $q_j(\tau)$ by the inverse propensity score weighted quantile regression. Let

$$\hat{q}_1(\tau) = \arg\min_q \frac{1}{n}\sum_{i=1}^n \frac{A_i}{\hat{\pi}(S_i)}\rho_\tau(Y_i - q) \quad \text{and} \quad \hat{q}_0(\tau) = \arg\min_q \frac{1}{n}\sum_{i=1}^n \frac{1-A_i}{1-\hat{\pi}(S_i)}\rho_\tau(Y_i - q).$$

Then, we estimate $q(\tau)$ by $\hat{q}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau)$.

Theorem 3.3. *If Assumptions 1 and 2 hold, then, uniformly over $\tau \in \Upsilon$,*

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

where $\mathcal{B}_{ipw}(\cdot)$ is a scalar Gaussian process with covariance kernel $\Sigma_{ipw}(\cdot,\cdot)$. The expression for $\Sigma_{ipw}(\cdot,\cdot)$ can be found in the Appendix.

In particular, the asymptotic variance for $\hat{q}(\tau)$ is

$$\zeta_Y^2(\pi,\tau) + \zeta_S^2(\tau).$$

Because both $\zeta_A^2(\pi,\tau)$ and $\zeta_A'^2(\pi,\tau)$ are nonnegative, $\hat{q}(\tau)$ is weakly more efficient than $\hat{\beta}_1(\tau)$ and $\hat{\beta}_{sfe,1}(\tau)$ for all randomization schemes that satisfy Assumption 1:

$$\Sigma_{ipw}(\tau,\tau) \leq \Sigma_{sqr}(\tau,\tau) \quad \text{and} \quad \Sigma_{ipw}(\tau,\tau) \leq \Sigma_{sfe}(\tau,\tau).$$

When the strong balance is achieved, both $\zeta_A^2(\pi,\tau)$ and $\zeta_A'^2(\pi,\tau)$ are zero. In this case, the three estimators are asymptotically first-order equivalent.

Analogously, we can also estimate the ATE by the inverse propensity score weighting method. Let $\theta = \mathbb{E}(Y(1) - Y(0))$ denote the true ATE. Then, we can estimate it by

$$\hat{\theta}_{ipw} = \frac{1}{n}\sum_{i=1}^n \frac{Y_i A_i}{\hat{\pi}(S_i)} - \frac{1}{n}\sum_{i=1}^n \frac{Y_i(1-A_i)}{1-\hat{\pi}(S_i)}.$$

Theorem 3.4. *If Assumption 1 holds, $\mathbb{E}(Y_i^2(1)+Y_i^2(0)) < \infty$, and, for some $s \in \mathcal{S}$ and $a \in \{0,1\}$, $\mathbb{V}(Y_i(a)|S_i = s) > 0$, then*

$$\sqrt{n}(\hat{\theta}_{ipw} - \theta) \rightsquigarrow \mathcal{N}(0, \sigma_{ipw}^2),$$

where $\mathcal{N}(0,\sigma_{ipw}^2)$ is a normal distribution with mean 0 and variance

$$\begin{aligned}\sigma_{ipw}^2 =& \left\{\frac{\mathbb{E}[Y_i(1) - \mathbb{E}(Y_i(1)|S_i)]^2}{\pi} + \frac{\mathbb{E}[Y_i(0) - \mathbb{E}(Y_i(0)|S_i)]^2}{1-\pi}\right\} + \left\{\mathbb{E}[\mathbb{E}(Y_i(1)|S_i) - \mathbb{E}(Y_i(0)|S_i)]^2\right\} \\ \equiv& \bar{\zeta}_Y^2(\pi) + \bar{\zeta}_S^2.\end{aligned}$$

The proof of Theorem 3.4 is basically the same as that of Theorem 3.3, and thus, is omitted for brevity. Denote the two sample means and the strata fixed effects estimators proposed in Bugni et al. (2018a) as $\hat{\theta}$ and $\hat{\theta}_{sfe}$, defining them as

$$\hat{\theta} = \frac{\sum_{i=1}^n A_i Y_i}{n_1(s)} - \frac{\sum_{i=1}^n (1-A_i) Y_i}{n(s) - n_1(s)}$$

and

$$\hat{\theta}_{sfe} = \frac{\sum_{i=1}^n \tilde{A}_i Y_i}{\sum_{i=1}^n \tilde{A}_i^2},$$

respectively. Further denote the asymptotic variances for $\hat{\theta}$ and $\hat{\theta}_{sfe}$ as σ_s^2 and σ_{sfe}^2, respectively. Then, by Bugni et al. (2018a, Theorems 4.1 and 4.3), we have

$$\sigma_s^2 = \bar{\zeta}_Y^2(\pi) + \bar{\zeta}_S^2 + \tilde{\zeta}_A^2 \quad \text{and} \quad \sigma_{sfe}^2 = \bar{\zeta}_Y^2(\pi) + \bar{\zeta}_S^2 + \tilde{\zeta}_\pi^2,$$

where both $\tilde{\zeta}_A^2$ and $\tilde{\zeta}_\pi^2$ are nonnegative. Therefore, the inverse propensity score estimator for ATE has the smallest asymptotic variance among the three:

$$\sigma_{ipw}^2 \leq \sigma_s^2 \quad \text{and} \quad \sigma_{ipw}^2 \leq \sigma_{sfe}^2.$$

Such efficiency is also achieved by the fully saturated linear regression proposed in Bugni et al. (2018b).

4 Weighted Bootstrap Inference

In this section, we consider the weighted bootstrap inference. Let $\{\xi_i\}_{i=1}^n$ be a sequence of bootstrap weights which will be specified later. Further denote $n_1^w(s) = \sum_{i=1}^n \xi_i A_i 1\{S_i = s\}$, $n^w(s) = \sum_{i=1}^n \xi_i 1\{S_i = s\}$, $\hat{\pi}^w(s) = n_1^w(s)/n^w(s)$, $\dot{A}_i = (1, A_i)'$, $\tilde{A}_i^w = A_i - \hat{\pi}^w(S_i)$, and $\dot{\tilde{A}}_i^w = (1, \tilde{A}_i^w)'$. Then, the weighted bootstrap counterparts of the three estimators studied in this paper can be written as

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

$$\hat{\beta}_{sfe}^w(\tau) = \arg\min_b \sum_{i=1}^n \xi_i \rho_\tau \left(Y_i - \dot{\tilde{A}}_i^w b \right),$$

and

$$\hat{q}^w(\tau) = \hat{q}^w_1(\tau) - \hat{q}^w_0(\tau),$$

where

$$\hat{q}^w_1(\tau) = \arg\min_q \sum_{i=1}^n \frac{\xi_i A_i}{\hat{\pi}^w(S_i)} \rho_\tau\left(Y_i - q\right) \quad \text{and} \quad \hat{q}^w_0(\tau) = \arg\min_q \sum_{i=1}^n \frac{\xi_i (1-A_i)}{1-\hat{\pi}^w(S_i)} \rho_\tau\left(Y_i - q\right).$$

In particular, the second elements $\hat{\beta}^w_1(\tau)$ and $\hat{\beta}^w_{sfe,1}(\tau)$ of vectors $\hat{\beta}^w(\tau)$ and $\hat{\beta}^w_{sfe}(\tau)$, respectively, and $\hat{q}^w(\tau)$ are the three bootstrap estimators of the τ-th QTE. Next, we specify the bootstrap weights.

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

The nonnegativity is required to maintain the convexity of the quantile regression objective function. The other conditions in Assumption 3 are common for the weighted bootstrap inference. In practice, we generate $\{\xi_i\}_{i=1}^n$ by the standard exponential distribution. In this case, the weighted bootstrap is also known as the Bayesian bootstrap.

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

$$\sqrt{n}\left(\hat{\beta}^w_1(\tau) - \hat{\beta}_1(\tau)\right) \rightsquigarrow \tilde{\mathcal{B}}_{sqr}(\tau),$$

$$\sqrt{n}\left(\hat{\beta}^w_{sfe,1}(\tau) - \hat{\beta}_{sfe,1}(\tau)\right) \rightsquigarrow \tilde{\mathcal{B}}_{sfe}(\tau),$$

and

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

where $\tilde{\mathcal{B}}_{sqr}(\tau)$ and $\tilde{\mathcal{B}}_{sfe}(\tau)$ are two Gaussian processes with covariance kernels being equal to those of $\mathcal{B}_{sqr}(\tau)$ and $\mathcal{B}_{sfe}(\tau)$ defined in Theorems 3.1 and 3.2, respectively, with $\gamma(s)$ being replaced by $\pi(1-\pi)$, and $\mathcal{B}_{ipw}(\tau)$ is the same Gaussian process defined in Theorem 3.3.

Five remarks are in order. First, the weighted bootstrap sample does not preserve the negative cross-sectional dependence in the original sample. Asymptotic variances of the weighted bootstrap estimators equal to those of their original sample counterparts as if simple random sampling generated the data. This asymptotic variance is intuitive as the weight ξ_i is independent with each other, which implies that, conditionally on data, the bootstrap sample observations are independent.

Second, for both the QR with or without strata fixed effects, the weighted bootstrap inference is conservative. In fact, the asymptotic variances for $\hat{\beta}_1^w(\tau)$ and $\hat{\beta}_{sfe,1}^w(\tau)$ are

$$\zeta_Y^2(\pi,\tau) + \tilde{\zeta}_A^2(\pi,\tau) + \zeta_S^2(\tau)$$

and

$$\zeta_Y^2(\pi,\tau) + \tilde{\zeta}_A'^2(\pi,\tau) + \zeta_S^2(\tau),$$

respectively, where

$$\tilde{\zeta}_A^2(\pi,\tau) = \mathbb{E}\pi(1-\pi)\left(\frac{m_1(S,\tau)}{\pi f_1(q_1(\tau))} + \frac{m_0(S,\tau)}{(1-\pi)f_0(q_0(\tau))}\right)^2$$

and

$$\begin{aligned}\tilde{\zeta}_A'^2(\pi,\tau) =& \mathbb{E}\pi(1-\pi)\bigg[(m_1(S,\tau)-m_0(S,\tau))\left(\frac{1-\pi}{\pi f_1(q_1(\tau))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right) \\ &+ q(\tau)\left(\frac{f_1(q_1(\tau)|S)}{f_1(q_1(\tau))} - \frac{f_0(q_0(\tau)|S)}{f_0(q_0(\tau))}\right)\bigg]^2.\end{aligned}$$

Because $\gamma(s) \leq \pi(1-\pi)$, we have

$$\zeta_A^2(\pi,\tau) \leq \tilde{\zeta}_A^2(\pi,\tau) \quad \text{and} \quad \zeta_A'^2(\pi,\tau) \leq \tilde{\zeta}_A'^2(\pi,\tau).$$

The inequalities are strict if the treatment assignment rule achieves strong balance.

Third, the weighted bootstrap inference has the exact asymptotic size for the inverse propensity score estimator. Theorem 3.3 shows that the asymptotic variance for $\hat{q}(\tau)$ is invariant against the treatment assignment rule applied. Therefore, even though the weighted bootstrap sample ignores the cross-sectional dependence and behaves as if the treatment status is generated randomly, the asymptotic variance for $\hat{q}^w(\tau)$ is still

$$\zeta_Y^2(\pi,\tau) + \zeta_S^2(\tau).$$

Fourth, by checking the proof of Theorem 4.1, the validity of weighted bootstrap for the inverse propensity score weighted estimator still holds if Assumption 1.3 is relaxed to

$$\sup_{s\in\mathcal{S}} |D_n(s)| = O_p(\sqrt{n}).$$

Fifth, it is also possible to consider the conventional nonparametric bootstrap which generates the bootstrap sample from the empirical distribution of the data. If the observations are i.i.d., van der Vaart and Wellner (1996, Section 3.6) shows that the conventional bootstrap is first-order equivalent to a weighted bootstrap with Poisson(1) weights. However, in the current setting, $\{A_i\}_{i\geq 1}$ is in general not independent. It is technically challenging to rigorously show that the above equivalence still holds. We leave it as an interesting topic for future research.

5 Covariate-Adaptive Bootstrap Inference

In this section, we consider the covariate-adaptive bootstrap procedure as follows:

1. Draw $\{S_i^*\}_{i=1}^n$ from the empirical distribution of $\{S_i\}_{i=1}^n$ with replacement.
2. Generate $\{A_i^*\}_{i=1}^n$ based on $\{S_i^*\}_{i=1}^n$ and the treatment assignment rule.
3. For $A_i^* = a$ and $S_i^* = s$, draw Y_i^* from the empirical distribution of Y_i given $A_i = a$ and $S_i = s$ with replacement.

First, step 1 is the conventional nonparametric bootstrap. The sample $\{S_i^*\}_{i=1}^n$ is obtained by drawing from the empirical distribution of $\{S_i\}_{i=1}^n$ with replacement n times. Second, step 2 repeats the treatment assignment rule, and thus preserves the cross-sectional dependence structure of the bootstrap sample, even after conditioning on data. The weighted bootstrap sample, by contrast, is cross-sectionally independent given data. Third, step 3 applies the conventional bootstrap procedure to the outcome Y_i in the cell $(S_i, A_i) = (s, a) \in \mathcal{S} \times \{0,1\}$. Given that the original data contain $n_a(s)$ observations in this cell, in this step, the bootstrap sample $\{Y_i^*\}_{i:A_i^*=a, S_i^*=s}$ is obtained by drawing from the empirical distribution of these $n_a(s)$ outcomes with replacement $n_a^*(s)$ times, where $n_a^*(s) = \sum_{i=1}^n 1\{A_i^* = a, S_i^* = s\}$. Unlike the conventional bootstrap, here both $n_a(s)$ and $n_a^*(s)$ are random and are not necessarily the same. Last, to implement the covariate-adaptive bootstrap, researchers need to know the treatment assignment rule for the original sample. Unlike in the observational study, in RCTs, such information is usually available. If one only knows that the assignment rule achieves strong balance, then Theorem 5.1 below still holds, provided that the bootstrap sample is generated from any assignment rule that achieves strong balance. Even worse, if no information on the treatment assignment rule is available, then one cannot implement the covariate-adaptive bootstrap inference. In this case, the weighted bootstrap for inverse propensity score weighted estimator can still provide a non-conservative t-test, as shown in Theorem 4.1.

Using the bootstrap sample $\{Y_i^*, A_i^*, S_i^*\}_{i=1}^n$, we can estimate QTE by the three methods considered in the paper, i.e., simple QR, QR with strata fixed effects, and inverse propensity score weighted QR. Let $n_1^*(s) = \sum_{i=1}^n A_i^* 1\{S_i^* = s\}$, $n^*(s) = \sum_{i=1}^n 1\{S_i^* = s\}$, $\hat{\pi}^*(s) = \frac{n_1^*(s)}{n^*(s)}$, $\dot{A}_i^* = (1, A_i^*)'$, $\tilde{A}_i^* = A_i^* - \hat{\pi}^*(S_i)$, and $\dot{\tilde{A}}_i^* = (1, \tilde{A}_i^*)'$. Then, the three bootstrap estimators can be written as

$$\hat{\beta}^*(\tau) = \arg\min_b \sum_{i=1}^n \rho_\tau \left(Y_i^* - \dot{A}_i^* b \right),$$

$$\hat{\beta}^*_{sfe}(\tau) = \arg\min_b \sum_{i=1}^n \rho_\tau \left(Y_i^* - \dot{\tilde{A}}_i^* b \right),$$

and

$$\hat{q}^*(\tau) = \hat{q}_1^*(\tau) - \hat{q}_0^*(\tau),$$

where

$$\hat{q}_1^* = \arg\min_q \sum_{i=1}^n \frac{A_i^*}{\hat{\pi}^*(S_i^*)} \rho_\tau(Y_i^* - q) \quad \text{and} \quad \hat{q}_0^* = \arg\min_q \sum_{i=1}^n \frac{1 - A_i^*}{1 - \hat{\pi}^*(S_i^*)} \rho_\tau(Y_i^* - q).$$

In particular, the second elements $\hat{\beta}_1^*(\tau)$ and $\hat{\beta}^*_{sfe,1}(\tau)$ of vectors $\hat{\beta}^*(\tau)$ and $\hat{\beta}^*_{sfe}(\tau)$, respectively, and $\hat{q}^*(\tau)$ are the three bootstrap estimators of the τ-th QTE. Parallel to Assumption 1, we make the following assumption on the bootstrap sample.

Assumption 4. *Let $D_n^*(s) = \sum_{i=1}^n (A_i^* - \pi) 1\{S_i^* = s\}$. Then, $\left\{ \left\{ \frac{D_n^*(s)}{\sqrt{n}} \right\}_{s \in \mathcal{S}} \middle| \{S_i^*\}_{i=1}^n \right\} \rightsquigarrow N(0, \Sigma_D)$ a.s.*

Assumption 4 is a high-level assumption. Obviously, it holds for SRS. For WEI, this condition holds by the same argument in Bugni et al. (2018a, Lemma B.12) with the fact that $\frac{n^*(s)}{n} \xrightarrow{p} p(s)$. For BCD, as shown in Bugni et al. (2018a, Lemma B.11),

$$D_n^*(s) | \{S_i^*\}_{i=1}^n = O_p(1).$$

Therefore, $D_n^*(s)/\sqrt{n} \xrightarrow{p} 0$ and Assumption 4 holds with $\gamma(s) = 0$. For SBR, it is clear that $|D_n^*(s)| \leq 1$. Therefore, Assumption 4 holds with $\gamma(s) = 0$ as well.

Theorem 5.1. *If Assumptions 1, 2, and 4 hold, then, uniformly over $\tau \in \Upsilon$ and conditionally on data,*

$$\sqrt{n} \left(\hat{\beta}_1^*(\tau) - \hat{q}(\tau) \right) \rightsquigarrow \mathcal{B}_{sqr}(\tau),$$

$$\sqrt{n} \left(\hat{\beta}^*_{sfe,1}(\tau) - \hat{q}(\tau) \right) \rightsquigarrow \mathcal{B}_{sfe}(\tau),$$

and

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

where $\mathcal{B}_{sqr}(\tau)$, $\mathcal{B}_{sfe}(\tau)$, and $\mathcal{B}_{ipw}(\tau)$ are three Gaussian processes defined in Theorem 3.1, 3.2, and 3.3, respectively.

Several remarks are in order. First, unlike the weighted bootstrap estimator, the covariate-adaptive bootstrap estimators are not centered on their corresponding counterparts from the original sample, but rather the inverse propensity score weighted estimator $\hat{q}(\tau)$. The reason is that the treatment status A_i^* is not generated by bootstrap. In the linear expansion of the bootstrap estimator, the part of the influence function that accounts for the variation generated by A_i^* need not be centered. We also know from the proof of Theorem 3.3 that the linear expansion of $\hat{q}(\tau)$ do not have the influence function that represents the variation generated by A_i. Therefore, it is natural to use it as the center.

Second, we do not necessarily need to estimate $\hat{q}(\tau)$ in order to make bootstrap inference. Note that the asymptotic distribution of the bootstrap estimator has the same dispersion as that of the original estimator. Therefore, we can use interdecile range[3] to estimate the standard error. Taking $\hat{\beta}^*_{sfe,1}(\tau)$ as an example, we first compute the bootstrap estimator B times and denote them as $\{\hat{\beta}^*_{sfe,1,b}(\tau)\}_{b=1}^B$. Then, the standard error estimator $\hat{\Sigma}_{sfe}(\tau,\tau)$ can be computed as

$$\hat{\Sigma}_{sfe}(\tau,\tau) = \frac{\hat{Q}(0.9) - \hat{Q}(0.9)}{\Phi^{-1}(0.9) - \Phi^{-1}(0.1)},$$

where $\hat{Q}(\tau)$ is the τ-th empirical quantile of the sequence $\left\{\hat{\beta}^*_{sfe,1,b}(\tau)\right\}_{b=1}^B$ and $\Phi(\cdot)$ is the standard normal CDF.

Third, for inferring the ATE, we can use the same bootstrap sample to compute the standard errors for $\hat{\theta}$, $\hat{\theta}_{sfe}$, and $\hat{\theta}_{ipw}$ and construct corresponding t-tests. We expect that all the conclusions for QTEs hold for the ATE as well. The simulation results in Section G of the Appendix provide some finite sample evidences.

6 Simulation

6.1 Data Generating Processes

We consider four DGPs with parameters $\gamma = 4$, $\sigma = 2$, and μ which will be specified later.

1. Let Z be the standardized $\beta(2,2)$ distributed, $S_i = \sum_{j=1}^4 \{Z_i \leq g_j\}$, and $(g_1, \cdots, g_4) = (-0.25\sqrt{20}, 0, 0.25\sqrt{20}, 0.5\sqrt{20})$. The outcome equation is

$$Y_i = A_i\mu + \gamma Z_i + \eta_i,$$

where $\eta_i = \sigma A_i \varepsilon_{i,1} + (1 - A_i)\varepsilon_{i,2}$ and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are jointly standard normal.

[3] It is valid to consider the difference of other two quantiles, such as 0.75 and 0.25. We recommend the interdecile range because it performs well in finite sample.

2. Let S be the same as in DGP1. The outcome equation is

$$Y_i = A_i\mu + \gamma Z_i A_i - \gamma(1 - A_i)(\log(Z_i + 3)1\{Z_i \leq 0.5\}) + \eta_i.$$

where $\eta_i = \sigma A_i \varepsilon_{i,1} + (1 - A_i)\varepsilon_{i,2}$ and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are jointly standard normal.

3. Let Z be uniformly distributed on $[-2, 2]$, $S_i = \sum_{j=1}^{4}\{Z_i \leq g_j\}$, and $(g_1, \cdots, g_4) = (-1, 0, 1, 2)$. The outcome equation is

$$Y_i = A_i\mu + A_i m_{i,1} + (1 - A_i)m_{i,0} + \eta_i,$$

where $m_{i,0} = \gamma Z_i^2 1\{|Z_i| \geq 1\} + \frac{\gamma}{4}(2 - Z_i^2)1\{|Z_i| < 1\}$, $\eta_i = \sigma(1+Z_i^2)A_i\varepsilon_{i,1} + (1+Z_i^2)(1-A_i)\varepsilon_{i,2}$, and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are mutually independent $T(3)/3$ distributed.

4. Let Z_i be normally distributed with mean 0 and variance 4, $S_i = \sum_{j=1}^{4}\{Z_i \leq g_j\}$, $(g_1, \cdots, g_4) = (2\Phi^{-1}(0.25), 2\Phi^{-1}(0.5), 2\Phi^{-1}(0.75), \infty)$, and $\Phi(\cdot)$ is the standard normal CDF. The outcome equation is

$$Y_i = A_i\mu + A_i m_{i,1} + (1 - A_i)m_{i,0} + \eta_i,$$

where $m_{i,0} = -\gamma Z_i^2/4$, $m_{i,1} = \gamma Z_i^2/4$,

$$\eta_i = \sigma(1 + 0.5\exp(-Z_i^2/2))A_i\varepsilon_{i,1} + (1 + 0.5\exp(-Z_i^2/2))(1 - A_i)\varepsilon_{i,2},$$

and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are jointly standard normal.

When $\pi = \frac{1}{2}$, for each DGP, we consider four randomization schemes:

1. SRS: Treatment assignment is generated as in Example 1.
2. WEI: Treatment assignment is generated as in Example 2 with $\phi(x) = (1 - x)/2$.
3. BCD: Treatment assignment is generated as in Example 3 with $\lambda = 0.75$.
4. SBR: Treatment assignment is generated as in Example 4.

When $\pi \neq 0.5$, WEI and BCD are not defined in the literature. Therefore, we only consider SRS and SBR as in Bugni et al. (2018a). We conduct the simulations with sample sizes $n = 200$ and 400. The numbers of simulation replications and bootstrap samples are 1000. Under the null, $\mu = 0$ and the true parameters of interest are computed by simulations with 10^6 sample size and 10^4 replications. Under the alternative, we perturb the true values by $\mu = 1$ and $\mu = 0.75$ for $n = 200$ and 400, respectively. Throughout this section, we focus on the median QTE. The simulation results for QTEs with $\tau = 0.25$ and 0.75 can be found in Section G in the Appendix. Section G also contains the simulation results for ATE. All the observations made in this section still apply.

6.2 QTE, $\pi = 0.5$

For the inference of QTEs, we consider eight t-tests with 95% nominal rate. For all of them, we construct the t-test statistics by one of the three point estimates studied in this paper and some estimate of the standard error. The null hypothesis will be rejected when the absolute value of the t-statistic is greater than 1.96. The details about the point estimates and standard errors are as follows:

1. "s/naive": the point estimator is computed by the simple QR and its standard error σ_{naive} is computed as

$$\begin{aligned}\sigma^2_{naive} =& \frac{\tau(1-\tau) - \frac{1}{n}\sum_{i=1}^n \hat{m}_1^2(S_i,\tau)}{\pi \hat{f}_1^2(\hat{q}_1(\tau))} + \frac{\tau(1-\tau) - \frac{1}{n}\sum_{i=1}^n \hat{m}_0^2(S_i,\tau)}{(1-\pi)\hat{f}_0^2(\hat{q}_0(\tau))} \\ &+ \frac{1}{n}\sum_{i=1}^n \pi(1-\pi)\left(\frac{\hat{m}_1(S_i,\tau)}{\pi \hat{f}_1(\hat{q}_1(\tau))} + \frac{\hat{m}_0(S_i,\tau)}{(1-\pi)\hat{f}_0(\hat{q}_0(\tau))}\right)^2 \\ &+ \frac{1}{n}\sum_{i=1}^n \left(\frac{\hat{m}_1(S_i,\tau)}{\hat{f}_1(\hat{q}_1(\tau))} - \frac{\hat{m}_0(S_i,\tau)}{\hat{f}_0(\hat{q}_0(\tau))}\right)^2, \end{aligned} \tag{6.1}$$

   where $\hat{q}_j(\tau)$ is the τ-the empirical quantile of $Y_i|A_i = j$,

$$\hat{m}_{i,1}(s,\tau) = \frac{\sum_{i=1}^n A_i 1\{S_i = s\}(\tau - 1\{Y_i \le \hat{q}_1(\tau)\})}{n_1(s)},$$

$$\hat{m}_{i,0}(s,\tau) = \frac{\sum_{i=1}^n (1-A_i) 1\{S_i = s\}(\tau - 1\{Y_i \le \hat{q}_0(\tau)\})}{n(s) - n_1(s)},$$

   and for $j = 0, 1$, $\hat{f}_j(\cdot)$ is computed by the kernel density estimation using the observations Y_i provided that $A_i = j$, bandwidth $h_j = 1.06\hat{\sigma}_j n_j^{-1/5}$, and the Gaussian kernel function, where $\hat{\sigma}_j$ is the standard deviation of the observations Y_i provided that $A_i = j$, and $n_j = \sum_{i=1}^n \{A_i = j\}$, $j = 0, 1$.

2. "s/adj": exactly the same as the "s/naive" method with one difference: replacing $\pi(1-\pi)$ in σ^2_{naive} by $\gamma(S_i)$.

3. "s/B": the point estimator is computed by the simple QR and its standard error σ_B is computed by the weighted bootstrap procedure. The bootstrap weights $\{\xi_i\}_{i=1}^n$ are generated from the standard exponential distribution. Denote $\{\hat{\beta}^w_{1,b}\}_{b=1}^B$ as the collection of B estimates obtained by the simple QR applied to the samples generated by the weighted bootstrap procedure. Then,

$$\sigma_B = \frac{\hat{Q}(0.9) - \hat{Q}(0.1)}{\Phi^{-1}(0.9) - \Phi^{-1}(0.1)},$$

where $\Phi(\cdot)$ is the standard normal CDF and $\hat{Q}(\tau)$ is the τ-th empirical quantile of $\{\hat{\beta}^w_{1,b}\}_{b=1}^B$.

4. "sfe/B": the same as above with one difference: the estimation method for both the original and bootstrap samples is the QR with strata fixed effects.

5. "ipw/B": the same as above with one difference: the estimation method for both the original and bootstrap samples is the inverse propensity score weighted QR.

6. "s/CA": the point estimator is computed by the simple QR and its standard error σ_{CA} is computed by the covariate-adaptive bootstrap procedure. Denote $\{\hat{\beta}^*_{1,b}\}_{b=1}^B$ as the collection of B estimates obtained by the simple QR applied to the samples generated by the covariate-adaptive bootstrap procedure. Then,

$$\sigma_{CA} = \frac{\hat{Q}(0.9) - \hat{Q}(0.1)}{\Phi^{-1}(0.9) - \Phi^{-1}(0.1)},$$

   where $\hat{Q}(\tau)$ is the τ-th empirical quantile of $\{\hat{\beta}^*_{1,b}\}_{b=1}^B$.

7. "sfe/CA": the same as above with one difference: the estimation method for both the original and bootstrap samples is the QR with strata fixed effects.

8. "ipw/CA": the same as above with one difference: the estimation method for both the original and bootstrap samples is the inverse propensity score weighted QR.

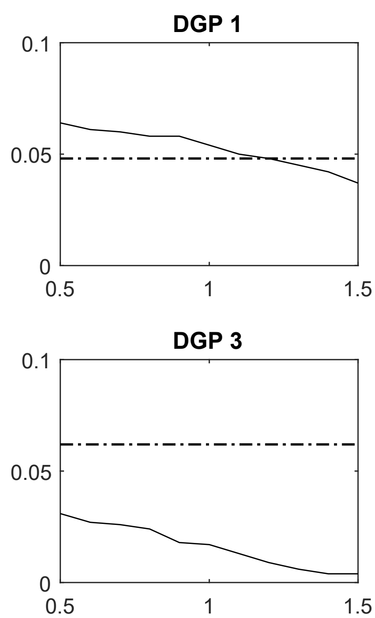
Tables 1 and 2 present the coverage rates for the eight t-tests under the null with sample sizes $n = 200$ and 400, respectively. In these two tables, column M and A represent DGPs and treatment assignment rules, respectively. We can make six observations. First, the naive t-test ("s/naive") is conservative for WEI, BCD, and SBR, which is consistent with the findings for ATE estimators discovered by Shao et al. (2010) and Bugni et al. (2018a). Second, although the adjusted t-test ("s/adj") is expected to have the exact asymptotic size, it does not perform well. The main reason is that, in order to analytically compute the standard error, one needs to compute nuisance parameters such as the unconditional densities of $Y(0)$ and $Y(1)$, which requires tuning parameters. We further compute the standard errors following (6.1) with $\pi(1-\pi)$ and the tuning parameter h_j replaced by $\gamma(S_i)$ and $1.06C_f\hat{\sigma}_j n_j^{-1/5}$, respectively, for some constant $C_f \in [0.5, 1.5]$. Figure 1 plots the rejection probabilities of the "s/adj" t-tests against C_f for the BCD assignment rule with $n = 200$, $\tau = 0.5$, and $\pi = 0.5$. We see that (1) the rejection probability is sensitive to the choice bandwidth, (2) there is no universal optimal bandwidth across different DGPs, and (3) the covariate-adaptive bootstrap t-tests ("s/CA") represented by the dotted dash lines are quite stable across different DGPs and close to the nominal rate of rejection. Third, the weighted bootstrap t-test for the simple QR estimator ("s/B") is conservative, especially for the first two DGPs with BCD and SBR assignment rules: both BCD and SBR achieve strong balance. Fourth, the weighted bootstrap t-test for the fixed effects estimator ("sfe/B") is slightly conservative, even for the assignment rule that achieve strong balance. For example, the rejection rates are 0.43 and 0.35 for DGP 2 and 4 with assignment rule SBR in Table 1. We will provide more evidence to illustrate the conservatism later. Fifth, the rejection probabilities of the weighted bootstrap t-test for the inverse propensity score weighted estimator ("ipw/B") is close to the nominal rate even for sample size $n = 200$. This is consistent with Theorem 4.1. Last, the rejection rates for the three covariate-adaptive bootstrap t-tests ("s/CA", "sfe/CA", and "ipw/CA") are close to the nominal rate, which is also consistent with Theorem 5.1.

Table 1: H_0, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.058	0.058	0.065	0.039	0.042	0.044	0.039	0.038
	WEI	0.009	0.055	0.019	0.042	0.039	0.045	0.039	0.038
	BCD	0.002	0.054	0.002	0.041	0.047	0.048	0.041	0.044
	SBR	0.002	0.057	0.001	0.050	0.051	0.049	0.047	0.049
2	SRS	0.049	0.049	0.046	0.054	0.059	0.044	0.054	0.060
	WEI	0.040	0.063	0.045	0.057	0.061	0.068	0.061	0.063
	BCD	0.014	0.055	0.022	0.048	0.058	0.056	0.061	0.058
	SBR	0.020	0.057	0.019	0.043	0.054	0.052	0.051	0.055
3	SRS	0.015	0.015	0.052	0.055	0.063	0.053	0.056	0.055
	WEI	0.012	0.012	0.052	0.051	0.060	0.058	0.054	0.059
	BCD	0.016	0.017	0.058	0.059	0.063	0.062	0.062	0.063
	SBR	0.013	0.015	0.054	0.052	0.057	0.057	0.057	0.058
4	SRS	0.021	0.021	0.064	0.052	0.063	0.061	0.053	0.061
	WEI	0.017	0.017	0.052	0.044	0.061	0.059	0.057	0.058
	BCD	0.018	0.019	0.057	0.048	0.060	0.062	0.062	0.061
	SBR	0.017	0.018	0.048	0.035	0.051	0.054	0.053	0.054

Table 2: H_0, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.040	0.040	0.048	0.031	0.030	0.039	0.029	0.031
	WEI	0.009	0.036	0.008	0.033	0.033	0.030	0.035	0.033
	BCD	0.000	0.041	0.000	0.035	0.036	0.037	0.034	0.035
	SBR	0.000	0.047	0.001	0.047	0.051	0.046	0.046	0.047
2	SRS	0.061	0.061	0.061	0.052	0.052	0.054	0.046	0.048
	WEI	0.027	0.060	0.030	0.058	0.064	0.059	0.065	0.065
	BCD	0.017	0.049	0.015	0.043	0.051	0.055	0.051	0.047
	SBR	0.015	0.047	0.016	0.047	0.055	0.057	0.056	0.056
3	SRS	0.016	0.016	0.045	0.049	0.054	0.048	0.048	0.054
	WEI	0.017	0.017	0.053	0.049	0.056	0.059	0.054	0.055
	BCD	0.012	0.012	0.055	0.056	0.055	0.060	0.057	0.058
	SBR	0.007	0.007	0.064	0.063	0.067	0.065	0.064	0.066
4	SRS	0.017	0.017	0.041	0.043	0.050	0.042	0.050	0.052
	WEI	0.020	0.020	0.073	0.054	0.073	0.072	0.064	0.072
	BCD	0.020	0.020	0.046	0.037	0.054	0.049	0.050	0.054
	SBR	0.020	0.020	0.035	0.036	0.039	0.037	0.042	0.041



Note: Rejection probabilities for BCD assignment rule with $n = 200$, $\pi = 0.5$, and $\tau = 0.5$. The X-axis is C_f. The solid lines are the rejection probabilities for "s/adj". The densities of Y_j is computed using the tuning parameters $h_j = 1.06C_f\hat{\sigma}_j n_j^{-1/5}$, for $j = 0, 1$. The dotted dash lines are the rejection probability for "s/CA".

Figure 1: Rejection Probabilities Across Different Bandwidth Values

Tables 3 and 4 show the powers of the eight t-tests for sample sizes $n = 200$ and 400, respectively.

We can make three observations. First, for DGPs 2 and 4, “sfe/B” has lower power than “ipw/B”, “sfe/CA” and “ipw/CA” for assignment rules “WEI”, “BCD”, and “SBR”, which illustrate that the weighted bootstrap inference for QR with strata fixed effects is conservative. Second, for BCD and SBR, the powers for “ipw/B”, “s/CA”, “sfe/CA”, and “ipw/CA” are close. This is because both BCD and SBR achieve strong balance. In this case, the three estimators proposed in this paper are asymptotically first-order equivalent. Third, for assignment rules SRS and WEI in Table 3, “ipw/CA” is more powerful than “sfe/CA” and “s/CA”. This confirms our theoretical finding that the inverse propensity score weighted estimator is *strictly* more efficient than the other two when the assignment rule does *not* achieve strong balance. In Table 4, “ipw/CA” is still more powerful than “s/CA”. However, the power of “sfe/CA” is close to that of “ipw/CA”.

Table 3: H_1, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.198	0.198	0.210	0.435	0.438	0.209	0.420	0.432
	WEI	0.132	0.303	0.150	0.442	0.443	0.320	0.431	0.440
	BCD	0.077	0.463	0.101	0.442	0.443	0.420	0.443	0.443
	SBR	0.068	0.483	0.089	0.445	0.452	0.446	0.448	0.444
2	SRS	0.256	0.256	0.253	0.361	0.372	0.263	0.361	0.372
	WEI	0.207	0.287	0.205	0.303	0.326	0.292	0.320	0.333
	BCD	0.205	0.359	0.212	0.318	0.343	0.333	0.343	0.342
	SBR	0.211	0.375	0.214	0.350	0.371	0.357	0.369	0.354
3	SRS	0.800	0.800	0.905	0.895	0.902	0.907	0.900	0.907
	WEI	0.805	0.809	0.905	0.902	0.908	0.912	0.901	0.904
	BCD	0.795	0.797	0.904	0.901	0.906	0.902	0.906	0.903
	SBR	0.808	0.817	0.913	0.916	0.916	0.913	0.914	0.914
4	SRS	0.170	0.170	0.295	0.266	0.319	0.307	0.296	0.315
	WEI	0.170	0.171	0.298	0.268	0.319	0.307	0.308	0.318
	BCD	0.180	0.187	0.312	0.291	0.319	0.318	0.318	0.318
	SBR	0.168	0.171	0.290	0.269	0.311	0.318	0.312	0.312

Table 4: H_1, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.216	0.216	0.227	0.500	0.502	0.227	0.495	0.493
	WEI	0.133	0.343	0.161	0.507	0.506	0.353	0.493	0.497
	BCD	0.093	0.503	0.098	0.491	0.489	0.476	0.482	0.479
	SBR	0.094	0.493	0.109	0.470	0.471	0.469	0.473	0.473
2	SRS	0.276	0.276	0.275	0.389	0.424	0.285	0.395	0.419
	WEI	0.266	0.370	0.254	0.398	0.415	0.359	0.414	0.409
	BCD	0.263	0.455	0.269	0.409	0.432	0.432	0.434	0.434
	SBR	0.257	0.443	0.268	0.417	0.438	0.438	0.447	0.436
3	SRS	0.897	0.897	0.954	0.949	0.953	0.953	0.948	0.952
	WEI	0.886	0.887	0.946	0.949	0.948	0.947	0.947	0.949
	BCD	0.888	0.889	0.945	0.941	0.945	0.945	0.943	0.944
	SBR	0.889	0.889	0.946	0.947	0.949	0.949	0.948	0.948
4	SRS	0.239	0.239	0.360	0.347	0.355	0.353	0.360	0.355
	WEI	0.245	0.245	0.356	0.341	0.373	0.360	0.368	0.369
	BCD	0.208	0.213	0.352	0.329	0.359	0.363	0.356	0.355
	SBR	0.222	0.223	0.344	0.330	0.351	0.358	0.362	0.359

6.3 QTE, $\pi = 0.7$

Tables 5–8 show the similar results with $\pi = 0.7$. In addition to the observations made previously, we want to highlight that, under the null, the rejection rates of "sfe/B" for DGP 4 and assignment rule SBR are 0.020 and 0.038 in Tables 5 and 6, respectively, which are below the nominal rate. Under the alternative, the rejection rates are 0.186 and 0.239, which are lower than those of "ipw/B", "s/CA", "sfe/CA", and "ipw/CA". Both observations indicate that "sfe/B" is conservative even for the assignment rule that achieves strong balance, such as SBR.

Table 5: H_0, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.052	0.052	0.057	0.042	0.049	0.040	0.044	0.044
	SBR	0.000	0.043	0.001	0.035	0.037	0.042	0.035	0.034
2	SRS	0.049	0.049	0.047	0.040	0.053	0.039	0.043	0.050
	SBR	0.020	0.056	0.022	0.050	0.050	0.052	0.052	0.050
3	SRS	0.010	0.010	0.052	0.052	0.062	0.053	0.056	0.058
	SBR	0.010	0.012	0.056	0.058	0.060	0.058	0.064	0.060
4	SRS	0.015	0.015	0.050	0.042	0.064	0.046	0.051	0.064
	SBR	0.007	0.012	0.035	0.020	0.051	0.047	0.045	0.047

Table 6: H_0, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.039	0.039	0.057	0.037	0.037	0.040	0.032	0.031
	SBR	0.000	0.047	0.001	0.043	0.046	0.045	0.043	0.041
2	SRS	0.056	0.056	0.054	0.053	0.053	0.044	0.051	0.050
	SBR	0.016	0.058	0.014	0.056	0.060	0.058	0.060	0.061
3	SRS	0.013	0.013	0.049	0.056	0.064	0.051	0.051	0.060
	SBR	0.006	0.006	0.045	0.050	0.055	0.051	0.053	0.051
4	SRS	0.014	0.014	0.048	0.039	0.055	0.050	0.048	0.057
	SBR	0.016	0.024	0.049	0.038	0.066	0.069	0.065	0.065

Table 7: H_1, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.186	0.186	0.195	0.407	0.419	0.181	0.397	0.400
	SBR	0.037	0.446	0.061	0.424	0.438	0.423	0.410	0.410
2	SRS	0.261	0.261	0.248	0.319	0.333	0.247	0.320	0.331
	SBR	0.228	0.381	0.225	0.359	0.380	0.353	0.363	0.368
3	SRS	0.774	0.774	0.883	0.866	0.881	0.885	0.866	0.879
	SBR	0.801	0.815	0.904	0.900	0.909	0.917	0.910	0.913
4	SRS	0.125	0.125	0.262	0.197	0.289	0.269	0.231	0.280
	SBR	0.104	0.142	0.274	0.186	0.323	0.322	0.323	0.321

Table 8: H_1, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.182	0.182	0.193	0.483	0.496	0.194	0.476	0.481
	SBR	0.055	0.513	0.076	0.485	0.496	0.503	0.491	0.491
2	SRS	0.328	0.328	0.298	0.420	0.409	0.290	0.416	0.410
	SBR	0.269	0.423	0.252	0.398	0.404	0.384	0.402	0.400
3	SRS	0.860	0.860	0.927	0.915	0.931	0.925	0.917	0.931
	SBR	0.860	0.867	0.932	0.930	0.934	0.936	0.933	0.936
4	SRS	0.200	0.200	0.333	0.278	0.372	0.348	0.296	0.371
	SBR	0.157	0.204	0.317	0.239	0.348	0.346	0.344	0.344

6.4 Summary

First, "s/naive", "s/B", and "sfe/B" are conservative while "s/adj", "ipw/B", "s/CA", "sfe/CA", and "ipw/CA" are not. Second, among the non-conservative t-tests, when the treatment assignment rule does not achieve strong balance (such as SRS and WEI), "ipw/B" and "ipw/CA" are strictly more powerful than "s/adj", "s/CA", and "sfe/CA". When the treatment assignment rule does achieve strong balance (such as BCD and SBR), "s/adj", "ipw/B", "s/CA", "sfe/CA", and "ipw/CA" are asymptotically first order equivalent. Third, the bootstrap based t-tests ("ipw/B", "s/CA", "sfe/CA", and "ipw/CA") have better finite sample performances than the analytically adjusted t-test ("s/adj").

7 Empirical Application

We illustrate our methods by estimating and inferring the average and quantile treatment effects of iron efficiency on educational attainment. The dataset we use is the same as the one analyzed by Chong, Cohen, Field, Nakasone, and Torero (2016) and Bugni et al. (2018a).

7.1 Data Description

The dataset consists of 215 students from one Peruvian secondary school during the 2009 school year. About two thirds of students were assigned as treatment group ($A = 1$ or $A = 2$). The other one third of students were assigned as control group ($A = 0$). One half of the students in the treatment group were exposed to a video of encouraging iron supplements by a physician ($A = 1$) and the other half were exposed to the same encouragement from a popular soccer player ($A = 2$). Those assignments were stratified by the number of years of secondary school completed ($\mathcal{S} = \{1, \cdots, 5\}$). The field experiment used a stratified block randomization scheme with fractions $(1/3, 1/3, 1/3)$ for each group, which achieves strong balance ($\gamma(s) = 0$).

In the following, we focus on the observations with $A = 0$ and $A = 1$, and estimate the treatment effect of the exposure to a video of encouraging iron supplements by a physician only. This practice was also implemented in Bugni et al. (2018a). In this case, the target proportions of treatment is $\pi = 1/2$. As in Chong et al. (2016), it is also possible to combine the two treatment groups, i.e., $A = 1$ and $A = 2$ and compute the treatment effects of exposure to a video of encouraging iron supplements by either a physician or a popular soccer player. Last, one can use the method developed in Bugni et al. (2018b) to estimate the treatment effects under multiple treatment status. However, in this setting, the validity of bootstrap inference has not been investigated yet and is an interesting topic for future research.

For each observation, we have three outcome variables: number of pills taken, grade point average, and cognitive ability measured by the average score across different Nintendo Wii games. For more details about the outcome variables, we refer interested readers to Chong et al. (2016). In the following, we focus on the grade point average only as the other two outcomes are discrete.

7.2 Test Statistics

Based on our theoretical and simulation results, we consider four non-conservative t-statistics: (1) the simple estimates (difference of the two sample means or sample quantiles) with covariate-adaptive bootstrap standard errors, (2) the strata fixed effects estimates with covariate-adaptive bootstrap standard errors, (3) the inverse propensity score weighted estimates with covariate-adaptive bootstrap standard errors, and (4) the inverse propensity score weighted estimates with weighted bootstrap standard errors. We denote them as "s/CA", "sfe/CA", "ipw/CA", "ipw/B", respectively. For the ATEs, we also compute the simple estimates with the adjusted standard errors based on the analytical formula derived in Bugni et al. (2018a), i.e., "s/adj". For QTE estimates, we consider quantile indexes $\{0.1, 0.15, \cdots, 0.90\}$. The number of replications for the two bootstrap methods is 1000. For the weighted bootstrap, we use the standard exponentially distributed weights.

7.3 Main Results

Table 9 shows the estimates with the corresponding standard errors in the parenthesis. From the table, we can make three remarks. First, for both ATE and QTE, the three estimates (simple, strata fixed effects, and inverse propensity score) and their standard errors computed via analytical formula, weighted bootstrap, and covariate-adaptive bootstrap are very close to each other. This is consistent with our theory that, under strong balance, all these estimators are first-order equivalent. Second, the four bootstrap-based p-values for the ATE are close to that of the adjust t-statistics computed in Bugni et al. (2018a, Table 6). Third, we do not compute the adjusted standard error for the QTEs as it requires tuning parameters. QTEs provide us a new insight that the impact of supplementation on grade promotion is only significantly positive at 25% among the three quantiles. This may imply that the policy of reducing iron deficiency is more effective for lower-ranked students.

Table 9: Grades Points Average

	s/adj	s/CA	sfe/CA	ipw/B	ipw/CA
ATE	0.35^{**}(0.16)	0.35^{**}(0.17)	0.37^{**}(0.17)	0.37^{**}(0.16)	0.37^{**}(0.17)
QTE,25%		0.43^{***} (0.15)	0.42^{***} (0.16)	0.43^{***} (0.15)	0.43^{***} (0.15)
QTE,50%		0.29(0.21)	0.30(0.22)	0.29(0.20)	0.29(0.21)
QTE,75%		0.35(0.24)	0.38(0.24)	0.36(0.26)	0.36(0.24)

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In order to provide more details on the QTE estimates, we plot the 95% point-wise confidence band in Figure 2 with quantile index ranging from 0.1 to 0.9. The blue line and the shadow area represent the point estimate and its 95% point-wise confidence interval, respectively. The confidence interval is constructed by

$$[\hat{\beta} - 1.96\hat{\sigma}(\hat{\beta}), \hat{\beta} + 1.96\hat{\sigma}(\hat{\beta})],$$

where $\hat{\beta}$ and $\hat{\sigma}(\hat{\beta})$ are computed in four combinations: "s/CA", "sfe/CA", "ipw/CA",and "ipw/B". As we expected, all the four figures look the same and the estimates are only significantly positive at low quantiles (15%-30%).

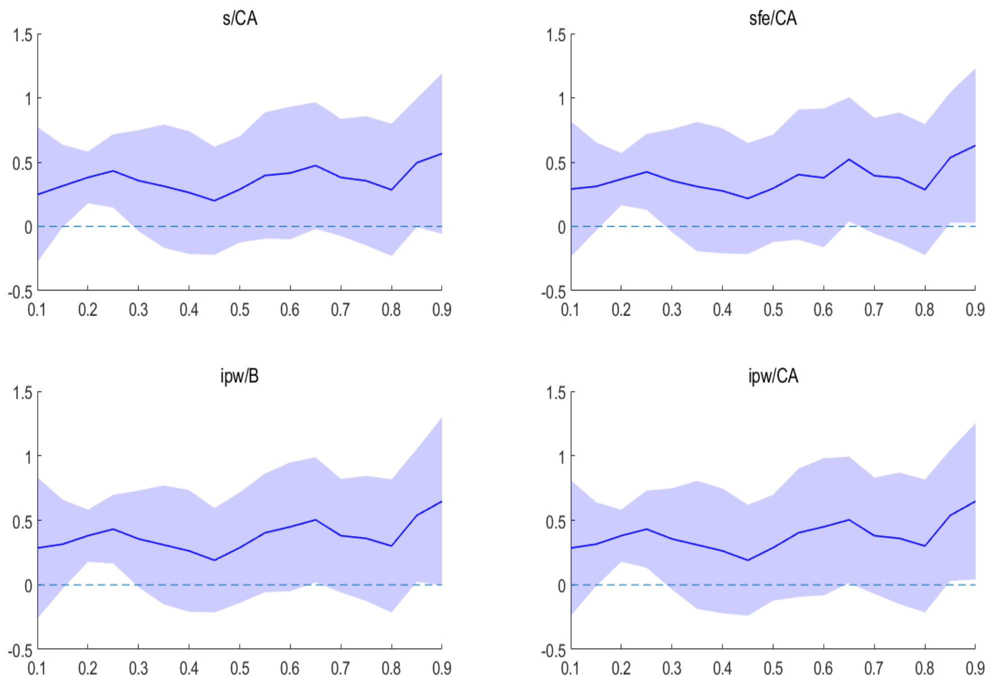


Figure 2: 95% Point-wise Confidence Interval for Quantile Treatment Effects

7.4 Subsample Results

Following Chong et al. (2016), we further split the sample into two based on whether the student is anemic, i.e., $Anem_i = 0$ or 1. We anticipate that there is no treatment effect for the nonanemic individuals and positive effects for anemic ones. In this subsample analysis, only the inverse propensity score weighted estimator with the weighted bootstrap is applicable. There are two reasons. First, the covariate-adaptive bootstrap is infeasible in the two subsamples, as the strong-balance condition may be lost and the treatment assignment rule is not necessarily the stratified block randomization anymore and is generally unknown. Second, however, the weighted bootstrap is still feasible as it does not require the knowledge of the treatment assignment rule. According to the fourth remark after Theorem 4.1, instead of Assumption 1.3, the weighted bootstrap for the inverse propensity score weighted estimator is valid if

$$\sup_{s \in \mathcal{S}} |D_n^{(1)}(s)| = \sum_{i=1}^{n} (A_i - \pi) 1\{S_i = s\} 1\{Anem_i = 1\} = O_p(\sqrt{n})$$

and

$$\sup_{s\in\mathcal{S}}|D_n^{(0)}(s)| = \sum_{i=1}^{n}(A_i-\pi)1\{S_i=s\}1\{Anem_i=0\} = O_p(\sqrt{n}).$$

We assume this weaker condition in this section.

From Table 10 and Figure 3, we see that the QTE estimates are significantly positive for the anemic students when the quantile index is between around 20%–75%, while are insignificant for nonanemic students.

Table 10: Grades Points Average

	Total	Anemic	Nonanemic
ATE	0.37**(0.16)	0.69***(0.19)	0.19(0.21)
QTE, 25%	0.43***(0.15)	0.76***(0.25)	0.22(0.27)
QTE, 50%	0.29(0.22)	1.05***(0.27)	-0.14(0.25)
QTE, 75%	0.36(0.25)	0.76**(0.32)	0.14(0.40)

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

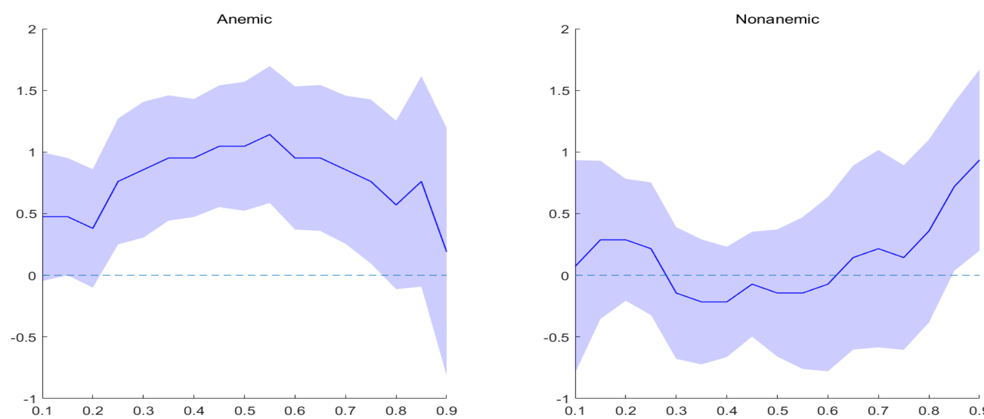


Figure 3: 95% Point-wise Confidence Interval for Anemic and Nonanemic Students

8 Conclusion

This paper studies the estimation and bootstrap inference for QTEs under covariate-adaptive randomization. We show that the weighted bootstrap inference is only valid for the inverse propensity score weighted estimator while the covariate-adaptive bootstrap is valid for all three estimators considered in the paper. In the empirical application, we find that the QTE of iron supplementation on grade promotion is trivial for nonanemic students, while the impact is significantly positive for middle-ranked anemic students.

A Proof of Theorem 3.1

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

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

Then, by the change of variable, we have that

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

Notice that $L_n(u,\tau)$ is convex in u for each τ and bounded in τ for each u. In the following, we aim to show that there exists

$$g_n(u,\tau) = -u'W_n(\tau) + \frac{1}{2}u'Q(\tau)u$$

such that (1) for each u,

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

(2) the maximum eigenvalue of $Q(\tau)$ is bounded from above and the minimum eigenvalue of $Q(\tau)$ is bounded away from 0, uniformly over $\tau \in \Upsilon$; (3) $W_n(\tau) \rightsquigarrow \tilde{\mathcal{B}}(\tau)$ uniformly over $\tau \in \Upsilon$, in which $\tilde{\mathcal{B}}(\cdot)$ is some Gaussian process. 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 $\sup_{\tau\in\Upsilon} ||r_n(\tau)|| = o_p(1)$. In addition, by (3), we have, uniformly over $\tau \in \Upsilon$,

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \rightsquigarrow [Q(\tau)]^{-1}\tilde{\mathcal{B}}(\tau) \equiv \mathcal{B}(\tau).$$

The second element of $\mathcal{B}(\tau)$ is $\mathcal{B}_{sqr}(\tau)$ stated in Theorem 3.1. In the following, we prove requirements (1)–(3) in three steps.

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

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

where

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

and

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

We first consider $Q_{n,1}(u,\tau)$. Following Bugni et al. (2018a), define $\{(Y_i^s(1), Y_i^s(0)) : 1 \le i \le n\}$ as a sequence of i.i.d. random variables with marginal distributions equal to the distribution of $(Y_i(1), Y_i(0))|S_i = s$. The distribution of $Q_{n,1}(u,\tau)$ is the same as the counterpart with units ordered by strata and then ordered by $A_i = 1$ first and $A_i = 0$ second within strata, i.e.,

$$\begin{aligned}
Q_{n,1}(u,\tau) \stackrel{d}{=}& \sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i^s(1) - q_1(\tau) \le v\} - 1\{Y_i^s(1) - q_1(\tau) \le 0\}\right) dv \\
=& \sum_{s\in\mathcal{S}} \left[\Gamma_n^s(N(s) + n_1(s), \tau) - \Gamma_n^s(N(s), \tau)\right], \qquad \text{(A.1)}
\end{aligned}$$

where $N(s) = \sum_{i=1}^n 1\{S_i < s\}$, $n_1(s) = \sum_{i=1}^n 1\{S_i = s\} A_i$, and

$$\Gamma_n^s(k,\tau) = \sum_{i=1}^{k} \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i^s(1) - q_1(\tau) \le v\} - 1\{Y_i^s(1) - q_1(\tau) \le 0\}\right) dv.$$

In addition, note that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in(0,1),\tau\in\Upsilon} |\Gamma_n^s(\lfloor nt \rfloor, \tau) - \mathbb{E}\Gamma_n^s(\lfloor nt \rfloor, \tau)| > \varepsilon\right) \\
=& \mathbb{P}\left(\max_{1\le k\le n} \sup_{\tau\in\Upsilon} |\Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)| > \varepsilon\right) \\
\le& 3 \max_{1\le k\le n} \mathbb{P}\left(\sup_{\tau\in\Upsilon} |\Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)| > \varepsilon/3\right) \\
\le& 9\mathbb{P}\left(\sup_{\tau\in\Upsilon} |\Gamma_n^s(n,\tau) - \mathbb{E}\Gamma_n^s(n,\tau)| > \varepsilon/30\right)
\end{aligned}$$

$$\leq \frac{270\mathbb{E}\sup_{\tau\in\Upsilon}|\Gamma_n^s(n,\tau)-\mathbb{E}\Gamma_n^s(n,\tau)|}{\varepsilon} = o(1). \tag{A.2}$$

The first inequality holds due to Lemma F.1 with $S_k = \Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)$ and $||S_k|| = \sup_{\tau\in\Upsilon}|\Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)|$. The second inequality holds due to Montgomery-Smith (1993, Theorem 1). For the last inequality of (A.2), consider the class of functions

$$\mathcal{F} = \left\{ \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \Big(1\{Y_i^s(1) - q_1(\tau) \leq v\} - 1\{Y_i^s(1) - q_1(\tau) \leq 0\} \Big) dv : \tau \in \Upsilon \right\}$$

with envelope $\frac{|u_0+u_1|}{\sqrt{n}}$ and

$$\sup_{f\in\mathcal{F}} \mathbb{E}f^2 \leq \sup_{\tau\in\Upsilon} \mathbb{E}\left[\frac{u_0+u_1}{\sqrt{n}} 1\left\{|Y_i^s(1) - q_1(\tau)| \leq \frac{u_0+u_1}{\sqrt{n}}\right\}\right]^2 \lesssim n^{-3/2}.$$

Note that $\mathcal{F}$ is a VC-class with a fixed VC index. Therefore, by Chernozhukov et al. (2014, Corollary 5.1),

$$\mathbb{E}\sup_{\tau\in\Upsilon}|\Gamma_n^s(n,\tau) - \mathbb{E}\Gamma_n^s(n,\tau)| = n||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} \lesssim n\left[\sqrt{\frac{\log(n)}{n^{5/2}}} + \frac{\log(n)}{n^{3/2}}\right] = o(1).$$

Therefore, (A.2) implies that

$$\sup_{\tau\in\Upsilon}\left| Q_n(u,\tau) - \sum_{s\in\mathcal{S}} \mathbb{E}\Big[\Gamma_n^s(\lfloor n(N(s)/n + n_1(s)/n)\rfloor,\tau) - \Gamma_n^s(\lfloor n(N(s)/n)\rfloor,\tau)\Big]\right| = o_p(1),$$

where following the convention in the empirical process literature,

$$\mathbb{E}\Big[\Gamma_n^s(\lfloor n(N(s)/n + n_1(s)/n)\rfloor,\tau) - \Gamma_n^s(\lfloor n(N(s)/n)\rfloor,\tau)\Big]$$

is interpreted as

$$\mathbb{E}\Big[\Gamma_n^s(\lfloor nt_2\rfloor,\tau) - \Gamma_n^s(\lfloor nt_1\rfloor,\tau)\Big]_{t_2=\frac{N(s)}{n}, t_2=\frac{N(s)+n_1(s)}{n}}.$$

In addition, $N(s)/n \xrightarrow{p} F(s) = F(S_i < s)$ and $n_1(s)/n \xrightarrow{p} \pi p(s)$. Thus, uniformly over $\tau \in \Upsilon$,

$$\mathbb{E}\Big[\Gamma_n^s(\lfloor n(N(s)/n + n_1(s)/n)\rfloor,\tau) - \Gamma_n^s(\lfloor n(N(s)/n)\rfloor,\tau)\Big] \xrightarrow{p} \frac{\pi f_1(q_1(\tau))(u_0+u_1)^2}{2}.$$

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

$$Q_{n,1}(u,\tau) \xrightarrow{p} \frac{\pi f_1(q_1(\tau))(u_0+u_1)^2}{2}.$$

Similarly, we can show that, uniformly over $\tau \in \Upsilon$,

$$Q_{n,0}(u,\tau) \xrightarrow{p} \frac{(1-\pi)f_0(q_0(\tau))u_0^2}{2},$$

and thus

$$Q_n(u,\tau) \xrightarrow{p} \frac{1}{2}u'Q(\tau)u,$$

where

$$Q(\tau) = \begin{pmatrix} \pi f_1(q_1(\tau)) + (1-\pi)f_0(q_0(\tau)) & \pi f_1(q_1(\tau)) \\ \pi f_1(q_1(\tau)) & \pi f_1(q_1(\tau)) \end{pmatrix}.$$

Then,

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

This concludes the first step.

Step 2. Note that $\det(Q(\tau)) = \pi(1-\pi)f_1(q_1(\tau))f_0(q_0(\tau))$, which is bounded and bounded away from zero. In addition, it can be shown that the two eigenvalues of Q are nonnegative. This leads to the desired result.

Step 3. Let $e_1 = (1,1)'$ and $e_0 = (1,0)'$. Then, we have

$$\begin{aligned} W_n(\tau) =& e_1 \sum_{s\in\mathcal{S}}\sum_{i=1}^n \frac{1}{\sqrt{n}}A_i 1\{S_i = s\}(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\ &+ e_0 \sum_{s\in\mathcal{S}}\sum_{i=1}^n \frac{1}{\sqrt{n}}(1-A_i)1\{S_i = s\}(\tau - 1\{Y_i(0) \le q_0(\tau)\}). \end{aligned}$$

Let $m_j(s,\tau) = \mathbb{E}(\tau - 1\{Y_i(j) \le q_j(\tau)\}|S_i = s)$ and $\eta_{i,j}(s,\tau) = (\tau - 1\{Y_i(j) \le q_j(\tau)\}) - m_j(s,\tau)$, $j = 0,1$. Then,

$$\begin{aligned} W_n(\tau) =& \left[e_1 \sum_{s\in\mathcal{S}}\sum_{i=1}^n \frac{1}{\sqrt{n}}A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) + e_0 \sum_{s\in\mathcal{S}}\sum_{i=1}^n \frac{1}{\sqrt{n}}(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)\right] \\ &+ \left[e_1 \sum_{s\in\mathcal{S}}\sum_{i=1}^n \frac{1}{\sqrt{n}}(A_i - \pi)1\{S_i = s\}m_1(s,\tau) - e_0 \sum_{s\in\mathcal{S}}\sum_{i=1}^n \frac{1}{\sqrt{n}}(A_i - \pi)1\{S_i = s\}m_0(s,\tau)\right] \end{aligned}$$

$$+\left[e_1\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\pi 1\{S_i=s\}m_1(s,\tau)+e_0\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}(1-\pi)1\{S_i=s\}m_0(s,\tau)\right]$$

$$\equiv W_{n,1}(\tau)+W_{n,2}(\tau)+W_{n,3}(\tau). \tag{A.3}$$

By Lemma F.2, uniformly over $\tau\in\Upsilon$,

$$(W_{n,1}(\tau),W_{n,2}(\tau),W_{n,3}(\tau))\rightsquigarrow(\mathcal{B}_1(\tau),\mathcal{B}_2(\tau),\mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau),\mathcal{B}_2(\tau),\mathcal{B}_3(\tau))$ are three independent two-dimensional Gaussian processes with covariance kernels $\Sigma_1(\tau_1,\tau_2)$, $\Sigma_2(\tau_1,\tau_2)$, and $\Sigma_3(\tau_1,\tau_2)$, respectively. Therefore, uniformly over $\tau\in\Upsilon$,

$$W_n(\tau)\rightsquigarrow\tilde{\mathcal{B}}(\tau),$$

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

$$\tilde{\Sigma}(\tau_1,\tau_2)=\sum_{j=1}^{3}\Sigma_j(\tau_1,\tau_2).$$

Consequently,

$$\sqrt{n}(\hat{\beta}(\tau)-\beta(\tau))\rightsquigarrow[Q(\tau)]^{-1}\tilde{\mathcal{B}}(\tau)\equiv\mathcal{B}(\tau),$$

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

$$\begin{aligned}
\Sigma(\tau_1,\tau_2)=&[Q(\tau_1)]^{-1}\tilde{\Sigma}(\tau_1,\tau_2)[Q(\tau_2)]^{-1}\\
=&\frac{1}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)]\begin{pmatrix}0&0\\0&1\end{pmatrix}\\
&+\frac{1}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)]\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\\
&+\sum_{s\in\mathcal{S}}p(s)\gamma(s)\left[\frac{m_1(s,\tau_1)m_1(s,\tau_2)}{\pi^2 f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix}0&0\\0&1\end{pmatrix}-\frac{m_1(s,\tau_1)m_0(s,\tau_2)}{\pi(1-\pi)f_1(q_1(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix}0&0\\1&-1\end{pmatrix}\right.\\
&\left.-\frac{m_0(s,\tau_1)m_1(s,\tau_2)}{\pi(1-\pi)f_0(q_0(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix}0&1\\0&-1\end{pmatrix}+\frac{m_0(s,\tau_1)m_0(s,\tau_2)}{(1-\pi)^2 f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}\right]\\
&+\frac{\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix}0&0\\0&1\end{pmatrix}+\frac{\mathbb{E}m_1(S,\tau_1)m_0(S,\tau_2)}{f_1(q_1(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix}0&0\\1&-1\end{pmatrix}\\
&+\frac{\mathbb{E}m_0(S,\tau_1)m_1(S,\tau_2)}{f_0(q_0(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix}0&1\\0&-1\end{pmatrix}+\frac{\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}.
\end{aligned}$$

Focusing on the $(2,2)$-element of $\Sigma(\tau_1,\tau_2)$, we can conclude that

$$\sqrt{n}(\hat{\beta}_1(\tau)-q(\tau)) \rightsquigarrow \mathcal{B}_{sqr}(\tau),$$

where the Gaussian process $\mathcal{B}_{sqr}(\tau)$ has a covariance kernel

$$\begin{aligned}
&\Sigma_{sqr}(\tau_1,\tau_2)\\
=&\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\\
&+\mathbb{E}\gamma(S)\Bigg[\frac{m_1(S,\tau_1)m_1(S,\tau_2)}{\pi^2 f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{m_1(S,\tau_1)m_0(S,\tau_2)}{\pi(1-\pi)f_1(q_1(\tau_1))f_0(q_0(\tau_2))}\\
&+\frac{m_0(S,\tau_1)m_1(S,\tau_2)}{\pi(1-\pi)f_0(q_0(\tau_1))f_1(q_1(\tau_2))}+\frac{m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)^2 f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\Bigg]\\
&+\mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}-\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}-\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right].
\end{aligned}$$

B Proof of Theorem 3.2

Define $\tilde{\beta}_1(\tau)=q(\tau)$, $\tilde{\beta}_0(\tau)=\pi q_1(\tau)+(1-\pi)q_0(\tau)$, $\tilde{\beta}(\tau)=(\tilde{\beta}_0(\tau),\tilde{\beta}_1(\tau))'$, and $\breve{A}_i=(1,A_i-\pi)'$. For arbitrary b_0 and b_1, let $u_0=\sqrt{n}(b_0-\tilde{\beta}_0(\tau))$, $u_1=\sqrt{n}(b_1-\tilde{\beta}_1(\tau))$, $u=(u_0,u_1)'\in\Re^2$, and

$$L_{sfe,n}(u,\tau)=\sum_{i=1}^{n}\left[\rho_\tau(Y_i-\breve{A}_i'\tilde{\beta}(\tau)-(\dot{\breve{A}}_i'b-\breve{A}_i'\tilde{\beta}(\tau)))-\rho_\tau(Y_i-\breve{A}_i'\tilde{\beta}(\tau))\right].$$

Then, by the change of variable, we have that

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau)-\tilde{\beta}(\tau))=\arg\min_u L_{sfe,n}(u,\tau).$$

Notice that $L_{sfe,n}(u,\tau)$ is convex in u for each τ and bounded in τ for each u. In the following, we aim to show that there exists

$$g_{sfe,n}(u,\tau)=-u'W_{sfe,n}(\tau)+\frac{1}{2}u'Q_{sfe}(\tau)u$$

such that (1) for each u,

$$\sup_{\tau\in\Upsilon}|L_{sfe,n}(u,\tau)-g_{sfe,n}(u,\tau)-h_{sfe,n}(\tau)|\xrightarrow{p}0,$$

where $h_{sfe,n}(\tau)$ does not depend on u; (2) the maximum eigenvalue of $Q_{sfe}(\tau)$ is bounded from above and the minimum eigenvalue of $Q_{sfe}(\tau)$ is bounded away from 0 uniformly over $\tau\in\Upsilon$; (3)

$W_{sfe,n}(\tau) \rightsquigarrow \tilde{\mathcal{B}}(\tau)$ uniformly over $\tau \in \Upsilon$ for some $\tilde{\mathcal{B}}(\tau)$.[4] Then by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau) - \tilde{\beta}(\tau)) = [Q_{sfe}(\tau)]^{-1} W_{sfe,n}(\tau) + r_{sfe,n}(\tau),$$

where $\sup_{\tau \in \Upsilon} ||r_{sfe,n}(\tau)|| = o_p(1)$. In addition, by (3), we have, uniformly over $\tau \in \Upsilon$,

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau) - \tilde{\beta}(\tau)) \rightsquigarrow [Q_{sfe}(\tau)]^{-1} \tilde{\mathcal{B}}(\tau) \equiv \mathcal{B}(\tau).$$

The second element of $\mathcal{B}(\tau)$ is $\mathcal{B}_{sfe}(\tau)$ stated in Theorem 3.2. Next, we prove requirements (1)–(3) in three steps.

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

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

Step 1.1. We first consider $L_{1,n}(u,\tau)$. Note that $\tilde{\beta}_1(\tau) = q(\tau)$ and

$$\begin{aligned}
& L_{1,n}(u,\tau) \\
= & \sum_{i=1}^{n}\sum_{s \in \mathcal{S}} A_i 1\{S_i = s\} \left(\frac{u_0}{\sqrt{n}} + (1-\hat{\pi}(s))\frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}(s))q(\tau)\right)(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
& + \sum_{i=1}^{n}\sum_{s \in \mathcal{S}} (1-A_i) 1\{S_i = s\} \left(\frac{u_0}{\sqrt{n}} - \hat{\pi}(s)\frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}(s))q(\tau)\right)(\tau - 1\{Y_i(0) \le q_0(\tau)\}) \\
\equiv & L_{1,1,n}(u,\tau) + L_{1,0,n}(u,\tau).
\end{aligned} \tag{B.1}$$

Let $\iota_1 = (1, 1-\pi)'$ and $\iota_0 = (1, -\pi)'$. Note that $\hat{\pi}(s) - \pi = \frac{D_n(s)}{n(s)}$. Then, for $L_{1,1,n}(u,\tau)$, we have

$$\begin{aligned}
& L_{1,1,n}(u,\tau) \\
= & \sum_{i=1}^{n}\sum_{s \in \mathcal{S}} A_i 1\{S_i = s\} \left[\frac{u'\iota_1}{\sqrt{n}} + (\pi - \hat{\pi}(s))\left(q(\tau) + \frac{u_1}{\sqrt{n}}\right)\right](\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
= & \frac{u'\iota_1}{\sqrt{n}} \sum_{i=1}^{n}\sum_{s \in \mathcal{S}} A_i 1\{S_i = s\}(\tau - 1\{Y_i(1) \le q_1(\tau)\})
\end{aligned}$$

[4] We abuse the notation and denote the weak limit of $W_{sfe,n}(\tau)$ as $\tilde{\mathcal{B}}(\tau)$. This limit is different from the weak limit of $W_n(\tau)$ in the proof of Theorem 3.1.

$$
\begin{aligned}
&-\sum_{s\in\mathcal{S}}\frac{D_n(s)}{\sqrt{n}}\frac{u_1}{n(s)}\sum_{i=1}^n A_i 1\{S_i=s\}\left(\tau-1\{Y_i(1)\le q_1(\tau)\}\right)\\
&+\sum_{s\in\mathcal{S}}(\pi-\hat{\pi}(s))q(\tau)\sum_{i=1}^n A_i 1\{S_i=s\}\left(\tau-1\{Y_i(1)\le q_1(\tau)\}\right)\\
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_1}{\sqrt{n}}\sum_{i=1}^n\Big[A_i1\{S_i=s\}\eta_{i,1}(s,\tau)+(A_i-\pi)1\{S_i=s\}m_1(s,\tau)+\pi 1\{S_i=s\}m_1(s,\tau)\Big]\\
&-\sum_{s\in\mathcal{S}}\frac{D_n(s)}{\sqrt{n}}\frac{u_1}{n(s)}\sum_{i=1}^n\Big[A_i1\{S_i=s\}\eta_{i,1}(s,\tau)+(A_i-\pi)1\{S_i=s\}m_1(s,\tau)+\pi 1\{S_i=s\}m_1(s,\tau)\Big]+h_{1,1}(\tau)\\
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_1}{\sqrt{n}}\sum_{i=1}^n\Big[A_i1\{S_i=s\}\eta_{i,1}(s,\tau)+(A_i-\pi)1\{S_i=s\}m_1(s,\tau)+\pi 1\{S_i=s\}m_1(s,\tau)\Big]\\
&-\sum_{s\in\mathcal{S}}\frac{u_1D_n(s)\pi m_1(s,\tau)}{\sqrt{n}}+h_{1,1}(\tau)+R_{sfe,1,1}(u,\tau),
\end{aligned}\tag{B.2}
$$

where

$$
h_{1,1}(\tau)=\sum_{s\in\mathcal{S}}(\pi-\hat{\pi}(s))q(\tau)\sum_{i=1}^n A_i1\{S_i=s\}\left(\tau-1\{Y_i(1)\le q_1(\tau)\}\right)
$$

and

$$
R_{sfe,1,1}(u,\tau)=-\sum_{s\in\mathcal{S}}\frac{u_1D_n(s)}{\sqrt{n}n(s)}\sum_{i=1}^n\Big[A_i1\{S_i=s\}\eta_{i,1}(s,\tau)+(A_i-\pi)1\{S_i=s\}m_1(s,\tau)\Big].
$$

By the same argument in Lemma F.2 and Assumption 1.3, we have for every $s\in\mathcal{S}$,

$$
\sup_{\tau\in\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n A_i1\{S_i=s\}\eta_{i,1}(s,\tau)\right|=O_p(1)\tag{B.3}
$$

and

$$
\sup_{\tau\in\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big[(A_i-\pi)1\{S_i=s\}m_1(s,\tau)\Big]\right|=\sup_{\tau\in\Upsilon}\left|\frac{D_n(s)m_1(s,\tau)}{\sqrt{n}}\right|=O_p(1).
$$

In addition, note that $n(s)/n\xrightarrow{p}p(s)$. Therefore,

$$
\sup_{\tau\in\Upsilon}|R_{sfe,1,1}(u,\tau)|=O_p(\frac{1}{\sqrt{n}})=o_p(1).
$$

Similarly, we have

$L_{1,0,n}(u,\tau)$

$$
\begin{aligned}
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_0}{\sqrt{n}}\sum_{i=1}^{n}\Big[(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)-(A_i-\pi)1\{S_i=s\}m_0(s,\tau)+(1-\pi)1\{S_i=s\}m_0(s,\tau)\Big]\\
&-\sum_{s\in\mathcal{S}}\frac{u_1D_n(s)(1-\pi)m_0(s,\tau)}{\sqrt{n}}+h_{1,0}(\tau)+R_{sfe,1,0}(u,\tau), \qquad \text{(B.4)}
\end{aligned}
$$

where

$$
h_{1,0}(\tau)=\sum_{s\in\mathcal{S}}(\pi-\hat{\pi}(s))q(\tau)\sum_{i=1}^{n}(1-A_i)1\{S_i=s\}\left(\tau-1\{Y_i(0)\le q_0(\tau)\}\right),
$$

$$
R_{sfe,1,0}(u,\tau)=-\sum_{s\in\mathcal{S}}\frac{u_1D_n(s)}{\sqrt{n}n(s)}\sum_{i=1}^{n}\Big[(1-A_i)1\{S_i=s\}\eta_{i,0}(\tau)-(A_i-\pi)1\{S_i=s\}m_0(s,\tau)\Big],
$$

and

$$
\sup_{\tau\in\Upsilon}|R_{sfe,1,0}(\tau)|=O_p(\frac{1}{\sqrt{n}})=o_p(1).
$$

Combining (B.1), (B.2), (B.4) and letting $\iota_2=(1,1-2\pi)'$, we have

$$
\begin{aligned}
L_{1,n}(u,\tau)=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\Big[u'\iota_1A_i1\{S_i=s\}\eta_{i,1}(s,\tau)+u'\iota_0(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)\Big]\\
&+\sum_{s\in\mathcal{S}}u'\iota_2\frac{D_n(s)}{\sqrt{n}}\left(m_1(s,\tau)-m_0(s,\tau)\right)\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(u'\iota_1\pi m_1(S_i,\tau)+u'\iota_0(1-\pi)m_0(S_i,\tau)\big)\\
&+R_{sfe,1,1}(u,\tau)+R_{sfe,1,0}(u,\tau)+h_{1,1}(\tau)+h_{1,0}(\tau). \qquad \text{(B.5)}
\end{aligned}
$$

Step 1.2. Next, we consider $L_{2,n}(u,\tau)$. Denote $E_n(s)=\sqrt{n}(\hat{\pi}(s)-\pi)$. Then,

$$
\{E_n(s)\}_{s\in\mathcal{S}}=\left\{\frac{D_n(s)}{\sqrt{n}}\frac{n}{n(s)}\right\}_{s\in\mathcal{S}}\rightsquigarrow\mathcal{N}(0,\Sigma_D')=O_p(1),
$$

where $\Sigma_D'=\text{diag}(\gamma(s)/p(s):s\in\mathcal{S})$. In addition,

$$
\begin{aligned}
&L_{2,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}A_i1\{S_i=s\}\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{E_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}\left(1\{Y_i(1)\le q_1(\tau)+v\}-1\{Y_i(1)\le q_1(\tau)\}\right)dv
\end{aligned}
$$

$$+\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}(1-A_i)1\{S_i=s\}\int_0^{\frac{u'\iota_0}{\sqrt{n}}-\frac{E_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}\left(1\{Y_i(0)\leq q_0(\tau)+v\}-1\{Y_i(0)\leq q_0(\tau)\}\right)dv$$

$$\equiv L_{2,1,n}(u,\tau)+L_{2,0,n}(u,\tau). \tag{B.6}$$

By the same argument in (A.1), we have

$$L_{2,1,n}(u,\tau)\stackrel{d}{=}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{E_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}\left(1\{Y_i^s(1)\leq q_1(\tau)+v\}-1\{Y_i^s(1)\leq q_1(\tau)\}\right)dv$$

$$\equiv\sum_{s\in\mathcal{S}}\left[\Gamma_n^s(N(s)+n_1(s),\tau,E_n(s))-\Gamma_n^s(N(s),\tau,E_n(s))\right], \tag{B.7}$$

where

$$\Gamma_n^s(k,\tau,e)=\sum_{i=1}^{k}\int_0^{\frac{u'\iota_1-e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}}\left(1\{Y_i^s(1)\leq q_1(\tau)+v\}-1\{Y_i^s(1)\leq q_1(\tau)\}\right)dv.$$

We want to show, for some any sufficiently large constant M,

$$\sup_{0<t\leq 1,\tau\in\Upsilon,|e|\leq M}|\Gamma_n^s(\lfloor nt\rfloor,\tau,e)-\mathbb{E}\Gamma_n^s(\lfloor nt\rfloor,\tau,e)|=o_p(1). \tag{B.8}$$

By the same argument in (A.2), it suffices to show that

$$\sup_{\tau\in\Upsilon,|e|\leq M}n||\mathbb{P}_n-\mathbb{P}||_{\mathcal{F}}=o_p(1),$$

where

$$\mathcal{F}=\left\{\int_0^{\frac{u'\iota_1-e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}}\left(1\{Y_i^s(1)\leq q_1(\tau)+v\}-1\{Y_i^s(1)\leq q_1(\tau)\}\right)dv:\tau\in\Upsilon,|e|\leq M\right\}$$

with an envelope $F=\frac{|u_0|+|u_1|+M\sup_{\tau\in\Upsilon}|q(\tau)|+\frac{|u_1|}{\sqrt{n}}}{\sqrt{n}}$. Note that

$$\sup_{f\in\mathcal{F}}\mathbb{E}f^2\leq\sup_{\tau\in\Upsilon}\mathbb{E}\left[\frac{|u_0|+|u_1|+M|q(\tau)|+\frac{|u_1|}{\sqrt{n}}}{\sqrt{n}}1\left\{|Y_i^s(1)-q_1(\tau)|\leq\frac{|u_0|+|u_1|+M|q(\tau)|+\frac{|u_1|}{\sqrt{n}}}{\sqrt{n}}\right\}\right]^2$$

$$\lesssim n^{-3/2},$$

and $\mathcal{F}$ is a VC-class with a fixed VC index. Then, by Chernozhukov et al. (2014, Corollary 5.1),

$$\mathbb{E}\sup_{\tau\in\Upsilon,|e|\leq M}|\Gamma_n^s(n,\tau,e)-\mathbb{E}\Gamma_n^s(n,\tau,e)|=n||\mathbb{P}_n-\mathbb{P}||_{\mathcal{F}}\lesssim n\left[\sqrt{\frac{\log(n)}{n^{5/2}}}+\frac{\log(n)}{n^{3/2}}\right]=o(1). \tag{B.9}$$

In addition, we have

$$\begin{aligned}\mathbb{E}\Gamma_n^s(\lfloor nt\rfloor,\tau,e)=&\lfloor nt\rfloor\int_0^{\frac{u'\iota_1-e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}}[F_1(q_1(\tau)+v|s)-F_1(q_1(\tau)|s)]dv\\=&t\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1-eq(\tau))^2+o(1),\end{aligned} \tag{B.10}$$

where $F_j(\cdot|s)$ and $f_j(\cdot|s)$, $j=0,1$ are the conditional CDF and PDF for $Y(j)$ given $S=s$, respectively, and the $o(1)$ term holds uniformly over $\{\tau\in\Upsilon,|e|\leq M\}$. Combining (B.8) and (B.10) with the fact that $\frac{n_1(s)}{n}\xrightarrow{p}\pi p(s)$, we have

$$\begin{aligned}L_{2,1,n}(u,\tau)=&\sum_{s\in\mathcal{S}}\pi p(s)\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1-E_n(s)q(\tau))^2+R'_{sfe,2,1}(u,\tau)\\=&\frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2-\sum_{s\in\mathcal{S}}f_1(q_1(\tau)|s)\frac{\pi D_n(s)u'\iota_1}{\sqrt{n}}q(\tau)+h_{2,1}(\tau)+R_{sfe,2,1}(u,\tau),\end{aligned} \tag{B.11}$$

where

$$\sup_{\tau\in\Upsilon}|R'_{sfe,2,1}(u,\tau)|=o_p(1),\quad\sup_{\tau\in\Upsilon}|R_{sfe,2,1}(u,\tau)|=o_p(1),$$

and

$$h_{2,1}(\tau)=\sum_{s\in\mathcal{S}}\frac{\pi f_1(q_1(\tau)|s)}{2}p(s)E_n^2(s)\tilde{\beta}_1^2(\tau).$$

Similarly, we have

$$\begin{aligned}L_{2,0,n}(u,\tau)=&\frac{(1-\pi)f_0(q_0(\tau))}{2}(u'\iota_0)^2-\sum_{s\in\mathcal{S}}(1-\pi)f_0(q_0(\tau)|s)\frac{D_n(s)u'\iota_0}{\sqrt{n}}q(\tau)\\&+h_{2,0}(\tau)+R_{sfe,2,0}(u,\tau),\end{aligned} \tag{B.12}$$

where

$$\sup_{\tau\in\Upsilon}|R_{sfe,2,0}(u,\tau)|=o_p(1)\quad\text{and}\quad h_{2,0}(\tau)=\sum_{s\in\mathcal{S}}\frac{(1-\pi)f_0(q_0(\tau)|s)}{2}p(s)E_n^2(s)\tilde{\beta}_1^2(\tau).$$

Combining (B.6), (B.11), and (B.12), we have

$$L_{2,n}(u,\tau) = \frac{1}{2}u'Q_{sfe}(\tau)u - \sum_{s\in\mathcal{S}} q(\tau)\left[f_1(q_1(\tau)|s)\pi u'\iota_1 + f_0(q_0(\tau)|s)(1-\pi)u'\iota_0\right]\frac{D_n(s)}{\sqrt{n}}$$
$$+ R_{sfe,2,1}(u,\tau) + R_{sfe,2,0}(u,\tau) + h_{2,1}(\tau) + h_{2,0}(\tau). \tag{B.13}$$

where

$$\begin{aligned} Q_{sfe} &= \pi f_1(q_1(\tau))\iota_1\iota_1' + (1-\pi)f_0(q_0(\tau))\iota_0\iota_0' \\ &= \begin{pmatrix} \pi f_1(q_1(\tau)) + (1-\pi)f_0(q_0(\tau)) & \pi(1-\pi)(f_1(q_1(\tau)) - f_0(q_0(\tau))) \\ \pi(1-\pi)(f_1(q_1(\tau)) - f_0(q_0(\tau))) & \pi(1-\pi)((1-\pi)f_1(q_1(\tau)) + \pi f_0(q_0(\tau))) \end{pmatrix}. \end{aligned}$$

Step 1.3. Last, by combining (B.5) and (B.13), we have

$$L_{sfe,n}(u,\tau) = -u'W_{sfe,n}(\tau) + \frac{1}{2}u'Q_{sfe}(\tau)u + R_{sfe}(u,\tau) + h_{sfe,n}(\tau),$$

where

$$\begin{aligned} & W_{sfe,n}(\tau) \\ =& \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\iota_1 A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) + \iota_0(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)\right] \\ & + \sum_{s\in\mathcal{S}}\left\{\iota_2\left(m_1(s,\tau) - m_0(s,\tau)\right) + q(\tau)\left[f_1(q_1(\tau)|s)\pi\iota_1 + f_0(q_0(\tau)|s)(1-\pi)\iota_0\right]\right\}\frac{D_n(s)}{\sqrt{n}} \\ & + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\iota_1\pi m_1(S_i,\tau) + \iota_0(1-\pi)m_0(S_i,\tau)\right) \\ \equiv& W_{sfe,n,1}(\tau) + W_{sfe,n,2}(\tau) + W_{sfe,n,3}(\tau), \end{aligned} \tag{B.14}$$

$$R_{sfe}(u,\tau) = R_{sfe,1,1}(u,\tau) + R_{sfe,1,0}(u,\tau) + R_{sfe,2,1}(u,\tau) + R_{sfe,2,0}(u,\tau)$$

such that $\sup_{\tau\in\Upsilon}|R_{sfe}(u,\tau)| = o_p(1)$, and

$$h_{sfe,n}(\tau) = h_{1,1}(\tau) + h_{1,0}(\tau) + h_{2,1}(\tau) + h_{2,0}(\tau).$$

This concludes the proof of Step 1.

Step 2. Note that $\det(Q_{sfe}(\tau)) = \pi(1-\pi)f_0(q_0(\tau))f_1(q_1(\tau))$, which is bounded and bounded away from zero. In addition, it can be shown that the two eigenvalues of $Q_{sfe}(\tau)$ are nonnegative. This leads to the desired result.

Step 3. Lemma F.3 establishes the weak convergence that

$$(W_{sfe,1,n}(\tau), W_{sfe,2,n}(\tau), W_{sfe,3,n}(\tau)) \rightsquigarrow (\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau)),$$

where $(\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau))$ are three independent two-dimensional Gaussian processes with covariance kernels $\Sigma_1(\tau_1,\tau_2)$, $\Sigma_2(\tau_1,\tau_2)$, and $\Sigma_3(\tau_1,\tau_2)$, respectively. Therefore, uniformly over $\tau \in \Upsilon$,

$$W_{sfe,n}(\tau) \rightsquigarrow \tilde{\mathcal{B}}(\tau),$$

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

$$\tilde{\Sigma}(\tau_1,\tau_2) = \sum_{j=1}^{3} \Sigma_j(\tau_1,\tau_2).$$

Consequently,

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau) - \tilde{\beta}(\tau)) \rightsquigarrow \mathcal{B}(\tau) \equiv Q_{sfe}^{-1}(\tau)\tilde{\mathcal{B}}(\tau),$$

where $\Sigma(\tau_1,\tau_2)$, the covariance kernel of $\mathcal{B}(\tau)$, has the expression that

$$\begin{aligned}
&\Sigma(\tau_1,\tau_2)\\
=&Q_{sfe}^{-1}(\tau_1)\tilde{\Sigma}(\tau_1,\tau_2)Q_{sfe}^{-1}(\tau_2)\\
=&\Bigg\{\frac{1}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\left[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)\right]\begin{pmatrix}\pi^2 & \pi\\ \pi & 1\end{pmatrix}\\
&+\frac{1}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\left[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)\right]\begin{pmatrix}(1-\pi)^2 & \pi-1\\ \pi-1 & 1\end{pmatrix}\Bigg\}\\
&+\Bigg\{\mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau_1)-m_0(S,\tau_1))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_1))}+\frac{1-\pi}{f_1(q_1(\tau_1))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_1))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\end{pmatrix} + q(\tau_1)\frac{f_1(q_1(\tau_1)|S)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\ 1\end{pmatrix}\\
&+q(\tau_1)\frac{f_0(q_0(\tau_1)|S)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\Bigg]\times\Bigg[(m_1(S,\tau_2)-m_0(S,\tau_2))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_2))}+\frac{1-\pi}{f_1(q_1(\tau_2))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_2))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\end{pmatrix}\\
&+q(\tau_2)\frac{f_1(q_1(\tau_2)|S)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\ 1\end{pmatrix} + q(\tau_2)\frac{f_0(q_0(\tau_2)|S)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\Bigg]\Bigg\}\\
&+\Bigg\{\mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\ 1\end{pmatrix}+\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\ 1\end{pmatrix}+\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\right]'\Bigg\}.
\end{aligned}$$

By checking the $(2,2)$-element of $\Sigma(\tau_1,\tau_2)$, we have

$$
\begin{aligned}
&\Sigma_{sfe}(\tau_1,\tau_2)\\
=&\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\\
&+\mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau_1)-m_0(S,\tau_1))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_1))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\right)+q(\tau_1)\left(\frac{f_1(q(\tau_1)|S)}{f_1(q_1(\tau_1))}-\frac{f_0(q(\tau_1)|S)}{f_0(q_0(\tau_1))}\right)\Bigg]\\
&\times\left[(m_1(S,\tau_2)-m_0(S,\tau_2))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_2))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\right)+q(\tau_2)\left(\frac{f_1(q(\tau_2)|S)}{f_1(q_2(\tau_2))}-\frac{f_0(q(\tau_2)|S)}{f_0(q_0(\tau_2))}\right)\right]\\
&+\mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}-\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}-\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right].
\end{aligned}
$$

C Proof of Theorem 3.3

By Knight's identity, we have

$$\sqrt{n}(\hat{q}_1(\tau)-q_1(\tau))=\arg\min_u L_n(u,\tau),$$

where

$$
\begin{aligned}
L_n(u,\tau)\equiv&\sum_{i=1}^n\frac{A_i}{\hat{\pi}(S_i)}\left[\rho_\tau(Y_i-q_1(\tau)-\frac{u}{\sqrt{n}})-\rho_\tau(Y_i-q_1(\tau))\right]\\
=&-L_{1,n}(\tau)u+L_{2,n}(u,\tau),
\end{aligned}
$$

$$L_{1,n}(\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{A_i}{\hat{\pi}(S_i)}(\tau-1\{Y_i\le q_1(\tau)\})$$

and

$$L_{2,n}(u,\tau)=\sum_{i=1}^n\frac{A_i}{\hat{\pi}(S_i)}\int_0^{\frac{u}{\sqrt{n}}}(1\{Y_i\le q_1(\tau)+v\}-1\{Y_i\le q_1(\tau)\})dv.$$

We aim to show that there exists

$$g_{ipw,n}(u,\tau)=-W_{ipw,n}(\tau)u+\frac{1}{2}Q_{ipw}(\tau)u^2 \tag{C.1}$$

such that (1) for each u,

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

(2) $Q_{ipw}(\tau)$ is bounded and bounded away from zero uniformly over $\tau \in \Upsilon$. In addition, as a corollary of claim (3) below, $\sup_{\tau\in\Upsilon} |W_{ipw,1,n}(\tau)| = O_p(1)$. Therefore, by Kato (2009, Theorme 2), we ave

$$\sqrt{n}(\hat{q}_1(\tau) - q_1(\tau)) = Q^{-1}_{ipw,1}(\tau)W_{ipw,1,n}(\tau) + R_{ipw,1,n}(\tau),$$

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

$$\sqrt{n}(\hat{q}_0(\tau) - q_0(\tau)) = Q^{-1}_{ipw,0}(\tau)W_{ipw,0,n}(\tau) + R_{ipw,0,n}(\tau),$$

where $\sup_{\tau\in\Upsilon} |R_{ipw,0,n}(\tau)| = o_p(1)$. Therefore,

$$\sqrt{n}(\hat{q}(\tau) - q(\tau)) = Q^{-1}_{ipw,1}(\tau)W_{ipw,1,n}(\tau) - Q^{-1}_{ipw,0}(\tau)W_{ipw,0,n}(\tau) + R_{ipw,1,n}(\tau) - R_{ipw,0,n}(\tau).$$

Last, we aim to show that, (3) uniformly over $\tau \in \Upsilon$,

$$Q^{-1}_{ipw,1}(\tau)W_{ipw,1,n}(\tau) - Q^{-1}_{ipw,0}(\tau)W_{ipw,0,n}(\tau) \rightsquigarrow \mathcal{B}_{ipw}(\tau),$$

where $\mathcal{B}_{ipw}(\tau)$ is a scalar Gaussian process with covariance kernel $\Sigma_{ipw}(\tau_1, \tau_2)$. We prove statements (1)–(3) in three steps.

Step 1. For $L_{1,n}(\tau)$, we have

$$\begin{aligned}
L_{1,n}(\tau) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{A_i}{\pi} 1\{S_i = s\}(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{A_i 1\{S_i = s\}(\hat{\pi}(s) - \pi)}{\sqrt{n}\hat{\pi}(s)\pi}(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{A_i}{\pi} 1\{S_i = s\}(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{A_i 1\{S_i = s\}D_n(s)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}} \frac{D_n(s)m_1(s,\tau)}{n(s)\sqrt{n}\hat{\pi}(s)\pi} D_n(s) - \sum_{s\in\mathcal{S}} \frac{D_n(s)m_1(s,\tau)}{\sqrt{n}\hat{\pi}(s)} \\
=& \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{A_i 1\{S_i = s\}}{\pi}\eta_{i,1}(s,\tau) + \sum_{s\in\mathcal{S}} \frac{D_n(s)}{\sqrt{n}\pi} m_1(s,\tau) + \sum_{i=1}^n \frac{m_1(S_i,\tau)}{\sqrt{n}} \\
& - \sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{A_i 1\{S_i = s\}D_n(s)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}} \frac{D_n(s)m_1(s,\tau)}{n(s)\sqrt{n}\hat{\pi}(s)\pi} D_n(s) - \sum_{s\in\mathcal{S}} \frac{D_n(s)m_1(s,\tau)}{\sqrt{n}\hat{\pi}(s)} \\
=& W_{ipw,1,n}(\tau) + R_{ipw}(\tau),
\end{aligned}$$

where

$$W_{ipw,1,n}(\tau) = \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{A_i 1\{S_i = s\}}{\pi} \eta_{i,1}(s,\tau) + \sum_{i=1}^{n} \frac{m_1(S_i,\tau)}{\sqrt{n}} \tag{C.2}$$

and

$$R_{ipw}(\tau) = -\sum_{i=1}^{n}\sum_{s\in\mathcal{S}} \frac{A_i 1\{S_i = s\} D_n(s)}{n(s)\sqrt{n}\hat{\pi}(s)\pi} \eta_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}} \frac{D_n(s) m_1(s,\tau)}{n(s)\sqrt{n}\hat{\pi}(s)\pi} D_n(s) + \sum_{s\in\mathcal{S}} \frac{D_n(s) m_1(s,\tau)}{\sqrt{n}} (\frac{1}{\pi} - \frac{1}{\hat{\pi}(s)}).$$

Because of (B.3) and the facts that $\frac{D_n(s)}{\sqrt{n}} = O_p(1)$, $\sup_{s\in\mathcal{S},\tau\in\Upsilon} |m_1(s,\tau)|$ is bounded, and

$$\frac{1}{\pi} - \frac{1}{\hat{\pi}(s)} = \frac{D_n(s)}{n(s)\pi\hat{\pi}(s)} = O_p(\frac{1}{\sqrt{n}}),$$

we have

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

For $L_{2,n}(u,\tau)$, by the same argument in (B.7), we have

$$\begin{aligned} L_{2,n}(u,\tau) =& \sum_{s\in\mathcal{S}} \frac{1}{\hat{\pi}(s)} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \int_0^{\frac{u}{\sqrt{n}}} (1\{Y_i^s(1) \le q_1(\tau) + v\} - 1\{Y_i^s(1) \le q_1(\tau) + v\}) dv \\ =& \sum_{s\in\mathcal{S}} \frac{1}{\hat{\pi}(s)} \left[\Gamma_n^s(N(s) + n_1(s),\tau) - \Gamma_n^s(N(s),\tau)\right], \end{aligned}$$

where

$$\Gamma_n^s(k,\tau) = \sum_{i=1}^{k} \int_0^{\frac{u}{\sqrt{n}}} (1\{Y_i^s(1) \le q_1(\tau) + v\} - 1\{Y_i^s(1) \le q_1(\tau) + v\}) dv.$$

By the same argument in (B.8), we can show that

$$\sup_{t\in(0,1),\tau\in\Upsilon} |\Gamma_n^s(\lfloor nt \rfloor,\tau) - \mathbb{E}\Gamma_n^s(\lfloor nt \rfloor,\tau)| = o_p(1).$$

In addition,

$$\mathbb{E}\Gamma_n^s(N(s) + n_1(s),\tau) - \mathbb{E}\Gamma_n^s(N(s),\tau) \xrightarrow{p} \frac{\pi p(s) f_1(q_1(\tau)|s) u^2}{2}.$$

Therefore,

$$\sup_{\tau\in\Upsilon}\left|L_{2,n}(u,\tau)-\frac{f_1(q_1(\tau))u^2}{2}\right|=o_p(1),$$

where we use the fact that $\hat{\pi}(s)\xrightarrow{p}\pi$ and

$$\sum_{s\in\mathcal{S}}p(s)f_1(q_1(\tau)|s)=f_1(q_1(\tau)).$$

This establishes (C.1) with $Q_{ipw,1}(\tau)=f_1(q_1(\tau))$ and $W_{ipw,n}(\tau)$ defined in (C.2).

Step 2. Statement (2) holds by Assumption 2.

Step 3. By a similar argument in Step 1, we have

$$W_{ipw,0,n}(\tau)=\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(1-A_i)1\{S_i=s\}}{1-\pi}\eta_{i,0}(s,\tau)+\sum_{i=1}^{n}\frac{m_0(S_i,\tau)}{\sqrt{n}}$$

and $Q_{ipw,0}(\tau)=f_0(q_0(\tau))$. Therefore,

$$\begin{aligned}\sqrt{n}(\hat{q}-q)=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]\\&+\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right]+R_{ipw,n}(\tau)\\=&\mathcal{W}_{n,1}(\tau)+\mathcal{W}_{n,2}(\tau)+R_{ipw,n}(\tau)\end{aligned}\tag{C.3}$$

where $\sup_{\tau\in\Upsilon}|R_{ipw,n}(\tau)|=o_p(1)$. Last, Lemma F.4 establishes that

$$(\mathcal{W}_{n,1}(\tau),\mathcal{W}_{n,2}(\tau))\rightsquigarrow(\mathcal{B}_{ipw,1}(\tau),\mathcal{B}_{ipw,2}(\tau)),$$

where $(\mathcal{B}_{ipw,1}(\tau),\mathcal{B}_{ipw,2}(\tau))$ are two mutually independent scalar Gaussian processes with covariance kernels

$$\Sigma_{ipw,1}(\tau_1,\tau_2)=\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}$$

and

$$\Sigma_{ipw,2}(\tau_1,\tau_2)=\mathbb{E}\left(\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}-\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right)\left(\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}-\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right),$$

respectively. In particular, the asymptotic variance for $\hat{q}$ is

$$\zeta_Y^2(\pi,\tau)+\zeta_S^2(\tau),$$

where $\zeta_Y^2(\pi,\tau)$ and $\zeta_S^2(\tau)$ are the same as those in the proof of Theorems 3.1 and 3.2.

D Proof of Theorem 4.1

We focus on the strata fixed effects estimator and derive the asymptotic distribution of its bootstrap counterpart. The argument for the simple QR estimator is similar but easier. Therefore, the details are omitted. Last, we will highlight parts of the derivation for the inverse propensity score estimator to show that why its bootstrap inference is valid.

Note that

$$\sqrt{n}(\hat{\beta}^w_{sfe} - \tilde{\beta}) = \arg\min_u L^w_{sfe,n}(u,\tau),$$

where

$$L^w_{sfe,n}(u,\tau) = \sum_{i=1}^n \xi_i \left[\rho_\tau(Y_i - \dot{\tilde{A}}_i^{w\prime}(\tilde{\beta}(\tau) + \frac{u}{\sqrt{n}})) - \rho_\tau(Y_i - \dot{\tilde{A}}_i^{w\prime}\tilde{\beta}(\tau))\right],$$

$\dot{\tilde{A}}_i^w = (1, \tilde{A}_i^w)'$, $\tilde{A}_i^w = A_i - \hat{\pi}^w(S_i)$, and

$$\hat{\pi}^w(s) = \frac{\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}}{\sum_{i=1}^n \xi_i 1\{S_i = s\}}.$$

Similar to the proof of Theorem 3.2, we divide the proof into two steps. In the first step, we show that there exists

$$g^w_{sfe,n}(u,\tau) = -u' W^w_{sfe,n}(\tau) + \frac{1}{2}u' Q_{sfe}(\tau) u$$

and $h^w_{sfe,n}(\tau)$ independent of u such that for each u

$$\sup_{\tau\in\Upsilon} |L^w_{sfe,n}(u,\tau) - g^w_{sfe,n}(u,\tau) - h^w_{sfe,n}(\tau)| \stackrel{p}{\longrightarrow} 0.$$

In addition, we will show that $\sup_{\tau\in\Upsilon} ||W^w_{sfe,n}(\tau)|| = O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^w_{sfe}(\tau) - \tilde{\beta}(\tau)) = [Q_{sfe}(\tau)]^{-1} W^w_{sfe,n}(\tau) + R^w_{sfe,n}(\tau),$$

where

$$\sup_{\tau\in\Upsilon} ||R^w_{sfe,n}(\tau)|| = o_p(1).$$

In the second step, we show that, conditionally on data,

$$\sqrt{n}(\hat{\beta}^w_{sfe,1}(\tau) - \hat{\beta}_{sfe,1}(\tau)) \rightsquigarrow \tilde{\mathcal{B}}_{sfe}(\tau).$$

Step 1. Following Step 1 in the proof of Theorem 3.2, we have

$$L^w_{sfe,n}(u,\tau) \equiv -L^w_{1,n}(u,\tau) + L^w_{2,n}(u,\tau),$$

where

$$\begin{aligned}
&L^w_{1,n}(u,\tau)\\
=&\sum_{i=1}^n \sum_{s\in\mathcal{S}} \xi_i A_i 1\{S_i = s\}\left(\frac{u_0}{\sqrt{n}} + (1-\hat{\pi}^w(s))\frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}^w(s))q(\tau)\right)(\tau - 1\{Y_i \le q_1(\tau)\})\\
&+\sum_{i=1}^n \sum_{s\in\mathcal{S}} \xi_i (1-A_i) 1\{S_i = s\}\left(\frac{u_0}{\sqrt{n}} - \hat{\pi}^w(s)\frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}^w(s))q(\tau)\right)(\tau - 1\{Y_i \le q_0(\tau)\})\\
\equiv& L^w_{1,1,n}(u,\tau) + L^w_{1,0,n}(u,\tau),
\end{aligned}$$

$$\begin{aligned}
&L^w_{2,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}\int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{E^w_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} (1\{Y_i \le q_1(\tau)+v\} - 1\{Y_i \le q_1(\tau)\})\,dv\\
&+\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i (1-A_i) 1\{S_i = s\}\int_0^{\frac{u'\iota_0}{\sqrt{n}} - \frac{E^w_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} (1\{Y_i \le q_0(\tau)+v\} - 1\{Y_i \le q_0(\tau)\})\,dv\\
\equiv& L^w_{2,1,n}(u,\tau) + L^w_{2,0,n}(u,\tau),
\end{aligned}$$

and $E^w_n(s) = \sqrt{n}(\hat{\pi}^w(s) - \pi)$.

Step 1.1. Recall that $\iota_1 = (1, 1-\pi)'$ and $\iota_0 = (1, -\pi)'$. In addition, denote $\hat{\pi}^w(s) - \pi = \frac{D^w_n(s)}{n^w(s)}$, where

$$D^w_n(s) = \sum_{i=1}^n \xi_i (A_i - \pi) 1\{S_i = s\} \quad \text{and} \quad n^w(s) = \sum_{i=1}^n \xi_i 1\{S_i = s\}.$$

Then, we have

$$\begin{aligned}
&L^w_{1,1,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}} \frac{u'\iota_1}{\sqrt{n}} \sum_{i=1}^n \xi_i \left[A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) + \pi 1\{S_i = s\} m_1(s,\tau)\right] + \sum_{s\in\mathcal{S}} \frac{u'\iota_2 D^w_n(s) m_1(s,\tau)}{\sqrt{n}}\\
&+ h^w_{1,1}(\tau) + R^w_{sfe,1,1}(u,\tau),
\end{aligned} \tag{D.1}$$

where $\eta_{i,1}(s,\tau) = (\tau - 1\{Y_i(1) \leq q_1(\tau)\}) - m_1(s,\tau)$,

$$h^w_{1,1}(\tau) = \sum_{s\in\mathcal{S}} (\pi - \hat{\pi}^w(s))q(\tau)\left(\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}(\tau - 1\{Y_i \leq q_1(\tau)\})\right),$$

and

$$R^w_{sfe,1,1}(u,\tau) = -\sum_{s\in\mathcal{S}} \frac{u_1 D^w_n(s)}{\sqrt{n}n^w(s)}\left\{\sum_{i=1}^n \xi_i \left[A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) + (A_i - \pi)1\{S_i = s\}m_1(s,\tau)\right]\right\}. \tag{D.2}$$

By Lemma F.5, we have

$$\sup_{\tau\in\Upsilon} |R^w_{sfe,1,1}(u,\tau)| = o_p(1).$$

Similarly, we have

$$\begin{aligned}
&L^w_{1,0,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i\left\{\frac{u'\iota_0}{\sqrt{n}}\left[(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau) + \pi 1\{S_i = s\}m_1(s,\tau)\right] - \frac{u'\iota_2}{\sqrt{n}}(A_i - \pi)1\{S_i = s\}m_0(s,\tau)\right\}\\
&+ h^w_{1,0}(\tau) + R^w_{sfe,1,0}(u,\tau),
\end{aligned} \tag{D.3}$$

where

$$\sup_{\tau\in\Upsilon} |R^w_{sfe,1,0}(u,\tau)| = o_p(1).$$

Combining (D.1) and (D.3), we have

$$\begin{aligned}
&L^w_{1,n}(u,\tau)\\
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i\Big[u'\iota_1 A_i 1\{S_i = s\}\eta_{i,1}(u,\tau) + u'\iota_0(1-A_i)1\{S_i = s\}\eta_{i,0}(u,\tau)\\
&+ u'\iota_2(A_i - \pi)1\{S_i = s\}(m_1(s,\tau) - m_0(s,\tau)) + 1\{S_i = s\}(u'\iota_1\pi m_1(s,\tau) + u'\iota_0(1-\pi)m_0(s,\tau))\Big]\\
&+ R^w_{sfe,1,1}(u,\tau) + R^w_{sfe,1,0}(u,\tau) + h^w_{1,1}(\tau) + h^w_{1,0}(\tau).
\end{aligned}$$

Furthermore, by Lemma F.6, we have

$$L^w_{2,1,n}(u,\tau) = \frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D^w_n(s)u'\iota_1}{\sqrt{n}}q(\tau) + h^w_{2,1}(\tau) + R^w_{sfe,2,1}(u,\tau) \tag{D.4}$$

and

$$L^w_{2,0,n}(u,\tau) = \frac{(1-\pi)f_0(q_0(\tau))}{2}(u'\iota_0)^2 - \sum_{s\in\mathcal{S}} f_0(q_0(\tau)|s)\frac{(1-\pi)D^w_n(s)u'\iota_0}{\sqrt{n}}q(\tau) + h^w_{2,0}(\tau) + R^w_{sfe,2,0}(u,\tau), \tag{D.5}$$

where

$$h^w_{2,1}(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2}p(s)(E^w_n(s))^2 q^2(\tau),$$

$$h^w_{2,0}(\tau) = \sum_{s\in\mathcal{S}} \frac{(1-\pi)f_0(q_0(\tau)|s)}{2}p(s)(E^w_n(s))^2 q^2(\tau),$$

$$\sup_{\tau\in\Upsilon}|R^w_{sfe,2,1}(u,\tau)| = o_p(1),$$

and

$$\sup_{\tau\in\Upsilon}|R^w_{sfe,2,0}(u,\tau)| = o_p(1).$$

Therefore,

$$\begin{aligned} L^w_{2,n}(u,\tau) =& \frac{1}{2}u'Q_{sfe}(\tau)u - \sum_{s\in\mathcal{S}} q(\tau)\left[f_1(q_1(\tau)|s)\pi u'\iota_1 + f_0(q_0(\tau)|s)(1-\pi)u'\iota_0\right]\frac{D^w_n(s)}{\sqrt{n}} \\ &+ R^w_{sfe,2,1}(u,\tau) + R^w_{sfe,2,0}(u,\tau) + h^w_{2,1}(\tau) + h^w_{2,0}(\tau). \end{aligned}$$

Combining (D.1), (D.3), (D.4), and (D.5), we have

$$L^w_{sfe,n}(u,\tau) = -u'\tilde{W}^w_{sfe,n}(\tau) + \frac{1}{2}u'Q_{sfe}u + \tilde{R}^w_{sfe,n}(u,\tau) + h^w_{sfe,n}(\tau),$$

where

$$\begin{aligned} &W^w_{sfe,n}(\tau) \\ =&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i\Big[\iota_1 A_i 1\{S_i=s\}\eta_{i,1}(s,\tau) + \iota_0(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)\Big] \\ &+ \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i\Big\{\iota_2(m_1(s,\tau)-m_0(s,\tau)) + q(\tau)\Big[f_1(q_1(\tau)|s)\pi\iota_1 + f_0(q_0(\tau)|s)(1-\pi)\iota_0\Big]\Big\} \end{aligned}$$

$$\times (A_i - \pi)1\{S_i = s\} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(\iota_1\pi m_1(S_i,\tau) + \iota_0(1-\pi)m_0(S_i,\tau)),$$

$$h^w_{sfe,n}(\tau) = h^w_{1,1}(\tau) + h^w_{1,0}(\tau) + h^w_{2,1}(\tau) + h^w_{2,0}(\tau),$$

and

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

In addition, by Lemma F.7, $\sup_{\tau\in\Upsilon}|W^w_{sfe,n}(\tau)| = O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^w_{sfe}(\tau) - \tilde{\beta}(\tau)) = [Q_{sfe}(\tau)]^{-1}W^w_{sfe,n}(\tau) + R^w_{sfe,n}(\tau),$$

where

$$\sup_{\tau\in\Upsilon}||R^w_{sfe,n}(\tau)|| = o_p(1).$$

This concludes Step 1.

Step 2. We now focus on the second element of $\hat{\beta}^w_{sfe}(\tau)$. From Step 1, we know that

$$\sqrt{n}(\hat{\beta}^w_{sfe,1}(\tau) - q(\tau)) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i\mathcal{J}_i(s,\tau) + R^w_{sfe,n,1}(\tau),$$

where

$$\begin{aligned}
\mathcal{J}_i(s,\tau) = &\left[\frac{A_i1\{S_i = s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] \\
&+ \left\{\left(\frac{1-\pi}{\pi f_1(q_1(\tau))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right)(m_1(s,\tau) - m_0(s,\tau)) \right.\\
&\left. + q(\tau)\left[\frac{f_1(q_1(\tau)|s)}{f_1(q_1(\tau))} - \frac{f_0(q_0(\tau)|s)}{f_0(q_0(\tau))}\right]\right\}(A_i - \pi)1\{S_i = s\} \\
&+ \left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)1\{S_i = s\}
\end{aligned}$$

and

$$\sup_{\tau\in\Upsilon}|R^w_{sfe,n,1}(\tau)| = o_p(1).$$

By (B.14), we have

$$\sqrt{n}(\hat{\beta}_{sfe,1}(\tau) - q(\tau)) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\mathcal{J}_i(s,\tau) + R_{sfe,n,1}(\tau),$$

where

$$\sup_{\tau\in\Upsilon}|R_{sfe,n,1}(\tau)| = o_p(1).$$

Taking the difference of the above two equations, we have

$$\sqrt{n}(\hat{\beta}^w_{sfe,1}(\tau) - \hat{\beta}_{sfe,1}(\tau)) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}(\xi_i - 1)\mathcal{J}_i(s,\tau) + R^w(\tau),$$

where

$$\sup_{\tau\in\Upsilon}|R^w(\tau)| = o_p(1).$$

Lemma F.8 shows that, conditionally on data,

$$\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}(\xi_i - 1)\mathcal{J}_i(s,\tau) \rightsquigarrow \tilde{\mathcal{B}}_{sfe}(\tau),$$

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

$$\begin{aligned}
&\tilde{\Sigma}_{sfe}(\tau_1,\tau_2)\\
=&\frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))} + \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\\
&+ \mathbb{E}\pi(1-\pi)\Bigg[(m_1(S,\tau_1) - m_0(S,\tau_1))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_1))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\right)\\
&+ q(\tau_1)\left(\frac{f_1(q(\tau_1)|S)}{f_1(q_1(\tau_1))} - \frac{f_0(q(\tau_1)|S)}{f_0(q_0(\tau_1))}\right)\Bigg]\\
&\times\left[(m_1(S,\tau_2) - m_0(S,\tau_2))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_2))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\right) + q(\tau_2)\left(\frac{f_1(q(\tau_2)|S)}{f_1(q_2(\tau_2))} - \frac{f_0(q(\tau_2)|S)}{f_0(q_0(\tau_2))}\right)\right]\\
&+ \mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))} - \frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))} - \frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right].
\end{aligned} \tag{D.6}$$

This concludes the proof for the strata fixed effects estimator. Next, we turn to the inverse propensity score weighted estimator. Denote $\hat{q}^w_j(\tau)$, $j = 0,1$ the weighted bootstrap counterpart of $\hat{q}_j(\tau)$.

We have

$$\sqrt{n}(\hat{q}_1^w(\tau) - q_1(\tau)) = \arg\min_u L_n^w(u,\tau),$$

where

$$\begin{aligned}
L_n^w(u,\tau) =& \sum_{i=1}^n \frac{\xi_i A_i}{\hat{\pi}^w(S_i)} \left[\rho_\tau(Y_i - q_1(\tau) - \frac{u}{\sqrt{n}}) - \rho_\tau(Y_i - q_1(\tau))\right] \\
\equiv& - L_{1,n}^w(\tau)u + L_{2,n}^w(u,\tau),
\end{aligned}$$

where

$$L_{1,n}^w(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{M_{ni}A_i}{\hat{\pi}^w(S_i)}(\tau - 1\{Y_i \leq q_1(\tau)\})$$

and

$$L_{2,n}^w(\tau) = \sum_{i=1}^n \frac{\xi_i A_i}{\hat{\pi}^w(S_i)} \int_0^{\frac{u}{\sqrt{n}}} (1\{Y_i \leq q_1(\tau) + v\} - 1\{Y_i \leq q_1(\tau)\})dv.$$

Recall

$$D_n^w(s) = \sum_{i=1}^n \xi_i(A_i - \pi)1\{S_i = s\}, \quad n^w(s) = \sum_{i=1}^n \xi_i 1\{S_i = s\},$$

and

$$\hat{\pi}^w(s) = \frac{\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}}{n^w(s)} = \pi + \frac{D_n^w(s)}{n^w(s)}.$$

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

$$\begin{aligned}
L_{1,n}^w(\tau) =& \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{\xi_i A_i}{\pi} 1\{S_i = s\}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\}(\hat{\pi}^w(s) - \pi)}{\sqrt{n}\hat{\pi}^w(s)\pi}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{\xi_i A_i}{\pi} 1\{S_i = s\}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s\in\mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\} D_n^w(s)}{n^w(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}} \frac{D_n^w(s) m_1(s,\tau)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} D_n^w(s) - \sum_{s\in\mathcal{S}} \frac{D_n^w(s) m_1(s,\tau)}{\sqrt{n}\hat{\pi}^w(s)} \\
=& \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{\pi}\eta_{i,1}(s,\tau) + \sum_{s\in\mathcal{S}} \frac{D_n^w(s)}{\sqrt{n}\pi} m_1(s,\tau) + \sum_{i=1}^n \frac{\xi_i m_1(S_i,\tau)}{\sqrt{n}}
\end{aligned}$$

$$
- \sum_{s \in \mathcal{S}} D_n^w(s) \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} \eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D_n^w(s) m_1(s,\tau)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} D_n^w(s) - \sum_{s \in \mathcal{S}} \frac{D_n^w(s) m_1(s,\tau)}{\sqrt{n}\hat{\pi}^w(s)}
$$

$$
= W_{ipw,1,n}^w(\tau) + R_{ipw}^w(\tau),
$$

where

$$
W_{ipw,1,n}^w(\tau) = \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{\pi} \eta_{i,1}(s,\tau) + \sum_{i=1}^n \frac{\xi_i m_1(S_i,\tau)}{\sqrt{n}} \tag{D.7}
$$

and

$$
\begin{aligned}
& R_{ipw}^w(\tau) \\
= & - \sum_{s \in \mathcal{S}} D_n^w(s) \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} \eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D_n^w(s) m_1(s,\tau)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} D_n^w(s) + \sum_{s \in \mathcal{S}} \frac{D_n^w(s) m_1(s,\tau)}{\sqrt{n}} (\frac{1}{\pi} - \frac{1}{\hat{\pi}^w(s)}).
\end{aligned}
$$

By Step 1 in the proof of F.5, we have $\sup_{s \in \mathcal{S}} |D_n^w(s)| = O_p(\sqrt{n})$. By the standard bootstrap argument, we have $\frac{n^w(s)}{n} \xrightarrow{p} p(s) > 0$. Therefore, $\hat{\pi}^w(s) - \pi \xrightarrow{p} 0$. In addition, by Step 2 in the proof of Lemma F.5, we have

$$
\sup_{\tau \in \Upsilon, s \in \mathcal{S}} |\sum_{i=1}^n \xi_i A_i 1\{S_i = s\} \eta_{i,1}(s,\tau)| = O_p(\sqrt{n}).
$$

Therefore,

$$
\sup_{\tau \in \Upsilon} |R_{ipw}^w(\tau)| = o_p(1).
$$

Similar to the strata fixed effects estimator, we can show that

$$
\sup_{\tau \in \Upsilon} |L_{2,n}^w(u,\tau) - \frac{f_1(q_1(\tau))u^2}{2}| = o_p(1).
$$

Therefore,

$$
\sqrt{n}(\hat{q}_1^w(\tau) - q_1(\tau)) = \frac{W_{ipw,1,n}^w(\tau)}{f_1(q_1(\tau))} + R_1^w(\tau),
$$

where $\sup_{\tau \in \Upsilon} |R_1^w(\tau)| = o_p(1)$. Similarly,

$$
\sqrt{n}(\hat{q}_0^w(\tau) - q_0(\tau)) = \frac{W_{ipw,0,n}^w(\tau)}{f_0(q_0(\tau))} + R_0^w(\tau),
$$

where

$$W^w_{ipw,0,n}(\tau) = \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\xi_i(1-A_i)1\{S_i=s\}}{1-\pi}\eta_{i,0}(s,\tau) + \sum_{i=1}^n \frac{\xi_i m_0(S_i,\tau)}{\sqrt{n}}$$

and $\sup_{\tau\in\Upsilon}|R^w_0(\tau)| = o_p(1)$. Therefore,

$$\begin{aligned}
&\sqrt{n}(\hat{q}^w(\tau)-\hat{q}(\tau)) \\
=&\sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i - 1)\Bigg\{\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))} \\
&+\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right]1\{S_i=s\}\Bigg\} + o_p(1),
\end{aligned}$$

where the $o_p(1)$ term holds uniformly over $\tau\in\Upsilon$. In order to show the conditional weak convergence, we only need to show the conditionally stochastic equicontinuity and finite-dimensional convergence. The former can be shown in the same manner as the strata fixed effects estimator. For the latter, we note that

$$\begin{aligned}
&\frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=1}^n\Bigg\{\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))} + \left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right]1\{S_i=s\}\Bigg\}^2 \\
=&\sum_{s\in\mathcal{S}}\frac{1}{n}\sum_{i=1}^n\left\{\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right\}^2 + \sum_{s\in\mathcal{S}}\frac{1}{n}\sum_{i=1}^n\left\{\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}\right\}^2 \\
&+\sum_{s\in\mathcal{S}}\frac{1}{n}\sum_{i=1}^n\left\{\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right]1\{S_i=s\}\right\}^2 \\
&+\sum_{s\in\mathcal{S}}\frac{2}{n}\sum_{i=1}^n\left\{\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}\right\}\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right] \\
&-\sum_{s\in\mathcal{S}}\frac{2}{n}\sum_{i=1}^n\left\{\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right\}\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right] \\
\xrightarrow{p}&\zeta_Y^2(\pi,\tau) + \zeta_S^2(\tau).
\end{aligned}$$

Note that the RHS of the above display is the same as the asymptotic variance of the original estimator $\hat{q}(\tau)$. This concludes the proof.

E Proof of Theorem 5.1

We focus on the strata fixed effects estimator and aim to show that, uniformly over $\tau\in\Upsilon$ and conditionally on data,

$$\sqrt{n}(\hat{\beta}^*_{sfe,1}(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}_{sfe}(\tau).$$

The results for $\hat{\beta}_1^*(\tau)$ and $\hat{q}^*(\tau)$ can be established by a similar but simpler argument.

Recall the definition of $\tilde{\beta}(\tau) = (\tilde{\beta}_0(\tau), \tilde{\beta}_1(\tau))'$ in the proof of Theorem 3.2. Let $u_0 = \sqrt{n}(b_0 - \tilde{\beta}_0(\tau))$, $u_1 = \sqrt{n}(b_1 - \tilde{\beta}_1(\tau))$ and $u = (u_0, u_1)' \in \Re^2$. Then,

$$\sqrt{n}(\hat{\beta}^*_{sfe}(\tau) - \tilde{\beta}(\tau)) = \arg\min_u L^*_{sfe,n}(u, \tau),$$

where

$$L^*_{sfe,n}(u,\tau) = \sum_{i=1}^{n} \left[\rho_\tau(Y_i^* - \dot{\breve{A}}_i^{*\prime}(\tilde{\beta}(\tau) + \frac{u}{\sqrt{n}})) - \rho_\tau(Y_i^* - \breve{A}_i^{*\prime}\tilde{\beta}(\tau)) \right]$$

and $\breve{A}_i^* = (1, A_i^* - \pi)'$. Following the proof of Theorem 3.2, we divide the current proof into two steps. In the first step, we show that there exist

$$g^*_{sfe,n}(u,\tau) = -u' W^*_{sfe,n}(\tau) + \frac{1}{2} u' Q_{sfe}(\tau) u$$

and $h^*_{sfe,n}(\tau)$ independent of u such that for each u

$$\sup_{\tau \in \Upsilon} |L^*_{sfe,n}(u,\tau) - g^*_{sfe,n}(u,\tau) - h^*_{sfe,n}(\tau)| \xrightarrow{p} 0.$$

In addition, we show that $\sup_{\tau \in \Upsilon} ||W^*_{sfe,n}(\tau)|| = O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^*_{sfe}(\tau) - \tilde{\beta}(\tau)) = [Q_{sfe}(\tau)]^{-1} W^*_{sfe,n}(\tau) + R^*_{sfe,n}(\tau),$$

where

$$\sup_{\tau \in \Upsilon} ||R^*_{sfe,n}(\tau)|| = o_p(1).$$

In the second step, we show that, conditionally on data,

$$\sqrt{n}(\hat{\beta}^*_{sfe,1}(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}_{sfe}(\tau).$$

Step 1. Following Step 1 in the proof of Theorem 3.2, we have

$$L^*_{sfe,n}(u,\tau) \equiv -L^*_{1,n}(u,\tau) + L^*_{2,n}(u,\tau),$$

where

$$L^*_{1,n}(u,\tau)$$

$$
\begin{aligned}
=&\sum_{i=1}^{n}\sum_{s\in\mathcal{S}} A_i^* 1\{S_i^*=s\}\left(\frac{u_0}{\sqrt{n}}+(1-\hat{\pi}^*(s))\frac{u_1}{\sqrt{n}}+(\pi-\hat{\pi}^*(s))q(\tau)\right)(\tau-1\{Y_i^*\le q_1(\tau)\})\\
&+\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}(1-A_i^*) 1\{S_i^*=s\}\left(\frac{u_0}{\sqrt{n}}-\hat{\pi}^*(s)\frac{u_1}{\sqrt{n}}+(\pi-\hat{\pi}^*(s))q(\tau)\right)(\tau-1\{Y_i^*\le q_0(\tau)\})\\
\equiv& L_{1,1,n}^*(u,\tau)+L_{1,0,n}^*(u,\tau),
\end{aligned}
$$

$$
\begin{aligned}
&L_{2,n}^*(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\sum_{i=1}^{n} A_i^* 1\{S_i^*=s\}\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{E_n^*(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}(1\{Y_i^*\le q_1(\tau)+v\}-1\{Y_i^*\le q_1(\tau)\})\,dv\\
&+\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}(1-A_i^*) 1\{S_i^*=s\}\int_0^{\frac{u'\iota_0}{\sqrt{n}}-\frac{E_n^*(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}(1\{Y_i^*\le q_0(\tau)+v\}-1\{Y_i^*\le q_0(\tau)\})\,dv\\
\equiv& L_{2,1,n}^*(u,\tau)+L_{2,0,n}^*(u,\tau),
\end{aligned}
$$

and $E_n^*(s)=\sqrt{n}(\hat{\pi}^*(s)-\pi)$.

Step 1.1. Recall that $\iota_1=(1,1-\pi)'$ and $\iota_0=(1,-\pi)'$. In addition, $\hat{\pi}^*(s)-\pi=\frac{D_n^*(s)}{n^*(s)}$. Then,

$$
\begin{aligned}
&L_{1,1,n}^*(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_1}{\sqrt{n}}\sum_{i=1}^{n}\left[A_i^* 1\{S_i^*=s\}\eta_{i,1}^*(s,\tau)+(A_i^*-\pi)1\{S_i^*=s\}m_1(s,\tau)+\pi 1\{S_i^*=s\}m_1(s,\tau)\right]\\
&-\sum_{s\in\mathcal{S}}\frac{u_1 D_n^*(s)\pi m_1(s,\tau)}{\sqrt{n}}+h_{1,1}^*(\tau)+R_{sfe,1,1}^*(u,\tau),
\end{aligned}
\tag{E.1}
$$

where $\eta_{i,1}^*(s,\tau)=(\tau-1\{Y_i^*(1)\le q_1(\tau)\})-m_1(s,\tau)$,

$$
h_{1,1}^*(\tau)=\sum_{s\in\mathcal{S}}(\pi-\hat{\pi}^*(s))q(\tau)\left(\sum_{i=1}^{n}A_i^* 1\{S_i^*=s\}(\tau-1\{Y_i^*\le q_1(\tau)\})\right),
$$

and

$$
R_{sfe,1,1}^*(u,\tau)=-\sum_{s\in\mathcal{S}}\frac{u_1 D_n^*(s)}{\sqrt{n}n^*(s)}\left\{\sum_{i=1}^{n}A_i^* 1\{S_i^*=s\}\eta_{i,1}^*(s,\tau)+(A_i^*-\pi)1\{S_i^*=s\}m_1(s,\tau)\right\}. \tag{E.2}
$$

Lemma F.9 shows that

$$
\sup_{\tau\in\Upsilon}|R_{sfe,1,1}^*(u,\tau)|=O_p(1/\sqrt{n}).
$$

Similarly, we have

$$
\begin{aligned}
& L^*_{1,0,n}(u,\tau) \\
=& \sum_{s\in\mathcal{S}} \frac{u'\iota_0}{\sqrt{n}} \sum_{i=1}^{n} \left[(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,1}(s,\tau) - (A_i^*-\pi)1\{S_i^*=s\}m_0(s,\tau) + (1-\pi)1\{S_i^*=s\}m_0(s,\tau)\right] \\
& - \sum_{s\in\mathcal{S}} \frac{u_1 D_n^*(s)(1-\pi)m_0(s,\tau)}{\sqrt{n}} + h^*_{1,0}(\tau) + R^*_{sfe,1,0}(u,\tau), \qquad \text{(E.3)}
\end{aligned}
$$

where

$$
h^*_{1,0}(\tau) = \sum_{s\in\mathcal{S}} (\pi - \hat{\pi}^*(s))q(\tau)\left(\sum_{i=1}^{n}(1-A_i^*)1\{S_i^*=s\}(\tau - 1\{Y_i^*\le q_0(\tau)\})\right),
$$

and

$$
R^*_{sfe,1,0}(u,\tau) = -\sum_{s\in\mathcal{S}} \frac{u_1 D_n^*(s)}{\sqrt{n}n^*(s)} \left\{\sum_{i=1}^{n}(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,0}(s,\tau) - (A_i^*-\pi)1\{S_i^*=s\}m_0(s,\tau)\right\}. \qquad \text{(E.4)}
$$

Lemma F.9 shows that

$$
\sup_{\tau\in\Upsilon} |R^*_{sfe,1,0}(u,\tau)| = O_p(1/\sqrt{n}).
$$

Therefore,

$$
\begin{aligned}
L^*_{1,n}(u,\tau) =& \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{S}} \sum_{i=1}^{n} \left[u'\iota_1 A_i^* 1\{S_i^*=s\}\eta^*_{i,1}(s,\tau) + u'\iota_0(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,0}(s,\tau)\right] \\
& + \sum_{s\in\mathcal{S}} u'\iota_2 \frac{D_n^*(s)}{\sqrt{n}} \left(m_1(s,\tau) - m_0(s,\tau)\right) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (u'\iota_1 \pi m_1(S_i^*,\tau) + u'\iota_0(1-\pi)m_0(S_i^*,\tau)) \\
& + R^*_{sfe,1,1}(u,\tau) + R^*_{sfe,1,0}(u,\tau) + h_{1,1}(\tau) + h_{1,0}(\tau).
\end{aligned}
$$

Furthermore, by Lemma F.10, we have

$$
L^*_{2,1,n}(u,\tau) = \frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s) \frac{\pi D_n^*(s) u'\iota_1}{\sqrt{n}} q(\tau) + h^*_{2,1}(\tau) + R^*_{sfe,2,1}(u,\tau) \qquad \text{(E.5)}
$$

and

$$L^*_{2,0,n}(u,\tau) = \frac{(1-\pi)f_0(q_0(\tau))}{2}(u'\iota_0)^2 - \sum_{s\in\mathcal{S}} f_0(q_0(\tau)|s)\frac{(1-\pi)D^*_n(s)u'\iota_0}{\sqrt{n}}q(\tau) + h^*_{2,0}(\tau) + R^*_{sfe,2,0}(u,\tau), \tag{E.6}$$

where

$$h^*_{2,1}(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2}p(s)(E^*_n(s))^2 q^2(\tau),$$

$$h^*_{2,0}(\tau) = \sum_{s\in\mathcal{S}} \frac{(1-\pi)f_0(q_0(\tau)|s)}{2}p(s)(E^*_n(s))^2 q^2(\tau),$$

$$\sup_{\tau\in\Upsilon} |R^*_{sfe,2,1}(u,\tau)| = o_p(1),$$

and

$$\sup_{\tau\in\Upsilon} |R^*_{sfe,2,0}(u,\tau)| = o_p(1).$$

Therefore,

$$\begin{aligned} L^*_{2,n}(u,\tau) =& \frac{1}{2}u'Q_{sfe}(\tau)u - \sum_{s\in\mathcal{S}} q(\tau)\left[f_1(q_1(\tau)|s)\pi u'\iota_1 + f_0(q_0(\tau)|s)(1-\pi)u'\iota_0\right]\frac{D^*_n(s)}{\sqrt{n}} \\ &+ R^*_{sfe,2,1}(u,\tau) + R^*_{sfe,2,0}(u,\tau) + h^*_{2,1}(\tau) + h^*_{2,0}(\tau). \end{aligned}$$

Combining (E.1), (E.3), (E.5), and (E.6), we have

$$L^*_{sfe,n}(u,\tau) = -u'W^*_{sfe,n}(\tau) + \frac{1}{2}u'Q_{sfe}u + \tilde{R}^*_{sfe,n}(u,\tau) + h^*_{sfe,n}(\tau),$$

where

$$\begin{aligned} &W^*_{sfe,n}(\tau) \\ =&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n\left[\iota_1 A^*_i 1\{S^*_i = s\}\eta^*_{i,1}(s,\tau) + \iota_0(1-A^*_i)1\{S^*_i = s\}\eta^*_{i,0}(s,\tau)\right] \\ &+ \sum_{s\in\mathcal{S}}\left\{\iota_2(m_1(s,\tau) - m_0(s,\tau)) + q(\tau)\left[f_1(q_1(\tau)|s)\pi\iota_1 + f_0(q_0(\tau)|s)(1-\pi)\iota_0\right]\right\}\frac{D^*_n(s)}{\sqrt{n}} \end{aligned}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\iota_1\pi m_1(S_i^*,\tau)+\iota_0(1-\pi)m_0(S_i^*,\tau)),$$

$$h^*_{sfe,n}(\tau)=h^*_{1,1}(\tau)+h^*_{1,0}(\tau)+h^*_{2,1}(\tau)+h^*_{2,0}(\tau),$$

and

$$\sup_{\tau\in\Upsilon}|\tilde{R}^*_{sfe,n}(u,\tau)|=o_p(1).$$

By Lemma F.11, $\sup_{\tau\in\Upsilon}|W^*_{sfe,n}(\tau)|=O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^*_{sfe}(\tau)-\tilde{\beta}(\tau))=[Q_{sfe}(\tau)]^{-1}W^*_{sfe,n}(\tau)+R^*_{sfe,n}(\tau),$$

where

$$\sup_{\tau\in\Upsilon}||R^*_{sfe,n}(\tau)||=o_p(1).$$

This concludes Step 1.

Step 2. We now focus on the second element of $\hat{\beta}^*_{sfe}(\tau)$. From Step 1, we know that

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}^*_{sfe,1}(\tau)-q(\tau))\\
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i^*1\{S_i^*=s\}\eta^*_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]\\
&+\sum_{s\in\mathcal{S}}\left\{\left(\frac{1-\pi}{\pi f_1(q_1(\tau))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right)(m_1(s,\tau)-m_0(s,\tau))+q(\tau)\left[\frac{f_1(q_1(\tau)|s)}{f_1(q_1(\tau))}-\frac{f_0(q_0(\tau)|s)}{f_0(q_0(\tau))}\right]\right\}\frac{D_n^*(s)}{\sqrt{n}}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m_1(S_i^*,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i^*,\tau)}{f_0(q_0(\tau))}\right)+R^*_{sfe,n,1}(\tau)\\
\equiv&W^*_{sfe,n,1}(\tau)+W^*_{sfe,n,2}(\tau)+W^*_{sfe,n,3}(\tau)+R^*_{sfe,n,1}(\tau),
\end{aligned}$$

where

$$\sup_{\tau\in\Upsilon}|R^*_{sfe,n,1}(\tau)|=o_p(1).$$

By (C.3), we have

$$\begin{aligned}
&\sqrt{n}(\hat{q}(\tau)-q(\tau))\\
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)+R_{ipw,n}(\tau)\\
\equiv&\mathcal{W}_{n,1}(\tau)+\mathcal{W}_{n,2}(\tau)+R_{ipw,n}(\tau),
\end{aligned}
$$

where

$$\sup_{\tau\in\Upsilon}|R_{ipw,n}(\tau)|=o_p(1).$$

Taking the difference of the above two equations, we have

$$\sqrt{n}(\hat{\beta}^*_{sfe,1}(\tau)-\hat{q}(\tau))=(W^*_{sfe,n,1}(\tau)-\mathcal{W}_{n,1}(\tau))+W^*_{sfe,n,2}(\tau)+(W^*_{sfe,n,3}(\tau)-\mathcal{W}_{n,2}(\tau))+R^*(\tau), \tag{E.7}$$

where

$$\sup_{\tau\in\Upsilon}|R^*(\tau)|=o_p(1).$$

Lemma F.11 shows that, conditionally on data,

$$(W^*_{sfe,n,1}(\tau)-\mathcal{W}_{n,1}(\tau)),W^*_{sfe,n,2}(\tau),(W^*_{sfe,n,3}(\tau)-\mathcal{W}_{n,2}(\tau))\rightsquigarrow(\mathcal{B}_1(\tau),\mathcal{B}_2(\tau),\mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau),\mathcal{B}_2(\tau),\mathcal{B}_3(\tau))$ are three independent Gaussian processes and $\sum_{j=1}^{3}\mathcal{B}_j(\tau)\stackrel{d}{=}\mathcal{B}_{sfe}(\tau)$. This concludes the proof.

F Technical Lemmas

Lemma F.1. *Let S_k be the k-th partial sum of Banach space valued independent identically distributed random variables, then*

$$\mathbb{P}(\max_{1\le k\le n}||S_k||\ge\varepsilon)\le 3\max_{1\le k\le n}\mathbb{P}(||S_k||\ge\varepsilon/3).$$

When S_k takes values on $\Re$, Lemma F.1 is Peña, Lai, and Shao (2008, Exercise 2.3).

Proof. First suppose $\max_k\mathbb{P}(||S_n-S_k||\ge 2\varepsilon/3)\le 2/3$. In addition, define

$$A_k=\{||S_k||\ge\varepsilon,||S_j||<\varepsilon,1\le j<k\}.$$

Then,

$$\begin{aligned}\mathbb{P}(\max_k ||S_k|| \geq \varepsilon) \leq & \mathbb{P}(||S_n|| \geq \varepsilon/3) + \sum_{k=1}^{n} \mathbb{P}(||S_n|| \leq \varepsilon/3, A_k) \\ \leq & \mathbb{P}(||S_n|| \geq \varepsilon/3) + \sum_{k=1}^{n} \mathbb{P}(||S_n - S_k|| \geq 2\varepsilon/3)\mathbb{P}(A_k) \\ \leq & \mathbb{P}(||S_n|| \geq \varepsilon/3) + \frac{2}{3}\mathbb{P}(\max_k ||S_k|| \geq \varepsilon).\end{aligned}$$

This implies,

$$\mathbb{P}(\max_k ||S_k|| \geq \varepsilon) \leq 3\mathbb{P}(||S_n|| \geq \varepsilon/3).$$

On the other hand, if $\max_k \mathbb{P}(||S_n - S_k|| \geq 2\varepsilon/3) > 2/3$, then there exists k_0 such that $\mathbb{P}(||S_n - S_{k_0}|| \geq 2\varepsilon/3) > 2/3$. Thus,

$$\mathbb{P}(||S_n|| \geq \varepsilon/3) + \mathbb{P}(||S_{k_0}|| \geq \varepsilon/3) \geq 2/3.$$

This implies,

$$3 \max_{1\leq k\leq n} \mathbb{P}(||S_k|| \geq \varepsilon/3) \geq 3\max(\mathbb{P}(||S_n|| \geq \varepsilon/3), \mathbb{P}(||S_{k_0}|| \geq \varepsilon/3)) \geq 1 \geq \mathbb{P}(\max_{1\leq k\leq n} ||S_k|| \geq \varepsilon).$$

This concludes the proof. □

Lemma F.2. *Let $W_{n,j}(\tau)$, $j = 1, 2, 3$ be defined as in (A.3). If Assumptions in Theorem 3.1 hold, then uniformly over $\tau \in \Upsilon$,*

$$(W_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) \rightsquigarrow (\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$ are three independent two-dimensional Gaussian processes with covariance kernels $\Sigma_1(\tau_1, \tau_2)$, $\Sigma_2(\tau_1, \tau_2)$, and $\Sigma_3(\tau_1, \tau_2)$, respectively. The expressions for the three kernels are derived in the proof below.

Proof. We follow the general argument in the proof of Bugni et al. (2018a, Lemma B.2). We divide the proof into two steps. In the first step, we show that

$$(W_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) \stackrel{d}{=} (W^{\star}_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) + o_p(1),$$

where the $o_p(1)$ term holds uniformly over $\tau \in \Upsilon$, $W^{\star}_{n,1}(\tau) \perp\!\!\!\perp (W_{n,2}(\tau), W_{n,3}(\tau))$, and, uniformly over $\tau \in \Upsilon$,

$$W^{\star}_{n,1}(\tau) \rightsquigarrow \mathcal{B}_1(\tau).$$

In the second step, we show that

$$(W_{n,2}(\tau), W_{n,3}(\tau)) \rightsquigarrow (\mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$$

uniformly over $\tau \in \Upsilon$ and $\mathcal{B}_2(\tau) \perp\!\!\!\perp \mathcal{B}_3(\tau)$.

Step 1. Let $\tilde{\eta}_{i,j}(s,\tau) = \tau - 1\{Y_i^s(j) \leq q_j(\tau)\} - m_j(s,\tau)$, for $j = 0,1$, where $\{Y_i^s(0), Y_i^s(1)\}_{i\geq 1}$ are the same as defined in Step 1 in the proof of Theorem 3.1. In addition, denote

$$\tilde{W}_{n,1}(\tau) = e_1 \sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{1}{\sqrt{n}}\tilde{\eta}_{i,1}(s,\tau) + e_0 \sum_{s\in\mathcal{S}} \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{1}{\sqrt{n}}\tilde{\eta}_{i,0}(s,\tau),$$

where $n(s) = \sum_{i=1}^n 1\{S_i = s\}$. Then, we have

$$\{W_{n,1}(\tau) | \{A_i, S_i\}_{i=1}^n\} \stackrel{d}{=} \{\tilde{W}_{n,1}(\tau) | \{A_i, S_i\}_{i=1}^n\}.$$

Because both $W_{n,2}(\tau)$ and $W_{n,3}(\tau)$ are only functions of $\{A_i, S_i\}_{i=1}^n$, we have

$$(W_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) \stackrel{d}{=} (\tilde{W}_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)).$$

Let

$$W^{\star}_{n,1}(\tau) = e_1 \sum_{s\in\mathcal{S}} \sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor} \frac{1}{\sqrt{n}}\tilde{\eta}_{i,1}(s,\tau) + e_0 \sum_{s\in\mathcal{S}} \sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} \frac{1}{\sqrt{n}}\tilde{\eta}_{i,0}(s,\tau).$$

Note that $W^{\star}_{n,1}(\tau)$ is a function of only $(Y_i^s(1), Y_i^s(0))_{i\geq 1}$, which is independent of $\{A_i, S_i\}_{i=1}^n$ by construction. Therefore, $W^{\star}_{n,1}(\tau) \perp\!\!\!\perp (W_{n,2}(\tau), W_{n,3}(\tau))$.

Furthermore, note that

$$\frac{N(s)}{n} \xrightarrow{p} F(s), \quad \frac{n_1(s)}{n} \xrightarrow{p} \pi p(s), \quad \text{and} \quad \frac{n(s)}{n} \xrightarrow{p} p(s).$$

Denote $\Gamma_{n,j}(s,t,\tau) = \sum_{i=1}^{\lfloor nt\rfloor} \frac{1}{\sqrt{n}}\tilde{\eta}_{i,j}(s,\tau)$. In order to show $\sup_{\tau\in\Upsilon} |\tilde{W}_{n,1}(\tau) - W^{\star}_{n,1}(\tau)| = o_p(1)$ and $W^{\star}_{n,1}(\tau) \rightsquigarrow \mathcal{B}_1(\tau)$, it suffices to show that, (1) for $j = 0,1$ and $s \in \mathcal{S}$, the stochastic processes

$$\{\Gamma_{n,j}(s,t,\tau) : t \in (0,1), \tau \in \Upsilon\}$$

in stochastically equicontinuous; and (2) $W^{\star}_{n,1}(\tau)$ converges to $\mathcal{B}_1(\tau)$ in finite dimension.

Claim (1). We want to bound

$$\sup |\Gamma_{n,j}(s,t_2,\tau_2) - \Gamma_{n,j}(s,t_1,\tau_1)|,$$

where supremum is taken over $0 < t_1 < t_2 < t_1+\varepsilon < 1$ and $\tau_1 < \tau_2 < \tau_1+\varepsilon$ such that $\tau_1, \tau_1+\varepsilon \in \Upsilon$. Note that,

$$\begin{aligned}
&\sup |\Gamma_{n,j}(s,t_2,\tau_2) - \Gamma_{n,j}(s,t_1,\tau_1)| \\
\leq &\sup_{0<t_1<t_2<t_1+\varepsilon<1, \tau\in\Upsilon} |\Gamma_{n,j}(s,t_2,\tau) - \Gamma_{n,j}(s,t_1,\tau)| + \sup_{t\in(0,1),\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon} |\Gamma_{n,j}(s,t,\tau_2) - \Gamma_{n,j}(s,t,\tau_1)|.
\end{aligned} \tag{F.1}$$

Let $m = \lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor \leq \lfloor n\varepsilon \rfloor + 1$. Then, for an arbitrary $\delta > 0$, by taking $\varepsilon = \delta^4$, we have

$$\begin{aligned}
&\mathbb{P}(\sup_{0<t_1<t_2<t_1+\varepsilon<1,\tau\in\Upsilon} |\Gamma_{n,j}(s,t_2,\tau) - \Gamma_{n,j}(s,t_1,\tau)| \geq \delta) \\
=&\mathbb{P}(\sup_{0<t_1<t_2<t_1+\varepsilon<1,\tau\in\Upsilon} |\sum_{i=\lfloor nt_1\rfloor+1}^{i=\lfloor nt_2\rfloor} \tilde{\eta}_{i,j}(s,\tau)| \geq \sqrt{n}\delta) \\
=&\mathbb{P}(\sup_{0<t\leq\varepsilon,\tau\in\Upsilon} |\sum_{i=1}^{\lfloor nt\rfloor} \tilde{\eta}_{i,j}(s,\tau)| \geq \sqrt{n}\delta) \\
\leq&\mathbb{P}(\max_{1\leq k\leq\lfloor n\varepsilon\rfloor} \sup_{\tau\in\Upsilon} |S_k(\tau)| \geq \sqrt{n}\delta) \\
\leq&\frac{270\mathbb{E}\sup_{\tau\in\Upsilon} |\sum_{i=1}^{\lfloor n\varepsilon\rfloor} \tilde{\eta}_{i,j}(s,\tau)|}{\sqrt{n}\delta} \\
\lesssim&\frac{\sqrt{n\varepsilon}}{\sqrt{n}\delta} \lesssim \delta,
\end{aligned}$$

where in the first inequality, $S_k(\tau) = \sum_{i=1}^{k} \tilde{\eta}_{i,j}(s,\tau)$ and the second inequality holds due to the same argument in (A.2). For the third inequality, denote

$$\mathcal{F} = \{\tilde{\eta}_{i,j}(s,\tau) : \tau \in \Upsilon\}$$

with an envelope function $F = 2$. In addition, because $\mathcal{F}$ is a VC-class with a fixed VC-index, we have

$$J(1,\mathcal{F}) < \infty,$$

where

$$J(\delta, \mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\varepsilon ||F||_{Q,2}, \mathcal{F}, L_2(Q))} d\varepsilon,$$

$N(\varepsilon ||F||_{Q,2}, \mathcal{F}, L_2(Q))$ is the covering number, and the supremum is taken over all discrete probability measures Q. Therefore, by van der Vaart and Wellner (1996, Theorem 2.14.1)

$$\frac{270\mathbb{E}\sup_{\tau\in\Upsilon} |\sum_{i=1}^{\lfloor n\varepsilon \rfloor} \tilde{\eta}_{i,j}(s,\tau)|}{\sqrt{n}\delta} \lesssim \frac{\sqrt{\lfloor n\varepsilon \rfloor}\left[\mathbb{E}\sqrt{\lfloor n\varepsilon \rfloor}||\mathbb{P}_{\lfloor n\varepsilon \rfloor} - \mathbb{P}||_{\mathcal{F}}\right]}{\sqrt{n}\delta} \lesssim \frac{\sqrt{\lfloor n\varepsilon \rfloor} J(1, \mathcal{F})}{\sqrt{n}\delta}.$$

For the second term on the RHS of (F.1), by taking $\varepsilon = \delta^4$, we have

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in(0,1),\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon} |\Gamma_{n,j}(s,t,\tau_2) - \Gamma_{n,j}(s,t,\tau_1)| \geq \delta) \\
=&\mathbb{P}(\max_{1\leq k\leq n} \sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon} |S_k(\tau_1,\tau_2)| \geq \sqrt{n}\delta) \\
\leq&\frac{270\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon} |\sum_{i=1}^n (\tilde{\eta}_{i,j}(s,\tau_2) - \tilde{\eta}_{i,j}(s,\tau_1))|}{\sqrt{n}\delta} \lesssim \delta\sqrt{\log(\frac{C}{\delta^2})},
\end{aligned}$$

where in the first equality, $S_k(\tau_1, \tau_2) = \sum_{i=1}^k (\tilde{\eta}_{i,j}(s,\tau_2) - \tilde{\eta}_{i,j}(s,\tau_1))$ and the first inequality follows the same argument as in (A.2). For the last inequality, denote

$$\mathcal{F} = \{\tilde{\eta}_{i,j}(s,\tau_2) - \tilde{\eta}_{i,j}(s,\tau_1) : \tau_1, \tau_2 \in \Upsilon, \tau_1 < \tau_2 < \tau_1 + \varepsilon\}$$

with a constant envelope function $F = C$ and

$$\sigma^2 = \sup_{f\in\mathcal{F}} \mathbb{E}f^2 \in [c_1\varepsilon, c_2\varepsilon],$$

for some constant $0 < c_1 < c_2 < \infty$. Last, $\mathcal{F}$ is nested by some VC class with a fixed VC index. Therefore, by Chernozhukov et al. (2014, Corollary 5.1),

$$\begin{aligned}
&\frac{270\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon} |\sum_{i=1}^n (\tilde{\eta}_{i,j}(s,\tau_2) - \tilde{\eta}_{i,j}(s,\tau_1))|}{\sqrt{n}\delta} \\
\lesssim&\frac{\sqrt{n}\mathbb{E}||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}}}{\delta} \lesssim \sqrt{\frac{\sigma^2\log(\frac{C}{\sigma})}{\delta^2}} + \frac{C\log(\frac{C}{\sigma})}{\sqrt{n}\delta} \lesssim \delta\sqrt{\log(\frac{C}{\delta^2})},
\end{aligned}$$

where the last inequality holds by letting n be sufficiently large. Note that $\delta\sqrt{\log(\frac{C}{\delta^2})} \to 0$ as $\delta \to 0$. This concludes the proof of Claim (1).

Claim (2). For a single τ, by the triangular CLT,

$$W^{\star}_{n,1}(\tau) \rightsquigarrow N(0, \Sigma_1(\tau)),$$

where $\Sigma_1(\tau) = \pi[\tau(1-\tau) - \mathbb{E}m_1^2(S,\tau)]e_1e_1' + (1-\pi)[\tau(1-\tau) - \mathbb{E}m_0^2(S,\tau)]e_0e_0'$. The convergence in finite dimension can be proved by using the Cramér-Wold device. In particular, we can show that the covariance kernel is

$$\begin{aligned}\Sigma_1(\tau_1,\tau_2) =&\pi[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)]e_1e_1' \\ &+ (1-\pi)[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)]e_0e_0'.\end{aligned}$$

This concludes the proof of Claim (2), and thus leads to the desired results in Step 1.

Step 2. We first consider the marginal distributions for $W_{n,2}(\tau)$ and $W_{n,3}(\tau)$. For $W_{n,2}(\tau)$, by Assumption 1 and the fact that $m_j(s,\tau)$ is continuous in $\tau \in \Upsilon$ $j = 0,1$, we have, conditionally on $\{S_i\}_{i=1}^n$,

$$W_{n,2}(\tau) = \sum_{s\in\mathcal{S}} \frac{D_n(s)}{\sqrt{n}}[e_1m_1(s,\tau) - e_0m_0(s,\tau)] \rightsquigarrow \mathcal{B}_2(\tau), \tag{F.2}$$

where $\mathcal{B}_2(\tau)$ is a two-dimensional Gaussian process with covariance kernel

$$\begin{aligned}&\Sigma_2(\tau_1,\tau_2) \\ =&\sum_{s\in\mathcal{S}} p(s)\gamma(s)\Big[e_1e_1'm_1(s,\tau_1)m_1(s,\tau_2) - e_1e_0'm_1(s,\tau_1)m_0(s,\tau_2) \\ &- e_0e_1'm_0(s,\tau_1)m_1(s,\tau_2) + e_0e_0'm_0(s,\tau_1)m_0(s,\tau_2)\Big].\end{aligned}$$

For $W_{n,3}(\tau)$, by the fact that $m_j(s,\tau)$ is continuous in $\tau \in \Upsilon$ $j = 0,1$, we have that, uniformly over $\tau \in \Upsilon$,

$$W_{n,3}(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [e_1\pi m_1(S_i,\tau) + e_0(1-\pi)m_0(S_i,\tau)] \rightsquigarrow \mathcal{B}_3(\tau), \tag{F.3}$$

where $\mathcal{B}_3(\tau)$ a two-dimensional Gaussian process with covariance kernel

$$\begin{aligned}\Sigma_3(\tau_1,\tau_2) =&e_1e_1'\pi^2\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2) + e_1e_0'\pi(1-\pi)\mathbb{E}m_1(S,\tau_1)m_0(S,\tau_2) \\ &+ e_0e_1'\pi(1-\pi)\mathbb{E}m_0(S,\tau_1)m_1(S,\tau_2) + e_0e_0'(1-\pi)^2\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2).\end{aligned}$$

In addition, we note that, for any fixed τ,

$$
\begin{aligned}
\mathbb{P}(W_{n,2}(\tau) \leq w_1, W_{n,3}(\tau) \leq w_2) =& \mathbb{E}\mathbb{P}(W_{n,2}(\tau) \leq w_1 | \{S_i\}_{i=1}^n) 1\{W_{n,3}(\tau) \leq w_2\} \\
=& \mathbb{E}\mathbb{P}(N(0, \Sigma_2(\tau,\tau)) \leq w_1) 1\{W_{n,3}(\tau) \leq w_2\} + o(1) \\
=& \mathbb{P}(N(0, \Sigma_3(\tau,\tau)) \leq w_2) \mathbb{P}(N(0, \Sigma_2(\tau,\tau)) \leq w_1) + o(1).
\end{aligned}
$$

This implies $\mathcal{B}_2(\tau) \perp\!\!\!\perp \mathcal{B}_3(\tau)$. By using the Cramér-Wold device, we can show that

$$
(W_{n,2}(\tau), W_{n,3}(\tau)) \rightsquigarrow (\mathcal{B}_2(\tau), \mathcal{B}_3(\tau))
$$

jointly in finite dimension, where by an abuse of notation, $\mathcal{B}_2(\tau)$ and $\mathcal{B}_3(\tau)$ have the same marginal distributions of those in (F.2) and (F.3), respectively, and $\mathcal{B}_2(\tau) \perp\!\!\!\perp \mathcal{B}_3(\tau)$. Last, because both $W_{n,2}(\tau)$ and $W_{n,3}(\tau)$ are tight marginally, so be the joint process $(W_{n,2}(\tau), W_{n,3}(\tau))$. This concludes the proof of Step 2, and thus the whole lemma. □

Lemma F.3. *Let $W_{sfe,n,j}(\tau)$, $j = 1, 2, 3$ be defined as in (B.14). If Assumptions in Theorem 3.2 hold, then uniformly over $\tau \in \Upsilon$,*

$$
(W_{sfe,n,1}(\tau), W_{sfe,n,2}(\tau), W_{sfe,n,3}(\tau)) \rightsquigarrow (\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau)),
$$

where $(\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau))$ are three independent two-dimensional Gaussian process with covariance kernels $\Sigma_{sfe,1}(\tau_1, \tau_2)$, $\Sigma_{sfe,2}(\tau_1, \tau_2)$, and $\Sigma_{sfe,3}(\tau_1, \tau_2)$, respectively. The expressions for the three kernels are derived in the proof below.

Proof. The proofs of weak convergence and the independence among $(\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau))$ are similar to that in Lemma F.2, and thus, are omitted. In the following, we focus on deriving the covariance kernels.

First, similar to the argument in the proof of Lemma F.2,

$$
W_{sfe,n,1}(\tau) \stackrel{d}{=} \iota_1 \sum_{s \in \mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,1}(s,\tau) + \iota_0 \sum_{s \in \mathcal{S}} \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,0}(s,\tau).
$$

Therefore,

$$
\begin{aligned}
\Sigma_1(\tau_1, \tau_2) =& \pi[\min(\tau_1, \tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S, \tau_1) m_1(S, \tau_2)] \iota_1 \iota_1' \\
&+ (1-\pi)[\min(\tau_1, \tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S, \tau_1) m_0(S, \tau_2)] \iota_0 \iota_0'.
\end{aligned}
$$

For $W_{sfe,n,2}(\tau)$, we have

$$
\Sigma_2(\tau_1, \tau_2) = \mathbb{E}\gamma(S) \Big[\iota_2 (m_1(S, \tau_1) - m_0(S, \tau_1)) + q(\tau_1) \Big(f_1(q_1(\tau_1)|S) \pi \iota_1 + f_0(q_0(\tau_1)|S)(1-\pi) \iota_0 \Big) \Big]
$$

$$\times\left[\iota_2(m_1(S,\tau_2)-m_0(S,\tau_2))+q(\tau_2)\bigg(f_1(q_1(\tau_2)|S)\pi\iota_1+f_0(q_0(\tau_2)|S)(1-\pi)\iota_0\bigg)\right]'.$$

Next, we have

$$\Sigma_3(\tau_1,\tau_2)=\mathbb{E}(\iota_1\pi m_1(S,\tau_1)+\iota_0(1-\pi)m_0(S,\tau_1))(\iota_1\pi m_1(S,\tau_2)+\iota_0(1-\pi)m_0(S,\tau_2))'.$$

In addition,

$$[Q_{sfe}(\tau)]^{-1}=\begin{pmatrix}\frac{1-\pi}{f_0(q_0(\tau))}+\frac{\pi}{f_1(q_1(\tau))} & \frac{1}{f_1(q_1(\tau))}-\frac{1}{f_0(q_0(\tau))}\\ \frac{1}{f_1(q_1(\tau))}-\frac{1}{f_0(q_0(\tau))} & \frac{1}{(1-\pi)f_0(q_0(\tau))}+\frac{1}{\pi f_1(q_1(\tau))}\end{pmatrix}.$$

Therefore,

$$\begin{aligned}
&\Sigma(\tau_1,\tau_2)\\
=&\bigg\{\frac{1}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\left[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)\right]\begin{pmatrix}\pi^2 & \pi\\ \pi & 1\end{pmatrix}\\
&+\frac{1}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\left[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)\right]\begin{pmatrix}(1-\pi)^2 & \pi-1\\ \pi-1 & 1\end{pmatrix}\bigg\}\\
&+\bigg\{\mathbb{E}\gamma(S)\bigg[(m_1(S,\tau_1)-m_0(S,\tau_1))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_1))}+\frac{1-\pi}{f_1(q_1(\tau_1))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_1))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\end{pmatrix}+q(\tau_1)\frac{f_1(q_1(\tau_1)|S)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\ 1\end{pmatrix}\\
&+q(\tau_1)\frac{f_0(q_0(\tau_1)|S)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\bigg]\times\bigg[(m_1(S,\tau_2)-m_0(S,\tau_2))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_2))}+\frac{1-\pi}{f_1(q_1(\tau_2))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_2))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\end{pmatrix}\\
&+q(\tau_2)\frac{f_1(q_1(\tau_2)|S)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\ 1\end{pmatrix}+q(\tau_2)\frac{f_0(q_0(\tau_2)|S)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\bigg]\bigg\}\\
&+\bigg\{\mathbb{E}\bigg[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\ 1\end{pmatrix}+\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\bigg]\bigg[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\ 1\end{pmatrix}+\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\ -1\end{pmatrix}\bigg]'\bigg\}.
\end{aligned}$$

□

Lemma F.4. *Let $\mathcal{W}_{n,j}(\tau)$, $j=1,2$ be defined as in (C.3). If Assumptions in Theorem 3.3 hold, then uniformly over $\tau\in\Upsilon$,*

$$(\mathcal{W}_{n,1}(\tau),\mathcal{W}_{n,2}(\tau))\rightsquigarrow(\mathcal{B}_{ipw,1}(\tau),\mathcal{B}_{ipw,2}(\tau)),$$

where $(\mathcal{B}_{ipw,1}(\tau),\mathcal{B}_{ipw,2}(\tau))$ are two independent two-dimensional Gaussian process with covariance kernels $\Sigma_{ipw,1}(\tau_1,\tau_2)$ and $\Sigma_{ipw,2}(\tau_1,\tau_2)$, respectively. The expressions for $\Sigma_{ipw,1}(\tau_1,\tau_2)$ and $\Sigma_{ipw,2}(\tau_1,\tau_2)$ are derived in the proof below.

Proof. The proof of weak convergence and the independence between $(\mathcal{B}_{ipw,1}(\tau),\mathcal{B}_{ipw,2}(\tau))$ are similar to that in Lemma F.2, and thus, are omitted. Next, we focus on deriving the covariance kernels.

First, similar to the argument in the proof of Lemma F.2,

$$\mathcal{W}_{n,1}(\tau) \stackrel{d}{=} \sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{1}{\sqrt{n}f_1(q_1(\tau))}\tilde{\eta}_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{1}{\sqrt{n}f_0(q_0(\tau))}\tilde{\eta}_{i,0}(s,\tau).$$

Because $(\tilde{\eta}_{i,1}(s,\tau),\tilde{\eta}_{i,0}(s,\tau))$ are independent across i, $n_1(s)/n \xrightarrow{p} \pi p(s)$, and $(n(s)-n_1(s))/n \xrightarrow{p} (1-\pi)p(s)$, we have

$$\Sigma_{ipw,1}(\tau_1,\tau_2) = \frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))} + \frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}.$$

Obviously,

$$\Sigma_{ipw,2}(\tau_1,\tau_2) = \mathbb{E}\left(\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))} - \frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right)\left(\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))} - \frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right),$$

□

Lemma F.5. *Recall the definition of $R^w_{sfe,1,1}(u,\tau)$ in (D.2). If Assumptions 1 and 2 hold, then*

$$\sup_{\tau\in\Upsilon}|R^w_{sfe,1,1}(u,\tau)| = o_p(1).$$

Proof. We divide the proof into two steps. In the first step, we show that $\sup_{s\in\mathcal{S}}|D^w_n(s)| = O_p(\sqrt{n})$. In the second step, we show that

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}|\sum_{i=1}^n \xi_i A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)| = O_p(\sqrt{n}). \tag{F.4}$$

Then,

$$\begin{aligned}
&\sup_{\tau\in\Upsilon}|R^w_{sfe,1,1}(u,\tau)| \\
\leq &\sum_{s\in\mathcal{S}}\frac{|u_1|}{n^w(s)}\sup_{s\in\mathcal{S}}\left|\frac{D^w_n(s)}{\sqrt{n}}\right|\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{i=1}^n \xi_i A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)\right| + \sup_{s\in\mathcal{S}}|D^w_n(s)|\right] \\
= &O_p(1/\sqrt{n}),
\end{aligned}$$

as $n^w(s)/n \xrightarrow{p} p(s) > 0$.

Step 1. Because

$$\sup_{s\in\mathcal{S}}|D_n(s)| = O_p(\sqrt{n}),$$

we only need to bound the difference $D_n^w(s) - D_n(s)$. Note that

$$n^{-1/2}D_n^w(s) - n^{-1/2}D_n(s) = n^{-1/2}\sum_{i=1}^{n}(\xi_i - 1)(A_i - \pi)1\{S_i = s\}. \tag{F.5}$$

We aim to prove that, conditionally on data, for $s \in \mathcal{S}$,

$$n^{-1/2}\sum_{i=1}^{n}(\xi_i - 1)(A_i - \pi)1\{S_i = s\} \rightsquigarrow N(0, p(s)\pi(1-\pi)) \tag{F.6}$$

and they are independent across $s \in \mathcal{S}$. The independence is straightforward because

$$\frac{1}{n}\sum_{i=1}^{n}(\xi_i - 1)^2(A_i - \pi)^2 1\{S_i = s\}1\{S_i = s'\} = 0 \quad \text{for} \quad s \neq s'.$$

For the limiting distribution, let $\mathcal{D}_n = \{Y_i, A_i, S_i\}_{i=1}^n$ denote data. According to the Lindeberg-Feller central limit theorem, (F.6) holds because (1)

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}\mathbb{E}[(\xi_i - 1)^2(A_i - \pi)^2 1\{S_i = s\}|\mathcal{D}_n] =& n^{-1}\sum_{i=1}^{n}(A_i - 2A_i\pi + \pi^2)1\{S_i = s\} \\
=& n^{-1}\sum_{i=1}^{n}(A_i - \pi - 2(A_i - \pi)\pi + \pi - \pi^2)1\{S_i = s\} \\
=& \frac{1-2\pi}{n}\sum_{i=1}^{n}(A_i - \pi)1\{S_i = s\} + \pi(1-\pi)\frac{n(s)}{n} \\
\xrightarrow{p}& \pi(1-\pi)p(s),
\end{aligned}$$

and (2) for every $\varepsilon > 0$,

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(A_i - \pi)^2 1\{S_i = s\}\mathbb{E}\left[(\xi_i - 1)^2 1\{|\xi_i - 1|(A_i - \pi)^2 1\{S_i = s\} > \varepsilon\sqrt{n}\}|\mathcal{D}_n\right] \\
\leq & 4\mathbb{E}(\xi_i - 1)^2 1\{2|\xi_i - 1| \geq \varepsilon\sqrt{n}\} \to 0,
\end{aligned}$$

where we use the fact that $|A_i - \pi|1\{S_i = s\} \leq 2$. This concludes the proof of Step 1.

Step 2. By the same rearrangement argument and the fact that $\{\xi_i\}_{i=1}^n \perp\!\!\!\perp \mathcal{D}_n$, we have

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{n}\sum_{i=1}^n \xi_i A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)\right| \stackrel{d}{=} \sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{n}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i\tilde{\eta}_{i,1}(s,\tau)\right|.$$

Let $\Gamma_{n,1}(s,t,\tau) = \sum_{i=1}^{\lfloor nt\rfloor} \frac{\xi_i\tilde{\eta}_{i,1}(s,\tau)}{\sqrt{n}}$ and $\mathcal{F} = \{\xi_i\tilde{\eta}_{i,1}(s,\tau) : \tau\in\Upsilon, s\in\mathcal{S}\}$ with envelope $F_i = C\xi_i$ and $||F_i||_{P,2} < \infty$. By Lemma F.1 and van der Vaart and Wellner (1996, Theorem 2.14.1), for any $\varepsilon > 0$, by , we can choose M sufficiently large such that

$$\begin{aligned}\mathbb{P}(\sup_{0<t\leq 1,\tau\in\Upsilon,s\in\mathcal{S}}|\Gamma_{n,1}(s,t,\tau)| \geq M) \leq & \frac{270\mathbb{E}\sup_{\tau\in\Upsilon,s\in\mathcal{S}}|\Gamma_{n,1}(s,1,\tau)|}{M} \\ = & \frac{270\mathbb{E}\sqrt{n}||\mathbb{P}_n-\mathbb{P}||_{\mathcal{F}}}{M} \lesssim \frac{J(1,\mathcal{F})||F_i||_{P,2}}{M} < \varepsilon.\end{aligned}$$

Therefore,

$$\sup_{0<t\leq 1,\tau\in\Upsilon,s\in\mathcal{S}}|\Gamma_{n,1}(s,t,\tau)| = O_p(1)$$

and

$$\begin{aligned}\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{n}\sum_{i=1}^n \xi_i A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)\right| \stackrel{d}{=} & \sup_{\tau\in\Upsilon,s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left|\Gamma_{n,1}\left(s,\frac{N(s)+n_1(s)}{n},\tau\right) - \Gamma_{n,1}\left(s,\frac{N(s)}{n},\tau\right)\right| \\ = & O_p(1/\sqrt{n}). \qquad \text{(F.7)}\end{aligned}$$

This concludes the proof of Step 2. □

Lemma F.6. *If Assumptions 1 and 2 hold, then D.4 and D.5 hold.*

Proof. We focus on (D.4). Note that

$$\begin{aligned}& L^w_{2,1,n}(u,\tau) \\ = & \sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i=s\}\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{E^w_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i(1)\leq q_1(\tau)+v\} - 1\{Y_i(1)\leq q_1(\tau)\}\right) dv \\ = & \sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i=s\}[\phi_i(u,\tau,s,E^w_n(s)) - \mathbb{E}\phi_i(u,\tau,s,E^w_n(s)|S_i=s)] \\ & + \sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i=s\}\mathbb{E}\phi_i(u,\tau,s,E^w_n(s)|S_i=s), \qquad \text{(F.8)}\end{aligned}$$

where by Lemma F.5, $E_n^w(s) = \sqrt{n}(\hat{\pi}^w(s) - \pi) = \frac{n}{n^w(s)}\frac{D_n^w(s)}{\sqrt{n}} = O_p(1)$,

$$\phi_i(u,\tau,s,e) = \int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{e}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i(1) \leq q_1(\tau)+v\} - 1\{Y_i(1) \leq q_1(\tau)\}\right) dv,$$

and $\mathbb{E}\phi_i(u,\tau,s,E_n^w(s)|S_i = s)$ is interpreted as $\mathbb{E}(\phi_i(u,\tau,s,e)|S_i = s)$ with e being evaluated at $E_n^w(s)$.

For the first term on the RHS of (F.8), by the rearrangement argument in Lemma F.2, we have

$$\begin{aligned}
&\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}[\phi_i(u,\tau,s,E_n^w(s)) - \mathbb{E}\phi_i(u,\tau,s,E_n^w(s)|S_i = s)] \\
\stackrel{d}{=}&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i[\phi_i^s(u,\tau,s,E_n^w(s)) - \mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s))],
\end{aligned}$$

where

$$\phi_i^s(u,\tau,s,e) = \int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{e}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i^s(1) \leq q_1(\tau)+v\} - 1\{Y_i^s(1) \leq q_1(\tau)\}\right) dv.$$

Similar to (B.9), we can show that, as $n \to \infty$,

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i \left[\phi_i^s(u,\tau,s,E_n^w(s)) - \mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s))\right] \right| = o_p(1). \tag{F.9}$$

For the second term in (F.8), we have

$$\begin{aligned}
&\sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}\mathbb{E}\phi_i(u,\tau,s,E_n^w(s)|S_i = s) \\
=&\sum_{s\in\mathcal{S}} \frac{\sum_{i=1}^n \xi_i \pi 1\{S_i = s\}}{n} n\mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s)) + \sum_{s\in\mathcal{S}} \frac{D_n^w(s)}{n} n\mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s)) \\
=&\sum_{s\in\mathcal{S}} \pi p(s)\left[\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1 - E_n^w(s)q(\tau))^2 + o_p(1)\right] + \sum_{s\in\mathcal{S}} \frac{D_n^w(s)}{n}\left[\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1 - E_n^w(s)q(\tau))^2 + o_p(1)\right] \\
=&\frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D_n^w(s)u'\iota_1}{\sqrt{n}}q(\tau) + h_{2,1}^w(\tau) + o_p(1),
\end{aligned} \tag{F.10}$$

where the $o_p(1)$ term holds uniformly over $(\tau,s) \in \Upsilon \times \mathcal{S}$. The second equality holds by the same calculation in (B.10) and the fact that $\sum_{i=1}^n \xi_i 1\{S_i = s\}/n \xrightarrow{p} p(s)$. The last inequality holds because $\frac{D_n^w(s)}{n} = o_p(1)$, $E_n^w(s) = \frac{n}{n^w(s)} \frac{D_n^w(s)}{\sqrt{n}} = O_p(1)$, $\frac{n}{n^w(s)} \xrightarrow{p} 1/p(s)$, and

$$h_{2,1}^w(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2} p(s)(E_n^w(s))^2 q^2(\tau).$$

Combining (F.8)–(F.10), we have

$$L_{2,1,n}^w(u,\tau) = \frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D_n^w(s)u'\iota_1}{\sqrt{n}} q(\tau) + h_{2,1}^w(\tau) + R_{sfe,2,1}^w(u,\tau),$$

where

$$h_{2,1}^w(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2} p(s)(E_n^w(s))^2 q^2(\tau)$$

and

$$\sup_{\tau\in\Upsilon} |R_{sfe,2,1}^w(u,\tau)| = o_p(1).$$

This concludes the proof.

□

Lemma F.7. *If Assumptions 1 and 2 hold, then* $\sup_{\tau\in\Upsilon} ||W_{sfe,n}^w(\tau)|| = O_p(1)$.

Proof. It suffices to show that

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) \right| = O_p(1) \tag{F.11}$$

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i (1-A_i) 1\{S_i = s\}\eta_{i,0}(s,\tau) \right| = O_p(1), \tag{F.12}$$

$$\sup_{s\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i (A_i - \pi) 1\{S_i = s\} \right| = O_p(1), \tag{F.13}$$

and

$$\sup_{\tau\in\Upsilon} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i (\iota_1 \pi m_1(S_i,\tau) + \iota_0(1-\pi) m_0(S_i,\tau)) \right\| = O_p(1). \tag{F.14}$$

Note that (F.11) holds by the argument in step 2 in the proof of Lemma F.5, (F.12) holds similarly, (F.13) holds by (F.5) and (F.6), and (F.14) holds by the usual maximal inequality, e.g., van der Vaart and Wellner (1996, Theorem 2.14.1). This concludes the proof. □

Lemma F.8. *If Assumptions 1 and 2 hold, then conditionally on data,*

$$\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}(\xi_i-1)\mathcal{J}_i(s,\tau)\rightsquigarrow\tilde{\mathcal{B}}_{sfe}(\tau),$$

where $\tilde{\mathcal{B}}_{sfe}(\tau)$ *is a Gaussian process with covariance kernel* $\tilde{\Sigma}_{sfe}(\cdot,\cdot)$ *defined in* (D.6).

Proof. In order to show the weak convergence, we only need to show (1) conditional stochastic equicontinuity and (2) conditional convergence in finite dimension. We divide the proof into two steps accordingly.

Step 1. In order to show the conditional stochastic equicontinuity, it suffices to show that, for any $\varepsilon>0$, as $n\to\infty$ followed by $\delta\to 0$,

$$\mathbb{P}_{\xi}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_i(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\stackrel{p}{\longrightarrow}0,$$

where $\mathbb{P}_{\xi}(\cdot)$ means that the probability operator is with respect to $\xi_1,\cdots,\xi_n$ and conditional on data. Note

$$\begin{aligned}
&\mathbb{E}\mathbb{P}_{\xi}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_i(s,\tau_1)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\\
=&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_i(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\\
\leq&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2)-\mathcal{J}_{i,1}(s,\tau_1))\right|\geq\varepsilon/3\right)\\
&+\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,2}(s,\tau_2)-\mathcal{J}_{i,2}(s,\tau_1))\right|\geq\varepsilon/3\right)\\
&+\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,3}(s,\tau_2)-\mathcal{J}_{i,3}(s,\tau_1))\right|\geq\varepsilon/3\right),
\end{aligned}$$

where

$$\mathcal{J}_{i,1}(s,\tau)=\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))},$$

$$\mathcal{J}_{i,2}(s,\tau)=F_1(s,\tau)(A_i-\pi)1\{S_i=s\},$$

$$F_1(s,\tau) = \left(\frac{1-\pi}{\pi f_1(q_1(\tau))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right)(m_1(s,\tau)-m_0(s,\tau))+q(\tau)\left[\frac{f_1(q_1(\tau)|s)}{f_1(q_1(\tau))} - \frac{f_0(q_0(\tau)|s)}{f_0(q_0(\tau))}\right],$$

$$\mathcal{J}_{i,3}(s,\tau) = \left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)1\{S_i = s\}.$$

Further note that

$$\sum_{i=1}^{n}(\xi_i - 1)\mathcal{J}_{i,1}(s,\tau) \stackrel{d}{=} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{(\xi_i-1)\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{(\xi_i-1)\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}$$

By the same argument in Claim (1) in the proof of Lemma F.2, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2) - \mathcal{J}_i(s,\tau_1))\right| \geq \varepsilon/3\right)\\
\leq&\frac{3\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2) - \mathcal{J}_{i,1}(s,\tau_1))\right|}{\varepsilon}\\
\leq&\frac{3\sqrt{c_2\delta\log(\frac{C}{c_1\delta})} + \frac{3C\log(\frac{C}{c_1\delta})}{\sqrt{n}}}{\varepsilon},
\end{aligned}$$

where C, $c_1 < c_2$ are some positive constants that are independent of (n,ε,δ). By letting $n\to\infty$ followed by $\delta\to 0$, the RHS vanishes.

For $\mathcal{J}_{i,2}$, we note that $F_1(s,\tau)$ is Lipschitz in τ. Therefore,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,2}(s,\tau_2) - \mathcal{J}_{i,2}(s,\tau_1))\right| \geq \varepsilon/3\right)\\
\leq&\sum_{s\in\mathcal{S}}\mathbb{P}\left(C\delta\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(A_i-\pi)1\{S_i=s\}\right|\geq\varepsilon/3\right)\to 0
\end{aligned}$$

as $n\to\infty$ followed by $\delta\to 0$, in which we use the fact that, by (F.6),

$$\sup_{s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(A_i-\pi)1\{S_i=s\}\right| = O_p(1).$$

Last, by the standard maximal inequality (e.g., van der Vaart and Wellner (1996, Theorem 2.14.1)) and the fact that

$$\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)$$

is Lipschitz in τ, we have, as $n \to \infty$ followed by $\delta \to 0$,

$$\mathbb{P}\left(\sup_{\tau_1,\tau_2 \in \Upsilon, \tau_1 < \tau_2 < \tau_1 + \delta, s \in \mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - 1)(\mathcal{J}_{i,3}(s,\tau_2) - \mathcal{J}_{i,3}(s,\tau_1)) \right| \geq \varepsilon/3 \right) \to 0$$

This concludes the proof of the conditional stochastic equicontinuity.

Step 2. We focus on the one-dimension case and aim to show that, conditionally on data, for fixed $\tau \in \Upsilon$,

$$\frac{1}{\sqrt{n}} \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} (\xi_i - 1)\mathcal{J}_i(s,\tau) \rightsquigarrow \mathcal{N}(0, \tilde{\Sigma}_{sfe}(\tau,\tau)).$$

The finite-dimensional convergence can be established similarly by the Cramér-Wold device. In view of Lindeberg-Feller central limit theorem, we only need to show that (1)

$$\frac{1}{n} \sum_{i=1}^{n} [\sum_{s \in \mathcal{S}} \mathcal{J}_i(s,\tau)]^2 \stackrel{p}{\longrightarrow} \zeta_Y^2(\pi,\tau) + \tilde{\xi}_A'^2(\pi,\tau) + \xi_S^2(\pi,\tau)$$

and (2)

$$\frac{1}{n} \sum_{i=1}^{n} [\sum_{s \in \mathcal{S}} \mathcal{J}_i(s,\tau)]^2 \mathbb{E}_\xi (\xi - 1)^2 1\{|\sum_{s \in \mathcal{S}} (\xi_i - 1)\mathcal{J}_i(s,\tau)| \geq \varepsilon\sqrt{n}\} \to 0.$$

(2) is obvious as $|\mathcal{J}_i(s,\tau)|$ is bounded. Next, we focus on (1). We have

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} [\sum_{s \in \mathcal{S}} \mathcal{J}_i(s,\tau)]^2 \\
=&\frac{1}{n} \sum_{i=1}^{n} \sum_{s \in \mathcal{S}} \Bigg\{ \left[\frac{A_i 1\{S_i = s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1 - A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)}{(1-\pi) f_0(q_0(\tau))} \right] \\
&+ F_1(s,\tau)(A_i - \pi)1\{S_i = s\} + \left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))} \right) 1\{S_i = s\} \right] \Bigg\}^2 \\
\equiv&\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\sigma_{12} + 2\sigma_{13} + 2\sigma_{23},
\end{aligned}$$

where

$$\sigma_1^2 = \frac{1}{n} \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \left[\frac{A_i 1\{S_i = s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1 - A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)}{(1-\pi) f_0(q_0(\tau))} \right]^2,$$

$$\sigma_2^2 = \frac{1}{n} \sum_{s \in \mathcal{S}} F_1^2(s,\tau) \sum_{i=1}^{n} (A_i - \pi)^2 1\{S_i = s\},$$

$$\sigma_3^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right]^2,$$

$$\sigma_{12} = \frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]F_1(s,\tau)(A_i-\pi)1\{S_i=s\},$$

$$\sigma_{13} = \frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right],$$

and

$$\sigma_{23} = \sigma_{12} = \frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}F_1(s,\tau)(A_i-\pi)1\{S_i=s\}\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right].$$

For σ_1^2, we have

$$\begin{aligned}
\sigma_1^2 =& \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}^2(s,\tau)}{\pi^2 f_1^2(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}^2(s,\tau)}{(1-\pi)^2 f_0^2(q_0(\tau))}\right] \\
\overset{d}{=}& \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{\tilde{\eta}_{i,1}^2(s,\tau)}{\pi^2 f_1^2(q_1(\tau))} + \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\frac{\tilde{\eta}_{i,0}^2(s,\tau)}{(1-\pi)^2 f_0^2(q_0(\tau))} \\
\xrightarrow{p}& \frac{\tau(1-\tau) - \mathbb{E}m_1^s(S,\tau)}{\pi f_1^2(q_1(\tau))} + \frac{\tau(1-\tau) - \mathbb{E}m_0^s(S,\tau)}{(1-\pi)f_0^2(q_0(\tau))} = \zeta_Y^2(\pi,\tau),
\end{aligned}$$

where the second equality holds due to the rearrangement argument in Lemma F.2 and the convergence in probability holds due to uniform convergence of the partial sum process.

For σ_2^2, by Assumption 1,

$$\sigma_2^2 = \frac{1}{n}\sum_{s\in\mathcal{S}}F_1^2(s,\tau)(D_n(s) - 2\pi D_n(s) + \pi(1-\pi)1\{S_i=s\}) \xrightarrow{p} \pi(1-\pi)\mathbb{E}F_1^2(S_i,\tau) = \tilde{\xi}_A'^2(\pi,\tau).$$

For σ_3^2, by the law of large number,

$$\sigma_3^2 \xrightarrow{p} \mathbb{E}\left[\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right]^2 = \xi_S^2(\pi,\tau).$$

For σ_{12}, we have

$$\sigma_{12} = \frac{1}{n}\sum_{s\in\mathcal{S}}(1-\pi)F_1(s,\tau)\sum_{i=1}^{n}\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{1}{n}\sum_{s\in\mathcal{S}}\pi F_1(s,\tau)\sum_{i=1}^{n}\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}$$

$$\stackrel{d}{=} \frac{1}{n}\sum_{s\in\mathcal{S}}(1-\pi)F_1(s,\tau)\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{1}{n}\sum_{s\in\mathcal{S}}\pi F_1(s,\tau)\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))} \stackrel{p}{\longrightarrow} 0,$$

where the last convergence holds because by Lemma F.2,

$$\frac{1}{n}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\tilde{\eta}_{i,1}(s,\tau) \stackrel{p}{\longrightarrow} 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\tilde{\eta}_{i,0}(s,\tau) \stackrel{p}{\longrightarrow} 0.$$

By the same argument, we can show that

$$\sigma_{13} \stackrel{p}{\longrightarrow} 0.$$

Last, for σ_{23}, by Assumption 1,

$$\sigma_{23} = \sum_{s\in\mathcal{S}} F_1(s,\tau)\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right]\frac{D_n(s)}{n} \stackrel{p}{\longrightarrow} 0.$$

Therefore, we have

$$\frac{1}{n}\sum_{i=1}^{n}[\sum_{s\in\mathcal{S}}\mathcal{J}_i(s,\tau)]^2 \stackrel{p}{\longrightarrow} \zeta_Y^2(\pi,\tau) + \tilde{\xi}_A'^2(\pi,\tau) + \xi_S^2(\pi,\tau).$$

□

Lemma F.9. *Recall $R^*_{sfe,1,1}(u,\tau)$ and $R^*_{sfe,1,0}(u,\tau)$ defined in (E.2) and (E.4), respectively. If Assumptions in Theorem 5.1 hold, then*

$$\sup_{\tau\in\Upsilon}|R^*_{sfe,1,1}(u,\tau)| = O_p(1/\sqrt{n}) \quad \textit{and} \quad \sup_{\tau\in\Upsilon}|R^*_{sfe,1,0}(u,\tau)| = O_p(1/\sqrt{n}).$$

Proof. We focus on $R^*_{sfe,1,1}(u,\tau)$. Note that

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon}|\sum_{i=1}^{n}(A_i^*-\pi)1\{S_i^*=s\}m_1(s,\tau)| = \sup_{s\in\mathcal{S},\tau\in\Upsilon}|D_n^*(s)m_1(s,\tau)| = O_p(\sqrt{n}).$$

If

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon}|\sum_{i=1}^{n}A_i^*1\{S_i^*=s\}\eta^*_{i,1}(s,\tau)| = O_p(\sqrt{n}), \tag{F.15}$$

then, we have

$$\sup_{\tau\in\Upsilon}|R^*_{sfe,1,1}(u,\tau)|$$

$$\leq \sum_{s\in\mathcal{S}} \sup_{s\in\mathcal{S}} \left|\frac{u_1 D_n^*(s)}{\sqrt{n}n^*(s)}\right| \left[\sup_{s\in\mathcal{S},\tau\in\Upsilon} |\sum_{i=1}^n A_i^* 1\{S_i^*=s\}\eta_{i,1}^*(s,\tau)| + \sup_{s\in\mathcal{S},\tau\in\Upsilon} |\sum_{i=1}^n (A_i^*-\pi)1\{S_i^*=s\}m_1(s,\tau)|\right]$$
$$=O_p(1/\sqrt{n}).$$

Therefore, it suffices to show (F.15). Recall $\{Y_i^s(0), Y_i^s(1)\}_{i=1}^n$ as defined in the proof of Theorem 3.1 and

$$\tilde{\eta}_{i,j}(s,\tau) = \tau - 1\{Y_i^s(j) \leq q_j(\tau)\} - m_j(s,\tau),$$

$j = 0, 1$. In addition, let $\Psi_n = \{\eta_{i,1}(s,\tau)\}_{i=1}^n$,

$$\mathbb{N}_n = \{n(s)/n, n_1(s)/n, n^*(s)/n, n_1^*(s)/n\}_{s\in\mathcal{S}}$$

and given $\mathbb{N}_n$, $\{M_{ni}\}_{i=1}^n$ be a sequence of random variables such that the $n_1(s) \times 1$ vector

$$M_n^1(s) = (M_{n,N(s)+1}, \cdots, M_{n,N(s)+n_1(s)})$$

and the $(n(s) - n_1(s)) \times 1$ vector

$$M_n^0(s) = (M_{n,N(s)+n_1(s)+1}, \cdots, M_{n,N(s)+n(s)})$$

satisfy:

1. $M_n^1(s) = \sum_{i=1}^{n_1^*(s)} m_i$ and $M_n^0(s) = \sum_{i=1}^{n^*(s)-n_1^*(s)} m_i'$, where $\{m_i\}_{i=1}^{n_1^*(s)}$ and $\{m_i'\}_{i=1}^{n^*(s)-n_1^*(s)}$ are $n_1^*(s)$ i.i.d. multinomial$(1, n_1^{-1}(s), \cdots, n_1^{-1}(s))$ random vectors and $n^*(s) - n_1^*(s)$ i.i.d. multinomial$(1, (n(s) - n_1(s))^{-1}, \cdots, (n(s) - n_1(s))^{-1})$ random vectors, respectively;
2. $M_n^0(s) \perp\!\!\!\perp M_n^1(s)|\mathbb{N}_n$; and
3. $\{M_n^0(s), M_n^1(s)\}_{s\in\mathcal{S}}$ are independent across s given $\mathbb{N}_n$ and are independent of Ψ_n.

Recall that, by Bugni et al. (2018a), the original observations can be rearranged according to $s \in \mathcal{S}$ and then within strata, treatment group first and then the control group. Then, given $\mathbb{N}_n$ which is determined in Step 1 and 2 of the covariate-adaptive bootstrap procedure, the Step 3 implies that the bootstrap observations $\{Y_i^*\}_{i=1}^n$ can be generated by drawing with replacement from the empirical distribution of the outcomes in each (s,a) cell for $(s,a) \in \mathcal{S}\times\{0,1\}$, $n_a^*(s)$ times, $a = 0, 1$, where $n_0^*(s) = n^*(s) - n_1^*(s)$. Therefore,

$$\sum_{i=1}^n A_i^* 1\{S_i^* = s\}\eta_{i,1}^*(s,\tau) \stackrel{d}{=} \sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\tilde{\eta}_{i,1}(s,\tau). \tag{F.16}$$

Following the standard approach in dealing with the nonparametric bootstrap, we want to approximate

$$M_{ni}, i = N(s) + 1, \cdots, N(s) + n_1(s)$$

by a sequence of i.i.d. Poisson(1) random variables. We construct this sequence as follows. Let $\widetilde{M}_n^1(s) = \sum_{i=1}^{N(n_1(s))} m_i$, where $N(k)$ is a Poisson number with mean k and is independent of $\mathbb{N}_n$. The $n_1(s)$ elements of vector $\widetilde{M}_n^1(s)$ is denoted as $\{\widetilde{M}_{ni}\}_{i=N(s)+1}^{N(s)+n_1(s)}$, which is a sequence of i.i.d. Poisson(1) random variables, given $\mathbb{N}_n$. Therefore,

$$\{\widetilde{M}_{ni}, i = N(s)+1, \cdots, N(s)+n_1(s)|\mathbb{N}_n\} \stackrel{d}{=} \{\xi_i^s, i = N(s)+1, \cdots, N(s)+n_1(s)|\mathbb{N}_n\}$$

where $\{\xi_i^s\}_{i=1}^n$, $s \in \mathcal{S}$ are i.i.d. sequences of Poisson(1) random variables such that $\{\xi_i^s\}_{i=1}^n$ are independent across $s \in \mathcal{S}$ and against $\mathbb{N}_n$.

Following the argument in van der Vaart and Wellner (1996, Section 3.6), given $n_1(s)$, $n_1^*(s)$, and $N(n_1(s)) = k$, $|\xi_i^s - M_{ni}|$ is binomially $(|k - n_1^*(s)|, n_1(s)^{-1})$-distributed. In addition, there exists a sequence $\ell_n = O(\sqrt{n})$ such that

$$\begin{aligned}
\mathbb{P}(|N(n_1(s)) - n_1^*(s)| \geq \ell_n) \leq & \mathbb{P}(|N(n_1(s)) - n_1(s)| \geq \ell_n/3) + \mathbb{P}(|n_1^*(s) - n_1(s)| \geq 2\ell_n/3) \\
\leq & \mathbb{E}\mathbb{P}(|N(n_1(s)) - n_1(s)| \geq \ell_n/3|n_1(s)) + \mathbb{P}(|n_1^*(s) - n_1(s)| \geq 2\ell_n/3) \\
\leq & \varepsilon/3 + \mathbb{P}(|n_1^*(s) - n_1(s)| \geq 2\ell_n/3) \\
\leq & \varepsilon/3 + \mathbb{P}(|D_n^*(s)| + |D_n(s)| + \pi|n^*(s) - n(s)| \geq 2\ell_n/3) \\
\leq & 2\varepsilon/3 + \mathbb{P}(\pi|n^*(s) - n(s)| \geq \ell_n/3) \\
\leq & \varepsilon,
\end{aligned}$$

where the first inequality holds due to the union bound inequality, the second inequality holds by the law of iterated expectation, the third inequality holds because (1) conditionally on $n_1(s)$, $N(n_1(s)) - n_1(s) = O_p(\sqrt{n_1(s)})$ and (2) $n_1(s)/n \to \pi p(s) > 0$, the fourth inequality holds by the fact that

$$n_1^*(s) - n_1(s) = D_n^*(s) - D_n(s) + \pi(n^*(s) - n(s)),$$

the fifth inequality holds because by Assumptions 1 and 4, $|D_n^*(s)| + |D_n(s)| = O_p(\sqrt{n})$, and the sixth inequality holds because $\{S_i^*\}_{i=1}^n$ is generated from $\{S_i\}_{i=1}^n$ by the standard bootstrap procedure, and thus, by van der Vaart and Wellner (1996, Theorem 3.6.1),

$$n^*(s) - n(s) = \sum_{i=1}^n (M_{ni}^w - 1)(1\{S_i = s\} - p(s)) = O_p(\sqrt{n}),$$

where $(M_{n1}^w, \cdots, M_{nn}^w)$ is independent of $\{S_i\}_{i=1}^n$ and multinomially distributed with parameters n

and (probabilities) $1/n, \cdots, 1/n$. Therefore, by direct calculation, as $n \to \infty$,

$$
\begin{aligned}
&\mathbb{P}(\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} |\xi_i^s - M_{ni}| > 2) \\
\leq &\mathbb{P}(\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} |\xi_i^s - M_{ni}| > 2, n_1(s) \geq n\varepsilon) + \mathbb{P}(n_1(s) \leq n\varepsilon) \\
\leq &\varepsilon + \mathbb{E} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \mathbb{P}(|\xi_i^s - M_{ni}| > 2, |N(n_1(s)) - n_1^*(s)| \leq \ell_n, n_1(s) \geq n\varepsilon | n_1(s), n_1^*(s)) + \varepsilon \\
\leq &2\varepsilon + n\mathbb{E}\mathbb{P}(\text{bin}(\ell_n, n_1^{-1}(s)) > 2 | n_1(s), n_1^*(s)) 1\{n_1(s) \geq n\varepsilon\} \to 2\varepsilon,
\end{aligned}
$$

where we use the fact that

$$
n\mathbb{P}(\text{bin}(\ell_n, n_1^{-1}(s)) > 2 | n_1(s), n_1^*(s)) 1\{n_1(s) \geq n\varepsilon\} \lesssim n \left(\frac{\ell_n}{n}\right)^3 \left(\frac{n}{n_1(s)}\right)^3 1\{n_1(s) \geq n\varepsilon\} \lesssim \frac{1}{\sqrt{n}\varepsilon^3}.
$$

Because ε is arbitrary, we have

$$
\mathbb{P}\left(\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} |\xi_i^s - M_{ni}| > 2\right) \to 0. \tag{F.17}
$$

Note that $|\xi_i^s - M_{ni}| = \sum_{j=1}^{\infty} 1\{|\xi_i^s - M_{ni}| \geq j\}$. Let $I_n^j(s)$ be the set of indexes $i \in \{N(s) + 1, \cdots, N(s) + n_1(s)\}$ such that $|\xi_i^s - M_{ni}| \geq j$. Then, $\xi_i^s - M_{ni} = \text{sign}(N(n_1(s)) - n_1(s)) \sum_{j=1}^{\infty} 1\{i \in I_n^j(s)\}$. Thus,

$$
\frac{1}{\sqrt{n}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} (\xi_i^s - M_{ni}) \tilde{\eta}_{i,1}(s,\tau) = \text{sign}(N(n_1(s)) - n_1(s)) \sum_{j=1}^{\infty} \left[\frac{\# I_n^j(s)}{\sqrt{n}} \frac{1}{\# I_n^j(s)} \sum_{i \in I_n^j(s)} \tilde{\eta}_{i,1}(s,\tau)\right]. \tag{F.18}
$$

In the following, we aim to show that the RHS of (F.18) converges to zero in probability uniformly over $s \in \mathcal{S}, \tau \in \Upsilon$. First, note that, by (F.17), $\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} |\xi_i^s - M_{ni}| \leq 2$ occurs with probability approaching one. In the event set that $\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} |\xi_i^s - M_{ni}| \leq 2$, only the first two terms of the first summation on the RHS of (F.18) can be nonzero. In addition, for any j, we have $j(\# I_n^j(s)) \leq |N(n_1(s)) - n_1(s)| = O_p(\sqrt{n})$, and thus, $\frac{\# I_n^j(s)}{\sqrt{n}} = O_p(1)$ for $j = 1, 2$. Therefore, it suffices to show that, for $j = 1, 2$,

$$
\sup_{s \in \mathcal{S}, \tau \in \Upsilon} \left| \frac{1}{\# I_n^j(s)} \sum_{i \in I_n^j(s)} \tilde{\eta}_{i,1}(s,\tau) \right| = o_p(1).
$$

Note that

$$\frac{1}{\# I_n^j(s)} \sum_{i \in I_n^j(s)} \tilde{\eta}_{i,1}(s,\tau) = \sum_{i=N(s)+1}^{N(s)+n_1(s)} \omega_{ni} \tilde{\eta}_{i,1}(s,\tau), \tag{F.19}$$

where $\omega_{ni} = \frac{1\{|\xi_i^s - M_{ni}| \geq j\}}{\# I_n^j(s)}$, $i = N(s)+1, \cdots, N(s)+n_1(s)$ and by construction, $\{\omega_{ni}\}_{i=N(s)+1}^{N(s)+n_1(s)}$ is independent of $\{\eta_{i,1}(s,\tau)\}_{i=1}^n$. In addition, because $\{\omega_{ni}\}_{i=N(s)+1}^{N(s)+n_1(s)}$ is exchangeable conditional on $\mathbb{N}_n$, so be it unconditionally. Third, $\sum_{i=N(s)+1}^{N(s)+n_1(s)} \omega_{ni} = 1$ and $\max_{i=N(s)+1,\cdots,N(s)+n_1(s)} |\omega_{ni}| \leq 1/\# I_n^j(s) \xrightarrow{p} 0$. Then, by the same argument in the proof of van der Vaart and Wellner (1996, Lemma 3.6.16), for some $r \in (0,1)$ and any $n_0 = N(s)+1, \cdots, N(s)+n_1(s)$, we have

$$\begin{aligned}
&\mathbb{E}\left(\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \sum_{i=N(s)+1}^{N(s)+n_1(s)} \omega_{ni} \tilde{\eta}_{i,1}(s,\tau) \right|^r |\Psi_n, \mathbb{N}_n \right) \\
\leq &(n_0 - 1)\mathbb{E}\left[\max_{N(s)+n_0 \leq i \leq N(s)+n_1(s)} \omega_{ni}^r |\mathbb{N}_n \right] \left[\frac{1}{n_1(s)} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \sup_{\tau \in \Upsilon, s \in \mathcal{S}} |\tilde{\eta}_{i,1}^r(s,\tau)| \right] \\
&+ (n_1(s)\mathbb{E}(\omega_{ni}|\mathbb{N}_n))^r \max_{n_0 \leq k \leq n_1(s)} \mathbb{E}\left[\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=N(s)+n_0}^{N(s)+k} \tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau) \right|^r |\mathbb{N}_n, \Psi_n \right],
\end{aligned} \tag{F.20}$$

where $(R_{k_1+1}(k_1,k_2), \cdots, R_{k_1+k_2}(k_1,k_2))$ is uniformly distributed on the set of all permutations of $k_1+1, \cdots, k_1+k_2$ and independent of $\mathbb{N}_n$ and Ψ_n. First note that $\sup_{s \in \mathcal{S}, \tau \in \Upsilon} |\eta_{i,1}(s,\tau)|$ is bounded and

$$\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} \omega_{ni}^r \leq 1/(\# I_n^j(s))^r \xrightarrow{p} 0.$$

Therefore, the first term on the RHS of (F.20) converges to zero in probability for every fixed n_0. For the second term, because $\omega_{ni}|\mathbb{N}_n$ is exchangeable,

$$n_1(s)\mathbb{E}(\omega_{ni}|\mathbb{N}_n) = \sum_{i=N(s)+1}^{N(s)+n_1(s)} \mathbb{E}(\omega_{ni}|\mathbb{N}_n) = 1.$$

In addition, let $\mathbb{S}_n(k_1,k_2)$ be the σ-field generated by all functions of $\{\tilde{\eta}_{i,1}(s,\tau)\}_{i \geq 1}$ that are symmetric in their k_1+1 to k_1+k_2 arguments. Then,

$$\begin{aligned}
&\max_{n_0 \leq k \leq n_1(s)} \mathbb{E}\left[\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=N(s)+n_0}^{N(s)+k} \tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau) \right|^r |\mathbb{N}_n, \Psi_n \right] \\
= &\max_{n_0 \leq k \leq n_1(s)} \mathbb{E}\left[\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=N(s)+n_0}^{N(s)+k} \tilde{\eta}_{j,1}(s,\tau) \right|^r |\mathbb{N}_n, \mathbb{S}_n(N(s), n_1(s)) \right]
\end{aligned}$$

$$\leq 2\mathbb{E}\left\{\max_{n_0\leq k}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+1}^{N(s)+k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right]|\mathbb{N}_n,\mathbb{S}_n(N(s),n_1(s))\right\}$$
$$=2\mathbb{E}\left\{\max_{n_0\leq k}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right]|\mathbb{N}_n,\mathbb{S}_n(0,n_1(s))\right\},$$

where the inequality holds by the Jansen's inequality and the triangle inequality and the last equality holds because $\{\tilde{\eta}_{j,1}(s,\tau)\}_{j\geq 1}$ is an i.i.d. sequence. Apply expectation on both sides, we obtain that

$$\mathbb{E}\max_{n_0\leq k\leq n_1(s)}\mathbb{E}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+n_0}^{N(s)+k}\tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau)\right|^r|\mathbb{N}_n,\Psi_n\right]$$
$$\leq 2\mathbb{E}\max_{n_0\leq k\leq n}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right]. \tag{F.21}$$

By the usual maximal inequality, as $k\to\infty$,

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|\xrightarrow{a.s.}0,$$

which implies that as $n_0\to\infty$

$$\max_{n_0\leq k\leq n}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right]\leq\max_{n_0\leq k}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right]\xrightarrow{a.s.}0.$$

In addition, $\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|$ is bounded. Then, by the bounded convergence theorem, we have, as $n_0\to\infty$,

$$\mathbb{E}\max_{n_0\leq k\leq n}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right]\to 0.$$

which implies that,

$$\mathbb{E}\max_{n_0\leq k\leq n_1(s)}\mathbb{E}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+n_0}^{N(s)+k}\tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau)\right|^r|\mathbb{N}_n,\Psi_n\right]\xrightarrow{p}0.$$

Therefore, the second term on the RHS of (F.20) converges to zero in probability as $n_0\to\infty$.

Then, as $n \to \infty$ followed by $n_0 \to \infty$,

$$\mathbb{E}\left(\sup_{\tau\in\Upsilon, s\in\mathcal{S}}\left|\sum_{i=N(s)+1}^{N(s)+n_1(s)}\omega_{ni}\tilde{\eta}_{i,1}(s,\tau)\right|^r |\Psi_n, \mathbb{N}_n\right) \xrightarrow{p} 0.$$

Hence, by the Markov inequality and (F.19), we have

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon}\left|\frac{1}{\# I_n^j(s)}\sum_{i\in I_n^j(s)}\tilde{\eta}_{i,1}(s,\tau)\right| \xrightarrow{p} 0.$$

Consequently, following (F.18)

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon}\left|\sum_{i=N(s)+1}^{N(s)+n_1(s)}(\xi_i^s - M_{ni})\tilde{\eta}_{i,1}(s,\tau)\right| = o_p(\sqrt{n}). \tag{F.22}$$

Furthermore,

$$\sum_{i=N(s)+1}^{N(s)+n_1(s)}\widetilde{M}_{ni}\tilde{\eta}_{i,1}(s,\tau) \stackrel{d}{=} \sum_{i=N(s)+1}^{N(s)+n(s)}\xi_i^s\tilde{\eta}_{i,1}(s,\tau) = \Gamma_n^*\left(s, \frac{N(s)+n_1(s)}{n}, \tau\right) - \Gamma_n^*\left(s, \frac{N(s)}{n}, \tau\right),$$

where

$$\Gamma_n^*(s,t,\tau) = \sum_{i=1}^{\lfloor nt\rfloor}\xi_i^s\tilde{\eta}_{i,1}(s,\tau) \quad \text{and} \quad \Gamma_n^*(s,0,\tau) = 0.$$

Then, for any $\varepsilon > 0$, we can choose a constant $M > 0$ sufficiently large such that

$$\mathbb{P}(\sup_{0<t\le 1,\tau\in\Upsilon,s\in\mathcal{S}}|\Gamma_n^*(s,t,\tau)| \ge \sqrt{n}M) \le \sum_{s\in\mathcal{S}}\frac{270\mathbb{E}\sup_{\tau\in\Upsilon}|\sum_{i=1}^n\xi_i^s\tilde{\eta}_{i,1}(s,\tau)|}{\sqrt{n}M} \le \varepsilon,$$

where the first inequality is due to Lemma F.1 and a similar argument in (A.2), and the second inequality follows Chernozhukov et al. (2014, Corollary 5.1) with the fact that ξ_i^s has an exponential tail. In addition, because, $n_1(s)/n \xrightarrow{p} p(s)\pi \in (0,1)$, we have

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}|\sum_{i=N(s)+1}^{N(s)+n_1(s)}\widetilde{M}_{ni}\tilde{\eta}_{i,1}(s,\tau)| = O_p(\sqrt{n}). \tag{F.23}$$

Combining (F.16), (F.22), and (F.23), we establish (F.15). This concludes the proof. □

Lemma F.10. *Recall $R^*_{sfe,2,1}(u,\tau)$ and $R^*_{sfe,2,0}(u,\tau)$ defined in (E.5) and (E.6), respectively. If*

Assumptions in Theorem 5.1 hold, then (E.5) and (E.6) hold and

$$\sup_{\tau\in\Upsilon}|R^*_{sfe,2,1}(u,\tau)| = o_p(1) \quad \textit{and} \quad \sup_{\tau\in\Upsilon}|R^*_{sfe,2,0}(u,\tau)| = o_p(1).$$

Proof. We focus on (E.5). Following the definition of M_{ni} in the proof of Lemma F.9 and the argument in the Step 1.2 of the proof of Theorem 3.2, we have

$$\begin{aligned}
&L^*_{2,1,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{E^*_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i^s(1)\le q_1(\tau)+v\}-1\{Y_i^s(1)\le q_1(\tau)\}\right)dv\\
=&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\left[\phi_i(u,\tau,s,E^*_n(s))-\mathbb{E}\phi_i(u,\tau,E^*_n(s))\right]+\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\mathbb{E}\phi_i(u,\tau,s,E^*_n(s)),
\end{aligned} \tag{F.24}$$

where $E^*_n(s)=\sqrt{n}(\hat{\pi}^*(s)-\pi)=\frac{n}{n^*(s)}\frac{D^*_n(s)}{\sqrt{n}}=O_p(1)$,

$$\phi_i(u,\tau,s,e)=\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{e}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}\left(1\{Y_i^s(1)\le q_1(\tau)+v\}-1\{Y_i^s(1)\le q_1(\tau)\}\right)dv,$$

and $\mathbb{E}\phi_i(u,\tau,s,E^*_n(s))$ is interpreted as $\mathbb{E}\phi_i(u,\tau,s,e)$ with e being evaluated at $E^*_n(s)$.

For the first term on the RHS of (F.24), similar to (F.22), we have

$$\begin{aligned}
&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\left[\phi_i(u,\tau,s,E^*_n(s))-\mathbb{E}\phi_i(u,\tau,s,E^*_n(s))\right]\\
=&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s\left[\phi_i(u,\tau,s,E^*_n(s))-\mathbb{E}\phi_i(u,\tau,s,E^*_n(s))\right]+\sum_{s\in\mathcal{S}}r_n(u,\tau,s,E^*_n(s)),
\end{aligned} \tag{F.25}$$

where $\{\xi_i^s\}_{i=1}^n$ is a sequence of i.i.d. Poisson(1) random variables and is independent of everything else, and

$$r_n(u,\tau,s,e)=\text{sign}(N(n_1(s))-n_1(s))\sum_{j=1}^{\infty}\frac{\#I_n^j(s)}{\sqrt{n}}\frac{1}{\#I_n^j(s)}\sum_{i\in I_n^j(s)}\sqrt{n}\left[\phi_i(u,\tau,s,e)-\mathbb{E}\phi_i(u,\tau,s,e)\right].$$

We aim to show

$$\sup_{|e|\le M,\tau\in\Upsilon,s\in\mathcal{S}}|r_n(u,\tau,s,e)|=o_p(1), \tag{F.26}$$

Recall that the proof of Lemma F.9 relies on (F.21) and the fact that

$$\mathbb{E} \sup_{n \geq k \geq n_0} \sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=1}^{k} \tilde{\eta}_{j,1}(s,\tau) \right| \to 0.$$

Using the same argument and replacing $\tilde{\eta}_{j,1}(s,\tau)$ by $\sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)]$, in order to show (F.26), we only need to verify that, as $n \to \infty$ followed by $n_0 \to \infty$,

$$\mathbb{E} \sup_{n \geq k \geq n_0} \sup_{|e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=1}^{k} \sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] \right| \to 0$$

Because $\sup_{|e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=1}^{k} \sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] \right|$ is bounded as shown below, it suffices to show that, for any $\varepsilon > 0$, as $n \to \infty$ followed by $n_0 \to \infty$,

$$\mathbb{P}\left(\sup_{n \geq k \geq n_0} \sup_{|e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=1}^{k} \sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] \right| \geq \varepsilon \right) \to 0. \tag{F.27}$$

Define the class of functions $\mathcal{F}_n$ as

$$\mathcal{F}_n = \{\sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] : |e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}\}.$$

Then, $\mathcal{F}_n$ is nested by a VC-class with fixed VC-index. In addition, for fixed u, $\mathcal{F}_n$ has a bounded (and independent of n) envelope function

$$F = |u'\iota_1| + M\left(\max_{\tau \in \Upsilon} |q(\tau)| + |u_1| \right).$$

Last, define $\mathcal{I}_l = \{2^l, 2^l + 1, \cdots, 2^{l+1} - 1\}$. Then,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{n \geq k \geq n_0} \sup_{|e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=1}^{k} \sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] \right| \geq \varepsilon \right) \\
\leq &\sum_{l=\lfloor \log_2(n_0) \rfloor}^{\lfloor \log_2(n) \rfloor + 1} \mathbb{P}\left(\sup_{k \in \mathcal{I}_l} \sup_{|e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{1}{k} \sum_{j=1}^{k} \sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] \right| \geq \varepsilon \right) \\
\leq &\sum_{l=\lfloor \log_2(n_0) \rfloor}^{\lfloor \log_2(n) \rfloor + 1} \mathbb{P}\left(\sup_{k \leq 2^{l+1}} \sup_{|e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}} \left| \sum_{j=1}^{k} \sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] \right| \geq \varepsilon 2^l \right) \\
\leq &\sum_{l=\lfloor \log_2(n_0) \rfloor}^{\lfloor \log_2(n) \rfloor + 1} 9\mathbb{P}\left(\sup_{|e| \leq M, \tau \in \Upsilon, s \in \mathcal{S}} \left| \sum_{j=1}^{2^{l+1}} \sqrt{n}\,[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)] \right| \geq \varepsilon 2^l / 30 \right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{l=\lfloor \log_2(n_0)\rfloor}^{\lfloor \log_2(n)\rfloor+1} \frac{270\mathbb{E}\sup_{|e|\leq M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{j=1}^{2^{l+1}}\sqrt{n}\left[\phi_i(u,\tau,s,e)-\mathbb{E}\phi_i(u,\tau,s,e)\right]\right|}{\varepsilon 2^l} \\
&\leq \sum_{l=\lfloor \log_2(n_0)\rfloor}^{\lfloor \log_2(n)\rfloor+1} \frac{C_1}{\varepsilon 2^{l/2}} \\
&\leq \frac{2C_1}{\varepsilon\sqrt{n_0}} \to 0,
\end{aligned}
$$

where the first inequality holds by the union bound, the second inequality holds because on $\mathcal{I}_l$, $2^{l+1} \geq k \geq 2^l$, the third inequality follows the same argument in the proof of Theorem 3.1, the fourth inequality is due to the Markov inequality, the fifth inequality follows the standard maximal inequality such as van der Vaart and Wellner (1996, Theorem 2.14.1) and the constant C_1 is independent of (l,ε,n), and the last inequality holds by letting $n \to \infty$. Because ε is arbitrary, we have established (F.27), and thus, (F.26), which further implies that

$$
\sup_{\tau\in\Upsilon,s\in\mathcal{S}} |r_n(u,\tau,s,E_n^*(s))| = o_p(1),
$$

For the leading term of (F.25), we have

$$
\begin{aligned}
&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s\left[\phi_i(u,\tau,s,E_n^*(s)) - \mathbb{E}\phi_i(u,\tau,s,E_n^*(s))\right] \\
=&\sum_{s\in\mathcal{S}}\left[\Gamma_n^{s*}(N(s),\tau,E_n^*(s)) - \Gamma_n^{s*}(N(s)+n_1(s),\tau,E_n^*(s))\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\Gamma_n^{s*}(k,\tau,e) =& \sum_{i=1}^{k}\xi_i^s \int_0^{\frac{u'\iota_1 - e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}} \left(1\{Y_i^s(1)\leq q_1(\tau)+v\} - 1\{Y_i^s(1)\leq q_1(\tau)\}\right)dv \\
&- k\mathbb{E}\left[\int_0^{\frac{u'\iota_1 - e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}} \left(1\{Y_i^s(1)\leq q_1(\tau)+v\} - 1\{Y_i^s(1)\leq q_1(\tau)\}\right)dv\right].
\end{aligned}
$$

By the same argument in (B.8), we can show that

$$
\sup_{0<t\leq 1,\tau\in\Upsilon,|e|\leq M} |\Gamma_n^{s*}(k,\tau,e)| = o_p(1),
$$

where we need to use the fact that the Poisson(1) random variable has an exponential tail and thus

$$\mathbb{E} \sup_{i \in \{1, \cdots, n\}, s \in \mathcal{S}} \xi_i^s = O(\log(n)).$$

Therefore,

$$\sup_{\tau \in \Upsilon} \left| \sum_{s \in \mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni} \left[\phi_i(u, \tau, E_n^*(s)) - \mathbb{E}\phi_i(u, \tau, E_n^*(s)) \right] \right| = o_p(1). \tag{F.28}$$

For the second term on the RHS of (F.24), we have

$$\begin{aligned} \sum_{s \in \mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni} \mathbb{E}\phi_i(u, \tau, s, e) &= \sum_{s \in \mathcal{S}} n_1^*(s) \mathbb{E}\phi_i(u, \tau, s, e) \\ &= \sum_{s \in \mathcal{S}} \pi p(s) \frac{f_1(q_1(\tau)|s)}{2} (u'\iota_1 - eq(\tau))^2 + o(1), \end{aligned} \tag{F.29}$$

where the $o(1)$ term holds uniformly over $(\tau, e) \in \Upsilon \times [-M, M]$, the first equality holds because $\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni} = n_1^*(s)$ and the second equality holds by the same calculation in (B.10) and the facts that $n^*(s)/n \xrightarrow{p} p(s)$ and

$$\frac{n_1^*(s)}{n} = \frac{D_n^*(s) + \pi n^*(s)}{n} \xrightarrow{p} \pi p(s).$$

Combining (E.5), (F.24), (F.28), (F.29), and the facts that $E_n^*(s) = \frac{n}{n^*(s)} \frac{D_n^*(s)}{\sqrt{n}}$ and $\frac{n}{n^*(s)} \xrightarrow{p} 1/p(s)$, we have

$$L_{2,1,n}^*(u, \tau) = \frac{\pi f_1(q_1(\tau))}{2} (u'\iota_1)^2 - \sum_{s \in \mathcal{S}} f_1(q_1(\tau)|s) \frac{\pi D_n^*(s) u' \iota_1}{\sqrt{n}} q(\tau) + h_{2,1}^*(\tau) + R_{sfe,2,1}^*(u, \tau),$$

where

$$h_{2,1}^*(\tau) = \sum_{s \in \mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2} p(s) (E_n^*(s))^2 q^2(\tau)$$

and

$$\sup_{\tau \in \Upsilon} |R_{sfe,2,1}^*(u, \tau)| = o_p(1).$$

This concludes the proof. $\square$

Lemma F.11. *Recall the definition of* $(W_{sfe,n,1}^*(\tau) - \mathcal{W}_{n,1}(\tau), W_{sfe,n,2}^*(\tau), W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau))$

in (E.7). *If Assumptions in Theorem 5.1 hold, then conditionally on data,*

$$(W^*_{sfe,n,1}(\tau) - \mathcal{W}_{n,1}(\tau), W^*_{sfe,n,2}(\tau), W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau)) \rightsquigarrow (\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$ *are three independent Gaussian processes with covariance kernels*

$$\Sigma_1(\tau_1, \tau_2) = \frac{\min(\tau_1, \tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S, \tau_1)m_1(S, \tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))} + \frac{\min(\tau_1, \tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S, \tau_1)m_0(S, \tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))},$$

$$\begin{aligned}
&\Sigma_2(\tau_1, \tau_2) \\
=&\mathbb{E}\gamma(S)\left[(m_1(S, \tau_1) - m_0(S, \tau_1))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_1))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\right) + q(\tau_1)\left(\frac{f_1(q(\tau_1)|S)}{f_1(q_1(\tau_1))} - \frac{f_0(q(\tau_1)|S)}{f_0(q_0(\tau_1))}\right)\right] \\
&\times \left[(m_1(S, \tau_2) - m_0(S, \tau_2))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_2))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\right) + q(\tau_2)\left(\frac{f_1(q(\tau_2)|S)}{f_1(q_2(\tau_2))} - \frac{f_0(q(\tau_2)|S)}{f_0(q_0(\tau_2))}\right)\right],
\end{aligned}$$

and

$$\Sigma_3(\tau_1, \tau_2) = \mathbb{E}\left[\frac{m_1(S, \tau_1)}{f_1(q_1(\tau_1))} - \frac{m_0(S, \tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S, \tau_2)}{f_1(q_1(\tau_2))} - \frac{m_0(S, \tau_2)}{f_0(q_0(\tau_2))}\right],$$

respectively.

Proof. Let $\mathcal{A}_n = \{(A_i^*, S_i^*, A_i, S_i) : i = 1, \cdots, n\}$. Following the definition of M_{ni} and arguments in the proof of Lemma F.9, we have

$$\begin{aligned}
&\{W^*_{sfe,n,1}(\tau) - \mathcal{W}_{n,1}(\tau)|\mathcal{A}_n\} \\
\overset{d}{=}&\left\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=N(s)+1}^{N(s)+n_1(s)}(M_{ni}-1)\left(\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}\right) - \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}(M_{ni}-1)\left(\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right)\right]\middle|\mathcal{A}_n\right\} \\
=&\left\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=N(s)+1}^{N(s)+n_1(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] + R_1(\tau)\middle|\mathcal{A}_n\right\},
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R_1(\tau)| = o_p(1)$ and $\{\xi_i^s\}_{i=1}^n$, $s \in \mathcal{S}$ are sequences of i.i.d. Poisson(1) random variables that are independent of $\mathcal{A}_n$ and across $s \in \mathcal{S}$. In addition, by the same argument in the proof of Lemma F.2, we have

$$\begin{aligned}
&\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=N(s)+1}^{N(s)+n_1(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] \\
=&\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor}(\xi_i^s-1)\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor}(\xi_i^s-1)\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] + R_2(\tau)
\end{aligned}$$

$$\equiv W_1^*(\tau) + R_2(\tau),$$

where $\sup_{\tau \in \Upsilon} |R_2(\tau)| = o_p(1)$. Because both $W_{sfe,n,2}^*(\tau)$ and $W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau)$ are in the σ-field generated by $\mathcal{A}_n$, we have

$$\begin{aligned}
&(W_{sfe,n,1}^*(\tau) - \mathcal{W}_{n,1}(\tau), W_{sfe,n,2}^*(\tau), W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau)) \\
\stackrel{d}{=}&(W_1^*(\tau) + R_1(\tau) + R_2(\tau), W_{sfe,n,2}^*(\tau), W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau)).
\end{aligned}$$

In addition, note that $\{\xi_i^s\}_{i=1}^n$ and $\{\tilde{\eta}_{i,1}(s,\tau), \tilde{\eta}_{i,1}(s,\tau)\}_{i=1}^n$ are independent of $\mathcal{A}_n$, therefore, $W_1^*(\tau) \perp\!\!\!\perp (W_{sfe,n,2}^*(\tau), W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau))$. Applying van der Vaart and Wellner (1996, Theorem 2.9.6) to each segment

$$\lfloor nF(s) \rfloor + 1, \cdots, \lfloor n(F(s) + \pi p(s)) \rfloor \quad \text{or} \quad \lfloor n(F(s) + \pi p(s)) \rfloor + 1, \cdots, \lfloor n(F(s) + p(s)) \rfloor$$

for $s \in \mathcal{S}$ and noticing that $\{\tilde{\eta}_{i,1}(s,\tau)\}_{i=1}^n$ and $\{\tilde{\eta}_{i,0}(s,\tau)\}_{i=1}^n$ are two i.i.d. sequences for each $s \in \mathcal{S}$, independent of each other, and independent across s, we have, conditionally on $\{\tilde{\eta}_{i,1}(s,\tau), \tilde{\eta}_{i,0}(s,\tau)\}_{i=1}^n$, $s \in \mathcal{S}$,

$$W_1^*(\tau) \rightsquigarrow \mathcal{B}_1(\tau)$$

with the covariance kernel $\Sigma_1(\tau_1, \tau_2)$.

For $W_{sfe,n,2}^*(\tau)$, we note that it depends on data only through $\{S_i^*\}_{i=1}^n$. By Assumption 4,

$$W_{sfe,n,2}^*(\tau) | \{S_i^*\}_{i=1}^n \rightsquigarrow \mathcal{B}_2(\tau)$$

with the covariance kernel $\Sigma_2(\tau_1, \tau_2)$.

Last, for $W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau)$, note that $\{S_i^*\}$ is sampled by the standard bootstrap procedure. Therefore, directly applying van der Vaart and Wellner (1996, Theorem 3.6.2), we have

$$W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i' - 1) \left[\frac{m_1(S_i, \tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i, \tau)}{f_0(q_0(\tau))} \right] + R_3(\tau)$$

where $\sup_{\tau \in \Upsilon} |R_3(\tau)| = o_p(1)$, $\{\xi_i'\}_{i=1}^n$ is a sequence of i.i.d. Poisson(1) random variables that is independent of data and $\{\xi_i^s\}_{i=1}^n$, $s \in \mathcal{S}$. By van der Vaart and Wellner (1996, Theorem 3.6.2), conditionally on data $\{S_i\}_{i=1}^n$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i' - 1) \left[\frac{m_1(S_i, \tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i, \tau)}{f_0(q_0(\tau))} \right] \rightsquigarrow \mathcal{B}_3(\tau),$$

where $\mathcal{B}_3(\tau)$ has the covariance kernel $\Sigma_3(\tau_1, \tau_2)$. Furthermore, $\mathcal{B}_2(\tau)$ and $\mathcal{B}_3(\tau)$ are independent as $\Sigma_2(\tau_1, \tau_2)$ is not a function of $\{S_i^*\}_{i=1}^n$. This concludes the proof.

□

G Additional Simulation Results

G.1 QTE, H_0, $\pi = 0.5$

Table 11: H_0, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.044	0.044	0.058	0.051	0.058	0.048	0.049	0.054
	WEI	0.007	0.035	0.014	0.044	0.045	0.043	0.039	0.042
	BCD	0.002	0.040	0.003	0.040	0.039	0.040	0.037	0.037
	SBR	0.003	0.036	0.004	0.033	0.035	0.037	0.037	0.034
2	SRS	0.046	0.046	0.064	0.063	0.061	0.055	0.059	0.057
	WEI	0.026	0.041	0.047	0.054	0.064	0.053	0.055	0.063
	BCD	0.016	0.037	0.032	0.048	0.051	0.053	0.053	0.051
	SBR	0.013	0.027	0.027	0.045	0.052	0.049	0.050	0.049
3	SRS	0.055	0.055	0.055	0.058	0.056	0.056	0.057	0.058
	WEI	0.049	0.052	0.054	0.053	0.054	0.056	0.054	0.053
	BCD	0.052	0.057	0.054	0.058	0.059	0.060	0.058	0.054
	SBR	0.046	0.048	0.053	0.058	0.056	0.062	0.059	0.057
4	SRS	0.075	0.075	0.060	0.059	0.053	0.055	0.054	0.057
	WEI	0.069	0.069	0.054	0.043	0.059	0.058	0.054	0.058
	BCD	0.063	0.066	0.055	0.041	0.061	0.063	0.063	0.060
	SBR	0.066	0.070	0.048	0.034	0.064	0.062	0.063	0.062

Table 12: H_0, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.051	0.051	0.058	0.042	0.039	0.041	0.045	0.037
	WEI	0.012	0.052	0.015	0.037	0.040	0.052	0.034	0.034
	BCD	0.000	0.047	0.002	0.046	0.042	0.045	0.036	0.037
	SBR	0.000	0.032	0.002	0.034	0.036	0.039	0.035	0.035
2	SRS	0.044	0.044	0.065	0.051	0.053	0.048	0.046	0.047
	WEI	0.035	0.062	0.041	0.061	0.059	0.068	0.062	0.058
	BCD	0.012	0.044	0.022	0.046	0.049	0.056	0.053	0.048
	SBR	0.009	0.043	0.011	0.036	0.047	0.039	0.042	0.042
3	SRS	0.065	0.065	0.066	0.072	0.062	0.064	0.071	0.058
	WEI	0.068	0.070	0.063	0.067	0.073	0.066	0.067	0.073
	BCD	0.058	0.061	0.051	0.054	0.059	0.055	0.058	0.056
	SBR	0.044	0.048	0.039	0.042	0.044	0.041	0.043	0.042
4	SRS	0.045	0.045	0.065	0.057	0.072	0.059	0.054	0.069
	WEI	0.039	0.040	0.050	0.028	0.063	0.064	0.049	0.063
	BCD	0.032	0.038	0.044	0.019	0.055	0.056	0.050	0.054
	SBR	0.028	0.029	0.036	0.015	0.047	0.041	0.039	0.040

Table 13: H_0, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.040	0.040	0.059	0.048	0.049	0.051	0.048	0.046
	WEI	0.012	0.043	0.017	0.038	0.037	0.046	0.042	0.039
	BCD	0.003	0.052	0.007	0.048	0.048	0.052	0.052	0.053
	SBR	0.003	0.046	0.007	0.048	0.046	0.045	0.047	0.044
2	SRS	0.038	0.038	0.052	0.050	0.050	0.047	0.046	0.044
	WEI	0.027	0.044	0.038	0.055	0.059	0.054	0.059	0.059
	BCD	0.018	0.034	0.029	0.042	0.045	0.044	0.048	0.051
	SBR	0.027	0.047	0.045	0.062	0.065	0.062	0.064	0.065
3	SRS	0.053	0.053	0.053	0.058	0.063	0.055	0.058	0.064
	WEI	0.061	0.061	0.054	0.054	0.058	0.060	0.057	0.060
	BCD	0.043	0.046	0.053	0.051	0.053	0.055	0.051	0.053
	SBR	0.053	0.057	0.042	0.044	0.049	0.047	0.047	0.051
4	SRS	0.075	0.075	0.052	0.062	0.058	0.056	0.052	0.062
	WEI	0.063	0.063	0.045	0.032	0.057	0.047	0.049	0.056
	BCD	0.058	0.062	0.050	0.036	0.059	0.052	0.055	0.058
	SBR	0.066	0.070	0.045	0.033	0.057	0.059	0.060	0.056

Table 14: H_0, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.047	0.047	0.049	0.039	0.037	0.039	0.039	0.038
	WEI	0.011	0.042	0.022	0.050	0.049	0.048	0.053	0.054
	BCD	0.003	0.052	0.004	0.050	0.048	0.046	0.046	0.045
	SBR	0.005	0.048	0.006	0.048	0.050	0.047	0.049	0.048
2	SRS	0.042	0.042	0.053	0.047	0.054	0.044	0.045	0.051
	WEI	0.016	0.049	0.029	0.059	0.064	0.057	0.065	0.062
	BCD	0.007	0.033	0.015	0.041	0.051	0.044	0.042	0.051
	SBR	0.012	0.047	0.021	0.055	0.067	0.065	0.063	0.064
3	SRS	0.052	0.052	0.044	0.041	0.052	0.050	0.048	0.048
	WEI	0.057	0.060	0.045	0.047	0.046	0.048	0.044	0.046
	BCD	0.051	0.053	0.050	0.051	0.053	0.048	0.049	0.050
	SBR	0.054	0.057	0.044	0.043	0.040	0.043	0.041	0.041
4	SRS	0.046	0.046	0.056	0.061	0.057	0.053	0.059	0.055
	WEI	0.053	0.056	0.057	0.048	0.063	0.064	0.066	0.064
	BCD	0.055	0.060	0.053	0.028	0.059	0.057	0.057	0.053
	SBR	0.031	0.032	0.038	0.016	0.048	0.041	0.039	0.040

G.2 QTE, H_1, $\pi = 0.5$

Table 15: H_1, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.197	0.197	0.218	0.364	0.366	0.230	0.361	0.360
	WEI	0.135	0.293	0.159	0.360	0.369	0.301	0.353	0.360
	BCD	0.097	0.372	0.120	0.359	0.370	0.347	0.357	0.358
	SBR	0.110	0.420	0.134	0.373	0.378	0.405	0.369	0.367
2	SRS	0.293	0.293	0.318	0.365	0.359	0.323	0.362	0.356
	WEI	0.239	0.282	0.268	0.322	0.329	0.319	0.344	0.330
	BCD	0.229	0.303	0.264	0.317	0.332	0.329	0.326	0.326
	SBR	0.260	0.335	0.292	0.330	0.352	0.365	0.347	0.344
3	SRS	0.719	0.719	0.686	0.716	0.706	0.699	0.713	0.702
	WEI	0.727	0.731	0.702	0.705	0.720	0.713	0.705	0.712
	BCD	0.723	0.727	0.702	0.712	0.713	0.726	0.711	0.712
	SBR	0.705	0.714	0.691	0.677	0.689	0.708	0.689	0.682
4	SRS	0.186	0.186	0.136	0.126	0.148	0.153	0.135	0.144
	WEI	0.193	0.200	0.149	0.099	0.154	0.161	0.148	0.151
	BCD	0.176	0.189	0.132	0.098	0.145	0.148	0.143	0.142
	SBR	0.196	0.203	0.145	0.103	0.162	0.173	0.164	0.164

Table 16: H_1, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.214	0.214	0.229	0.392	0.388	0.231	0.395	0.383
	WEI	0.135	0.296	0.161	0.369	0.369	0.317	0.367	0.366
	BCD	0.097	0.390	0.125	0.392	0.395	0.373	0.387	0.389
	SBR	0.092	0.383	0.119	0.402	0.398	0.370	0.406	0.391
2	SRS	0.252	0.252	0.291	0.385	0.412	0.313	0.387	0.407
	WEI	0.234	0.345	0.289	0.401	0.436	0.392	0.404	0.429
	BCD	0.197	0.381	0.272	0.404	0.424	0.418	0.425	0.428
	SBR	0.187	0.382	0.256	0.418	0.438	0.420	0.439	0.429
3	SRS	0.693	0.693	0.626	0.603	0.621	0.639	0.613	0.610
	WEI	0.695	0.698	0.620	0.608	0.636	0.638	0.617	0.634
	BCD	0.706	0.711	0.635	0.641	0.647	0.651	0.640	0.641
	SBR	0.679	0.686	0.601	0.639	0.646	0.618	0.647	0.648
4	SRS	0.165	0.165	0.173	0.131	0.194	0.187	0.139	0.201
	WEI	0.162	0.174	0.171	0.101	0.187	0.190	0.177	0.185
	BCD	0.167	0.179	0.183	0.105	0.206	0.205	0.189	0.200
	SBR	0.145	0.153	0.172	0.097	0.203	0.193	0.204	0.204

Table 17: H_1, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.222	0.222	0.228	0.420	0.420	0.244	0.416	0.416
	WEI	0.137	0.289	0.166	0.407	0.410	0.308	0.405	0.410
	BCD	0.136	0.435	0.151	0.432	0.427	0.414	0.429	0.428
	SBR	0.117	0.428	0.127	0.398	0.403	0.429	0.403	0.401
2	SRS	0.324	0.324	0.350	0.405	0.406	0.355	0.401	0.400
	WEI	0.300	0.343	0.328	0.380	0.397	0.380	0.390	0.399
	BCD	0.322	0.382	0.324	0.394	0.405	0.402	0.405	0.400
	SBR	0.323	0.395	0.344	0.391	0.398	0.416	0.407	0.398
3	SRS	0.806	0.806	0.777	0.766	0.781	0.785	0.769	0.778
	WEI	0.771	0.775	0.738	0.745	0.752	0.749	0.744	0.744
	BCD	0.777	0.781	0.753	0.746	0.756	0.759	0.751	0.754
	SBR	0.796	0.799	0.770	0.752	0.761	0.775	0.760	0.754
4	SRS	0.180	0.180	0.146	0.128	0.156	0.153	0.134	0.158
	WEI	0.203	0.206	0.151	0.123	0.151	0.156	0.155	0.157
	BCD	0.187	0.197	0.143	0.097	0.162	0.155	0.153	0.157
	SBR	0.216	0.230	0.164	0.121	0.184	0.180	0.178	0.179

Table 18: H_1, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.228	0.228	0.234	0.430	0.432	0.251	0.430	0.432
	WEI	0.156	0.324	0.177	0.412	0.414	0.327	0.409	0.409
	BCD	0.112	0.421	0.137	0.412	0.412	0.407	0.408	0.407
	SBR	0.102	0.433	0.132	0.438	0.439	0.431	0.449	0.444
2	SRS	0.298	0.298	0.328	0.456	0.482	0.325	0.464	0.481
	WEI	0.260	0.369	0.302	0.423	0.446	0.416	0.448	0.454
	BCD	0.264	0.467	0.311	0.460	0.487	0.486	0.486	0.481
	SBR	0.258	0.472	0.317	0.488	0.510	0.495	0.516	0.508
3	SRS	0.759	0.759	0.697	0.693	0.701	0.702	0.691	0.697
	WEI	0.730	0.736	0.675	0.677	0.683	0.690	0.682	0.673
	BCD	0.754	0.759	0.710	0.700	0.713	0.717	0.704	0.708
	SBR	0.742	0.747	0.695	0.712	0.719	0.704	0.718	0.712
4	SRS	0.212	0.212	0.218	0.151	0.227	0.229	0.172	0.236
	WEI	0.192	0.195	0.199	0.117	0.212	0.218	0.191	0.213
	BCD	0.180	0.189	0.202	0.117	0.218	0.227	0.214	0.222
	SBR	0.171	0.176	0.178	0.104	0.211	0.203	0.213	0.213

G.3 QTE, H_0, $\pi = 0.7$

Table 19: H_0, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.027	0.027	0.044	0.031	0.037	0.030	0.026	0.037
	SBR	0.002	0.026	0.002	0.029	0.029	0.027	0.027	0.027
2	SRS	0.042	0.042	0.055	0.047	0.046	0.053	0.047	0.051
	SBR	0.019	0.024	0.038	0.043	0.051	0.055	0.049	0.052
3	SRS	0.053	0.053	0.062	0.059	0.056	0.063	0.058	0.060
	SBR	0.047	0.059	0.048	0.054	0.054	0.062	0.058	0.056
4	SRS	0.076	0.076	0.054	0.061	0.070	0.055	0.052	0.058
	SBR	0.059	0.073	0.045	0.024	0.061	0.062	0.063	0.063

Table 20: H_0, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.044	0.044	0.058	0.050	0.047	0.043	0.047	0.049
	SBR	0.002	0.045	0.008	0.052	0.060	0.058	0.053	0.055
2	SRS	0.047	0.047	0.053	0.052	0.053	0.050	0.050	0.051
	SBR	0.013	0.041	0.021	0.044	0.054	0.046	0.048	0.046
3	SRS	0.068	0.068	0.062	0.078	0.067	0.063	0.079	0.066
	SBR	0.052	0.053	0.050	0.048	0.056	0.058	0.056	0.056
4	SRS	0.029	0.029	0.062	0.058	0.063	0.063	0.066	0.066
	SBR	0.023	0.028	0.056	0.046	0.059	0.059	0.057	0.059

Table 21: H_0, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.041	0.041	0.059	0.049	0.049	0.052	0.050	0.051
	SBR	0.005	0.051	0.004	0.045	0.044	0.042	0.043	0.043
2	SRS	0.042	0.042	0.061	0.068	0.060	0.059	0.065	0.057
	SBR	0.022	0.026	0.043	0.047	0.054	0.050	0.050	0.049
3	SRS	0.047	0.047	0.043	0.037	0.046	0.045	0.037	0.046
	SBR	0.039	0.044	0.043	0.044	0.045	0.048	0.047	0.050
4	SRS	0.070	0.070	0.048	0.060	0.047	0.052	0.045	0.048
	SBR	0.070	0.081	0.035	0.021	0.056	0.051	0.055	0.052

Table 22: H_0, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.045	0.045	0.052	0.041	0.044	0.039	0.044	0.042
	SBR	0.002	0.040	0.004	0.041	0.042	0.039	0.042	0.042
2	SRS	0.039	0.039	0.044	0.046	0.049	0.042	0.045	0.049
	SBR	0.018	0.046	0.025	0.048	0.061	0.061	0.064	0.065
3	SRS	0.087	0.087	0.060	0.060	0.074	0.063	0.059	0.078
	SBR	0.061	0.063	0.045	0.040	0.047	0.051	0.051	0.051
4	SRS	0.024	0.024	0.039	0.039	0.042	0.046	0.040	0.046
	SBR	0.037	0.039	0.056	0.048	0.059	0.059	0.057	0.057

G.4 QTE, H_1, $\pi = 0.7$

Table 23: H_1, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.150	0.150	0.173	0.324	0.309	0.178	0.318	0.323
	SBR	0.060	0.178	0.100	0.341	0.346	0.367	0.337	0.337
2	SRS	0.273	0.273	0.312	0.334	0.336	0.315	0.338	0.349
	SBR	0.269	0.301	0.312	0.327	0.352	0.361	0.353	0.352
3	SRS	0.725	0.725	0.677	0.679	0.678	0.702	0.689	0.707
	SBR	0.738	0.749	0.711	0.719	0.728	0.736	0.731	0.729
4	SRS	0.130	0.130	0.100	0.090	0.118	0.102	0.090	0.123
	SBR	0.138	0.167	0.112	0.046	0.150	0.143	0.145	0.142

Table 24: H_1, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.178	0.178	0.200	0.355	0.362	0.190	0.340	0.336
	SBR	0.063	0.319	0.074	0.342	0.348	0.326	0.334	0.339
2	SRS	0.280	0.280	0.311	0.423	0.446	0.320	0.415	0.436
	SBR	0.195	0.378	0.250	0.388	0.417	0.400	0.422	0.412
3	SRS	0.669	0.669	0.586	0.572	0.602	0.599	0.573	0.598
	SBR	0.671	0.679	0.617	0.605	0.633	0.634	0.633	0.636
4	SRS	0.145	0.145	0.198	0.172	0.211	0.203	0.180	0.213
	SBR	0.137	0.146	0.179	0.155	0.194	0.197	0.204	0.194

Table 25: H_1, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.208	0.208	0.229	0.402	0.398	0.231	0.390	0.395
	SBR	0.075	0.372	0.100	0.352	0.359	0.373	0.350	0.349
2	SRS	0.345	0.345	0.381	0.404	0.408	0.382	0.396	0.414
	SBR	0.343	0.376	0.391	0.406	0.425	0.425	0.421	0.420
3	SRS	0.786	0.786	0.756	0.758	0.763	0.758	0.756	0.763
	SBR	0.785	0.800	0.749	0.758	0.766	0.771	0.761	0.765
4	SRS	0.173	0.173	0.113	0.081	0.136	0.118	0.089	0.133
	SBR	0.134	0.167	0.082	0.037	0.118	0.120	0.126	0.125

Table 26: H_1, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.195	0.195	0.209	0.384	0.384	0.216	0.377	0.378
	SBR	0.086	0.375	0.099	0.396	0.394	0.385	0.387	0.391
2	SRS	0.296	0.296	0.337	0.452	0.471	0.351	0.440	0.463
	SBR	0.315	0.478	0.356	0.491	0.507	0.502	0.510	0.503
3	SRS	0.737	0.737	0.690	0.649	0.691	0.697	0.664	0.696
	SBR	0.717	0.721	0.670	0.641	0.678	0.673	0.672	0.671
4	SRS	0.169	0.169	0.235	0.221	0.224	0.238	0.222	0.227
	SBR	0.162	0.164	0.204	0.171	0.218	0.219	0.213	0.223

G.5 ATE, $\pi = 0.5$

Table 27: H_0, $n = 200$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.058	0.057	0.050	0.061	0.055	0.053	0.055	0.049	0.044
	WEI	0.003	0.058	0.064	0.004	0.067	0.068	0.061	0.059	0.062
	BCD	0.000	0.069	0.054	0.000	0.057	0.058	0.054	0.053	0.053
	SBR	0.000	0.063	0.053	0.000	0.058	0.057	0.061	0.056	0.056
2	SRS	0.053	0.054	0.058	0.051	0.052	0.051	0.053	0.053	0.046
	WEI	0.031	0.059	0.064	0.032	0.061	0.064	0.062	0.064	0.064
	BCD	0.014	0.062	0.058	0.015	0.060	0.061	0.060	0.055	0.054
	SBR	0.008	0.045	0.046	0.010	0.047	0.048	0.045	0.047	0.047
3	SRS	0.059	0.059	0.063	0.064	0.070	0.067	0.061	0.062	0.061
	WEI	0.053	0.053	0.056	0.056	0.062	0.062	0.058	0.056	0.055
	BCD	0.062	0.063	0.064	0.063	0.066	0.064	0.061	0.061	0.061
	SBR	0.051	0.054	0.053	0.054	0.056	0.056	0.057	0.056	0.056
4	SRS	0.072	0.072	0.077	0.077	0.075	0.076	0.074	0.075	0.074
	WEI	0.073	0.073	0.075	0.074	0.075	0.078	0.077	0.073	0.075
	BCD	0.071	0.073	0.073	0.074	0.073	0.075	0.073	0.072	0.072
	SBR	0.070	0.070	0.070	0.069	0.069	0.071	0.070	0.070	0.070

Table 28: H_1, $n = 200$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.384	0.386	0.937	0.389	0.941	0.941	0.384	0.942	0.942
	WEI	0.321	0.667	0.935	0.331	0.939	0.939	0.674	0.940	0.941
	BCD	0.256	0.922	0.937	0.265	0.944	0.944	0.894	0.938	0.938
	SBR	0.271	0.947	0.955	0.278	0.959	0.958	0.944	0.954	0.955
2	SRS	0.523	0.518	0.745	0.529	0.747	0.762	0.525	0.742	0.752
	WEI	0.504	0.642	0.725	0.513	0.719	0.724	0.639	0.723	0.725
	BCD	0.497	0.744	0.737	0.498	0.740	0.746	0.729	0.741	0.740
	SBR	0.508	0.737	0.741	0.521	0.740	0.745	0.731	0.741	0.741
3	SRS	0.772	0.774	0.774	0.781	0.778	0.782	0.773	0.771	0.771
	WEI	0.786	0.791	0.793	0.791	0.794	0.799	0.788	0.786	0.788
	BCD	0.785	0.786	0.784	0.790	0.794	0.795	0.787	0.785	0.786
	SBR	0.765	0.769	0.772	0.766	0.774	0.772	0.771	0.773	0.773
4	SRS	0.201	0.199	0.216	0.209	0.206	0.210	0.198	0.203	0.208
	WEI	0.207	0.211	0.213	0.212	0.208	0.220	0.204	0.206	0.211
	BCD	0.213	0.215	0.216	0.222	0.223	0.232	0.213	0.214	0.215
	SBR	0.220	0.221	0.221	0.220	0.219	0.229	0.222	0.221	0.221

Table 29: H_0, $n = 400$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.063	0.061	0.042	0.063	0.043	0.045	0.055	0.042	0.042
	WEI	0.005	0.050	0.050	0.006	0.052	0.052	0.052	0.050	0.050
	BCD	0.000	0.067	0.052	0.000	0.059	0.059	0.051	0.059	0.059
	SBR	0.000	0.059	0.058	0.000	0.057	0.057	0.063	0.060	0.060
2	SRS	0.061	0.057	0.055	0.058	0.055	0.054	0.061	0.054	0.051
	WEI	0.018	0.051	0.064	0.019	0.063	0.064	0.052	0.064	0.064
	BCD	0.009	0.045	0.046	0.006	0.046	0.047	0.043	0.049	0.049
	SBR	0.014	0.062	0.060	0.016	0.065	0.065	0.063	0.063	0.063
3	SRS	0.050	0.049	0.050	0.050	0.049	0.051	0.052	0.048	0.048
	WEI	0.046	0.047	0.049	0.047	0.046	0.047	0.048	0.047	0.046
	BCD	0.049	0.049	0.049	0.049	0.050	0.050	0.050	0.050	0.050
	SBR	0.055	0.056	0.056	0.059	0.058	0.059	0.055	0.056	0.056
4	SRS	0.057	0.057	0.055	0.056	0.056	0.059	0.054	0.051	0.056
	WEI	0.051	0.051	0.053	0.052	0.054	0.054	0.051	0.051	0.052
	BCD	0.056	0.056	0.056	0.054	0.056	0.056	0.054	0.053	0.053
	SBR	0.056	0.058	0.058	0.055	0.056	0.057	0.057	0.057	0.057

Table 30: H_1, $n = 400$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.453	0.453	0.965	0.452	0.966	0.966	0.454	0.963	0.963
	WEI	0.361	0.710	0.954	0.367	0.960	0.958	0.704	0.953	0.953
	BCD	0.322	0.964	0.973	0.331	0.975	0.974	0.955	0.971	0.971
	SBR	0.325	0.956	0.956	0.328	0.958	0.959	0.956	0.954	0.954
2	SRS	0.578	0.576	0.811	0.580	0.803	0.814	0.576	0.803	0.812
	WEI	0.593	0.707	0.798	0.595	0.794	0.799	0.705	0.797	0.795
	BCD	0.624	0.821	0.828	0.625	0.830	0.830	0.810	0.824	0.824
	SBR	0.586	0.800	0.808	0.586	0.804	0.807	0.802	0.805	0.806
3	SRS	0.817	0.819	0.829	0.826	0.823	0.827	0.811	0.817	0.815
	WEI	0.801	0.802	0.804	0.802	0.803	0.807	0.803	0.800	0.801
	BCD	0.804	0.806	0.807	0.809	0.814	0.814	0.807	0.806	0.806
	SBR	0.798	0.800	0.800	0.808	0.810	0.809	0.805	0.805	0.805
4	SRS	0.211	0.211	0.220	0.213	0.214	0.223	0.220	0.224	0.224
	WEI	0.223	0.223	0.219	0.222	0.219	0.220	0.224	0.220	0.220
	BCD	0.212	0.215	0.217	0.221	0.220	0.221	0.217	0.215	0.215
	SBR	0.226	0.227	0.228	0.231	0.228	0.231	0.228	0.230	0.230

G.6 ATE, $\pi = 0.7$

Table 31: H_0, $n = 200$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.050	0.047	0.054	0.052	0.054	0.061	0.050	0.054	0.056
	SBR	0.000	0.054	0.043	0.000	0.051	0.056	0.053	0.049	0.049
2	SRS	0.050	0.050	0.079	0.047	0.053	0.058	0.045	0.052	0.056
	SBR	0.014	0.045	0.044	0.013	0.022	0.045	0.046	0.041	0.041
3	SRS	0.061	0.064	0.074	0.067	0.065	0.069	0.061	0.062	0.061
	SBR	0.054	0.057	0.056	0.058	0.050	0.063	0.060	0.061	0.058
4	SRS	0.055	0.054	0.055	0.059	0.056	0.069	0.058	0.056	0.064
	SBR	0.055	0.059	0.057	0.055	0.041	0.063	0.060	0.059	0.061

Table 32: H_1, $n = 200$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.343	0.344	0.935	0.347	0.945	0.945	0.341	0.938	0.938
	SBR	0.159	0.941	0.936	0.166	0.944	0.947	0.939	0.942	0.941
2	SRS	0.539	0.540	0.656	0.550	0.608	0.734	0.539	0.592	0.723
	SBR	0.547	0.733	0.736	0.556	0.647	0.753	0.724	0.736	0.743
3	SRS	0.762	0.764	0.747	0.775	0.738	0.777	0.766	0.734	0.766
	SBR	0.785	0.788	0.788	0.790	0.775	0.803	0.792	0.793	0.790
4	SRS	0.159	0.153	0.115	0.162	0.149	0.178	0.161	0.143	0.162
	SBR	0.139	0.161	0.160	0.147	0.103	0.166	0.168	0.164	0.164

Table 33: H_0, $n = 400$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.041	0.036	0.055	0.043	0.050	0.049	0.040	0.050	0.048
	SBR	0.000	0.053	0.046	0.000	0.047	0.050	0.054	0.051	0.051
2	SRS	0.047	0.048	0.072	0.046	0.050	0.060	0.048	0.051	0.055
	SBR	0.024	0.053	0.055	0.023	0.029	0.052	0.049	0.053	0.048
3	SRS	0.068	0.067	0.075	0.067	0.065	0.067	0.065	0.066	0.066
	SBR	0.044	0.047	0.047	0.047	0.041	0.051	0.048	0.048	0.048
4	SRS	0.054	0.051	0.050	0.055	0.053	0.059	0.052	0.053	0.058
	SBR	0.049	0.056	0.056	0.051	0.032	0.058	0.055	0.054	0.055

Table 34: H_1, $n = 400$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/B	sfe/B	ipw/B	s/CA	sfe/CA	ipw/CA
1	SRS	0.361	0.364	0.965	0.371	0.966	0.968	0.361	0.967	0.967
	SBR	0.211	0.961	0.956	0.222	0.960	0.960	0.963	0.958	0.958
2	SRS	0.628	0.630	0.735	0.629	0.675	0.785	0.628	0.674	0.789
	SBR	0.665	0.806	0.810	0.667	0.731	0.818	0.810	0.816	0.822
3	SRS	0.804	0.802	0.799	0.810	0.785	0.825	0.809	0.786	0.818
	SBR	0.808	0.812	0.813	0.810	0.804	0.823	0.822	0.821	0.821
4	SRS	0.174	0.175	0.132	0.181	0.160	0.205	0.181	0.156	0.195
	SBR	0.164	0.183	0.184	0.168	0.126	0.191	0.187	0.187	0.188

References

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

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 (2018a). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association* (Forthcoming).

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2018b). Inference under covariate-adaptive randomization with multiple treatments. *working paper*.

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

Chong, A., I. Cohen, E. Field, E. Nakasone, and M. Torero (2016). Iron deficiency and schooling attainment in peru. *American Economic Journal: Applied Economics 8*(4), 222–55.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics 2*(2), 267–277.

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics 4*, 3895–3962.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika 58*(3), 403–417.

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

Hu, Y. and F. Hu (2012). Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics 40*(3), 1794–1815.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

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. *Annals of statistics 26*(2), 755–770.

Ma, W., Y. Qin, Y. Li, and F. Hu (2018). Statistical inference of covariate-adjusted randomized experiments. *arXiv preprint arXiv:1807.09678*.

Montgomery-Smith, S. J. (1993). Comparison of sums of independent identically distributed random variables. *Probability and Mathematical Statistics 14*(2), 281–285.

Peña, V. H., T. L. Lai, and Q.-M. Shao (2008). *Self-normalized processes: Limit theory and Statistical Applications*. Springer Science & Business Media.

Shao, J. and X. Yu (2013). Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics 69*(4), 960–969.

Shao, J., X. Yu, and B. Zhong (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika 97*(2), 347–360.

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

Wei, L. (1978). An application of an urn model to the design of sequential controlled clinical trials. *Journal of the American Statistical Association 73*(363), 559–563.