5-2021

Regression-adjusted estimation of quantile treatment effects under covariate-adaptive randomizations

Liang JIANG

Peter C.B. PHILLIPS

Yubo TAO

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

Citation

JIANG, Liang; PHILLIPS, Peter C.B.; TAO, Yubo; and ZHANG, Yichong. Regression-adjusted estimation of quantile treatment effects under covariate-adaptive randomizations. (2021). *Journal of Econometrics*. 1-110.

Available at: https://ink.library.smu.edu.sg/soe_research/2494

Regression-Adjusted Estimation of Quantile Treatment Effects under Covariate-Adaptive Randomizations*

Liang Jiang† Peter C.B. Phillips‡ Yubo Tao§ Yichong Zhang¶

July 6, 2021

Abstract

Datasets from field experiments with covariate-adaptive randomizations (CARs) usually contain extra baseline covariates in addition to the strata indicators. We propose to incorporate these extra covariates via auxiliary regressions in the estimation and inference of unconditional QTEs under CARs. We establish the consistency, limiting distribution, and validity of the multiplier bootstrap of the regression-adjusted QTE estimator. The auxiliary regression may be estimated parametrically, nonparametrically, or via regularization when the data are high-dimensional. Even when the auxiliary regression is misspecified, the proposed bootstrap inferential procedure still achieves the nominal rejection probability in the limit under the null. When the auxiliary regression is correctly specified, the regression-adjusted estimator achieves the minimum asymptotic variance. We also derive the optimal pseudo true values for the potentially misspecified parametric model that minimize the asymptotic variance of the corresponding QTE estimator. We demonstrate the finite sample performance of the new estimation and inferential methods using simulations and provide an empirical application to a well-known dataset in education.

Keywords: Covariate-adaptive randomization, high-dimensional data, regression adjustment, quantile treatment effects.

JEL codes: C14, C21, I21

*We are grateful to Michael Qingliang Fan and participants at 2021 Nanyang Econometrics workshop for their valuable comments. Yichong Zhang acknowledges financial support from the Singapore Ministry of Education under Tier 2 grant No. MOE2018-T2-2-169 and a Lee Kong Chian fellowship. Peter C.B. Phillips acknowledges support from NSF Grant No. SES 18-50860, a Kelly Fellowship at the University of Auckland, and a Lee Kong Chian Fellowship.

†Fudan University. 220 Handan Rd., Shanghai 200433. E-mail address: jiangliang@fudan.edu.cn.

‡Yale University, University of Auckland, University of Southampton, Singapore Management University. New Haven, Connecticut USA 06520-8281. E-mail: peter.phillips@yale.edu

§Corresponding author: Singapore Management University. 90 Stamford Rd, Singapore 178903. E-mail: ybtao@smu.edu.sg.

¶Singapore Management University. 90 Stamford Rd, Singapore 178903. E-mail address: yczhang@smu.edu.sg.

1 Introduction

Covariate-adaptive randomizations (CARs) have recently seen growing use in a wide variety of randomized experiments in economic research. Examples include Chong, Cohen, Field, Nakasone, and Torero (2016), Greaney, Kaboski, and Van Leemput (2016), Jakiela and Ozier (2016), Burchardi, Gulesci, Lerva, and Sulaiman (2019), Anderson and McKenzie (2021), among many others. In CAR modeling, units are first stratified using some baseline covariates, and then, within each stratum, the treatment status is assigned (independent of covariates) to achieve the balance between the numbers of treated and control units.

In many empirical studies, apart from the average treatment effect (ATE), researchers are often interested in using the randomized experiments to estimate quantile treatment effects (QTEs). The QTE has a useful role as a robustness check for the ATE and characterizes any heterogeneity that may be present in the sign and magnitude of the treatment effects according to their position within the distribution of outcomes. See, for example, Bitler, Gelbach, and Hoynes (2006), Muralidharan and Sundararaman (2011), Duflo, Greenstone, Pande, and Ryan (2013), Banerjee, Duflo, Glennerster, and Kinnan (2015), Crépon, Devoto, Duflo, and Parienté (2015), and Campos, Frese, Goldstein, Iacovone, Johnson, McKenzie, and Mensmann (2017).

Two practical issues arise in estimation and inference concerning QTEs under CARs. First, other covariates in addition to the strata indicators are collected during the experiment. In the estimation of ATE, the usual practice is to run a simple linear OLS regression of the outcome on treatment status, strata indicators, additional covariates, and interaction terms, as in the analysis of covariance (ANCOVA). However, because the quantile function is a nonlinear operator, even when the treatment status is completely randomly assigned, a similar linear quantile regression is unable to consistently estimate the *unconditional* QTE. Second, in order to achieve balance in the respective number of treated and control units within each stratum, treatment statuses under CARs usually exhibit a (negative) cross-sectional dependence. Standard inference procedures that rely on cross-sectional independence are therefore conservative and lack power. These two issues raise questions of how to use the additional covariates in the estimation of QTE and how to conduct valid statistical procedures that mitigate conservatism in inference.

The present paper addresses these issues by including additional covariates via auxiliary regressions, deriving the limit theory, and establishing the validity of multiplier bootstrap inference for the corresponding regression-adjusted QTE estimator under CARs. Even under potential misspecification of the auxiliary regressions, the QTE estimator is shown to maintain consistency and the multiplier bootstrap procedure to have asymptotic size equal to the nominal level under the null. When the auxiliary regression is correctly specified, the QTE estimator achieves minimum asymptotic variance. These results are built on high-level conditions concerning the estimates of the auxiliary regressions. The conditions are then verified for auxiliary regressions that are estimated (1) parametrically, (2) nonparametrically, or (3) via regularization in high dimensional cases. Specifically, for (1) we derive the pseudo true values of the parameters that minimize the asymptotic variance of the QTE estimator and verify the validity of quasi maximum likelihood estimation of the auxiliary regressions. For (2) and (3), we consider logistic sieve regression and logistic regression under ℓ_1 penalization when the additional covariates are high-dimensional. The limit theory holds uniformly over a compact set of quantile indexes, implying that our multiplier bootstrap procedure can be used to conduct inference on QTEs involving single, multiple, or a continuum of quantile indexes.

From a practical perspective, our estimation and inferential methods have four advantages. First, they allow for common choices of auxiliary regressions such as linear probability, logit, and probit regressions, even though these regressions may be misspecified. Second, the methods can be implemented without tuning parameters. Third, our (bootstrap) estimator can be directly computed via the subgradient condition, and the auxiliary regressions need not be re-estimated in the bootstrap procedure, both of which save considerable computation time. Last, our estimation and inference methods can be implemented without the knowledge of the exact treatment assignment rule used in the experiment. Such information may not be available when researchers are using an experiment that was run in the past and the randomization procedure may not have been fully described. It also occurs in subsample analysis, where sub-groups are defined using variables that may have not been used to form the strata and the treatment assignment rule for each sub-group becomes unknown. See, for example, the anemic subsample analysis in Chong et al. (2016).

The contributions in the present paper relate to two strands of the present literature. The first is causal inference under CARs. Shao, Yu, and Zhong (2010) showed that the usual two-sample t-test for the ATE is conservative under CARs and proposed a covariate-adaptive bootstrap to conduct inference. Bugni, Canay, and Shaikh (2018) extended the framework to the nonparametric setup, proposed an adjusted standard error for the ATE estimator, and considered a permutation test. Zhang and Zheng (2020) showed that, under CARs, the standard bootstrap inference for the QTE is conservative under the null and lacks power under the alternative. Instead, they suggested multiplier bootstrapping the inverse propensity score weighted (IPW) estimator of the QTE and showed that the distribution of the bootstrap estimator can mimic the original one under CARs. Additional studies in this literature include Hu and Hu (2012); Shao and Yu (2013); Ma, Hu, and Zhang (2015); Ma, Qin, Li, and Hu (2018); Ye (2018); Bugni, Canay, and Shaikh (2019); Ye and Shao (2020); Bugni and Gao (2021); Olivares (2021). Our work complements these studies by considering estimation and bootstrap inference of the QTE under CARs with regression adjustments.

A second strand of research deals with regression adjustment of treatment effect estimation in randomized experiments. Freedman (2008a,b) pointed out that ordinary least squares regression adjustment in randomized experiments can degrade the precision of the ATE estimator. Lin (2013) reexamined Freedman's critique and pointed out that linear regression adjustment does not lead to loss of precision when a full set of interactions between the treatment status and covariates is included. Lei and Ding (2021) extended Lin (2013) to situations where the dimension of the covariates can grow with the sample size at a certain rate. Bloniarz, Liu, Zhang, Sekhon, and Yu (2016) provided a further extension to cases where the dimension of the covariates exceeds the sample size, proposing the use of Lasso in the auxiliary regression. Ye, Yi, and Shao (2020) and Ma, Tu, and Liu (2020) studied regression adjustment of ATE estimates under CARs. Liu, Tu, and Ma (2020) established a general theory for regression-adjusted ATE estimation under CARs, incorporating OLS and Lasso estimation as special cases. The other works in this branch include, but are limited to, Lu (2016), Fogarty (2018), Li and Ding (2020), Liu and Yang (2020), Negi and Wooldridge (2020), and Zhao and Ding (2020). All of these papers focus on ATE. Zhang, Tsiatis, and Davidian (2008) considered the regression adjustment for a semiparametric model and linked the estimation method with the theory of semiparametric efficiency. The present paper complements the above works by studying QTE regression, which is nonparametrically specified, with both linear (linear probability model) and nonlinear (logit and probit models) regression adjustments. Similar to Liu et al. (2020), we establish a general theory for regression adjustment that allows for parametric, nonparametric, and regularized estimation of the auxiliary regressions. A further contribution to the literature is to establish the validity of a multiplier bootstrap inferential procedure that does not suffer from conservatism under CARs.

The present paper also comes under the umbrella of a growing literature that studies estimation and inference in randomized experiments. In this connection, we mention the work of Hahn, Hirano, and Karlan (2011); Athey and Imbens (2017); Abadie, Chingos, and West (2018); Tabord-Meehan (2018); Bai, Shaikh, and Romano (2021); Bai (2019); Jiang, Liu, Phillips, and Zhang (2021) among many others.

The remainder of the paper is organized as follows. Section 2 describes the model setup and notation. Section 3 develops the asymptotic properties of our regression-adjusted QTE estimator. Section 4 studies the validity of the multiplier bootstrap inference. Section 5 considers parametric, nonparametric, and regularized estimation of the auxiliary regressions. Section 6 provides computational details and recommendations for practitioners. Section 7 reports simulation results, and an empirical application of our methods to the impact of child health and nutrition on educational outcomes is provided in Section 8. Section 9 concludes. Proofs of all results and some additional simulations are given in the Online Supplement.

2 Setup and Notation

Potential outcomes for treated and control groups are denoted by $Y(1)$ and $Y(0)$, respectively. Treatment status is denoted by A, with $A = 1$ indicating treated and $A = 0$ untreated. The stratum indicator is denoted by S, based on which the researcher implements the covariate-adaptive randomization. The support of S is denoted by $\mathcal{S}$, a finite set. After randomization, the researcher can observe the data $\{Y_i, S_i, A_i, X_i\}_{i=1}^n$ where $Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$ is the observed outcome and X_i contains extra covariates besides S_i in the dataset. The support of X is denoted Supp(X). In this paper, we allow X_i and S_i to be dependent. For $1 \leq i \leq n$ and $[n] = \{1, 2, ...n\}$, let $p(s) = \mathbb{P}(S_i = s)$, $n(s) = \sum_{i\in[n]} 1\{S_i = s\}$, $n_1(s) = \sum_{i\in[n]} A_i 1\{S_i = s\}$, and $n_0(s) = n(s) - n_1(s)$. We make the following assumptions on the data generating process (DGP) and the treatment assignment rule.

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

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

(iii) Suppose $p(s)$ *is fixed w.r.t.* n *and is positive for every* $s \in \mathcal{S}$.

(iv) Let $\pi(s)$ *denote the propensity score for stratum* s*. Then,* $c < \min_{s\in\mathcal{S}} \pi(s) \leq \max_{s\in\mathcal{S}} \pi(s) < 1 - c$ *for some constant* $c \in (0, 0.5)$ *and* $\frac{D_n(s)}{n(s)} = o_p(1)$ *for* $s \in \mathcal{S}$*, where* $D_n(s) = \sum_{i=1}^n (A_i - \pi(s))1\{S_i = s\}$.

Several remarks are in order. First, Assumption 1(i) allows for the cross-sectional dependence among treatment statuses ($\{A_i\}_{i\in[n]}$), thereby accommodating many covariate-adaptive randomization schemes as discussed below. Second, although treatment statuses are cross-sectionally dependent, they are independent of the potential outcomes and additional covariates conditional on the stratum indicator S. Therefore, data are still experimental rather than observational. Third, Assumption 1(iii) requires the size of each stratum to be proportional to the sample size. Fourth, we can view$\pi(s)$ as the target fraction of treated units in stratum s. Similar to Bugni et al. (2019), we allow the target fractions to differ across strata. Just as for the overlapping support condition in an observational study, the target fractions are assumed to be bounded away from zero and one. In randomized experiments, this condition usually holds because investigators can determine $\pi(s)$ in the design stage. In fact, in most CARs, $\pi(s) = 0.5$ for $s \in \mathcal{S}$. Fifth, $D_n(s)$ represents the degree of imbalance between the real and target factions of treated units in the sth stratum. Bugni et al. (2018) show that Assumption 1(iv) holds under several covariate-adaptive treatment assignment rules such as simple random sampling (SRS), biased-coin design (BCD), adaptive biased-coin design (WEI), and stratified block randomization (SBR). For completeness, we briefly repeat their descriptions below. Note we only require $D_n(s)/n(s) = o_p(1)$, which is weaker than the assumption imposed by Bugni et al. (2018) but the same as that imposed by Bugni et al. (2019) and Zhang and Zheng (2020).

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

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

Then, Assumption 1(iv) holds.

Example 2 (WEI)**.** *This design was first proposed by Wei (1978). Let $n_{k-1}(S_k) = \sum_{i=1}^{k-1} 1\{S_i = S_k\}$, $D_{k-1}(s) = \sum_{i=1}^{k-1}\left(A_i - \frac{1}{2}\right) 1\{S_i = s\}$, 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. (2018) show that Assumption 1(iv) holds with $\pi(s) = \frac{1}{2}$. Recently, Hu (2016) generalized the adaptive biased-coin design to multiple treatment values and unequal target factions.

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. (2018) show that Assumption 1(iv) holds with $\pi = \frac{1}{2}$.

Example 4 (SBR)**.** *For each stratum, $\lfloor \pi(s)n(s) \rfloor$ units are assigned to treatment and the rest are assigned to control. Then obviously Assumption 1(iv) holds as $n(s) \to \infty$.*

Denote the τth quantile of $Y(a)$ by $q_a(\tau)$ for $a = 0, 1$. We are interested in estimating and inferring the τth quantile treatment effect defined as $q(\tau) = q_1(\tau) - q_0(\tau)$. The testing problems of interest involve single, multiple, or even a continuum of quantile indexes, as in the following null hypotheses

$$\mathcal{H}_0 : q(\tau) = \underline{q} \quad \text{versus} \quad q(\tau) \neq \underline{q},$$

$$\mathcal{H}_0 : q(\tau_1) - q(\tau_2) = \underline{q} \quad \text{versus} \quad q(\tau_1) - q(\tau_2) \neq \underline{q}, \text{ and}$$

$$\mathcal{H}_0 : q(\tau) = \underline{q}(\tau) \; \forall \tau \in \Upsilon \quad \text{versus} \quad q(\tau) \neq \underline{q}(\tau) \text{ for some } \tau \in \Upsilon,$$

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

3 Estimation

Our estimator is based on the doubly robust moments for the quantiles of $Y(1)$ and $Y(0)$. Define $m_a(\tau, s, x) = \tau - \mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i = x)$. Firpo (2007), Belloni, Chernozhukov, Fernández-Val, and Hansen (2017), and Kallus, Mao, and Uehara (2019) showed that the doubly robust moment for $q_1(\tau)$ is

$$\mathbb{E}\left[\frac{A_i(\tau - 1\{Y_i(1) \leq q\})}{\overline{\pi}(S_i)} - \frac{A_i - \overline{\pi}(S_i)}{\overline{\pi}(S_i)}\overline{m}_1(\tau, S_i, X_i)\right] = 0, \tag{3.1}$$

where $\overline{\pi}(s)$ and $\overline{m}_1(\tau, s, x)$ are the working models for the propensity score ($\pi(s)$) and conditional probability ($m_1(\tau, s, x)$), respectively. In CAR, the propensity score is usually known or can be consistently estimated by $\hat{\pi}(s) = n_1(s)/n(s)$. Therefore, $\overline{\pi}(s)$ is correctly specified as $\pi(s)$. Then, due to the double robustness, even though the working model $\overline{m}_1(\tau, s, x)$ is misspecified, (3.1) still identifies the true unconditional quantile $q_1(\tau)$. Therefore, it is expected that the estimator of $q_1(\tau)$ based on the sample analogue of (3.1) is still consistent. In addition, (3.1) also satisfies the Neyman orthogonality condition w.r.t. the working models $\overline{\pi}(\cdot)$ and $\overline{m}_1(\cdot)$. This implies our QTE estimator still has the usual parametric rate of convergence even when the estimators of the working models converge at a nonparametric rate.

To proceed, we note that (3.1) is the first-order condition of the optimization problem

$$\min_q \mathbb{E}\left[\frac{A_i}{\pi(S_i)}\rho_\tau(Y_i - q) + \frac{A_i - \pi(S_i)}{\pi(S_i)}\overline{m}_1(\tau, S_i, X_i)q\right], \tag{3.2}$$

where $\rho_\tau(u) = u(\tau - 1\{u \leq 0\})$ is the usual check function. Denote $\widehat{m}_1(\cdot)$ as a consistent estimator of $\overline{m}_1(\cdot)$ obtained via an auxiliary regression. By the sample analog of (3.2), our regression-adjusted estimator of $q_1(\tau)$, denoted as $\hat{q}_1^{adj}(\tau)$, can be defined as

$$\hat{q}_1^{adj}(\tau) = \arg\min_q \sum_{i \in [n]} \left[\frac{A_i}{\hat{\pi}(S_i)}\rho_\tau(Y_i - q) + \frac{(A_i - \hat{\pi}(S_i))}{\hat{\pi}(S_i)}\widehat{m}_1(\tau, S_i, X_i)q\right]. \tag{3.3}$$

Here, we emphasize that $\widehat{m}_1(\cdot)$ may not consistently estimate the true specification $m_1(\cdot)$. Similarly, we can define

$$\hat{q}_0^{adj}(\tau) = \arg\min_q \sum_{i\in[n]} \left[\frac{1-A_i}{1-\hat{\pi}(S_i)} \rho_\tau(Y_i - q) - \frac{(A_i - \hat{\pi}(S_i))}{1-\hat{\pi}(S_i)} \widehat{m}_0(\tau, S_i, X_i) q \right], \tag{3.4}$$

where $\widehat{m}_0(\cdot)$ is a consistent estimator of $\overline{m}_0(\cdot)$. Then, our regression adjusted QTE estimator is

$$\hat{q}^{adj}(\tau) = \hat{q}_1^{adj}(\tau) - \hat{q}_0^{adj}(\tau). \tag{3.5}$$

We use the estimated propensity score $\hat{\pi}(s)$ even when $\pi(s)$ is known. Consider the special case of our setup for which there is no regression adjustment ($\widehat{m}_a(\cdot) = 0$) and $\pi(s) = 1/2$ for $s \in \mathcal{S}$. In the design stage, researchers can use simple random sampling (SRS) or CAR (such as BCD, WEI, or SBR defined above). In the analysis stage, researchers can choose the difference in quantile estimator (DIQ) which is just the difference of the τth quantiles of the treated and control groups or the IPW estimator with $\hat{\pi}(s)$ defined in (3.3)–(3.5). Note when $\pi(s) = 1/2$, the DIQ estimator can be viewed as an IPW estimator in which $\hat{\pi}(s)$ is replaced by the true propensity score $\pi(s) = 1/2$. In this simplified setting, Zhang and Zheng (2020) derived the asymptotic variances of two estimators under SRS and CAR, which can be summarized below.

Table 1: Asymptotic Variance of DIQ and IPW estimators

	DIQ	IPW
SRS	$\Sigma(1/4)$	$\Sigma(0)$
CAR(δ)	$\Sigma(\delta)$	$\Sigma(0)$

In Table 1, the asymptotic variance $\Sigma(\delta)$ is a monotone increasing function of δ and $\delta \in [0, 1/4]$ indicates the balance level achieved by the CAR scheme. When $\delta = 0$, the CAR achieves the strong balance. For the four examples of CARs mentioned earlier, we have $\delta(\text{SRS}) = \frac{1}{4}$, $\delta(\text{WEI}) = \frac{1}{4}(1 - \phi'(0))^{-1}$, $\delta(\text{BCD}) = 0$, and $\delta(\text{SBR}) = 0$. From Table 1, we can make two observations. First, the asymptotic variance of the QTE estimator with $\hat{\pi}(s)$ (i.e., IPW) is no larger than that with $\pi(s)$ (i.e., DIQ). Second, when we conduct multiplier bootstrap inference, the bootstrap weights are independent cross-sectionally. Therefore, conditionally on the data, the observations in the bootstrap sample mimic the SRS scheme. The bootstrap variance for the DIQ estimator is conservative because $\Sigma(1/4) \geq \Sigma(\delta)$ while it is not for the IPW estimator. This motivates us to use the IPW estimator with the estimated propensity score $\hat{\pi}(s)$ for both estimation and bootstrap inference, even when $\pi(s)$ is known. In the simulation study in Section 7, we show that, comparing with the multiplier bootstrap inference with the estimated propensity score, a similar inference procedure with the true propensity score is conservative under the null (except for SRS) and lacks power under the alternative.

In this and the next sections, we establish the limit distribution and validity of bootstrap inference for $\hat{q}^{adj}(\tau)$ under high-level conditions on $(\overline{m}_a(\cdot), \widehat{m}_a(\cdot))_{a=0,1}$. In Section 5, we verify these conditions when $(\overline{m}_a(\cdot))_{a=0,1}$ are correctly or incorrectly specified and $(\widehat{m}_a(\cdot))_{a=0,1}$ are computed parametrically, nonparametrically, or via regularization when the covariate X is high-dimensional.

Assumption 2. *For $a = 0, 1$, denote $f_a(\cdot)$, $f_a(\cdot|s)$, and $f_a(\cdot|x, s)$ as the PDFs of $Y_i(a)$, $Y_i(a)|S_i = s$, and $Y_i(a)|S_i = s, X_i = x$, respectively.*

(i) $f_a(q_a(\tau))$ and $f_a(q_a(\tau)|s)$ are bounded and bounded away from zero uniformly over $\tau \in \Upsilon$ and $s \in \mathcal{S}$, where Υ is a compact subset of $(0, 1)$.

(ii) $f_a(\cdot)$ and $f_a(\cdot|s)$ are Lipschitz over $\{q_j(\tau) : \tau \in \Upsilon\}$.

(iii) $\sup_{y\in\Re, x\in Supp(X), s\in\mathcal{S}} f_a(y|x, s) < \infty$.

Assumption 3. *(i) For $a = 0, 1$, there exists a function $\overline{m}_a(\tau, x, s)$ such that for $\overline{\Delta}_a(\tau, s, X_i) = \widehat{m}_a(\tau, s, X_i) - \overline{m}_a(\tau, s, X_i)$, we have*

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{\sum_{i\in I_1(s)} \overline{\Delta}_a(\tau, s, X_i)}{n_1(s)} - \frac{\sum_{i\in I_0(s)} \overline{\Delta}_a(\tau, s, X_i)}{n_0(s)} \right| = o_p(n^{-1/2}),$$

where $I_a(s) = \{i \in [n] : A_i = a, S_i = s\}$.

(ii) For $a = 0, 1$, let $\mathcal{F}_a = \{\overline{m}_a(\tau, S_i, X_i) : \tau \in \Upsilon\}$ with an envelope $F_{a,i}$. Then, $\max_{s\in\mathcal{S}} \mathbb{E}(|F_{a,i}|^q|S_i = s) < \infty$ for $q > 2$ and there exist fixed constants $(\alpha, v) > 0$ such that

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

where $N(\cdot)$ denotes the covering number, $e_Q(f, g) = ||f - g||_{Q,2}$, and the supremum is taken over all finitely discrete probability measures Q.

(iii) For any $\tau_1, \tau_2 \in \Upsilon$, there exists a constant $C > 0$ such that

$$\mathbb{E}((\overline{m}_a(\tau_2, S_i, X_i) - \overline{m}_a(\tau_1, S_i, X_i))^2|S_i = s) \leq C|\tau_2 - \tau_1|.$$

Several remarks are in order. First, Assumption 2 is standard in the quantile regression literature. We do not need $f_a(y|x, s)$ to be bounded away from zero because we are interested in the unconditional quantile $q_a(\tau)$, which is uniquely defined as long as the unconditional density $f_a(q_a(\tau))$ is positive. Second, Assumption 3(i) is high-level. If we consider a linear probability model such that $\overline{m}_a(\tau, s, X_i) = \tau - X_i^\top \theta_{a,s}(\tau)$ and $\widehat{m}_a(\tau, s, X_i) = \tau - X_i^\top \hat{\theta}_{a,s}(\tau)$, then Assumption 3(i) is equivalent to

$$\sup_{\tau\in\Upsilon, a=0,1, s\in\mathcal{S}} \left| \left(\frac{\sum_{i\in I_1(s)} X_i}{n_1(s)} - \frac{\sum_{i\in I_0(s)} X_i}{n_0(s)} \right)^\top \left(\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau)\right) \right| = o_p(n^{-1/2}),$$

which is similar to Liu et al. (2020, Assumption 3) and holds intuitively if $\hat{\theta}_{a,s}(\tau)$ is a consistent estimator of the pseudo true value $\theta_{a,s}(\tau)$. Third, Assumptions 3(ii) and 3(iii) impose mild regularity conditions on $\overline{m}_a(\cdot)$. They hold if

$$\sup_{a=0,1, s\in\mathcal{S}, x\in \text{Supp}(X)} |\overline{m}_a(\tau_2, s, x) - \overline{m}_a(\tau_1, s, x)| \leq L|\tau_2 - \tau_1|.$$

for some constant $L > 0$. Such Lipschitz continuity holds for the true specification ($\overline{m}_a(\cdot) = m_a(\cdot)$) under Assumption 2. Fourth, we provide primitive sufficient conditions for Assumption 3 in Section 5.

Theorem 3.1. *Suppose Assumptions 1–3 hold. Then, uniformly over $\tau \in \Upsilon$,*

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

where $\mathcal{B}(\tau)$ is a tight Gaussian process with covariance kernel $\Sigma(\tau, \tau')$ defined in Section F of the Online Supplement. In addition, for any finite set of quantile indexes $(\tau_1, \cdots, \tau_K)$, the asymptotic covariance matrix of $(\hat{q}^{adj}(\tau_1), \cdots, \hat{q}^{adj}(\tau_K))$ is denoted as $[\Sigma(\tau_k, \tau_l)]_{k,l\in[K]}$, where we use $[A_{kl}]_{k,l\in[K]}$ to denote a $K \times K$ matrix whose (k, l)th entry is A_{kl}. Then, $[\Sigma(\tau_k, \tau_l)]_{k,l\in[K]}$ is minimized in the matrix sense[1] when the auxiliary regressions are correctly specified at $(\tau_1, \cdots, \tau_K)$, i.e., $\overline{m}_a(\tau_k, s, x) = m_a(\tau_k, s, x)$ for $a = 0, 1$, $k \in [K]$, and all (s, x) in the joint support of (S_i, X_i).

Three remarks are in order. First, the expression for the asymptotic variance of $\hat{q}^{adj}(\tau)$ can be found in the proof of Theorem 3.1. It is the same whether the randomization scheme achieves strong balance or not. This robustness is due to the use of the estimated propensity score ($\hat{\pi}(s)$). The same phenomenon was discovered in the simplified setting by Zhang and Zheng (2020) as shown in the $(2, 2)$ entry of Table 1. Second, although our estimator is still consistent and asymptotically normal when the auxiliary regression is misspecified, it is meaningful to pursue the correct specification as it achieves the minimum variance. Third, the asymptotic variance of $\hat{q}^{adj}(\tau)$ depends on $(f_a(q_a(\tau)), m_a(\tau, s, x))_{a=0,1}$, which are infinite-dimensional nuisance parameters. To conduct analytic inference, it is necessary to nonparametrically estimate these nuisance parameters, which

[1] For two symmetric matrices A and B, we say A is greater than or equal to B if $A - B$ is positive semidefinite.

requires tuning parameters. Nonparametric estimation can be sensitive to the choice of tuning parameters and rule-of-thumb tuning parameter selection may not be appropriate for every data generating process (DGP) or every quantile. Use of cross-validation in selecting the tuning parameters is possible in principle but in practice time-consuming. These practical difficulties of analytic methods of inference provide strong motivation to investigate bootstrap inference procedures that are much less reliant on tuning parameters.

4 Multiplier Bootstrap Inference

We approximate the asymptotic distributions of $\hat{q}^{adj}(\tau)$ via the multiplier bootstrap. Let $\{\xi_i\}_{i=1}^n$ be a sequence of bootstrap weights which will be specified later. Define $n_1^w(s) = \sum_{i=1}^n \xi_i A_i 1\{S_i = s\}$, $n_1^w(s) = \sum_{i=1}^n \xi_i (1 - A_i) 1\{S_i = s\}$, $n^w(s) = \sum_{i=1}^n \xi_i 1\{S_i = s\} = n_1^w(s) + n_0^w(s)$, and $\hat{\pi}^w(s) = n_1^w(s)/n^w(s)$. The multiplier bootstrap counterpart of $\hat{q}^{adj}(\tau)$ is denoted by $\hat{q}^w(\tau)$ and defined as

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

where

$$\hat{q}_1^w(\tau) = \arg\min_q \sum_{i \in [n]} \xi_i \left[\frac{A_i}{\hat{\pi}^w(S_i)} \rho_\tau(Y_i - q) + \frac{(A_i - \hat{\pi}^w(S_i))}{\hat{\pi}^w(S_i)} \widehat{m}_1(\tau, S_i, X_i) q \right], \tag{4.1}$$

and

$$\hat{q}_0^w(\tau) = \arg\min_q \sum_{i \in [n]} \xi_i \left[\frac{1 - A_i}{1 - \hat{\pi}^w(S_i)} \rho_\tau(Y_i - q) - \frac{(A_i - \hat{\pi}^w(S_i))}{1 - \hat{\pi}^w(S_i)} \widehat{m}_0(\tau, S_i, X_i) q \right]. \tag{4.2}$$

Two comments on implementation are noted here: (i) we do not re-estimate $\widehat{m}_a(\cdot)$ in the bootstrap sample, which is similar to the multiplier bootstrap procedure proposed by Belloni et al. (2017); and (ii) in Section 6 we propose a way to directly compute $(\hat{q}_a^w(\tau))_{a=0,1}$ from the subgradient conditions of (4.1) and (4.2), thereby avoiding the optimization. Both features considerably reduce computation time of our bootstrap procedure.

Next, we specify the bootstrap weights.

Assumption 4. *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.*

Assumption 5. *Recall $\overline{\Delta}_a(\tau, s, x)$ defined in Assumption 3. We have, for $a = 0, 1$,*

$$\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_1(s)} \xi_i \overline{\Delta}_a(\tau, s, X_i)}{n_1^w(s)} - \frac{\sum_{i \in I_0(s)} \xi_i \overline{\Delta}_a(\tau, s, X_i)}{n_0^w(s)} \right| = o_p(n^{-1/2}).$$

We require the bootstrap weights to be nonnegative so that the objective functions in (4.1) and (4.2) are convex. In practice, we generate ξ_i independently from the standard exponential distribution. Assumption 5 is the bootstrap counterpart of Assumption 3. Continuing with the linear model example considered after Assumption 3, Assumption 5 requires

$$\sup_{\tau \in \Upsilon, a=0,1, s \in \mathcal{S}} \left| \left(\frac{\sum_{i \in I_1(s)} \xi_i X_i}{n_1(s)} - \frac{\sum_{i \in I_0(s)} \xi_i X_i}{n_0(s)} \right)^\top \left(\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau) \right) \right| = o_p(n^{-1/2}),$$

which holds if $\hat{\theta}_{a,s}(\tau)$ is a uniformly consistent estimator of $\theta_{a,s}(\tau)$.

Theorem 4.1. *Suppose Assumptions 1–5 hold. Then, uniformly over $\tau \in \Upsilon$ and conditionally on data,*

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

where $\mathcal{B}(\tau)$ is the same Gaussian process defined in Theorem 3.1.[2]

Theorem 4.1 shows the limit distribution of the bootstrap estimator conditional on data can approximate that of the original estimator uniformly over $\tau \in \Upsilon$. This provides the theoretical foundation for our bootstrap confidence intervals and bands described in Section 6. We also emphasize that the distribution of the bootstrap estimator conditional on data can consistently approximate that of the original estimator under CAR because the estimated propensity score, i.e., $\hat{\pi}(s)$, is used in (4.1) and (4.2).

[2]We view $\sqrt{n}(\hat{q}^w(\tau) - \hat{q}^{adj}(\tau))$ and $\mathcal{B}(\tau)$ as two processes indexed by $\tau \in \Upsilon$ and denote them as G_n and G, respectively. Then, following van der Vaart and Wellner (1996, Chapter 2.9), we say G_n weakly converges to G conditionally on data and uniformly over $\tau \in \Upsilon$ if

$$\sup_{h \in \mathrm{BL}_1} |\mathbb{E}_\xi h(G_n) - \mathbb{E} h(G)| \xrightarrow{p} 0,$$

where BL_1 is the set of all functions $h : \ell^\infty(\Upsilon) \mapsto [0,1]$ such that $|h(z_1) - h(z_2)| \le |z_1 - z_2|$ for every $z_1, z_2 \in \ell^\infty(\Upsilon)$, and $\mathbb{E}_\xi$ denotes expectation with respect to the bootstrap weights $\{\xi\}_{i=1}^n$.

5 Auxiliary Regressions

In this section, we consider three approaches to estimation for the auxiliary regressions: (1) a parametric method, (2) a nonparametric method, and (3) a regularization method. For the parametric method, we do not require the model to be correctly specified. We propose ways to estimate the pseudo true value of the auxiliary regression. For the other two methods, we (nonparametrically) estimate the true model so that the asymptotic variance of $\hat{q}^{adj}(\tau)$ achieves its minimum based on Theorem 3.1. For all three methods, we verify Assumptions 3 and 5.

5.1 Parametric method

In this section, we consider the case where X_i is finite-dimensional. Recall $m_a(\tau, S_i, X_i) \equiv \tau - \mathbb{P}\left(Y_i(a) \leq q_a(\tau)|X_i, S_i\right)$ for $a = 0, 1$. We propose to model $\mathbb{P}\left(Y_i(a) \leq q_a(\tau)|X_i, S_i = s\right)$ as $\Lambda_{a,s}(X_i, \theta_{a,s}(\tau))$ so that our model for $m_a(\tau, S_i, X_i)$ is

$$\overline{m}_a(\tau, S_i, X_i) = \tau - \sum_{s \in \mathcal{S}} 1\{S_i = s\}\Lambda_{a,s}(X_i, \theta_{a,s}(\tau)) \tag{5.1}$$

We propose to estimate $\theta_{a,s}(\tau)$ by $\hat{\theta}_{a,s}(\tau)$. The corresponding $\widehat{m}(\tau, S_i, X_i)$ can be written as

$$\widehat{m}_a(\tau, S_i, X_i) = \tau - \sum_{s \in \mathcal{S}} 1\{S_i = s\}\Lambda_{a,s}(X_i, \hat{\theta}_{a,s}(\tau)) \tag{5.2}$$

Assumption 6. *(i) Suppose there exist a positive random variable L_i and a positive constant $C > 0$ such that*

$$\sup_{\tau \in \Upsilon, a=0,1, s\in \mathcal{S}} ||\partial_\theta \Lambda_{a,s}(X_i, \theta_{a,s}(\tau))||_2 \leq L_i,$$
$$\sup_{\tau \in \Upsilon, a=0,1, s\in \mathcal{S}} |\Lambda_{a,s}(X_i, \theta_{a,s}(\tau))| \leq L_i,$$

and $\mathbb{E}(L_i^q|S_i = s) < \infty$.

(ii) Suppose $\sup_{\tau_1,\tau_2 \in \Upsilon, a=0,1, s\in\mathcal{S}} |\theta_{a,s}(\tau_1) - \theta_{a,s}(\tau_2)| \leq C|\tau_1 - \tau_2|$.

(iii) There exists $\hat{\theta}_{a,s}(\tau)$ such that $\sup_{\tau \in \Upsilon, a=0,1, s\in\mathcal{S}} ||\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau)||_2 \xrightarrow{p} 0$.

Theorem 5.1. *Denote $\hat{q}^{par}(\tau)$ and $\hat{q}^{par,w}(\tau)$ as the τth QTE estimator and its multiplier bootstrap counterpart defined in Sections 3 and 4, respectively, with $\overline{m}_a(\tau, S_i, X_i)$ and $\widehat{m}_a(\tau, S_i, X_i)$ defined in (5.1) and (5.2), respectively. Suppose Assumptions 1, 2, 4, and 6 hold. Then, Assumptions 3 and 5 hold, which further implies Theorems 3.1 and 4.1 hold for $\hat{q}^{par}(\tau)$ and $\hat{q}^{par,w}(\tau)$, respectively.*

Several remarks are in order. First, common choices for auxiliary regressions are linear probability, logistic, and probit regressions, corresponding to $\Lambda_{a,s}(X_i, \theta_{a,s}(\tau)) = X_i^\top \theta_{a,s}(\tau)$, $\lambda(\vec{X}_i^\top \theta_{a,s}(\tau))$, and $\Phi(\vec{X}_i^\top \theta_{a,s}(\tau))$, respectively, where $\lambda(\cdot)$ and $\Phi(\cdot)$ are the logistic and standard normal CDFs, respectively, and $\vec{X}_i = (1, X_i^\top)^\top$. For the linear regression case, we do not include the intercept because our regression adjusted estimators ((3.3) and (3.4)) and their bootstrap counterparts ((4.1) and (4.2)) are numerically invariant to location shift of the auxiliary models. Second, Theorem 5.1 shows that, as long as the estimator of the pseudo true value ($\hat{\theta}_{a,s}(\tau)$) is uniformly consistent, under mild regularity conditions, all the general estimation and bootstrap inference results established in Sections 3 and 4 hold. Third, it is still in question what pseudo true value one should target. We note that the asymptotic variance (denoted as σ^2) of the $\hat{q}^{adj}(\tau)$ is a function of the working model ($\overline{m}_a(\tau, s, \cdot)$), which is further indexed by its parameters (denoted as $\{t_{a,s}(\tau)\}_{a=0,1,s\in\mathcal{S}}$), i.e., $\sigma^2 = \sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}})$. One option is to target the pseudo true value that minimizes $\sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}})$, i.e.,

$$\{\theta_{a,s}(\tau)\}_{a=0,1,s\in\mathcal{S}} \in \mathop{\arg\min}_{t_{a,s}:a=0,1,s\in\mathcal{S}} \sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}}).$$

Theorem 5.2 below characterizes such $\{\theta_{a,s}(\tau)\}_{a=0,1,s\in\mathcal{S}}$. We then consider the consistent estimator of these pseudo true values for linear and logistic models in Propositions 5.1 and 5.2, respectively. Fourth, as the working model may be misspecified, the pseudo true value that minimizes $\sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}})$ may not be the best parameter value that fits the data. On the other hand, the minimum of $\sigma^2(\{\overline{m}_a(\tau, s, \cdot; t_{a,s})\}_{a=0,1,s\in\mathcal{S}})$ w.r.t. $\{t_{a,s}\}_{a=0,1,s\in\mathcal{S}}$ is still no smaller than the one attained by minimizing the whole working model, i.e., $\min_{\overline{m}_a(\tau,s,\cdot),a=0,1,s\in\mathcal{S}} \sigma^2(\{\overline{m}_a(\tau, s, \cdot)\}_{a=0,1,s\in\mathcal{S}})$. The latter minimum is attained if $\overline{m}_a(\tau, s, x) = m_a(\tau, s, x)$ for $a = 0, 1$ and (s, x) in the joint support of (S_i, X_i), as shown in Theorem 3.1. Therefore, it is still meaningful to pursue a working model that can closely approximate the true model. In Proposition 5.3, we propose to estimate a parametric but flexible distribution regression model via quasi maximum likelihood estimation (QMLE). We give our practical recommendation after that.

Theorem 5.2. *Suppose Assumptions 1, 2, 4, 6 hold, and $\Lambda_{a,s}(X_i, \theta_{a,s}(\tau))$ is differentiable in $\theta_{a,s}(\tau)$. Then, the asymptotic variance of $\hat{q}^{par}(\tau)$ is minimized at $(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau)$, where for $s \in \mathcal{S}$ and $\tau \in \Upsilon$,*

$$\Theta_s(\tau) = \mathop{\arg\min}_{\theta_1,\theta_0} Q(s, \tau, \theta_1, \theta_0), \text{ where}$$

$$Q(s, \tau, \theta_1, \theta_0) = \mathbb{E}\Bigg\{ \left(\frac{g_{1,s}(X_i, \theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i, \theta_0)}{f_0(q_0(\tau))} \right)^2$$

$$
\begin{aligned}
&- 2\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\left(\frac{m_1(\tau,s,X_i)-m_1(\tau,s)}{f_1(q_1(\tau))}\right) \\
&- \frac{2\pi(s)}{1-\pi(s)}\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\left(\frac{m_0(\tau,s,X_i)-m_0(\tau,s)}{f_0(q_0(\tau))}\right)\Bigg| S_i = s\Bigg\},
\end{aligned}
\tag{5.3}
$$

and $g_{a,s}(X_i,\theta_a) = \mathbb{E}(\Lambda_{a,s}(X_i,\theta_a)|S_i = s) - \Lambda_{a,s}(X_i,\theta_a)$. If we further assume a linear probability model, i.e., $\Lambda_{a,s}(X_i,\theta_{a,s}(\tau)) = X_i^\top\theta_{a,s}(\tau)$, and denote the asymptotic covariance matrix of $(\hat{q}^{par}(\tau_1),\cdots,\hat{q}^{par}(\tau_K))$ for any finite set of quantile indexes $(\tau_1,\cdots,\tau_K)$ as $[\Sigma^{LP}(\tau_k,\tau_l)]_{k,l\in[K]}$. Then, $[\Sigma^{LP}(\tau_k,\tau_l)]_{k,l\in[K]}$ is minimized in the matrix sense when the pseudo true values $(\theta_{1,s}(\tau_k),\theta_{0,s}(\tau_k))_{k\in[K]}$ satisfy

$$
\frac{\theta_{1,s}(\tau_k)}{f_1(q_1(\tau_k))} + \frac{\pi(s)\theta_{0,s}(\tau_k)}{(1-\pi(s))f_0(q_0(\tau_k))} = \frac{\theta^{LP}_{1,s}(\tau_k)}{f_1(q_1(\tau_k))} + \frac{\pi(s)\theta^{LP}_{0,s}(\tau_k)}{(1-\pi(s))f_0(q_0(\tau_k))},\ k \in [K],
$$

where $[K] = \{1,2,...,K\}$ and

$$
\begin{aligned}
\theta^{LP}_{1,s}(\tau_k) &= \left[\mathbb{E}(\tilde{X}_{i,s}\tilde{X}_{i,s}^\top|S_i = s, A_i = 1)\right]^{-1}\mathbb{E}\left[\tilde{X}_{i,s}1\{Y_i \le q_1(\tau_k)\}|S_i = s, A_i = 1\right], \\
\theta^{LP}_{0,s}(\tau_k) &= \left[\mathbb{E}(\tilde{X}_{i,s}\tilde{X}_{i,s}^\top|S_i = s, A_i = 0)\right]^{-1}\mathbb{E}\left[\tilde{X}_{i,s}1\{Y_i \le q_0(\tau_k)\}|S_i = s, A_i = 0\right],
\end{aligned}
$$

and $\tilde{X}_{i,s} = X_i - \mathbb{E}(X_i|S_i = s) = X_i - \mathbb{E}(X_i|S_i = s, A_i = 1) = X_i - \mathbb{E}(X_i|S_i = s, A_i = 0)$.

Two remarks are in order. First, we propose consistent estimators of $\theta_{a,s}(\tau)$ for linear probability and logistic models below. The corresponding results for the probit model are similar to those of the logistic model and are therefore omitted. Second, when $\mathbb{E}(||X_i||_2^q|S_i = s) < \infty$ for some $q > 2$, Assumption 6(i) holds for these two models. We maintain Assumption 6(ii) as it usually holds under the regularity conditions imposed in Assumption 2. In the following, we propose estimators for the pseudo true values and establish Assumption 6(iii).

5.1.1 Linear probability model

Consider the linear probability model $\Lambda_{a,s}(X_i,\theta_a) = X_i^\top\theta_a$. From Theorem 5.2, we see that $\Theta_s(\tau)$ is not a singleton. In order to achieve the minimal variance, we only need to consistently estimate one point in the set $\Theta_s(\tau)$. We choose

$$
(\theta_{1,s}(\tau),\theta_{0,s}(\tau)) = (\theta^{LP}_{1,s}(\tau),\theta^{LP}_{0,s}(\tau)),\ s \in \mathcal{S},
$$

as this choice avoids estimation of the densities $f_1(q_1(\tau))$ and $f_0(q_0(\tau))$.

Note $\theta^{LP}_{a,s}(\tau)$ is the projection coefficient of $1\{Y_i \le q_a(\tau)\}$ on $\tilde{X}_{i,s}$ for the sub-population with

$S_i = s$ and $A_i = a$. We estimate them by the sample analog. The parameter $q_a(\tau)$ is unknown, and thus, is replaced by some $\sqrt{n}$-consistent estimator denoted as $\hat{q}_a(\tau)$.

Assumption 7. *Assume that* $\sup_{\tau \in \Upsilon, a=0,1} |\hat{q}_a(\tau) - q_a(\tau)| = O_p(n^{-1/2})$.

In practice, we compute $\{\hat{q}_a(\tau)\}_{a=0,1}$ based on (3.3) and (3.4) by setting $\hat{m}_a(\tau, S_i, X_i) \equiv 0$. Then, Assumption 7 holds automatically by Theorem 3.1 with $\widehat{m}_a(\tau, S_i, X_i) = \overline{m}_a(\tau, S_i, X_i) = 0$.

Next, we define the estimator of $\theta_{a,s}^{LP}(\tau)$. Let

$$\overline{m}_a(\tau, S_i, X_i) = \tau - \sum_{s \in \mathcal{S}} 1\{S_i = s\} X_i^\top \theta_{a,s}^{LP}(\tau), \tag{5.4}$$

$$\widehat{m}_a(\tau, S_i, X_i) = \tau - \sum_{s \in \mathcal{S}} 1\{S_i = s\} X_i^\top \hat{\theta}_{a,s}^{LP}(\tau), \tag{5.5}$$

$$\hat{X}_{i,a,s} = X_i - \frac{1}{n_a(s)} \sum_{i \in I_a(s)} X_i, \tag{5.6}$$

and

$$\hat{\theta}_{a,s}^{LP}(\tau) = \left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} \hat{X}_{i,a,s} \hat{X}_{i,a,s}^\top \right]^{-1} \left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} \hat{X}_{i,a,s} 1\{Y_i \leq \hat{q}_a(\tau)\} \right]. \tag{5.7}$$

Assumption 8. *There exist constants* $0 < c < C < \infty$ *such that*

$$c < \min_{s \in \mathcal{S}} \lambda_{\min}(\mathbb{E}\tilde{X}_{i,s}\tilde{X}_{i,s}^\top | S_i = s) \leq \max_{s \in \mathcal{S}} \lambda_{\max}(\mathbb{E}\tilde{X}_{i,s}\tilde{X}_{i,s}^\top | S_i = s) \leq C$$

and $\mathbb{E}(||\tilde{X}_{i,s}||_2^q | S_i = s) \leq C$ *for some* $q > 2$*, where for a generic symmetric matrix* A*,* $\lambda_{\min}(A)$ *and* $\lambda_{\max}(A)$ *denote the minimal and maximal eigenvalues of* A*, respectively.*

Proposition 5.1. *Suppose Assumptions 1, 2, 7, and 8 hold. Then Assumptions 6(ii) and 6(iii) hold for* $(\theta_{a,s}(\tau), \hat{\theta}_{a,s}(\tau)) = (\theta_{a,s}^{LP}(\tau), \hat{\theta}_{a,s}^{LP}(\tau)),\ a = 0, 1, s \in \mathcal{S}, \tau \in \Upsilon$.

By construction, the regression-adjusted QTE estimator with a linear auxiliary regression and the pseudo true value $\theta_{a,s}^{LP}(\tau)$) is weakly more efficient than the estimator without any adjustments because the latter corresponds to the regression-adjusted estimator when the linear coefficients are all zero.

5.1.2 Logistic model

Next consider the logistic model. Set $\lambda(u) = \exp(u)/(1+\exp(u))$ to be the logistic CDF and then $\Lambda_{a,s}(X_i, \theta_{a,s}(\tau)) = \lambda(\vec{X}_i^\top \theta_{a,s}(\tau))$ and

$$\widehat{m}_a(\tau, S_i, X_i) = \tau - \sum_{s \in \mathcal{S}} 1\{S_i = s\} \lambda(\vec{X}_i^\top \hat{\theta}_{a,s}^{LG}(\tau)), \tag{5.8}$$

where $\hat{\theta}_{a,s}^{LG}(\tau)$ is computed as the minimizer of the sample analogue of the objective function in (5.3) and $\vec{X}_i = (1, X_i^\top)^\top$. Specifically, let $I(s) = I_1(s) \cup I_0(s)$, $\hat{g}_{a,s}(X_i, \theta_a) = \frac{1}{n_a(s)} \sum_{i \in I_a(s)} \lambda(\vec{X}_i^\top \theta_a) - \lambda(\vec{X}_i^\top \theta_a)$, and let $\hat{f}_a(\cdot)$ be an estimator of the density of $Y(a)$ (i.e., $f_a(\cdot)$). Define $\Theta = \Theta_1 \times \Theta_0$, where Θ_a is a compact set in $\Re^{d_x}$ for $a = 0, 1$, and then

$$\begin{aligned}
&(\hat{\theta}_{1,s}^{LG}(\tau), \hat{\theta}_{0,s}^{LG}(\tau)) \in \operatorname*{arg\,min}_{(\theta_1, \theta_0) \in \Theta} Q_n(s, \tau, \theta_1, \theta_0), \text{ with} \\
&Q_n(s, \tau, \theta_1, \theta_0) = \frac{1}{n(s)} \sum_{i \in I(s)} \left(\frac{\hat{g}_{1,s}(X_i, \theta_1)}{\hat{f}_1(\hat{q}_1(\tau))} + \frac{\hat{\pi}(s)}{1 - \hat{\pi}(s)} \frac{\hat{g}_{0,s}(X_i, \theta_0)}{\hat{f}_0(\hat{q}_0(\tau))} \right)^2 \\
&+ \frac{2}{n_1(s)} \sum_{i \in I_1(s)} \left(\frac{\hat{g}_{1,s}(X_i, \theta_1)}{\hat{f}_1(\hat{q}_1(\tau))} + \frac{\hat{\pi}(s)}{1 - \hat{\pi}(s)} \frac{\hat{g}_{0,s}(X_i, \theta_0)}{\hat{f}_0(\hat{q}_0(\tau))} \right) \frac{1\{Y_i \le \hat{q}_1(\tau)\}}{\hat{f}_1(\hat{q}_1(\tau))} \\
&+ \frac{2\hat{\pi}(s)}{n_0(s)(1 - \hat{\pi}(s))} \sum_{i \in I_0(s)} \left(\frac{\hat{g}_{1,s}(X_i, \theta_1)}{\hat{f}_1(\hat{q}_1(\tau))} + \frac{\hat{\pi}(s)}{1 - \hat{\pi}(s)} \frac{\hat{g}_{0,s}(X_i, \theta_0)}{\hat{f}_0(\hat{q}_0(\tau))} \right) \frac{1\{Y_i \le \hat{q}_0(\tau)\}}{\hat{f}_0(\hat{q}_0(\tau))}.
\end{aligned} \tag{5.9}$$

Assumption 9. *Suppose $\Theta_s(\tau) \cap \Theta$ is a singleton, denoted as $(\theta_{1,s}^{LG}(\tau), \theta_{0,s}^{LG}(\tau))$, and*

$$\sup_{\tau \in \Upsilon} |\hat{f}_a(\hat{q}_a(\tau)) - f_a(q_a(\tau))| \xrightarrow{p} 0.$$

Proposition 5.2. *Suppose Assumptions 1, 2, 7–9 hold. Then Assumption 6(iii) holds for $(\theta_{a,s}(\tau), \hat{\theta}_{a,s}(\tau)) = (\theta_{a,s}^{LG}(\tau), \hat{\theta}_{a,s}^{LG}(\tau))$, $a = 0, 1, s \in \mathcal{S}, \tau \in \Upsilon$.*

By construction, the regression-adjusted QTE estimator with a logistic auxiliary regression and the pseudo true value $\theta_{a,s}^{LG}(\tau))$ is weakly more efficient than the estimator without any adjustments because the latter corresponds to the regression-adjusted estimator when the logistic regression coefficients are all zero.[3] In Section D of the Online Supplement, we propose a way to consistently estimate one point in $\Theta_s(\tau) \cap \Theta$ when it is not a singleton.

[3] This implies the adjustment is a constant. In addition, note the regression-adjusted QTE estimator is variant to the location shift in the adjustment term.

5.1.3 Quasi maximum likelihood estimation

It is also common to estimate the logistic regression by maximum likelihood. The main goal for the working model is to approximate the true model as closely as possible. It is therefore useful to include additional technical regressors such as interactions in the logistic regression. The set of regressors used is defined as $H_i = H(X_i)$. Let $\hat{\theta}^{ML}_{a,s}(\tau)$ and $\theta^{ML}_{a,s}(\tau)$ be the quasi-ML estimator and its corresponding pseudo true value, respectively, i.e.,

$$\hat{\theta}^{ML}_{a,s}(\tau) = \arg\max_{\theta_a} \frac{1}{n_a(s)} \sum_{i \in I_a(s)} [1\{Y_i \leq \hat{q}_a(\tau)\} \log(\lambda(H_i^\top \theta_a)) + 1\{Y_i > \hat{q}_a(\tau)\} \log(1 - \lambda(H_i^\top \theta_a))] \tag{5.10}$$

and

$$\theta^{ML}_{a,s}(\tau) = \arg\max_{\theta_a} \mathbb{E}[1\{Y_i(a) \leq q_a(\tau)\} \log(\lambda(H_i^\top \theta_a)) + 1\{Y_i(a) > q_a(\tau)\} \log(1 - \lambda(H_i^\top \theta_a)) | S_i = s]. \tag{5.11}$$

We then define

$$\widehat{m}_a(\tau, S_i, X_i) = \tau - \sum_{s \in \mathcal{S}} 1\{S_i = s\} \lambda(H_i^\top \hat{\theta}^{ML}_{a,s}(\tau)). \tag{5.12}$$

In addition to the inclusion of technical regressors we allow the pseudo true value to vary across quantiles τ, giving another layer of flexibility of the model. Such a model is called distribution regression and was first proposed by Chernozhukov, Fernández-Val, and Melly (2013). We emphasize here that, although we aim to make the regression model as flexible as possible, our theoretical results do not require the model to be correctly specified.

Assumption 10. *Suppose $\theta^{ML}_{a,s}(\tau)$ is the unique minimizer defined in (5.11) for $a = 0, 1$.*

Theorem 5.3. *Suppose Assumptions 1, 2, 7, 8, 10 hold, then Assumption 6(iii) holds for $(\theta_{a,s}(\tau), \hat{\theta}_{a,s}(\tau)) = (\theta^{ML}_{a,s}(\tau), \hat{\theta}^{ML}_{a,s}(\tau))$, $a = 0, 1, s \in \mathcal{S}, \tau \in \Upsilon$.*

5.1.4 Remarks

From a practical perspective we prefer the linear probability (LP) and quasi maximum likelihood (ML) estimators to the logistic model estimator that minimizes the asymptotic variance of $\hat{q}^{adj}(\tau)$ (LG) for two reasons. First, the LP and ML estimators are free of tuning parameters whereas estimator LG requires nonparametric estimation of the densities, which involves tuning parameters.

Second, the estimators LP and ML are easy to compute: the LP estimator has a closed-form expression and the ML estimator is just simple logistic regression which can be implemented in standard statistical softwares. On the other hand, the LG estimator requires non-convex optimization and is therefore computationally costly.

From a theory perspective, we continue to prefer the LP and ML estimators. The LP estimator minimizes (over the class of linear models) not only the asymptotic variance of $\hat{q}^{par}(\tau)$ but also the covariance matrix of $(\hat{q}^{par}(\tau_1), \cdots, \hat{q}^{par}(\tau_K))$ for any finite-dimension quantile indexes $(\tau_1, \cdots, \tau_K)$. This implies we can use the same LP estimator for hypothesis testing involving single, multiple, or even a continuum of quantile indexes. In contrast, the LG estimator only aims to minimize the asymptotic variance for a single quantile index, and may not be optimal for tests that involve multiple quantile indexes.[4] Although the ML estimator is not guaranteed to be optimal, it is near optimal if the logistic model is close to the true model $m_a(\tau, S_i, X_i)$. To achieve this benefit we suggest including additional technical terms in the regression and allowing the regression coefficients to vary across τ. The next section justifies the use of the ML estimator with a flexible logistic model by letting the number of technical terms (or equivalently, the dimension of H_i) diverge to infinity, showing by this means that the ML estimator can indeed consistently estimate the true model and thereby achieve the minimum covariance matrix of the adjusted QTE estimator.

5.2 Nonparametric method

This section considers nonparametric estimation of $m_a(\tau, s, X_i)$ when the dimension of X_i is fixed as d_x. For ease of notation we assume all coordinates of X_i are continuously distributed. If in an application some elements of X are discrete the dimension d_x is interpreted as the dimension of the continuous covariates. All results in this section can then be extended in a conceptually straightforward manner by using the continuous covariates only within samples that are homogeneous in discrete covariates.

As $m_a(\tau, s, X_i)$ is nonparametrically estimated, we have $\overline{m}_a(\tau, s, X_i) = m_a(\tau, s, X_i) = \tau - \mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i)$. We estimate $\mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i)$ by the sieve method of fitting a logistic model, as studied in Hirano, Imbens, and Ridder (2003). Specifically, recall $\lambda(\cdot)$ is the logistic CDF and denote the number of sieve bases by h_n, which depends on the sample size n and can grow to infinity as $n \to \infty$. Let $H_{h_n}(x) = (b_{1n}(x), \cdots, b_{h_n n}(x))^\top$ where $\{b_{hn}(x)\}_{h \in [h_n]}$ is an h_n dimensional basis of a linear sieve space. We provide more details on the sieve space in Section 6.

[4] Section C of the supplement derives the logistic model estimator that minimizes the asymptotic variance of $\hat{q}^{par}(\tau_1) - \hat{q}^{par}(\tau_2)$.

Denote

$$\widehat{m}_a(\tau, s, X_i) = \tau - \lambda(H_{h_n}^\top(X_i)\hat{\theta}_{a,s}^{NP}(\tau)) \quad \text{and} \tag{5.13}$$

$$\hat{\theta}_{a,s}^{NP}(\tau) = \arg\max_{\theta_a} \frac{1}{n_a(s)} \sum_{i \in I_a(s)} \Big[1\{Y_i \le \hat{q}_a(\tau)\} \log(\lambda(H_{h_n}^\top(X_i)\theta_a)) + 1\{Y_i > \hat{q}_a(\tau)\} \log(1 - \lambda(H_{h_n}^\top(X_i)\theta_a)) \Big] \tag{5.14}$$

Assumption 11. *(i) There exist constants $0 < \kappa_1 < \kappa_2 < \infty$ such that with probability approaching one,*

$$\kappa_1 \le \lambda_{\min}\left(\frac{1}{n_a(s)} \sum_{i \in I_a(s)} H_{h_n}(X_i)H_{h_n}^\top(X_i)\right) \le \lambda_{\max}\left(\frac{1}{n_a(s)} \sum_{i \in I_a(s)} H_{h_n}(X_i)H_{h_n}^\top(X_i)\right) \le \kappa_2$$

and

$$\kappa_1 \le \lambda_{\min}\left(\mathbb{E}(H_{h_n}(X_i)H_{h_n}^\top(X_i)|S_i = s)\right) \le \lambda_{\max}\left(\mathbb{E}(H_{h_n}(X_i)H_{h_n}^\top(X_i)|S_i = s)\right) \le \kappa_2.$$

(ii) There exists an $h_n \times 1$ vector $\theta_{a,s}^{NP}(\tau)$ such that for $R_a(\tau, s, x) = \mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i = x) - \lambda(H_{h_n}^\top(x)\theta_{a,s}^{NP}(\tau))$, we have $\sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in Supp(X)} |R_a(\tau, s, x)| = o(1)$,

$$\sup_{a=0,1,\tau\in\Upsilon,s\in\mathcal{S}} \frac{1}{n_a(s)} \sum_{i \in I_a(s)} R_a^2(\tau, s, X_i) = O_p\left(\frac{h_n \log(n)}{n}\right),$$

and

$$\sup_{a=0,1,\tau\in\Upsilon,s\in\mathcal{S}} \mathbb{E}(R_a^2(\tau, s, X_i)|S_i = s) = O\left(\frac{h_n \log(n)}{n}\right).$$

(iii) There exists a constant $c \in (0, 0.5)$ such that

$$\begin{aligned} c &\le \inf_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in Supp(X)} \mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i = x) \\ &\le \sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in Supp(X)} \mathbb{P}(Y_i(a) \le q_a(\tau)|S_i = s, X_i = x) \le 1 - c. \end{aligned}$$

(iv) Suppose $\mathbb{E}(H_{h_n,h}^2(X_i)|S_i = s) \le C < \infty$ for some constant $C > 0$, $\sup_{x\in Supp(X)} ||H_{h_n}(x)||_2 \le \zeta(h_n)$, $\zeta^2(h_n)h_n \log(n) = o(n)$, and $h_n^2 \log^2(n) = o(n)$, where $H_{h_n,h}(X_i)$ denotes the hth coordinate of $H_{h_n}(X_i)$.

Theorem 5.4. *Denote $\hat{q}^{NP}(\tau)$ and $\hat{q}^{NP,w}(\tau)$ as the τth QTE estimator and its multiplier boot-*

strap counterpart defined in Sections 3 and 4, respectively, with $\overline{m}_a(\tau, S_i, X_i) = m_a(\tau, S_i, X_i)$ *and* $\widehat{m}_a(\tau, S_i, X_i)$ *defined in* (5.13). *Further suppose Assumptions 1, 2, 4, 7, and 11 hold. Then, Assumptions 3 and 5 hold, which further implies that Theorems 3.1 and 4.1 hold for* $\hat{q}^{NP}(\tau)$ *and* $\hat{q}^{NP,w}(\tau)$*, respectively. In addition, for any finite-dimensional quantile indexes* $(\tau_1, \cdots, \tau_K)$*, the covariance matrix of* $(\hat{q}^{NP}(\tau_1), \cdots, \hat{q}^{NP}(\tau_K))$ *achieves the minimum (in the matrix sense) as characterized in Theorem 3.1.*

5.3 Regularization method

This section considers estimation of $m_a(\tau, s, X)$ in a high-dimensional environment. Let $H_{p_n}(X_i)$ be the regressors with dimension p_n, which may exceed the sample size. When the number of raw controls is comparable to or exceeds the sample size, we can just let $H_{p_n}(X_i) = X_i$. On the other hand, $H_{p_n}(X_i)$ may be composed of a large dictionary of sieve bases derived from a fixed dimensional vector X_i through suitable transformations such as powers and interactions. Thus, high dimensionality in $H_{p_n}(X_i)$ can arise from a desire flexibly approximate nuisance functions. In our approach we follow Belloni et al. (2017) and implement a logistic regression with ℓ_1-penalization. In their notation we view $m_a(\tau, s, x)$ as a function of $q_a(\tau)$, i.e,. $m_a(\tau, s, x) = \tau - \mathcal{M}_a(q_a(\tau), s, x)$, where $\mathcal{M}_a(q, s, x) = \mathbb{P}(Y_i(a) \leq q | S_i = s, X_i = x)$. We estimate $\mathcal{M}_a(q_a(\tau), s, x)$ as $\lambda(H_{p_n}(X_i)^\top \hat{\theta}^{HD}_{a,s}(\hat{q}_a(\tau)))$, where $\hat{q}_a(\tau)$ is defined in Assumption 7,

$$\begin{aligned}\hat{\theta}^{HD}_{a,s}(q) = \arg\min_{\theta_a} \frac{-1}{n_a(s)} \sum_{i \in I_a(s)} \Big[& 1\{Y_i \leq q\} \log(\lambda(H_{p_n}(X_i)^\top \theta_a)) \\ & + 1\{Y_i > q\} \log(1 - \lambda(H_{p_n}(X_i)^\top \theta_a)) \Big] + \frac{\varrho_{n,a}(s)}{n_a(s)} ||\hat{\Omega}\theta_a||_1,\end{aligned}$$

$\varrho_{n,a}(s)$ is a tuning parameter, and $\hat{\Omega} = \text{diag}(\hat{\omega}_1, \cdots, \hat{\omega}_{p_n})$ is a diagonal matrix of data-dependent penalty loadings. We specify $\varrho_{n,a}(s)$ and $\hat{\Omega}$ in Section 6. Post-Lasso estimation is also considered. Let $\hat{\mathbb{S}}_{a,s}(q)$ be the support of $\hat{\theta}^{HD}_{a,s}(q) = \{h \in [p_n] : \hat{\theta}^{HD}_{a,s,h}(q) \neq 0\}$, where $\hat{\theta}^{HD}_{a,s,h}(q)$ is the hth coordinate of $\hat{\theta}^{HD}_{a,s}(q)$. We can complement $\hat{\mathbb{S}}_{a,s}(q)$ with additional variables in $\hat{\mathbb{S}}^+_{a,s}(q)$ that researchers want to control for and define the enlarged set of variables as $\tilde{\mathbb{S}}_{a,s}(q) = \hat{\mathbb{S}}_{a,s}(q) \cup \hat{\mathbb{S}}^+_{a,s}(q)$. We compute the post-Lasso estimator $\hat{\theta}^{post}_{a,s}(q)$ as

$$\begin{aligned}\hat{\theta}^{post}_{a,s}(q) = \arg\min_{\theta_a \in \tilde{\mathbb{S}}_{a,s}(q)} \frac{-1}{n_a(s)} \sum_{i \in I_a(s)} \Big[& 1\{Y_i \leq q\} \log(\lambda(H_{p_n}(X_i)^\top \theta_a)) \\ & + 1\{Y_i > q\} \log(1 - \lambda(H_{p_n}(X_i)^\top \theta_a)) \Big]. \end{aligned} \tag{5.15}$$

Finally, we compute the auxiliary model as

$$\widehat{m}_a(\tau, s, X_i) = \lambda(H_{p_n}^\top(X_i)\hat{\theta}_{a,s}^{HD}(\hat{q}_a(\tau))) \quad \text{or} \quad \widehat{m}_a(\tau, s, X_i) = \lambda(H_{p_n}^\top(X_i)\hat{\theta}_{a,s}^{post}(\hat{q}_a(\tau))). \tag{5.16}$$

Assumption 12. *(i) Let* $\mathcal{Q}_a^\varepsilon = \{q : \inf_{\tau\in\Upsilon}|q - q_a(\tau)| \leq \varepsilon\}$. *Suppose* $\mathbb{P}(Y_i(a) \leq q|X, S_i = s) = \lambda(H_{p_n}(X_i)^\top\theta_{a,s}^{HD}(q)) + r_a(q, s, X)$ *such that* $\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S}}||\theta_{a,s}^{HD}(q)||_0 \leq h_n$.

(ii) Suppose $\sup_{i\in[n]}||H_{p_n}(X_i)||_\infty \leq \zeta_n$ *and* $\sup_{h\in[p_n]}\mathbb{E}(H_{p_n,h}^q(X_i)|S_i = s) < \infty$ *for* $q > 2$.

(iii) Suppose

$$\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S}} \frac{1}{n_a(s)}\sum_{i\in I_a(s)} r_a^2(q, s, X_i) = O_p(h_n\log(p_n)/n),$$

$$\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S}} \mathbb{E}(r_a^2(q, s, X_i)|S_i = s) = O(h_n\log(p_n)/n),$$

and

$$\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S},x\in\mathcal{X}} |r_a(q, s, X)| = O(\sqrt{\xi_n^2 h_n^2\log(p_n)/n}).$$

(iv) $\frac{\log(p_n)\xi_n^2h_n^2}{n} \to 0$, $\frac{\log^2(p_n)\log^2(n)h_n^2}{n} \to 0$, $\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S}}|\hat{\mathbb{S}}_{a,s}^+(q)| = O_p(h_n)$, *where* $|\hat{\mathbb{S}}_{a,s}^+(q)|$ *denotes the number of elements in* $\hat{\mathbb{S}}_{a,s}^+(q)$.

(v) There exists a constant $c \in (0, 0.5)$ *such that*

$$\begin{aligned} c &\leq \inf_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in Supp(X)} \mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i = x) \\ &\leq \sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in Supp(X)} \mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i = x) \leq 1 - c. \end{aligned}$$

(vi) Let ℓ_n *be a sequence that diverges to infinity. Then, there exist two constants* κ_1 *and* κ_2 *such that with probability approaching one,*

$$\begin{aligned} 0 < \kappa_1 &\leq \inf_{a=0,1,s\in\mathcal{S},|v|_0\leq h_n\ell_n} \frac{v^T\left(\frac{1}{n_a(s)}\sum_{i\in I_a(s)} H_{p_n}(X_i)H_{p_n}(X_i)^\top\right)v}{||v||_2^2} \\ &\leq \sup_{a=0,1,s\in\mathcal{S},|v|_0\leq h_n\ell_n} \frac{v^T\left(\frac{1}{n_a(s)}\sum_{i\in I_a(s)} H_{p_n}(X_i)H_{p_n}(X_i)^\top\right)v}{||v||_2^2} \leq \kappa_2 < \infty, \end{aligned}$$

and

$$0 < \kappa_1 \leq \inf_{a=0,1,s\in\mathcal{S},|v|_0\leq h_n\ell_n} \frac{v^T\mathbb{E}(H_{p_n}(X_i)H_{p_n}(X_i)^\top|S_i = s)v}{||v||_2^2}$$

$$\leq \sup_{a=0,1,s\in\mathcal{S},|v|_0\leq h_n\ell_n} \frac{v^T\mathbb{E}(H_{p_n}(X_i)H_{p_n}(X_i)^\top|S_i=s)v}{||v||_2^2} \leq \kappa_2 < \infty,$$

where $|v|_0$ denotes the number of nonzero components in v.

(vii) $\varrho_{n,a}(s) = c\sqrt{n_a(s)}\Phi^{-1}(1-0.1/(\log(n_a(s))4p_n))$ *where $\Phi(\cdot)$ is the standard normal CDF and $c>0$ is a constant.*

Assumption 12 is standard in the literature and we refer interested readers to Belloni et al. (2017) for more discussion. Assumption 12(i) implies the logistic model is approximately correctly specified. As the approximation is assumed to be sparse, the condition is not innocuous in the high-dimensional setting. As our method is valid even when the auxiliary model is misspecified, we conjecture that Assumption 12(i) can be relaxed, which links to the recent literature on the study of regularized estimation in the high-dimensional setting under misspecification: see, for example, Bradic, Wager, and Zhu (2019) and Tan (2020) and the references therein. An interesting topic for future work is to study misspecification-robust high-dimensional estimators of the conditional probability model and their use to adjust the QTE estimator under CAR based on (3.3) and (3.4).

Theorem 5.5. *Denote $\hat{q}^{HD}(\tau)$ and $\hat{q}^{HD,w}(\tau)$ as the τth QTE estimator and its multiplier bootstrap counterpart defined in Sections 3 and 4, respectively, with $\overline{m}_a(\tau,S_i,X_i) = m_a(\tau,S_i,X_i)$ and $\widehat{m}_a(\tau,S_i,X_i)$ defined in* (5.16)*. Further suppose Assumptions 1, 2, 4, 7, and 12 hold. Then, Assumptions 3 and 5 hold, which further imply Theorems 3.1 and 4.1 hold for $\hat{q}^{HD}(\tau)$ and $\hat{q}^{HD,w}(\tau)$, respectively. In addition, for any finite-dimensional quantile indexes $(\tau_1,\cdots,\tau_K)$, the covariance matrix of $(\hat{q}^{HD}(\tau_1),\cdots,\hat{q}^{HD}(\tau_K))$ achieves the minimum (in the matrix sense) as characterized in Theorem 3.1.*

6 Practical Guidance and Computation

6.1 Procedures for estimation and bootstrap inference

We can compute $(\hat{q}_1^{adj},\hat{q}_0^{adj})$ by solving the subgradient conditions of (3.3) and (3.4), respectively. Specifically, we have $(\hat{q}_1^{adj},\hat{q}_0^{adj}) = (Y_{i_1},Y_{i_0})$ such that $A_{i_1}=1$, $A_{i_0}=0$,

$$\tau\left(\sum_{i\in[n]}\frac{A_i}{\hat{\pi}(S_i)}\right) - \sum_{i\in[n]}\left(\frac{(A_i-\hat{\pi}(S_i))}{\hat{\pi}(S_i)}\widehat{m}_1(\tau,S_i,X_i)\right)$$
$$\geq \sum_{i\in[n]}\frac{A_i}{\hat{\pi}(S_i)}1\{Y_i<Y_{i_1}\}$$

$$\geq \tau\left(\sum_{i\in[n]}\frac{A_i}{\hat{\pi}(S_i)}\right) - \frac{1}{\hat{\pi}(S_{i_1})} - \sum_{i\in[n]}\left(\frac{(A_i-\hat{\pi}(S_i))}{\hat{\pi}(S_i)}\widehat{m}_1(\tau,S_i,X_i)\right), \tag{6.1}$$

and

$$\begin{aligned}
&\tau\left(\sum_{i\in[n]}\frac{1-A_i}{1-\hat{\pi}(S_i)}\right) + \sum_{i\in[n]}\left(\frac{(A_i-\hat{\pi}(S_i))}{1-\hat{\pi}(S_i)}\widehat{m}_0(\tau,S_i,X_i)\right)\\
\geq &\sum_{i\in[n]}\frac{1-A_i}{1-\hat{\pi}(S_i)}1\{Y_i < Y_{i_0}\}\\
\geq &\tau\left(\sum_{i\in[n]}\frac{1-A_i}{1-\hat{\pi}(S_i)}\right) - \frac{1}{1-\hat{\pi}(S_{i_0})} + \sum_{i\in[n]}\left(\frac{(A_i-\hat{\pi}(S_i))}{1-\hat{\pi}(S_i)}\widehat{m}_0(\tau,S_i,X_i)\right).
\end{aligned} \tag{6.2}$$

We note (i_1, i_0) are uniquely defined as long as all the inequalities in (6.1) and (6.2) are strict, which is usually the case. If we have

$$\tau\left(\sum_{i\in[n]}\frac{A_i}{\hat{\pi}(S_i)}\right) - \sum_{i\in[n]}\left(\frac{(A_i-\hat{\pi}(S_i))}{\hat{\pi}(S_i)}\widehat{m}_1(\tau,S_i,X_i)\right) = \sum_{i\in[n]}\frac{A_i}{\hat{\pi}(S_i)}1\{Y_i \leq Y_{i_1}\},$$

then both i_1 and i_1' satisfy (6.1), where i_1' is the index such that $A_{i_1'} = 1$ and $Y_{i_1'}$ is the smallest observation in the treatment group that is larger than Y_{i_1}. In this case, we let $\hat{q}_1^{adj} = Y_{i_1}$. Similarly, by solving the subgradient conditions of (4.1) and (4.2), we have $(\hat{q}_1^w, \hat{q}_0^w) = (Y_{i_1^w}, Y_{i_0^w})$ such that $A_{i_1^w} = 1$, $A_{i_0^w} = 0$,

$$\begin{aligned}
&\tau\left(\sum_{i\in[n]}\frac{\xi_i A_i}{\hat{\pi}^w(S_i)}\right) - \sum_{i\in[n]}\left(\frac{\xi_i(A_i-\hat{\pi}^w(S_i))}{\hat{\pi}^w(S_i)}\widehat{m}_1(\tau,S_i,X_i)\right)\\
\geq &\sum_{i\in[n]}\frac{\xi_i A_i}{\hat{\pi}^w(S_i)}1\{Y_i < Y_{i_1^w}\}\\
\geq &\tau\left(\sum_{i\in[n]}\frac{\xi_i A_i}{\hat{\pi}^w(S_i)}\right) - \frac{\xi_{i_1^w}}{\hat{\pi}^w(S_{i_1^w})} - \sum_{i\in[n]}\left(\frac{\xi_i(A_i-\hat{\pi}^w(S_i))}{\hat{\pi}^w(S_i)}\widehat{m}_1(\tau,S_i,X_i)\right)
\end{aligned} \tag{6.3}$$

and

$$\begin{aligned}
&\tau\left(\sum_{i\in[n]}\frac{\xi_i(1-A_i)}{1-\hat{\pi}^w(S_i)}\right) + \sum_{i\in[n]}\left(\frac{\xi_i(A_i-\hat{\pi}(S_i))}{1-\hat{\pi}^w(S_i)}\widehat{m}_0(\tau,S_i,X_i)\right)\\
\geq &\sum_{i\in[n]}\frac{\xi_i(1-A_i)}{1-\hat{\pi}^w(S_i)}1\{Y_i < Y_{i_0^w}\}
\end{aligned}$$

$$\geq \tau\left(\sum_{i\in[n]}\frac{\xi_i(1-A_i)}{1-\hat{\pi}^w(S_i)}\right)-\frac{\xi_{i_0^w}}{1-\hat{\pi}^w(S_{i_0})}+\sum_{i\in[n]}\left(\frac{\xi_i(A_i-\hat{\pi}^w(S_i))}{1-\hat{\pi}^w(S_i)}\widehat{m}_0(\tau,S_i,X_i)\right). \tag{6.4}$$

The inequalities in (4.1) and (4.2) are strict with probability one if ξ_i is continuously distributed. In this case, $(\hat{q}_1^w, \hat{q}_0^w)$ is uniquely defined with probability one.

We summarize the steps in the bootstrap procedure as follows.

1. Let $\mathcal{G}$ be a set of quantile indexes. For $\tau \in \mathcal{G}$, compute $\hat{q}_1(\tau)$ and $\hat{q}_0(\tau)$ following (6.1) and (6.2) with $\widehat{m}_1(\tau, S_i, X_i)$ and $\widehat{m}_0(\tau, S_i, X_i)$ replaced by zero.
2. Compute $\widehat{m}_a(\tau, S_i, X_i)$ for $a = 0, 1$ and $\tau \in \mathcal{G}$ using $\hat{q}_1(\tau)$ and $\hat{q}_0(\tau)$.
3. Compute the original estimator $\hat{q}^{adj}(\tau) = \hat{q}_1^{adj}(\tau) - \hat{q}_0^{adj}(\tau)$, following (6.1) and (6.2) for $\tau \in \mathcal{G}$.
4. Let B be the number of bootstrap replications. For $b \in [B]$, generate $\{\xi_i\}_{i\in[n]}$. Compute $\hat{q}^{w,b}(\tau) = \hat{q}_1^{w,b}(\tau) - \hat{q}_0^{w,b}(\tau)$ for $\tau \in \mathcal{G}$ following (6.3) and (6.4). Obtain $\{\hat{q}^{w,b}(\tau)\}_{\tau\in\mathcal{G}}$.
5. Repeat the above step for $b \in [B]$ and obtain B bootstrap estimators of the QTE, denoted as $\{\hat{q}^{w,b}(\tau)\}_{b\in[B],\tau\in\mathcal{G}}$.

6.2 Bootstrap confidence intervals

Given the bootstrap estimates, we discuss how to conduct bootstrap inference for the null hypotheses with single, multiple, and a continuum of quantile indexes.

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

$$CI(\alpha) = (\hat{q}^{adj}(\tau) + \mathcal{C}(\alpha/2)\hat{\sigma}, \hat{q}^{adj}(\tau) + \mathcal{C}(1 - \alpha/2)\hat{\sigma}),$$

and $1\{\left|\frac{\hat{q}^{adj}(\tau)-\underline{q}}{\hat{\sigma}}\right| \geq \mathcal{C}(1 - \alpha/2)\}$, respectively.[5]

Case (2). We test the null hypothesis that $\mathcal{H}_0 : q(\tau_1) - q(\tau_2) = \underline{q}$ vs. $q(\tau_1) - q(\tau_2) \neq \underline{q}$. In this case, we have $\mathcal{G} = \{\tau_1, \tau_2\}$ in the procedure described in Section 6.1. Further, let $\widehat{\mathcal{C}}(\nu)$ be the νth empirical quantile of the sequence $\{\hat{q}^{w,b}(\tau_1) - \hat{q}^{w,b}(\tau_2)\}_{b\in[B]}$, and let $\alpha \in (0, 1)$ be the significance

[5] It is asymptotically valid to use standard and percentile bootstrap confidence intervals. In simulations we found that the confidence interval proposed in the paper has better finite-sample performance.

level. We suggest using the bootstrap standard error to construct the valid confidence interval and Wald test as

$$CI(\alpha) = (\hat{q}^{adj}(\tau_1) - \hat{q}^{adj}(\tau_2) + \mathcal{C}(\alpha/2)\hat{\sigma}, \hat{q}^{adj}(\tau_1) - \hat{q}^{adj}(\tau_2) + \mathcal{C}(1-\alpha/2)\hat{\sigma}),$$

and $1\{|\frac{\hat{q}^{adj}(\tau_1)-\hat{q}^{adj}(\tau_2)-\underline{q}}{\hat{\sigma}}| \geq \mathcal{C}(1-\alpha/2)\}$, respectively, where $\hat{\sigma} = \frac{\widehat{\mathcal{C}}(0.975)-\widehat{\mathcal{C}}(0.025)}{\mathcal{C}(0.975)-\mathcal{C}(0.025)}$.

Case (3). We test the null hypothesis that

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

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

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

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

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

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

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

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

6.3 Computation of auxiliary regressions

Parametric regressions. For the linear probability model, we compute the LP estimator via (5.7). For the logistic model, we consider both LG and ML estimators. First, we compute the LG estimator $\hat{\theta}_{a,s}^{LG}(\tau)$ as in (5.9), which minimizes the sample variance of our adjusted QTE estimator $\hat{q}^{adj}(\tau)$.[6] Second, we propose to compute the ML estimator $\hat{\theta}_{a,s}^{ML}(\tau)$ as in (5.11), which is the quasi

[6]When the null hypothesis in Case (2) is tested, the optimal pseudo true value is defined to minimize the asymptotic variance of $\hat{q}^{par}(\tau_1) - \hat{q}^{par}(\tau_2)$, which is not necessarily equal to the pseudo true values that minimize the asymptotic variances of $\hat{q}^{par}(\tau_1)$ and $\hat{q}^{par}(\tau_2)$, i.e., $(\theta_{a,s}^{LG}(\tau_1), \theta_{a,s}^{LG}(\tau_2))$. We propose a way to compute the new pseudo true value

maximum likelihood estimator of a flexible distribution regression.

Sieve logistic regressions. We provide more details on the sieve basis. Recall $H_{h_n}(x) \equiv (b_{1n}(x), \cdots, b_{h_n n}(x))^\top$, where $\{b_{hn}(\cdot)\}_{h\in[h_n]}$ are h_n basis functions of a linear sieve space, denoted as $\mathbb{B}$. Given that all d_x elements of X are continuously distributed, the sieve space $\mathbb{B}$ can be constructed as follows.

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

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

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

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

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

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

We refer interested readers to Hirano et al. (2003) and Chen (2007) for more details about the implementation of sieve estimation. Given the sieve basis, we can compute the $\widehat{m}_a(\tau, s, X_i)$ following (5.13) and (5.14).

Logistic regressions with an ℓ_1 penalization. We follow the estimation procedure and the choice of tuning parameter proposed by Belloni et al. (2017). We provide details below for completeness. Recall $\varrho_{n,a}(s) = c\sqrt{n_a(s)}\Phi^{-1}(1 - 1/(p_n \log(n_a(s))))$. We set $c = 1.1$ following Belloni et al. (2017). We then implement the following algorithm to estimate $\hat{\theta}^{HD}_{a,s}(\hat{q}_a(\tau))$ for $\tau \in \Upsilon$:

(i) Let $\hat{\sigma}_h^{(0)} = \frac{1}{n_a(s)} \sum_{i\in I_a(s)} (1\{Y_i \le \hat{q}_a(\tau)\} - \bar{Y}_{a,s}(\tau))^2 H^2_{p_n,h}$ for $h \in [p_n]$, where $\bar{Y}_{a,s}(\tau) = \frac{1}{n_a(s)} \sum_{i\in I_a(s)} 1\{Y_i \le \hat{q}_a(\tau)\}$. Estimate

$$\hat{\theta}^{HD,0}_{a,s}(\hat{q}_a(\tau)) = \arg\min_{\theta_a} \frac{-1}{n_a(s)} \sum_{i\in I_a(s)} \Big[1\{Y_i \le \hat{q}_a(\tau)\} \log(\lambda(H_{p_n}(X_i)^\top \theta_a))$$

in Section C.

$$+ 1\{Y_i > \hat{q}_a(\tau)\} \log(1 - \lambda(H_{p_n}(X_i)^\top \theta_a))\Big] + \frac{\varrho_{n,a}(s)}{n_a(s)} \sum_{h \in [p_n]} \hat{\sigma}_h^{(0)} |\theta_{a,h}|.$$

(ii) For $k = 1, \cdots, K$, obtain $\hat{\sigma}_h^{(k)} = \sqrt{\frac{1}{n} \sum_{i \in [n]} (H_{p_n,h} \hat{\varepsilon}_i^{(k)})^2}$, where $\hat{\varepsilon}_i^{(k)} = 1\{Y_i \leq \hat{q}_a(\tau)\} - \lambda(H_{p_n}^T \hat{\theta}_{a,s}^{HD,k-1}(\hat{q}_a(\tau)))$. Estimate

$$\hat{\theta}_{a,s}^{HD,k}(\hat{q}_a(\tau)) = \arg\min_{\theta_a} \frac{-1}{n_a(s)} \sum_{i \in I_a(s)} \Big[1\{Y_i \leq \hat{q}_a(\tau)\} \log(\lambda(H_{p_n}(X_i)^\top \theta_a))$$
$$+ 1\{Y_i > \hat{q}_a(\tau)\} \log(1 - \lambda(H_{p_n}(X_i)^\top \theta_a))\Big] + \frac{\varrho_{n,a}(s)}{n_a(s)} \sum_{h \in [p_n]} \hat{\sigma}_h^{(k)} |\theta_{a,h}|.$$

(iii) Let $\hat{\theta}_{a,s}^{HD}(\hat{q}_a(\tau)) = \hat{\theta}_{a,s}^{HD,K}(\hat{q}_a(\tau))$.

(iv) Repeat the above procedure for $\tau \in \mathcal{G}$.

7 Simulation

7.1 Data generating processes

Three data generating processes (DGPs) are used to assess the finite-sample performance of the estimation and inference methods introduced in the paper. We consider the outcomes equation

$$Y_i = \alpha(X_i) + \gamma Z_i + \mu(X_i) A_i + \eta_i, \tag{7.1}$$

where $\gamma = 4$ for all cases while $\alpha(X_i)$, $\mu(X_i)$, and η_i are separately specified as follows.

(i) Let Z be standardized Beta(2, 2) distributed, $S_i = \sum_{j=1}^4 1\{Z_i \leq g_j\}$, and $(g_1, \cdots, g_4) = (-0.25\sqrt{20}, 0, 0.25\sqrt{20}, 0.5\sqrt{20})$. X_i contains two covariates $(X_{1i}, X_{2i})^\top$, where X_{1i} follows a uniform distribution on $[-2, 2]$, X_{2i} follows a standard normal distribution, and X_{1i} and X_{2i} are independent. Further define $\alpha(X_i) = 1 + X_{2i}$, $\mu(X_i) = 1 + X_i^\top \beta$, $\beta = (3, 3)^\top$, and $\eta_i = (0.25 + X_{1i}^2) A_i \varepsilon_{1i} + (1 - A_i)\varepsilon_{2i}$, where $(\varepsilon_{1i}, \varepsilon_{2i})$ are jointly standard normal.

(ii) Let Z be uniformly distributed on $[-2, 2]$, $S_i = \sum_{j=1}^4 1\{Z_i \leq g_j\}$, and $(g_1, \cdots, g_4) = (-1, 0, 1, 2)$. Let $X_i = (X_{1i}, X_{2i})^\top$ be the same as defined in DGP (i). Further define $\alpha(X_i) = 1 + X_{1i} + X_{2i}$, $\mu(X_i) = 1 + X_{1i} + X_{2i} + \frac{1}{4}(X_i^\top \beta)^2$ with $\beta = (2, 2)^\top$, and $\eta_i = 2(1 + Z_i^2) A_i \varepsilon_{1i} + (1 + Z_i^2)(1 - A_i)\varepsilon_{2i}$, where $(\varepsilon_{1i}, \varepsilon_{2i})$ are mutually independently $T(5)/\sqrt{5}$ distributed.

(iii) Let Z be standardized Beta(2, 2) distributed, $S_i = \sum_{j=1}^{4} 1\{Z_i \leq g_j\}$, and $(g_1, \cdots, g_4) = (-0.5\sqrt{5}, 0, 0.5\sqrt{5}, \sqrt{5})$. Further suppose that X_i contains twenty covariates $(X_{1i}, \cdots, X_{20,i})^\top$, where $X = \Phi(Z)$ with $Z \sim N(0_{20\times 1}, \Omega)$ and the variance-covariance matrix Ω is the Toeplitz matrix

$$\Omega = \begin{pmatrix} 0.5 & 0.5^2 & 0.5^3 & \cdots & 0.5^{19} \\ 0.5^2 & 0.5 & 0.5^2 & \cdots & 0.5^{18} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.5^{19} & 0.5^{18} & 0.5^{17} & \cdots & 0.5 \end{pmatrix}$$

Further define $\alpha(X_i) = 1$, $\mu(X_i) = 1 + \sum_{k=1}^{20} X_{ki}\beta_k$ with $\beta_k = 4/k^2$, and $\eta_i = 2A_i\varepsilon_{1i} + (1 - A_i)\varepsilon_{2i}$, where $(\varepsilon_{1i}, \varepsilon_{2i})$ are jointly standard normal.

For each data generating process, we consider the following four randomization schemes as in Zhang and Zheng (2020) with $\pi(s) = 0.5$ for $s \in \mathcal{S}$:

(i) SRS: Treatment assignment is generated as in Example 1.

(ii) WEI: Treatment assignment is generated as in Example 2 with $\phi(x) = (1 - x)/2$.

(iii) BCD: Treatment assignment is generated as in Example 3 with $\lambda = 0.75$.

(iv) SBR: Treatment assignment is generated as in Example 4.

We assess the empirical size and power of the tests for $n = 200$ and $n = 400$. All simulations are replicated 10,000 times with bootstrap sample size being 1,000. We compute the true QTEs or QTE differences by simulations with 10,000 sample size and 1,000 replications. To compute power, we perturb the true values by 1.5 for DGPs (i) and (ii) and 2 for DGP (iii). As discussed in Section 6, we examine three null hypotheses:

(i) Pointwise test

$$H_0 : q(\tau) = \text{truth} \quad \text{v.s.} \quad H_1 : q(\tau) = \text{truth} + \text{perturbation}, \quad \tau = 0.25, 0.5, 0.75;$$

(ii) Test for the difference

$$H_0 : q(0.75) - q(0.25) = \text{truth} \quad \text{v.s.} \quad H_1 : q(0.75) - q(0.25) = \text{truth} + \text{perturbation};$$

(iii) Uniform test

$$H_0 : q(\tau) = \text{truth}(\tau) \quad v.s. \quad H_1 : q(\tau) = \text{truth}(\tau) + \text{perturbation}, \quad \tau \in [0.25, 0.75].$$

For the pointwise test, we report the results for the median ($\tau = 0.5$) in the main text and give the cases $\tau = 0.25$ and $\tau = 0.75$ to the Online Supplement.

7.2 Estimation methods

For DGPs (i) and (ii), we consider the following estimation methods of the auxiliary regression.

(i) NA: the estimator without any adjustments, i.e., setting $\hat{m}_a(\cdot) = \overline{m}_a(\cdot) = 0$.

(ii) LP: the linear probability model with regressors X_i and the pseudo true value is estimated by $\hat{\theta}^{LP}_{a,s}(\tau)$ defined in (5.7).

(iii) LG: the logistic model with regressors $(1, X_{1i}, X_{2i})$ and the pseudo true value is estimated by $\hat{\theta}^{LG}_{a,s}(\tau)$ defined in (5.9).

(iv) ML: the logistic model with regressor $H_i = (1, X_{1i}, X_{2i}, X_{1i}X_{2i})^\top$ and the pseudo true value is estimated by $\hat{\theta}^{ML}_{a,s}(\tau)$ defined in (5.10).

(v) NP: the logistic model with regressor $H_{h_n}(X_i) = (1, X_{1i}, X_{2i}, X_{1i}X_{2i}, X_{1i}1\{X_{1i} > t_1\}X_{2i}1\{X_{2i} > t_2\})^\top$ where t_1 and t_2 are the sample medians of $\{X_{1i}\}_{i\in[n]}$ and $\{X_{2i}\}_{i\in[n]}$, respectively. The pseudo true value is estimated by $\hat{\theta}^{NP}_{a,s}(\tau)$ defined in (5.14).

For DGP (iii), we consider the post-Lasso estimator $\hat{\theta}^{post}_{a,s}(\hat{q}_a(\tau))$ as defined in (5.15) with $H_{p_n}(X_i) = (1, X_i^\top)^\top$ and $\hat{\mathbb{S}}^+_{a,s}(q) = \{2\}$. The choice of tuning parameter and the estimation procedure are detailed in Section 6.3.

7.3 Simulation results

Table 2 presents the empirical size and power for the pointwise test with $\tau = 0.5$ under DGPs (i) and (ii). We make five observations. First, none of the auxiliary models is correctly specified but test sizes are all close to the nominal level 5%, confirming that estimation and inference are robust to misspecification. Second, the inclusion of auxiliary regressions improves efficiency of the QTE estimator as the powers for method "NA" are the smallest among all the methods for both DGPs and all randomization schemes. This finding is consistent with theory because methods "LP", "LG", "NP" are guaranteed to be weakly more efficient than "NA". Method "ML" fits a flexible distribution regression which can approximate the true DGP well. Therefore, its performance is close to the method "NP" and, thus, better than "NA", "LP", and "LG". Third, the powers of method "NP" are better than those for "LP", "LG", and "ML" because this method estimates the true specification and achieves the minimum asymptotic variance of $\hat{q}^{adj}(\tau)$ as shown in Theorem 5.1. Fourth, from a theory perspective the performance of methods "LP" and "LG" is not comparable as both are misspecified. In simulations the powers of method "LG" are worse for two reasons. First, computation of the objective function requires nonparametric estimation of the densities, which requires tuning parameters[7] and has a slow convergence rate. Second, since the objective function is not convex, the computation of the optimal values $\hat{\theta}^{LG}_{a,s}(\tau)$ is sometimes unstable.[8] Fifth, when the sample size is 200, the method "NP" slightly overrejects but size becomes closer to nominal when the sample size increases to 400.

Table 2: Pointwise Test ($\tau = 0.5$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.051	0.052	0.050	0.054	0.055	0.055	0.051	0.054
LP	0.050	0.047	0.055	0.053	0.050	0.052	0.053	0.054
LG	0.060	0.063	0.061	0.057	0.058	0.058	0.060	0.058
ML	0.060	0.058	0.060	0.055	0.051	0.055	0.050	0.052
NP	0.059	0.061	0.063	0.066	0.053	0.055	0.055	0.054
A.2: Power								
NA	0.403	0.402	0.405	0.402	0.681	0.675	0.679	0.681
LP	0.495	0.495	0.499	0.494	0.791	0.791	0.788	0.782
LG	0.466	0.478	0.462	0.466	0.760	0.757	0.769	0.756
ML	0.508	0.511	0.533	0.511	0.808	0.813	0.810	0.809
NP	0.526	0.537	0.535	0.530	0.813	0.815	0.813	0.809
				Panel B: DGP (ii)				
B.1: Size								
NA	0.049	0.049	0.044	0.045	0.051	0.047	0.047	0.046
LP	0.049	0.049	0.048	0.049	0.053	0.052	0.051	0.045
LG	0.055	0.056	0.055	0.058	0.054	0.054	0.054	0.057
ML	0.062	0.062	0.057	0.058	0.055	0.053	0.055	0.056
NP	0.070	0.063	0.059	0.066	0.059	0.057	0.055	0.057
B.2: Power								
NA	0.490	0.494	0.497	0.485	0.779	0.772	0.768	0.762
LP	0.579	0.567	0.569	0.573	0.851	0.850	0.852	0.852
LG	0.524	0.526	0.524	0.521	0.819	0.823	0.821	0.819
ML	0.612	0.601	0.609	0.623	0.868	0.875	0.877	0.871
NP	0.612	0.618	0.612	0.618	0.878	0.876	0.880	0.876

Tables 3 and 4 present sizes and powers of inference on $q(0.75) - q(0.25)$ and on $q(\tau)$ uniformly

[7] We use the Gaussian kernel and bandwidth $1.06\hat{\sigma}_{a,s}(n_a(s))^{1/5}$ where $\hat{\sigma}_{a,s}$ and $n_a(s)$ are the standard error and sample size of the observations $\{Y_i\}_{i \in I_a(s)}$, respectively, for $a = 0, 1$ and $s \in \mathcal{S}$.

[8] In simulations we added a ridge penalty $\log^{-1}(n(s))||\theta||_2^2$ to the objective function to stabilize the solution. Such a penalty term does not affect the proof of consistency of $\hat{\theta}^{LG}_{a,s}(\tau)$ because it will vanish as $n \to \infty$.

over $\tau \in [0.25, 0.75]$, respectively, for DGPs (i) and (ii) and four randomization schemes. All the observations made above apply to these results.

Table 3: Test for Differences ($\tau_1 = 0.25$, $\tau_2 = 0.75$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.039	0.039	0.041	0.043	0.043	0.045	0.042	0.045
LP	0.046	0.045	0.045	0.047	0.045	0.047	0.046	0.047
LG	0.047	0.052	0.042	0.046	0.045	0.050	0.046	0.046
ML	0.048	0.047	0.048	0.047	0.048	0.047	0.047	0.048
NP	0.050	0.048	0.049	0.050	0.050	0.049	0.046	0.053
A.2: Power								
NA	0.193	0.194	0.201	0.218	0.369	0.376	0.377	0.397
LP	0.211	0.201	0.207	0.207	0.400	0.395	0.401	0.398
LG	0.227	0.235	0.237	0.236	0.462	0.468	0.457	0.453
ML	0.245	0.242	0.242	0.241	0.464	0.454	0.455	0.455
NP	0.240	0.239	0.246	0.240	0.455	0.464	0.462	0.464
				Panel B: DGP (ii)				
B.1: Size								
NA	0.040	0.038	0.040	0.042	0.042	0.043	0.044	0.043
LP	0.042	0.044	0.041	0.048	0.045	0.047	0.043	0.041
LG	0.047	0.046	0.046	0.048	0.050	0.046	0.052	0.052
ML	0.051	0.052	0.049	0.051	0.051	0.049	0.048	0.049
NP	0.060	0.057	0.053	0.050	0.050	0.053	0.056	0.051
B.2: Power								
NA	0.207	0.208	0.212	0.237	0.403	0.401	0.397	0.423
LP	0.236	0.241	0.244	0.233	0.444	0.444	0.443	0.433
LG	0.236	0.241	0.244	0.233	0.444	0.444	0.443	0.433
ML	0.255	0.261	0.267	0.252	0.471	0.475	0.467	0.475
NP	0.266	0.265	0.267	0.267	0.465	0.480	0.467	0.477

Next, we assess the empirical size and power of all three testing scenarios under the high-dimensional setting. In Table 5, we compare the methods "NA" and our post-Lasso estimator. In general, all sizes for both methods approach the nominal level as the sample size increases. The post-Lasso method dominates "NA" in all tests with superior power performance. This result is consistent with theory given in Theorem 5.5.

Last, Table 6 reports the size and power of our regression-adjusted estimator for the median QTE when we replace $\hat{\pi}(s)$ by the true propensity score $1/2$. Comparing it with Table 2, we see that using the true, instead of the estimated, propensity score, the multiplier bootstrap inference

Table 4: Uniform Test ($\tau \in [0.25, 0.75]$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.044	0.047	0.043	0.048	0.047	0.047	0.046	0.046
LP	0.047	0.047	0.047	0.044	0.048	0.045	0.046	0.047
LG	0.049	0.048	0.051	0.051	0.053	0.056	0.053	0.053
ML	0.050	0.051	0.051	0.050	0.051	0.055	0.050	0.049
NP	0.059	0.054	0.058	0.052	0.052	0.053	0.053	0.054
A.2: Power								
NA	0.448	0.459	0.456	0.448	0.771	0.769	0.774	0.772
LP	0.594	0.580	0.599	0.596	0.895	0.897	0.904	0.899
LG	0.567	0.559	0.582	0.568	0.883	0.887	0.882	0.883
ML	0.614	0.621	0.616	0.613	0.918	0.916	0.923	0.918
NP	0.630	0.630	0.634	0.639	0.920	0.919	0.921	0.917
				Panel B: DGP (ii)				
B.1: Size								
NA	0.035	0.039	0.039	0.038	0.044	0.044	0.042	0.044
LP	0.040	0.046	0.041	0.040	0.042	0.044	0.045	0.042
LG	0.050	0.049	0.047	0.044	0.051	0.056	0.050	0.055
ML	0.055	0.053	0.055	0.052	0.052	0.053	0.049	0.054
NP	0.066	0.061	0.055	0.062	0.053	0.056	0.053	0.055
B.2: Power								
NA	0.560	0.572	0.570	0.582	0.876	0.878	0.874	0.880
LP	0.700	0.705	0.707	0.707	0.948	0.955	0.954	0.951
LG	0.647	0.654	0.654	0.650	0.936	0.939	0.936	0.935
ML	0.765	0.770	0.770	0.771	0.973	0.974	0.974	0.977
NP	0.771	0.770	0.769	0.778	0.971	0.972	0.975	0.976

Table 5: Empirical Size and Power for High-dimensional Covariates (DGP 3)

	$N = 200$				$N = 400$			
	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: Size				
A.1. NA								
$\tau = 0.25$	0.043	0.044	0.043	0.043	0.041	0.047	0.044	0.046
$\tau = 0.5$	0.045	0.046	0.049	0.046	0.051	0.043	0.046	0.045
$\tau = 0.75$	0.046	0.047	0.049	0.049	0.049	0.046	0.046	0.052
Diff(0.25-0.75)	0.042	0.041	0.038	0.042	0.043	0.045	0.043	0.041
Uniform	0.045	0.044	0.048	0.043	0.049	0.055	0.050	0.052
A.2. Post-Lasso								
$\tau = 0.25$	0.055	0.057	0.058	0.057	0.049	0.048	0.055	0.055
$\tau = 0.5$	0.061	0.056	0.060	0.058	0.057	0.047	0.050	0.051
$\tau = 0.75$	0.056	0.062	0.058	0.052	0.053	0.052	0.051	0.053
Diff(0.25-0.75)	0.055	0.051	0.052	0.053	0.051	0.047	0.045	0.048
Uniform	0.070	0.066	0.066	0.070	0.060	0.064	0.058	0.055
				Panel B: Power				
B.1. NA								
$\tau = 0.25$	0.399	0.403	0.409	0.410	0.689	0.690	0.698	0.696
$\tau = 0.5$	0.342	0.342	0.363	0.367	0.601	0.602	0.599	0.622
$\tau = 0.75$	0.336	0.338	0.342	0.328	0.596	0.593	0.604	0.589
Diff(0.25-0.75)	0.183	0.186	0.183	0.207	0.340	0.347	0.349	0.365
Uniform	0.529	0.533	0.521	0.536	0.846	0.846	0.845	0.843
B.2. Post-Lasso								
$\tau = 0.25$	0.430	0.428	0.428	0.425	0.718	0.709	0.722	0.713
$\tau = 0.5$	0.375	0.359	0.370	0.359	0.619	0.621	0.626	0.635
$\tau = 0.75$	0.355	0.366	0.364	0.363	0.606	0.617	0.618	0.604
Diff(0.25-0.75)	0.210	0.207	0.211	0.202	0.365	0.363	0.364	0.368
Uniform	0.580	0.585	0.586	0.582	0.870	0.870	0.878	0.870

becomes conservative for randomization schemes "WEI", "BCD", and "SBR". Specifically, the sizes are much smaller than the nominal rate (5%) while the powers are smaller than their counterparts in Table 2. This is consistent with the findings in Bugni et al. (2018) and Zhang and Zheng (2020) that the naive inference methods under CARs are conservative. Similar results for $\tau = 0.25$, $\tau = 0.75$, the difference of the 0.25 and 0.75 QTEs, and QTEs uniformly over $\tau \in [0.25, 0.75]$ are reported in Section B in the Online Supplement. All the remarks here still apply.

Table 6: Pointwise Test with Naïve Estimator ($\tau = 0.5$, $\hat{\pi}(s) = 0.5$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.049	0.025	0.014	0.017	0.047	0.028	0.015	0.016
LP	0.046	0.008	0.000	0.000	0.050	0.008	0.000	0.000
LG	0.053	0.017	0.008	0.006	0.051	0.019	0.008	0.006
ML	0.052	0.015	0.009	0.004	0.050	0.015	0.006	0.006
NP	0.053	0.020	0.007	0.006	0.056	0.018	0.006	0.006
A.2: Power								
NA	0.287	0.257	0.249	0.245	0.495	0.495	0.511	0.496
LP	0.180	0.103	0.057	0.036	0.307	0.253	0.194	0.150
LG	0.260	0.212	0.185	0.151	0.476	0.446	0.433	0.414
ML	0.267	0.242	0.213	0.183	0.511	0.504	0.509	0.485
NP	0.293	0.254	0.219	0.191	0.502	0.505	0.507	0.486
				Panel B: DGP (ii)				
B.1: Size								
NA	0.050	0.017	0.006	0.005	0.050	0.017	0.006	0.006
LP	0.050	0.005	0.000	0.000	0.048	0.006	0.000	0.000
LG	0.055	0.012	0.005	0.006	0.059	0.016	0.005	0.005
ML	0.050	0.021	0.013	0.016	0.046	0.025	0.022	0.020
NP	0.060	0.022	0.019	0.018	0.055	0.029	0.027	0.024
B.2: Power								
NA	0.290	0.245	0.220	0.221	0.491	0.496	0.494	0.501
LP	0.170	0.093	0.041	0.019	0.301	0.234	0.154	0.118
LG	0.260	0.181	0.137	0.125	0.476	0.446	0.413	0.399
ML	0.350	0.279	0.241	0.218	0.625	0.607	0.575	0.559
NP	0.365	0.300	0.263	0.244	0.645	0.621	0.599	0.577

7.4 Practical recommendations

When X is finite-dimensional, we suggest using the ML estimator to construct the fitted auxiliary model when the logistic model includes additional technical terms and the regression coefficients are allowed to depend on (τ, a, s). When X is high-dimensional, we suggest using the logistic post-Lasso estimation of the auxiliary model.

8 Empirical Application

Evaluating the impact of child health and nutrition on educational outcomes has been found to provide a key to understanding the links between health and economic development in less developed countries (Glewwe and Miguel, 2007; Dupas and Miguel, 2017). In particular, Chong et al. (2016) conducted a randomized experiment in Peru with a CAR to study the impact of an iron supplement encouragement program on schooling attainment. In their paper, the authors focused on the estimating the ATEs of that program. This section reports an application of our methods to the same dataset to examine the QTEs of the program on student academic achievement – a central outcome of interest in their study.

The sample consists of 215 students from a rural secondary school in Peru during the 2009 academic year. Within each of 5 secondary school grade levels, one-third of students were randomly assigned to each of three groups: two treatment arms and one control group. This is a block stratified randomization design with 5 strata, which satisfies Assumption 1 in Section 2.[9] Students in the first treatment arm were shown a video in which a popular soccer player encouraged iron supplements; students in the second treatment arm were shown a video in which a doctor encouraged iron supplements; students in the control group were shown a video without mentioning iron supplements at all. In this section, we focus on the impact of iron supplements on the academic achievement measured by a standardized average of a student's grades from five subjects (See Chong et al., 2016 for a detailed description). Throughout the analysis, students in the two treatment arms are grouped into the treatment group as in Chong et al. (2016).

One of the key findings in Chong et al. (2016) is that iron supplements have a significant positive impact on grades for anemic students, which is shown in Table 7 where we repeat their ATE estimates for easier comparison between ATEs and QTEs.[10] The columns "Full 1x" and "Anemic 1x" refer to the ATE estimates obtained from the regressions on the full sample and the anemic subsample respectively, controlling for only one additional baseline variable (male) besides strata indicators. The columns "Full 4x" and "Anemic 4x" refer to the ATE estimates obtained from similar regressions while controlling for four additional baseline variables (male, monthly income, electricity in home, and mother's years of schooling).

[9] It is trivial to see that statements (i), (ii), and (iii) in Assumption 1 are satisfied. Because $\sup_{s\in\mathcal{S}} |D_n(s)| \approx 0.03$ in both the full sample and the anemic subsample, it is plausible to claim that Assumption 1(iv) is also satisfied in our analysis.

[10] The ATE results on the anemic subsample are directly retrieved from Table 5 in Chong et al. (2016). The ATE results on the full sample are obtained by applying the same Stata code provided by them to the full sample.

Tables 8 and 9 present the QTE estimates and their standard errors (in parentheses) estimated by different methods at quantile indexes 0.25, 0.5, and 0.75. The description of these estimators is similar to that in Section 7.[11] Following the lead of Chong et al. (2016), throughout the analysis, we focus on two sets of additional baseline variables: male only (one auxiliary regressor) and male, monthly income, electricity in home, and mother's years of schooling (four auxiliary regressors). Table 8 reports the results with only one auxiliary regressor and Table 9 reports the results with four auxiliary regressors.

Table 7: ATEs of Iron Supplement Encouragement on Grades in Chong et al. (2016)

	Full 1x	Full 4x	Anemic 1x	Anemic 4x
ATE	0.091	0.081	0.455	0.434
	(0.140)	(0.140)	(0.218)	(0.215)

Notes: The table repeats the ATE estimates of the effect of the iron supplement encouragement program on grades reported in Chong et al. (2016). Standard errors are in parentheses.

The results in Tables 8-9 lead to two observations. First, consistent with the theoretical and simulation results, the standard errors for the regression-adjusted QTEs are mostly lower than those for the QTE estimate without adjustment. This observation holds generally, irrespective of the specification and estimation methods of the auxiliary regressions. For example in Table 8, in the anemic subsample, the standard errors for the "ML" QTE estimates are 14.1% and 3.5% less than those for the QTE estimate without adjustment at the median and the 75th percentile, respectively. For another example in Table 9, at the 25th percentile, the standard errors for the "Lasso" QTE estimates are 24.8% and 14.1% less than those for the QTE estimate without adjustment in the

[11] Specifically,

(i) NA: the estimator without any adjustments.

(ii) LP: linear probability model.

(iii) LG: logistic probability model with regressors $(1, X_i^\top)^\top$ and coefficient estimator $\hat{\theta}_{a,s}^{LG}(\tau)$. In line with the simulation, we add an ridge penalty $\log^{-1}(n(s))||\theta||_2^2$ to the objective function to stabilize the solution.

(iv) ML: logistic probability model with regressor H and coefficient estimator $\hat{\theta}_{a,s}^{ML}(\tau)$. When there is only one auxiliary regressor, $H_i = (1, X_{1i})^\top$, and when there are four auxiliary regressors, $H_i = (1, X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{3i}X_{4i})^\top$, where $X_{1i}, X_{2i}, X_{3i}, X_{4i}$, represent these variables for student i: male, monthly income, electricity in home, and mother's years of schooling, respectively.

(v) NP: logistic probability model with regressor H_{h_n} and coefficient estimator $\hat{\theta}_{a,s}^{NP}(\tau)$. NP is only applied to the case when there are four auxiliary regressors with $H_{h_n} = (1, X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{2i}1\{X_{2i} > t_1\}X_{4i}1\{X_{4i} > t_2\})^\top$ where t_1 and t_2 are the sample medians of $\{X_{2i}\}_{i\in[n]}$ and $\{X_{4i}\}_{i\in[n]}$, respectively.

(vi) Lasso: logistic probability model with regressor H_{p_n} and post-Lasso coefficient estimator $\hat{\theta}_{a,s}^{post}(\tau)$. Lasso is only applied to the case when there are four auxiliary regressors with $H_{p_n}(X_i) = (1, X_i^\top)^\top$ and $\hat{\mathbb{S}}_{a,s}^{+}(q) = \{2\}$. The post-Lasso estimator $\hat{\theta}_{a.s}^{post}(\tau)$ is defined in (5.15). The choice of tuning parameter and the estimation procedure are detailed in Section 6.3.

Table 8: QTEs of Iron Supplement Encouragement on Grades (one auxiliary regressor)

	NA	LP	LG	ML
Panel A. Full Sample				
25%	0.317	0.317	0.238	0.317
	(0.121)	(0.142)	(0.121)	(0.121)
50%	0.079	0.000	0.000	0.000
	(0.182)	(0.162)	(0.162)	(0.162)
75%	0.159	-0.079	0.000	-0.079
	(0.243)	(0.223)	(0.223)	(0.223)
Panel B. Anemic Subsample				
25%	0.381	0.381	0.381	0.381
	(0.170)	(0.146)	(0.158)	(0.146)
50%	0.761	0.761	0.666	0.761
	(0.243)	(0.243)	(0.243)	(0.243)
75%	0.666	0.666	0.666	0.666
	(0.340)	(0.316)	(0.340)	(0.328)

Notes: The table presents the QTE estimates of the effect of the iron supplement encouragement program across the distribution of grade when only one auxiliary regressor (male) is used in the regression adjustment models. Standard errors are in parentheses.

Table 9: QTEs of Iron Supplement Encouragement on Grades (four auxiliary regressors)

	NA	LP	LG	ML	NP	Lasso
Panel A. Full Sample						
25%	0.317	0.317	0.317	0.317	0.317	0.317
	(0.121)	(0.142)	(0.121)	(0.101)	(0.101)	(0.091)
50%	0.079	-0.079	0.000	-0.079	0.000	0.000
	(0.182)	(0.182)	(0.162)	(0.142)	(0.162)	(0.142)
75%	0.159	-0.159	0.079	-0.079	-0.079	-0.079
	(0.243)	(0.233)	(0.223)	(0.223)	(0.243)	(0.223)
Panel B. Anemic Subsample						
25%	0.381	0.476	0.381	0.381	0.381	0.381
	(0.170)	(0.381)	(0.146)	(0.146)	(0.146)	(0.146)
50%	0.761	0.666	0.666	0.761	0.666	0.666
	(0.243)	(0.267)	(0.243)	(0.219)	(0.194)	(0.219)
75%	0.666	0.571	0.571	0.666	0.666	0.857
	(0.340)	(0.303)	(0.340)	(0.316)	(0.291)	(0.291)

Notes: The table presents the QTE estimates of the effect of the iron supplement encouragement program on academic grades at quantiles 25%, 50%, and 75% when four auxiliary regressor (male, monthly income, electricity in home, and mother's years of schooling) are used in the regression adjustment models. Standard errors are in parentheses.

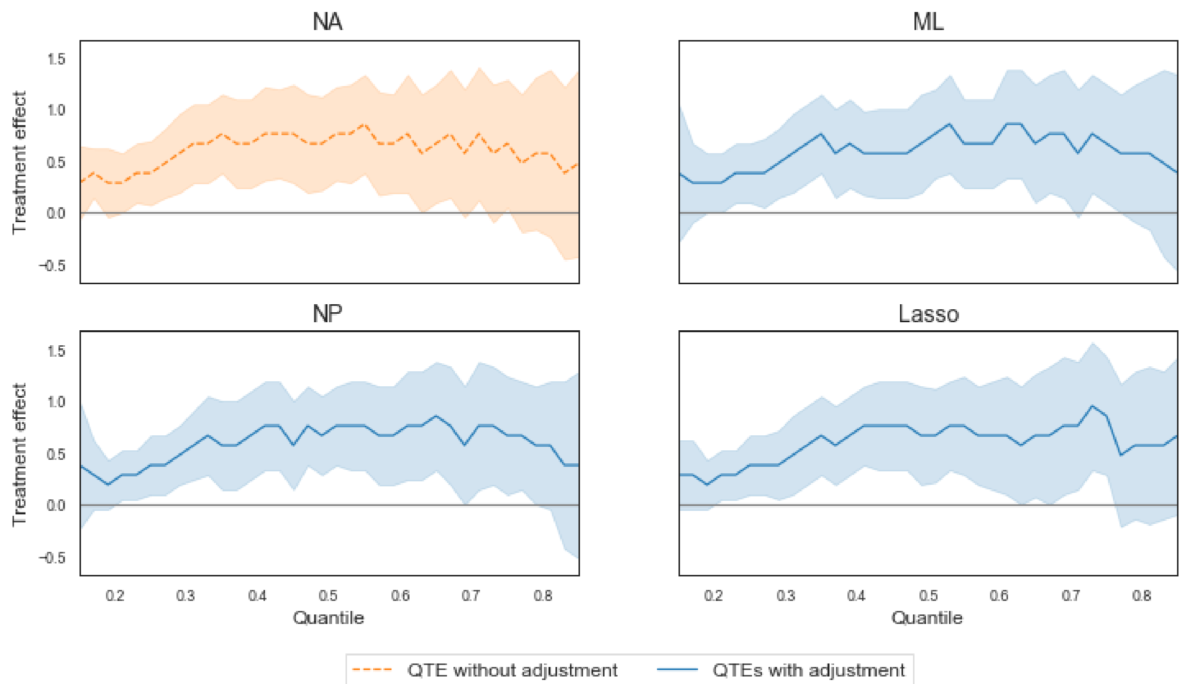

Figure 1: Quantile Treatment Effects on the Distribution of Grades (Anemic subsample)

Notes: The figure plots the QTE estimates of the effect of the iron supplement encouragement program on the distribution of grades in the anemic subsample when there are four auxiliary regressors: male, monthly income, electricity at home, mother education. The shadow areas represent the bounds of 95% confidence intervals.

full sample and the anemic subsample, respectively.

Second, there seems to be heterogeneity in the impact of the iron supplement encouragement. Specifically, in the full sample, only the treatment effects at the 25th percentile are significantly positive; in the anemic subsample, the treatment effects increase by about 50% from the 25th percentile to the median and seem to stay constant afterwards.

How does the variation in the impact of the iron supplement encouragement appear across the distribution of grades? The QTEs on the distribution of grades for the anemic subsample are plotted in Fig 1, where the shaded areas represent the 95% confidence region. The figure shows that the QTEs seem insignificantly different from zero below about the 20% percentile. At higher levels to around 80% percentile the treatment group grades exceed the control group grades, yielding significant positive QTEs. Beyond the 80% percentile, the QTEs again become insignificantly different from zero. These findings point to notable heterogeneity in the QTEs of the impact of the iron supplement encouragement program on the distribution of grades.

9 Conclusion

This paper proposes the use of auxiliary regression to incorporate additional covariates into estimation and inference relating to unconditional QTEs under CARs. The auxiliary regression model may be estimated parametrically, nonparametrically, or via regularization if there are high-dimensional covariates. Both estimation and bootstrap inference methods are robust to potential misspecification of the auxiliary model and do not suffer from the conservatism due to the CAR. It is shown that efficiency can be improved when including extra covariates and when the auxiliary regression is correctly specified the regression-adjusted estimator achieves the minimum asymptotic variance. In both the simulations and the empirical application, the proposed regression-adjusted QTE estimator performs well. These results and the robustness of the methods to auxiliary model misspecification reflect the aphorism widespread in scientific modeling that all models may be wrong, but some are useful.[12]

[12]The aphorism "all models are wrong but some are useful" is often attributed to the statistician George Box (1976). But the notion has many antecedents, including a particularly apposite remark made in 1947 by John von Neumann (2019) in an essay on the empirical origins of mathematical ideas to the effect that "truth ... is much too complicated to allow anything but approximations".

A Additional Simulation Results for Pointwise Tests

This section gives additional simulation results for pointwise tests at 25% and 75% quantiles. The results are summarized in Tables 10 and 11. The simulation settings are the same as the pointwise test simulations in Section 7 of the original paper.

Table 10: Pointwise Test ($\tau = 0.25$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.055	0.052	0.051	0.051	0.054	0.054	0.053	0.053
LP	0.055	0.054	0.058	0.054	0.055	0.053	0.052	0.052
LG	0.052	0.048	0.054	0.051	0.055	0.051	0.053	0.053
ML	0.060	0.059	0.060	0.059	0.052	0.053	0.055	0.056
NP	0.063	0.064	0.060	0.061	0.053	0.058	0.061	0.060
A.2: Power								
NA	0.341	0.329	0.346	0.352	0.591	0.593	0.589	0.611
LP	0.404	0.406	0.402	0.397	0.688	0.696	0.691	0.695
LG	0.386	0.392	0.399	0.390	0.692	0.684	0.688	0.697
ML	0.435	0.423	0.428	0.426	0.718	0.729	0.733	0.724
NP	0.443	0.437	0.434	0.443	0.730	0.723	0.737	0.728
				Panel B: DGP (ii)				
B.1: Size								
NA	0.047	0.050	0.048	0.047	0.046	0.049	0.048	0.048
LP	0.052	0.050	0.049	0.055	0.049	0.047	0.052	0.049
LG	0.046	0.047	0.040	0.041	0.049	0.042	0.047	0.042
ML	0.063	0.062	0.057	0.065	0.046	0.051	0.049	0.051
NP	0.066	0.067	0.064	0.063	0.058	0.057	0.048	0.052
B.2: Power								
NA	0.447	0.457	0.460	0.487	0.741	0.738	0.740	0.759
LP	0.528	0.523	0.539	0.538	0.812	0.822	0.819	0.819
LG	0.462	0.465	0.467	0.467	0.789	0.790	0.791	0.796
ML	0.569	0.578	0.564	0.575	0.841	0.846	0.844	0.838
NP	0.576	0.567	0.574	0.572	0.843	0.843	0.843	0.845

Table 11: Pointwise Test ($\tau = 0.75$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.055	0.056	0.058	0.050	0.054	0.055	0.051	0.056
LP	0.053	0.053	0.049	0.050	0.050	0.050	0.053	0.052
LG	0.065	0.071	0.069	0.069	0.065	0.067	0.065	0.065
ML	0.052	0.060	0.058	0.061	0.058	0.057	0.058	0.056
NP	0.062	0.067	0.063	0.063	0.059	0.054	0.057	0.056
A.2: Power								
NA	0.337	0.341	0.343	0.332	0.583	0.594	0.590	0.571
LP	0.426	0.434	0.430	0.441	0.685	0.694	0.696	0.698
LG	0.415	0.416	0.415	0.414	0.662	0.668	0.666	0.671
ML	0.425	0.430	0.424	0.426	0.697	0.708	0.709	0.708
NP	0.438	0.437	0.433	0.443	0.714	0.717	0.714	0.704
				Panel B: DGP (ii)				
B.1: Size								
NA	0.051	0.053	0.048	0.047	0.055	0.052	0.053	0.051
LP	0.057	0.054	0.053	0.055	0.052	0.056	0.055	0.055
LG	0.073	0.076	0.072	0.075	0.070	0.070	0.073	0.066
ML	0.062	0.068	0.063	0.064	0.058	0.055	0.056	0.059
NP	0.069	0.070	0.065	0.067	0.058	0.062	0.058	0.056
B.2: Power								
NA	0.318	0.334	0.332	0.308	0.550	0.551	0.554	0.531
LP	0.383	0.393	0.398	0.391	0.633	0.629	0.637	0.621
LG	0.405	0.404	0.409	0.403	0.633	0.632	0.639	0.622
ML	0.411	0.428	0.416	0.427	0.658	0.672	0.666	0.663
NP	0.413	0.417	0.416	0.417	0.662	0.660	0.657	0.671

B Additional Simulation Results for Tests with Naïve Estimator

This section contains the addition simulation results for the pointwise tests with $\tau = 0.25$ and 0.75 (Tables 12 and 13), test for differences (Table 14), and uniform test (Table 15), when in the estimation and bootstrap inference, the estimated propensity score $\hat{\pi}(s)$ is replaced by its true value 1/2.

Table 12: Pointwise Test with Naïve Estimator ($\tau = 0.25$, $\hat{\pi}(s) = 0.5$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.051	0.029	0.023	0.021	0.052	0.030	0.022	0.022
LP	0.048	0.019	0.010	0.007	0.054	0.016	0.007	0.006
LG	0.048	0.029	0.023	0.023	0.052	0.039	0.033	0.033
ML	0.050	0.033	0.033	0.032	0.051	0.045	0.037	0.040
NP	0.057	0.040	0.038	0.038	0.057	0.049	0.047	0.040
A.2: Power								
NA	0.253	0.235	0.217	0.237	0.465	0.447	0.441	0.459
LP	0.222	0.177	0.141	0.125	0.410	0.379	0.362	0.327
LG	0.313	0.285	0.264	0.251	0.608	0.605	0.597	0.597
ML	0.367	0.339	0.314	0.298	0.683	0.665	0.662	0.638
NP	0.394	0.356	0.344	0.323	0.690	0.670	0.673	0.661
				Panel B: DGP (ii)				
B.1: Size								
NA	0.051	0.021	0.009	0.008	0.049	0.020	0.011	0.009
LP	0.048	0.011	0.002	0.002	0.051	0.012	0.002	0.002
LG	0.037	0.025	0.018	0.016	0.039	0.024	0.018	0.019
ML	0.049	0.045	0.048	0.049	0.046	0.049	0.050	0.050
NP	0.061	0.056	0.058	0.060	0.055	0.053	0.053	0.048
B.2: Power								
NA	0.283	0.244	0.228	0.246	0.503	0.510	0.505	0.525
LP	0.234	0.193	0.155	0.119	0.439	0.402	0.397	0.368
LG	0.342	0.323	0.309	0.322	0.647	0.655	0.665	0.666
ML	0.529	0.530	0.535	0.531	0.826	0.830	0.830	0.833
NP	0.557	0.546	0.557	0.562	0.840	0.843	0.845	0.843

Table 13: Pointwise Test with Naïve Estimator ($\tau = 0.75$, $\hat{\pi}(s) = 0.5$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.054	0.034	0.023	0.025	0.054	0.034	0.018	0.023
LP	0.017	0.001	0.000	0.000	0.029	0.003	0.000	0.000
LG	0.040	0.006	0.001	0.001	0.049	0.006	0.000	0.001
ML	0.030	0.003	0.000	0.001	0.041	0.006	0.000	0.001
NP	0.031	0.004	0.000	0.000	0.043	0.006	0.001	0.000
A.2: Power								
NA	0.259	0.243	0.237	0.217	0.461	0.455	0.453	0.439
LP	0.090	0.025	0.002	0.000	0.157	0.071	0.012	0.005
LG	0.132	0.067	0.024	0.012	0.221	0.135	0.071	0.056
ML	0.124	0.049	0.014	0.007	0.225	0.137	0.072	0.049
NP	0.125	0.057	0.016	0.011	0.231	0.142	0.067	0.053
				Panel B: DGP (ii)				
B.1: Size								
NA	0.053	0.028	0.018	0.017	0.049	0.026	0.017	0.017
LP	0.010	0.000	0.000	0.000	0.023	0.002	0.000	0.000
LG	0.041	0.005	0.000	0.000	0.051	0.006	0.000	0.000
ML	0.024	0.003	0.001	0.000	0.042	0.004	0.000	0.000
NP	0.025	0.003	0.001	0.000	0.044	0.004	0.000	0.000
B.2: Power								
NA	0.232	0.215	0.191	0.176	0.405	0.389	0.389	0.373
LP	0.066	0.020	0.001	0.000	0.144	0.055	0.008	0.003
LG	0.121	0.052	0.017	0.010	0.207	0.122	0.053	0.039
ML	0.105	0.040	0.009	0.004	0.197	0.113	0.041	0.029
NP	0.106	0.044	0.011	0.005	0.196	0.116	0.048	0.032

Table 14: Test for Differences with Naïve Estimator ($\tau_1 = 0.25$, $\tau_2 = 0.75$, $\hat{\pi}(s) = 0.5$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
				Panel A: DGP (i)				
A.1: Size								
NA	0.041	0.034	0.028	0.027	0.041	0.032	0.029	0.033
LP	0.010	0.007	0.003	0.002	0.031	0.010	0.003	0.003
LG	0.030	0.006	0.001	0.002	0.034	0.008	0.002	0.002
ML	0.027	0.008	0.002	0.002	0.038	0.010	0.002	0.002
NP	0.028	0.007	0.002	0.002	0.043	0.010	0.003	0.003
A.2: Power								
NA	0.187	0.179	0.171	0.157	0.344	0.341	0.343	0.311
LP	0.082	0.058	0.037	0.027	0.197	0.150	0.110	0.093
LG	0.095	0.048	0.027	0.020	0.172	0.113	0.068	0.056
ML	0.106	0.059	0.031	0.023	0.191	0.127	0.079	0.066
NP	0.103	0.058	0.030	0.021	0.187	0.126	0.078	0.066
				Panel B: DGP (ii)				
B.1: Size								
NA	0.037	0.029	0.022	0.023	0.042	0.026	0.023	0.024
LP	0.007	0.003	0.000	0.001	0.023	0.005	0.001	0.001
LG	0.024	0.003	0.001	0.000	0.037	0.005	0.000	0.000
ML	0.021	0.005	0.000	0.000	0.039	0.005	0.001	0.001
NP	0.023	0.004	0.001	0.000	0.035	0.007	0.001	0.000
B.2: Power								
NA	0.175	0.151	0.153	0.125	0.323	0.312	0.314	0.283
LP	0.056	0.036	0.016	0.008	0.168	0.108	0.064	0.046
LG	0.073	0.028	0.009	0.004	0.139	0.066	0.030	0.019
ML	0.082	0.033	0.010	0.006	0.177	0.093	0.042	0.033
NP	0.085	0.037	0.014	0.007	0.168	0.097	0.044	0.037

Table 15: Uniform Test with Naïve Estimator ($\tau \in [0.25, 0.75]$, $\hat{\pi}(s) = 0.5$)

	$N = 200$				$N = 400$			
Methods	SRS	WEI	BCD	SBR	SRS	WEI	BCD	SBR
Panel A: DGP (i)								
A.1: Size								
NA	0.045	0.020	0.011	0.012	0.049	0.022	0.014	0.012
LP	0.031	0.004	0.001	0.001	0.036	0.005	0.001	0.001
LG	0.037	0.010	0.005	0.004	0.049	0.015	0.007	0.006
ML	0.036	0.011	0.007	0.008	0.044	0.015	0.011	0.011
NP	0.038	0.014	0.010	0.010	0.047	0.016	0.013	0.011
A.2: Power								
NA	0.298	0.269	0.250	0.244	0.562	0.560	0.577	0.564
LP	0.174	0.090	0.059	0.039	0.350	0.291	0.231	0.188
LG	0.282	0.210	0.174	0.146	0.602	0.554	0.536	0.505
ML	0.320	0.248	0.207	0.184	0.668	0.642	0.620	0.592
NP	0.342	0.280	0.238	0.206	0.682	0.649	0.630	0.609
Panel B: DGP (ii)								
B.1: Size								
NA	0.043	0.012	0.005	0.004	0.047	0.015	0.004	0.005
LP	0.028	0.001	0.000	0.000	0.041	0.003	0.000	0.000
LG	0.036	0.009	0.004	0.004	0.045	0.009	0.004	0.005
ML	0.033	0.016	0.013	0.017	0.042	0.024	0.023	0.022
NP	0.041	0.023	0.024	0.023	0.047	0.026	0.032	0.027
B.2: Power								
NA	0.302	0.258	0.221	0.219	0.562	0.574	0.577	0.572
LP	0.167	0.087	0.042	0.025	0.343	0.282	0.215	0.182
LG	0.337	0.258	0.228	0.221	0.688	0.653	0.626	0.623
ML	0.532	0.503	0.482	0.471	0.896	0.886	0.880	0.880
NP	0.591	0.552	0.535	0.526	0.918	0.910	0.896	0.905

C The Optimal Pseudo True Value when Inferring the Difference of Two QTEs

The pseudo true value $\theta_{a,s}^{LG}(\tau)$ is defined to achieve the minimum asymptotic variance of $\hat{q}^{par}(\tau)$ under the logistic model. However, it does not necessarily minimize the asymptotic variance of $\hat{q}^{par}(\tau_1) - \hat{q}^{par}(\tau_2)$, which is used to construct the test statistic for the second null hypothesis in Section 6.2.[13] In this section we derive the pseudo true value that minimizes the asymptotic variance of $\hat{q}^{par}(\tau_1) - \hat{q}^{par}(\tau_2)$ and its estimator for the logistic model. Proof of the consistency of

[13]For the linear probability model Theorem 5.2 has shown that $(\theta_{a,s}^{LG}(\tau_1), \theta_{a,s}^{LP}(\tau_2))_{a=0,1,s\in\mathcal{S}}$ still minimizes the asymptotic variance of $\hat{q}^{par}(\tau_1) - \hat{q}^{par}(\tau_2)$.

the estimator and verification of Assumptions 3 and 5 are similar to those for $\hat{\theta}^{LP}_{a,s}(\tau)$ and $\hat{\theta}^{LG}_{a,s}(\tau)$ studied in Section 5.1 and are, therefore, omitted for brevity. We first state a general result that is parallel to Theorem 5.2.

Theorem C.1. *Suppose Assumptions 1, 2, 4, 6 hold, and $\Lambda_{a,s}(X_i, \theta_{a,s}(\tau_1, \tau_2))$ is differentiable in $\theta_{a,s}(\tau_1, \tau_2)$. Then, the asymptotic variance of $\hat{q}^{par}(\tau_1) - \hat{q}^{par}(\tau_2)$ is minimized at*

$$(\theta_{1,s}(\tau_1), \theta_{0,s}(\tau_1), \theta_{1,s}(\tau_2), \theta_{0,s}(\tau_2)) \in \Theta_s(\tau_1, \tau_2),$$

where for $s \in \mathcal{S}$ and $\tau_1, \tau_2, \in \Upsilon$,

$$\begin{aligned}
\Theta_s(\tau_1, \tau_2) = & \mathop{\arg\min}_{\theta_{1,1},\theta_{0,1},\theta_{1,2},\theta_{0,2}} Q(s, \tau_1, \theta_{1,1}, \theta_{0,1}) + Q(s, \tau_2, \theta_{1,2}, \theta_{0,2}) - 2\tilde{Q}(s, \tau_1, \tau_2, \theta_{1,1}, \theta_{0,1}, \theta_{1,2}, \theta_{0,2}), \\
& \tilde{Q}(s, \tau_1, \tau_2, \theta_{1,1}, \theta_{0,1}, \theta_{1,2}, \theta_{0,2}) \qquad \text{(C.1)} \\
& = \mathbb{E}\Bigg\{ \left(\frac{g_{1,s}(X_i, \theta_{1,1})}{f_1(q_1(\tau_1))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i, \theta_{0,1})}{f_0(q_0(\tau_1))} \right) \left(\frac{g_{1,s}(X_i, \theta_{1,2})}{f_1(q_1(\tau_2))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i, \theta_{0,2})}{f_0(q_0(\tau_2))} \right) \\
& - \left(\frac{g_{1,s}(X_i, \theta_{1,1})}{f_1(q_1(\tau_1))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i, \theta_{0,1})}{f_0(q_0(\tau_1))} \right) \left(\frac{m_1(\tau_2, s, X_i) - m_1(\tau_2, s)}{f_1(q_1(\tau_2))} \right) \\
& - \left(\frac{g_{1,s}(X_i, \theta_{1,2})}{f_1(q_1(\tau_2))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i, \theta_{0,2})}{f_0(q_0(\tau_2))} \right) \left(\frac{m_1(\tau_1, s, X_i) - m_1(\tau_1, s)}{f_1(q_1(\tau_1))} \right) \\
& - \frac{\pi(s)}{1-\pi(s)} \left(\frac{g_{1,s}(X_i, \theta_{1,1})}{f_1(q_1(\tau_1))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i, \theta_{0,1})}{f_0(q_0(\tau_1))} \right) \left(\frac{m_0(\tau_2, s, X_i) - m_0(\tau_2, s)}{f_0(q_0(\tau_2))} \right) \\
& - \frac{\pi(s)}{1-\pi(s)} \left(\frac{g_{1,s}(X_i, \theta_{1,2})}{f_1(q_1(\tau_2))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i, \theta_{0,2})}{f_0(q_0(\tau_2))} \right) \left(\frac{m_0(\tau_1, s, X_i) - m_0(\tau_1, s)}{f_0(q_0(\tau_1))} \right) \Bigg| S_i = s \Bigg\}, \qquad \text{(C.2)}
\end{aligned}$$

$Q(s, \tau, \theta_1, \theta_0)$ is defined in (5.3), and $g_{a,s}(X_i, \theta_a) = \mathbb{E}(\Lambda_{a,s}(X_i, \theta_a)|S_i = s) - \Lambda_{a,s}(X_i, \theta_a)$.

For the logistic model, we need to assume $\Theta_s(\tau_1, \tau_2)$ is a singleton, denoted as

$$(\theta^{LG}_{1,s}(\tau_1), \theta^{LG}_{0,s}(\tau_1), \theta^{LG}_{1,s}(\tau_2), \theta^{LG}_{0,s}(\tau_2)).$$

We estimate the pseudo true value by minimizing the sample analogue of the objective function in Theorem C.1. Specifically,

$$\begin{aligned}
& (\tilde{\theta}^{LG}_{1,s}(\tau_1), \tilde{\theta}^{LG}_{0,s}(\tau_1), \tilde{\theta}^{LG}_{1,s}(\tau_2), \tilde{\theta}^{LG}_{0,s}(\tau_2)) \\
= & \mathop{\arg\min}_{\theta_{1,1},\theta_{0,1},\theta_{1,1},\theta_{0,1}} Q_n(s, \tau_1, \theta_{1,1}, \theta_{0,1}) + Q_n(s, \tau_2, \theta_{1,2}, \theta_{0,2}) - 2\tilde{Q}_n(s, \tau_1, \tau_2, \theta_{1,1}, \theta_{0,1}, \theta_{1,1}, \theta_{0,1}),
\end{aligned}$$

where $Q_n(s,\tau,\theta_1,\theta_0)$ is defined in (5.9) and

$$
\begin{aligned}
&\tilde{Q}_n(s,\tau_1,\tau_2,\theta_{1,1},\theta_{0,1},\theta_{1,2},\theta_{0,2})\\
=&\frac{1}{n(s)}\sum_{i\in I(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_{1,1})}{\hat{f}_1(\hat{q}_1(\tau_1))}+\frac{\hat{\pi}(s)}{1-\hat{\pi}(s)}\frac{\hat{g}_{0,s}(X_i,\theta_{0,1})}{\hat{f}_0(\hat{q}_0(\tau_1))}\right)\left(\frac{\hat{g}_{1,s}(X_i,\theta_{1,2})}{\hat{f}_1(\hat{q}_1(\tau_2))}+\frac{\hat{\pi}(s)}{1-\hat{\pi}(s)}\frac{\hat{g}_{0,s}(X_i,\theta_{0,2})}{\hat{f}_0(\hat{q}_0(\tau_2))}\right)\\
&+\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_{1,1})}{\hat{f}_1(\hat{q}_1(\tau_1))}+\frac{\hat{\pi}(s)}{1-\hat{\pi}(s)}\frac{\hat{g}_{0,s}(X_i,\theta_{0,1})}{\hat{f}_0(\hat{q}_0(\tau_1))}\right)\frac{1\{Y_i\le\hat{q}_1(\tau_2)\}}{\hat{f}_1(\hat{q}_1(\tau_2))}\\
&+\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_{1,2})}{\hat{f}_1(\hat{q}_1(\tau_2))}+\frac{\hat{\pi}(s)}{1-\hat{\pi}(s)}\frac{\hat{g}_{0,s}(X_i,\theta_{0,2})}{\hat{f}_0(\hat{q}_0(\tau_2))}\right)\frac{1\{Y_i\le\hat{q}_1(\tau_1)\}}{\hat{f}_1(\hat{q}_1(\tau_1))}\\
&+\frac{\hat{\pi}(s)}{n_0(s)(1-\hat{\pi}(s))}\sum_{i\in I_0(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_{1,1})}{\hat{f}_1(\hat{q}_1(\tau_1))}+\frac{\hat{\pi}(s)}{1-\hat{\pi}(s)}\frac{\hat{g}_{0,s}(X_i,\theta_{0,1})}{\hat{f}_0(\hat{q}_0(\tau_1))}\right)\frac{1\{Y_i\le\hat{q}_0(\tau_2)\}}{\hat{f}_0(\hat{q}_0(\tau_2))}\\
&+\frac{\hat{\pi}(s)}{n_0(s)(1-\hat{\pi}(s))}\sum_{i\in I_0(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_{1,2})}{\hat{f}_1(\hat{q}_1(\tau_2))}+\frac{\hat{\pi}(s)}{1-\hat{\pi}(s)}\frac{\hat{g}_{0,s}(X_i,\theta_{0,2})}{\hat{f}_0(\hat{q}_0(\tau_2))}\right)\frac{1\{Y_i\le\hat{q}_0(\tau_1)\}}{\hat{f}_0(\hat{q}_0(\tau_1))}.
\end{aligned}
$$

D When $\Theta_s(\tau)$ Is Not a Singleton

Recall the definition of $\Theta_s(\tau)$ in Assumption 9 for the logistic model. In this section we relax the requirement that $\Theta_s(\tau)\cap\Theta$ is a singleton and propose a way to consistently estimate one point, denoted as $\theta_{a,s}(\tau)$, that belongs to the set of optimizers. We work with a fixed τ. The extension of the results to multiple τ's is straightforward. The extension to a continuum of τ's is left for future research.

Let ε_n be some deterministic sequence such that $\varepsilon_n \downarrow 0$ and

$$\hat{\Theta}_s^{\varepsilon_n}(\tau)=\{(\theta_1,\theta_0)\in\Theta: Q_n(s,\tau,\theta_1,\theta_0)\le\inf_{(\theta_1,\theta_0)\in\Theta}Q_n(s,\tau,\theta_1,\theta_0)+\varepsilon_n\}.$$

Define

$$(\hat{\theta}^*_{1,s}(\tau),\hat{\theta}^*_{0,s}(\tau))=\operatorname*{arg\,min}_{(\theta_1,\theta_0)\in\hat{\Theta}_s^{\varepsilon_n}(\tau)}(||\theta_1||_2^2+||\theta_0||_2^2)$$

and

$$(\theta^*_{1,s}(\tau),\theta^*_{0,s}(\tau))=\operatorname*{arg\,min}_{(\theta_1,\theta_0)\in\Theta_s(\tau)\cap\Theta}(||\theta_1||_2^2+||\theta_0||_2^2).$$

Assumption 13. *Suppose* $(\theta_{1,s}(\tau),\theta_{0,s}(\tau))$ *is uniquely defined and* $|\hat{f}_a(\hat{q}_a(\tau))-f_a(q_a(\tau))|=o_p(\varepsilon_n)$ *for* $a=0,1$.

Proposition D.1. *Suppose Assumptions 1, 2, 7, 8, and 13 hold. Then, pointwise in τ,*

$$\hat{\theta}^*_{a,s}(\tau) \xrightarrow{p} \theta^*_{a,s}(\tau).$$

E Additional Notation

Throughout the supplement we denote $(\xi_i^s, X_i^s, Y_i^s(1), Y_i^s(0))_{i\in[n]}$ as an i.i.d. sequence with marginal distribution equal to the conditional distribution of $(\xi_i, S_i, Y_i(1), Y_i(0))$ given $S_i = s$. In addition, $\{(\xi_i^s, X_i^s, Y_i^s(1), Y_i^s(0))_{i\in[n]}\}_{s\in\mathcal{S}}$ are independent across s and with $\{A_i, S_i\}_{i\in[n]}$. We further denote $\mathcal{F}$ as a generic class of functions which differs in different contexts. The envelope of $\mathcal{F}$ is denoted as F_i. We say $\mathcal{F}$ is of VC-type with coefficients (α_n, v_n) if

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

where $N(\cdot)$ denote the covering number, $e_Q(f,g) = ||f-g||_{Q,2}$, and the supremum is taken over all finitely discrete probability measures.

F Proof of Theorem 3.1

We first derive the linear expansion of $\hat{q}_1^{adj}(\tau)$. By Knight's identity (Knight (1998)), we have

$$\begin{aligned} L_n(u,\tau) &= \sum_{i\in[n]}\left[\frac{A_i}{\hat{\pi}(S_i)}\left[\rho_\tau(Y_i - q_1(\tau) - u/\sqrt{n}) - \rho_\tau(Y_i - q_1(\tau))\right] + \frac{(A_i - \hat{\pi}(S_i))}{\hat{\pi}(S_i)\sqrt{n}}\widehat{m}_1(\tau, S_i, X_i)u\right] \\ &\equiv -L_{1,n}(\tau)u + L_{2,n}(u,\tau), \end{aligned}$$

where

$$L_{1,n}(\tau) = \frac{1}{\sqrt{n}}\sum_{i\in[n]}\left[\frac{A_i}{\hat{\pi}(S_i)}(\tau - 1\{Y_i \leq q_1(\tau)\}) - \frac{(A_i - \hat{\pi}(S_i))}{\hat{\pi}(S_i)}\widehat{m}_1(\tau, S_i, X_i)\right]$$

and

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

By change of variables, we have

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

Note that $L_{2,n}(\tau)$ is exactly the same as that considered in the proof of Theorem 3.2 in Zhang and Zheng (2020) and by their result we have

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

Next, consider $L_{1,n}(\tau)$. Denote $m_1(\tau,s) = \mathbb{E}(m_1(\tau,S_i,X_i)|S_i=s)$, $\eta_{i,1}(s,\tau) = \tau - 1\{Y_i \le q_1(\tau)\} - m_1(\tau,s)$, and

$$\begin{aligned}
L_{1,n}(\tau) =& \frac{1}{\sqrt{n}}\sum_{i\in[n]}\left[\frac{A_i}{\hat{\pi}(S_i)}(\tau - 1\{Y_i \le q_1(\tau)\})\right] - \frac{1}{\sqrt{n}}\sum_{i\in[n]}\left[\frac{(A_i - \hat{\pi}(S_i))}{\hat{\pi}(S_i)}\widehat{m}_1(\tau,S_i,X_i)\right] \\
\equiv& L_{1,1,n}(\tau) - L_{1,2,n}(\tau).
\end{aligned}$$

First, note that $\hat{\pi}(s) - \pi(s) = \frac{D_n(s)}{n(s)}$. Therefore,

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

where

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

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

In addition, note that

$$\{\tau-1\{Y_i(1)\leq q_1(\tau)\}-m_1(\tau,S_i):\tau\in\Upsilon\}$$

is of the VC-type with fixed coefficients (α,v) and bounded envelope, and $\mathbb{E}(\tau-1\{Y_i(1)\leq q_1(\tau)\}-m_1(\tau,S_i)|S_i=s)=0$. Therefore, Lemma P.2 implies

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

By Assumption 1 we have $\max_{s\in\mathcal{S}}|D_n(s)/n(s)|=o_p(1)$, $\max_{s\in\mathcal{S}}|\hat{\pi}(s)-\pi(s)|=o_p(1)$, and $\min_{s\in\mathcal{S}}\pi(s)>c>0$, which imply $\sup_{\tau\in\Upsilon}|R_{1,1}(\tau)|=o_p(1)$.

Next, denote $\overline{m}_1(\tau,s)=\mathbb{E}(\overline{m}_1(q_1(\tau),X_i,S_i)|S_i=s)$. Then

$$L_{1,2,n}=\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\frac{A_i}{\hat{\pi}(s)}\overline{m}_1(\tau,s,X_i)1\{S_i=s\}-\frac{1}{\sqrt{n}}\sum_{i\in[n]}\overline{m}_1(\tau,S_i,X_i)$$
$$+\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}(s)}\sum_{i\in[n]}(A_i-\hat{\pi}(s))\left(\widehat{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s,X_i)\right)1\{S_i=s\}$$
$$=\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\frac{A_i}{\hat{\pi}(s)}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}$$
$$-\frac{1}{\sqrt{n}}\sum_{i\in[n]}\left(\overline{m}_1(\tau,S_i,X_i)-\overline{m}_1(\tau,S_i)\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}(s)}\sum_{i\in[n]}(A_i-\hat{\pi}(s))\left(\widehat{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s,X_i)\right)1\{S_i=s\}$$
$$=\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\frac{A_i}{\pi(s)}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}$$
$$-\frac{1}{\sqrt{n}}\sum_{i\in[n]}\left(\overline{m}_1(\tau,S_i,X_i)-\overline{m}_1(\tau,S_i)\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\left(\frac{\pi(s)-\hat{\pi}(s)}{\hat{\pi}(s)\pi(s)}\right)\left(\sum_{i\in[n]}A_i(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}\right)$$
$$+\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}(s)}\sum_{i\in[n]}(A_i-\hat{\pi}(s))\left(\widehat{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s,X_i)\right)1\{S_i=s\}$$

$$\equiv \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{S}} \sum_{i\in[n]} \frac{A_i}{\pi(s)} (\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)) 1\{S_i = s\}$$

$$- \frac{1}{\sqrt{n}} \sum_{i\in[n]} (\overline{m}_1(\tau, S_i, X_i) - \overline{m}_1(\tau, S_i)) + R_2(\tau), \tag{F.2}$$

where the second equality holds because

$$\sum_{s\in\mathcal{S}} \sum_{i\in[n]} \frac{A_i}{\hat{\pi}(s)} \overline{m}_1(\tau, s) 1\{S_i = s\} = \sum_{s\in\mathcal{S}} n(s) \overline{m}_1(\tau, s) = \sum_{i\in[n]} \overline{m}_1(\tau, S_i).$$

For the first term of $R_2(\tau)$, we have

$$\sup_{\tau\in\Upsilon} \left| \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{S}} \left(\frac{\pi(s) - \hat{\pi}(s)}{\hat{\pi}(s)\pi(s)} \right) \left(\sum_{i\in[n]} A_i (\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)) 1\{S_i = s\} \right) \right|$$

$$\leq \sum_{s\in\mathcal{S}} \left| \frac{D_n(s)}{n_1(s)\pi(s)} \right| \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i\in[n]} A_i 1\{S_i = s\} (\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)) \right|.$$

Assumption 3 implies

$$\mathcal{F} = \{\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s) : \tau \in \Upsilon\}$$

is of the VC-type with fixed coefficients (α, v) and an envelope F_i such that $\mathbb{E}(|F_i|^q | S_i = s) < \infty$ for $q > 2$. Therefore,

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i\in[n]} A_i 1\{S_i = s\} (\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)) \right| = O_p(n^{-1/2}).$$

It is also assumed that $D_n(s)/n(s) = o_p(1)$ and $n(s)/n_1(s) \xrightarrow{p} 1/\pi(s) < \infty$. Therefore, we have

$$\sup_{\tau\in\Upsilon} \left| \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{S}} \left(\frac{\pi(s) - \hat{\pi}(s)}{\hat{\pi}(s)\pi(s)} \right) \left(\sum_{i\in[n]} A_i (\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)) 1\{S_i = s\} \right) \right| = o_p(1).$$

Recall $\overline{\Delta}_1(\tau, s, X_i) = \widehat{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s, X_i)$. Then, for the second term of $R_2(\tau)$, we have

$$\left| \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{S}} \frac{1}{\hat{\pi}(s)} \sum_{i\in[n]} (A_i - \hat{\pi}(s)) \overline{\Delta}_1(\tau, s, X_i) 1\{S_i = s\} \right|$$

$$= \frac{1}{\sqrt{n}} \sum_{s\in\mathcal{S}} n_0(s) \sup_{\tau\in\Upsilon} \left| \frac{\sum_{i\in I_1(s)} \overline{\Delta}_1(\tau, s, X_i)}{n_1(s)} - \frac{\sum_{i\in I_0(s)} \overline{\Delta}_1(\tau, s, X_i)}{n_0(s)} \right| = o_p(1),$$

where the last equality holds by Assumption 3(i). Therefore, we have

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

Combining (F.1) and (F.2), we have

$$\begin{aligned}
L_{1,n}(\tau) = & \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i 1\{S_i = s\} \left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)} + \left(1 - \frac{1}{\pi(s)}\right) \left(\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)\right) \right] \\
& + \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - A_i) 1\{S_i = s\} \left(\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)\right) \\
& + \sum_{i=1}^{n} \frac{m_1(\tau, S_i)}{\sqrt{n}} + R_{1,1}(\tau) - R_{1,2}(\tau).
\end{aligned}$$

Note by Assumption 3 that the classes of functions

$$\left\{ \frac{\eta_{i,1}(s,\tau)}{\pi(s)} + \left(1 - \frac{1}{\pi(s)}\right) \left(\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)\right) : \tau \in \Upsilon \right\}$$

and

$$\{\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s) : \tau \in \Upsilon\}$$

are of the VC-type with fixed coefficients and envelopes belonging to $L_{\mathbb{P},q}$. In addition,

$$\mathbb{E}\left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)} + \left(1 - \frac{1}{\pi(s)}\right) \left(\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)\right) | S_i = s \right] = 0$$

and

$$\mathbb{E}(\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s) | S_i = s) = 0.$$

Therefore, Lemma P.2 implies,

$$\sup_{\tau \in \Upsilon} \left| \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i 1\{S_i = s\} \left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)} + \left(1 - \frac{1}{\pi(s)}\right) \left(\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)\right) \right] \right| = O_p(1)$$

and

$$\sup_{\tau \in \Upsilon} \left| \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - A_i) 1\{S_i = s\} \left(\overline{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s)\right) \right| = O_p(1).$$

This implies $\sup_{\tau\in\Upsilon}|L_{1,n}(\tau)| = O_p(1)$. Then by Kato (2009, Theorem 2), we have

$$
\begin{aligned}
&\sqrt{n}(\hat{q}_1^{adj}(\tau) - q_1(\tau)) \\
&= \frac{1}{f_1(q_1(\tau))}\Bigg\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i 1\{S_i = s\}\left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)} + \left(1 - \frac{1}{\pi(s)}\right)(\overline{m}_1(\tau,s,X_i) - \overline{m}_1(\tau,s))\right] \\
&+ \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-A_i)1\{S_i = s\}(\overline{m}_1(\tau,s,X_i) - \overline{m}_1(\tau,s)) + \sum_{i=1}^{n}\frac{m_1(\tau,S_i)}{\sqrt{n}}\Bigg\} + R_{q,1}(\tau),
\end{aligned}
$$

where $\sup_{\tau\in\Upsilon}|R_{q,1}(\tau)| = o_p(1)$. Similarly, we have

$$
\begin{aligned}
&\sqrt{n}(\hat{q}_0^{adj}(\tau) - q_0(\tau)) \\
=&\frac{1}{f_0(q_0(\tau))}\Bigg\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-A_i)1\{S_i = s\}\left[\frac{\eta_{i,0}(s,\tau)}{1-\pi(s)} + \left(1 - \frac{1}{1-\pi(s)}\right)(\overline{m}_0(\tau,s,X_i) - \overline{m}_0(\tau,s))\right] \\
&+ \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i 1\{S_i = s\}(\overline{m}_0(\tau,s,X_i) - \overline{m}_0(\tau,s)) + \sum_{i=1}^{n}\frac{m_0(\tau,S_i)}{\sqrt{n}}\Bigg\} + R_{q,0}(\tau),
\end{aligned}
$$

where $\eta_{i,0}(s,\tau) = \tau - 1\{Y_i(0) \le q_0(\tau)\} - m_0(\tau,s)$ and $\sup_{\tau\in\Upsilon}|R_{q,0}(\tau)| = o_p(1)$. Taking the difference of the above two displays gives

$$
\begin{aligned}
&\sqrt{n}(\hat{q}^{adj}(\tau) - q(\tau)) \\
&= \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i\in[n]} A_i 1\{S_i = s\}\left[\frac{\eta_{i,1}(s,\tau) - (1-\pi(s))(\overline{m}_1(\tau,s,X_i) - \overline{m}_1(\tau,s))}{\pi(s) f_1(q_1(\tau))} - \frac{(\overline{m}_0(\tau,s,X_i) - \overline{m}_0(\tau,s))}{f_0(q_0(\tau))}\right] \\
&- \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i\in[n]}(1-A_i)1\{S_i = s\}\left[\frac{\eta_{i,0}(s,\tau) - \pi(s)(\overline{m}_0(\tau,s,X_i) - \overline{m}_0(\tau,s))}{(1-\pi(s)) f_0(q_0(\tau))} - \frac{(\overline{m}_1(\tau,s,X_i) - \overline{m}_1(\tau,s))}{f_1(q_1(\tau))}\right] \\
&+ \frac{1}{\sqrt{n}}\sum_{i\in[n]}\left(\frac{m_1(\tau,S_i)}{f_1(q_1(\tau))} - \frac{m_0(\tau,S_i)}{f_0(q_0(\tau))}\right) + R_q(\tau) \\
&\equiv \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i\in[n]} A_i 1\{S_i = s\}\phi_1(\tau,s,Y_i(1),X_i) - \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i\in[n]}(1-A_i)1\{S_i = s\}\phi_0(\tau,s,Y_i(0),X_i) \\
&+ \frac{1}{\sqrt{n}}\sum_{i\in[n]}\phi_s(\tau,S_i) + R_q(\tau),
\end{aligned}
$$

where $\sup_{\tau\in\Upsilon}|R_q(\tau)| = o_p(1)$. Lemma P.3 shows that, uniformly over $\tau\in\Upsilon$,

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

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

$$\begin{aligned}\Sigma(\tau,\tau') = &\mathbb{E}\pi(S_i)\phi_1(\tau,S_i,Y_i(1),X_i)\phi_1(\tau',S_i,Y_i(1),X_i)\\ &+\mathbb{E}(1-\pi(S_i))\phi_0(\tau,S_i,Y_i(0),X_i)\phi_0(\tau',S_i,Y_i(0),X_i)\\ &+\mathbb{E}\phi_s(\tau,S_i)\phi_s(\tau',S_i).\end{aligned}$$

For the second result in Theorem 3.1, we denote

$$\delta_a(\tau,S_i,X_i) = m_a(\tau,S_i,X_i) - m_a(\tau,S_i) \quad \text{and} \quad \overline{\delta}_a(\tau,S_i,X_i) = (\overline{m}_a(\tau,S_i,X_i) - \overline{m}_a(\tau,S_i)),\ a=0,1. \tag{F.3}$$

Then

$$\begin{aligned}&\mathbb{E}\pi(S_i)\phi_1(\tau,S_i,Y_i(1),X_i)\phi_1(\tau',S_i,Y_i(1),X_i)\\ &=\mathbb{E}\frac{1}{\pi(S_i)}\left[\frac{(\tau-1\{Y_i(1)\le q_1(\tau)\}-m_1(\tau,S_i,X_i))}{f_1(q_1(\tau))}\right]\left[\frac{(\tau'-1\{Y_i(1)\le q_1(\tau')\}-m_1(\tau',S_i,X_i))}{f_1(q_1(\tau'))}\right]\\ &+\mathbb{E}\pi(S_i)\left[\frac{\delta_1(\tau,S_i,X_i)-\overline{\delta}_1(\tau,S_i,X_i)}{\pi(S_i)f_1(q_1(\tau))}+\left(\frac{\overline{\delta}_1(\tau,S_i,X_i)}{f_1(q_1(\tau))}-\frac{\overline{\delta}_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right)\right]\\ &\times\left[\frac{\delta_1(\tau',S_i,X_i)-\overline{\delta}_1(\tau',S_i,X_i)}{\pi(S_i)f_1(q_1(\tau'))}+\left(\frac{\overline{\delta}_1(\tau',S_i,X_i)}{f_1(q_1(\tau'))}-\frac{\overline{\delta}_0(\tau',S_i,X_i)}{f_0(q_0(\tau'))}\right)\right],\end{aligned}$$

$$\begin{aligned}&\mathbb{E}(1-\pi(S_i))\phi_0(\tau,S_i,Y_i(1),X_i)\phi_0(\tau',S_i,Y_i(1),X_i)\\ &=\mathbb{E}\frac{1}{1-\pi(S_i)}\left[\frac{(\tau-1\{Y_i(0)\le q_0(\tau)\}-m_0(\tau,S_i,X_i))}{f_1(q_1(\tau))}\right]\left[\frac{(\tau'-1\{Y_i(0)\le q_0(\tau')\}-m_0(\tau',S_i,X_i))}{f_1(q_1(\tau'))}\right]\\ &+\mathbb{E}(1-\pi(S_i))\left[\frac{\delta_0(\tau,S_i,X_i)-\overline{\delta}_0(\tau,S_i,X_i)}{(1-\pi(S_i))f_0(q_0(\tau))}-\left(\frac{\overline{\delta}_1(\tau,S_i,X_i)}{f_1(q_1(\tau))}-\frac{\overline{\delta}_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right)\right]\\ &\times\left[\frac{\delta_0(\tau',S_i,X_i)-\overline{\delta}_0(\tau',S_i,X_i)}{(1-\pi(S_i))f_0(q_0(\tau'))}-\left(\frac{\overline{\delta}_1(\tau',S_i,X_i)}{f_1(q_1(\tau'))}-\frac{\overline{\delta}_0(\tau',S_i,X_i)}{f_0(q_0(\tau'))}\right)\right],\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\phi_s(\tau,S_i)\phi_s(\tau',S_i) &= \mathbb{E}\left(\frac{m_1(\tau,S_i)}{f_1(q_1(\tau))}-\frac{m_0(\tau,S_i)}{f_0(q_0(\tau))}\right)\left(\frac{m_1(\tau',S_i)}{f_1(q_1(\tau'))}-\frac{m_0(\tau',S_i)}{f_0(q_0(\tau'))}\right)\\ &=\mathbb{E}\left(\frac{m_1(\tau,S_i,X_i)}{f_1(q_1(\tau))}-\frac{m_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right)\left(\frac{m_1(\tau',S_i,X_i)}{f_1(q_1(\tau'))}-\frac{m_0(\tau',S_i,X_i)}{f_0(q_0(\tau'))}\right)\\ &+\mathbb{E}\left(\frac{\delta_1(\tau,S_i,X_i)}{f_1(q_1(\tau))}-\frac{\delta_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right)\left(\frac{\delta_1(\tau',S_i,X_i)}{f_1(q_1(\tau'))}-\frac{\delta_0(\tau',S_i,X_i)}{f_0(q_0(\tau'))}\right).\end{aligned}$$

Let

$$
\begin{aligned}
&\Sigma^*(\tau,\tau') \\
&= \mathbb{E}\frac{1}{\pi(S_i)}\left[\frac{(\tau - 1\{Y_i(1) \le q_1(\tau)\} - m_1(\tau, S_i, X_i))}{f_1(q_1(\tau))}\right]\left[\frac{(\tau' - 1\{Y_i(1) \le q_1(\tau')\} - m_1(\tau', S_i, X_i))}{f_1(q_1(\tau'))}\right] \\
&+ \mathbb{E}\frac{1}{1-\pi(S_i)}\left[\frac{(\tau - 1\{Y_i(0) \le q_0(\tau)\} - m_0(\tau, S_i, X_i))}{f_1(q_1(\tau))}\right]\left[\frac{(\tau' - 1\{Y_i(0) \le q_0(\tau')\} - m_0(\tau', S_i, X_i))}{f_1(q_1(\tau'))}\right] \\
&+ \mathbb{E}\left(\frac{m_1(\tau, S_i, X_i)}{f_1(q_1(\tau))} - \frac{m_0(\tau, S_i, X_i)}{f_0(q_0(\tau))}\right)\left(\frac{m_1(\tau', S_i, X_i)}{f_1(q_1(\tau'))} - \frac{m_0(\tau', S_i, X_i)}{f_0(q_0(\tau'))}\right),
\end{aligned}
$$

which does not rely on the working models. Then,

$$
\begin{aligned}
&\Sigma(\tau,\tau') - \Sigma^*(\tau,\tau') \\
&= \mathbb{E}\pi(S_i)\left[\frac{\delta_1(\tau, S_i, X_i) - \overline{\delta}_1(\tau, S_i, X_i)}{\pi(S_i)f_1(q_1(\tau))} + \left(\frac{\overline{\delta}_1(\tau, S_i, X_i)}{f_1(q_1(\tau))} - \frac{\overline{\delta}_0(\tau, S_i, X_i)}{f_0(q_0(\tau))}\right)\right] \\
&\times \left[\frac{\delta_1(\tau', S_i, X_i) - \overline{\delta}_1(\tau', S_i, X_i)}{\pi(S_i)f_1(q_1(\tau'))} + \left(\frac{\overline{\delta}_1(\tau', S_i, X_i)}{f_1(q_1(\tau'))} - \frac{\overline{\delta}_0(\tau', S_i, X_i)}{f_0(q_0(\tau'))}\right)\right] \\
&+ \mathbb{E}(1-\pi(S_i))\left[\frac{\delta_0(\tau, S_i, X_i) - \overline{\delta}_0(\tau, S_i, X_i)}{(1-\pi(S_i))f_0(q_0(\tau))} - \left(\frac{\overline{\delta}_1(\tau, S_i, X_i)}{f_1(q_1(\tau))} - \frac{\overline{\delta}_0(\tau, S_i, X_i)}{f_0(q_0(\tau))}\right)\right] \\
&\times \left[\frac{\delta_0(\tau', S_i, X_i) - \overline{\delta}_0(\tau', S_i, X_i)}{(1-\pi(S_i))f_0(q_0(\tau'))} - \left(\frac{\overline{\delta}_1(\tau', S_i, X_i)}{f_1(q_1(\tau'))} - \frac{\overline{\delta}_0(\tau', S_i, X_i)}{f_0(q_0(\tau'))}\right)\right] \\
&- \mathbb{E}\left(\frac{\delta_1(\tau, S_i, X_i)}{f_1(q_1(\tau))} - \frac{\delta_0(\tau, S_i, X_i)}{f_0(q_0(\tau))}\right)\left(\frac{\delta_1(\tau', S_i, X_i)}{f_1(q_1(\tau'))} - \frac{\delta_0(\tau', S_i, X_i)}{f_0(q_0(\tau'))}\right) \\
&= \mathbb{E}\left[\sqrt{\frac{1-\pi(S_i)}{\pi(S_i)}}\frac{\delta_1(\tau, S_i, X_i) - \overline{\delta}_1(\tau, S_i, X_i)}{f_1(q_1(\tau))} + \sqrt{\frac{\pi(S_i)}{1-\pi(S_i)}}\frac{\delta_0(\tau, S_i, X_i) - \overline{\delta}_0(\tau, S_i, X_i)}{f_0(q_0(\tau))}\right] \\
&\times \left[\sqrt{\frac{1-\pi(S_i)}{\pi(S_i)}}\frac{\delta_1(\tau', S_i, X_i) - \overline{\delta}_1(\tau', S_i, X_i)}{f_1(q_1(\tau'))} + \sqrt{\frac{\pi(S_i)}{1-\pi(S_i)}}\frac{\delta_0(\tau', S_i, X_i) - \overline{\delta}_0(\tau', S_i, X_i)}{f_0(q_0(\tau'))}\right] \\
&\equiv \mathbb{E}a_i(\tau)a_i(\tau'),
\end{aligned}
$$

where

$$
a_i(\tau) = \sqrt{\frac{1-\pi(S_i)}{\pi(S_i)}}\frac{\delta_1(\tau, S_i, X_i) - \overline{\delta}_1(\tau, S_i, X_i)}{f_1(q_1(\tau))} + \sqrt{\frac{\pi(S_i)}{1-\pi(S_i)}}\frac{\delta_0(\tau, S_i, X_i) - \overline{\delta}_0(\tau, S_i, X_i)}{f_0(q_0(\tau))}.
$$

Further, denote $\vec{a}_i = (a_i(\tau_1), \cdots, a_i(\tau_K))^\top$, the asymptotic variance covariance matrix of $(\hat{q}^{adj}(\tau_1), \cdots, \hat{q}^{adj}(\tau_K))$ as $[\Sigma_{kl}]_{k,l\in[K]}$, and the optimal variance covariance matrix as $[\Sigma^*_{kl}]_{k,l\in[K]}$. We have

$$
[\Sigma_{kl}]_{k,l\in[K]} - [\Sigma^*_{kl}]_{k,l\in[K]} = [\mathbb{E}a_i(\tau_k)a_i(\tau_l)]_{k,l\in[K]} = \mathbb{E}\vec{a}_i\vec{a}_i^\top,
$$

which is positive semidefinite. In addition, $\mathbb{E}\vec{a}_i\vec{a}_i^\top = 0$ if $\overline{m}_a(\tau,s,x) = m_a(\tau,s,x)$ for $a = 0,1$, $\tau \in \{\tau_1, \cdots, \tau_K\}$, and (s,x) in the joint support of (S_i, X_i). This concludes the proof.

G Proof of Theorem 4.1

We focus on deriving the linear expansion of $\hat{q}_1^w(\tau)$. Let

$$\begin{aligned} L_n^w(u,\tau) =& \sum_{i\in[n]} \xi_i \left[\frac{A_i}{\hat{\pi}^w(S_i)}\left[\rho_\tau(Y_i - q_1(\tau) - u/\sqrt{n}) - \rho_\tau(Y_i - q_1(\tau))\right] + \frac{(A_i - \hat{\pi}^w(S_i))}{\hat{\pi}^w(S_i)\sqrt{n}}\widehat{m}_1(\tau, S_i, X_i)u\right] \\ \equiv& - L_{1,n}^w(\tau)u + L_{2,n}^w(u,\tau), \end{aligned}$$

where

$$L_{1,n}^w(\tau) = \frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_i\left[\frac{A_i}{\hat{\pi}^w(S_i)}(\tau - 1\{Y_i \le q_1(\tau)\}) - \frac{(A_i - \hat{\pi}^w(S_i))}{\hat{\pi}^w(S_i)}\widehat{m}_1(\tau, S_i, X_i)\right],$$

and

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

By the change of variables, we have

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

Note that $L_{2,n}^w(\tau)$ is exactly the same as that considered in the proof of Theorem 3.2 in Zhang and Zheng (2020) and by their result we have

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

Next consider $L_{1,n}^w(\tau)$. Recall $m_1(\tau,s) = \mathbb{E}(m_1(\tau,S_i,X_i)|S_i = s)$ and $\eta_{i,1}(s,\tau) = \tau - 1\{Y_i \le q_1(\tau)\} - m_1(\tau,s)$. Denote

$$\begin{aligned} L_{1,n}^w(\tau) =& \frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_i\left[\frac{A_i}{\hat{\pi}^w(S_i)}(\tau - 1\{Y_i \le q_1(\tau)\})\right] - \frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_i\left[\frac{(A_i - \hat{\pi}^w(S_i))}{\hat{\pi}^w(S_i)}\widehat{m}_1(\tau, S_i, X_i)\right] \\ \equiv& L_{1,1,n}^w(\tau) - L_{1,2,n}^w(\tau). \end{aligned}$$

First, note that

$$
\begin{aligned}
& L^w_{1,1,n}(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i}{\pi(s)} 1\{S_i = s\}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\}(\hat{\pi}^w(s) - \pi(s))}{\sqrt{n}\hat{\pi}^w(s)\pi(s)} (\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
& = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i}{\pi(s)} 1\{S_i = s\}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\} D^w_n(s)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi(s)} \eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(\tau,s)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi(s)} D^w_n(s) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(\tau,s)}{\sqrt{n}\hat{\pi}^w(s)} \\
& = \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{\pi(s)} \eta_{i,1}(s,\tau) + \sum_{s \in \mathcal{S}} \frac{D^w_n(s)}{\sqrt{n}\pi(s)} m_1(\tau,s) + \sum_{i=1}^n \frac{\xi_i m_1(\tau,S_i)}{\sqrt{n}} \\
& - \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\} D^w_n(s)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi(s)} \eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(s,\tau)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi(s)} D^w_n(s) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(\tau,s)}{\sqrt{n}\hat{\pi}^w(s)} \\
& = \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{\pi(s)} \eta_{i,1}(s,\tau) + \sum_{i=1}^n \frac{\xi_i m_1(\tau,S_i)}{\sqrt{n}} + R^w_{1,1}(\tau), \qquad \text{(G.1)}
\end{aligned}
$$

where $D^w_n(s) = \sum_{i \in [n]} \xi_i (A_i - \pi(S_i)) 1\{S_i = s\} = (\pi^w(s) - \pi(s)) n^w(s)$,

$$
\begin{aligned}
R^w_{1,1}(\tau) = & - \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\} D^w_n(s)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi(s)} \eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(\tau,s)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi(s)} D^w_n(s) \\
& + \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(\tau,s)}{\sqrt{n}} \left(\frac{1}{\pi(s)} - \frac{1}{\hat{\pi}^w(s)} \right) = - \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\} D^w_n(s)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi(s)} \eta_{i,1}(s,\tau).
\end{aligned}
$$

Note that

$$
\{\xi_i(\tau - 1\{Y_i(1) \leq q_1(\tau)\} - m_1(\tau,S_i)) : \tau \in \Upsilon\}
$$

is of the VC-type with fixed coefficients (α, v) and the envelope $F_i = \xi_i$, and

$$
\mathbb{E}\left[\xi_i(\tau - 1\{Y_i(1) \leq q_1(\tau)\} - m_1(\tau,S_i)) | S_i = s\right] = 0.
$$

We can also let $\sigma^2_n = \mathbb{E}(F^2_i | S_i = s) \leq C < \infty$ for some constant C. Then, Lemma P.2 implies

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

In addition, Lemma P.4 implies $\max_{s \in \mathcal{S}} |D^w_n(s)/n^w(s)| = o_p(1)$, which further implies $\max_{s \in \mathcal{S}} |\hat{\pi}^w(s) -$

$\pi(s)| = o_p(1)$. Combining these results, we have

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

Next, recall $\overline{m}_1(\tau,s) = \mathbb{E}(\overline{m}_1(q_1(\tau),X_i,S_i)|S_i=s)$. Then

$$\begin{aligned}
L^w_{1,2,n} &= \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\frac{\xi_i A_i}{\hat{\pi}^w(s)}\overline{m}_1(\tau,s,X_i)1\{S_i=s\} - \frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_i\overline{m}_1(\tau,S_i,X_i)\\
&+ \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}^w(s)}\sum_{i\in[n]}\xi_i(A_i-\hat{\pi}^w(s))\left(\widehat{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s,X_i)\right)1\{S_i=s\}\\
&= \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\frac{\xi_i A_i}{\hat{\pi}^w(s)}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}\\
&- \frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_i\left(\overline{m}_1(\tau,S_i,X_i)-\overline{m}_1(\tau,S_i)\right)\\
&+ \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}^w(s)}\sum_{i\in[n]}\xi_i(A_i-\hat{\pi}^w(s))\left(\widehat{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s,X_i)\right)1\{S_i=s\}\\
&= \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\frac{\xi_i A_i}{\pi(s)}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}\\
&- \frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_i\left(\overline{m}_1(\tau,S_i,X_i)-\overline{m}_1(\tau,S_i)\right)\\
&+ \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\left(\frac{\pi(s)-\hat{\pi}^w(s)}{\hat{\pi}^w(s)\pi(s)}\right)\left(\sum_{i\in[n]}\xi_i A_i(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}\right)\\
&+ \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}^w(s)}\sum_{i\in[n]}\xi_i(A_i-\hat{\pi}^w(s))\left(\widehat{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s,X_i)\right)1\{S_i=s\}\\
&\equiv \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\frac{\xi_i A_i}{\pi(s)}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}\\
&- \frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_i\left(\overline{m}_1(\tau,S_i,X_i)-\overline{m}_1(\tau,S_i)\right) + R^w_{1,2}(\tau),
\end{aligned}\tag{G.2}$$

where the second equality holds because

$$\begin{aligned}
\sum_{s\in\mathcal{S}}\sum_{i\in[n]}\xi_i A_i 1\{S_i=s\}\frac{\overline{m}_1(\tau,s)}{\hat{\pi}^w(s)} &= \sum_{s\in\mathcal{S}} n^w_1(s)\frac{\overline{m}_1(\tau,s)}{\hat{\pi}^w(s)} = \sum_{s\in\mathcal{S}} n^w(s)\overline{m}_1(\tau,s)\\
&= \sum_{i\in[n]}\sum_{s\in\mathcal{S}}\xi_i 1\{S_i=s\}\overline{m}_1(\tau,S_i) = \sum_{i\in[n]}\xi_i\overline{m}_1(\tau,S_i).
\end{aligned}$$

For the first term in $R_2^w(\tau)$, we have

$$\sup_{\tau\in\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\left(\frac{\pi(s)-\hat{\pi}^w(s)}{\hat{\pi}^w(s)\pi(s)}\right)\left(\sum_{i\in[n]}A_i\xi_i(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))1\{S_i=s\}\right)\right|$$
$$\leq \sum_{s\in\mathcal{S}}\left|\frac{D_n^w(s)}{n^w(s)\hat{\pi}^w(s)\pi(s)}\right|\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i\in[n]}\xi_iA_i1\{S_i=s\}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))\right|=o_p(1),$$

where the last equality holds due to Lemmas P.2 and P.4, and the fact that $\mathcal{F}=\{\xi(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s)):\tau\in\Upsilon\}$ is of the VC-type with fixed coefficients (α,v) and envelope ξ_iF_i such that $\mathbb{E}((\xi_iF_i)^q|S_i=s)<\infty$ for $q>2$.

For the second term in $R_{1,2}^w(\tau)$, recall $\overline{\Delta}_1(\tau,s,X_i)=\widehat{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s,X_i)$. Then

$$\left|\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}^w(s)}\sum_{i\in[n]}\xi_i(A_i-\hat{\pi}^w(s))\overline{\Delta}_1(\tau,s,X_i)1\{S_i=s\}\right|$$
$$=\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}n_0^w(s)\sup_{\tau\in\Upsilon}\left|\frac{\sum_{i\in I_1(s)}\xi_i\overline{\Delta}_1(\tau,s,X_i)}{n_1^w(s)}-\frac{\sum_{i\in I_0(s)}\xi_i\overline{\Delta}_1(\tau,s,X_i)}{n_0^w(s)}\right|=o_p(1),$$

where the last equality holds by Assumption 3. Therefore, we have

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

Combining (G.1) and (G.2), we have

$$\begin{aligned}L_{1,n}^w(\tau)=&\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_iA_i1\{S_i=s\}\left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)}+\left(1-\frac{1}{\pi(s)}\right)(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))\right]\\&+\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i(1-A_i)1\{S_i=s\}\left(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s)\right)\\&+\sum_{i=1}^n\xi_i\frac{m_1(\tau,S_i)}{\sqrt{n}}+R_{1,1}^w(\tau)-R_{1,2}^w(\tau),\end{aligned}$$

where $\sup_{\tau\in\Upsilon}(|R_{1,1}^w(\tau)|+|R_{1,2}^w(\tau)|)=o_p(1)$. In addition, Assumption 3 implies the classes of functions

$$\left\{\xi_i\left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)}+\left(1-\frac{1}{\pi(s)}\right)(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))\right]:\tau\in\Upsilon\right\}$$

and

$$\{\xi_i[\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s)]:\tau\in\Upsilon\}$$

are of the VC-type with fixed coefficients and envelopes belonging to $L_{\mathbb{P},q}$. In addition,

$$\mathbb{E}\left[\xi_i\left(\frac{\eta_{i,1}(s,\tau)}{\pi(s)}+\left(1-\frac{1}{\pi(s)}\right)(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))\right)\bigg|S_i=s\right]=0,$$

and

$$\mathbb{E}\left[\xi_i(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))|S_i=s\right]=0.$$

Therefore, Lemma P.2 implies,

$$\sup_{\tau\in\Upsilon}\left|\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n A_i 1\{S_i=s\}\left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)}+\left(1-\frac{1}{\pi(s)}\right)(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))\right]\right|=O_p(1),$$

and

$$\sup_{\tau\in\Upsilon}\left|\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-A_i)1\{S_i=s\}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))\right|=O_p(1).$$

This implies $\sup_{\tau\in\Upsilon}|L^w_{1,n}(\tau)|=O_p(1)$. Then by Kato (2009, Theorem 2) we have

$$\begin{aligned}
&\sqrt{n}(\hat{q}^w_1(\tau)-q_1(\tau))\\
&=\frac{1}{f_1(q_1(\tau))}\Bigg\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_iA_i1\{S_i=s\}\left[\frac{\eta_{i,1}(s,\tau)}{\pi(s)}+\left(1-\frac{1}{\pi(s)}\right)(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))\right]\\
&+\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i(1-A_i)1\{S_i=s\}(\overline{m}_1(\tau,s,X_i)-\overline{m}_1(\tau,s))+\sum_{i=1}^n\xi_i\frac{m_1(\tau,S_i)}{\sqrt{n}}\Bigg\}+R^w_{q,1}(\tau),
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R^w_{q,1}(\tau)|=o_p(1)$. Similarly, we have

$$\begin{aligned}
&\sqrt{n}(\hat{q}^w_0(\tau)-q_0(\tau))\\
&=\frac{1}{f_0(q_0(\tau))}\Bigg\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i(1-A_i)1\{S_i=s\}\left[\frac{\eta_{i,0}(s,\tau)}{1-\pi(s)}+\left(1-\frac{1}{1-\pi(s)}\right)(\overline{m}_0(\tau,s,X_i)-\overline{m}_0(\tau,s))\right]\\
&+\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_iA_i1\{S_i=s\}(\overline{m}_0(\tau,s,X_i)-\overline{m}_0(\tau,s))+\sum_{i=1}^n\xi_i\frac{m_0(\tau,S_i)}{\sqrt{n}}\Bigg\}+R_{q,0}(\tau),
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R^w_{q,0}(\tau)| = o_p(1)$. Taking the difference of the above two displays we obtain

$$\begin{aligned}
&\sqrt{n}(\hat{q}^w(\tau) - q(\tau)) \\
&= \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} \xi_i A_i 1\{S_i = s\} \\
&\times \left[\frac{\eta_{i,1}(s,\tau) - (1-\pi(s))\left(\overline{m}_1(\tau,s,X_i) - \overline{m}_1(\tau,s)\right)}{\pi(s) f_1(q_1(\tau))} - \frac{\left(\overline{m}_0(\tau,s,X_i) - \overline{m}_0(\tau,s)\right)}{f_0(q_0(\tau))}\right] \\
&- \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} \xi_i (1-A_i) 1\{S_i = s\} \\
&\times \left[\frac{\eta_{i,0}(s,\tau) - \pi(s)\left(\overline{m}_0(\tau,s,X_i) - \overline{m}_0(\tau,s)\right)}{(1-\pi(s)) f_0(q_0(\tau))} - \frac{\left(\overline{m}_1(\tau,s,X_i) - \overline{m}_1(\tau,s)\right)}{f_1(q_1(\tau))}\right] \\
&+ \frac{1}{\sqrt{n}} \sum_{i\in[n]} \xi_i \left(\frac{m_1(\tau,S_i)}{f_1(q_1(\tau))} - \frac{m_0(\tau,S_i)}{f_0(q_0(\tau))}\right) + R^w_q(\tau) \\
&= \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} \xi_i A_i 1\{S_i = s\} \phi_1(\tau,s,Y_i(1),X_i) - \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} \xi_i (1-A_i) 1\{S_i = s\} \phi_0(\tau,s,Y_i(0),X_i) \\
&+ \frac{1}{\sqrt{n}} \sum_{i\in[n]} \xi_i \phi_s(\tau,S_i) + R^w_q(\tau),
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R^w_q(\tau)| = o_p(1)$ and $(\phi_1(\cdot), \phi_0(\cdot), \phi_s(\cdot))$ are defined in Section F. Recalling the linear expansion of $\sqrt{n}(\hat{q}^{adj}(\tau) - q(\tau))$ established in Section F, we have

$$\begin{aligned}
&\sqrt{n}(\hat{q}^w(\tau) - \hat{q}^{adj}(\tau)) = \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1) A_i 1\{S_i = s\} \phi_1(\tau,s,Y_i(1),X_i) \\
&- \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1)(1-A_i) 1\{S_i = s\} \phi_0(\tau,s,Y_i(0),X_i) + \frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1) \phi_s(\tau,S_i) + R^d_q(\tau) \\
&= W^w_{n,1}(\tau) - W^w_{n,2}(\tau) + R^d_q(\tau),
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R^d_q(\tau)| = o_p(1)$,

$$\begin{aligned}
W^w_{n,1}(\tau) = &\sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1) A_i 1\{S_i = s\} \phi_1(\tau,s,Y_i(1),X_i) \\
&- \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1)(1-A_i) 1\{S_i = s\} \phi_0(\tau,s,Y_i(0),X_i),
\end{aligned}$$

and

$$W^w_{n,2}(\tau) = \frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1) \phi_s(\tau,S_i).$$

Lemma P.5 shows that, uniformly over $\tau \in \Upsilon$ and conditionally on data,

$$W^w_{n,1}(\tau) + W^w_{n,2}(\tau) \rightsquigarrow \mathcal{B}(\tau),$$

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

$$\Sigma(\tau, \tau') = \mathbb{E}\pi(S_i)\phi_1(\tau, S_i, Y_i(1), X_i)\phi_1(\tau', S_i, Y_i(1), X_i) + \mathbb{E}(1 - \pi(S_i))\phi_0(\tau, S_i, Y_i(0), X_i)\phi_0(\tau', S_i, Y_i(0), X_i) + \mathbb{E}\phi_s(\tau, S_i)\phi_s(\tau', S_i),$$

as defined in Theorem 3.1. This concludes the proof.

H Proof of Theorem 5.1

The proof is divided into two steps. In the first step, we show Assumption 5. Assumption 3(i) can be shown in the same manner and is omitted. In the second step, we establish Assumptions 3(ii) and 3(iii).

Step 1. Recall

$$\overline{\Delta}_a(\tau, s, X_i) = \hat{m}_a(\tau, s, X_i) - \overline{m}_a(\tau, s, X_i) = \Lambda_{a,s}(X_i, \theta_{a,s}(\tau)) - \Lambda_{a,s}(X_i, \hat{\theta}_{a,s}(\tau)),$$

and $\{X_i^s, \xi_i^s\}_{i \in [n]}$ is generated independently from the joint distribution of (X_i, ξ_i) given $S_i = s$, and so is independent of $\{A_i, S_i\}_{i \in [n]}$. Let $H_{a,s}(\theta_1, \theta_2) = \mathbb{E}[\Lambda_{a,s}(X_i, \theta_1) - \Lambda_{a,s}(X_i, \theta_2)|S_i = s] = \mathbb{E}[\Lambda_{a,s}(X_i^s, \theta_1) - \Lambda_{a,s}(X_i^s, \theta_2)]$. We have

$$\begin{aligned}
&\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_1(s)} \xi_i \overline{\Delta}_1(\tau, s, X_i)}{n_1^w(s)} - \frac{\sum_{i \in I_0(s)} \xi_i \overline{\Delta}_1(\tau, s, X_i)}{n_0^w(s)} \right| \\
&\leq \left(\max_{s \in \mathcal{S}} \frac{n_1(s)}{n_1^w(s)} \right) \sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_1(s)} \xi_i [\overline{\Delta}_1(\tau, s, X_i) - H_{1,s}(\theta_{a,s}(\tau), \hat{\theta}_{a,s}(\tau))]}{n_1(s)} \right| \\
&+ \left(\max_{s \in \mathcal{S}} \frac{n_0(s)}{n_0^w(s)} \right) \sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_0(s)} \xi_i [\overline{\Delta}_1(\tau, s, X_i) - H_{1,s}(\theta_{a,s}(\tau), \hat{\theta}_{a,s}(\tau))]}{n_0(s)} \right| \\
&= o_p(n^{-1/2}).
\end{aligned} \tag{H.1}$$

To see the last equality, we note that, for any $\varepsilon > 0$, with probability approaching one (w.p.a.1), we have

$$\sup_{\tau \in \Upsilon, s \in \mathcal{S}} ||\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau)|| \leq \varepsilon.$$

Therefore, on the event $\mathcal{A}_n(\varepsilon) \equiv \{\sup_{\tau\in\Upsilon,s\in\mathcal{S}} ||\hat{\theta}_{a,s}(\tau) - \theta_{a,s}(\tau)|| \leq \varepsilon, \min_{s\in\mathcal{S}} n_1(s) \geq \varepsilon n\}$ we have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i) - H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_1(s)}\right|\Bigg|\{A_i,S_i\}_{i\in[n]} \\
&\stackrel{d}{=} \sup_{\tau\in\Upsilon}\left|\frac{\sum_{i=N(s)+1}^{N(s)+n_1(s)}\xi_i^s[\overline{\Delta}_1(\tau,s,X_i^s) - H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_1(s)}\right|\Bigg|\{A_i,S_i\}_{i\in[n]} \\
&\leq ||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}}\Big|\{A_i,S_i\}_{i\in[n]},
\end{aligned}$$

where

$$\mathcal{F} = \{\xi_i^s[\Lambda_{a,s}(X_i^s,\theta_1) - \Lambda_{a,s}(X_i^s,\theta_2) - H_{1,s}(\theta_1,\theta_2)] : ||\theta_1 - \theta_2|| \leq \varepsilon\}.$$

By Assumption 6, $\mathcal{F}$ is a VC-class with a fixed VC index and envelope L_i. In addition,

$$\sup_{f\in\mathcal{F}} \mathbb{P}f^2 \leq \mathbb{E}L_i^2(\theta_1-\theta_2)^2 \leq C\varepsilon^2.$$

Therefore, for any $\delta > 0$ we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i) - H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_1(s)}\right| \geq \delta n^{-1/2}\right) \\
&\leq \mathbb{P}\left(\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i) - H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_1(s)}\right| \geq \delta n^{-1/2}, \mathcal{A}_n(\varepsilon)\right) + \mathbb{P}(\mathcal{A}_n^c(\varepsilon)) \\
&\leq \mathbb{E}\left[\mathbb{P}\left(\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i) - H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_1(s)}\right| \geq \delta n^{-1/2}, \mathcal{A}_n(\varepsilon)\Bigg|\{A_i,S_i\}_{i\in[n]}\right)\right] \\
&+ \mathbb{P}(\mathcal{A}_n^c(\varepsilon)) \\
&\leq \sum_{s\in\mathcal{S}}\mathbb{E}\left[\mathbb{P}\left(||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}} \geq \delta n^{-1/2}\Big|\{A_i,S_i\}_{i\in[n]}\right)1\{n_1(s)\geq n\varepsilon\}\right] + \mathbb{P}(\mathcal{A}_n^c(\varepsilon)) \\
&\leq \sum_{s\in\mathcal{S}}\mathbb{E}\left\{\frac{n^{1/2}\mathbb{E}\left[||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}}|\{A_i,S_i\}_{i\in[n]}\right]1\{n_1(s)\geq n\varepsilon\}}{\delta}\right\} + \mathbb{P}(\mathcal{A}_n^c(\varepsilon)).
\end{aligned}$$

By Chernozhukov, Chetverikov, and Kato (2014, Corollary 5.1),

$$\begin{aligned}
&n^{1/2}\mathbb{E}\left[||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}}|\{A_i,S_i\}_{i\in[n]}\right]1\{n_1(s)\geq n\varepsilon\} \\
&\leq C(\sqrt{\frac{n}{n_1(s)}\varepsilon^2} + n^{1/2}n_1^{1/q-1}(s))1\{n_1(s)\geq n\varepsilon\} \\
&\leq C(\varepsilon^{1/2} + n^{1/q-1/2}\varepsilon^{1/q-1}).
\end{aligned}$$

Therefore,

$$\mathbb{E}\left\{\frac{n^{1/2}\mathbb{E}\left[||\mathbb{P}_{n_1(s)}-\mathbb{P}||_{\mathcal{F}}|\{A_i,S_i\}_{i\in[n]}\right]1\{n_1(s)\geq n\varepsilon\}}{\delta}\right\}\leq C\mathbb{E}\left(\varepsilon^{1/2}+n^{1/q-1/2}\varepsilon^{1/q-1}\right)/\delta.$$

By letting $n\to\infty$ followed by $\varepsilon\to 0$, we have

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i)-H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_1(s)}\right|\geq\delta n^{-1/2}\right)=0,$$

In addition,

$$\max_{s\in\mathcal{S}}|n_1^w(s)/n_1(s)-1|=\max_{s\in\mathcal{S}}|(D_n^w(s)-D_n(s))/(\pi(s)n(s)+D_n(s))|\stackrel{p}{\longrightarrow}1,$$

as Lemma P.4 shows that $\max_{s\in\mathcal{S}}|(D_n^w(s)-D_n(s))/n(s)|=o_p(1)$.

Therefore,

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i)-H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_1^w(s)}\right|=o_p(n^{-1/2}).$$

For the same reason, we have

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_0(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i)-H_{1,s}(\theta_{a,s}(\tau),\hat{\theta}_{a,s}(\tau))]}{n_0^w(s)}\right|=o_p(n^{-1/2}),$$

and (H.1) holds.

Step 2. By Assumption 6,

$$\begin{aligned}&|\overline{m}_a(\tau_2,S_i,X_i)-\overline{m}_a(\tau_1,S_i,X_i)|\\&\leq|\tau_2-\tau_1|+|\Lambda_{a,s}(X_i,\theta_{a,s}(\tau_1))-\Lambda_{a,s}(X_i,\theta_{a,s}(\tau_2))|\\&\leq|\tau_2-\tau_1|+L_i|\theta_{a,s}(\tau_1)-\theta_{a,s}(\tau_2)|\leq(CL_i+1)|\tau_2-\tau_1|.\end{aligned}$$

This implies Assumption 3(iii). Furthermore, by Assumption 6 we can let the envelope for the class of functions $\mathcal{F}=\{\overline{m}_a(\tau_2,S_i,X_i):\tau\in\Upsilon\}$ be $F_i=\max(C,1)L_i+1$ where the constant C is the one in the above display. Then, we have

$$\sup_Q N(\mathcal{F},e_Q,\varepsilon||F||_{Q,2})\leq N(\Upsilon,d,\varepsilon)\leq 1/\varepsilon,$$

where $d(\tau_1,\tau_2)=|\tau_1-\tau_2|$. This verifies Assumption 3(ii).

I Proof of Theorem 5.2

Recall $\Sigma(\tau,\tau)$ is the asymptotic variance of $\hat{q}^{adj}(\tau)$, and $(\delta_a(\tau,S_i,X_i),\overline{\delta}_a(\tau,S_i,X_i))$ are defined in (F.3). Following the proof of the second part of Theorem 3.1, we have

$$
\begin{aligned}
&\Sigma(\tau,\tau)\\
&=\Bigg\{\mathbb{E}\left[\frac{(\tau-1\{Y_i(1)\le q_1(\tau)\}-m_1(\tau,S_i,X_i))^2}{\pi(S_i)f_1^2(q_1(0))}\right]+\mathbb{E}\left[\frac{(\tau-1\{Y_i(0)\le q_0(\tau)\}-m_0(\tau,S_i,X_i))^2}{(1-\pi(S_i))f_0^2(q_0(0))}\right]\\
&+\mathbb{E}\left(\frac{m_1(\tau,S_i,X_i)}{f_1(q_1(\tau))}-\frac{m_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right)^2\Bigg\}\\
&+\Bigg\{\mathbb{E}\left[\sqrt{\frac{1-\pi(S_i)}{\pi(S_i)}}\frac{\delta_1(\tau,S_i,X_i)-\overline{\delta}_1(\tau,S_i,X_i)}{f_1(q_1(\tau))}+\sqrt{\frac{\pi(S_i)}{1-\pi(S_i)}}\frac{\delta_0(\tau,S_i,X_i)-\overline{\delta}_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right]^2\Bigg\}\\
&=\Bigg\{\mathbb{E}\left[\frac{(\tau-1\{Y_i(1)\le q_1(\tau)\}-m_1(\tau,S_i,X_i))^2}{\pi(S_i)f_1^2(q_1(0))}\right]+\mathbb{E}\left[\frac{(\tau-1\{Y_i(0)\le q_0(\tau)\}-m_0(\tau,S_i,X_i))^2}{(1-\pi(S_i))f_0^2(q_0(0))}\right]\\
&+\mathbb{E}\left(\frac{m_1(\tau,S_i,X_i)}{f_1(q_1(\tau))}-\frac{m_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right)^2\Bigg\}\\
&+\sum_{s\in\mathcal{S}}p(s)\mathbb{E}\Bigg\{\Bigg[\sqrt{\frac{1-\pi(s)}{\pi(s)}}\frac{m_1(\tau,s,X_i)-m_1(\tau,s)-(\mathbb{E}(\Lambda_{1,s}(X_i,\theta_{1,s}(\tau))|S_i=s)-\Lambda_{1,s}(X_i,\theta_{1,s}(\tau)))}{f_1(q_1(\tau))}\\
&+\sqrt{\frac{\pi(s)}{1-\pi(s)}}\frac{m_0(\tau,s,X_i)-m_0(\tau,s)-(\mathbb{E}(\Lambda_{0,s}(X_i,\theta_{0,s}(\tau))|S_i=s)-\Lambda_{0,s}(X_i,\theta_{0,s}(\tau)))}{f_0(q_0(\tau))}\Bigg]^2\Bigg|S_i=s\Bigg\}.
\end{aligned}
$$

In addition, because the three terms in the first curly braces on the RHS of the above display do not depend on $\theta_{a,s}(\tau)$, to minimize $\Sigma(\tau,\tau)$ we only need to minimize the last term. Then, we have

$$
\begin{aligned}
(\theta_{1,s}(\tau),\theta_{0,s}(\tau))\in\ &\arg\min_{\theta_1,\theta_0}\mathbb{E}\Bigg\{\Bigg[\sqrt{\frac{1-\pi(s)}{\pi(s)}}\frac{m_1(\tau,s,X_i)-m_1(\tau,s)-g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))}\\
&+\sqrt{\frac{\pi(s)}{1-\pi(s)}}\frac{m_0(\tau,s,X_i)-m_0(\tau,s)-g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\Bigg]^2\Bigg|S_i=s\Bigg\}\\
\in\ &\arg\min_{\theta_1,\theta_0}\mathbb{E}\Bigg\{\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))}+\frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)^2\\
&-2\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))}+\frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\left(\frac{m_1(\tau,s,X_i)-m_1(\tau,s)}{f_1(q_1(\tau))}\right)\\
&-\frac{2\pi(s)}{1-\pi(s)}\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))}+\frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\left(\frac{m_0(\tau,s,X_i)-m_0(\tau,s)}{f_0(q_0(\tau))}\right)\Bigg|S_i=s\Bigg\},
\end{aligned}
$$

where $g_{a,s}(X_i,\theta_a)=\mathbb{E}(\Lambda_{a,s}(X_i,\theta_a)|S_i=s)-\Lambda_{a,s}(X_i,\theta_a)$.

Next we turn to the second part of Theorem 5.2. Following the previous proof, under the linear probability model with pseudo true values $(\theta_{1,s}(\tau_k), \theta_{0,s}(\tau_k))_{k\in[K]}$, we have

$$\Sigma^{LP}(\tau_k,\tau_l) = \Sigma_1^{LP}(\tau_k,\tau_l) + \sum_{s\in\mathcal{S}} p(s)\mathbb{E}\left[(\tilde{X}_{i,s}\beta_s(\tau_k) - \overline{y}_{i,s}(\tau_k))(\tilde{X}_{i,s}\beta_s(\tau_l) - \overline{y}_{i,s}(\tau_l))|S_i = s\right],$$

where

$$\begin{aligned}
\Sigma_1^{LP}(\tau_k,\tau_l) = &\left\{\mathbb{E}\left[\frac{(\tau - 1\{Y_i(1)\le q_1(\tau)\} - m_1(\tau,S_i,X_i))^2}{\pi(S_i)f_1^2(q_1(0))}\right]\right.\\
&+ \mathbb{E}\left[\frac{(\tau - 1\{Y_i(0)\le q_0(\tau)\} - m_0(\tau,S_i,X_i))^2}{(1-\pi(S_i))f_0^2(q_0(0))}\right]\\
&\left.+ \mathbb{E}\left(\frac{m_1(\tau,S_i,X_i)}{f_1(q_1(\tau))} - \frac{m_0(\tau,S_i,X_i)}{f_0(q_0(\tau))}\right)^2\right\},\\
\beta_s(\tau) = &\sqrt{\frac{1-\pi(s)}{\pi(s)}}\frac{\theta_{1,s}(\tau)}{f_1(q_1(\tau))} + \sqrt{\frac{\pi(s)}{1-\pi(s)}}\frac{\theta_{0,s}(\tau)}{f_0(q_0(\tau))}, \quad \text{and}\\
\overline{y}_{i,s}(\tau) = &\sqrt{\frac{1-\pi(s)}{\pi(s)}}\frac{[\mathbb{P}(Y_i(1)\le q_1(\tau)|X_i,S_i=s) - \mathbb{P}(Y_i(1)\le q_1(\tau)|S_i=s)]}{f_1(q_1(\tau))}\\
&+ \sqrt{\frac{\pi(s)}{1-\pi(s)}}\frac{[\mathbb{P}(Y_i(0)\le q_0(\tau)|X_i,S_i=s) - \mathbb{P}(Y_i(0)\le q_0(\tau)|S_i=s)]}{f_0(q_0(\tau))}.
\end{aligned}$$

To minimize $[\Sigma^{LP}(\tau_k,\tau_l)]_{k,l\in[K]}$ (in the matrix sense) is the same as minimizing

$$\left[\mathbb{E}\left[(\tilde{X}_{i,s}\beta_s(\tau_k) - \overline{y}_{i,s}(\tau_k))(\tilde{X}_{i,s}\beta_s(\tau_l) - \overline{y}_{i,s}(\tau_l))|S_i = s\right]\right]_{k,l\in[K]}$$

for each $s \in \mathcal{S}$, which is achieved if

$$\beta_s(\tau_k) = [\mathbb{E}\tilde{X}_{i,s}\tilde{X}_{i,s}'|S_i = s]^{-1}\mathbb{E}[\tilde{X}_{i,s}\overline{y}_{i,s}(\tau_k)|S_i = s]. \tag{I.1}$$

Because $\mathbb{E}[\tilde{X}_{i,s}\mathbb{P}(Y_i(a)\le q_a(\tau)|S_i = s)|S_i = s] = 0$ for $a = 0, 1$, (I.1) implies

$$\begin{aligned}
&\sqrt{\frac{1-\pi(s)}{\pi(s)}}\frac{\theta_{1,s}(\tau_k)}{f_1(q_1(\tau_k))} + \sqrt{\frac{\pi(s)}{1-\pi(s)}}\frac{\theta_{0,s}(\tau_k)}{f_0(q_0(\tau_k))}\\
&= \sqrt{\frac{1-\pi(s)}{\pi(s)}}\frac{\theta_{1,s}^{LP}(\tau_k)}{f_1(q_1(\tau_k))} + \sqrt{\frac{\pi(s)}{1-\pi(s)}}\frac{\theta_{0,s}^{LP}(\tau_k)}{f_0(q_0(\tau_k))},
\end{aligned}$$

or equivalently,

$$\frac{\theta_{1,s}(\tau_k)}{f_1(q_1(\tau_k))} + \frac{\pi(s)}{1-\pi(s)}\frac{\theta_{0,s}(\tau_k)}{f_0(q_0(\tau_k))} = \frac{\theta_{1,s}^{LP}(\tau_k)}{f_1(q_1(\tau_k))} + \frac{\pi(s)}{1-\pi(s)}\frac{\theta_{0,s}^{LP}(\tau_k)}{f_0(q_0(\tau_k))}.$$

This concludes the proof.

J Proof of Proposition 5.1

By Assumption 8, we see that $\sup_{\tau\in\Upsilon}|\partial_\tau\theta^{LP}_{a,s}(\tau)|<\infty$. This implies Assumption 6(ii). Next, we aim to show

$$\sup_{\tau\in\Upsilon,a=0,1,s\in\mathcal{S}}|\hat{\theta}^{LP}_{a,s}(\tau)-\theta^{LP}_{a,s}(\tau)|=O_p(n^{-1/2}).$$

Focusing on $\hat{\theta}^{LP}_{1,s}(\tau)$ we have

$$\hat{\theta}^{LP}_{1,s}(\tau)-\theta^{LP}_{1,s}(\tau)=\left[\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\hat{X}_{i,1,s}\hat{X}_{i,1,s}^\top\right]^{-1}\left[\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\hat{X}_{i,1,s}(1\{Y_i\le\hat{q}_1(\tau)\}-\hat{X}_{i,1,s}\theta^{LP}_{1,s}(\tau))\right]. \tag{J.1}$$

For the first term in (J.1), we have

$$\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\hat{X}_{i,1,s}\hat{X}_{i,1,s}^\top\stackrel{d}{=}\frac{1}{n_1(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\hat{X}^s_i\hat{X}^{s\prime}_i,$$

where $\hat{X}^s_i=X^s_i-\frac{1}{n_1(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)}X^s_i$ and X^s_i is as defined in Section E. As $\{X^s_i\}_{i=N(s)+1}^{N(s)+n_1(s)}$ is a sequence of i.i.d random variables that is independent of $N(s),n_1(s)$, we have

$$\frac{1}{n_1(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\hat{X}^s_i\hat{X}^{s\top}_i\stackrel{p}{\longrightarrow}\mathbb{E}(X^s_i-\mathbb{E}X^s_i)(X^s_i-\mathbb{E}X^s_i)^\top=\mathbb{E}(\tilde{X}_{i,s}\tilde{X}_{i,s}^\top|S_i=s),$$

where $\tilde{X}_{i,s}=X_i-\mathbb{E}(X_i|S_i=s)$. For the second term in (J.1), we have

$$\begin{aligned}
&\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\hat{X}_{i,1,s}\left(1\{Y_i\le\hat{q}_1(\tau)\}-\hat{X}_{i,1,s}\theta^{LP}_{1,s}(\tau)\right)\\
&=\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}\left(1\{Y_i\le\hat{q}_1(\tau)\}-\hat{X}_{i,1,s}\theta^{LP}_{1,s}(\tau)\right)+R_1(\tau)\\
&=\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}\left(1\{Y_i\le\hat{q}_1(\tau)\}-\tilde{X}_{i,s}\theta^{LP}_{1,s}(\tau)\right)+R_1(\tau)+R_2(\tau)\\
&=\left[\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}(1\{Y_i\le\hat{q}_1(\tau)\}-1\{Y_i\le q_1(\tau)\})\right]
\end{aligned}$$

$$+\left[\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}(1\{Y_i\le q_1(\tau)\}-\tilde{X}_{i,s}\theta_{1,s}^{LP}(\tau))\right]+R_1(\tau)+R_2(\tau)$$

$$\equiv I(\tau)+II(\tau)+R_1(\tau)+R_2(\tau),$$

where

$$R_1(\tau)=-\left(\frac{1}{n_1(s)}\sum_{i\in I_1(s)}X_i-\mathbb{E}(X_i|S_i=s)\right)\left(\frac{1}{n_1(s)}\sum_{i\in I_1(s)}(1\{Y_i\le\hat{q}_1(\tau)\}-\hat{X}_{i,1,s}\theta_{1,s}^{LP}(\tau))\right),$$

and

$$R_2(\tau)=\left(\frac{1}{n_1(s)}\sum_{i\in I_1(s)}X_i-\mathbb{E}(X_i|S_i=s)\right)\left(\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}\theta_{1,s}^{LP}(\tau)\right).$$

By Assumption 8 we can show that $\sup_{\tau\in\Upsilon}|\theta_{1,s}^{LP}(\tau)|\le C<\infty$ for some constant $C>0$. Therefore, we have

$$\sup_{\tau\in\Upsilon}|R_1(\tau)|=\sup_{\tau\in\Upsilon}\left|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}X_i-\mathbb{E}(X_i|S_i=s)\right|\left|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}(1\{Y_i\le\hat{q}_1(\tau)\}-\hat{X}_{i,1,s}\theta_{1,s}^{LP}(\tau))\right|$$
$$=O_p(n^{-1/2}),$$

and

$$\sup_{\tau\in\Upsilon}|R_2(\tau)|=\sup_{\tau\in\Upsilon}\left|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}X_i-\mathbb{E}(X_i|S_i=s)\right|\left|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}\theta_{1,s}^{LP}(\tau)\right|$$
$$=O_p(n^{-1/2}),$$

where we use the fact that

$$\left|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}X_i-\mathbb{E}(X_i|S_i=s)\right|\stackrel{d}{=}\left|\frac{1}{n_1(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)}(X_i^s-\mathbb{E}X_i^s)\right|=O_p(n^{-1/2}).$$

Next, note that $\sup_{\tau\in\Upsilon}|\hat{q}(\tau)-q(\tau)|=O_p(n^{-1/2})$, which means for any $\varepsilon>0$, there exists a constant $M>0$ such that $\sup_{\tau\in\Upsilon}|\hat{q}(\tau)-q(\tau)|\le Mn^{-1/2}$ with probability greater than $1-\varepsilon$. On the event set that $\sup_{\tau\in\Upsilon}|\hat{q}(\tau)-q(\tau)|\leq Mn^{-1/2}$, we have

$$
\begin{aligned}
&\sup_{\tau\in\Upsilon}|I(\tau)|\leq\sup_{\tau\in\Upsilon}\bigg|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}\bigg(1\{Y_i(1)\leq\hat{q}_1(\tau)\}-1\{Y_i(1)\leq q_1(\tau)\}\\
&-\mathbb{P}(Y_i(1)\leq\hat{q}_1(\tau)|X_i,S_i=s)+\mathbb{P}(Y_i(1)\leq q_1(\tau)|X_i,S_i=s)\bigg)\bigg|\\
&+\sup_{\tau\in\Upsilon}\bigg|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}(\mathbb{P}(Y_i(1)\leq\hat{q}_1(\tau)|X_i,S_i=s)-\mathbb{P}(Y_i(1)\leq q_1(\tau)|X_i,S_i=s))\bigg|\\
&\leq\sup_{|q-q'|\leq Mn^{-1/2}}\bigg|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}\bigg(1\{Y_i(1)\leq q\}-1\{Y_i(1)\leq q'\}\\
&-\mathbb{P}(Y_i(1)\leq q|X_i,S_i=s)+\mathbb{P}(Y_i(1)\leq q'|X_i,S_i=s)\bigg)\bigg|+C\sup_{\tau\in\Upsilon}|\hat{q}(\tau)-q(\tau)|\\
&\leq\sup_{|q-q'|\leq Mn^{-1/2}}\bigg|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\tilde{X}_{i,s}\bigg(1\{Y_i(1)\leq q\}-1\{Y_i(1)\leq q'\}\\
&-\mathbb{P}(Y_i(1)\leq q|X_i,S_i=s)+\mathbb{P}(Y_i(1)\leq q'|X_i,S_i=s)\bigg)\bigg|+Cn^{-1/2}\\
&=O_p(n^{-1/2}),
\end{aligned}
$$

where the first inequality is due to the triangle inequality, the second inequality is due to the fact that $\sup_{\tau\in\Upsilon}|\hat{q}(\tau)-q(\tau)|\leq Mn^{-1/2}$, and the third inequality is due to the fact that $f_1(\cdot|X_i,S_i=s)$ is assumed to be bounded. To see the last equality in the above display, we define

$$
\mathcal{F}=\left\{\begin{array}{c}(X_{i,j}-\mathbb{E}X_{i,j}|S_i=s)\bigg(1\{Y_i(1)\leq q\}-1\{Y_i(1)\leq q'\}\\ -\mathbb{P}(Y_i(1)\leq q|X_i,S_i=s)+\mathbb{P}(Y_i(1)\leq q'|X_i,S_i=s)\bigg):|q-q'|\leq Mn^{-1/2}\end{array}\right\}
$$

with envelope $F_i=2|X_{i,j}-\mathbb{E}(X_{i,j}|S_i=s)|\in L_{\mathbb{P},q}$ for some $q>2$, where $X_{i,j}$ is the j-th coordinate of X_i. Clearly $\mathcal{F}$ is of the VC-type with fixed coefficients (α,v). In addition,

$$
\sup_{f\in\mathcal{F}}\mathbb{P}f^2\leq Cn^{-1/2}\equiv\sigma_n^2.
$$

Therefore, Lemma P.2 implies that $\sup_{\tau\in\Upsilon}|I(\tau)|=O_p(n^{-1/2})$. By the usual maximal inequality (e.g. van der Vaart and Wellner, 1996, Theorem 2.14.1), we can show that

$$
\sup_{\tau\in\Upsilon}|II(\tau)|=O_p(n^{-1/2}).
$$

Combining these results, we conclude that

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}\hat{X}_{i,1,s}(1\{Y_i\leq\hat{q}_1(\tau)\}-\hat{X}_{i,1,s}\theta^{LP}_{1,s}(\tau))\right|=O_p(n^{-1/2}),$$

and hence

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}|\hat{\theta}^{LP}_{1,s}(\tau)-\theta^{LP}_{1,s}(\tau)|=O_p(n^{-1/2}).$$

K Proof of Proposition 5.2

Note $\hat{\theta}^{LG}_{a,s}(\tau)\in\Theta_a$, which is a compact subset of $\Re^{d_x}$. In addition, by Assumption 9

$$\sup_{\tau\in\Upsilon}||\theta^{LG}_{a,s}(\tau)||_2<\infty.$$

Therefore, there exists a constant $M>0$ such that

$$\sup_{\tau\in\Upsilon}||\hat{\theta}^{LG}_{a,s}(\tau)-\theta^{LG}_{a,s}(\tau)||_2\leq M.$$

Recall $Q(\tau,s,\theta_1,\theta_0)$ defined in Theorem 5.2 with $\Lambda_{a,s}(X_i,\theta_a)=\lambda(\vec{X}_i^\top\theta_a)$. For some $\varepsilon_0>0$, let $\mathcal{D}=\{(\delta_1,\delta_0)\in\Re^{2dx},\sqrt{||\delta_1||_2^2+||\delta_0||_2^2}\in[\varepsilon_0,M]\}$, which is compact and

$$\eta=\inf_{(\tau,\delta)\in\Upsilon\times\mathcal{D}}(Q(s,\tau,\theta^{LG}_{1,s}(\tau)+\delta_1,\theta^{LG}_{0,s}(\tau)+\delta_0)-Q(s,\tau,\theta^{LG}_{1,s}(\tau),\theta^{LG}_{0,s}(\tau))).$$

Because $Q(s,\tau,\theta^{LG}_{1,s}(\tau)+\delta_1,\theta^{LG}_{0,s}(\tau)+\delta_0)-Q(s,\tau,\theta^{LG}_{1,s}(\tau),\theta^{LG}_{0,s}(\tau))$ is a continuous function of (δ,τ), $\Upsilon\times\mathcal{D}$ is compact, and $(\theta^{LG}_{1,s}(\tau),\theta^{LG}_{0,s}(\tau))$ is the unique minimizer of $Q(s,\tau,\theta_1,\theta_0)$, we have $\eta>0$. On the other hand

$$\begin{aligned}&(Q(s,\tau,\theta^{LG}_{1,s}(\tau)+\delta_1,\theta^{LG}_{0,s}(\tau)+\delta_0)-Q(s,\tau,\theta^{LG}_{1,s}(\tau),\theta^{LG}_{0,s}(\tau)))\\&\leq(Q_n(s,\tau,\theta^{LG}_{1,s}(\tau)+\delta_1,\theta^{LG}_{0,s}(\tau)+\delta_0)-Q_n(s,\tau,\theta^{LG}_{1,s}(\tau),\theta^{LG}_{0,s}(\tau)))+2\Delta_n,\end{aligned}$$

where

$$\Delta_n=\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta}|Q_n(s,\tau,\theta_1,\theta_0)-Q(s,\tau,\theta_1,\theta_0)|.$$

If $\sup_{\tau\in\Upsilon}\sqrt{||\hat{\theta}_{1,s}^{LG}(\tau)-\theta_{1,s}^{LG}(\tau)||_2^2+||\hat{\theta}_{0,s}^{LG}(\tau)-\theta_{0,s}^{LG}(\tau)||_2^2}\geq\delta_0$, then there exists some $\tau\in\Upsilon$ and $\delta_1=\hat{\theta}_{1,s}^{LG}(\tau)-\theta_{1,s}^{LG}(\tau)$, and $\delta_0=\hat{\theta}_{0,s}^{LG}(\tau)-\theta_{0,s}^{LG}(\tau)$ such that

$$Q_n(s,\tau,\theta_{1,s}^{LG}(\tau)+\delta_1,\theta_{0,s}^{LG}(\tau)+\delta_0)-Q_n(s,\tau,\theta_{1,s}^{LG}(\tau),\theta_{1,s}^{LG}(\tau))\leq 0.$$

This implies

$$\mathbb{P}(\sup_{\tau\in\Upsilon}\sqrt{||\hat{\theta}_{1,s}^{LG}(\tau)-\theta_{1,s}^{LG}(\tau)||_2^2+||\hat{\theta}_{0,s}^{LG}(\tau)-\theta_{0,s}^{LG}(\tau)||_2^2}\geq\varepsilon_0)\leq\mathbb{P}(\eta\leq 2\Delta_n)\to 0,$$

where the last step holds because Lemma P.6 has established that $\Delta_n=o_p(1)$. As ε_0 is arbitrary, we have

$$\sup_{\tau\in\Upsilon}||\hat{\theta}_{1,s}^{LG}(\tau)-\theta_{1,s}^{LG}(\tau)||_2=o_p(1)\quad\text{and}\quad\sup_{\tau\in\Upsilon}||\hat{\theta}_{0,s}^{LG}(\tau)-\theta_{0,s}^{LG}(\tau)||_2=o_p(1).$$

L Proof of Proposition 5.3

Let

$$Q_n(\tau,s,q,\theta_a)=\frac{1}{n_a(s)}\sum_{i\in I_a(s)}[1\{Y_i\leq q\}\log(\lambda(H_i^\top\theta_a))+1\{Y_i>q\}\log(1-\lambda(H_i^\top\theta_a))],$$

and

$$Q(\tau,s,q,\theta_a)=\mathbb{E}[1\{Y_i(a)\leq q\}\log(\lambda(H_i^\top\theta_a))+1\{Y_i(a)>q\}\log(1-\lambda(H_i^\top\theta_a))|S_i=s].$$

Following the same argument in the proof of Lemma P.6 (replacing $\vec{X}_i$ by H_i), we can show

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S},q\in\Re,\theta_a\in\Re^{d_x}}|Q_n(\tau,s,q,\theta_a)-Q(\tau,s,q,\theta_a)|=o_p(1).$$

In addition,

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S},q\in\Re,\theta_a\in\Re^{d_x}}|\partial_q Q(\tau,s,q,\theta_a)|\leq C,$$

and $\sup_{\tau\in\Upsilon}|\hat{q}_a(\tau)-q_a(\tau)|=O_p(n^{-1/2})$. Therefore,

$$\Delta_n\equiv\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_a\in\Re^{d_x}}|Q_n(\tau,s,\hat{q}_a(\tau),\theta_a)-Q(\tau,s,q_a(\tau),\theta_a)|$$

$$\leq\sup_{\tau\in\Upsilon,s\in\mathcal{S},q\in\Re,\theta_a\in\Re^{d_x}}|Q_n(\tau,s,q,\theta_a)-Q(\tau,s,q,\theta_a)|$$

$$+\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_a\in\Re^{d_x}}|Q(\tau,s,\hat{q}_a(\tau),\theta_a)-Q(\tau,s,q_a(\tau),\theta_a)|=o_p(1). \quad \text{(L.1)}$$

In addition, note that $Q_n(\tau,s,\hat{q}_a(\tau),\theta_a)$ is concave in θ_a for fixed τ. Therefore, for $u\in S^{d_x-1}$ and $l>\delta$

$$Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau)+\delta u)\geq\frac{\delta}{l}Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau)+lu)+(1-\frac{\delta}{l})Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau)),$$

which implies

$$\begin{aligned}
&\frac{\delta}{l}\left(Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau)+lu)-Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau))\right)\\
&\leq Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau)+\delta u)-Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau))\\
&\leq Q(\tau,s,q_a(\tau),\theta^{ML}_{a,s}(\tau)+\delta u)-Q(\tau,s,q_a(\tau),\theta^{ML}_{a,s}(\tau))+2\Delta_n.
\end{aligned}$$

Because $Q(\tau,s,q_a(\tau),\theta^{ML}_{a,s}(\tau)+\delta u)-Q(\tau,s,q_a(\tau),\theta^{ML}_{a,s}(\tau))$ is continuous in $(\tau,u)\in\Upsilon\times S^{d_x-1}$, $\Upsilon\times S^{d_x-1}$ is compact, and $\theta^{ML}_{a,s}(\tau)$ is the unique maximizer of $Q(\tau,s,q_a(\tau),\theta_a)$, we have

$$\sup_{(\tau,u)\in\Upsilon\times S^{d_x-1}}Q(\tau,s,q_a(\tau),\theta^{ML}_{a,s}(\tau)+\delta u)-Q(\tau,s,q_a(\tau),\theta^{ML}_{a,s}(\tau))\leq-\eta,$$

for some $\eta>0$. In addition, if $\sup_{\tau\in\Upsilon}||\hat{\theta}^{ML}_{a,s}(\tau)-\theta^{ML}_{a,s}(\tau)||_2>\delta$, then there exists $(\tau,l,u)\in\Upsilon\times(\delta,\infty)\times S^{d_x-1}$ such that

$$\frac{\delta}{l}(Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau)+lu)-Q_n(\tau,s,\hat{q}_a(\tau),\theta^{ML}_{a,s}(\tau)))\geq 0.$$

Therefore,

$$\mathbb{P}\left(\sup_{\tau\in\Upsilon}||\hat{\theta}^{ML}_{a,s}(\tau)-\theta^{ML}_{a,s}(\tau)||_2>\delta\right)\leq\mathbb{P}(\eta\leq 2\Delta_n)\to 0,$$

where the last step is due to (L.1). This implies

$$\sup_{\tau\in\Upsilon}||\hat{\theta}^{ML}_{a,s}(\tau)-\theta^{ML}_{a,s}(\tau)||_2=o_p(1).$$

M Proof of Theorem 5.4

The proof strategy follows Belloni et al. (2017). We provide details here just for completeness. We divide the proof into three steps. In the first step, we show

$$\sup_{\tau \in \Upsilon} ||\hat{\theta}_{a,s}^{NP}(\tau) - \theta_{a,s}^{NP}(\tau)||_2 = O_p(\sqrt{h_n \log(n)/n}).$$

In the second step, we establish Assumption 5. By a similar argument, we can establish Assumption 3(i). In the third step, we establish Assumptions 3(ii) and 3(iii).

Step 1. Let $\hat{U}_\tau = \hat{\theta}_{a,s}^{NP}(\tau) - \theta_{a,s}^{NP}(\tau)$,

$$\begin{aligned} Q_n(\tau, s, q, \theta) &= \frac{-1}{n_a(s)} \sum_{i \in I_a(s)} [1\{Y_i \leq q\} \log(\lambda(H_{h_n}^\top(X_i)\theta_a)) + 1\{Y_i > q\} \log(1 - \lambda(H_{h_n}^\top(X_i)\theta_a))] \\ &= \frac{1}{n_a(s)} \sum_{i \in I_a(s)} [\log\Big(1 + \exp(H_{h_n}^\top(X_i)\theta_a)\Big) - 1\{Y_i \leq q\} H_{h_n}^\top(X_i)\theta_a], \end{aligned}$$

and for an arbitrary $U_\tau \in \Re^{h_n}$,

$$\ell_i(t) = \log(1 + \exp(H_{h_n}^\top(X_i)(\theta_{a,s}^{NP}(\tau) + tU_\tau))).$$

Then, we have

$$\hat{U}_\tau = \arg\max_{U_\tau} Q_n(\tau, s, \hat{q}_a(\tau), \theta_{a,s}^{NP}(\tau) + U_\tau) - Q_n(\tau, s, \hat{q}_a(\tau), \theta_{a,s}^{NP}(\tau)),$$

$$\partial_t Q_n(\tau, s, \hat{q}_a(\tau), \theta_{a,s}^{NP}(\tau) + tU_\tau)|_{t=0} = \frac{-1}{n_a(s)} \sum_{i \in I_a(s)} \Big(1\{Y_i \leq \hat{q}_a(\tau)\} - \lambda(H_{h_n}^\top(X_i)\theta_{a,s}^{NP}(\tau))\Big) H_{h_n}(X_i),$$

and

$$\begin{aligned} & Q_n(\tau, s, \hat{q}_a(\tau), \theta_{a,s}^{NP}(\tau) + U_\tau) - Q_n(\tau, s, \hat{q}_a(\tau), \theta_{a,s}^{NP}(\tau)) - \partial_t Q_n^\top(\tau, s, \hat{q}_a(\tau), \theta_{a,s}^{NP}(\tau) + tU_\tau)|_{t=0} U_\tau \\ =& \frac{1}{n_a(s)} \sum_{i \in I_a(s)} \left[\ell_i(1) - \ell_i(0) - \ell_i'(0)\right]. \end{aligned}$$

In addition

$$|\ell_i^{'''}(t)| \leq |\ell_i^{''}(t)||H_{h_n}^\top(X_i)U_\tau|.$$

Therefore, there exists a constant $\underline{c} > 0$ such that

$$
\begin{aligned}
&\ell_i(1) - \ell_i(0) - \ell_i'(0) \\
&\geq \frac{\ell_i''(0)}{(H_{h_n}^\top(X_i)U_\tau)^2}\left[\exp(-|H_{h_n}^\top(X_i)U_\tau|) + |H_{h_n}^\top(X_i)U_\tau| - 1\right] \\
&= \lambda(H_{h_n}^\top(X_i)\theta_{a,s}^{NP}(\tau))(1-\Lambda(H_{h_n}^\top(X_i)\theta_{a,s}^{NP}(\tau)))\left[\exp(-|H_{h_n}^\top(X_i)U_\tau|) + |H_{h_n}^\top(X_i)U_\tau| - 1\right] \\
&\geq \underline{c}\left[\exp(-|H_{h_n}^\top(X_i)U_\tau|) + |H_{h_n}^\top(X_i)U_\tau| - 1\right] \\
&\geq \underline{c}\left(\frac{(H_{h_n}^\top(X_i)U_\tau)^2}{2} - \frac{|H_{h_n}^\top(X_i)U_\tau|^3}{6}\right),
\end{aligned}
$$

where the first inequality is due to Bach (2010, Lemma 1) and the third inequality holds because

$$
e^{-x} + x - 1 \geq \frac{x^2}{2} - \frac{x^3}{6},\ x > 0.
$$

To see the second inequality, note that $e^{-x} + x - 1 \geq 0$ for $x \geq 0$ and by Assumption 11,

$$
\begin{aligned}
&\inf_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in\text{Supp}(X)} \lambda(H_{h_n}^\top(x)\theta_{a,s}^{NP}(\tau)) \\
&= \inf_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in\text{Supp}(X)} (\mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i = x) - R_a(\tau,s,x)) \geq c/2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in\text{Supp}(X)} \lambda(H_{h_n}^\top(x)\theta_{a,s}^{NP}(\tau)) \\
&= \sup_{a=0,1,s\in\mathcal{S},\tau\in\Upsilon,x\in\text{Supp}(X)} (\mathbb{P}(Y_i(a) \leq q_a(\tau)|S_i = s, X_i = x) + R_a(\tau,s,x)) \leq 1 - c/2.
\end{aligned}
$$

This implies

$$
\inf_{\tau\in\Upsilon} \lambda(H_{h_n}^\top(X_i)\theta_{a,s}^{NP}(\tau))(1-\Lambda(H_{h_n}^\top(X_i)\theta_{a,s}^{NP}(\tau))) \geq \underline{c} > 0,
$$

and thus,

$$
\begin{aligned}
G_n(U_\tau) &\equiv Q_n(\tau,s,\hat{q}_a(\tau),\theta_{a,s}^{NP}(\tau)+U_\tau) - Q_n(\tau,s,\hat{q}_a(\tau),\theta_{a,s}^{NP}(\tau)) - \partial_t Q_n^\top(\tau,s,\hat{q}_a(\tau),\theta_{a,s}^{NP}(\tau)+tU_\tau)|_{t=0}U_\tau \\
&\geq \frac{\underline{c}}{n_a(s)}\sum_{i\in I_a(s)}\left(\frac{(H_{h_n}^\top(X_i)U_\tau)^2}{2} - \frac{|H_{h_n}^\top(X_i)U_\tau|^3}{6}\right).
\end{aligned}
$$

Let

$$\overline{\ell} = \inf_{U \in \Re^{h_n}} \frac{\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)U)^2\right]^{3/2}}{\frac{1}{n_a(s)} \sum_{i \in I_a(s)} |H_{h_n}^\top(X_i)U|^3}. \tag{M.1}$$

If $\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2} \leq \overline{\ell}$, then

$$\begin{aligned}
&\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2 \\
&= \left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{-1/2} \frac{\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{3/2}}{\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^3} \frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^3 \\
&\geq \frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^3,
\end{aligned}$$

and thus

$$G_n(\hat{U}_\tau) \geq \frac{\underline{c}}{n_a(s)} \sum_{i \in I_a(s)} \left(\frac{(H_{h_n}^\top(X_i)U_\tau)^2}{2} - \frac{|H_{h_n}^\top(X_i)U_\tau|^3}{6} \right) \geq \frac{\underline{c}}{n_a(s)} \sum_{i \in I_a(s)} \frac{(H_{h_n}^\top(X_i)\hat{U}_\tau)^2}{3}.$$

On the other hand, if $\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2} > \overline{\ell}$, we can denote $\overline{U}_\tau = \frac{\overline{\ell}\hat{U}_\tau}{\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2}}$ such that

$$\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\overline{U}_\tau)^2\right]^{1/2} \leq \overline{\ell}.$$

Further, because $G_n(U_\tau)$ is convex in U_τ we have

$$\begin{aligned}
G_n(\hat{U}_\tau) &= G_n\left(\frac{\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2}}{\overline{\ell}} \overline{U}_\tau \right) \\
&\geq \frac{\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2}}{\overline{\ell}} G_n(\overline{U}_\tau) \\
&\geq \frac{\left[\frac{1}{n_a(s)} \sum_{i \in I_a(s)} (H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2}}{\overline{\ell}} \frac{\underline{c}}{n_a(s)} \sum_{i \in I_a(s)} \frac{(H_{h_n}^\top(X_i)\overline{U}_\tau)^2}{3}
\end{aligned}$$

$$= \frac{\underline{c}\overline{\ell}}{3}\left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)}(H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2}.$$

Therefore, for some constant $\overline{c}$ that only depends on $\underline{c}$ and κ_1, we have

$$\begin{aligned}
G_n(\hat{U}_\tau) &\geq \min\left(\frac{\underline{c}}{n_a(s)}\sum_{i\in I_a(s)}\frac{(H_{h_n}^\top(X_i)\hat{U}_\tau)^2}{3}, \frac{\underline{c}\overline{\ell}}{3}\left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)}(H_{h_n}^\top(X_i)\hat{U}_\tau)^2\right]^{1/2}\right) \\
&\geq \frac{\overline{c}}{3}\min(||\hat{U}_\tau||_2^2, \overline{\ell}||\hat{U}_\tau||_2). \qquad \text{(M.2)}
\end{aligned}$$

In addition, by construction,

$$\begin{aligned}
G_n(\hat{U}_\tau) &\leq |\partial_t Q_n^\top(\tau, s, \hat{q}_a(\tau), \theta_{a,s}^{NP}(\tau) + tU_\tau)|_{t=0}\hat{U}_\tau| \\
&= \left|\frac{1}{n_a(s)}\sum_{i\in I_a(s)}(1\{Y_i \leq \hat{q}_a(\tau)\} - \lambda(H_{h_n}^\top(X_i)\theta_{a,s}^{NP}(\tau)))H_{h_n}^\top(X_i)\hat{U}_\tau\right| \\
&\leq \left\|\frac{1}{n_a(s)}\sum_{i\in I_a(s)}(1\{Y_i \leq \hat{q}_a(\tau)\} - m_a(\tau, s, X_i))H_{h_n}(X_i)\right\|_\infty \left\|\hat{U}_\tau\right\|_1 \\
&+ \left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)}R_a^2(\tau, s, X_i)\right]^{1/2}\left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)}(H_{h_n}^\top(X_i)\hat{U}_\tau)\right]^{1/2} \\
&\leq \left\|\frac{h_n^{1/2}}{n_a(s)}\sum_{i\in I_a(s)}(1\{Y_i \leq \hat{q}_a(\tau)\} - m_a(\tau, s, X_i))H_{h_n}(X_i)\right\|_\infty \left\|\hat{U}_\tau\right\|_2 \\
&+ \left[\frac{\kappa_2}{n_a(s)}\sum_{i\in I_a(s)}R_a^2(\tau, s, X_i)\right]^{1/2}||\hat{U}_\tau||_2. \qquad \text{(M.3)}
\end{aligned}$$

Combining (M.2) and (M.3), we have

$$\frac{\overline{c}}{3}\min(||\hat{U}_\tau||_2, \overline{\ell}) \leq \left\|\frac{h_n^{1/2}}{n_a(s)}\sum_{i\in I_a(s)}(1\{Y_i \leq \hat{q}_a(\tau)\} - m_a(\tau, s, X_i))H_{h_n}(X_i)\right\|_\infty + \left[\frac{\kappa_2}{n_a(s)}\sum_{i\in I_a(s)}R_a^2(\tau, s, X_i)\right]^{1/2}.$$

Taking $\sup_{\tau\in\Upsilon}$ on both sides, we have

$$\begin{aligned}
&\frac{\overline{c}}{3}\min(\sup_{\tau\in\Upsilon}||\hat{U}_\tau||_2, \overline{\ell}) \\
&\leq \sup_{\tau\in\Upsilon}\left\|\frac{h_n^{1/2}}{n_a(s)}\sum_{i\in I_a(s)}(1\{Y_i \leq \hat{q}_a(\tau)\} - m_a(\tau, s, X_i))H_{h_n}(X_i)\right\|_\infty + \sup_{\tau\in\Upsilon}\left[\frac{\kappa_2}{n_a(s)}\sum_{i\in I_a(s)}R_a^2(\tau, s, X_i)\right]^{1/2}
\end{aligned}$$

$$= O_p(\sqrt{\frac{h_n \log(n)}{n}}),$$

where the last line holds due to Assumption 11 and Lemma P.7. Finally, Lemma P.8 shows that $\overline{\ell}/\sqrt{\frac{h_n \log(n)}{n}} \to \infty$, which implies

$$\sup_{\tau \in \Upsilon} ||\hat{U}_\tau||_2 = O_p\left(\sqrt{\frac{h_n \log(n)}{n}}\right).$$

Step 2. Recall

$$\begin{aligned}
\overline{\Delta}_a(\tau, s, X_i) &= \hat{m}_a(\tau, s, X_i) - \overline{m}_a(\tau, s, X_i) \\
&= \mathbb{P}(Y_i(a) \leq q_a(\tau)|X_i, S_i = s) - \lambda(H_{h_n}^\top(X_i)\hat{\theta}_{a,s}^{NP}(\tau)) \\
&= \lambda(H_{h_n}^\top(X_i)\theta_{a,s}^{NP}(\tau)) - \lambda(H_{h_n}^\top(X_i)\hat{\theta}_{a,s}^{NP}(\tau)) + R_a(\tau, s, X_i),
\end{aligned}$$

and $\{X_i^s, \xi_i^s\}_{i \in [n]}$ is generated independently from the joint distribution of (X_i, ξ_i) given $S_i = s$, and so is independent of $\{A_i, S_i\}_{i \in [n]}$. Let

$$H(\theta_1, \theta_2) = \mathbb{E}[\lambda(H_{h_n}^\top(X_i)\theta_1) - \lambda(H_{h_n}^\top(X_i)\theta_2)|S_i = s] = \mathbb{E}[\lambda(H_{h_n}^\top(X_i^s)\theta_1) - \lambda(H_{h_n}^\top(X_i^s)\theta_2)].$$

We have

$$\begin{aligned}
&\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_1(s)} \xi_i \overline{\Delta}_1(\tau, s, X_i)}{n_1^w(s)} - \frac{\sum_{i \in I_0(s)} \xi_i \overline{\Delta}_1(\tau, s, X_i)}{n_0^w(s)} \right| \\
&\leq \sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_1(s)} \xi_i [\overline{\Delta}_1(\tau, s, X_i) - H(\theta_{1,s}^{NP}(\tau), \hat{\theta}_{1,s}^{NP}(\tau)) - \mathbb{E}(R_1(\tau, s, X_i)|S_i = s)]}{n_1^w(s)} \right| \\
&+ \sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_0(s)} \xi_i [\overline{\Delta}_1(\tau, s, X_i) - H(\theta_{1,s}^{NP}(\tau), \hat{\theta}_{1,s}^{NP}(\tau)) - \mathbb{E}(R_1(\tau, s, X_i)|S_i = s)]}{n_0^w(s)} \right| \qquad \text{(M.4)}
\end{aligned}$$

We aim to bound the first term on the RHS of (M.4). Note for any $\varepsilon > 0$, there exists a constant $M > 0$ such that

$$\mathbb{P}\left(\sup_{\tau \in \Upsilon} ||\hat{\theta}_{a,s}^{NP}(\tau) - \theta_{a,s}^{NP}(\tau)||_2 \leq M\sqrt{h_n \log(n)/n}\right) \geq 1 - \varepsilon.$$

On the set $\mathcal{A}(\varepsilon) = \{\sup_{\tau \in \Upsilon} ||\hat{\theta}_{a,s}^{NP}(\tau) - \theta_{a,s}^{NP}(\tau)||_2 \leq M\sqrt{h_n \log(n)/n}\}$, we have

$$\sup_{\tau \in \Upsilon, s \in \mathcal{S}} \left| \frac{\sum_{i \in I_1(s)} \xi_i [\overline{\Delta}_1(\tau, s, X_i) - H(\theta_{1,s}^{NP}(\tau), \hat{\theta}_{1,s}^{NP}(\tau)) - \mathbb{E}(R_1(\tau, s, X_i)|S_i = s)]}{n_1^w(s)} \right|$$

$$
\begin{aligned}
&\leq \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{\sum_{i\in I_1(s)} \xi_i [\lambda(H_{h_n}^\top(X_i)\theta_{1,s}^{NP}(\tau)) - \lambda(H_{h_n}^\top(X_i)\hat{\theta}_{1,s}^{NP}(\tau)) - H(\theta_{1,s}^{NP}(\tau), \hat{\theta}_{1,s}^{NP}(\tau))]}{n_1(s)} \right| \\
&+ \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{\sum_{i\in I_1(s)} \xi_i \left[R_1(\tau,s,X_i) - \mathbb{E}(R_1(\tau,s,X_i)|S_i=s)\right]}{n_1^w(s)} \right| \\
&\leq \frac{n_1(s)}{n_1^w(s)} \Bigg[\sup_{s\in\mathcal{S}, \theta_1,\theta_2\in\Re^{h_n}, ||\theta_1-\theta_2||_2 \leq M\sqrt{h_n \log(n)/n}} \left| \frac{\sum_{i\in I_1(s)} \xi_i [\lambda(H_{h_n}^\top(X_i)\theta_1) - \lambda(H_{h_n}^\top(X_i)\theta_2) - H(\theta_1,\theta_2)]}{n_1(s)} \right| \\
&+ \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{\sum_{i\in I_1(s)} \xi_i \left[R_1(\tau,s,X_i) - \mathbb{E}(R_1(\tau,s,X_i)|S_i=s)\right]}{n_1(s)} \right| \Bigg] \\
&\equiv \frac{n_1(s)}{n_1^w(s)}(D_1 + D_2).
\end{aligned}
$$

For D_1, we have

$$
\begin{aligned}
& D_1 | \{A_i, S_i\}_{i\in[n]} \\
& \stackrel{d}{=} \sup \left| \frac{\sum_{i=N(s)}^{N(s)+n_1(s)} \xi_i^s [\lambda(H_{h_n}^\top(X_i^s)\theta_1) - \lambda(H_{h_n}^\top(X_i^s)\theta_2) - H(\theta_1,\theta_2)]}{n_1(s)} \right| |\{A_i, S_i\}_{i\in[n]} \\
& \stackrel{d}{=} ||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}} | \{A_i, S_i\}_{i\in[n]},
\end{aligned}
$$

where the supremum in the first equality is taken over $\{s \in \mathcal{S}, \theta_1, \theta_2 \in \Re^{h_n}, ||\theta_1 - \theta_2||_2 \leq M\sqrt{h_n \log(n)/n}\}$ and

$$
\mathcal{F} = \left\{ \begin{array}{c} \xi_i^s [\lambda(H_{h_n}^\top(X_i^s)\theta_1) - \lambda(H_{h_n}^\top(X_i^s)\theta_2) - H(\theta_1,\theta_2)] : \\ s \in \mathcal{S}, \theta_1, \theta_2 \in \Re^{h_n}, ||\theta_1 - \theta_2||_2 \leq M\sqrt{h_n \log(n)/n} \end{array} \right\}
$$

with the envelope $F = 2\xi_i^s$. We further note that $|| \max_{i\in[n]} 2\xi_i^s||_{\mathbb{P},2} \leq C\log(n)$,

$$
\sup_{f\in\mathcal{F}} \mathbb{E} f^2 \leq \sup \mathbb{E}(H_{h_n}^\top(X_i^s)(\theta_1 - \theta_2))^2 \leq \kappa_2 M^2 h_n \log(n)/n,
$$

and

$$
\sup_Q N(\mathcal{F}, e_Q, \varepsilon ||F||_{Q,2}) \leq \left(\frac{a}{\varepsilon}\right)^{ch_n},
$$

where a, c are two fixed constants. Therefore, by Chernozhukov et al. (2014, Corollary 5.1),

$$
\mathbb{E}\left[||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}} | \{A_i, S_i\}_{i\in[n]}\right] = O_p\left(h_n \log(n)/n + \frac{h_n \log^2(n)}{n}\right) = o_p(n^{-1/2}),
$$

which implies $D_1 = o_p(n^{-1/2})$.

Similarly, we have

$$
\begin{aligned}
D_2|\{A_i,S_i\}_{i\in[n]} \stackrel{d}{=} & \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{\sum_{i=N(s)}^{N(s)+n_1(s)} \xi_i^s \left[R_1(\tau,s,X_i^s) - \mathbb{E}(R_1(\tau,s,X_i^s))\right]}{n_1(s)} \right| |\{A_i,S_i\}_{i\in[n]} \\
= & ||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}}|\{A_i,S_i\}_{i\in[n]}|\{A_i,S_i\}_{i\in[n]},
\end{aligned}
$$

where $\mathcal{F} = \{\xi_i^s[\tau - m_1(\tau,s,X_i^s) - \lambda(H_{h_n}^\top(X_i^s)\theta_{1,s}^{NP}(\tau))] : \tau\in\Upsilon\}$ with an envelope $F = \xi_i^s$. In addition, we note $\mathcal{F}$ is nested in

$$
\widetilde{\mathcal{F}} = \{\xi_i^s[\tau - m_1(\tau,s,X_i^s) - \lambda(H_{h_n}^\top(X_i^s)\theta_1)] : \tau\in\Upsilon, \theta_1\in\Re^{h_n}\},
$$

so that

$$
\sup_Q N(\mathcal{F}, e_Q, \varepsilon||F||_{Q,2}) \le \sup_Q N(\widetilde{\mathcal{F}}, e_Q, \varepsilon||F||_{Q,2}) \le \left(\frac{a}{\varepsilon}\right)^{ch_n}.
$$

Last,

$$
\sup_{f\in\mathcal{F}} \mathbb{E}f^2 = \sup_{\tau\in\Upsilon, a=0,1, s\in\mathcal{S}} \mathbb{E}R_1^2(\tau,s,X_i^s) = O(h_n\log(n)/n).
$$

by Chernozhukov et al. (2014, Corollary 5.1),

$$
\mathbb{E}\left[||\mathbb{P}_{n_1(s)} - \mathbb{P}||_{\mathcal{F}}|\{A_i,S_i\}_{i\in[n]}\right] = O_p\left(h_n\log(n)/n + \frac{h_n\log^2(n)}{n}\right) = o_p(n^{-1/2}),
$$

which implies $D_2 = o_p(n^{-1/2})$. This leads to (M.4).

Step 3. Note $|m_a(\tau_1,s,X_i)| \le 1$ and

$$
\begin{aligned}
& |m_a(\tau_1,s,X_i) - m_a(\tau_2,s,X_i)| \\
\le & |\tau_1-\tau_2| + |\mathbb{P}(Y_i(a)\le q_a(\tau_1)|X_i,S_i=s) - \mathbb{P}(Y_i(a)\le q_a(\tau_2)|X_i,S_i=s)| \\
\le & \left(1 + \frac{\sup_y f_a(y|X_i,S_i=s)}{\inf_{\tau\in\Upsilon} f_a(q_a(\tau))}\right)|\tau_1-\tau_2| \\
\le & C|\tau_1-\tau_2|.
\end{aligned}
$$

This implies Assumptions 3(ii) and 3(iii).

N Proof of Theorem 5.5

We focus on the case with $a = 1$. Note

$$\{X_i, Y_i(1)\}_{i\in I_1(s)}|\{A_i, S_i\}_{i\in[n]} \stackrel{d}{=} \{X_i^s, Y_i^s(1)\}_{i=N(s)+1}^{N(s)+n_1(s)}|\{A_i, S_i\}_{i\in[n]}.$$

where $\{X_i^s, Y_i(1)^s\}_{i\in[n]}$ is an i.i.d. sequence that is independent of $\{A_i, S_i\}_{i\in[n]}$. Therefore,

$$\begin{aligned}\hat{\theta}_{1,s}^{HD}(q)|\{A_i, S_i\}_{i\in[n]} \stackrel{d}{=} \arg\min_{\theta_a} \frac{-1}{n_1(s)} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \Big[1\{Y_i^s(1) \le q\} \log(\lambda(H_{p_n}(X_i)^\top \theta_a)) \\ + 1\{Y_i^s(1) > q\} \log(1 - \lambda(H_{p_n}^\top(X_i^s)\theta_a)) \Big] + \frac{\varrho_{n,1}(s)}{n_1(s)} ||\hat{\Omega}\theta_a||_1 \Big| \{A_i, S_i\}_{i\in[n]},\end{aligned}$$

and Assumption 12(vi) implies

$$\begin{aligned}0 < \kappa_1 \le &\inf_{a=0,1,s\in\mathcal{S},|v|_0\le h_n\ell_n} \frac{v^T\left(\frac{1}{n_1(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)} H_{p_n}(X_i^s)H_{p_n}^\top(X_i^s)\right)v}{||v||_2^2} \\ \le &\sup_{a=0,1,s\in\mathcal{S},|v|_0\le h_n\ell_n} \frac{v^T\left(\frac{1}{n_1(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)} H_{p_n}(X_i^s)H_{p_n}^\top(X_i^s)\right)v}{||v||_2^2} \le \kappa_2 < \infty,\end{aligned}$$

and

$$\begin{aligned}0 < \kappa_1 \le &\inf_{a=0,1,s\in\mathcal{S},|v|_0\le h_n\ell_n} \frac{v^T\mathbb{E}(H_{p_n}(X_i^s)H_{p_n}^\top(X_i^s))v}{||v||_2^2} \\ \le &\sup_{a=0,1,s\in\mathcal{S},|v|_0\le h_n\ell_n} \frac{v^T\mathbb{E}(H_{p_n}(X_i^s)H_{p_n}^\top(X_i^s))v}{||v||_2^2} \le \kappa_2 < \infty.\end{aligned}$$

In addition, we have $n_1(s)/n \stackrel{a.s.}{\to} \pi(s)p(s) > 0$. Therefore, based on the results established by Belloni et al. (2017), we have, conditionally on $\{A_i, S_i\}_{i\in[n]}$, and thus, unconditionally,

$$\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S}} ||\hat{\theta}_{a,s}^{HD}(q) - \theta_{a,s}^{HD}(q)||_2 = O_p\left(\sqrt{\frac{h_n\log(p_n)}{n}}\right),$$

$$\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S}} ||\hat{\theta}_{a,s}^{post}(q) - \theta_{a,s}^{HD}(q)||_2 = O_p\left(\sqrt{\frac{h_n\log(p_n)}{n}}\right),$$

$$\sup_{a=0,1,q\in\mathcal{Q}_a^\varepsilon,s\in\mathcal{S}} ||\hat{\theta}_{a,s}^{HD}(q)|| = O_p(h_n),$$

and

$$\sup_{a=0,1,q\in\mathcal{Q}_a^{\varepsilon},s\in\mathcal{S}}||\hat{\theta}_{a,s}^{post}(q)|| = O_p(h_n).$$

In the following, we prove the results when $\hat{\theta}_{a,s}^{HD}(q)$ is used. The results corresponding to $\hat{\theta}_{a,s}^{post}(q)$ can be proved in the same manner and are therefore omitted. Recall

$$\begin{aligned}
&\overline{\Delta}_1(\tau, s, X_i) = \hat{m}_1(\tau, s, X_i) - \overline{m}_1(\tau, s, X_i) \\
&= \mathbb{P}(Y_i(1) \leq q_1(\tau)|X_i, S_i = s) - \lambda(H_{p_n}(X_i)^\top \hat{\theta}_{1,s}^{HD}(\hat{q}_1(\tau))) \\
&= [\mathcal{M}_1(q_1(\tau), s, X_i) - \mathcal{M}_1(\hat{q}_1(\tau), s, X_i) + r_a(\hat{q}_1(\tau), s, X_i)] \\
&+ \left[\lambda(H_{p_n}(X_i)^\top \theta_{1,s}^{HD}(\hat{q}_1(\tau))) - \lambda(H_{p_n}(X_i)^\top \hat{\theta}_{1,s}^{HD}(\hat{q}_1(\tau)))\right] \\
&\equiv \mathcal{R}_{a,s}(q_1(\tau), \hat{q}_1(\tau), X_i) + \lambda(H_{p_n}(X_i)^\top \theta_{1,s}^{HD}(\hat{q}_1(\tau))) - \lambda(H_{p_n}(X_i)^\top \hat{\theta}_{1,s}^{HD}(\hat{q}_1(\tau))),
\end{aligned}$$

where

$$\mathcal{R}_{a,s}(q, q', X_i) = \mathcal{M}_1(q, s, X_i) - \mathcal{M}_1(q', s, X_i) + r_a(q', s, X_i).$$

Let

$$H_\lambda(\theta_1, \theta_2, s) = \mathbb{E}[\lambda(H_{p_n}(X_i)^\top \theta_1) - \lambda(H_{p_n}(X_i)^\top \theta_2)|S_i = s],$$

and

$$H_R(q, q', s) = \mathbb{E}(\mathcal{R}_{a,s}(q, q', X_i)|S_i = s).$$

Then, we have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i\overline{\Delta}_1(\tau, s, X_i)}{n_1^w(s)} - \frac{\sum_{i\in I_0(s)}\xi_i\overline{\Delta}_1(\tau, s, X_i)}{n_0^w(s)}\right| \\
&\leq \sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau, s, X_i) - H_\lambda(\hat{\theta}_{1,s}^{HD}(\hat{q}_1(\tau)), \theta_{1,s}^{HD}(\hat{q}_1(\tau)), s) - H_R(q_1(\tau), \hat{q}_1(\tau), s)]}{n_1^w(s)}\right| \\
&+ \sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_0(s)}\xi_i[\overline{\Delta}_1(\tau, s, X_i) - H_\lambda(\hat{\theta}_{1,s}^{HD}(\hat{q}_1(\tau)), \theta_{1,s}^{HD}(\hat{q}_1(\tau)), s) - H_R(q_1(\tau), \hat{q}_1(\tau), s)]}{n_0^w(s)}\right| \quad \text{(N.1)}
\end{aligned}$$

We aim to bound the first term on the RHS of (N.1). Note for any $\varepsilon > 0$, there exists a constant

$M > 0$ such that

$$\mathbb{P}\left(\begin{matrix}\sup_{q\in\mathcal{Q}_1^\varepsilon}||\hat{\theta}_{1,s}^{HD}(q)-\theta_{1,s}^{HD}(q)||_2 \leq M\sqrt{\frac{h_n\log(p_n)}{n}}, \ \sup_{q\in\mathcal{Q}_1^\varepsilon}||\hat{\theta}_{1,s}^{HD}(q)||_0 \leq Mh_n, \\ \sup_{\tau\in\Upsilon}|\hat{q}_1(\tau)-q_1(\tau)| \leq Mn^{-1/2}\end{matrix}\right) \geq 1-\varepsilon.$$

On the set

$$\mathcal{A}(\varepsilon) = \left\{\begin{matrix}\sup_{q\in\mathcal{Q}_1^\varepsilon}||\hat{\theta}_{1,s}^{HD}(q)-\theta_{1,s}^{HD}(q)||_2 \leq M\sqrt{\frac{h_n\log(p_n)}{n}}, \ \sup_{q\in\mathcal{Q}_1^\varepsilon}||\hat{\theta}_{1,s}^{HD}(q)||_0 \leq Mh_n, \\ \sup_{\tau\in\Upsilon}|\hat{q}_1(\tau)-q_1(\tau)| \leq Mn^{-1/2}\end{matrix}\right\},$$

we have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\overline{\Delta}_1(\tau,s,X_i)-H_\lambda(\hat{\theta}_{1,s}^{HD}(\hat{q}_1(\tau)),\theta_{1,s}^{HD}(\hat{q}_1(\tau)),s)-H_R(q_1(\tau),\hat{q}_1(\tau),s)]}{n_1^w(s)}\right| \\
&\leq \sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i[\lambda(H_{h_n}^\top(X_i)\theta_{1,s}^{NP}(\tau))-\lambda(H_{h_n}^\top(X_i)\hat{\theta}_{1,s}^{NP}(\tau))-H_\lambda(\theta_{1,s}^{NP}(\tau),\hat{\theta}_{1,s}^{NP}(\tau),s)]}{n_1(s)}\right| \\
&+ \sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i\left[\mathcal{R}_{1,s}(q_1(\tau),\hat{q}_1(\tau),X_i)-\mathbb{E}(\mathcal{R}_{1,s}(q_1(\tau),\hat{q}_1(\tau),X_i)|S_i=s)\right]}{n_1^w(s)}\right| \\
&\leq \frac{n_1(s)}{n_1^w(s)}\left[\sup\left|\frac{\sum_{i\in I_1(s)}\xi_i[\lambda(H_{h_n}^\top(X_i)\theta_1)-\lambda(H_{h_n}^\top(X_i)\theta_2)-H_\lambda(\theta_1,\theta_2,s)]}{n_1(s)}\right|\right. \\
&+ \left.\sup_{q,q'\in\mathcal{Q}_1^\varepsilon,|q-q'|\leq Mn^{-1/2},s\in\mathcal{S}}\left|\frac{\sum_{i\in I_1(s)}\xi_i\left[\mathcal{R}_{1,s}(q,q',X_i)-\mathbb{E}(\mathcal{R}_{1,s}(q,q',X_i)|S_i=s)\right]}{n_1(s)}\right|\right] \\
&\equiv \frac{n_1(s)}{n_1^w(s)}(D_1+D_2),
\end{aligned}$$

where the first supremum in the second inequality is taken over $\{s\in\mathcal{S},\theta_1,\theta_2\in\Re^{h_n},||\theta_1-\theta_2||_2 \leq M\sqrt{h_n\log(n)/n},||\theta_1||_0+||\theta_2||_0 \leq Mh_n\}$. Denote

$$\mathcal{F} = \left\{\begin{matrix}\xi_i^s[\lambda(H_{h_n}^\top(X_i^s)\theta_1)-\lambda(H_{h_n}^\top(X_i^s)\theta_2)-H_\lambda(\theta_1,\theta_2,s)]: \\ s\in\mathcal{S},\theta_1,\theta_2\in\Re^{h_n},||\theta_1-\theta_2||_2\leq M\sqrt{h_n\log(n)/n},||\theta_1||_0+||\theta_2||_0\leq Mh_n\end{matrix}\right\}$$

with the envelope $F = 2\xi_i^s$. We further note that $||\max_{i\in[n]}2\xi_i^s||_{\mathbb{P},2} \leq C\log(n)$,

$$\sup_{f\in\mathcal{F}}\mathbb{E}f^2 \leq \sup\mathbb{E}(H_{h_n}^\top(X_i^s)(\theta_1-\theta_2))^2 \leq \kappa_2M^2h_n\log(p_n)/n,$$

and

$$\sup_Q N(\mathcal{F},e_Q,\varepsilon||F||_{Q,2}) \leq \left(\frac{ap_n}{\varepsilon}\right)^{ch_n},$$

where a, c are two fixed constants. Therefore, Lemma P.2 implies

$$D_1 = O_p\left(\frac{h_n \log(p_n)}{n} + \frac{h_n \log(n)\log(p_n)}{n}\right) = o_p(n^{-1/2}).$$

Similarly, denote

$$\mathcal{F} = \left\{\xi_i^s\left[\mathcal{M}_1(q, s, X_i) - \mathcal{M}_1(q', s, X_i) + r_a(q', s, X_i)\right] : q, q' \in \mathcal{Q}_1^{\varepsilon}, |q - q'| \le Mn^{-1/2}, s \in \mathcal{S},\right\}$$

with an envelope $F = \xi_i^s$. In addition, note that $\mathcal{F}$ is nested in

$$\widetilde{\mathcal{F}} = \{\xi_i^s[\mathcal{M}_1(q, s, X_i^s) - \lambda(H_{h_n}^{\top}(X_i^s)\theta_1)] : q \in \mathcal{Q}_1^{\varepsilon}, \theta_1 \in \Re^{p_n}, ||\theta_1||_0 \le h_n\},$$

with the same envelope. Hence,

$$\sup_Q N(\mathcal{F}, e_Q, \varepsilon ||F||_{Q,2}) \le \sup_Q N(\widetilde{\mathcal{F}}, e_Q, \varepsilon ||F||_{Q,2}) \le \left(\frac{ap_n}{\varepsilon}\right)^{ch_n}.$$

Last,

$$\sup_{f \in \mathcal{F}} \mathbb{E}f^2 \le C \sup_{q, q' \in \mathcal{Q}_1^{\varepsilon}, |q - q'| \le Mn^{-1/2}, s \in \mathcal{S}} (|q - q'|^2 + \mathbb{E}r_a^2(q', s, X_i^s)) = O(h_n \log(p_n)/n).$$

Therefore, Lemma P.2 implies

$$D_2 = O_p\left(\frac{h_n \log(p_n)}{n} + \frac{h_n \log(n)\log(p_n)}{n}\right) = o_p(n^{-1/2}).$$

This leads to (N.1). We can establish Assumption 3(i) in the same manner. Assumptions 3(ii) and 3(iii) can be established by the same argument used in Step 3 of the proof of Theorem 5.4. This concludes the proof of Theorem 5.5.

O Proof of Proposition D.1

We divide the proof into two steps. In the first step, we show $d_H(\hat{\Theta}_s^{\varepsilon_n}(\tau), \Theta_s(\tau) \cap \Theta) = o_p(1)$ where $d_H(\cdot, \cdot)$ denotes the Hausdorff distance. In the second step, we show $\hat{\theta}_{a,s}^*(\tau) \xrightarrow{p} \theta_{a,s}^*(\tau)$.

Step 1. For some $\delta_0 > 0$, let

$$\eta = \inf_{(\theta_1, \theta_0) \in \Theta, d_H((\theta_1, \theta_0), \Theta_s(\tau) \cap \Theta) \ge \delta_0} (Q(s, \tau, \theta_1, \theta_0) - \inf_{(\theta_1, \theta_0) \in \Theta_s(\tau) \cap \Theta} Q(s, \tau, \theta_1, \theta_0)).$$

Because for a fixed τ, $\{(\theta_1, \theta_0) \in \Theta, d_H((\theta_1, \theta_0), \Theta_s(\tau)) \ge \delta_0\}$ is compact and $Q(s, \tau, \theta_1, \theta_0)$ is continuous in (θ_1, θ_0), we have $\eta > 0$. On the other hand, for any $(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta$,

$$\begin{aligned}&(Q(s, \tau, \theta_1, \theta_0) - Q(s, \tau, \theta_{1,s}(\tau), \theta_{1,s}(\tau)))\\&\leq (Q_n(s, \tau, \theta_1, \theta_0) - Q_n(s, \tau, \theta_{1,s}(\tau), \theta_{1,s}(\tau))) + 2\Delta_n,\end{aligned}$$

where

$$\Delta_n = \sup_{\tau \in \Upsilon, s \in \mathcal{S}, \theta_1, \theta_0 \in \Theta} |Q_n(s, \tau, \theta_1, \theta_0) - Q(s, \tau, \theta_1, \theta_0)|.$$

Taking $\inf_{(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta}$ on both sides, we have

$$\begin{aligned}&(Q(s, \tau, \theta_1, \theta_0) - \inf_{(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta} Q(s, \tau, \theta_{1,s}(\tau), \theta_{1,s}(\tau)))\\&\leq (Q_n(s, \tau, \theta_1, \theta_0) - \inf_{(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta} Q(s, \tau, \theta_{1,s}(\tau), \theta_{1,s}(\tau))) + 2\Delta_n,\end{aligned}$$

Suppose there exist $(\hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau)) \in \hat{\Theta}_s^{\varepsilon_n}(\tau)$ such that $d_H((\hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau), \Theta_s(\tau) \cap \Theta) \geq \delta_0$. Then, because $\Delta_n = o_p(\varepsilon_n)$ as shown in Lemma P.6, we have

$$Q_n(s, \tau, \hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau)) - \inf_{(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta} Q_n(s, \tau, \theta_{1,s}(\tau), \theta_{1,s}(\tau)) \leq \varepsilon_n,$$

and

$$\begin{aligned}\eta &\leq (Q(s, \tau, \hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau)) - \inf_{(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta} Q(s, \tau, \theta_{1,s}(\tau), \theta_{1,s}(\tau)))\\&\leq Q_n(s, \tau, \hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau)) - \inf_{(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta} Q_n(s, \tau, \theta_{1,s}(\tau), \theta_{1,s}(\tau)) + 2\Delta_n \leq 2\Delta_n + \varepsilon_n.\end{aligned}$$

Therefore,

$$\mathbb{P}\left(\exists (\hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau)) \in \hat{\Theta}_s^{\varepsilon_n}(\tau), d_H((\hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau), \Theta_s(\tau) \cap \Theta) \geq \delta_0)\right) \to 0,$$

or equivalently,

$$\sup_{(\hat{\theta}_{1,s}(\tau), \hat{\theta}_{0,s}(\tau)) \in \hat{\Theta}_s^{\varepsilon_n}(\tau)} \inf_{(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta} \sqrt{||\hat{\theta}_{1,s}(\tau) - \theta_{1,s}(\tau)||_2^2 + ||\hat{\theta}_{0,s}(\tau) - \theta_{0,s}(\tau)||_2^2} \xrightarrow{p} 0.$$

Next, note that for any $(\theta_{1,s}(\tau), \theta_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta$ and $(\theta'_{1,s}(\tau), \theta'_{0,s}(\tau)) \in \Theta$, we have

$$Q_n(s, \tau, \theta_{1,s}(\tau), \theta_{0,s}(\tau)) - Q_n(s, \tau, \theta'_{1,s}(\tau), \theta'_{0,s}(\tau))$$

$$\leq Q(s,\tau,\theta_{1,s}(\tau),\theta_{0,s}(\tau)) - Q(s,\tau,\theta'_{1,s}(\tau),\theta'_{0,s}(\tau)) + 2\Delta_n \leq 2\Delta_n = o_p(\varepsilon_n),$$

where the last equality is due to the second part of Lemma P.6. Therefore, by the definition of $\hat{\Theta}_s^{\varepsilon_n}(\tau)$,

$$\mathbb{P}(\Theta_s(\tau) \cap \Theta \subset \hat{\Theta}_s^{\varepsilon_n}(\tau)) \to 1,$$

which implies

$$\sup_{(\theta_{1,s}(\tau),\theta_{0,s}(\tau))\in\Theta_s(\tau)\cap\Theta} \inf_{(\hat{\theta}_{1,s}(\tau),\hat{\theta}_{0,s}(\tau))\in\hat{\Theta}_s^{\varepsilon_n}(\tau)} \sqrt{||\hat{\theta}_{1,s}(\tau) - \theta_{1,s}(\tau)||_2^2 + ||\hat{\theta}_{0,s}(\tau) - \theta_{0,s}(\tau)||_2^2} \xrightarrow{p} 0.$$

This concludes the first step of the proof.

Step 2. Because $\Theta_s(\tau) \cap \Theta \subset \hat{\Theta}_s^{\varepsilon_n}(\tau)$ with probability approaching one, we have

$$\sqrt{||\hat{\theta}^*_{1,s}(\tau)||_2^2 + ||\hat{\theta}^*_{0,s}(\tau)||_2^2} \leq \sqrt{||\theta^*_{1,s}(\tau)||_2^2 + ||\theta^*_{0,s}(\tau)||_2^2} \;\; w.p.a.1.$$

On the other hand, for $\hat{\theta}^*_{a,s}(\tau)$, we can find $(\tilde{\theta}^*_{1,s}(\tau), \tilde{\theta}^*_{0,s}(\tau)) \in \Theta_s(\tau) \cap \Theta$ such that

$$\sqrt{||\hat{\theta}^*_{1,s}(\tau) - \tilde{\theta}^*_{1,s}(\tau)||_2^2 + ||\hat{\theta}^*_{0,s}(\tau) - \tilde{\theta}^*_{0,s}(\tau)||_2^2} \leq d_H(\hat{\Theta}_s^{\varepsilon_n}(\tau), \Theta_s(\tau) \cap \Theta) = o_p(1), \tag{O.1}$$

and thus,

$$\begin{aligned}\sqrt{||\hat{\theta}^*_{1,s}(\tau)||_2^2 + ||\hat{\theta}^*_{0,s}(\tau)||_2^2} \geq & \sqrt{||\tilde{\theta}^*_{1,s}(\tau)||_2^2 + ||\tilde{\theta}^*_{0,s}(\tau)||_2^2} - d_H(\hat{\Theta}_s^{\varepsilon_n}(\tau), \Theta_s(\tau) \cap \Theta) \\ = & \sqrt{||\theta^*_{1,s}(\tau)||_2^2 + ||\theta^*_{0,s}(\tau)||_2^2} - o_p(1).\end{aligned}$$

Therefore,

$$\sqrt{||\hat{\theta}^*_{1,s}(\tau)||_2^2 + ||\hat{\theta}^*_{0,s}(\tau)||_2^2} - \sqrt{||\theta^*_{1,s}(\tau)||_2^2 + ||\theta^*_{0,s}(\tau)||_2^2} = o_p(1),$$

and

$$\sqrt{||\tilde{\theta}^*_{1,s}(\tau)||_2^2 + ||\tilde{\theta}^*_{0,s}(\tau)||_2^2} \xrightarrow{p} \sqrt{||\theta^*_{1,s}(\tau)||_2^2 + ||\theta^*_{0,s}(\tau)||_2^2}. \tag{O.2}$$

Last, note that $\Theta_s(\tau) \cap \Theta$ is compact because $Q(s,\tau,\theta_1,\theta_2)$ is continuous in (θ_1,θ_2). We also note that the Euclidean distance $d(\theta_1,\theta_0) = \sqrt{||\theta_1||_2^2 + ||\theta_0||_2^2}$ is a continuous function and

$(\theta^*_{1,s}(\tau), \theta^*_{0,s}(\tau))$ is uniquely defined over $\Theta_s(\tau) \cap \Theta$ by Assumption 13. Then, for any $\delta_0 > 0$,

$$\eta = \inf_{(\theta_1,\theta_0) \in \Theta_s(\tau)\cap\Theta, d(\theta_1 - \theta^*_{1,s}(\tau), \theta_0 - \theta^*_{0,s}(\tau)) \geq \delta_0} d(\theta_1, \theta_0) - d(\theta^*_{1,s}(\tau), \theta^*_{0,s}(\tau)) > 0,$$

and

$$\mathbb{P}(d(\tilde{\theta}^*_{1,s}(\tau) - \theta^*_{1,s}(\tau), \tilde{\theta}^*_{0,s}(\tau) - \theta^*_{0,s}(\tau)) \geq \delta_0) \leq \mathbb{P}\left(d(\tilde{\theta}^*_{1,s}(\tau), \tilde{\theta}^*_{0,s}(\tau)) - d(\theta^*_{1,s}(\tau), \theta^*_{0,s}(\tau)) \geq \eta\right) \to 0,$$

where the last step is due to (O.2). Therefore, we have

$$(\tilde{\theta}^*_{1,s}(\tau), \tilde{\theta}^*_{0,s}(\tau)) \xrightarrow{p} \theta^*_{1,s}(\tau), \theta^*_{0,s}(\tau),$$

which, along with (O.1), further implies

$$(\hat{\theta}^*_{1,s}(\tau), \hat{\theta}^*_{0,s}(\tau)) \xrightarrow{p} \theta^*_{1,s}(\tau), \theta^*_{0,s}(\tau)$$

P Technical Lemmas

The first lemma was established in Zhang and Zheng (2020).

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

$$\mathbb{P}(\max_{1\leq k\leq n} ||S_k|| \geq \varepsilon) \leq 3 \max_{1\leq k\leq n} \mathbb{P}(||S_k|| \geq \varepsilon/3) \leq 9\mathbb{P}(||S_n|| \geq \varepsilon/30).$$

Proof. The first inequality is due to Zhang and Zheng (2020, Lemma E.1) and the second inequality is due to Montgomery-Smith (1993, Theorem 1). □

The next lemma is due to Chernozhukov et al. (2014) with our modification of their maximal inequality to the case with covariate-adaptive randomization.

Lemma P.2. *Suppose Assumption 1 holds. Let $w_i = 1$ or ξ_i defined in Assumption 4. Denote $\mathcal{F}$ as a class of functions of the form $f(x, y_1, y_0)$ where, $f(x, y_1, y_0)$ is a measurable function and $\mathbb{E}(f(X_i, Y_i(1), Y_i(0))|S_i = s) = 0$. Further suppose* $\max_{s\in\mathcal{S}} \mathbb{E}(|F_i|^q|S_i = s) < \infty$ *for some $q \geq 2$, where*

$$F_i = \sup_{f\in\mathcal{F}} |w_i f(X_i, Y_i(1), Y_i(0))|,$$

$\mathcal{F}$ is of the VC-type with coefficients $(\alpha_n, v_n) > 0$, and $\sup_{f\in\mathcal{F}} \mathbb{E}(f^2|S=s) \leq \sigma_n^2$. Then,

$$\sup_{f\in\mathcal{F}, s\in\mathcal{S}} \frac{1}{\sqrt{n}} \left| \sum_{i\in[n]} A_i 1\{S_i = s\} w_i f(X_i, Y_i(1), Y_i(0)) \right|$$
$$= O_p\left(\sqrt{v_n \sigma_n^2 \log\left(\frac{\alpha_n ||F||_{\mathbb{P},2}}{\sigma}\right)} + \frac{v_n || \max_{i\in[n]} F_i||_{\mathbb{P},2} \log\left(\frac{\alpha_n ||F||_{\mathbb{P},2}}{\sigma}\right)}{\sqrt{n}} \right),$$

and

$$\sup_{f\in\mathcal{F}, s\in\mathcal{S}} \frac{1}{\sqrt{n}} \left| \sum_{i\in[n]} (1-A_i) 1\{S_i = s\} w_i f(X_i, Y_i(1), Y_i(0)) \right|$$
$$= O_p\left(\sqrt{v_n \sigma_n^2 \log\left(\frac{\alpha_n ||F||_{\mathbb{P},2}}{\sigma}\right)} + \frac{v_n || \max_{i\in[n]} F_i||_{\mathbb{P},2} \log\left(\frac{\alpha_n ||F||_{\mathbb{P},2}}{\sigma}\right)}{\sqrt{n}} \right).$$

Proof. We focus on establishing the first statement. The proof of the second statement is similar and is omitted. Following Bugni et al. (2018), we define the sequence of i.i.d. random variables $\{(w_i^s, X_i^s, Y_i^s(1), Y_i^s(0)) : 1 \leq i \leq n\}$ with marginal distributions equal to the distribution of $(w_i, X_i, Y_i(1), Y_i(0))|S_i = s$. The distribution of $\sum_{i\in[n]} A_i 1\{S_i = s\} w_i f(X_i, Y_i(1), Y_i(0))$ 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 each stratum, i.e.,

$$\sum_{i\in[n]} A_i 1\{S_i = s\} w_i f(X_i, Y_i(1), Y_i(0)) \stackrel{d}{=} \sum_{i=N(s)+1}^{N(s)+n_1(s)} w_i^s f(X_i^s, Y_i^s(1), Y_i^s(0))$$
$$\equiv \Gamma_n^s(N(s) + n_1(s), f) - \Gamma_n^s(N(s) + 1, f),$$

where $N(s) = \sum_{i\in[n]} 1\{S_i < s\}$ and

$$\Gamma_n^s(k, f) = \sum_{i\in[k]} w_i^s f(X_i^s, Y_i^s(1), Y_i^s(0)).$$

Let $\mu_n = \sqrt{v_n \sigma_n^2 \log\left(\frac{\alpha_n ||F||_{\mathbb{P},2}}{\sigma}\right)} + \frac{v_n || \max_{i\in[n]} F_i||_{\mathbb{P},2} \log\left(\frac{\alpha_n ||F||_{\mathbb{P},2}}{\sigma}\right)}{\sqrt{n}}$. Then, for some constant $C > 0$, we have

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}, s\in\mathcal{S}} \frac{1}{\sqrt{n}} \left| \sum_{i\in[n]} A_i 1\{S_i = s\} w_i f(X_i, Y_i(1), Y_i(0)) \right| \geq t\mu_n \right)$$

$$
\begin{aligned}
&= \mathbb{P}\left(\sup_{f\in\mathcal{F}, s\in\mathcal{S}} \frac{1}{\sqrt{n}} \left|\Gamma_n^s(N(s)+n_1(s), f) - \Gamma_n^s(N(s)+1, f)\right| \geq t\mu_n\right) \\
&\leq \sum_{s\in\mathcal{S}} \mathbb{P}\left(\max_{1\leq k\leq n} \sup_{f\in\mathcal{F}} \left|\frac{1}{\sqrt{n}}\Gamma_n^s(k,f)\right| \geq t\mu_n/2\right) \\
&\leq \sum_{s\in\mathcal{S}} 9\ \mathbb{P}\left(\sup_{f\in\mathcal{F}} \left|\frac{1}{\sqrt{n}}\Gamma_n^s(n,f)\right| \geq t\mu_n/60\right) \\
&\leq \sum_{s\in\mathcal{S}} \frac{540\ \mathbb{E}\left(\sup_{f\in\mathcal{F}} \left|\frac{1}{\sqrt{n}}\Gamma_n^s(n,f)\right|\right)}{t\mu_n} \\
&= \sum_{s\in\mathcal{S}} \frac{540\mathbb{E}\left(\sqrt{n}||\mathbb{P}_n^s - \mathbb{P}^s||_{\mathcal{F}}\right)}{t\mu_n} \\
&\leq C/t,
\end{aligned}
$$

where $\mathbb{P}_n^s$ and $\mathbb{P}^s$ are the empirical process and expectation w.r.t. i.i.d. data $\{w_i^s, X_i^s, Y_i^s(1), Y_i^s(0)\}_{i\in[n]}$, respectively, the second inequality is due to Lemma P.1, the last equality is due to the fact that

$$
\mathbb{E}w_i^s f(X_i^s, Y_i^s(1), Y_i^s(0)) = \mathbb{E}\left(w_i f(X_i, Y_i(1), Y_i(0)) | S_i = s\right) = 0,
$$

and the last inequality is due to the fact that, by Chernozhukov et al. (2014, Corollary 5.1),

$$
\mathbb{E}\left(\sqrt{n}||\mathbb{P}_n^s - \mathbb{P}^s||_{\mathcal{F}}\right) \leq C\mu_n.
$$

Then, for any $\varepsilon > 0$, we can choose $t \geq C/\varepsilon$ so that

$$
\mathbb{P}\left(\sup_{f\in\mathcal{F}, s\in\mathcal{S}} \frac{1}{\sqrt{n}} \left|\sum_{i\in[n]} A_i 1\{S_i = s\} w_i f(X_i, Y_i(1), Y_i(0))\right| \geq t\mu_n\right) \leq \varepsilon,
$$

which implies the desired result. □

Lemma P.3. *Suppose Assumptions in Theorem 3.1 hold. Denote*

$$
\begin{aligned}
W_{n,1}(\tau) = &\sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} A_i 1\{S_i = s\}\phi_1(\tau, s, Y_i(1), X_i) \\
&- \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i\in[n]} (1 - A_i) 1\{S_i = s\}\phi_0(\tau, s, Y_i(0), X_i),
\end{aligned}
$$

and

$$W_{n,2}(\tau) = \frac{1}{\sqrt{n}} \sum_{i \in [n]} \phi_s(\tau, S_i).$$

Then, uniformly over $\tau \in \Upsilon$*,*

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

where $(\mathcal{B}_1(\tau), \mathcal{B}_2(\tau))$ *are two independent Gaussian processes with covariance kernels* $\Sigma_1(\tau, \tau')$ *and* $\Sigma_2(\tau, \tau')$*, respectively, such that*

$$\Sigma_1(\tau, \tau') = \mathbb{E}\pi(S_i)\phi_1(\tau, S_i, Y_i(1), X_i)\phi_1(\tau', S_i, Y_i(1), X_i)$$
$$+ \mathbb{E}(1 - \pi(S_i))\phi_0(\tau, S_i, Y_i(0), X_i)\phi_0(\tau', S_i, Y_i(0), X_i)$$

and

$$\Sigma_2(\tau, \tau') = \mathbb{E}\phi_s(\tau, S_i)\phi_s(\tau', S_i).$$

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

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

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

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

In the second step, we show that

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

uniformly over $\tau \in \Upsilon$.

Step 1. Recall that we define $\{(X_i^s, Y_i^s(1), Y_i^s(0)) : 1 \leq i \leq n\}$ as a sequence of i.i.d. random variables with marginal distributions equal to the distribution of $(X_i, Y_i(1), Y_i(0))|S_i = s$ and $N(s) = \sum_{i \in [n]} 1\{S_i < s\}$. The distribution of $W_{n,1}(\tau)$ 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 each stratum, i.e.,

$$W_{n,1}(\tau)|\{(A_i, S_i)_{i\in[n]}\} \stackrel{d}{=} \widetilde{W}_{n,1}(\tau)|\{(A_i, S_i)_{i\in[n]}\}$$

where

$$\begin{aligned}\widetilde{W}_{n,1}(\tau) \equiv \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} &\sum_{i=N(s)+1}^{N(s)+n_1(s)} \phi_1(\tau, s, Y_i^s(1), X_i^s) \\ - \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} &\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \phi_0(\tau, s, Y_i^s(0), X_i^s),\end{aligned}$$

with

$$\begin{aligned}\phi_1(\tau, s, Y_i^s(1), X_i^s) =& \frac{\tau - 1\{Y_i^s(1) \le q_1(\tau)\} - m_1(\tau, s) - (1-\pi(s))\left(\overline{m}_1(\tau, s, X_i^s) - \overline{m}_1(\tau, s)\right)}{\pi(s) f_1(q_1(\tau))} \\ &- \frac{(\overline{m}_0(\tau, s, X_i^s) - \overline{m}_0(\tau, s))}{f_0(q_0(\tau))},\end{aligned}$$

and

$$\begin{aligned}\phi_0(\tau, s, Y_i^s(0), X_i^s) =& \frac{\tau - 1\{Y_i^s(0) \le q_0(\tau)\} - m_0(\tau, s) - \pi(s)\left(\overline{m}_0(\tau, s, X_i^s) - \overline{m}_0(\tau, s)\right)}{(1-\pi(s)) f_0(q_0(\tau))} \\ &- \frac{(\overline{m}_1(\tau, s, X_i^s) - \overline{m}_1(\tau, s))}{f_1(q_1(\tau))}.\end{aligned}$$

As $W_{n,2}(\tau)$ is only a function of $\{S_i\}_{i\in[n]}$, we have

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

Let $F(s) = \mathbb{P}(S_i < s)$, $p(s) = \mathbb{P}(S_i = s)$, and

$$\begin{aligned}W^{\star}_{n,1}(\tau) \equiv \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} &\sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi(s)p(s))\rfloor} \phi_1(\tau, s, Y_i^s(1), X_i^s) \\ - \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} &\sum_{i=\lfloor n(F(s)+\pi(s)p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} \phi_0(\tau, s, Y_i^s(0), X_i^s).\end{aligned}$$

Note $W^{\star}_{n,1}(\tau)$ is a function of $(Y_i^s(1), Y_i^s(0), X_i^s)_{i\in[n], s\in\mathcal{S}}$, which is independent of $\{A_i, S_i\}_{i\in[n]}$ by construction. Therefore,

$$W^{\star}_{n,1}(\tau) \perp\!\!\!\perp W_{n,2}(\tau).$$

Note that

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

Denote $\Gamma_{n,a}(s,t,\tau) = \sum_{i=1}^{\lfloor nt \rfloor} \frac{1}{n}\phi_a(\tau, s, Y_i^s(a), X_i^s)$ for $a = 0, 1$. In order to show that $\sup_{\tau\in\Upsilon} |\widetilde{W}_{n,1}(\tau) - W^\star_{n,1}(\tau)| = o_p(1)$ and $W^\star_{n,1}(\tau) \rightsquigarrow \mathcal{B}_1(\tau)$, it suffices to show that (1) for $a = 0, 1$ and $s \in \mathcal{S}$, the stochastic process

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

is stochastically equicontinuous and (2) $W^\star_{n,1}(\tau) \rightsquigarrow \mathcal{B}_1(\tau)$ converges to $B_1(\tau)$ in finite dimension.

Claim (1). We want to bound

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

where the 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,

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

Then, for an arbitrary $\delta > 0$, by taking $\varepsilon = \delta^4$, we have

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

where in the first inequality, $S_k(\tau) = \sum_{i=1}^k \phi_a(\tau, s, Y_i^s(a), X_i^s)$ and the second inequality holds due to Lemma P.1. To see the third inequality, denote

$$\mathcal{F} = \{\phi_a(\tau, s, Y_i^s(a), X_i^s) : \tau \in \Upsilon\}$$

with an envelope function F_i such that by Assumption 3, $||F_i||_{\mathbb{P},q} < \infty$. In addition, by Assumption 3 again and the fact that

$$\{\tau - 1\{Y_i^s(a) \leq q_a(\tau)\} - m_a(\tau, s) : \tau \in \Upsilon\}$$

is of the VC-type with fixed coefficients (α, v), so is $\mathcal{F}$. Then, we have

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

where

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

$N(\mathcal{F}, L_2(Q), \varepsilon||F||_{Q,2})$ 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)

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

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

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

where in the first equality, $S_k(\tau_1, \tau_2) = \sum_{i=1}^k(\phi_a(\tau_2, s, Y_i^s(a), X_i^s) - \phi_a(\tau_1, s, Y_i^s(a), X_i^s))$ and the first inequality is due to Lemma P.1. To see the last inequality, denote

$$\mathcal{F} = \{\phi_a(\tau_2, s, Y_i^s(a), X_i^s) - \phi_a(\tau_1, s, Y_i^s(a), X_i^s) : \tau_1, \tau_2 \in \Upsilon, \tau_1 < \tau_2 < \tau_1 + \varepsilon\}$$

with a constant envelope function F_i such that $||F_i||_{\mathbb{P},q} < \infty$. In addition, due to Assumptions 2.2 and 3.3, one can show that

$$\sup_{f \in \mathcal{F}} \mathbb{E} f^2 \leq c\varepsilon \equiv \sigma^2$$

for some constant $c > 0$. Last, due to Assumption 3.2, $\mathcal{F}$ is of the VC-type with fixed coefficients (α, v). Therefore, by Chernozhukov et al. (2014, Corollary 5.1),

$$\frac{270\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon}|\sum_{i=1}^n(\phi_a(\tau_2, s, Y_i^s(a), X_i^s) - \phi_a(\tau_1, s, Y_i^s(a), X_i^s))|}{\sqrt{n}\delta}$$
$$\lesssim \frac{\sqrt{n}\mathbb{E}||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}}}{\delta} \lesssim \sqrt{\frac{\sigma^2 \log(\frac{C}{\sigma})}{\delta^2}} + \frac{C\log(\frac{C}{\sigma})}{\sqrt{n}\delta} \lesssim \delta\sqrt{\log(\frac{C}{\delta^2})},$$

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 array CLT,

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

where

$$\begin{aligned}
\Sigma_1(\tau,\tau) = &\lim_{n\to\infty}\sum_{s\in\mathcal{S}}\frac{(\lfloor n(F(s)+\pi(s)p(s))\rfloor - \lfloor nF(s)\rfloor)}{n}\mathbb{E}\phi_1^2(\tau, s, Y_i^s(1), X_i^s) \\
&+ \lim_{n\to\infty}\sum_{s\in\mathcal{S}}\frac{(\lfloor n(F(s)+p(s))\rfloor - \lfloor n(F(s)+p(s)\pi(s))\rfloor)}{n}\mathbb{E}\phi_0^2(\tau, s, Y_i^s(0), X_i^s) \\
&= \sum_{s\in\mathcal{S}} p(s)\mathbb{E}(\pi(s)\phi_1^2(\tau, S_i, Y_i(1), X_i) + (1-\pi(s))\phi_0^2(\tau, S_i, Y_i(0), X_i)|S_i = s) \\
&= \mathbb{E}\pi(S_i)\phi_1^2(\tau, S_i, Y_i(1), X_i) + \mathbb{E}(1-\pi(S_i))\phi_0^2(\tau, S_i, Y_i(0), X_i).
\end{aligned}$$

Finite dimensional convergence is proved by the Cramér-Wold device. In particular, we can show that the covariance kernel is

$$\Sigma_1(\tau, \tau') = \mathbb{E}\pi(S_i)\phi_1(\tau, S_i, Y_i(1), X_i)\phi_1(\tau', S_i, Y_i(1), X_i)$$

$$+\mathbb{E}(1-\pi(S_i))\phi_0(\tau,S_i,Y_i(0),X_i)\phi_0(\tau',S_i,Y_i(0),X_i).$$

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

Step 2. As $m_a(\tau,S_i)=\tau-\mathbb{P}(Y_i(a)\leq q_a(\tau)|S_i)$ is Lipschitz continuous in τ with a bounded Lipschitz constant, $\{m_a(\tau,S_i):\tau\in\Upsilon\}$ is of the VC-type with fixed coefficients (α,v) and a constant envelope function. Therefore, $\{\frac{m_1(\tau,S_i)}{f_1(q_1(\tau))}-\frac{m_0(\tau,S_i)}{f_0(q_0(\tau))}:\tau\in\Upsilon\}$ is a Donsker class and we have

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

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

$$\Sigma_2(\tau,\tau')=\mathbb{E}\left(\frac{m_1(\tau,S_i)}{f_1(q_1(\tau))}-\frac{m_0(\tau,S_i)}{f_0(q_0(\tau))}\right)\left(\frac{m_1(\tau',S_i)}{f_1(q_1(\tau'))}-\frac{m_0(\tau',S_i)}{f_0(q_0(\tau'))}\right)\equiv\mathbb{E}\phi_s(\tau,S_i)\phi_s(\tau',S_i).$$

This concludes the proof. □

Lemma P.4. *Suppose Assumptions in Theorem 4.1 hold and recall* $D_n^w(s)=\sum_{i\in[n]}\xi_i(A_i-\pi(S_i))1\{S_i=s\}$. *Then,* $\max_{s\in\mathcal{S}}|(D_n^w(s)-D_n(s))/n(s)|=o_p(1)$ *and* $\max_{s\in\mathcal{S}}|D_n^w(s)/n^w(s)|=o_p(1)$.

Proof. We note that $n^w(s)/n(s)\xrightarrow{p}1$ and $D_n(s)/n(s)\xrightarrow{p}0$. Therefore, we only need to show

$$\frac{D_n^w(s)-D_n(s)}{n(s)}=\sum_{i=1}^n\frac{(\xi_i-1)(A_i-\pi(s))1\{S_i=s\}}{n(s)}\xrightarrow{p}0.$$

As $n(s)\to\infty$ a.s., given data,

$$\begin{aligned}\frac{1}{n(s)}\sum_{i=1}^n(A_i-\pi(s))^21\{S_i=s\}=&\frac{1}{n}\sum_{i=1}^n\left(A_i-\pi(s)-2\pi(s)(A_i-\pi(s))+\pi(s)-\pi^2(s)\right)1\{S_i=s\}\\=&\frac{D_n(s)-2\pi(s)D_n(s)}{n(s)}+\pi(s)(1-\pi(s))\xrightarrow{p}\pi(s)(1-\pi(s)).\end{aligned}$$

Then, by the Lindeberg CLT, conditionally on data,

$$\frac{1}{\sqrt{n(s)}}\sum_{i=1}^n(\xi_i-1)(A_i-\pi(s))1\{S_i=s\}\rightsquigarrow N(0,\pi(s)(1-\pi(s)))=O_p(1),$$

and thus

$$\frac{D_n^w(s)-D_n(s)}{n(s)}=O_p(n^{-1/2}(s))=o_p(1).$$

□

Lemma P.5. *Suppose Assumptions in Theorem 4.1 hold. Then, uniformly over $\tau \in \Upsilon$ and conditionally on data,*

$$W^w_{n,1}(\tau) + W^w_{n,2}(\tau) \rightsquigarrow \mathcal{B}(\tau),$$

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

$$\begin{aligned}\Sigma(\tau, \tau') &= \mathbb{E}\pi(S_i)\phi_1(\tau, S_i, Y_i(1), X_i)\phi_1(\tau', S_i, Y_i(1), X_i) \\ &+ \mathbb{E}(1-\pi(S_i))\phi_0(\tau, S_i, Y_i(0), X_i)\phi_0(\tau', S_i, Y_i(0), X_i) + \mathbb{E}\phi_s(\tau, S_i)\phi_s(\tau', S_i).\end{aligned}$$

Proof. We divide the proof into two steps. In the first step, we show the conditional stochastic equicontinuity of $W^w_{n,1}(\tau)$ and $W^w_{n,2}(\tau)$. In the second step, we show the finite-dimensional convergence of $W^w_{n,1}(\tau) + W^w_{n,2}(\tau)$ conditional on data.

Step 1. Following the same idea in the proof of Lemma P.3, we define $\{(\xi_i^s, X_i^s, Y_i^s(1), Y_i^s(0)) : 1 \leq i \leq n\}$ as a sequence of i.i.d. random variables with marginal distributions equal to the distribution of $(\xi_i, X_i, Y_i(1), Y_i(0))|S_i = s$ and $N(s) = \sum_{i\in[n]} 1\{S_i < s\}$. The distribution of $W_{n,1}(\tau)$ 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 each stratum, i.e.,

$$W^w_{n,1}(\tau)|\{(A_i, S_i)_{i\in[n]}\} \stackrel{d}{=} \widetilde{W}^w_{n,1}(\tau)|\{(A_i, S_i)_{i\in[n]}\},$$

and thus,

$$W^w_{n,1}(\tau) \stackrel{d}{=} \widetilde{W}^w_{n,1}(\tau), \tag{P.2}$$

where

$$\begin{aligned}\widetilde{W}^w_{n,1}(\tau) \equiv &\sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} (\xi_i^s - 1)\phi_1(\tau, s, Y_i^s(1), X_i^s) \\ &- \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} (\xi_i^s - 1)\phi_0(\tau, s, Y_i^s(0), X_i^s).\end{aligned}$$

In addition, let

$$W^{w\star}_{n,1}(\tau) \equiv \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi(s)p(s))\rfloor} (\xi_i^s - 1)\phi_1(\tau, s, Y_i^s(1), X_i^s)$$
$$- \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=\lfloor n(F(s)+\pi(s)p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} (\xi_i^s - 1)\phi_0(\tau, s, Y_i^s(0), X_i^s).$$

Following exactly the same argument as in the proof of Lemma P.3, we have

$$\sup_{\tau\in\Upsilon} |\widetilde{W}^w_{n,1}(\tau) - W^{w\star}_{n,1}(\tau)| = o_p(1). \tag{P.3}$$

and $W^{w\star}_{n,1}(\tau)$ is *unconditionally* stochastically equicontinuous, i.e., for any $\varepsilon > 0$, as $n \to \infty$ followed by $\delta \to 0$, we have

$$\mathbb{E}\mathbb{P}_\xi \left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta} |W^{w\star}_{n,1}(\tau_1) - W^{w\star}_{n,1}(\tau_2)| \geq \varepsilon \right)$$
$$= \mathbb{P} \left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta} |W^{w\star}_{n,1}(\tau_1) - W^{w\star}_{n,1}(\tau_2)| \geq \varepsilon \right) \to 0,$$

where $\mathbb{P}_\xi$ means the probability operator is with respect to the bootstrap weights $\{\xi_i\}_{i\in[n]}$ and is conditional on data. This implies the *unconditional* stochastic equicontinuity of $W^w_{n,1}(\tau)$ due to (P.2) and (P.3), which further implies the *conditional* stochastic equicontinuity of $W^w_{n,1}(\tau)$, i.e., for any $\varepsilon > 0$, as $n \to \infty$ followed by $\delta \to 0$,

$$\mathbb{P}_\xi \left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta} |W^{w\star}_{n,1}(\tau_1) - W^{w\star}_{n,1}(\tau_2)| \geq \varepsilon \right) \xrightarrow{p} 0.$$

By a similar but simpler argument, the *conditional* stochastic equicontinuity of $W^w_{n,2}(\tau)$ holds as well. This concludes the first step.

Step 2. We first show the asymptotic normality of $W^w_{n,1}(\tau) + W^w_{n,2}(\tau)$ conditionally on data for a fixed τ. Note

$$\begin{aligned}
&W^w_{n,1}(\tau) + W^w_{n,2}(\tau) \\
=&\frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1) \left[A_i 1\{S_i = s\}\phi_1(\tau, S_i, Y_i(1), X_i) - (1 - A_i)1\{S_i = s\}\phi_1(\tau, S_i, Y_i(1), X_i) + \phi_s(\tau, S_i)\right] \\
\equiv&\frac{1}{\sqrt{n}} \sum_{i\in[n]} (\xi_i - 1)\mathcal{J}_i(s, \tau).
\end{aligned}$$

Conditionally on data, $\{(\xi_i - 1)\mathcal{J}_i(\tau)\}_{i\in[n]}$ is a sequence of i.n.i.d. random variables. In order to apply the Lindeberg-Feller central limit theorem, we only need to show that (1)

$$\frac{1}{n}\sum_{i\in[n]}\mathcal{J}_{n,i}^2(\tau) \xrightarrow{p} \Sigma(\tau,\tau),$$

where $\Sigma(\tau,\tau)$ is defined in Theorem 3.1, and (2) the Lindeberg condition holds, i.e.,

$$\frac{1}{n}\sum_{i\in[n]}\mathcal{J}_{n,i}^2(\tau)\mathbb{E}(\xi_i-1)^2 1\{|(\xi_i-1)\mathcal{J}_{n,i}(\tau)| \geq \sqrt{n}\varepsilon\} \xrightarrow{p} 0.$$

For part (1), we have

$$\frac{1}{n}\sum_{i\in[n]}\mathcal{J}_{n,i}^2(\tau) = \sigma_1^2 + 2\sigma_{12} + \sigma_2^2,$$

where

$$\sigma_1^2 = \frac{1}{n}\sum_{i\in[n]}\left[A_i 1\{S_i = s\}\phi_1(\tau, S_i, Y_i(1), X_i) - (1-A_i)1\{S_i = s\}\phi_0(\tau, S_i, Y_i(1), X_i)\right]^2,$$

$$\sigma_{12} = \frac{1}{n}\sum_{i\in[n]}\left[A_i 1\{S_i = s\}\phi_1(\tau, S_i, Y_i(1), X_i) - (1-A_i)1\{S_i = s\}\phi_1(\tau, S_i, Y_i(1), X_i)\right]\phi_s(\tau, S_i),$$

and

$$\sigma_2^2 = \frac{1}{n}\sum_{i\in[n]}\phi_s^2(\tau, S_i).$$

Note

$$\begin{aligned}
\sigma_1^2 &= \frac{1}{n}\sum_{i\in[n]}A_i 1\{S_i = s\}\phi_1^2(\tau, S_i, Y_i(1), X_i) + \frac{1}{n}\sum_{i\in[n]}(1-A_i)1\{S_i = s\}\phi_0^2(\tau, S_i, Y_i(1), X_i) \\
&\stackrel{d}{=} \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\phi_1^2(\tau, s, Y_i^s(1), X_i^s) + \frac{1}{n}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\phi_0^2(\tau, s, Y_i^s(1), X_i^s) \\
&\xrightarrow{p} \sum_{s\in\mathcal{S}}[\pi(s)\mathbb{E}\phi_1^2(\tau, s, Y_i^s(1), X_i^s) + (1-\pi(s))\mathbb{E}\phi_0^2(\tau, s, Y_i^s(1), X_i^s)] \\
&= \mathbb{E}\left[\pi(S_i)\phi_1^2(\tau, S_i, Y_i(1), X_i) + (1-\pi(s))\phi_0^2(\tau, S_i, Y_i(1), X_i)\right],
\end{aligned}$$

where the convergence holds due to the fact that $N(s)/n \to F(s)$, $n_1(s)/n \xrightarrow{p} \pi(s)p(s)$, $n(s)/n \xrightarrow{p}$

$p(s)$, and the uniform convergence of the partial sum process. Similarly,

$$\sigma_{12} \overset{d}{=} \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \phi_1(\tau, s, Y_i^s(1), X_i^s)\phi_s(\tau, s) + \frac{1}{n}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \phi_0(\tau, s, Y_i^s(1), X_i^s)\phi_s(\tau, s)$$
$$\xrightarrow{p} \sum_{s\in\mathcal{S}}[\pi(s)\mathbb{E}\phi_1(\tau, s, Y_i^s(1), X_i^s) + (1-\pi(s))\mathbb{E}\phi_0(\tau, s, Y_i^s(1), X_i^s)]\phi_s(\tau, s) = 0,$$

where we use the fact that

$$\mathbb{E}\phi_1(\tau, s, Y_i^s(1), X_i^s) = \mathbb{E}\phi_0(\tau, s, Y_i^s(1), X_i^s) = 0.$$

By the standard weak law of large numbers, we have

$$\sigma_2^2 \xrightarrow{p} \mathbb{E}\phi_s^2(\tau, S_i).$$

Therefore,

$$\frac{1}{n}\sum_{i\in[n]} \mathcal{J}_{n,i}^2(\tau) \xrightarrow{p} \mathbb{E}\left[\pi(S_i)\phi_1^2(\tau, S_i, Y_i(1), X_i) + (1-\pi(s))\phi_0^2(\tau, S_i, Y_i(1), X_i)\right] + \mathbb{E}\phi_s^2(\tau, S_i) = \Sigma(\tau, \tau).$$

To verify the Lindeberg condition, we note that

$$\begin{aligned}
&\frac{1}{n}\sum_{i\in[n]} \mathcal{J}_{n,i}^2(\tau)\mathbb{E}(\xi_i - 1)^2 1\{|(\xi_i - 1)\mathcal{J}_{n,i}(\tau)| \geq \sqrt{n}\varepsilon\} \\
&\leq \frac{1}{n(\sqrt{n}\varepsilon)^{q-2}} \sum_{i\in[n]} \mathcal{J}_{n,i}^q(\tau)\mathbb{E}(\xi_i - 1)^q \\
&\leq \frac{c}{n(\sqrt{n}\varepsilon)^{q-2}} \sum_{i\in[n]} [\phi_1^q(\tau, S_i, Y_i(1), X_i) + \phi_0^q(\tau, S_i, Y_i(1), X_i) + \phi_s^q(\tau, S_i)] = o_p(1),
\end{aligned}$$

where the last equality is due to Assumption 3(ii) and the fact that $\eta_{i,a}(s, \tau)$ is bounded.

The finite dimensional convergence of $W_{n,1}^w(\tau) + W_{n,2}^w(\tau)$ across τ can be established in the same manner using the Cramér-Wold device and details are omitted. By the same calculation given above the covariance kernel is shown to be

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\sum_{i\in[n]} \mathcal{J}_{n,i}(\tau_1)\mathcal{J}_{n,i}(\tau_2) \\
&= \mathbb{E}\left[\pi(S_i)\phi_1(\tau_1, S_i, Y_i(1), X_i)\phi_1(\tau_2, S_i, Y_i(1), X_i)\right] \\
&+ \mathbb{E}\left[(1-\pi(s))\phi_0(\tau_1, S_i, Y_i(1), X_i)\phi(\tau_2, S_i, Y_i(1), X_i)\right] \\
&+ \mathbb{E}\phi_s(\tau_1, S_i)\phi_s(\tau_2, S_i) = \Sigma(\tau_1, \tau_2),
\end{aligned}$$

which concludes the proof. $\square$

Lemma P.6. *Suppose Assumptions in Proposition 5.2 hold. Then,*

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta} |Q_n(s,\tau,\theta_1,\theta_0) - Q(s,\tau,\theta_1,\theta_0)| = o_p(1),$$

where $Q(\cdot)$ and $Q_n(\cdot)$ are defined in (5.3) *and* (5.9)*, respectively, with $\Lambda_{a,s}(X_i,\theta_a) = \lambda(\vec{X}_i^\top \theta_a)$. In addition, if $\sup_{\tau\in\Upsilon}(|\hat{f}_1(\hat{q}_1(\tau)) - f_1(q_1(\tau))| + |\hat{f}_0(\hat{q}_0(\tau)) - f_0(q_0(\tau))|) = o_p(\varepsilon_n)$ for some $\varepsilon_n \downarrow 0$ such that $\varepsilon_n\sqrt{n} \to \infty$. Then,*

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta} |Q_n(s,\tau,\theta_1,\theta_0) - Q(s,\tau,\theta_1,\theta_0)| = o_p(\varepsilon_n).$$

Proof. Recall $\Theta = \Theta_1 \times \Theta_0$, where Θ_1 and Θ_0 are two compact sets in $\Re^{d_x}$. Note that

$$\sup_{\theta_a\in\Theta_a} |\hat{g}_{a,s}(X_i,\theta_a)| \leq C < \infty,$$

for some constant C. By Assumption 9 and the fact that $\hat{\pi}(s) \xrightarrow{p} \pi(s) > 0$, we have

$$Q_n(s,\tau,\theta_1,\theta_0) = \tilde{Q}_n(s,\tau,\theta_1,\theta_0) + R_{n,1}(s,\tau,\theta_1,\theta_0),$$

where

$$\begin{aligned}\tilde{Q}_n(s,\tau,\theta_1,\theta_0) = &\frac{1}{n(s)}\sum_{i\in I(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)}\frac{\hat{g}_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)^2 \\ &+ \frac{2}{n_1(s)}\sum_{i\in I_1(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)}\frac{\hat{g}_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\frac{1\{Y_i \leq \hat{q}_1(\tau)\}}{f_1(q_1(\tau))} \\ &+ \frac{2\pi(s)}{n_0(s)(1-\pi(s))}\sum_{i\in I_0(s)}\left(\frac{\hat{g}_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)}\frac{\hat{g}_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\frac{1\{Y_i \leq \hat{q}_0(\tau)\}}{f_0(q_0(\tau))}\end{aligned}$$

and

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta} |R_{n,1}(s,\tau,\theta_1,\theta_0)| = o_p(1).$$

In addition,

$$
\sup_{\theta_1 \in \Theta_1} |g_{1,s}(X_i,\theta_1) - \hat{g}_{1,s}(X_i,\theta_1)| = \sup_{\theta_1 \in \Theta_1} \left| \frac{1}{n_1(s)} \sum_{i \in I_1(s)} \lambda(\vec{X}_i^\top \theta_1) - \mathbb{E}\lambda(\vec{X}_i^\top \theta_1 | S_i = s) \right|
$$

$$
\stackrel{d}{=} \sup_{\theta_1 \in \Theta_1} \left| \frac{1}{n_1(s)} \sum_{i=N(s)+1}^{N(s)+n_1(s)} (\lambda((\vec{X}_i^s)^\top \theta_1) - \mathbb{E}\lambda((\vec{X}_i^s)^\top \theta_1)) \right| = O_p(n^{-1/2}), \tag{P.4}
$$

where the last equality is due to the fact that $\{X_i^s\}_{i \in [n]}$ is an i.i.d. sequence independent of $\{A_i, S_i\}_{i \in [n]}$ and the usual maximal inequality such as that in van der Vaart and Wellner (1996, Theorem 2.14.1) applies. Therefore,

$$
\tilde{Q}_n(s,\tau,\theta_1,\theta_0) = \breve{Q}_n(s,\tau,\theta_1,\theta_0) + R_{n,2}(s,\tau,\theta_1,\theta_0),
$$

where

$$
\begin{aligned}
\breve{Q}_n(s,\tau,\theta_1,\theta_0) = {} & \frac{1}{n(s)} \sum_{i \in I(s)} \left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))} \right)^2 \\
& + \frac{2}{n_1(s)} \sum_{i \in I_1(s)} \left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))} \right) \frac{1\{Y_i \le \hat{q}_1(\tau)\}}{f_1(q_1(\tau))} \\
& + \frac{2\pi(s)}{n_0(s)(1-\pi(s))} \sum_{i \in I_0(s)} \left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))} \right) \frac{1\{Y_i \le \hat{q}_0(\tau)\}}{f_0(q_0(\tau))} \\
& \equiv \breve{Q}_{n,1}(s,\tau,\theta_1,\theta_0) + \breve{Q}_{n,2}(s,\tau,\theta_1,\theta_0) + \breve{Q}_{n,3}(s,\tau,\theta_1,\theta_0)
\end{aligned}
$$

and

$$
\sup_{\tau \in \Upsilon, s \in \mathcal{S}, \theta_1, \theta_0 \in \Theta} |R_{n,2}(s,\tau,\theta_1,\theta_0)| = O_p(n^{-1/2}).
$$

By the same argument as (P.4), we can show that

$$
\sup_{\tau \in \Upsilon, s \in \mathcal{S}, \theta_1, \theta_0 \in \Theta} \left| \breve{Q}_{n,1}(s,\tau,\theta_1,\theta_0) - \mathbb{E}\left[\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))} \right)^2 \middle| S_i = s \right] \right| = O_p(n^{-1/2}).
$$

Next, we show

$$
\sup_{\tau \in \Upsilon, s \in \mathcal{S}, \theta_1, \theta_0 \in \Theta} \left| \breve{Q}_{n,2}(s,\tau,\theta_1,\theta_0) - 2\mathbb{E}\left[\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))} + \frac{\pi(s)}{1-\pi(s)} \frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))} \right) \frac{1\{Y_i(1) \le q_1(\tau)\}}{f_1(q_1(\tau))} | S_i = s \right] \right|
$$

$$
= O_p(n^{-1/2}).
$$

The proof of

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta}\left|\breve{Q}_{n,3}(s,\tau,\theta_1,\theta_0)-\frac{2\pi(s)}{(1-\pi(s))}\mathbb{E}\left[\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))}+\frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\frac{1\{Y_i\leq q_0(\tau)\}}{f_0(q_0(\tau))}\middle|S_i=s\right]\right|$$
$$=O_p(n^{-1/2})$$

is similar and is omitted.

Denote

$$\phi_i(s,\tau,\theta_1,\theta_0,q)=\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))}+\frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\frac{1\{Y_i(1)\leq q\}}{f_1(q_1(\tau))}.$$

Because $\sup_{\tau\in\Upsilon}|\hat{q}_1(\tau)-q_1(\tau)|=O_p(n^{-1/2})$, for any $\varepsilon>0$, there exists a constant $M>0$ such that with probability greater than $1-\varepsilon$, we have

$$\sup_{\tau\in\Upsilon}|\hat{q}_1(\tau)-q_1(\tau)|\leq n^{-1/2}M.$$

Therefore, with probability greater than $1-\varepsilon$, we have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta}\left|\breve{Q}_{n,2}(s,\tau,\theta_1,\theta_0)-2\mathbb{E}\left[\left(\frac{g_{1,s}(X_i,\theta_1)}{f_1(q_1(\tau))}+\frac{\pi(s)}{1-\pi(s)}\frac{g_{0,s}(X_i,\theta_0)}{f_0(q_0(\tau))}\right)\frac{1\{Y_i(1)\leq q_1(\tau)\}}{f_1(q_1(\tau))}|S_i=s\right]\right|\\
\leq&\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta,|q_1-q_2|\leq n^{-1/2}M}\left|\frac{2}{n_1(s)}\sum_{i\in I_1(s)}[\phi_i(s,\tau,\theta_1,\theta_0,q_1)-\mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q_2)|S_i=s)]\right|\\
\leq&\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta,q_1\in\Re}\left|\frac{2}{n_1(s)}\sum_{i\in I_1(s)}[\phi_i(s,\tau,\theta_1,\theta_0,q_1)-\mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q_1)|S_i=s)]\right|\\
+&\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta,|q_1-q_2|\leq n^{-1/2}M}\left|2\left[\mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q_1)|S_i=s)-\mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q_2)|S_i=s)\right]\right|.
\end{aligned} \tag{P.5}$$

For the first term on the RHS of (P.5), we note that

$$\mathcal{F}=\{\phi_i(s,\tau,\theta_1,\theta_0,q_1):\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta,q_1\in\Re\}$$

is of the VC-type with fixed coefficients (α,v) and a bounded envelope. Therefore, Lemma P.2 implies

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S},\theta_1,\theta_0\in\Theta,q_1\in\Re}\left|\frac{2}{n_1(s)}\sum_{i\in I_1(s)}[\phi_i(s,\tau,\theta_1,\theta_0,q_1)-\mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q_1)|S_i=s)]\right|=O_p(n^{-1/2}).$$

For the second term on the RHS of (P.5), we note that for some constant $C > 0$,

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta, q_1\in\Re} |\partial_q \mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q)|S_i = s)| \leq C \sup_{q\in\Re} f_1(q|X_i,S_i) < \infty.$$

Therefore,

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta, |q_1-q_2|\leq n^{-1/2}M} |2\left[\mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q_1)|S_i = s) - \mathbb{E}(\phi_i(s,\tau,\theta_1,\theta_0,q_2)|S_i = s)\right]| = O(n^{-1/2}).$$

Combining these bounds, we have shown that

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta} |Q_n(s,\tau,\theta_1,\theta_0) - Q(s,\tau,\theta_1,\theta_0)| = o_p(1).$$

For the second result, we note that

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta} |R_{n,1}(s,\tau,\theta_1,\theta_0)| \leq C \sup_{\tau\in\Upsilon} (|\hat{f}_1(\hat{q}_1(\tau)) - f_1(q_1(\tau))| + |\hat{f}_0(\hat{q}_0(\tau)) - f_0(q_0(\tau))|) = o_p(\varepsilon_n).$$

The other terms converge at the $n^{1/2}$-rate. This implies

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}, \theta_1,\theta_0\in\Theta} |Q_n(s,\tau,\theta_1,\theta_0) - Q(s,\tau,\theta_1,\theta_0)| = o_p(\varepsilon_n).$$

□

Lemma P.7. *Suppose Assumptions in Theorem 5.4 hold. Then,*

$$\sup_{\tau\in\Upsilon, a=0,1, s\in\mathcal{S}} \left\| \frac{1}{n_a(s)} \sum_{i\in I_a(s)} (1\{Y_i \leq \hat{q}_a(\tau)\} - m_a(\tau,s,X_i)) H_{h_n}(X_i) \right\|_\infty = O_p\left(\sqrt{\frac{\log(n)}{n}}\right).$$

Proof. We focus on $a = 1$. We have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon} \left\| \frac{1}{n_1(s)} \sum_{i\in I_1(s)} (1\{Y_i \leq \hat{q}_1(\tau)\} - m_1(\tau,s,X_i)) H_{h_n}(X_i) \right\|_\infty \\
&\leq \sup_{\tau\in\Upsilon} \left\| \frac{1}{n_1(s)} \sum_{i\in I_1(s)} (1\{Y_i \leq \hat{q}_1(\tau)\} - \mathbb{P}(Y_i(1) \leq \hat{q}_1(\tau)|X_i, S_i = s)) H_{h_n}(X_i) \right\|_\infty \\
&+ \sup_{\tau\in\Upsilon} \left\| \frac{1}{n_1(s)} \sum_{i\in I_1(s)} (\mathbb{P}(Y_i(1) \leq \hat{q}_1(\tau)|X_i, S_i = s) - m_1(\tau,s,X_i)) H_{h_n}(X_i) \right\|_\infty
\end{aligned}$$

$$\leq \sup_{q\in\Re}\left\|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}(1\{Y_i\leq q\}-\mathbb{P}(Y_i(1)\leq q|X_i,S_i=s))H_{h_n}(X_i)\right\|_\infty$$
$$+\sup_{\tau\in\Upsilon}\left\|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}(\mathbb{P}(Y_i(1)\leq \hat{q}_1(\tau)|X_i,S_i=s)-m_1(\tau,s,X_i))H_{h_n}(X_i)\right\|_\infty. \quad \text{(P.6)}$$

Denote

$$\mathcal{F}_h=\{1\{Y_i^s(1)\leq q\}H_{h_n,h}(X_i):q\in\Re\},\quad \mathcal{F}=\cup_{h\in[h_n]}\mathcal{F}_{h_n},$$

and $H_{h_n,h}(X_i)$ is the h-th coordinate of $H_{h_n}(X_i)$. For each $h\in[h_n]$, $\mathcal{F}_h$ is of the VC-type with fixed coefficients (α,v) and a common envelope $F_i=||H_{h_n}(X_i)||_2\leq\zeta(h_n)$, i.e.,

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

where the supremum is taken over all finitely discrete probability measures. This implies

$$\sup_Q N(\mathcal{F},e_Q,\varepsilon||F||_{Q,2})\leq\sum_{h\in[h_n]}\sup_Q N(\mathcal{F}_h,e_Q,\varepsilon||F||_{Q,2})\leq\left(\frac{\alpha h_n}{\varepsilon}\right)^v\quad \forall\varepsilon\in(0,1],$$

i.e., $\mathcal{F}$ is also of the VC-type with coefficients $(\alpha h_n,v)$. In addition,

$$\sup_{f\in\mathcal{F}}\mathbb{E}f^2\leq\max_{h\in[h_n]}\mathbb{E}H^2_{h_n,h}(X_i)\leq C<\infty.$$

Then, Lemma P.2 implies

$$\sup_{q\in\Re}\left\|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}(1\{Y_i\leq q\}-\mathbb{P}(Y_i(1)\leq q|X_i,S_i=s))H_{h_n}(X_i)\right\|_\infty$$
$$=O_p\left(\sqrt{\frac{\log(h_n\zeta(h_n))}{n}}+\frac{\zeta(h_n)\log(\zeta(h_n))}{n}\right)=O_p\left(\sqrt{\frac{\log(n)}{n}}\right).$$

For the second term of (P.6), because $\sup_{q\in\Re,x\in\text{Supp}(X),s\in\mathcal{S}}f_1(q|x,s)<\infty$, we have

$$\sup_{\tau\in\Upsilon}\left\|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}(\mathbb{P}(Y_i(1)\leq\hat{q}_1(\tau)|X_i,S_i=s)-m_1(\tau,s,X_i))H_{h_n}(X_i)\right\|_\infty$$
$$\leq\sup_{\tau\in\Upsilon}|\hat{q}_1(\tau)-q_1(\tau)|\left\|\frac{1}{n_1(s)}\sum_{i\in I_1(s)}|H_{h_n}(X_i)|\right\|_\infty$$

$$
\begin{aligned}
&\leq \sup_{\tau\in\Upsilon} |\hat{q}_1(\tau) - q_1(\tau)| \left\| \frac{1}{n_1(s)} \sum_{i\in I_s(1)} [|H_{h_n}(X_i^s)| - \mathbb{E}(|H_{h_n}(X_i)||S_i = s)] \right\|_\infty \\
&+ \sup_{\tau\in\Upsilon} |\hat{q}_1(\tau) - q_1(\tau)| \, \|\mathbb{E}(|H_{h_n}(X_i)||S_i = s)\|_\infty \\
&= \sup_{\tau\in\Upsilon} |\hat{q}_1(\tau) - q_1(\tau)| \left\| \frac{1}{n_1(s)} \sum_{i\in I_s(1)} [|H_{h_n}(X_i^s)| - \mathbb{E}(|H_{h_n}(X_i)||S_i = s)] \right\|_\infty + O_p(n^{-1/2}) \\
&= O_p(n^{-1/2}),
\end{aligned}
$$

where the second to last inequality holds because of Assumption 7 and $\|\mathbb{E}(|H_{h_n}(X_i)||S_i = s)\|_\infty \leq C < \infty$, and the last inequality holds because by a similar argument to the one used in bounding the first term on the RHS of (P.6), we can show that

$$
\left\| \frac{1}{n_1(s)} \sum_{i\in I_s(1)} [|H_{h_n}(X_i^s)| - \mathbb{E}(|H_{h_n}(X_i)||S_i = s)] \right\|_\infty = O_p\left(\sqrt{\frac{\log(n)}{n}}\right).
$$

This concludes the proof. □

Lemma P.8. *Suppose Assumptions in Theorem 5.4 hold and recall* $\overline{\ell}$ *defined in* (M.1). *We have* $\overline{\ell}/(\sqrt{h_n \log(n)}/n) \to \infty$, *w.p.a.1.*

Proof. Note that w.p.a.1,

$$
\begin{aligned}
\overline{\ell} &= \inf_{U\in\Re^{h_n}} \frac{\left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)} (H_{h_n}^\top(X_i)U)^2\right]^{3/2}}{\frac{1}{n_a(s)}\sum_{i\in I_a(s)} |H_{h_n}^\top(X_i)U|^3} \\
&\geq \inf_{U\in\Re^{h_n}} \frac{\left[\frac{1}{n_a(s)}\sum_{i\in I_a(s)} (H_{h_n}^\top(X_i)U)^2\right]^{1/2}}{\sup_{x\in\mathcal{X}} ||H_{h_n}(x)||_2||U||_2} \geq \frac{\kappa_1^{1/2}}{\zeta(h_n)},
\end{aligned}
$$

where the last inequality is due to Assumption 11. Therefore,

$$
\overline{\ell}/(\sqrt{h_n \log(n)}/n) \geq \sqrt{\frac{\kappa_1 n}{\zeta^2(h_n) h_n \log(n)}} \to \infty \; w.p.a.1.
$$

□

References

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

Anderson, S. J. and D. McKenzie (2021). Improving business practices and the boundary of the entrepreneur: a randomized experiment comparing training, consulting, insourcing and outsourcing. *Journal of Political Economy, forthcoming.*

Athey, S. and G. W. Imbens (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, Volume 1, pp. 73–140. Elsevier.

Bach, F. (2010). Self-concordant analysis for logistic regression. *Electronic Journal of Statistics 4*, 384–414.

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

Bai, Y., A. Shaikh, and J. P. Romano (2021). Inference in experiments with matched pairs. *Journal of the American Statistical Association, forthcoming.*

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

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017). Program evaluation with high-dimensional data. *Econometrica 85*(1), 233–298.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006). What mean impacts miss: distributional effects of welfare reform experiments. *American Economic Review 96*(4), 988–1012.

Bloniarz, A., H. Liu, C.-H. Zhang, J. S. Sekhon, and B. Yu (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences 113*(27), 7383–7390.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association 71*(356), 791–799.

Bradic, J., S. Wager, and Y. Zhu (2019). Sparsity double robust inference of average treatment effects. *arXiv preprint arXiv:1905.00744*.

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

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

Bugni, F. A. and M. Gao (2021). Inference under covariate-adaptive randomization with imperfect compliance. *arXiv preprint arXiv:2102.03937*.

Burchardi, K. B., S. Gulesci, B. Lerva, and M. Sulaiman (2019, February). Moral hazard: experimental evidence from tenancy contracts. *Quarterly Journal of Economics 134*(1), 281–347.

Campos, F., M. Frese, M. Goldstein, L. Iacovone, H. C. Johnson, D. McKenzie, and M. Mensmann (2017). Teaching personal initiative beats traditional training in boosting small business in west africa. *Science 357*(6357), 1287–1290.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, Volume 6, pp. 5549–5632. Elsevier.

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

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

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.

Crépon, B., F. Devoto, E. Duflo, and W. Parienté (2015). Estimating the impact of microcredit on those who take it up: evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics 7*(1), 123–50.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2013). Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India. *Quarterly Journal of Economics 128*(4), 1499–1545.

Dupas, P. and E. Miguel (2017). Impacts and determinants of health levels in low-income countries. In *Handbook of Economic Field Experiments*, Volume 2, pp. 3–93. Elsevier.

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

Fogarty, C. B. (2018). Regression-assisted inference for the average treatment effect in paired experiments. *Biometrika 105*(4), 994–1000.

Freedman, D. A. (2008a). On regression adjustments in experiments with several treatments. *Annals of Applied Statistics 2*(1), 176–196.

Freedman, D. A. (2008b). On regression adjustments to experimental data. *Advances in Applied Mathematics 40*(2), 180–193.

Glewwe, P. and E. A. Miguel (2007). The impact of child health and nutrition on education in less developed countries. In *Handbook of Development Economics*, Volume 4, pp. 3561–3606. Elsevier.

Greaney, B. P., J. P. Kaboski, and E. Van Leemput (2016). Can self-help groups really be "self-help"? *Review of Economic Studies 83*(4), 1614–1644.

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

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

Hu, Y. (2016). Generalized Efron's biased coin design and its theoretical properties. *Journal of Applied Probability 53*(2), 327–340.

Hu, Y. and F. Hu (2012). Asymptotic properties of covariate-adaptive randomization. *Annals of Statistics 40*(3), 1794–1815.

Jakiela, P. and O. Ozier (2016). Does Africa need a rotten kin theorem? Experimental evidence from village economies. *Review of Economic Studies 83*(1), 231–268.

Jiang, L., X. Liu, P. C. B. Phillips, and Y. Zhang (2021). Bootstrap inference for quantile treatment effects in randomized experiments with matched pairs. *Review of Economics and Statistics, forthcoming.*

Kallus, N., X. Mao, and M. Uehara (2019). Localized debiased machine learning: efficient inference on quantile treatment effects and beyond. *arXiv preprint arXiv:1912.12945*.

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

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Annals of Statistics 26*(2), 755–770.

Lei, L. and P. Ding (2021). Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika, forthcoming.*

Li, X. and P. Ding (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82*(1), 241–268.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique. *Annals of Applied Statistics 7*(1), 295–318.

Liu, H., F. Tu, and W. Ma (2020). A general theory of regression adjustment for covariate-adaptive randomization: OLS, Lasso, and beyond. *arXiv preprint arXiv:2011.09734*.

Liu, H. and Y. Yang (2020). Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika 107*(4), 935–948.

Lu, J. (2016). Covariate adjustment in randomization-based causal inference for 2k factorial designs. *Statistics & Probability Letters 119*, 11–20.

Ma, W., F. Hu, and L. Zhang (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association 110*(510), 669–680.

Ma, W., Y. Qin, Y. Li, and F. Hu (2018). Statistical inference of covariate-adjusted randomized experiments. *arXiv preprint arXiv:1807.09678*.

Ma, W., F. Tu, and H. Liu (2020). Regression analysis for covariate-adaptive randomization: a robust and efficient inference perspective. *arXiv preprint arXiv:2009.02287*.

Montgomery-Smith, S. J. (1993). Comparison of sums of independent identically distributed random variables. *Probability and Mathematical Statistics 14*(2), 281–285.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: experimental evidence from India. *Journal of Political Economy 119*(1), 39–77.

Negi, A. and J. M. Wooldridge (2020). Revisiting regression adjustment in experiments with heterogeneous treatment effects. *Econometric Reviews*, 1–31.

Olivares, M. (2021). Robust permutation test for equality of distributions under covariate-adaptive randomization. *Working paper, University of Illinois at Urbana Champaign*.

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.

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

Tan, Z. (2020). Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data. *Annals of Statistics 48*(2), 811–837.

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

von Neumann, J. (2019). The mathematician. In *Mathematics: People, Problems, Results* (2 ed.). Chapman and Hall/CRC.

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.

Ye, T. (2018). Testing hypotheses under covariate-adaptive randomisation and additive models. *Statistical Theory and Related Fields 2*(1), 96–101.

Ye, T. and J. Shao (2020). Robust tests for treatment effect in survival analysis under covariate-adaptive randomization. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82*(5), 1301–1323.

Ye, T., Y. Yi, and J. Shao (2020). Inference on average treatment effect under minimization and other covariate-adaptive randomization methods. *arXiv preprint arXiv:2007.09576*.

Zhang, M., A. A. Tsiatis, and M. Davidian (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics 64*(3), 707–715.

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

Zhao, A. and P. Ding (2020). Covariate-adjusted fisher randomization tests for the average treatment effect. *arXiv preprint arXiv:2010.14555*.