1-2018

Extremal quantile treatment effects

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

DOI: https://doi.org/10.1214/17-AOS1673

Extremal Quantile Treatment Effects*

Yichong Zhang†

June 7, 2016

Abstract

This paper establishes an asymptotic theory and inference method for quantile treatment effect estimators when the quantile index is close or equal to zero. Such quantile treatment effects are of interest in many economic applications, such as the effect of maternal smoking on an infant's adverse birth outcomes. When the quantile index is close to zero, the sparsity of data jeopardizes conventional asymptotic theory and bootstrap inference. When the quantile index is zero, there are no existing inference methods directly applicable in the treatment effect context. This paper establishes new estimation and inference theory for cases close or equal to zero. In addition, finite sample properties of the new procedures are illustrated through both simulation studies and empirical applications.

Keywords: Extreme quantile, Intermediate quantile

JEL codes: C21, I19

*I am deeply grateful to my co-advisors Shakeeb Khan and Arnaud Maurel, as well as my committee members Federico Bugni and Matt Masten for their guidance and encouragement. I also thank Xavier D'Haultfœuille, Jia Li, and participants in the microeconometrics lunch at Duke and the 11th World Congress of the Econometric Society for their comments and Peter Arcidiacono for sharing the UCOP data with me. All remaining errors are mine.

†Duke University. E-mail address: yz98@duke.edu.

1 Introduction

Economic theory usually predicts that the sign and magnitude of treatment effects vary depending on one's place in the overall distribution of outcomes, a heterogeneity captured by quantile treatment effects (QTEs). In many economic applications, the populations of interest are located at the tail of the outcome distribution, such as infants with low birth weights or students with low scores. Thus researchers encounter not only the usual missing counterfactual, but also data sparsity because there are not many observations at the tails. While previous literature has considered the two problems separately, how to cope with both at the same time while conducting proper statistical inferences remains unanswered.

This paper addresses both issues simultaneously. I establish a new asymptotic theory and inference method for an estimator of the QTE for low-rank populations. To deal with the usual missing counterfactual problem, I assume unconfoundedness and rely on the propensity score to identify QTEs. To address the data sparsity, I model a small quantile index τ as a drifting object with sample size n; that is, $\tau := \tau_n \to 0$ as $n \to \infty$. Then, I use the device of extremal quantiles to derive a new asymptotic approximation for the finite sample distribution of the QTE estimator when the quantile index τ is close to zero.

My paper addresses the problem of missing counterfactual and data sparsity jointly. I build on the previous literature that address only one issue at a time. For the treatment effect literature addressing the missing counterfactual problem, I adapt the same unconfoundedness assumption as Bitler, Gelbach, and Hoynes (2006), Chernozhukov, Fernández-Val, and Melly (2013), Firpo (2007), and Hirano, Imbens, and Ridder (2003). For further applications of QTEs, see Card (1996) and DiNardo, Fortin, and Lemieux (1996), for example.

For the extremal quantile literature addressing the data sparsity problem, Chernozhukov (2005), Chernozhukov and Fernández-Val (2011), Feigin and Resnick (1994), Knight (2001), Portnoy and Jurečková (1999), and Smith (1994) assume that the conditional quantile is linear. In particular, the extremal QTE considered in this paper is closely related to the linear extremal quantile regression (LEQR) investigated in Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011), but substantially differs in two aspects. First, the QTE considered in this paper has a causal interpretation by addressing the problem of missing counterfactuals, while the causal interpretation for the coefficient in the LEQR relies on the assumption that the treatment variable is exogenous at the tails. Second, I allow for heterogeneous quantile treatment response, while the linear model implies that two individuals, who are observationally equivalent, will have the same quantile treatment effect. In fact, since the QTE is an unconditional object, I do not assume the linearity of the conditional quantiles of the outcome variable given covariates.

The literature on extremal percentiles also addresses the data sparsity problem. See, for example, Bertail, Haefke, Politis, and White (2004), Bickel and Sakov (2008), and Dekkers and De Haan (1989). The key difference between these papers and mine is that I include additional covariates X and use propensity score $P(X)$ to correct the selection bias.

Last, my paper is related to the concept of drifting sequence asymptotics. This concept goes back to Pitman (1949) using Pitman drift to characterize power functions. Recently, the concept has been used in the context of weak instruments by, for example, Stock J (2008), Stock and Yogo (2005), and other various models by Andrews and Cheng (2012), Andrews and Cheng (2013), Chen, Ponomareva, and Tamer (2014), and Khan and Nekipelov (2013).

I establish the asymptotic properties for extremal QTE estimators when $\tau_n \to 0$. I find that there are two asymptotic distributions of the estimator of τ_n-th QTE, depending on how fast τ_n approaches zero. Following the terminology used in Chernozhukov (2005), I say τ_n is *intermediate* when $\tau_n \to 0$ and $\tau_n n \to \infty$. In this case, I show that the asymptotic distribution for the proposed estimator of QTE is still Gaussian. Again, following Chernozhukov (2005), when $\tau_n \to 0$, $\tau_n n \to k$, for some $k > 0$, I say τ_n is *extreme*. In this case, I show that the asymptotic distribution is non-Gaussian. For completeness, a quantile index is called regular if it is fixed strictly between zero and one. In this case, Firpo (2007) showed that the QTE estimator is asymptotically normal. Figure 1 summarizes the evolution of asymptotic behaviors of the estimator of QTE.

Figure 1: Asymptotic distribution over the quantile index

For inference, when the quantile index is intermediate, I show that the standard bootstrap confidence interval (CI) for the QTE estimator is consistent. For the extreme-order quantile case, I first prove that the conventional bootstrap CI does not control size. I then propose a resampling method that is uniformly consistent over a range of quantile indices. Last, by considering a linear combination of extreme QTE estimators with carefully chosen weights, I construct a consistent CI for the 0-th QTE without imposing additional restrictions or extrapolations.

To choose among different categories of quantile index, I propose a quantile-order-category-selection procedure similar to the identification-category-selection procedure used in Andrews and Cheng (2012). The difference here is that I have two thresholds while they only have one. When the quantile index is smaller than the first threshold, the extreme-order quantile asymptotic distribution is expected to approximate the finite sample distribution of the QTE estimator better than the normal approximation. In this case, I suggest using the new resampling CI developed in this paper to conduct inference. In simulation, I examine the performance of this threshold in 16 simulation designs with small, moderate, and large size samples. In all cases, I find that when the criterion is satisfied, the new resampling CI controls size while the standard bootstrap CI undercovers (that is, over-rejects) by as much as 18 absolute percentage points. When the quantile index is greater than the second threshold, I prove that the standard bootstrap CI is consistent. Last, when the quantile index is in between the first and second threshold, I construct a robust CI which is conservative.

My resampling inference method gives empirical researchers tools to estimate, infer, and test QTEs for low-rank populations. This method can be used in a number of economics applications. For instance, when focusing on the population of admitted university students, the college preparation index of low-rank students reflects the tolerance of low academic performance in the college's admission policy. My methods allow researchers to estimate the college preparation index gap between low-scoring minority and non-minority students while controlling for family background. This gap measures the magnitude of racial preference in college's admission. In another example, the extremely low or lower boundary of babies' birth weights represents the severity of adverse birth outcomes, which have been found to result in large economic costs. See, for example, Abrevaya (2001). My methods allow researchers to make inferences about the effect of maternal smoking on the lower tail of the distribution of infant birth weights.

The rest of the paper is organized as follows. Section 2 defines the parameters of interest, introduces additional notation, and provides relevant background on extreme value theory. Section 3 considers the asymptotic properties of the estimator for intermediate QTEs while Section 4 considers the asymptotic properties of the estimator for extreme QTEs. Section 5 establishes the inference theory and provides a step-by-step description of implementation. Sections 6 and 7 explore the finite sample properties of the new inferences methods through a simulation study, and applications, respectively. A supplement collects preliminary conditions for a high-level assumption in Section 4, numerical examples, all tables and figures in the Simulation section, additional simulation results, more detail on the Application section, and all theoretical proofs.

2 Definition, extreme value theory, and notation

First, I denote the outcomes for treated and control groups as Y_1 and Y_0, respectively. The treatment status is denoted as D, where $D = 1$ means treated and $D = 0$ means untreated.

The econometrician can only observe (Y, X, D) where $Y = Y_1 D + Y_0(1 - D)$, and X is a collection of confounders. The propensity score $P(D = 1|X = x)$ is denoted as $P(x)$. The parameters of interest are the τ-th QTE defined as

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

and the τ-th quantile treatment effect on treated (QTT) defined as

$$q_{|D=1}(\tau) := q_{1|D=1}(\tau) - q_{0|D=1}(\tau),$$

in which $q_j(\tau)$ and $q_{j|D=1}(\tau)$ denote the τ-th quantile of random variables Y_j and $Y_j|D = 1$, respectively.

Next, I introduce some extreme value theory, which will be used when I characterize the asymptotic theories in Section 3 and 4. The cumulative distribution function (CDF) F *belongs to the domain of attraction of generalized extreme value distributions* if there exist sequences $(\alpha_n)_{n\in\mathbb{N}}$, $(\beta_n)_{n\in\mathbb{N}}$ and a CDF G indexed by a parameter ξ, such that, for any independent draws $(U_1, ..., U_n)$ from F, $\alpha_n(\min(U_1, ..., U_n) - \beta_n)$ converges in distribution to G. Here, F belongs to the domain of attraction of generalized extreme value distributions with a parameter ξ called the extreme value (EV) index. Define $a(z) := \int_{s_l}^{z} F(v)dv/F(z)$ for some $z > s_l$, in which s_l is the lower end point of the support of U. In addition, for two generic functions $f_1(\cdot)$ and $f_2(\cdot)$, I write $f_1(z) \sim f_2(z)$ if

$$\frac{f_1(z)}{f_2(z)} \to 1, \text{ as } z \to s_l.$$

Then based on the value of ξ, F has three types of tail:

$$\begin{aligned}
&\text{type 1 tails } (\xi = 0): && \text{as } z \to s_l && F(z + va(z)) \sim F(z)e^v, && \forall v \in \mathbb{R},\\
&\text{type 2 tails } (\xi > 0): && \text{as } z \to s_l = -\infty && F(vz) \sim v^{-1/\xi}F(z), && \forall v > 0,\\
&\text{type 3 tails } (\xi < 0): && \text{as } z \to s_l > -\infty && F(vz) \sim v^{-1/\xi}F(z), && \forall v > 0.
\end{aligned}$$

For example, normal, T, and Beta distributions have type 1, 2, and 3 tails, respectively.

Last, I provide two weak convergence concepts this paper will rely on. $U_n \rightsquigarrow U$ indicates weak convergence as defined by Van der Vaart and Wellner (1996). When U_n and U are k-dimensional elements, the space of the sample path is $\mathbb{R}^k$ equipped with the Euclidean metric. When U_n and U are stochastic processes, the space of the sample path will be specified later in each different context. For this paper, the space is either $l^\infty(\{v \in \mathbb{R} : |v| < B\})$, for some positive B equipped with the sup norm or the Skorohod space $\mathcal{D}([-B, B])$, for some positive

B equipped with the Skorohod metric[1].

3 Intermediate quantile treatment effects

Theorems 3.1 and 3.2 establish the asymptotic theory for τ_n-th QTE when τ_n is intermediate. These theorems give the first main theoretical result of the paper: that the asymptotic distribution of the estimator of an intermediate QTE is still Gaussian. The asymptotic theory established here can be used to construct a uniform confidence band for both intermediate and extreme QTE, to estimate the EV index (which is analyzed in detail in Section 3.2), and to deal with the sample selection problem as in D'Haultfoeuille, Maurel, and Zhang (2015).

3.1 The main result

Recall the setup in Section 2. I further assume:

Assumption 1.
(1) (random sample): $\{Y_i, D_i, X_i\}_{i=1}^n$ *is i.i.d.*
(2) (unconfoundedness): $(Y_1, Y_0) \perp\!\!\!\perp D|X$.
(3) (common support): Supp(X)*, the support of* X*, is compact. For some* $c > 0$, $c < P(x) < 1 - c$, $\forall x \in$ *Supp*(X).

The unconfoundedness assumption states that the potential outcomes are independent of the treatment status conditional on additional covariates X. Although strong, this assumption has been widely used in both theoretical investigations and empirical studies. See, for example, Bitler et al. (2006), Chernozhukov et al. (2013), Firpo (2007), Hirano et al. (2003), Rosenbaum and Rubin (1983). For extremal QTEs, it is natural to first start with this unconfoundedness condition. When the quantile index is regular, that is, bounded away from 0 and 1, papers such as Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005), Chernozhukov and Hansen (2008), and Frölich and Melly (2013) extend the assumption to allow for endogenous treatment status and rely on an instrumental variable to correct the selection bias. Similar strategies can be applied here to the extremal quantile case. While important, I leave the problem of establishing the corresponding asymptotic theory to future research.

Assumption 2. τ_n *is intermediate. This is,*
(1) $\tau_n \to 0$ *as* $n \to \infty$.
(2) $\tau_n n \to \infty$ *as* $n \to \infty$.

[1] To differentiate, D is reserved for the binary treatment status and $\{\mathcal{D}_{i,j}\}_{i=1}^{\infty}$, $j = 0, 1$ are the sets of random variables defined in the limiting objective function in Section 4.

I define $\hat{q}(\tau_n)$, the estimator of the τ_n-th QTE, as $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$ and $\hat{q}_{|D=1}(\tau_n)$, the estimator of τ_n-th QTT, as $\hat{q}_{|D=1}(\tau_n) := \hat{q}_{1|D=1}(\tau_n) - \hat{q}_{0|D=1}(\tau_n)$. Under Assumption 1, Firpo (2007) found that the four quantiles $q_1(\tau)$, $q_0(\tau)$, $q_{1|D=1}(\tau)$, and $q_{0|D=1}(\tau)$ for any $\tau \in (0,1)$ are identified based on the following four moment equalities:

$$\mathbb{E}\left[\frac{D}{P(X)}\Big(\tau - \mathbb{1}\{Y \le q_1(\tau)\}\Big)\right] = 0,\ \mathbb{E}\left[\left(\frac{1-D}{1-P(X)}\right)\Big(\tau - \mathbb{1}\{Y \le q_0(\tau)\}\Big)\right] = 0,$$

and

$$\mathbb{E}\left[D(\tau - \mathbb{1}\{Y \le q_{1|D=1}(\tau)\})\right] = 0,\ \mathbb{E}\left[\frac{(1-D)P(X)}{1-P(X)}\Big(\tau - \mathbb{1}\{Y \le q_{0|D=1}(\tau)\}\Big)\right] = 0,$$

respectively.

Therefore, despite the extremal feature of the quantile index, the natural sample estimator $\hat{q}_1(\tau_n)$ for the τ_n-th quantile of Y_1 can be computed through an inverse propensity score weighted quantile regression:

$$\hat{q}_1(\tau_n) := \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)}(Y_i - q)(\tau_n - \mathbb{1}\{Y_i \le q\}). \tag{3.1}$$

Similarly, $\hat{q}_0(\tau_n)$, an estimator of the τ_n-th quantile of Y_0, can be computed as

$$\hat{q}_0(\tau_n) := \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \frac{1-D_i}{1-\hat{P}(X_i)}(Y_i - q)(\tau_n - \mathbb{1}\{Y_i \le q\}). \tag{3.2}$$

For estimating the QTT, $\hat{q}_{1|D=1}(\tau_n)$ and $\hat{q}_{0|D=1}(\tau_n)$ can be computed as

$$\hat{q}_{1|D=1}(\tau_n) := \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \frac{D_i}{\frac{1}{n}\sum_{i=1}^{n} D_i}(Y_i - q)(\tau_n - \mathbb{1}\{Y_i \le q\}),$$

and

$$\hat{q}_{0|D=1}(\tau_n) := \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \frac{1-D_i}{\frac{1}{n}\sum_{i=1}^{n} D_i} \frac{\hat{P}(X_i)}{1-\hat{P}(X_i)}(Y_i - q)(\tau_n - \mathbb{1}\{Y_i \le q\}).$$

Following Firpo (2007) and Hirano et al. (2003), $\hat{P}(X)$, the propensity score, is estimated by the sieve method of fitting a series logistic model. I denote the logistic CDF by $L(a) := \exp(a)/(1+\exp(a))$. $H_h(x) := (r_{1h}(x), \cdots, r_{hh}(x))'$ is a h-vector of power series of x. Then $\hat{P}(x) := L(H_h(x)'\hat{\pi}_h)$ with

$$\hat{\pi}_h := \arg\max_{\pi\in\mathbb{R}^h} \sum_{i=1}^{n} \left(D_i \log L(H_h(X_i)'\pi) + (1-D_i)\log(1 - L(H_h(X_i)'\pi))\right).$$

For brevity, the rest of the paper only considers the estimation of $\hat{q}_1(\tau_n)$, $\hat{q}_0(\tau_n)$, and $\hat{q}(\tau_n)$. The asymptotic results for $\hat{q}_{1|D=1}(\tau_n)$, $\hat{q}_{0|D=1}(\tau_n)$, and $\hat{q}_{|D=1}(\tau_n)$ can be derived in a similar manner.

Furthermore, instead of only one quantile index τ_n, I focus on a range of them. That is, $k\tau_n$, $k \in [\kappa_1, \kappa_2]$ for some fixed and known constants κ_1 and κ_2 such that $0 < \kappa_1 < \kappa_2 < \infty$. This is because I will derive a uniform asymptotic theory for the process $\{(\hat{q}_1(k\tau_n), \hat{q}_0(k\tau_n)) : k \in [\kappa_1, \kappa_2]\}$. For each k,

$$\hat{q}(k\tau_n) := \hat{q}_1(k\tau_n) - \hat{q}_0(k\tau_n)$$

where

$$\hat{q}_1(k\tau_n) := \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)}(Y_i - q)(k\tau_n - \mathbb{1}\{Y_i \leq q\})$$

and

$$\hat{q}_0(k\tau_n) := \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \frac{1 - D_i}{1 - \hat{P}(X_i)}(Y_i - q)(k\tau_n - \mathbb{1}\{Y_i \leq q\}).$$

The following sufficient regularity conditions are adapted from Assumptions A.1 and A.2 of Firpo (2007):

Assumption 3.
(1) The density of X is bounded above and bounded away from 0 over its support.
(2) The propensity score $P(x)$ is s-times continuously differentiable with all the derivatives bounded.
(3) $\mathbb{E}(k\tau_n - \mathbb{1}\{Y_j \leq q_j(k\tau_n)\}|x)$ is t-times continuously differentiable in x with all derivatives bounded by M_n uniformly over $(x, k) \in Supp(X) \times [\kappa_1, \kappa_2]$.
(4) The order of the series is $h = CN^c$ for some constants C and c such that $c < \frac{1}{6}$, $\tau_n n^{1+c(6-\frac{s}{r})} \to 0$, $\frac{M_n n^{(1-\frac{t}{r})}}{\tau_n} \to 0$, and $n^{11c-1}\tau_n \to 0$, where r is the dimension of X.

Assumptions 3(1) and 3(2) are common in the sieve estimation literature. Assumptions 3(3) and 3(4) are tailored to fit the special case in which the quantile index is intermediate and the derivative of the quantile varies with the sample size. In fact, the magnitude of M_n depends on the tail behavior of Y_j conditional on X. When the density of $Y_j|X$ vanishes on its lower tail, M_n decreases to zero. When the density of $Y_j|X$ diverges on its lower tail (such as a beta distribution with the first shape parameter less than 1), M_n diverges to infinity. Last, Assumptions 3(3) and 3(4) can be further relaxed by using the doubly robust estimation method as illustrated in Firpo and Rothe (2014).

Next, I impose regularity conditions on the tails of Y_1 and Y_0.

Assumption 4. *For $j = 0, 1$*
(1) Y_j, $Y_j|X$ are continuously distributed with density $f_j(\cdot)$ and $f_j(\cdot|X)$, respectively.

(2) $f_j(\cdot)$ is monotone at its lower tails.
(3) The CDF of Y_j belongs to the domain of attraction of generalized EV distributions with the EV index ξ_j.

These restrictions are mild. Assumption 4(1) is common in quantile regression literature. Assumption 4(2) refers to the tail of the distribution, which is satisfied by most well-known continuous distributions. Assumption 4(3) is a standard condition in extreme value theory and is satisfied by almost all continuous distributions.

Before stating the first main theoretical result of the paper, I introduce the normalizing factor $\lambda_{j,n}(k)$ for $\hat{q}_j(k\tau_n)$:

$$\lambda_{j,n}(k) := \sqrt{\frac{n}{k\tau_n}} f_j(q_j(k\tau_n)) \text{ for } j = 0, 1 \text{ and } k \in [\kappa_1, \kappa_2]. \tag{3.3}$$

Recall that for the regular quantile estimation, the convergence rate is $\sqrt{n}$ and the asymptotic variance is $\frac{\tau(1-\tau)}{f_j^2(q_j(\tau))}$. By moving the asymptotic standard deviation to the same side of the convergence rate, we obtain a normalizing factor

$$\sqrt{\frac{n}{\tau(1-\tau)}} f_j(q_j(\tau)).$$

Then letting $\tau := \tau_n \to 0$, we heuristically obtain the normalizing factor for the intermediate-order quantile estimators defined in (3.3) with $k = 1$.

Theorem 3.1. *If Assumptions 1–4 hold, then*

$$\Big(\lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)), \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n))\Big)$$

as a two-dimensional stochastic process indexed by k is asymptotically tight under the uniform metric. In addition, if there exist functions $H_1(k_1, k_2)$, $H_0(k_1, k_2)$, and $H_{10}(k_1, k_2)$ on $(k_1, k_2) \in [\kappa_1, \kappa_2] \times [\kappa_1, \kappa_2]$ such that, as $\tau_n \to 0$,

$$\frac{1}{\tau_n}\mathbb{E}\left[\frac{P(Y_1 \leq q_1(\min(k_1, k_2)\tau_n)|X)}{P(X)} - \frac{1 - P(X)}{P(X)} P(Y_1 \leq q_1(k_1\tau_n)|X) P(Y_1 \leq q_1(k_2\tau_n)|X)\right]$$
$$\to H_1(k_1, k_2),$$
$$\frac{1}{\tau_n}\mathbb{E}\left[\frac{P(Y_0 \leq q_0(\min(k_1, k_2)\tau_n)|X)}{1 - P(X)} - \frac{P(X)}{1 - P(X)} P(Y_0 \leq q_0(k_1\tau_n)|X) P(Y_0 \leq q_0(k_2\tau_n)|X)\right]$$
$$\to H_0(k_1, k_2),$$
$$\text{and} \quad \frac{1}{\tau_n}\mathbb{E}P(Y_0 \leq q_0(k_1\tau_n)|X) P(Y_0 \leq q_0(k_2\tau_n)|X) \to H_{10}(k_1, k_2),$$

then for $k \in [\kappa_1, \kappa_2]$,

$$\Big(\lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)),\ \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n))\Big) \rightsquigarrow \mathcal{B}(k)$$

where $\mathcal{B}(k)$ *is a Brownian bridge with covariance kernel*

$$\mathcal{H}(k_1,k_2) := \begin{pmatrix} \frac{H_1(k_1,k_2)}{\sqrt{k_1k_2}} & \frac{H_{1,0}(k_1,k_2)}{\sqrt{k_1k_2}} \\ \frac{H_{1,0}(k_1,k_2)}{\sqrt{k_1k_2}} & \frac{H_0(k_1,k_2)}{\sqrt{k_1k_2}} \end{pmatrix}.$$

Theorem 3.1 shows that the asymptotic distribution of the intermediate QTE estimator is still Gaussian, just as when the quantile index is regular. Intuitively, this is because for $j = 0,1$, $\hat{q}_j(\tau_n)$ can be interpreted as a cutoff for which the number of $\{Y_{i,j}\}_{i=1}^n$ below and above the cutoff are of the same order of $n\tau_n$ and $n(1-\tau_n)$, respectively. When τ_n is intermediate, both orders diverge to infinity, which is the same as the case in which τ is regular. Thus the shapes of asymptotic distributions under regular and intermediate-order quantile indices are the same.

The difference between the regular and intermediate-order quantile asymptotic properties is that for the intermediate case, nonparametrically estimating the propensity score $P(x)$ provides no additional information. From the proof of Theorem 3.1, the influence function for $\hat{q}_j$ is

$$\phi_{i,1,n} := \frac{1}{\sqrt{\tau_n}}\left[\frac{D_i}{P(X_i)}T_{i,1,n} - \frac{\mathbb{E}(T_{i,1,n}|X_i)}{P(X_i)}(D_i - P(X_i))\right]$$

where

$$T_{i,1,n} := \tau_n - \mathbb{1}\{Y_{i,1} \le q_1(\tau_n)\}.$$

In $\phi_{i,1,n}$, the second term

$$\frac{\mathbb{E}(T_{i,1,n}|X_i)}{P(X_i)}(D_i - P(X_i))$$

represents the information gain and is asymptotically negligible compared to the first term $\frac{D_i}{P(X_i)}T_{i,1,n}$.

I next turn to the asymptotic theory of $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$. From Theorem 3.1, I can make two observations: (1) the normalizing factors proposed in Theorem 3.1 are not feasible, and (2) the tail behaviors of Y_1 and Y_0, and thus the convergence rates for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$, are not necessarily the same. To address the first point, I follow Chernozhukov (2005) and build a feasible normalizing factor based on quantile difference with spacing parameter $m > 1$. To address the second point, I use the following assumption.

Assumption 5.

$$\frac{q_1(m\tau_n) - q_1(\tau_n)}{q_0(m\tau_n) - q_0(\tau_n)} \to \rho \in [0, +\infty].$$

Assumption 5 aims to bridge the normalizing factors of $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ by ρ. When $\rho = 0$, the convergence rate for $\hat{q}_0$ is slower so the estimation error of $\hat{q}_1(\tau_n)$ is asymptotically negligible. On the other hand, if $\rho = \infty$, $\hat{q}_0(\tau_n)$ is super-consistent compared to $\hat{q}_1(\tau_n)$ and thus can be treated as known. Last, when $\rho \in (0, \infty)$, the convergence rates for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are the same. For analytical inference, when τ_n is intermediate, ρ can be estimated by

$$\hat{\rho} = \frac{\hat{q}_1(m\tau_n) - \hat{q}_1(\tau_n)}{\hat{q}_0(m\tau_n) - \hat{q}_0(\tau_n)}.$$

Under Assumption 5, I define the feasible normalizing factor for $\hat{q}(\tau_n)$ as

$$\hat{\lambda}_n := \frac{\sqrt{n\tau_n}}{\max\Big\{(\hat{q}_1(m\tau_n) - \hat{q}_1(\tau_n)), (\hat{q}_0(m\tau_n) - \hat{q}_0(\tau_n))\Big\}}.$$

The next theorem shows that the intermediate QTE estimator is asymptotically normal with the feasible normalizing factor $\hat{\lambda}_n$.

Theorem 3.2. *Let* $C_1(\rho, m) := (\frac{1-m^{-\xi_1}}{\xi_1})^{-1}\frac{\rho}{\max(1,\rho)}$, $C_0(\rho, m) := (\frac{1-m^{-\xi_0}}{\xi_0})^{-1}\frac{1}{\max(\rho,1)}$[2], *and*

$$\Sigma_n := Var(C_1(\rho, m)\phi_{i,1,n} - C_0(\rho, m)\phi_{0,n,i})/\tau_n.$$

If Assumptions 1–5 hold, then

$$\Sigma_n^{-1/2}\hat{\lambda}_n(\hat{q}(\tau_n) - q(\tau_n)) \rightsquigarrow \mathcal{N}(0, 1).$$

Based on Theorem 3.2, I can conduct inference by estimating Σ_n and referring to the standard normal critical value.

In addition, the next theorem shows that the standard bootstrap inference for the intermediate QTE is consistent. Let $\hat{q}^*(\tau_n)$ be the estimator from the bootstrap sample and $\widetilde{C}_a^{nn}(\tau_n)$ be the a-th quantile of $\hat{q}^*(\tau_n) - \hat{q}(\tau_n)$ conditional on data. The two-sided $1-a$-th bootstrap

[2]Here I adapt the convention that $\frac{c}{\infty} = 0$, $\frac{c}{0} = \text{sign}(c)\infty$ for any real number c, and $\frac{1-m^{-\xi}}{\xi} = \log(m)$ when $\xi = 0$.

CI for any $a \in (0,1)$ can be written as

$$\text{CI}^{boot}(\tau_n) = \left(\hat{q}(\tau_n) - \widetilde{C}^{nn}_{1-a/2}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^{nn}_{a/2}(\tau_n)\right).$$

Theorem 3.3. *If Assumptions 1–5 hold, then*

$$\lim_{n\to\infty} P(q(\tau_n) \in CI^{boot}(\tau_n)) = 1 - a.$$

Falk (1991) has already proven the validity of bootstrap inference for the intermediate-order percentiles. For the regression case, Chernozhukov (2000) points out that the bootstrap inference is valid for linear intermediate-order quantile regressions. Recently, D'Haultfoeuille et al. (2015) proves that the bootstrap inference for intermediate-order quantile regression is valid in sample selection models. Here, I show that the bootstrap inference is also valid for the intermediate-order QTE estimator.

3.2 Estimation of the extreme value index

In this section, I focus on the estimation of EV indices ξ_j for $j = 0, 1$. A consistent estimator of the EV index will be used in Section 5.4 to construct a consistent CI for the 0-th QTE. The result is also of independent interest because it contributes to the statistics literature on estimating the EV index when the data are missing randomly conditional on covariates. Previous literature has focused on estimating the EV index for the observable Y. See Chapter 4 of Resnick (2007) for a textbook treatment on this topic. By contrast, here the potential outcomes (Y_1, Y_0) are not fully observed. Theorem 3.4 addresses this issue, proposes estimators of the EV indices for Y_1 and Y_0, and establishes their asymptotic properties.

The proposed EV index estimator follows the Pickands type as described in Section 4.5 of Resnick (2007). For some positive integer R, $\{w_r\}_{r=1}^R$ is a set of weights which sum to one. I estimate ξ_j, the EV index of Y_j, for $j = 0, 1$ by

$$\hat{\xi}_j := \sum_{r=1}^{R} \frac{-w_r}{\log(l)} \log\left(\frac{\hat{q}_j(ml^r\tau_n) - \hat{q}_j(l^r\tau_n)}{\hat{q}_j(ml^{r-1}\tau_n) - \hat{q}_j(l^{r-1}\tau_n)}\right),$$

in which l is some positive constant and τ_n is intermediate.

The intuition for the estimator is straightforward. If Y_j has EV index ξ_j, $q_j(\tau)$ behaves as

$\tau^{-\xi_j}$ as $\tau \to 0$. Then

$$\log\left(\frac{q_j(ml^r\tau_n) - q_j(l^r\tau_n)}{q_j(ml^{r-1}\tau_n) - q_j(l^{r-1}\tau_n)}\right)$$

behaves as

$$\log\left(\frac{(ml)^{-\xi_j} - l^{-\xi_j}}{(m)^{-\xi_j} - 1}\right) = -\xi_j \log(l).$$

The next theorem establishes the consistency and asymptotic normality of the estimator. For this purpose, I first extend the definition of the influence function in Theorem 3.1. In particular, for any positive constant k, write

$$\tilde{\phi}_{i,1,n}(k) := \frac{D_i}{P(X_i)} T_{i,1,n}(k) - \frac{\mathbb{E}(T_{i,1,n}(k)|X_i)}{P(X_i)}(D_i - P(X_i))$$

and

$$\tilde{\phi}_{i,0,n}(k) := \frac{1 - D_i}{1 - P(X_i)} T_{i,0,n}(k) + \frac{\mathbb{E}(T_{i,0,n}(k)|X_i)}{1 - P(X_i)}(D_i - P(X_i))$$

where

$$T_{i,1,n}(k) := k\tau_n - \mathbb{1}\{Y_{i,1} \le q_1(k\tau_n)\}$$

and

$$T_{i,0,n}(k) := k\tau_n - \mathbb{1}\{Y_{i,0} \le q_0(k\tau_n)\}, \text{ respectively.}$$

Theorem 3.4. *Under the assumptions in Theorem 3.1, for $j = 0, 1$,*
(1) $\hat{\xi}_j \xrightarrow{p} \xi_j$.
(2) In addition, if

$$\sqrt{\tau_n n}\left(\frac{-1}{\log(l)} \log\left(\frac{q_j(ml^r\tau_n) - q_j(l^r\tau_n)}{q_j(ml^{r-1}\tau_n) - q_j(l^{r-1}\tau_n)}\right) - \xi_j\right) \to 0$$

as $n \to \infty$ for all $r = 1, 2, \cdots, R$, then, for $b_r := \frac{(w_r - w_{r+1}) l^{r\xi_j}(1 - m^{-\xi_j})}{log(l)\xi_j}$ and $w_{R+1} = w_0 := 0$, I have

$$\sqrt{\tau_n n}(\hat{\xi}_j - \xi_j) = -\frac{1}{\sqrt{\tau_n n}} \sum_{i=1}^{n} \left(\sum_{r=0}^{R} b_r \left(\tilde{\phi}_{j,n,i}(ml^r) - \tilde{\phi}_{j,n,i}(l^r)\right)\right) + o_p(1).$$

Denote $\sigma_{j,n}^2 := Var\left(\sum_{r=1}^{R} b_r \left(\tilde{\phi}_{j,n,i}(ml^r) - \tilde{\phi}_{j,n,i}(l^r)\right)\right)/\tau_n$, then

$$\sqrt{\tau_n n}\sigma_j^{-1}(\hat{\xi}_j - \xi_j) \rightsquigarrow \mathcal{N}(0, 1).$$

This theorem proves that the Pickands type estimator of the EV index is consistent. Under an additional assumption, its asymptotic normality also holds. The latter result can be used to test the type of tails of both Y_1 and Y_0.

4 Extreme quantile treatment effects

Section 4.1 establishes asymptotic theory for the τ_n-th QTE when τ_n is extreme. It serves as the foundation for the inference theory built in Sections 5.1 and 5.2. In addition, I will infer the 0-th QTE by a linear combination of extreme QTEs. Hence the asymptotic theory also contributes to the inference of 0-th QTE in Section 5.4. Appendix A verifies Assumption 8, a high-level assumption for the asymptotic theories of extreme QTE established in Section 4.1. Section 4.2 considers the asymptotic distribution of the extreme QTE estimator with a feasible normalizing factor. This permits inference through a resampling method proposed in Section 5.2.

4.1 The main result

First, assume the following,

Assumption 6. *τ_n is extreme; that is,*
(1) $\tau_n \to 0$ as $n \to \infty$,
(2) $\tau_n n \to k$ for some positive constant k as $n \to \infty$.

Define the estimator $\hat{q}(\tau_n)$ of the τ_n-th QTE $q(\tau_n)$ as:

$$\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n) \tag{4.1}$$

where $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are computed from (3.1) and (3.2), respectively.

In fact, I use the same objective functions as those used to compute the regular and intermediate QTE. On the practical side, this implies that researchers can compute them in a unified manner without pre-specifying a category for the quantile index. On the theoretical side, I will show that the asymptotic behavior of $\hat{q}_j(\tau_n)$ is no longer normal compared to the ones with intermediate and regular quantile indices. This is because the number of observations below $q_j(\tau_n)$ are of the same order of magnitude of $\tau_n n$, which does not diverge to infinity (Assumption 6). Furthermore, from this assumption, I only need consistency of the propensity score estimator $\hat{P}(x)$.

Assumption 7. $\sup_{x \in Supp(X)} |\hat{P}(x) - P(x)| = o_p(1)$.

This assumption does not require that the convergence rate for the nonparametric propensity score estimator is faster than $n^{1/4}$, as usually assumed. See, e.g. Newey and McFadden (1994). The reason is similar to the non-normality of the limiting distribution: there are only a finite number of observations below the estimator of $\hat{q}_j(\tau_n)$, which are thus counted in the summation of (3.1) and (3.2). This prevents the accumulation of first order approximation error $\hat{P}(X_i) - P(X_i)$.

Next, I state a high-level assumption that determines the shape of the asymptotic distribution of the extreme QTE estimator.

Assumption 8. *For $j = 0, 1$,*
(1) $P(X \in \cdot|Y_j = y)$, the conditional distribution of X given $Y_j = y$, weakly converges to the CDF of a random variable $\mathcal{X}_j$ as $y \to q_j(0)$. The CDF of $\mathcal{X}_j$ is denoted as $P_j^+(\mathcal{X}_j \in \cdot|Y_j = q_j(0))$.
(2) $P_j^+(\mathcal{X}_j \in \cdot|Y_j = q_j(0))$ has finite mass points.
(3) Let $\mathcal{S}$ be the discontinuity of $P(x)$. Then $P_j^+(\mathcal{X}_j \in \mathcal{S}|Y_j = q_j(0)) = 0$.

Assumption 8(1) is high-level. Appendix A provides primitive sufficient conditions for Assumption 8(1) to hold. Appendix B contains more numerical illustrations. In general, $P_j^+(\mathcal{X}_j \in \cdot|_j = q_j(0))$ depends on the structure of conditional boundary of Y_j on X. The phenomenon that the asymptotic distribution depends on boundary conditions, is common in nonregular estimations. See, for example, Hirano and Porter (2003), Chernozhukov and Hong (2004), and Lee and Seo (2008). For Assumption 8(2), the number of mass points depends on the number of discrete minimizers of the conditional boundary of Y_j given X which is usually finite. Also, Assumption 8(2) holds when $\mathcal{X}_j$ is continuous, in which there is no mass point.

Theorem 4.1, the main theoretical result of this section, establishes the joint asymptotic distribution of $\hat{q}_j(\tau_n), j = 0, 1$ by showing that a normalized version of $\hat{q}_j(\tau_n), j = 0, 1$ weakly converges to the minimizer of an asymptotic objective function. I first state the normalized version of $\hat{q}_j(\tau_n), j = 0, 1$ below.

For $j = 0, 1$, the normalized versions of $\hat{q}_j(\tau_n)$ with or without centering are

$$\hat{Z}_{j,n}^c(k) := \alpha_{j,n}(\hat{q}_j(\tau_n) - q_j(\tau_n))$$

and

$$\hat{Z}_{j,n}(k) := \alpha_{j,n}(\hat{q}_j(\tau_n) - q_j^* - \beta_{j,n}),$$

respectively. Here, q_j^* is an auxiliary constant so that $U_j = Y_j - q_j^*$ has lower endpoint 0 or $-\infty$. In particular, if $q_j(0) > -\infty$, then $q_j^* = q_j(0)$, otherwise, q_j^* is arbitrary. The normalizing constants $(\alpha_{j,n}, \beta_{j,n})$ for $j = 0, 1$ are given by

$$
\begin{aligned}
&\text{for type 1 tails } (\xi_j = 0): && \alpha_{j,n} = 1/(a(F_{u_j}^{-1}(1/n))), && \beta_{j,n} = F_{u_j}^{-1}(1/n),\\
&\text{for type 2 tails } (\xi_j > 0): && \alpha_{j,n} = -1/(F_{u_j}^{-1}(1/n)), && \beta_{j,n} = 0,\\
&\text{for type 3 tails } (\xi_j < 0): && \alpha_{j,n} = 1/(F_{u_j}^{-1}(1/n)), && \beta_{j,n} = 0,
\end{aligned}
$$

in which F_{u_j} is the CDF of U_j.

Now I turn to the second part, the asymptotic objective function. The asymptotic objective function of the local parameter z takes the following form:

$$
-kz + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j}))l_\delta(\mathcal{J}_{i,j}, z), \tag{4.2}
$$

in which $W_1(d, P) = \frac{d}{p}$ and $W_0(d, P) = \frac{1-d}{1-p}$. To see the meaning of each term in (4.2), I denote, for $j = 0, 1$,

$$
\begin{aligned}
&\text{for type 1 tails } (\xi_j = 0): && h_j(l) = \exp(l), \text{ for } l \in \mathbb{R}, && \eta_j(k) = \log(k),\\
&\text{for type 2 tails } (\xi_j > 0): && h_j(l) = (-l)^{-1/\xi_j}, \text{ for } l < 0, && \eta_j(k) = (-k)^{-\xi_j},\\
&\text{for type 3 tails } (\xi_j < 0): && h_j(l) = (l)^{-1/\xi_j}, \text{ for } l > 0, && \eta_j(k) = k^{-\xi_j}.
\end{aligned}
$$

Then $\{\mathcal{E}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}\}$ is an i.i.d. sequence such that $\{\mathcal{E}_{i,1}, \mathcal{D}_{i,1}, \mathcal{X}_{i,1}\} \perp\!\!\!\perp \{\mathcal{E}_{i,0}, \mathcal{D}_{i,0}, \mathcal{X}_{i,0}\}$ and for $j = 0, 1$, $\mathcal{X}_{i,j}$ is governed by the law $P_j^+(\mathcal{X}_j \in \cdot | Y_j = q_j(0))$. $\mathcal{D}_{i,j}$ is Bernoulli distributed with success probability $P(\mathcal{X}_{i,j})$ conditional on $\mathcal{X}_{i,j}$ and $\mathcal{E}_{i,j}$ is standard exponentially distributed independently of both $(\mathcal{X}_{i,j}, \mathcal{D}_{i,j})$. In addition, $\mathcal{J}_{i,j} := h_j^{-1}(\sum_{l=1}^{i} \mathcal{E}_{l,j})$ and $l_\delta(u, v) := \mathbb{1}\{u < v\}(v - u) - \mathbb{1}\{u \leq -\delta\}(-\delta - u)$ for an arbitrary $\delta > 0$. The same function of $l_\delta(u, v)$ is first used in Chernozhukov (2005).

Assumption 9. *For $j = 0, 1$ and a generic fixed constant $k > 0$,*

$$
-kz + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j}))l_\delta(\mathcal{J}_{i,j}, z)
$$

has a unique minimizer almost surely.

Assumption 9 indicates that the asymptotic objective function has a unique minimizer which is necessary for applying the argmin theory. This type of assumption is common in non-regular estimation literature. See, for example, Chernozhukov and Fernández-Val (2011), Chernozhukov and Hong (2004), and Lee and Seo (2008). Lemma E.6 provides a sufficient condition for this assumption to hold. In general, the assumption holds when $\mathcal{X}_j$ is absolutely continuous. If $\mathcal{X}_j$ has a mass point at x_0, the sufficient condition requires that $kP(x_0)$ is not an integer, where $P(x)$ is the propensity score. Since integers are sparse on the real line, I consider this sufficient condition mild.

Theorem 4.1. *If Assumptions 1, 4, 6–8 hold, there exist* κ_1 *and* κ_2 *such that* $0 < \kappa_1 < \kappa_2 < \infty$ *and* (κ_1, κ_2) *satisfy Assumption 9, then* $(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k)) \rightsquigarrow (Z_{1,\infty}(k), Z_{0,\infty}(k))$ *in* $\mathcal{D}^2([\kappa_1, \kappa_2])$*, where*

$$(Z_{1,\infty}(k), Z_{0,\infty}(k)) := \underset{(z_1,z_0)\in\mathbb{R}^2}{\arg\min} \sum_{j=0,1} \left[-kz_j + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j})) l_\delta(\mathcal{J}_{i,j}, z_j) \right].$$

In addition, $(\hat{Z}^c_{1,n}(k), \hat{Z}^c_{1,n}(k)) \rightsquigarrow (Z^c_{1,\infty}(k), Z^c_{0,\infty}(k)) := (Z_{1,\infty}(k) - \eta_1(k), Z_{0,\infty}(k) - \eta_0(k))$ *in* $\mathcal{D}^2([\kappa_1, \kappa_2])$*.*

The immediate corollary of Theorem 4.1 is the finite dimensional convergence. Due to the lack of continuity of the sample path of $(Z_{1,\infty}(\cdot), Z_{0,\infty}(\cdot))$, the projection mapping is only continuous when index k is not at the discontinuity.

Corollary 4.1. *If the assumptions in Theorem 4.1 hold and Assumption 9 is satisfied for* $k \in \{k_l\}_{l=1}^L$*, then*

$$\begin{aligned}(\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l))_{l=1}^L \rightsquigarrow & (Z_{1,\infty}(k_l), Z_{0,\infty}(k_l))_{l=1}^L \\ & := \underset{(z_{1,l},z_{0,l})_{l=1}^L}{\arg\min} \sum_{j=0,1} \sum_{l=1}^{L} \left\{ -k_l z_{j,l} + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j})) l_\delta(\mathcal{J}_{i,j}, z_{j,l}) \right\},\end{aligned}$$

and

$$(\hat{Z}^c_{1,n}(k_l), \hat{Z}^c_{0,n}(k_l))_{l=1}^L \rightsquigarrow (Z^c_{1,\infty}(k_l), Z^c_{0,\infty}(k_l))_{l=1}^L := (Z_{1,\infty}(k_l) - \eta_1(k_l), Z_{0,\infty}(k_l) - \eta_0(k_l))_{l=1}^L.$$

First, Theorem 4.1 and Theorem 3.1 (for the intermediate-order quantile), along with Theorem 1 in Firpo (2007) (for the regular quantile), characterizes the evolution of the asymptotic distribution of the QTE estimator when the quantile index ranges from 0 to 1. Starting with the regular quantile, the asymptotic distribution is normal. Estimating the unknown propensity score provides additional information. When the quantile index is intermediate, the shape of the asymptotic distribution remains normal, but the additional information from estimating the propensity score becomes asymptotically negligible. When the quantile index moves even closer to the origin so that it is extreme, the shape of the asymptotic distribution becomes non-Gaussian, and the information from estimating the propensity score is asymptotically negligible. Figure 1 in Section 1 shows the evolution of the asymptotic distribution over quantile index τ.

Second, I do not impose any parametric restriction on the conditional quantile of Y_j given X, in contrast to Chernozhukov (2005), which considered linear extreme-order quantile regressions. The parameters considered in linear quantile regressions are conditional objects, while QTEs in this paper are unconditional objects. In order to deal with conditional quantiles, Chernozhukov (2005) proposed an innovative solution: use the asymptotic independence between residuals and covariates X at tails in addition to linearity to regulate the conditional tail behavior. On the other hand, in this paper, I only need Assumption 8, which is weaker than the combination of linearity and asymptotic independence. Appendix A verifies Assumption 8 under three different conditional boundary conditions.

Third, Theorem 4.1 has shown that $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are asymptotically independent because, by construction, $\{\mathcal{J}_{i,1}, \mathcal{X}_{i,1}, \mathcal{D}_{i,1}\}_{i\geq 1} \perp\!\!\!\perp \{\mathcal{J}_{i,0}, \mathcal{X}_{i,0}, \mathcal{D}_{i,0}\}_{i\geq 1}$. Thus the joint asymptotic distribution of $(\hat{q}_0(\tau_n), \hat{q}_1(\tau_n))$ is fully characterized by the marginals. In Appendix B, I compute the marginal distribution of $\hat{q}_1(\tau)$ under various boundary conditions.

Fourth, directly computing the critical value of the asymptotic distribution of $\hat{q}(\tau_n)$ is infeasible. Note that the ultimate parameter of interest is $q(\tau_n) := q_1(\tau_n) - q_0(\tau_n)$. Although the joint asymptotic distribution of $(\hat{q}_0(\tau_n), \hat{q}_1(\tau_n))$ has been established by Theorem 4.1, the convergences depend on the tails of Y_1 and Y_0 and are hard to be estimated consistently. Furthermore, the asymptotic distributions of $\hat{q}_0(\tau_n)$ and $\hat{q}_1(\tau_n)$ are complicated and depend on unknown boundary conditions. In Section 5, I propose to use a b out of n bootstrap with or without replacement to construct a CI and to draw inferences.

Last, as pointed out in the first remark after Theorem 4.1, the shape of the asymptotic distribution changes as the quantile index moves from the intermediate region to the extreme region. So the extreme-order quantile asymptotics proposed in Theorem 4.1 are valid only if $k = \tau_n n$ is not large, i.e., $\tau_n \leq \tau_{n,1}$. I will explain $\tau_{n,1}$ in Section 5.3.

4.2 Feasible normalizing factor

This section considers the next missing piece needed for the resampling inference method: the feasible normalizing factor. I propose a feasible normalizing factor that is not a consistent estimator but has the same order of magnitude as the infeasible one and establish the corresponding asymptotic theory.

The normalizing factor for the τ_n-th QTE estimator when τ_n is extreme has not been obvious. First, the estimator of τ_n-th QTE is $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$. Due to the different tail behaviors, the normalizing factors for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are not necessarily the same.

In addition, by Theorem 4.1, the normalizing factors for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are first-order statistics that are unknown and hard to estimate.

I propose the following feasible normalizing factor:

$$\hat{\alpha}_n := \frac{\sqrt{\tau_{n,l'} n}}{\max\left\{\hat{q}_1(m\tau_{n,l'}) - \hat{q}_1(\tau_{n,l'}), \hat{q}_0(m\tau_{n,l'}) - \hat{q}_0(\tau_{n,l'})\right\}}, \tag{4.3}$$

where m is a spacing parameter and $\tau_{n,l'}$ is a quantile index selected by the researcher. How to choose $\tau_{n,l'}$ will be discussed later. The feasible normalizing factor uses the smaller of the two factors for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$. In addition, the proposed factor has the same order but is not a consistent estimator of the infeasible order statistic. This is possible by the following assumption:

Assumption 10.
(1) $\tau_{n,l'} n \to k_{l'}$.
(2) $k_{l'}$ *satisfies the condition in Lemma E.7 as well as Assumption 9.*
(3) Both Y_1 *and* Y_0 *have type 2 or 3 tails.*

Assumption 10(3) is valid in many economic applications. First, type 2 or 3 tails are also called Pareto-type tails, which are prevalent in economic data such as wealth and incomes, as argued in Section 2.2 of Chernozhukov and Fernández-Val (2011). Second, the assumption holds if and only if the EV index is non-zero, which is testable based on Theorem 3.4. In practice, it implies that the CDF of the two potential outcomes decay polynomially as $\tau \to 0$. Last, 10(3) implies that the feasible and infeasible normalizing factors are of the same order of magnitude. To see this, with $n \to \infty$, I have

$$\frac{1}{\alpha_{j,n}(q_j(m\tau_{n,l'}) - q_j(\tau_{n,l'}))} = \frac{F_{u_j}^{-1}(\frac{1}{n})}{q_j(m\tau_{n,l'}) - q_j(\tau_{n,l'})} \sim \frac{k_{l'}^{\xi_j}}{m^{-\xi_j} - 1}.$$

Theoretically, the choice of $\tau_{n,l'}$ in $\hat{\alpha}_n$ does not impact the asymptotic validity of the normalizing factor. However, in finite samples, this choice involves a trade-off between bias and variance. If $n\tau_{n,l'}$ is small, there are fewer observations used for estimating $\hat{q}_j(\tau_{n,l'})$, which produces a large variance. On the other hand, if $n\tau_{n,l'}$ is large, it can introduce bias in two ways. First, as the increase of $n\tau_{n,l'}$, the estimation error of the propensity score will accumulate and contaminate the CI. In addition, since I use a b out of n bootstrap method with subsample size b to construct the CI, if $mn\tau_{n,l'}/b$ is large, then this quantile index cannot be interpreted as extreme-order. Both imply that the EV asymptotic theory is not suitable. To address all the issues aforementioned, the rule of thumb I use to choose the index $\tau_{n,l'}$

is $\tau_{n,l'} = \min(\frac{C_1}{n}, \frac{C_2 b}{mn})$. The simulation study in Appendix C.1 shows that this rule with $(C_1, C_2) = (10, 0.1)$ performs well in finite samples.

Similar to Assumption 5, I have to bridge the two normalizing factors.

Assumption 11. $\frac{q_1(\frac{mk_{l'}}{n}) - q_1(\frac{k_{l'}}{n})}{q_0(\frac{mk_{l'}}{n}) - q_0(\frac{k_{l'}}{n})} \to \rho \in [0, \infty]$.

Since ρ can be 0 and ∞, the assumption incorporates the case when one convergence rate dominates another.

The next theorem characterizes the weak convergence of the extreme QTE estimator with the feasible normalizing factor.

Theorem 4.2. *The assumptions in Theorem 4.1 and Assumptions 10 and 11 hold. Denote*

$$\tilde{\rho} := k_{l'}^{\xi_0 - \xi_1} \frac{m^{-\xi_1} - 1}{\rho(m^{-\xi_0} - 1)} \quad \text{and} \quad \hat{Z}_n^c(k) := \hat{\alpha}_n(\hat{q}(\tau_n) - q(\tau_n))$$

for any $\tau_n n \to k$. Then for $k_{l'}$ fixed,

$$\hat{Z}_n^c(k) \rightsquigarrow Z_\infty^c(k) \text{ in } \mathcal{D}[\kappa_1, \kappa_2],$$

in which

$$Z_\infty^c(k) := \frac{\sqrt{k_{l'}}(Z_{1,\infty}^c(k) - \tilde{\rho} Z_{0,\infty}^c(k))}{\max\left\{ Z_{1,\infty}(mk_{l'}) - Z_{1,\infty}(k_{l'}), \tilde{\rho}(Z_{0,\infty}(mk_{l'}) - Z_{0,\infty}(k_{l'})) \right\}}.$$

An immediate corollary from the above theorem is the weak convergence of a linear combination of $\hat{Z}_n^c(k)$'s. In Section 5.4, I use the linear combination of extreme QTE estimators to construct a point estimator and a CI for the 0-th QTE. Proposition 4.2 establishes the theoretical foundation for this construction. The key here is to choose a proper set of weights $\{\hat{r}_l\}_{l=1}^L$. More details can be found in Section 5.4.

Assumption 12. *Let $\{\hat{r}_l\}_{l=1}^L$ be a set of weights that can be random, and*
(1) $\sum_{l=1}^L \hat{r}_l = 1$,
(2) $\hat{r}_l \xrightarrow{p} r_l$ for all $l = 1, \cdots, L$ and $\{r_l\}_{l=1}^L$ a set of constant real numbers.
(3) $\tau_{n,l} n \to k_l$ where $\{k_l\}_{l=1}^L$ satisfies Assumption 9.

Corollary 4.2. *The assumptions in Theorem 4.2 and Assumption 12 hold. Then*

$$\hat{\alpha}_n \left(\sum_{l=1}^L \hat{r}_l \hat{q}(\tau_{n,l}) - \sum_{l=1}^L r_l q(\tau_{n,l}) \right) \rightsquigarrow \sum_{l=1}^L r_l Z_\infty^c(k_l).$$

5 Inference

This section establishes inference theory for extreme QTE estimators that I then apply in Section 7. Section 5.1 shows that the conventional bootstrap CI does not control size. Section 5.2 establishes a new uniformly consistent CI over a range of quantile indices. Section 5.3 considers a robust confidence interval over different categories of quantile indices. Section 5.4 proposes to infer the 0-th QTE by combining a set of extreme QTE estimators with carefully chosen weights. Last, Section 5.5 considers the two-sample inference.

5.1 Inconsistency of the standard bootstrap inference method

I first define the bootstrap estimator with proper normalizations:

$$
\begin{aligned}
(\hat{Z}^*_{1,n}(k), \hat{Z}^*_{0,n}(k)) := \underset{(z_1,z_2)\in\mathbb{R}^2}{\arg\min} \sum_{j=0,1} \Bigg\{ &- \sum_{i=1}^{n} \left(\sum_{l=1}^{n} \mathbb{1}\{I_l = i\} \right) W_j(D_i, \hat{P}(X_i)) \tau_n z_j \\
&+ \sum_{i=1}^{n} \left(\sum_{l=1}^{n} \mathbb{1}\{I_l = i\} \right) W_j(D_i, \hat{P}(X_i)) l_\delta(\alpha_{j,n}(U_{i,j} - q_j(0)), z_j) \Bigg\}
\end{aligned}
$$

in which $\hat{Z}^*_{j,n}(k) := \alpha_{j,n}(\hat{q}^*_{j,n}(\tau_n) - q_j(0))$ for $\tau_n n \to k$. $\hat{q}^*_{j,n}(\tau_n)$ is the point estimator computed from (3.1) and (3.2) using the bootstrap sample. Similarly, $\hat{Z}^{c*}_{j,n}(k) := \alpha_{j,n}(\hat{q}^*_{j,n}(\tau_n) - q_j(\tau_n))$. Here, $(I_{n,1}, I_{n,2}, \cdots, I_{n,n})$ is a multinomial vector with parameter n and probabilities $(\frac{1}{n}, \cdots, \frac{1}{n})$. The data is denoted as Φ_n and $(I_{n,1}, I_{n,2}, \cdots, I_{n,n}) \perp \Phi_n$.

Theorem 5.1. *The Assumptions in Theorem 4.1 hold. Then*

$$
(\hat{Z}^*_{1,n}(k), \hat{Z}^*_{0,n}(k)) \rightsquigarrow (Z^*_{1,\infty}(k), Z^*_{0,\infty}(k)),
$$

in which

$$
(Z^*_{1,\infty}(k), Z^*_{0,\infty}(k)) := \underset{(z_1,z_0)\in\mathbb{R}^2}{\arg\min} \sum_{j=0,1} \left[-kz_j + \sum_{i=1}^{\infty} \Gamma_{i,j} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j})) l_\delta(\mathcal{J}_{i,j}, z_j) \right]
$$

and

$$
(\hat{Z}^{c*}_{1,n}(k), \hat{Z}^{c*}_{1,n}(k)) \rightsquigarrow (Z^{c*}_{1,\infty}(k), Z^{c*}_{0,\infty}(k)) := (Z^*_{1,\infty}(k) - \eta_1(k), Z^*_{0,\infty}(k) - \eta_0(k)).
$$

Here, $\{\mathcal{J}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}\}_{i\geq 1, j=0,1}$ *are the same as in Theorem 4.1 and* $\{\Gamma_{i,j}\}_{i\geq 1}$ *is a sequence of i.i.d. Poisson random variables with unit mean such that*

$$\{\Gamma_{i,j}\}_{i\geq 1, j=0,1} \perp \{\mathcal{J}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}\}_{i\geq 1, j=0,1}$$

and $\Gamma_{i,1} \perp \Gamma_{i,0}$.

The asymptotic distribution of the bootstrap estimator of extreme QTE is different from the original estimator. Compared with the limiting process in Theorem 4.1, there is an additional Poisson random variable term. Since the asymptotic objective function is not quadratic, $Z^*_{j,\infty}$, $j = 0, 1$ are not linear in $\Gamma_{i,j}$ which causes the invalidity of the bootstrap inference. Furthermore, due to the lack of linear expansion of the estimator, $\hat{Z}^*_{j,n}(k) - \hat{Z}_{j,n}(k)$ does not share the same limiting distribution with $\hat{Z}_{j,n}(k)$.

The intuition behind the invalidity of standard bootstrap is similar to the case of order statistics. When there are no missing counterfactuals or the data are fully missing at random, the extreme-order quantile estimator considered in this paper degenerates to an order statistic. However, Bickel and Freedman (1981) have already shown that the standard n out of n bootstrap inference is not consistent for order statistics.

5.2 Consistency of the b out of n bootstrap inference

We have just seen that the conventional bootstrap CI is inconsistent. In this section, I establish the uniform consistency of a b out of n bootstrap CI (BN-CI) both with and without replacement in which b is the subsample size with $b \to \infty$, $\frac{b}{n} \to 0$. This third main theoretical result of the paper allows empirical researchers to do uniformly consistent inferences over a range of extreme-order quantile indices. Section 6 confirms the consistency of BN-CI as well as the inconsistency of NN-CI through an extensive numerical study.

Let the quantile index for the subsample be τ_b. The key insight for the b out of n bootstrap inference is to align $\tau_b b$ with $\tau_n n$. Theorem 4.2 shows that the asymptotic distribution of the τ_n-th QTE is indexed by k. Letting $\tau_b b = \tau_n n = k$ ensures that the subsample estimator can mimic the same asymptotic distribution of the full sample estimator.

I consider the b out of n bootstrap inference for extreme QTEs both with and without replacement. Not allowing for replacement (subsampling), Bertail et al. (2004) studied the validity of inference for extreme-order statistics without covariates. Chernozhukov and Fernández-Val (2011) considered a similar inference procedure in linear extreme-order quantile regressions. Allowing for replacement, Bickel and Sakov (2008) considered the b out of n bootstrap inference in extreme-order statistics without covariates. Theorem 5.2 proves the consistency of b out of n bootstrap inference both with and without replacement for the extreme QTE.[3]

Before stating the main theorem of this section, I introduce the resampling version of the feasible normalizing factor for the subsample:

$$\hat{\alpha}_b^* := \frac{\sqrt{\tau_{b,l'} b}}{\max\left\{\hat{q}_1^*(m\tau_{b,l'}) - \hat{q}_1^*(\tau_{b,l'}), \hat{q}_0^*(m\tau_{b,l'}) - \hat{q}_0^*(\tau_{b,l'})\right\}}$$

where $\tau_{b,l'} b = \tau_{n,l'} n$, $\tau_{n,l'}$ satisfies Assumption 10. Then, the normalized estimator is

$$\hat{Z}_n^{c*}(k) := \hat{\alpha}_b^*(\hat{q}^*(\tau_b) - \hat{q}(\tau_b)).$$

In the above two equations, $\hat{q}^*(\tau) := \hat{q}_1^*(\tau) - \hat{q}_0^*(\tau)$ where $\hat{q}_j^*(\tau)$ is computed by (3.1) and (3.2) with τ_n replaced by $\tau = \tau_b$ or $\tau_{b,l'}$ and using only the data from the subsample, which is generated either with or without replacement. Without the star symbol, $\hat{q}(\tau_b) := \hat{q}_1(\tau_b) - \hat{q}_0(\tau_b)$ where $\hat{q}_j(\tau_b)$ is computed by (3.1) and (3.2) with τ_n replaced by τ_b and using the full sample.

Theorem 5.2. *If the assumptions in Theorem 4.2 hold and as $n \to \infty$, $\frac{b}{n} \to 0$, $b \to \infty$ at a polynomial rate in n, then $\hat{Z}_n^{c*}(k) \rightsquigarrow Z_\infty^c(k)$ in $\mathcal{D}([\kappa_1, \kappa_2])$.*

Theorem 5.2 builds the theoretical foundation for constructing the uniform confidence band for the extreme QTE over $k \in [\kappa_1, \kappa_2]$, in which κ_1, κ_2 are not at the discontinuity of the limiting process with probability 1. To construct a uniformly consistent confidence band, I next want to studentize the process $\hat{Z}_n^{c*}(k)$. When the limiting process is Gaussian, it is common to studentize the process by the point-wise standard deviation first and then to approximate the studentized limit. Here, I consider the same studentization in the non-Gaussian case. Let $S_n(k)$ and $\sigma(k)$ be the feasible and infeasible studentizing factors.

Assumption 13. *For a (random) scale function $S_n(k)$, there exists $\sigma(k) > 0$, a deterministic function of k, such that*

$$\sup_{k \in [\kappa_1, \kappa_2]} \left| \frac{S_n(k)}{\sigma(k)} - 1 \right| = o_p(1).$$

In addition, with probability approaching one, $\sigma(k)$, $S_n(k)$ are both continuous in k and uniformly bounded and bounded away from zero over $k \in [\kappa_1, \kappa_2]$.

$S_n(k)$ can be $S_n(k) := 1$ or $S_n(k) := k^{-\hat{\xi}_1} + k^{-\hat{\xi}_0}$ with corresponding $\sigma(k) := 1$ or $\sigma(k) := k^{-\xi_1} + k^{-\xi_0}$, respectively. In the later case, ξ_j, $j = 0, 1$ are unknown. So I replace them by their consistent estimators $\hat{\xi}_j$, $j = 0, 1$. The choice of studentizing factors will not affect the

[3] I suggest using the b out of n bootstrap with replacement because it performs better in simulation.

size of the uniform confidence band, but will rather affect its power. Unlike the Gaussian limit in which using $\sigma(k)$ as the point-wise standard deviation is natural, the best choice for the studentizing factor in this non-Gaussian case is still an open question and should be the focus of future research.

Corollary 5.1. *Let $\widehat{C}_{1-a}$ denote the $(1-a)$-th quantile of $\max_{k\in[\kappa_1,\kappa_2]} |\hat{Z}_n^{c*}(k)/S_n(k)|$. If the assumptions in Theorem 5.2 and Lemma E.7 as well as Assumption 13 hold, then*

$$P\left(q\left(\frac{k}{n}\right) \in \left[\hat{q}\left(\frac{k}{n}\right) - S_n(k)\widehat{C}_{1-a}/\hat{\alpha}_n, \hat{q}\left(\frac{k}{n}\right) + S_n(k)\widehat{C}_{1-a}/\hat{\alpha}_n\right] : k \in [\kappa_1, \kappa_2]\right) \to 1-a.$$

Let $\{k_l\}_{l=1}^L$ be a fine grid. $\tau_{n,l} = \frac{k_l}{n}$, $\tau_{b,l} = \frac{k_l}{b}$, $\tau_{n,l'} = \frac{k_{l'}}{n}$, and $\tau_{b,l'} = \frac{k_{l'}}{b}$. The number of subsamples is B_n, which is as large as computationally possible. Researchers can compute the uniform confidence band (CB_α) based on the following procedure.

1. Compute $\hat{q}(\tau_{n,l})$ and $\hat{q}(\tau_{b,l})$ as in (4.1). Compute $\hat{\alpha}_n$, $S_n(k)$, and the propensity score using the full sample.

2. For the i-th subsample, compute $\hat{q}^*(\tau_{b,l})$ for $l = 1, \cdots, L$ as in (4.1). Denote

$$\hat{\alpha}_b^* := \frac{\sqrt{\tau_{b,l'} b}}{\max\left\{\hat{q}_1^*(m\tau_{b,l'}) - \hat{q}_1^*(\tau_{b,l'}), \hat{q}_0^*(m\tau_{b,l'}) - \hat{q}_0^*(\tau_{b,l'})\right\}}$$

   where for $j = 0, 1$, $\hat{q}_j^*(\tau)$ is computed as in (3.1) and (3.2), respectively, using the subsample data and the propensity score estimated in the first step. Denote

$$\widehat{V}_{i,b}^* := \max_{l=1,\cdots,L} \hat{\alpha}_b^* \left|\left(\hat{q}^*(\tau_{b,l}) - \hat{q}(\tau_{b,l})\right)/S_n(k)\right|.$$

3. Repeat the above step for $i = 1, \cdots, B_n$. Compute $\widehat{C}_{1-a}$ as the $(1-a)$-th quantile of the $\{\widehat{V}_{i,b}^*\}_{i=1}^{B_n}$.

4. $\mathrm{CB}_\alpha = \left\{\left[\hat{q}\left(\frac{k}{n}\right) - S_n(k)\widehat{C}_{1-a}/\hat{\alpha}_n, \hat{q}\left(\frac{k}{n}\right) + S_n(k)\widehat{C}_{1-a}/\hat{\alpha}_n\right] : k \in [\kappa_1, \kappa_2]\right\}$.

Next I consider the b out of n inference for a linear combination of extreme QTEs. By carefully choosing the weights, in Section 5.4, I show that the linear combination of extreme QTE estimators can be utilized to infer the 0-th QTE.

Let C_a be the $a-$th quantile of $\sum_{l=1}^{L} \gamma_r Z_\infty^c(k_l)$ and $\widehat{C}_a$ be the a-th quantile of

$$\hat{\alpha}_b^* \left(\sum_{l=1}^{L} \hat{\gamma}_l \hat{q}^*(\tau_{b,l}) - \sum_{l=1}^{L} \hat{\gamma}_l \hat{q}(\tau_{b,l}) \right).$$

Given that $\sum_{l=1}^{L} \gamma_r Z_\infty^c(k_l)$ is continuously distributed,[4] Proposition 5.1 shows that $\widehat{C}_a$ is a consistent estimator of C_a. Denote

$$\sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \widehat{C}_{0.5}/\hat{\alpha}_n \text{ and } \left[\sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \widehat{C}_{1-a/2}/\hat{\alpha}_n, \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \widehat{C}_{a/2}/\hat{\alpha}_n \right]$$

the median-unbiased estimator and a $(1-a) \times 100\%$ CI for $\hat{q}(\tau)$, respectively.

Proposition 5.1. *Under the assumptions in Theorem 5.2 and Assumption 12, I have*

$$\hat{\alpha}_b^* \left(\sum_{l=1}^{L} \hat{\gamma}_l \hat{q}^*(\tau_{b,l}) - \sum_{l=1}^{L} \hat{\gamma}_l \hat{q}(\tau_{b,l}) \right) \rightsquigarrow \sum_{l=1}^{L} \gamma_r Z_\infty^c(k_l), \tag{5.1}$$

$$\lim_{n\to\infty} P \left(\sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \widehat{C}_{0.5}/\hat{\alpha}_n \le \sum_{l=1}^{L} r_l q(\tau_{n,l}) \right) = 0.5, \tag{5.2}$$

and

$$\lim_{n\to\infty} P \left(\sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \widehat{C}_{1-a/2}/\hat{\alpha}_n \le \sum_{l=1}^{L} r_l q(\tau_{n,l}) \le \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \widehat{C}_{a/2}/\hat{\alpha}_n \right) = 1-a. \tag{5.3}$$

(5.1) shows the weak convergence of the linear combination of extreme QTE estimators, (5.2) shows the median-unbiased estimator is asymptotically median-unbiased, and (5.3) implies that the CI asymptotically controls size.

To implement, let B_n denote the number of subsamples. I use the following steps to compute $\widehat{C}_a$.

1. Compute $\{\hat{r}_l\}_{l=1}^{L}$, $\hat{q}(\tau_{b,l})$, $\hat{q}(\tau_{n,l})$, and the propensity score estimator $\hat{P}(x)$ using the full sample.

2. For the i-th subsample, compute $\hat{q}_{i,b}^*(\tau_{b,l})$ for $l = 1, \cdots, L$ as in (4.1). Denote

$$\hat{\alpha}_b^* := \frac{\sqrt{\tau_{b,l'} b}}{\max\left\{ \hat{q}_1^*(m\tau_{b,l'}) - \hat{q}_1^*(\tau_{b,l'}), \hat{q}_0^*(m\tau_{b,l'}) - \hat{q}_0^*(\tau_{b,l'}) \right\}}$$

[4] This is shown in Lemma E.7 in the appendix.

where for $j = 0, 1$, $\hat{q}_j^*(\tau_b)$ is computed as in (4.1) for each subsample. Denote

$$\widehat{V}_{i,b}^* := \hat{\alpha}_b^* \left[\sum_{l=1}^{L} \hat{r}_l \left(\hat{q}^*(\tau_{b,l}) - \hat{q}(\tau_{b,l}) \right) \right].$$

3. Repeat the above step for $i = 1, \cdots, B_n$. Compute $\widehat{C}_{1-a}$ as the $(1-a)$-th quantile of the $\{\widehat{V}_{i,b}^*\}_{i=1}^{B_n}$.

When $L = 1$, I can use this procedure to construct the CI for $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$, the estimator of the τ_n-th QTE. The finite sample performance of the CI is examined in Section 6.

5.3 A robust confidence interval

The inference methods for intermediate and extreme QTE estimators are different. This difference raises the practical issue of how to choose the inference method in a given dataset with a small but given quantile index. Note that for $a \in (0, 1)$, any two-sided $(1-a)$-th CI can be written as

$$\text{CI} = \left(\hat{q}(\tau_n) - \widetilde{C}_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}_{\frac{a}{2}}(\tau_n) \right) \tag{5.4}$$

where $\widetilde{C}_a(\tau_n)$ is some critical value. However, the choice of $\widetilde{C}_a(\tau_n)$ depends on the order of τ_n.

Ideally, for extreme-order quantile index,

$$\widetilde{C}_a(\tau_n) = \widetilde{C}_a^{bn}(\tau_n) := \widehat{C}_a(\tau_n)/\hat{\alpha}_n$$

where $\widehat{C}_a(\tau_n)$ is the critical value computed by a b out of n bootstrap procedure for τ_n. For the intermediate and regular order quantile indices, $\widetilde{C}_a(\tau_n) = \widetilde{C}_a^{nn}(\tau_n)$ where $\widetilde{C}_a^{nn}(\tau_n)$ is the critical value computed by a standard bootstrap procedure. But in practice, it is impossible to determine the order of any quantile index because the size of the dataset is finite. The ideal procedure is not feasible.

Andrews and Cheng (2012) faced a similar problem because the model they considered can be either weakly, semi-strongly, or strongly identified. What they propose is an identification-category-selection (ICS) procedure based on the strength of identification. Similarly, I propose an order-category-selection (OCS) procedure based on the quantile index of interest and construct a robust CI.

Let $\tau_{n,1} := \min(\frac{40}{n}, \frac{0.2b}{mn})$, $\tau_{n,2} = \frac{b}{n\sqrt{\log(n)}}$, and for any $a \in (0,1)$,

$$\widetilde{C}^{lf}_{a/2}(\tau_n) = \max(\widetilde{C}^{bn}_{a/2}(\tau_n), \widetilde{C}^{nn}_{a/2}(\tau_n)) \quad \text{and} \quad \widetilde{C}^{lf}_{1-a/2}(\tau_n) = \min(\widetilde{C}^{bn}_{1-a/2}(\tau_n), \widetilde{C}^{nn}_{1-a/2}(\tau_n)).$$

The robust CI is constructed based on a hybrid critical value $\widetilde{C}^{h}_{a}(\tau_n)$ defined as follows.

$$\widetilde{C}^{h}_{a}(\tau_n) = \begin{cases} \widetilde{C}^{bn}_{a}(\tau_n) & \text{if } \tau_n \leq \tau_{n,1} \\ \widetilde{C}^{lf}_{a}(\tau_n) & \text{if } \tau_n \in (\tau_{n,1}, \tau_{n,2}) \\ \widetilde{C}^{nn}_{a}(\tau_n) & \text{if } \tau_n \geq \tau_{n,2}. \end{cases}$$

$\tau_{n,1}$, in general, takes the form of $\tau_{n,1} = \min(\frac{C_1}{n}, \frac{C_2 b}{mn})$, where C_1 and C_2 are two positive constants. If $k := \tau n$ is large, the approximation error from estimating the propensity score will contaminate the asymptotic approximation. This contamination inspires the requirement that $n\tau \leq C_1$. Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011) suggest to use $C_1 \in [40, 80]$. To be cautious, I choose $C_1 = 40$.

Second, the EV-law asymptotic approximation is only valid in the subsample with subsample size b if the quantile index used in the subsample, $m\tau_b := \frac{mk}{b} = \frac{m\tau n}{b}$, is close to zero. This inspires the second requirement that

$$m\tau_b \leq C_2.$$

Based on the simulations, the quantile index $m\tau_b$ is small enough if it is less than $C_2 = 0.2$. Combining these two requirements, I obtain $\tau_{n,1}$.

For n large enough, $\tau_{n,1} = \frac{40}{n}$. If $\tau \leq \tau_{n,1}$, $n\tau \leq 40 < \infty$. For such τ, it is expected that the extreme-order asymptotic distribution can approximate the finite distribution of the τ-th QTE estimator better than the standard normal distribution. In this case, the robust CI equals BN-CI.

On the other hand, if $\tau \geq \tau_{n,2}$,

$$\tau n \geq \frac{b}{\sqrt{\log(n)}} \to \infty$$

because $b \to \infty$ polynomially in n. For such τ, it is expected that the finite sample distribution of the τ-th QTE estimator is well approximated by the intermediate or regular order quantile asymptotic distribution. In both cases, the standard bootstrap CI is consistent. In addition, $\tau \geq \tau_{n,2}$ implies that

$$\tau_b := \frac{n\tau}{b} \geq \frac{1}{\sqrt{\log(n)}}.$$

It means that the quantile index τ_b used in computing the b out of n CI is not small. Thus to view τ_b in the subsample as close to zero is inappropriate and BN-CI constructed using τ_b may not be valid. For both reasons, when $\tau \geq \tau_{n,2}$, I suggest using only the standard bootstrap critical value.

When $\tau \in (\tau_{n,1}, \tau_{n,2})$, whether normal or EV approximation works better is not clear. In this case, the robust CI uses the least favorable critical value which is conservative.

The OCR procedure is different from the ICS procedure used in Andrews and Cheng (2012) because here I have two thresholds and when the quantile index is less than the first threshold, the asymptotic size is exact, while in Andrews and Cheng (2012), they only have one threshold and when the strength of identification is less than the threshold, their asymptotic size is conservative.

Let

$$\Gamma_{ex} := \left\{ \{\tau_n\}_{n\geq 1} : \tau_n \to 0, n\tau_n \to k \in (0,\infty), k \text{ satisfies Assumption 9} \right\},$$

$$\Gamma_{int} := \left\{ \{\tau_n\}_{n\geq 1} : \tau_n \to 0, n\tau_n \to \infty \right\},$$

and

$$\Gamma_{reg} := \left\{ \{\tau_n\}_{n\geq 1} : \tau_n = k \in (0,1) \right\}$$

denote the collections of extreme, intermediate, and regular order sequences of quantile indices. The next theorem shows that the robust CI is indeed robust over $\Gamma := \Gamma_{ex} \cup \Gamma_{int} \cup \Gamma_{reg}$.

Theorem 5.3. *Assumptions 1, 3–5, and 7–8 hold. Subsample size $b \to \infty$ polynomially in n and $\frac{b}{n} \to 0$. The standard bootstrap inference is consistent for regular quantile indices. Then, for any $a \in (0,1)$,*

$$\inf_{\{\tau_n\}_{n\geq 1} \in \Gamma} \lim_{n\to\infty} P\left(q(\tau_n) \in \left(\hat{q}(\tau_n) - \widetilde{C}^h_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^h_{\frac{a}{2}}(\tau_n) \right) \right) = 1 - a.$$

Unlike Andrews and Cheng (2012), in which the parameters and thus the DGPs are drifting, in my case, the DGP is fixed and the quantile index is drifting. So the above result mainly focuses on the robustness of CI's over different categories of quantile orders but does not speak to the uniformity over different DGPs.

5.4 Inference theory for the 0-th QTE

This section constructs a consistent CI for the 0-th QTE when the lower boundaries of Y_1 and Y_0 are bounded. The estimator for the 0-th QTE is a linear combination of extreme-order

QTE estimators with a set of carefully chosen weights. For inference, the same procedure of the inference method proposed for the extreme QTE in Section 5.2 can be directly applied.

I use a linear combination of extreme QTE estimators to infer the 0-th QTE so that the estimation bias cancels out. To see the source of bias, first recall that, when the lower end point is bounded and Assumption 10 holds, the tail is Type 3. This implies that $q_j^* = q_j(0)$ and $\beta_{n,j} = 0$. Hence I have

$$\hat{q}(\tau_n) - (q_1(0) - q_0(0)) = \hat{q}(\tau_n) - q(\tau_n) + \frac{k^{-\xi_1} + o(1)}{\alpha_{1,n}} - \frac{k^{-\xi_0} + o(1)}{\alpha_{0,n}}. \tag{5.5}$$

I can approximate the critical value of the asymptotic distribution for $\hat{q}(\tau_n) - q(\tau_n)$ based on the procedure after Proposition 5.1. The second term on the RHS of (5.5) is the bias caused by the fact that the parameter of interest is $q(0)$, instead of $q(\tau_n)$.

To get rid of this bias, I propose a feasible estimator $\hat{q}(0) := \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l})$ in which the weights $\{\hat{r}_l\}_{l=1}^{L}$ solve the following system of equations:

$$\sum_{l=1}^{L} \hat{r}_l = 1, \quad \sum_{l=1}^{L} \hat{r}_l k_l^{-\hat{\xi}_1} = 0, \quad \sum_{l=1}^{L} \hat{r}_l k_l^{-\hat{\xi}_0} = 0. \tag{5.6}$$

Here, $(\hat{\xi}_0, \hat{\xi}_1)$, the consistent estimators of (ξ_0, ξ_1), can be computed by Theorem 3.4.
To implement, I compute $\hat{q}(0)$ using only three different values of $\tau_{n,l}$, that is, $L = 3$. The reason is twofold: (1) I do not have a selection rule for choosing among solutions of weights that satisfies (5.6) if the solution is not unique, and (2) by fixing the upper and lower bound $\tau_{n,1}$ and $\tau_{n,L}$, the more quantile indices I use, the higher the weights, which will widen the implied CI.

Proposition 5.2. *Let $\hat{\xi}_j$ be consistent estimates of ξ_j for $j = 0, 1$, $L = 3$, $(\hat{r}_1, \hat{r}_2, \hat{r}_3)$ be computed as in (5.6), $\hat{q}(0) := \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l})$, and $\widehat{C}_a$ be computed as in the procedure after Proposition 5.1. If the assumptions in Theorem 4.2 hold and $q_j(0)$ is bounded for $j = 0, 1$, then*

$$\lim_{n\to\infty} P\left(\hat{q}(0) - \widehat{C}_{1-a/2}/\hat{\alpha}_n \leq q(0) \leq \hat{q}(0) - \widehat{C}_{a/2}/\hat{\alpha}_n\right) = 1 - a.$$

There are two alternative methods by which to infer the 0-th QTE, each of which has its own restriction. The first alternative is to analytically compute $\frac{k^{-\xi_1}}{\alpha_{1,n}} - \frac{k^{-\xi_0}}{\alpha_{0,n}}$, the leading term of the bias in (5.5). This requires the estimation of the infeasible convergence rate $\alpha_{j,n}$. However, computing an estimator $\tilde{\alpha}_{j,n}$ of $\alpha_{j,n}$ such that $\frac{\tilde{\alpha}_{j,n}}{\alpha_{j,n}} \to 1$ is harder than simply estimating the EV index ξ_j. Usually, in order to compute $\tilde{\alpha}_{j,n}$, distributional assumptions, such as $\alpha_{j,n} = C_j n^{\xi_j}$ for some constant C_j, are imposed. See, for example, the discussion in Chernozhukov and Fernández-Val (2011) on the distributional assumption and Bertail, Politis, and Romano (1999) on the point of conductin subsampling inference when the convergence rate is unknown. These distributional assumptions are not needed in Proposition 5.2.

The second alternative is to rely on asymptotics to ensure that the bias is asymptotically negligible and small in the finite sample. To be more specific, combining Theorems 4.1 and 4.2, it is clear that for $\tau_n n \to k$,

$$\hat{\alpha}_n(\hat{q}(\tau_n) - q(0))$$

converges weakly to a non-degenerate limiting distribution. I can then approximate the critical value of the limiting distribution by computing

$$\hat{Z}_n^*(k) := \hat{\alpha}_b^*(\hat{q}^*(\tau_b) - \hat{q}(\tau_n))$$

for $\tau_b b = \tau_n n$. Comparing $\hat{Z}_n^*(k)$ with $\hat{Z}_n^{c*}(k)$ in (5.2), the only difference is that the subsample estimator $\hat{q}_b^*(\tau_b)$ is now centered by $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$, the full sample QTE estimator at τ_n, instead of $\hat{q}(\tau_b)$. The reason is that for the subsample, $\hat{q}(\tau_b)$ and $\hat{q}(\tau_n)$ can be viewed as proxies for $q(\tau_b)$ and $q(0)$, respectively. Then, after I obtain an estimator of the critical value of the limiting distribution of $\hat{Z}_n^*(k)$ by a similar b out of n bootstrap procedure, I can construct a median-unbiased estimator and a consistent CI for $q(0)$. For this method to work, I rely on the fact that the bias of using $\hat{q}(\tau_n)$ as a proxy of $q(0)$ vanishes asymptotically. Since econometricians have no control of the magnitude of the bias in a finite sample, this method is passive. The properties of the implied CI in finite samples can be sensitive to both the choice of $k = \tau_n n$ and the subsample size b. Therefore, the passive method is less robust than the one proposed in Proposition 5.2.

5.5 Two-sample inference

Given two independent samples (1) and (2) with sample sizes n_1 and n_2, the $\tau_{n_1}^{(1)}$-th and $\tau_{n_2}^{(2)}$-th QTEs for the two samples are denoted as $q^{(1)}(\tau_{n_1}^{(1)})$ and $q^{(2)}(\tau_{n_2}^{(2)})$, respectively. In application, researchers are also interested in inferring the difference of the QTE at tails between two samples. In particular, they are interested in testing $q^{(1)}(\tau_{n_1}^{(1)}) = q^{(2)}(\tau_{n_2}^{(2)})$ for $\tau_{n_1}^{(1)} n_1 = \tau_{n_2}^{(2)} n_2 = k$. The following procedure constructs the median-unbiased point estimator and the CI for $q^{(1)}(\frac{k}{n_1}) - q^{(2)}(\frac{k}{n_2})$.

1. For the first sample, compute the propensity score, $(\hat{q}^{(1)}(\frac{k}{n_1}), \hat{q}^{(1)}(\frac{k}{b_1}))$ as in (4.1), and $\hat{\alpha}_{n_1}^{(1)}$ as in (4.3).

2. Let $b_2 := \lfloor \frac{b_1 n_2}{n_1} \rfloor$. For the second sample, compute the propensity score, $(\hat{q}^{(2)}(\frac{k}{n_2}), \hat{q}^{(2)}(\frac{k}{b_2}))$, and $\hat{\alpha}_{n_2}^{(2)}$ in the same manner. Denote

$$\hat{\alpha}_n = \min(\hat{\alpha}_{n_1}^{(1)}, \hat{\alpha}_{n_2}^{(2)}).$$

3. For the i-th step, generate subsample 1 with size b_1 from the first sample and subsample 2 with size b_2 from the second sample. Compute $\hat{q}^{(1)*}(\frac{k}{b_1})$ as in (4.1) and

$$\hat{\alpha}_{b_1}^{(1)*} := \frac{\sqrt{k_{l'}}}{\max\left\{\hat{q}_1^{(1)^*}(\frac{mk_{l'}}{b_1}) - \hat{q}_1^{(1)^*}(\frac{k_{l'}}{b_1}), \hat{q}_0^{(1)^*}(\frac{mk_{l'}}{b_1}) - \hat{q}_0^{(1)^*}(\frac{k_{l'}}{b_1})\right\}},$$

with some $k_{l'}$ specified by researchers, using the data from the first subsample. On the RHS of the above equation, $\hat{q}_j^{(1)*}(\tau)$ and $\hat{q}_j^{(1)*}(\tau)$, for $j = 0, 1$ are computed as in (3.1) and (3.2), respectively, with the propensity score computed using the full sample. Similarly, from the second subsample, compute $\hat{q}^{(2)*}(\frac{k}{b_2})$ and

$$\hat{\alpha}_{b_2}^{(2)*} := \frac{\sqrt{k_{l'}}}{\max\left\{\hat{q}_1^{(2)^*}(\frac{mk_{l'}}{b_2}) - \hat{q}_1^{(2)^*}(\frac{k_{l'}}{b_2}), \hat{q}_0^{(2)^*}(\frac{mk_{l'}}{b_2}) - \hat{q}_0^{(2)^*}(\frac{k_{l'}}{b_2})\right\}}$$

Denote

$$\hat{\alpha}_b^* = \min(\hat{\alpha}_{b_1}^{(1)*}, \hat{\alpha}_{b_2}^{(2)*}), \quad \widehat{V}_{i,b}^* := \hat{\alpha}_b^*\left[\left(\hat{q}^{(1)*}(\frac{k}{b_1}) - \hat{q}^{(1)}(\frac{k}{b_1})\right) - \left(\hat{q}^{(2)*}(\frac{k}{b_2}) - \hat{q}^{(2)}(\frac{k}{b_2})\right)\right].$$

4. Repeat the above step for $i = 1, \cdots, B_n$. Compute $\widehat{C}_{1-a}$ as the $(1-a)$-th quantile of the $\{\widehat{V}_{i,b}^*\}_{i=1}^{B_n}$.

5. Construct the $(1-a)$-CI as

$$\text{CI}_a = \left[\hat{q}^{(1)}(\frac{k}{n_1}) - \hat{q}^{(2)}(\frac{k}{n_2}) - \widehat{C}_{1-a/2}/\hat{\alpha}_n, \hat{q}^{(1)}(\frac{k}{n_1}) - \hat{q}^{(2)}(\frac{k}{n_2}) - \widehat{C}_{a/2}/\hat{\alpha}_n\right].$$

Theorem 5.4. *$\{Y_i^{(1)}, D_i^{(1)}, X_i^{(1)}\}_{i=1}^{n_1}$ and $\{Y_i^{(2)}, D_i^{(2)}, X_i^{(2)}\}_{i=1}^{n_2}$ are two independent samples which satisfy all the assumptions in Theorem 4.2. Let $b_2 := \lfloor \frac{b_1 n_2}{n_1} \rfloor$. As $n_1 \to \infty$, $\frac{b_1}{n_1} \to 0$, $b_1 \to \infty$ at a polynomial rate in n_1, and there exists constants $\upsilon \in (0, \infty)$ and $(\rho_0, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5) \in$*

$[0,\infty]^6$, *such that* $\frac{n_2}{n_1} \to \upsilon$,

$$\frac{q_0^{(1)}(\frac{mk_{l'}}{n_1}) - q_0^{(1)}(\frac{k_{l'}}{n_1})}{q_0^{(2)}(\frac{mk_{l'}}{n_1}) - q_0^{(2)}(\frac{k_{l'}}{n_1})} \to \rho_0, \quad \frac{q_1^{(1)}(\frac{mk_{l'}}{n_1}) - q_1^{(1)}(\frac{k_{l'}}{n_1})}{q_0^{(1)}(\frac{mk_{l'}}{n_1}) - q_0^{(1)}(\frac{k_{l'}}{n_1})} \to \rho_1, \quad \frac{q_1^{(2)}(\frac{mk_{l'}}{n_1}) - q_1^{(2)}(\frac{k_{l'}}{n_1})}{q_0^{(2)}(\frac{mk_{l'}}{n_1}) - q_0^{(2)}(\frac{k_{l'}}{n_1})} \to \rho_2,$$

$$\frac{q_1^{(1)}(\frac{mk_{l'}}{n_1}) - q_1^{(1)}(\frac{k_{l'}}{n_1})}{q_1^{(2)}(\frac{mk_{l'}}{n_1}) - q_1^{(2)}(\frac{k_{l'}}{n_1})} \to \rho_3, \quad \frac{q_1^{(1)}(\frac{mk_{l'}}{n_1}) - q_1^{(1)}(\frac{k_{l'}}{n_1})}{q_0^{(2)}(\frac{mk_{l'}}{n_1}) - q_0^{(2)}(\frac{k_{l'}}{n_1})} \to \rho_4, \quad \textit{and} \quad \frac{q_0^{(1)}(\frac{mk_{l'}}{n_1}) - q_0^{(1)}(\frac{k_{l'}}{n_1})}{q_1^{(2)}(\frac{mk_{l'}}{n_1}) - q_1^{(2)}(\frac{k_{l'}}{n_1})} \to \rho_5.$$

Then

$$\lim_{n_1\to\infty} P\left(q^{(1)}(\frac{k}{n_1}) - q^{(2)}(\frac{k}{n_2}) \le \widehat{C}_{0.5}/\hat{\alpha}_n\right) = 0.5 \quad \textit{and} \quad \lim_{n_1\to\infty} P\left(q^{(1)}(\frac{k}{n_1}) - q^{(2)}(\frac{k}{n_2}) \in CI_a\right) = 1-a.$$

In Section 7, I will rely on the above procedure and the theorem to infer the difference of racial gaps in college preparation index prior to and following a policy change.

6 Simulations

6.1 Limiting distributions

I first verify the asymptotic distributions of $\hat{q}_1(\tau_n)$ established in Section 4. Figure 6 plots the quantiles of the normalized sample distribution of $\hat{q}_1(\tau_n)$ against the quantiles of its limiting distribution established in Theorem 4.1 with four different boundary structures: single minimizer, finite minimizers, continuum minimizers, and mixture minimizers. Since the plots are all close to the diagonal line, the new asymptotic distributions based established in Theorem 4.1 approximate the finite sample distributions very well.

Figure 7, on the other hand, plots the exact same quantiles for the estimators against the quantiles of the standard normal distribution. The plots are all non-linear, which indicates that the shape of the finite sample distributions is not normal. Any inference method based asymptotic normality will fail to produce a consistent CI.

6.2 Inference for the extreme QTE

Table 2 and 3 illustrate that the standard bootstrap CI undercovers as much as 18.2 absolute percentage points while the BN-CI's coverage is very close to the nominal 95% when τ is less than 2% or correspondingly, $k := \tau n \le 40$. In addition, the length of the BN-CI is larger but still comparable to one with the standard bootstrap CI, which ensures the practical value of BN-CI.

Figure 8 shows that when the quantile index is less than the threshold, the BN-CI has an accurate coverage while the standard bootstrap CI (NN-CI) undercovers substantially. As the quantile index increases, BN-CI usually overcovers, which means that the BN-CI is conservative, while the NN-CI still undercovers, but the coverage gradually converges to the nominal rate. In addition, Figure 9 shows that the BN-CI is insensitive to the choice of subsample size b over a reasonable range.

6.3 The robust confidence interval

Figure 10 shows the finite sample performance of the robust CI proposed in Section 5.3. When $\tau \leq \tau_{n,1}$ or $\tau \geq \tau_{n,2}$, the coverage is close to the 95% nominal rate while when $\tau \in (\tau_{n,1}, \tau_{n,2})$, the CI overcovers and thus is conservative. All sixteen models exhibit this same pattern. For details, please see Appendix F.3.

6.4 Inference for the 0-th QTE

Table 4 shows that the coverages of BN-CI for the 0-th QTE estimator proposed in Section 5.4 are all close to the nominal rate and median length of the CI's are reasonable. Figure 11 plots the coverage of BN-CI against the subsample size b for $b \in [500, 1,000]$. It shows that the coverages for the BN-CI are not sensitive to the choice of subsample size.

7 Empirical applications

7.1 Effect of maternal status on extremely low birth weights

The lower tail of the birth weight distribution reflects severely adverse birth outcomes, which is the main research interest in health economics. Adverse birth outcomes, particularly low birth weight, are the leading causes of infant mortality, a main concern of public health research. In addition, adverse birth outcomes result in large economic costs in not only direct newborn care costs, but also long-term developmental costs like delayed entry into kindergarten, repeated grades, and the consequent labor market outcomes. For literature on maternal smoking and birth weights, see, for example, Abrevaya (2001), Abrevaya (2006), Abrevaya and Dahl (2008), Chernozhukov and Fernández-Val (2011), Evans and Lien (2005), Evans and Ringel (1999), Permutt and Hebel (1989), Rosenzweig and Wolpin (1991), and the references therein.

Despite the large literature on the effect of maternal smoking on birth weights, there is no consensus on its magnitude. Various research papers, using different estimation tools and data, find that the negative effect of maternal smoking is about 189-600 grams decrease in birth weight[5]. See Abrevaya (2006) for a summary. But in order to draw these conclusions, empirical researchers usually consider small but regular quantile estimates or subsamples of low-weight infants and refer to the asymptotic normality to draw inferences. The only exception is Chernozhukov and Fernández-Val (2011), who looked at extremely low birth weight and referred to the EV distribution to draw inferences. Figure 8 of Chernozhukov and Fernández-Val (2011) shows that the extremal quantile regression coefficient of maternal smoking is close to zero and statistically insignificant.

I estimate the QTE of maternal smoking on extremely low birth weight infants. The QTE is distinct from the linear regression coefficient of smoking status estimated in Chernozhukov and Fernández-Val (2011) in four aspects. First, the extreme QTE is an unconditional parameter while the regression coefficient is a conditional one. The extreme QTE estimated here differs empirically from the linear regression coefficient because the conditional quantile is heterogeneous as shown in Figure 8 and 9 in Chernozhukov and Fernández-Val (2011). To recover the unconditional QTE from a conditional coefficient is also hard because inverse CDF is a nonlinear operator. Second, I control for covariates in a more flexible way than the linear regression, which makes the QTE estimator robust to misspecifications. Third, the paradigm of QTE, given a fixed quantile index τ, still allows for two observationally equivalent babies to have different treatment responses to maternal smoking, while the QTE estimated by linear regression relies on the implicit assumption that the treatment effect is homogeneous. Last, I also estimate the exact 0-th unconditional QTE, which measures the effect of maternal smoking on the lower boundary of babies birth weight and is new to the literature.

I use the same dataset as in Chernozhukov and Fernández-Val (2011). It was collected based on June 1997 Detailed Natality Data published by the National Center for Health Statistics and has been previously investigated by Abrevaya (2001) and Koenker and Hallock (2001). I concentrate on African American mothers only, with 31,912 observations, because Figure 7 of Chernozhukov and Fernández-Val (2011) shows that low birth weights for black mothers have a heavy lower tail. Economically, it suggests a severe adverse birth outcome which is the main target of this analysis. Theoretically, the heavy lower tail of the birth weights distribution is consistent with Assumption 10(3), which is the key to conducting the b out of n bootstrap inference for the extreme QTE.

Table 5 reports the median-unbiased point estimates and the CI for the extreme QTE of maternal smoking. In all quantile indices, I cannot reject that maternal smoking has no negative impact on either extremal quantile or the lower bound of infants' birth weights

[5]The average birth weight for an infant is about 3,400 grams.

under 90% confidence level. A potential explanation for this result is that the catastrophic birth outcome may be due to more severe diseases rather than maternal smoking. On the other hand, the BN-CI is more than two times wider than the standard bootstrap CI. This indicates that the standard bootstrap CI potentially undercovers which is consistent with the simulation study. Last, the median-unbiased estimator for the 0-th QTE implies that if a pregnant mother smokes, with 50% probability, her child's lowest possible birth weight is 137.32 grams lighter than it would be if she did not smoke.

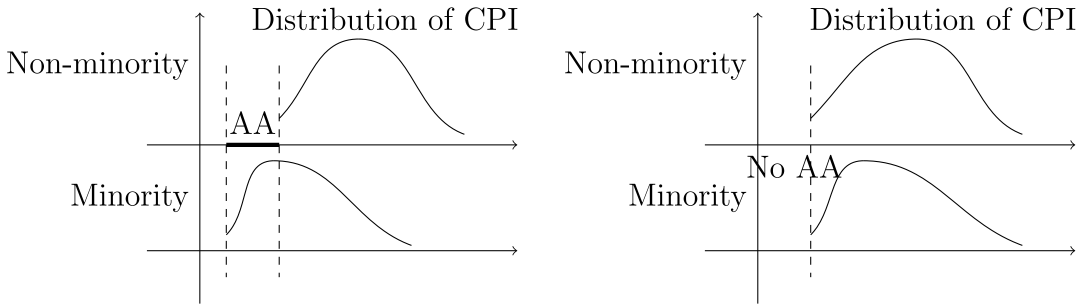
Although estimating the extreme QTE is one step forward in the direction of causal inference, the existence of unobserved confounders can jeopardize the selection on observables. For example, mothers who smoke during pregnancy are more likely to adopt other behaviors (drinking, poor nutritional intake, etc.) that could have a negative impact on birth weight. Evans and Lien (2005) and Evans and Ringel (1999) address this problem by using large cigarette taxation change as an instrumental variable (IV) for maternal smoking. Extending the current theory to incorporate IV and conduct inference for the extremal QTE for the compliers would be a useful research direction.

7.2 Effect of minority status on college preparation index

This section considers the effect of minority status on the college preparation index (CPI) for low-scoring college students with equivalent family backgrounds. Minority status can impact the distributions of CPI directly through universities' admission policy, and indirectly through the "backdoor" channel: minority students may live in a less favorable family environment with low parental income and education level, which causes minority students to be less prepared for college than their majority peers. After controlling for family backgrounds, the CPI gap can be viewed as a measure of affirmative action in colleges' admission selections in the dimension of academic performances. See, for example, Arcidiacono, Aucejo, Coate, and Hotz (2014). Throughout the application, I control for parental income and parental education as confounders when computing the causal gap of minority status.

I focus on students with low CPI because they are the *marginal* population who will be affected by the change of admission selection criteria. If a college's admission is purely meritocratic, then Proposition 1 of Bhattacharya, Kanaya, and Stevens (2016) shows that the optimal admission protocol is a simple threshold-crossing form. Given the population of enrolled students, the threshold can be identified as the lower boundary of the CPI distribution, which is just the zero-th quantile. The gap of zero-th quantile of the distributions of CPI for minority and majority students can then be viewed as a measure of the magnitude of racial preference in college admission in the dimension of academic performance, or in other words, a measure of the deviation of college admission rule from pure meritocracy. See

Figure 2a for an illustration. Furthermore, Figure 2b shows that it is common to have zero *marginal* gap at the tail, but non-zero gap on average.



(a) Different lower boundaries (b) Same boundaries, but different averages

In reality, the admission criteria in U.S. is multidimensional. Therefore, no simple threshold for CPI can be identified from the data. However, based on the intuition built by Bhattacharya et al. (2016), students with low CPI are the *marginal* population who are more likely to be affected by the policy change on racial preferences in colleges' admission selections, and thus is the population of research interest. In addition, Arcidiacono, Aucejo, and Hotz (2016) pointed out that CPI is related to racial inequality in terms of schooling achievement, and thus also later economic outcomes. Hence, even without the theoretical justification above, the racial gap in the tail of the distribution of academic performance of admitted students provides another measure of affirmative action other than the average gap, and is of its own interest.

The analysis here focuses on *marginal* admits which is the same as Bhattacharya et al. (2016), but is in contrast with many other studies which focus on *average* pre-admission test-scores (e.g. Zimdars, Sullivan, and Heath (2009)) or *average* post-admission test-scores (e.g. Keith, Bell, Swanson, and Williams (1985), Kane (1998), and Sackett, Kuncel, Arneson, Cooper, and Waters (2009)). See Hoxby (2009) for historical perspective on selectivity in US college admission and Arcidiacono, Lovenheim, and Zhu (2015) for a recent survey.

7.2.1 Pre-Prop 209

The UC campuses were subject to a ban on the use racial preference in admissions enacted under Proposition 209 (Prop 209) which took effect in 1998. I use the UCOP data for minority and non-minority students who first enrolled at one of the UC campuses in periods both pre- and post-Prop 209, to compute the racial CPI gap at tails.[6]

Table 6 shows that, prior-Prop 209, after controlling for family background, the gaps at the lower tail are almost all negative and statistically significant, except for students with

[6]For more details on Prop 209, the data, and the implementation, please see Appendix D.

science major in UC Santa Cruz and students with non-science major in UC San Diego. It suggests that prior-Prop 209, almost all UC campuses implemented racial preferences in the dimension of academic performances during admission. In addition, the gaps at the tail are larger for higher ranked campuses such as Berkeley and Los Angeles than that for the rest of the campuses. It suggests that minority students and their majority peers have more similar levels of college preparation in lower ranked campuses from the start. This provides a partial explanation for the empirical finding in Arcidiacono et al. (2016) that less-prepared minority students may have higher graduation probabilities at less-selective schools.

7.2.2 Post-Prop 209

Table 7 shows that the average CPI gaps for all campuses remain significant post-Prop 209. But this does not necessarily reflect that there still exist racial preference in college admission post-Prop 209 as argued by Figure 2b. In fact, Table 7 also shows that the tail gaps of CPI become insignificant for several campuses, which suggests that the racial preference in the corresponding campuses is insignificant.
Comparing Table 6 and 7, I find heterogeneous responses of UC campuses to Prop 209. The racial gaps in UC Berkeley and UCLA for students with science major and in UC Berkeley, UC Santa Cruz, and UC Riverside for students with non-science major remained significant after Prop 209. For UC Santa Cruz science major, the gap became significant post-Prop 209. These two results suggest that racial preferences in admission did not decrease post-Prop 209 for several campuses (especially Berkeley and Los Angeles). One possible explanation is that, post-Prop 209, colleges modified their admission rules to implicitly favor minority students. This is consistent with the finding in Antonovics and Backes (2014) that some campuses responded to the ban of the race-based affirmative action by lowering weights on academic credentials such as SAT scores and increasing weights on family backgrounds in determining admissions. Because minority students are more likely to have less favorable family backgrounds, by putting more weights on family background, the admission rule implicitly favor minority students.

7.2.3 Pre- and post-Prop 209 comparison

The median-unbiased point estimators pre- and post-Prop 209 differ most for admitted students majoring science at UC Berkeley and UC San Diego. The difference can be summarized in Figure 12.

I also test whether the differences of racial gaps pre- and post-Prop for UC Berkeley and UC San Diego are significant by using the two-sample test established in Section 5.5. Table 8 shows that, for UC Berkeley, we cannot reject that racial gaps remained the same level prior and post-Prop 209. In addition, the median-unbiased point estimator for the difference of racial gap among Berkeley students with science major pre- and post-Prop 209 is positive, which implies that the racial gap in UC Berkeley may actually increase with more than half of the probability. Again, these findings support the empirical results in Antonovics and Backes (2014), which suggest that UC Berkeley might have modified its admission protocol to maintain the same level of racial preference in the dimension of CPI. For students majoring science at UC San Diego, by contrast, the CPI gap deceases significantly post-Prop 209. This provides evidence that UC San Diego modified the college admission rule according to Prop 209.

8 Conclusion

This paper establishes asymptotic theory and inference procedures for an estimator of the unconditional QTE when the quantile index is close or equal to zero. There are two main difficulties: missing data and data sparsity. I address them simultaneously by relying on the unconfoundedness assumption and extremal quantile asymptotics, respectively. When the quantile index is close or equal to zero, I derive a new asymptotic approximation of the finite sample estimator of the QTE and show that standard bootstrap inference is inconsistent. Based on my new asymptotic theory, I propose a new way to construct a uniformly consistent confidence band for extreme QTEs. Last, by using a linear combination of extreme QTE estimators, I propose a median-unbiased estimator and consistent CI for the 0-th QTE.

I then apply the new inference method to estimate the effect of maternal smoking of African American mothers for the lower tail of infants' birth weights and the racial gap of CPI in college admissions. For the first application, while I cannot reject that maternal smoking has no effect on the lower tail of birth weights at the 90% confidence level, I find that the standard bootstrap CI is two times narrower than the new resampling CI developed in this paper. The difference suggests that the standard bootstrap CI potentially over-rejects.

For the second application, I find evidence that pre-Prop 209, most UC campuses implemented racial preference in academic performances and post-Prop 209, UC campuses modified their admission selection criteria in a heterogeneous manner.

References

Abadie, A., Angrist, J., and Imbens, G. (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Abrevaya, J. (2001), "The Effects of Demographics and Maternal Behavior on the Distribution of Birth Outcomes," *Empirical Economics*, 26 (1), 247–257.

Abrevaya, J. (2006), "Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach," *Journal of Applied Econometrics*, 21, 489–519.

Abrevaya, J. and Dahl, C. M. (2008), "The Effects of Birth Inputs on Birthweight: Evidence from Quantile Estimation on Panel Data," *Journal of Business & Economic Statistics*, 26, 379–397.

Andrews, D. W. and Cheng, X. (2012), "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

Andrews, D. W. and Cheng, X. (2013), "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," *Journal of Econometrics*, 173, 36–56.

Antonovics, K. and Backes, B. (2014), "The Effect of Banning Affirmative Action on College Admissions Policies and Student Quality," *Journal of Human Resources*, 49, 295–322.

Arcidiacono, P., Aucejo, E., Coate, P., and Hotz, V. J. (2014), "Affirmative Action and University Fit: Evidence from Proposition 209," *Journal of Labor Economics*, special issue on affirmative action.

Arcidiacono, P., Lovenheim, M., and Zhu, M. (2015), "Affirmative Action in Undergraduate Education," *Annu. Rev. Econ.*, 7, 487–518.

Arcidiacono, P., Aucejo, E. M., and Hotz, V. J. (2016), "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California," *American Economic Review*, 106(3), 525–562.

Bertail, P., Politis, D. N., and Romano, J. P. (1999), "On Subsampling Estimators with Unknown Rate of Convergence," *Journal of the American Statistical Association*, 94, 569–579.

Bertail, P., Haefke, C., Politis, D. N., and White, H. (2004), "Subsampling the Distribution of Diverging Statistics with Applications to Finance," *Journal of Econometrics*, 120, 295–326.

Bhattacharya, D., Kanaya, S., and Stevens, M. (2016), "Are University Admissions Academically Fair?" *Review of Economics and Statistics*, forthcoming.

Bickel, P. J. and Freedman, D. A. (1981), "Some Asymptotic Theory for the Bootstrap," *The Annals of Statistics*, 9, 1196–1217.

Bickel, P. J. and Sakov, A. (2008), "On the Choice of m in the m out of n Bootstrap and its Application to Confidence Bounds for Extreme Percentiles," *Statistica Sinica*, 18(3), 967–985.

Billingsley, P. (1999), *Convergence of Probability Measures*, vol. 493, John Wiley & Sons, second edition.

Bitler, M., Gelbach, J., and Hoynes, H. (2006), "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988–1012.

Card, D. (1996), "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957–979.

Chen, X., Ponomareva, M., and Tamer, E. (2014), "Likelihood Inference in Some Finite Mixture Models," *Journal of Econometrics*, 182, 87–99.

Chernozhukov, V. (2005), "Extremal quantile regression," *The Annals of Statistics*, 33, 806–839.

Chernozhukov, V. and Fernández-Val, I. (2011), "Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks," *The Review of Economic Studies*, 78, 559–589.

Chernozhukov, V. and Hansen, C. (2005), "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V. and Hansen, C. (2008), "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 142, 379–398.

Chernozhukov, V. and Hong, H. (2004), "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72, 1445–1480.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013), "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014), "Gaussian Approximation of Suprema of Empirical Processes," *The Annals of Statistics*, 42, 1564–1597.

Chernozhukov, V. V. (2000), "Conditional Extremes and Near-extremes: Concepts, Asymptotic Theory, and Economic Applications," Ph.D. thesis, stanford university.

Dekkers, A. L. and De Haan, L. (1989), "On the Estimation of the Extreme-value Index and Large Quantile Estimation," *The Annals of Statistics*, 17, 1795–1832.

D'Haultfoeuille, X., Maurel, A., and Zhang, Y. (2015), "Extremal Quantile Regressions for Selection Models and the Black-White Wage Gap," *NBER Working Paper 20257.*

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64(5), 1001–1044.

Durrett, R. (2010), *Probability: Theory and Examples*, Cambridge university press.

Evans, W. N. and Lien, D. S. (2005), "The Benefits of Prenatal Care: Evidence from the PAT Bus Strike," *Journal of Econometrics*, 125, 207–239.

Evans, W. N. and Ringel, J. S. (1999), "Can Higher Cigarette Taxes Improve Birth Outcomes?" *Journal of Public Economics*, 72, 135–154.

Falk, M. (1991), "A Note on the Inverse Bootstrap Process for Large Quantiles," *Stochastic processes and their applications*, 38, 359–363.

Feigin, P. D. and Resnick, S. I. (1994), "Limit Distributions for Linear Programming Time Series Estimators," *Stochastic Processes and their Applications*, 51, 135–165.

Firpo, S. (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

Firpo, S. and Rothe, C. (2014), "Semiparametric Estimation and Inference Using Doubly Robust Moment Conditions," Tech. rep., IZA Discussion Paper.

Frölich, M. and Melly, B. (2013), "Unconditional Quantile Treatment Effects under Endogeneity," *Journal of Business & Economic Statistics*, 31, 346–357.

Hirano, K. and Porter, J. R. (2003), "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," *Econometrica*, 71, 1307–1338.

Hirano, K., Imbens, G. W., and Ridder, G. (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated propensity score," *Econometrica*, 71, 1161–1189.

Hoxby, C. M. (2009), "The Changing Selectivity of American Colleges," .

Kane, T. J. (1998), "Racial and Ethnic Preferences in College Admissions," *Ohio St. LJ*, 59, 971.

Keith, S. N., Bell, R. M., Swanson, A. G., and Williams, A. P. (1985), "Effects of Affirmative Action in Medical Schools: A Study of the Class of 1975," *New England Journal of Medicine*, 313, 1519–1525.

Khan, S. and Nekipelov, D. (2013), "On Uniform Inference in Nonlinear Models with Endogeneity," *Economic Research Initiatives at Duke (ERID) Working Paper*.

Knight, K. (2001), "Limiting Distributions of Linear Programming Estimators," *Extremes*, 4, 87–103.

Koenker, R. (2005), *Quantile Regression*, no. 38, Cambridge university press.

Koenker, R. and Hallock, K. (2001), "Quantile Regression: An Introduction," *Journal of Economic Perspectives*, 15, 43–56.

Lee, S. and Seo, M. H. (2008), "Semiparametric Estimation of a Binary Response Model with a Change-point due to a Covariate Threshold," *Journal of Econometrics*, 144, 492–499.

Newey, W. K. and McFadden, D. (1994), "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

Permutt, T. and Hebel, J. R. (1989), "Simultaneous-equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight," *Biometrics*, 45(2), 619–622.

Pitman, E. (1949), "Lecture Notes on Nonparametric Statistical Inference," .

Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, pp. 186–199.

Portnoy, S. and Jurečková, J. (1999), "On Extreme Regression Quantiles," *Extremes*, 2, 227–243.

Resnick, S. (1987), *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York.

Resnick, S. I. (2007), *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*, Springer.

Rosenbaum, P. R. and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rosenzweig, M. R. and Wolpin, K. I. (1991), "Inequality at Birth: The Scope for Policy Intervention," *Journal of Econometrics*, 50, 205–228.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., and Waters, S. D. (2009), "Socioeconomic Status and the Relationship Between the SAT and Rreshman GPA: An Analysis of Data from 41 Colleges and Universities," *College Board, New York*.

Smith, R. L. (1994), "Nonregular Regression," *Biometrika*, 81, 173–183.

Stock, J. and Yogo, M. (2005), *Testing for Weak Instruments in Linear IV Regression*, pp. 80–108, Cambridge University Press, New York.

Stock J, S. D. (2008), "Semiparametric Estimation of a Binary Response Model with a Change-point due to a Covariate Threshold," *Journal of Econometrics*, 144, 492–499.

Van der Vaart, A. W. (2000), *Asymptotic Statistics*, vol. 3, Cambridge university press.

Van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.

Zimdars, A., Sullivan, A., and Heath, A. (2009), "Elite Higher Education Admissions in the Arts and Sciences: Is Cultural Capital the Key?" *Sociology*, 43, 648–666.

A Asymptotic distribution under various boundary conditions

This section verifies Assumption 8 under three different boundary conditions. I demonstrate that the asymptotic distribution for the extreme QTE is nonregular and depends on complications in boundary conditions. More numerical illustrations are in Appendix B. Since the boundary condition is unknown and is usually hard to estimate, analytical inference is difficult. Instead, in Section 5, I will focus on resampling based inference, which does not require knowledge of the boundary.

First, I give another representation of the asymptotic objective function established in Theorem 4.1. In fact,

$$-kz + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j}))l_\delta(\mathcal{J}_{i,j}, z) = -kz + \int_{E_j} W_j(d, P(x))l_\delta(u_j, z)dN_j(u_j, d, x),$$

where $N_j(u_j, d, x)$ is a Poisson random measure on E_j with mean measure μ_j (PRM(μ_j)) and

$$\begin{aligned}
&\text{for type 1 tails } (\xi_j = 0): && E_j = E^1 = [-\infty, +\infty) \times \{0,1\} \times \text{Supp}(\mathcal{X}),\\
&\text{for type 2 tails } (\xi_j > 0): && E_j = E^2 = [-\infty, 0) \times \{0,1\} \times \text{Supp}(\mathcal{X}),\\
&\text{for type 3 tails } (\xi_j < 0): && E_j = E^3 = [0, +\infty) \times \{0,1\} \times \text{Supp}(\mathcal{X}).
\end{aligned}$$

Let $\mathcal{F}$ be a basis of relatively compact open sets of $\mathbb{R}^r$ such that $\mathcal{F}$ is closed under finite unions and intersections[7] and for any $F \in \mathcal{F}$,

$$P_j^+(\mathcal{X}_j \in \text{Bd}(F)|Y = q_j(0)) = 0,$$

in which $\text{Bd}(F)$ is the boundary of the set F. Then the mean measure μ_j, which uniquely determines the distribution of a Poisson random measure, is defined as

$$\mu_j((a,b) \times \{d\} \times F) := \int_F (dP(x) + (1-d)(1-P(x)))P_j^+(dx|Y_j = q_j(0))(h_j(b) - h_j(a)). \quad \text{(A.1)}$$

Next, I establish the asymptotic distribution of $\hat{q}_j(\tau_n)$ by deriving the close-form expressions for the mean measure μ_j under three different boundary conditions: the conditional boundary of Y_j given X having finite minimizers, continuum minimizers, and mixture minimizers. I will restrict my attention to the marginal distribution of $\hat{q}_1(\tau_n)$ because of the asymptotic independence between $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$.

[7] r is the dimension of X.

A.1 Finite minimizers

When the lower endpoint of Y_1 is bounded, I denote $\varpi(x)$ as Y_1's conditional boundary given $X = x$. If $\varpi(x)$ is uniquely minimized at x_0, then as $Y_1 \to q_1(0)$, $X \to x_0$. So I expect $P_1^+(\mathcal{X}_1 \in \cdot | Y_1 = q_1(0))$ to be $\mathbb{1}\{x_0 \in \cdot\}$. This implies that the mean measure μ_1 in the asymptotic distribution of $Z_{1,\infty}(k)$ defined in (A.1) takes the following form:

$$\mu_1((a,b) \times \{d\} \times F) = (dP(x_0) + (1-d)(1-P(x_0)))(h_1(b) - h_1(a))\mathbb{1}\{x_0 \in F\},$$

for any $F \in \mathcal{F}$ in which

$$\mathcal{F} := \text{a basis generated by all open sets in } \mathbb{R}^r \text{ containing } x_0 \text{ as an interior point.}$$

Next, I will make the argument rigorous and generalize it to the scenario in which $\varpi(x)$ achieves its minimum on finite points of the support of X. See Figure 3 for an illustration of this type of boundary.

The Skorohod representation in Lemma 7.11 of Van der Vaart (2000) provides a measurable map g on $\mathbb{R}^r \times [0,1]$ and a random variable ε which is uniformly distributed on $[0,1]$, such that $Y_1 = g(X, \varepsilon)$, $X \perp\!\!\!\perp \varepsilon$. On top of this, I assume:

Assumption 14. *The measurable map g is lower semi-continuous.*

The conditional boundary obtains a finite set of minimizers; that is,

Assumption 15. *$\varpi(x) > -\infty$ and is minimized at $S_0 = \{x_t\}_{t=1}^T$ for some positive integer $T < +\infty$.*

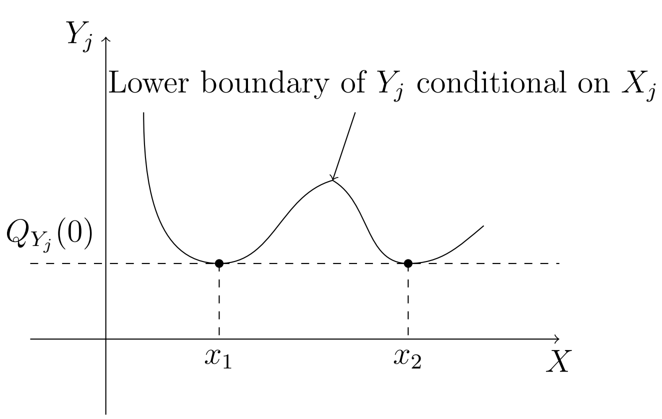


Figure 3: Finite minimizers

Now I characterize the weak limit $P_j^+(\mathcal{X}_j \in \cdot | Y_j = q_j(0))$ in Assumption 8 under Assumption 14 and 15. For each y, let S_y be the support of random variable $\lambda(X, y)$ where $\lambda(x,y) := \Pr(g(x,\varepsilon) \le y)$.

For a fixed y_0, define $S_{y_0} := \cup_t S_{y_0,t}$ where $\{S_{y_0,t}\}$ is a partition of S_{y_0} such that for $t' \neq t$, $x_t \in S_{y_0,t}$ and $d(x_{t'}, S_{y_0,t}) > 0$. For $y \leq y_0$, $S_{y,t} := S_{y_0,t} \cap S_y$ and $p_{y,t} := \frac{E\mathbb{1}\{X \in S_{y,t}\}\frac{\partial \lambda(X,y)}{\partial y}}{E\mathbb{1}\{X \in S_y\}\frac{\partial \lambda(X,y)}{\partial y}}$.

Assumption 16. *$\lim_{y \to q_1(0)} p_{y,t}$ exists and is equal to p_t.*

If Assumption 15 holds with $T = 1$, Assumption 16 holds with $p_1 = 1$ automatically. Given Assumption 16, the asymptotic objective function becomes

$$-kz + \sum_{i=1}^{\infty} \frac{\mathcal{D}_{i,1,f}}{P(\mathcal{X}_{i,1,f})} l_\delta(\mathcal{J}_{i,1,f}, z),$$

in which $\{\mathcal{E}_{i,1,f}, \mathcal{D}_{i,1,f}, \mathcal{X}_{i,1,f}\}$ is a sequence of i.i.d. random vectors, $\mathcal{E}_{i,1,f}$ is standard exponentially distributed, independent of $(\mathcal{X}_{i,1,f}, \mathcal{D}_{i,1,f})$, $\mathcal{J}_{i,1,f} := h_1^{-1}(\sum_{l=1}^{i} \mathcal{E}_{l,1,f})$, $\mathcal{D}_{i,1,f}$ is a Bernoulli distributed random variable with success probability $P(\mathcal{X}_{i,1,f})$ conditional on $\mathcal{X}_{i,1,f}$, $P(\cdot)$ is the propensity score, and $\mathcal{X}_{i,1,f}$ is supported by S_0 with corresponding point mass probabilities $\{p_t\}_{t=1}^{T}$.

Corollary A.1. *If Assumptions 1, 4, 6, 7, and 9-16 hold, then*

$$\hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg\min_{z \in \mathbb{R}} -kz + \sum_{i=1}^{\infty} \frac{\mathcal{D}_{i,1,f}}{P(\mathcal{X}_{i,1,f})} l_\delta(\mathcal{J}_{i,1,f}, z).$$

Examples 1 and 2 in Appendix B demonstrate the asymptotic distributions of this type.

A.2 Continuum minimizers

Next, I consider the conditional boundary in a case when it has continuum of minimizers; that is, a case in which it is flat over X. See Figure 4 for an illustration of the boundary. Recall $U_1 = Y_1 - q_1^*$. Then, I have

$$P(X \in F | Y_1 = y) = \frac{\int_F f_{U_1}(y - q_1^* | x) dF_X(x)}{\int f_{U_1}(y - q_1^* | x) dF_X(x)},$$

in which f_{U_1} is the conditional density of U_1. If $\varpi(x)$ is flat, I can adapt the independence at infinity condition assumed in both Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011).

Assumption 17. *$\varpi(x) \geq -\infty$ is flat, i.e. $\varpi(x) = q_1(0)$ for $x \in Supp(X)$ and there exists a random variable ε_1 such that*
(1) for $u \to 0$, uniformly over X, $F_{U_1}(u|X) \sim F_{\varepsilon_1}(\frac{u}{\sigma_1(X)})$ and $f_{U_1}(u|X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1}(\frac{u}{\sigma_1(X)})$,
(2) $\inf_x \sigma_1(x) > 0$,
(3) ξ_1, the EV index of both U_1 and ε_1, is nonzero.

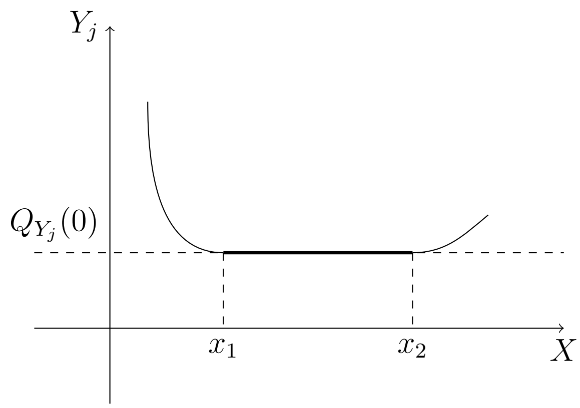


Figure 4: Continuum of minimizers

I allow the lower endpoint to be $-\infty$. Assumption 17(1) means U_1 behaves as $\sigma_1(X)\varepsilon_1$ at its lower tail and $X \perp\!\!\!\perp \varepsilon_1$. Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011) propose exactly this independence-at-tail condition. Resnick (1987) Proposition 0.7 shows that

$$f_{U_1}(u|X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1}\left(\frac{u}{\sigma_1(X)}\right)$$

holds point-wise by taking derivatives on both sides of $F_{U_1}(u|X) \sim F_{\varepsilon_1}(\frac{u}{\sigma_1(X)})$. Assumption 17(1) goes one-step further than Resnick (1987) Proposition 0.7, requires that

$$f_{U_1}(u|X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1}\left(\frac{u}{\sigma_1(X)}\right)$$

holds uniformly. The uniformity is not strong, given that $\text{Supp}(X)$ is compact. It can be relaxed to hold point-wisely with an envelope condition as illustrated in D'Haultfoeuille et al. (2015). Based on Assumption 17,

$$f_{U_1}(y - q_1^*|X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1}\left(\frac{y - q_1^*}{\sigma_1(X)}\right) \sim \sigma_1(X)^{1/\xi_1} f_{\varepsilon_1}(y - q_1^*)$$

uniformly over X.

Under the conditional independence at the tail, as $y \to q_1(0)$, I have

$$P(X \in F|Y_1 = y) \to \frac{\int_F \sigma_1(x)^{1/\xi_1} dF_X(x)}{\int_{\text{Supp}(X)} \sigma_1(x)^{1/\xi_1} dF_X(x)}.$$

Then, the asymptotic objective function becomes

$$-kz + \sum_{i=1}^{\infty} \frac{\mathcal{D}_{i,1,c}}{P(\mathcal{X}_{i,1,c})} l_\delta(\mathcal{J}_{i,1,c}, z),$$

in which $\{\mathcal{X}_{i,1,c}, \mathcal{D}_{i,1,c}, \mathcal{E}_{i,1,c}\}$ is i.i.d. sequence of random vectors, $\mathcal{X}_{i,1,c}$ is generated from the density

$$\frac{\sigma_1(x)^{1/\xi} dF_X(x)}{\int_{\text{Supp}(X)} \sigma_1(x) s^{1/\xi} dF_X(x)},$$

$\mathcal{D}_{i,1,c}$ is Bernoulli distributed with success probability $P(\mathcal{X}_{i,1,c})$ conditional on $\mathcal{X}_{i,1,c}$, $\mathcal{E}_{i,1,c}$ is a standard exponentially distributed random variable that is independent of $\mathcal{X}_{i,1,c}$ and $\mathcal{D}_{i,1,c}$, and $\mathcal{J}_{i,1,c} := h_1^{-1}(\sum_{l=1}^{i} \mathcal{E}_{l,1,c})$.

Corollary A.2. *If Assumptions 1, 4, 6, 7, 9 and 17 hold,*

$$\hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg\min_{z \in \mathbb{R}} -kz + \sum_{i=1}^{\infty} \frac{\mathcal{D}_{i,1,c}}{P(\mathcal{X}_{i,1,c})} l_\delta(\mathcal{J}_{i,1,c}, z).$$

Example 3 in Appendix B illustrates this type of asymptotic distribution.

A.3 Mixture Minimizers

Last, I combine the above two types of boundary structures and consider the case in which the minimizers of the conditional boundary is a mixture of discrete points and continuum intervals. See Figure 5 for an illustration. For two positive integers T and R, let $\varpi(x) > -\infty$ achieve its minimum on

$$x \in \{x_1, \cdots, x_R\} \cup (\cup_{t=1}^{T} S_{0,t}).$$

For each y, let S_y be the support of random variable $\lambda(X, y)$ where

$$\lambda(x, y) := \Pr(g(x, \varepsilon) \leq y).$$

For fixed y_0, let

$$\{\{S_{y_0,r}^d\}_{r=1}^{R}, \{S_{y_0,t}^c\}_{t=1}^{T}\}$$

be a partition of S_{y_0} such that (1) for all integers $r, r' = 1, 2, \cdots, R$ and $t, t' = 1, 2, \cdots, T$,

$$x_r \in S_{y_0,r}^d, \;\; S_{0,t} \subset S_{y,t}^c;$$

(2) for $r \neq r'$, $d(x_r, S_{y_0,r'}^d) > 0$; (3) for all t and r, $d(S_{y_0,t}^c, S_{y_0,r}^d) > 0$; and (4) for $t \neq t'$, $d(S_{y_0,t}^c, S_{y_0,t'}^c) > 0$. Finally, let

$$S_{y,r}^d := S_{y_0,r}^d \cap S_y, \quad p_{y,r}^d := \frac{E\mathbb{1}\{X \in S_{y,r}^d\} \frac{\partial \lambda(X,y)}{\partial y}}{E\mathbb{1}\{X \in S_y\} \frac{\partial \lambda(X,y)}{\partial y}}, \quad S_{y,t}^c := S_{y_0,t}^c \cap S_y,$$

and

$$p^c_{y,t} := \frac{E\mathbb{1}\{X \in S^c_{y,t}\}\frac{\partial \lambda(X,y)}{\partial y}}{E\mathbb{1}\{X \in S_y\}\frac{\partial \lambda(X,y)}{\partial y}}.$$

Assumption 18.
(1) $d(\cdot,\cdot)$ is the Euclidean distance between sets or between points and sets. Then

$$\min_{r\neq r'} d(x_r, x_{r'}) \wedge \min_{t\neq t'} d(S_{0,t}, S_{0,t'}) \wedge \min_{r\leq R, t\leq T} d(x_r, S_{0,t}) > \delta_0$$

for some positive δ_0.
(2) As $y \to q_1(0)$, $p^d_{y,r} \to p^d_r$ for $r = 1, 2, \cdots, R$ and $p^c_{y,t} \to p^c_t$ for $t = 1, 2, \cdots, T$.
(3) Let S^δ denote the δ-enlargement set $\{x|d(x,S) \leq \delta\}$; there then exists a positive constant δ such that for each $t = 1, 2, \cdots, T$, on $(S_{0,t})^\delta$, there exist ε_t with EV index $\xi_t < 0$ and σ_t such that, uniformly in $x \in (S_{0,t})^\delta$,

$$f_{U_1}(y - q_1(0)|X = x) \sim \frac{1}{\sigma_t(x)} f_{\varepsilon_t}(\frac{y - q_1(0)}{\sigma_t(x)}) \sim \sigma_t(x)^{-1/\xi_t} f_{\varepsilon_t}(y - q_1(0)).$$

(4) $\min_{t\leq T} \inf_x \sigma_t(x) > 0$.

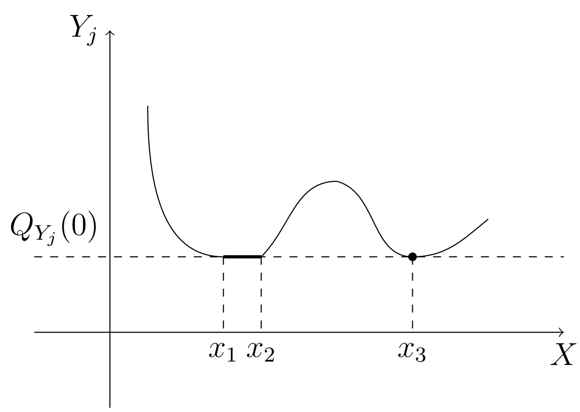


Figure 5: Mixture of minimizers

Next I define the asymptotic objective function for the mixture boundary case:

$$-kz + \sum_{i=1}^{\infty} \frac{\mathcal{D}_{i,1,m}}{P(\mathcal{X}_{i,1,m})} l_\delta(\mathcal{J}_{i,1,m}, z),$$

in which $\{\mathcal{E}_{i,1,m}, \mathcal{D}_{i,1,m}, \mathcal{X}_{i,1,m}\}$ is an i.i.d. sequence of random vectors, $\mathcal{E}_{i,1,m}$ is standard exponentially distributed, independent of both $\mathcal{X}_{i,1,m}$ and $\mathcal{D}_{i,1,m}$, $\mathcal{J}_{i,1,m} := h_1^{-1}(\sum_{l=1}^i \mathcal{E}_{l,1,m})$, $\mathcal{D}_{i,1,m}$ is Bernoulli distributed with success probability $P(\mathcal{X}_{i,1,m})$ conditional on $\mathcal{X}_{i,1,m}$, $\mathcal{X}_{i,1,m}$ is supported on $\{x_1, \cdots, x_R\} \cup (\cup_{t=1}^T S_{0,t})$, with its distribution being that, for any Borel set B,

$$P(\mathcal{X}_{i,1,m} \in B) = \sum_{r=1}^{R} \mathbb{1}\{x_r \in B\} P_r^d + \sum_{t=1}^{T} p_t^c \int_{S_{0,t}\cap B} \frac{\sigma_t(x)^{1/\xi_t} dF_X(x)}{\int_{S_{0,t}} \sigma_t(x)^{1/\xi_t} dF_X(x)}.$$

Corollary A.3. *If Assumptions 1, 4, 6, 7, 9, 14 and 18 hold, then*

$$\hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg\min_{z\in\mathbb{R}} -kz + \sum_{i=1}^{\infty} \frac{\mathcal{D}_{i,1,m}}{P(\mathcal{X}_{i,1,m})} l_\delta(\mathcal{J}_{i,1,m}, z).$$

Example 4 in Appendix B describes this type of asymptotic distribution.

B Illustrative examples

In this section, I consider four different types of conditional boundaries of Y_1 given X: single minimizer, multiple minimizers, continuum minimizers, and mixture minimizers. For each of the boundary behavior, I compute the limiting objective function based on the theoretical results in Appendix A. The results derived in this section are further used as the baseline models for the simulation study.

Example 1 (Single minimizer):

$$Y_1 = 0.5 + (X - 0.2)^2 + \varepsilon,\ D = \mathbb{1}\{\eta \le P(x)\},\ P(x) = 0.25 + x^2/2,$$

in which $X \sim$ Uniform$[0,1]$, $\varepsilon \sim$ Beta$(1,2)$, $\eta \sim$ Uniform$[0,1]$, X, ε, η are independent.

In this example, $\varpi(x)$, the conditional boundary of Y, is equal to $0.5 + (X - 0.2)^2$ and has a unique minimizer at $x = 0.2$. In addition, the EV index for Y is $-1/1.5$.[8] Hence by Corollary A.1, sequence $(\mathcal{D}_i, \mathcal{E}_i)$ is i.i.d, $\mathcal{D}_i$ is Bernoulli distributed with success probability $P(0.2)$, $\mathcal{E}_i \perp\!\!\!\perp \mathcal{D}_i$, $\mathcal{J}_i = (\sum_{l=1}^i \mathcal{E}_i)^{1/1.5}$ in which $\mathcal{E}_i$ is standard exponentially distributed, and

$$\hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg\min_{z\in\mathbb{R}} -kz + \sum_{i} \frac{\mathcal{D}_i}{P(0.2)} l_\delta(\mathcal{J}_i, z).$$

Example 2: (Multiple minimizers)

$$Y_1 = 0.5 + (|X - 0.3| - 0.1)^2 + \varepsilon,\ D = \mathbb{1}\{\eta \le P(x)\},\ P(x) = 0.25 + x^2/2,$$

in which $X \sim$ Uniform$[0,1]$, $\varepsilon \sim$ Beta$(1,2)$, $\eta \sim$ Uniform$[0,1]$, X, ε, η are independent.

In this example, $\varpi(x)$, the conditional boundary of Y, is $0.5 + (|X - 0.3| - 0.1)^2$ and has two minimizers $x_1 = 0.2$ and $x_2 = 0.4$. In addition, $S_{y,1} = [0.2 - \sqrt{y - 0.5}, 0.2 + \sqrt{y - 0.5}]$, $S_{y,2} =$

[8] In general, the EV index is $-1/(\alpha + 0.5)$ where α is the first parameter of the Beta distribution.

$[0.4 - \sqrt{y-0.5}, 0.4 + \sqrt{y-0.5}]$, and $p_1 = p_2 = 1/2$. Again, the EV index for Y is $-1/1.5$.[9] Hence by Corollary A.1, sequence $(\mathcal{D}_i, \mathcal{X}_i, \mathcal{E}_i)$ is i.i.d, $\mathcal{D}_i$ is Bernoulli distributed with success probability $P(\mathcal{X}_i)$ conditional on $\mathcal{X}_i$, $\mathcal{X}_i$ is equal to $x_1 = 0.2$ or $x_2 = 0.4$ with equal probability, $\mathcal{E}_i \perp\!\!\!\perp (\mathcal{X}_i, \mathcal{D}_i)$, $\mathcal{J}_i = (\sum_{l=1}^{i} \mathcal{E}_i)^{1/1.5}$ where $\mathcal{E}_i$ is standard exponentially distributed, and

$$\hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg\min_{z\in\mathbb{R}} -kz + \sum_i \frac{\mathcal{D}_i}{P(\mathcal{X}_i)} l_\delta(\mathcal{J}_i, z).$$

Example 3: (Continuum minimizers)

$$Y_1 = 0.5 + (X+0.5)\varepsilon,\ D = \mathbb{1}\{\eta \le P(x)\},\ P(x) = 0.25 + x^2/2,$$

in which $X \sim$ Uniform$[0,1]$, $\varepsilon \sim$ Beta$(1,2)$, $\eta \sim$ Uniform$[0,1]$, X, ε, η are independent.

In this example, $\varpi(x)$, the conditional boundary of Y is flat. It is easy to compute that the EV index of Y is -1 ($-1/\alpha$ in general where α is the first parameter of the Beta distribution). Hence by Corollary A.2, sequence $(\mathcal{D}_i, \mathcal{X}_i, \mathcal{E}_i)$ is i.i.d, $\mathcal{D}_i$ is Bernoulli distributed with success probability $P(\mathcal{X}_i)$ conditional on $\mathcal{X}_i$, $\mathcal{X}_i$ is continuously distributed over $[0,1]$ with density $x+0.5$.[10] $\mathcal{E}_i \perp\!\!\!\perp (\mathcal{D}_i, \mathcal{X}_i)$, $\mathcal{J}_i = \sum_{l=1}^{i} \mathcal{E}_i$ where $\mathcal{E}_i$ is standard exponentially distributed, and

$$\hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg\min_{z\in\mathbb{R}} -kz + \sum_i \frac{\mathcal{D}_i}{P(\mathcal{X}_i)} l_\delta(\mathcal{J}_i, z).$$

Example 4: (Mixture minimizers)

$$Y_1 = 0.5 + (|X-0.3| - 0.1)^2 \mathbb{1}\{X \in [0, 0.6)\} + (\mathbb{1}\{X > 0.5\} - \mathbb{1}\{X \in [0.7, 0.8]\}) + (X+0.5)\varepsilon,$$

$$D = \mathbb{1}\{\eta \le P(x)\},\ P(x) = 0.25 + x^2/2,$$

in which X takes value 0.2 with probability 0.1, 0.4 with probability 0.1 and is uniformly distributed on $[0.5, 1]$. $\varepsilon \sim$ Beta$(1,2)$, $\eta \sim$ Uniform$[0,1]$, X, ε, η are independent.

In this example, $\varpi(x)$, the conditional boundary of Y, is

$$(|X-0.3| - 0.1)^2 \mathbb{1}\{X \in [0, 0.6)\} + (\mathbb{1}\{X > 0.5\} - \mathbb{1}\{X \in [0.7, 0.8]\}).$$

$\varpi(x)$ achieves its minimum at $x_1 = 0.2$, $x_2 = 0.4$ and $x \in [0.7, 0.8]$. It is easy to compute that $p_1^d = 1/3.6$, $p_2^d = 1/3.6$, $p_1^c = 1.6/3.6$. Further more, the EV index for Y is -1.[11] Hence by

[9] In general, the EV index is $-1/(\alpha+0.5)$ where α is the first parameter of the Beta distribution.

[10] In general, the density is

$$\frac{(\frac{1}{\alpha}+1)(x+0.5)^{1/\alpha}}{1.5^{\frac{1}{\alpha}+1} - 0.5^{\frac{1}{\alpha}+1}}$$

where α is the first parameter of the Beta distribution.

[11] In general, the EV index is $-1/\alpha$, where α is the first parameter of the Beta distribution.

Corollary A.3, sequence $(\mathcal{D}_i, \mathcal{X}_i, \mathcal{E}_i)$ is i.i.d, $\mathcal{D}_i$ is Bernoulli distributed with success probability $P(\mathcal{X}_i)$ conditional on $\mathcal{X}_i$. $\mathcal{X}_i$ is a mixture distribution which has mass 1/3.6 at point 0.2, mass 1/3.6 at point 0.4, and is continuously distributed on $[0.7, 0.8]$ with density $\frac{32}{9}(x+0.5)$.[12] $\mathcal{E}_i \perp\!\!\!\perp (\mathcal{X}_i, \mathcal{D}_i)$, $\mathcal{J}_i = \sum_{l=1}^{i} \mathcal{E}_i$ where $\mathcal{E}_i$ is standard exponentially distributed, and

$$\hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg\min_{z \in \mathbb{R}} -kz + \sum_i \frac{\mathcal{D}_i}{P(\mathcal{X}_i)} l_\delta(\mathcal{J}_i, z).$$

C Simulation results

C.1 Details of simulation designs

For all DGPs, the error term ε_1 is generated from a Beta distribution with parameter $(1,2)$ and ε_0 is generated from a Beta distribution with parameter $(1.5, 2)$. They are independent of each other as well as covariate X. The treatment status $D = \mathbb{1}\{U \le P(x)\}$ where U is a uniformly distributed random variable independent of $(\varepsilon_1, \varepsilon_0, X)$ and $P(x)$ is the propensity score that takes the form of $0.25 + 0.5x^2$. The potential outcomes (Y_1, Y_0) are generated based on one of the following four models. For $j = 0, 1$,

1. Model (A_j):
$$Y_j = a_{1,j} + (X - a_{2,j})^2 + \varepsilon_j,$$
X is uniformly distributed on $[0,1]$, $(a_{1,1}, a_{2,1}) = (0.5, 0.2)$, and $(a_{1,0}, a_{2,0}) = (0.2, 0.3)$,

2. Model (B_j):
$$Y_j = b_{1,j} + (|X - b_{2,j}| - b_{3,j})^2 + \varepsilon_j,$$
X is uniformly distributed on $[0,1]$, $(b_{1,1}, b_{2,1}, b_{3,1}) = (0.5, 0.3, 0.1)$, and $(b_{1,0}, b_{2,0}, b_{3,0}) = (0.3, 0.2, 0.15)$.

3. Model (C_j):
$$Y_j = c_{1,j} + (X + c_{2,j})\varepsilon_j,$$
X is uniformly distributed on $[0,1]$, $(c_{1,1}, c_{2,1}) = (0.5, 0.5)$, and $(c_{1,0}, c_{2,0}) = (0.3, 0.2)$.

4. Model (D_j):
$$Y_j = d_{1,j} + (|X - d_{2,j}| - d_{3,j})^2 \mathbb{1}\{X < 0.6\} + (\mathbb{1}\{X > 0.5\} - \mathbb{1}\{0.7 < X < 0.8\}) + (X + 0.5)\varepsilon_j,$$
X takes values 0.2 or 0.4 with 0.1 probability and is uniform over $[0.5, 1]$, $(d_{1,1}, d_{2,1}, d_{3,1}) = (0.5, 0.3, 0.1)$, and $(d_{1,0}, d_{2,0}, d_{3,0}) = (0.3, 0.3, 0.1)$.

[12] In general, the density is
$$\frac{4(\frac{1}{\alpha}+1)(x+0.5)^{1/\alpha}}{9(1.3^{\frac{1}{\alpha}+1} - 1.2^{\frac{1}{\alpha}+1})}$$
where α is the first parameter of the Beta distribution.

The 16 simulation designs considered in Section 6 can be summarized in the following table where the first coordinate represents Y_1 and the second coordinate represents Y_0.

(A_1, A_0)	(A_1, B_0)	(A_1, C_0)	(A_1, D_0)
(B_1, A_0)	(B_1, B_0)	(B_1, C_0)	(B_1, D_0)
(C_1, A_0)	(C_1, B_0)	(C_1, C_0)	(C_1, D_0)
(D_1, A_0)	(D_1, B_0)	(D_1, C_0)	(D_1, D_0)

Table 1: Simulation designs used in Section 6.

C.2 Limiting distributions

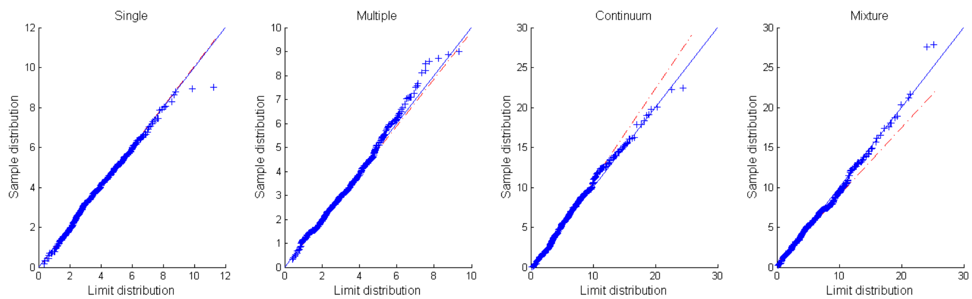


Figure 6: QQplot against EV law

To compute the sample estimator, I generate random samples with size 1,000 and repeat both the estimation and the minimization of the asymptotic objective function 400 times. $k := \tau_n n$ is set to 5. The propensity score is estimated in a sieve approach by fitting a series logistic model with ordinary polynomial basis to the fourth order.

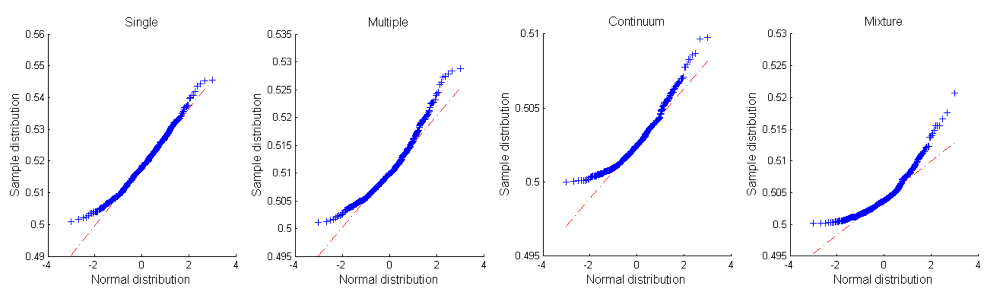


Figure 7: QQplot against Normal law

C.3 Inference for the extreme QTE

In the simulation, $n = 5,000$, k is fixed at $(5, 10, 20, 40)$, and the corresponding quantile indices are $\tau_n = (0.1\%, 0.2\%, 0.4\%, 0.8\%)$. The subsample size used in Table 2 and Figure 8 is 1,000. In Table 2, 3, Figure 8, and Figure 9, I consider four simulation designs corresponding to four different boundary conditions for both Y_1 and Y_0: (1) single minimizer, (2) multiple minimizers, (3) continuum minimizers, and (4) mixture minimizers. Table 2 and 3 report the coverages of BN-CI and NN-CI, respectively. The number in the parentheses is the median length of the CI. Figure 8 plots the coverages of BN-CI and NN-CI against τ for $\tau \in [0.1\%, 2\%]$. Figure 9 plots the coverage of BN-CI against b for $b \in [500, 1, 500]$.

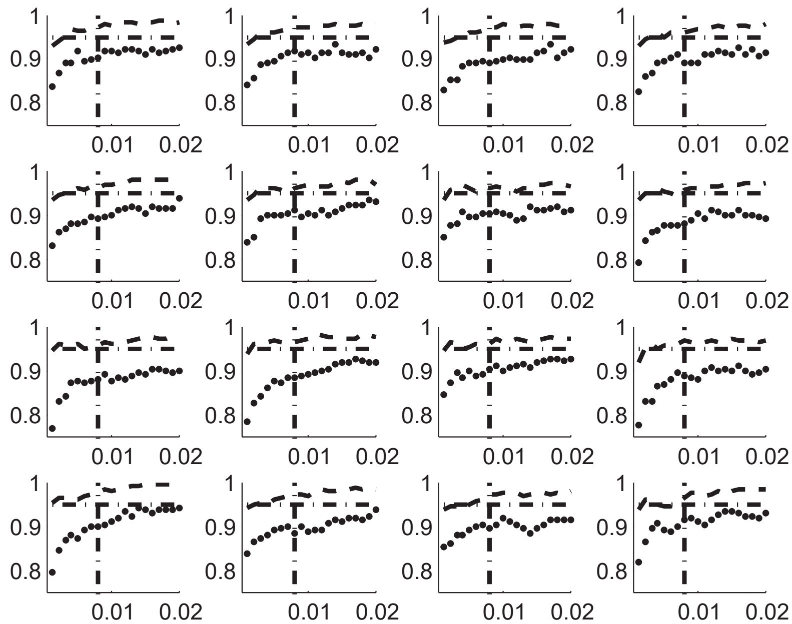
τ_n = 0.1%, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.2%, k = 10	(1)	(2)	(3)	(4)
(1)	0.941 (0.027)	0.936 (0.023)	0.941 (0.017)	0.940 (0.022)	(1)	0.955 (0.029)	0.948 (0.025)	0.95 (0.019)	0.949 (0.024)
(2)	0.948 (0.026)	0.944 (0.019)	0.943 (0.012)	0.941 (0.020)	(2)	0.953 (0.028)	0.942 (0.021)	0.961 (0.014)	0.949 (0.023)
(3)	0.957 (0.025)	0.948 (0.018)	0.947 (0.006)	0.939 (0.012)	(3)	0.959 (0.026)	0.957 (0.020)	0.966 (0.007)	0.956 (0.014)
(4)	0.954 (0.024)	0.938 (0.018)	0.940 (0.014)	0.935 (0.015)	(4)	0.959 (0.028)	0.949 (0.021)	0.941 (0.017)	0.950 (0.019)
τ_n = 0.4%, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.8%, k = 40	(1)	(2)	(3)	(4)
(1)	0.968 (0.030)	0.962 (0.026)	0.949 (0.020)	0.956 (0.027)	(1)	0.979 (0.037)	0.974 (0.032)	0.977 (0.026)	0.967 (0.033)
(2)	0.956 (0.029)	0.967 (0.023)	0.968 (0.015)	0.953 (0.027)	(2)	0.969 (0.034)	0.968 (0.027)	0.965 (0.020)	0.953 (0.033)
(3)	0.960 (0.028)	0.960 (0.022)	0.951 (0.008)	0.947 (0.016)	(3)	0.963 (0.031)	0.966 (0.025)	0.968 (0.011)	0.970 (0.021)
(4)	0.962 (0.030)	0.949 (0.022)	0.939 (0.018)	0.945 (0.021)	(4)	0.983 (0.041)	0.972 (0.033)	0.972 (0.027)	0.973 (0.031)

Table 2: Coverage of 95% b out of n bootstrap CI, sample size = 5,000

τ_n = 0.1%, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.2%, k = 10	(1)	(2)	(3)	(4)
(1)	0.835	0.842	0.829	0.825	(1)	0.869	0.855	0.853	0.86
	(0.023)	(0.019)	(0.014)	(0.017)		(0.024)	(0.020)	(0.016)	(0.019)
(2)	0.830	0.835	0.850	0.790	(2)	0.861	0.848	0.875	0.841
	(0.020)	(0.016)	(0.010)	(0.015)		(0.021)	(0.017)	(0.012)	(0.017)
(3)	0.768	0.783	0.844	0.775	(3)	0.828	0.824	0.873	0.830
	(0.018)	(0.013)	(0.004)	(0.009)		(0.019)	(0.015)	(0.005)	(0.011)
(4)	0.793	0.835	0.852	0.819	(4)	0.846	0.865	0.858	0.863
	(0.018)	(0.014)	(0.011)	(0.012)		(0.020)	(0.016)	(0.013)	(0.014)

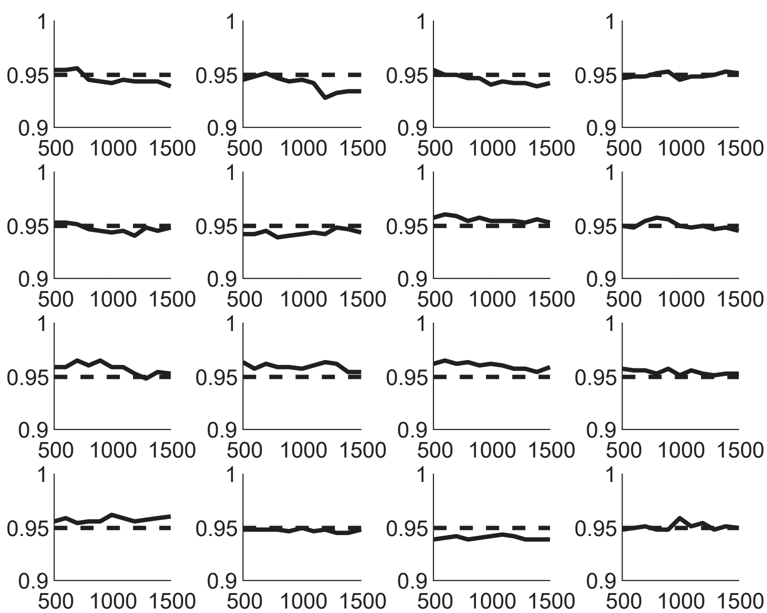
τ_n = 0.4%, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.8%, k = 40	(1)	(2)	(3)	(4)
(1)	0.891	0.891	0.882	0.891	(1)	0.903	0.919	0.890	0.892
	(0.025)	(0.022)	(0.018)	(0.021)		(0.027)	(0.024)	(0.020)	(0.025)
(2)	0.878	0.898	0.906	0.864	(2)	0.889	0.909	0.903	0.877
	(0.022)	(0.019)	(0.014)	(0.020)		(0.023)	(0.020)	(0.015)	(0.023)
(3)	0.871	0.860	0.882	0.865	(3)	0.879	0.881	0.903	0.885
	(0.020)	(0.016)	(0.006)	(0.012)		(0.020)	(0.017)	(0.007)	(0.015)
(4)	0.880	0.879	0.880	0.904	(4)	0.899	0.881	0.894	0.912
	(0.021)	(0.018)	(0.015)	(0.017)		(0.023)	(0.021)	(0.019)	(0.021)

Table 3: Coverage of 95% n out of n bootstrap CI, sample size $= 5,000$



Each (i, j)-th subplot represents the (i, j)-th model. The dashed line is the coverage of BN-CI with $b = 1,000$ and $n = 5,000$ for quantile index $\tau \in [0.1\%, 2\%]$. The dotted line is the coverage of NN-CI. The horizontal dotted dashed line is the 95% nominal coverage rate, and the vertical dotted dashed line is $\tau = \min(\frac{40}{n}, \frac{0.2b}{mn})$.

Figure 8: Coverage across quantiles

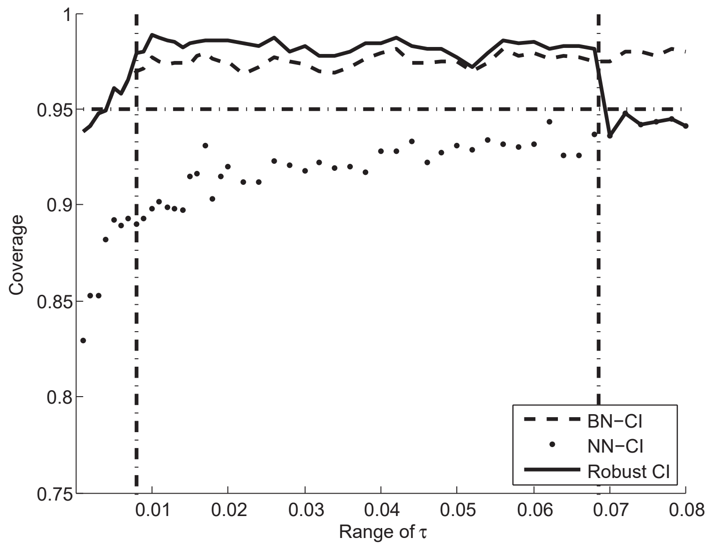


Each (i, j)-th subplot represents the (i, j)-th model. The solid line is the coverage for b out of n bootstrap CI at $\tau = 0.2\%$ in which $b \in [500, 1, 500]$.

Figure 9: Coverage across subsample size

Here I only report the results for sample size 5,000. The same simulation designs with sample size 300 and 1,000 can be found in the Appendix F.3. In Appendix F.3, I also show the mean bias (bias), root mean square error (rMSE), median bias (mbias), and mean absolute error (MAE) of the median-unbiased point estimator for small, moderate and large sample. The performance of the median-unbiased point estimator is satisfying in all samples.

C.4 The robust confidence interval



$b = 1,000$, $n = 5,000$, and $\tau \in [0.1\%, 8\%]$. The horizontal dotted dashed line is the 95% nominal coverage rate. $\tau_{n,1} = 0.8\%$ and $\tau_{n,2} = 6.85\%$.

Figure 10: Coverage across quantiles

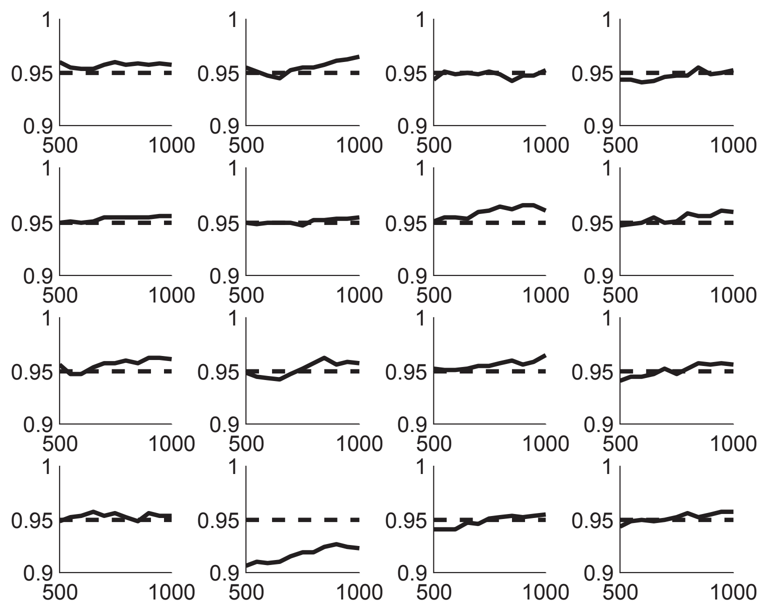
To produce Figure 10, the full sample size and subsample size are $n = 5,000$ and $b = 1,000$, respectively. Y_1 has a single minimizer and Y_0 has continuum minimizers. The quantile index $\tau \in [0.1\%, 8\%]$. For computing $\widetilde{C}_a^{bn}(\tau)$, when $\tau \le 2\%$ or equivalently, $k := \tau n \le 100$, I set the spacing parameter $m = 2$ and $k_l' = 10$.[13] When $\tau > 2\%$, I set $m = 1.2$ and $k_l' = 20$. Here I only report the simulation results for one model. In fact, all sixteen models exhibit this same pattern. For details, please see Appendix F.3.

[13] k_l' is used to compute the normalizing factor $\hat{\alpha}_n$.

C.5 Inference for the 0-th QTE

	(1)	(2)	(3)	(4)
(1)	0.960	0.955	0.951	0.948
	(0.118)	(0.086)	(0.058)	(0.077)
(2)	0.954	0.947	0.961	0.951
	(0.105)	(0.073)	(0.042)	(0.072)
(3)	0.957	0.952	0.955	0.948
	(0.082)	(0.056)	(0.017)	(0.036)
(4)	0.956	0.919	0.951	0.953
	(0.086)	(0.053)	(0.039)	(0.044)

Table 4: Coverage of 95% CI. Sample size is 5,000.



The solid line is the coverage for b out of n bootstrap CI at $\tau = 0$ in which $b \in [500, 1,000]$.

Figure 11: Coverage across subsample size

Here again I only focus on $n = 5,000$. The same simulation with $n = 300, 1,000$ can be found in the appendix. All the findings in Section 6 still hold.

There are two issues worth-mentioning when implementing the BN-CI for 0-th QTE. The first issue is that I use three extreme QTE estimators with $k = (5, 17.5, 30)$ to compute the linear combination. The choice of k invokes two concerns. First, the rule of thumb for $k = \tau n$ is $k \leq \min(40, \frac{0.2b}{m})$. Second, the space among k's must not be narrow, otherwise the weights will be large in absolute value, which will widen the CI.

The second issue is the estimatation of EV indices. I follow Theorem 3.4 with $R = 2$, $m = 2$, $l = 2$, and equal weights. The set of quantile indices I use to compute the EV indices are $\tau_n = (0.002, 0.004, \cdots, 0.01)$. Then for $j = 0, 1$, the two EV index estimators used to compute the weights $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3)$ are the median of the estimators computed using each of the quantile indices for $j = 0$ and 1, respectively.

The rest of the simulation details are the same as the ones in the previous subsection. The subsample size for Table 4 is 1,000.

D Data, implementation, and application results

D.1 Effect of maternal status on extremely low birth weights

To fit the notation in the paper, let D be an indicator of maternal smoking. The observed outcome variable Y is birth weight measured in grams, while $Y = DY_1 + (1 - D)Y_0$ where Y_1 is the infant's potential birth weight when the mother smokes and Y_0 is the infant's potential birth weight when the mother does not smoke. Covariates X are demographic variables which include mother's age, mother's education level[14], an indicator of whether the mother had parental care visit in the first and second trimester, mother's marriage status, the infant's sex, and mother's weight gain during pregnancy. The key unconfoundedness assumption in this context means that, maternal smoking is independent from the potential birth weights conditional on all the demographic variables.

Following the experience collected from Section 6, I set the subsample size to 3,000 and repeat the b out of n bootstrap with replacement 20,000 times. Also, I nonparametrically estimate the propensity by fitting a series logistic model with a set of second-order polynomial basis, and the spacing parameter m is set to 2.[15] When computing the 0-th QTE, I use a linear combination of extreme-order estimates with $k = (5, 20, 40)$. A set of estimators of EV index are computed following Theorem 3.4 with $R = 2$, $l = 2$ and $\tau_n = (0.0005, 0.001, 0.0015, 0.002, 0.0025)$. The final EV index estimators used are the median of the five estimators for Y_0 and Y_1, respectively.

D.2 Effect of minority status on college preparation index

In 1996, the voters of California approved Prop 209 which stipulates that: "The state shall not discriminate against, or grant preferential treatment to, any individual or group on the basis of race, sex, color, ethnicity, or national origin in the operation of public employment, public education, or public contracting." The proposition took effect in 1998. I use the same data as in Arcidiacono

[14] The education level equals 0 if the mother has less than a high school education, 1 if she completed high school, 2 if she obtained some college education, and 3 if she graduated from college.

[15] Here I implicitly assume that the sufficient condition for the spacing parameter in Lemma E.7 holds. In practice, neither the full sample nor any subsample estimation encounters the zero denominator error. Hence $m = 2$ behaves well in this data analysis.

	MU point estimates	90% BN-CI		90% NN-CI	
k=0	-137.32	-605.77	193.71		
k=5	-0.21	-198.08	87.09	-51.00	97.00
k=10	-5.57	-193.49	121.43	-82.00	84.00
k=15	30.64	-143.51	182.04	-63.00	108.00
k=20	16.12	-144.21	187.52	-72.00	107.00
k=25	-14.81	-179.51	163.69	-115.00	60.00
k=30	-19.11	-171.56	167.01	-139.00	45.00
k=35	10.87	-138.83	189.23	-68.00	114.00
k=40	-12.30	-169.21	153.74	-108.50	85.00

Table 5: Extreme order unconditional QTE of smoking status.

et al. (2016), the University of California Office of the President (UCOP) data for minority and non-minority students who first enrolled at one of the UC campuses in periods both prior and post-Prop 209, to compute the racial CPI gap at tails. The pre- and post-Prop 209 period data consist of students admitted between 1995 and 1997 and between 1998 and 2000, respectively.

The data for each UC campuses consist of all their admitted students. The outcome variable Y is normalized CPI.[16] The treatment status D is the indicator of under-represented minority groups in the dataset. X are two family background variables: family income percentage and two parents' highest education degree. Minority students may live in a less favorable family environment with low parental income and education level. This difference can cause minority students to be less prepared for college than their majority peers.

[16] As described in Arcidiacono et al. (2016), the raw preparation score (Y_i^{raw}) for student i is a weighted average of student's high school GPA (GPA_i) and their combined verbal and math SAT score (SAT_i): $Y_i^{raw} = \frac{3}{8} \cdot SAT_i + 400 \cdot GPA_i$. The CPI Y_i is the standardized version of Y_i^{raw} such that it has mean 0 and standard deviation 1 for the pool of applications to one or more of the UC campuses.

D.2.1 Pre-Prop 209

Campus	Berkeley	UCLA	San Diego	Davis	Irvine	Santa Barbara	Santa Cruz	Riverside
Science								
ATE	-0.893^{***}	-0.724^{***}	-0.510^{***}	-0.443^{***}	-0.312^{***}	-0.420^{***}	-0.351^{***}	-0.477^{***}
k=5	-0.732^{*}	-1.217^{***}	-0.298^{**}	-0.901^{**}	-0.136^{*}	-0.595^{**}	-0.276	-0.525^{**}
k=10	-0.857^{*}	-0.961^{***}	-0.397^{***}	-0.306^{*}	-0.288^{*}	-0.342^{**}	-0.300	-0.398^{***}
k=15	-1.023^{**}	-1.057^{***}	-0.421^{***}	-0.304^{*}	-0.338^{*}	-0.463^{***}	-0.285	-0.431^{**}
k=20	-0.886^{*}	-0.907^{***}	-0.449^{***}	-0.146^{*}	-0.401^{*}	-0.414^{**}	-0.590^{*}	-0.478^{**}
k=25	-0.927^{**}	-0.943^{***}	-0.466^{***}	-0.224^{*}	-0.449^{*}	-0.368^{**}	-0.505	-0.563^{***}
k=30	-0.952^{**}	-0.825^{***}	-0.438^{***}	-0.298^{**}	-0.472^{*}	-0.326^{**}	-0.573	-0.396^{**}
k=35	-0.986^{**}	-0.716^{***}	-0.212^{***}	-0.373^{**}	-0.508^{*}	-0.350^{**}	-0.539	-0.379^{***}
k=40	-0.997^{***}	-0.673^{***}	-0.188^{***}	-0.379^{**}	-0.433^{*}	-0.365^{**}	-0.529	-0.399^{***}
Non-Science								
ATE	-0.987^{***}	-0.761^{***}	-0.502^{***}	-0.539^{***}	-0.466^{***}	-0.466^{***}	-0.478^{***}	-0.424^{***}
k=5	0.183	-0.284	-0.647	-0.096	-0.347^{*}	0.169	-0.424	-0.459
k=10	-0.283	-0.869^{***}	-0.479	-0.450^{*}	-0.343^{*}	-0.321^{**}	-0.581^{**}	-0.529^{**}
k=15	-0.383^{*}	-0.988^{***}	-0.170	-0.227	-0.377^{*}	-0.359^{**}	-0.526^{***}	-0.464^{**}
k=20	-0.462^{**}	-0.949^{***}	-0.197	-0.299	-0.349^{*}	-0.419^{***}	-0.527^{***}	-0.540^{**}
k=25	-0.569^{**}	-0.878^{***}	-0.203	-0.371^{*}	-0.413^{**}	-0.458^{***}	-0.549^{***}	-0.570^{**}
k=30	-0.647^{***}	-0.861^{***}	-0.231	-0.360^{*}	-0.475^{**}	-0.459^{***}	-0.559^{***}	-0.544^{***}
k=35	-0.630^{***}	-0.886^{***}	-0.193	-0.392^{**}	-0.481^{***}	-0.402^{***}	-0.668^{***}	-0.578^{***}
k=40	-0.722^{***}	-0.869^{***}	-0.251	-0.386^{**}	-0.547^{***}	-0.424^{***}	-0.671^{***}	-0.567^{***}

The sample size (subsample size) for students with a science major and campus from Berkeley to Riverside are 4126 (700), 4204 (700), 4122 (700), 4298 (700), 3877 (700), 2704 (600), 1345 (350), 1641 (375). For students with non-science major, they are 4990 (750), 5837 (775), 3749 (650), 5105 (750), 4154 (650), 6674 (800), 3775 (650), 2784 (500). $*$, $**$, and $***$ indicate 90%, 95%, and 99% significance level, respectively. I use standard bootstrap CI for the inference of ATE and BN-CI for extreme QTE.

Table 6: Index gap across campus and initial major

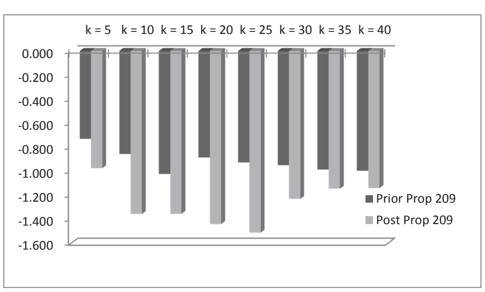
D.2.2 Post-Prop 209

Campus	Berkeley	UCLA	San Diego	Davis	Irvine	Santa Barbara	Santa Cruz	Riverside
Science								
ATE	-0.681^{***}	-0.416^{***}	-0.133^{**}	-0.273^{***}	-0.271^{***}	-0.262^{***}	-0.321^{***}	-0.260^{***}
k=5	-0.976^{**}	-0.460^{**}	-0.113	0.207	0.108	-0.710^{*}	-0.673^{**}	-0.214
k=10	-1.357^{**}	-0.583^{**}	-0.065	0.061	0.125	-0.716^{*}	-0.558^{**}	-0.159
k=15	-1.356^{**}	-0.637^{**}	-0.076	0.002	-0.090	-0.689	-0.450^{**}	-0.114
k=20	-1.441^{**}	-0.680^{**}	-0.044	-0.078	-0.116	-0.433	-0.440^{**}	-0.160
k=25	-1.512^{**}	-0.706^{**}	-0.038	-0.063	-0.144	-0.500	-0.469^{**}	-0.187
k=30	-1.232^{**}	-0.758^{**}	-0.057	-0.050	-0.196	-0.476	-0.375^{**}	-0.142
k=35	-1.146^{**}	-0.676^{**}	-0.111	-0.124	-0.156	-0.484	-0.385^{**}	-0.118
k=40	-1.141^{**}	-0.616^{**}	-0.117	-0.143	-0.172	-0.399	-0.367^{**}	-0.097
Non-Science								
ATE	-0.671^{***}	-0.607^{***}	-0.149^{***}	-0.264^{***}	-0.304^{***}	-0.388^{***}	-0.263^{***}	-0.302^{***}
k=5	-0.548^{***}	-0.541	-0.141	0.076	-0.182	-0.590^{*}	-0.254	-0.478^{***}
k=10	-0.628^{***}	-0.637	-0.176	0.045	-0.374	-0.217	-0.525^{***}	-0.385^{***}
k=15	-0.552^{***}	-0.544	-0.044	-0.050	-0.403	-0.307	-0.460^{**}	-0.390^{***}
k=20	-0.541^{***}	-0.344	-0.018	-0.165	-0.267	-0.285	-0.426^{**}	-0.417^{***}
k=25	-0.633^{***}	-0.522	-0.066	-0.146	-0.362	-0.285	-0.396^{*}	-0.403^{***}
k=30	-0.703^{***}	-0.552	-0.064	-0.125	-0.427^{*}	-0.303	-0.374^{**}	-0.413^{***}
k=35	-0.705^{***}	-0.645	-0.064	-0.147	-0.486^{**}	-0.297	-0.377^{**}	-0.428^{***}
k=40	-0.704^{***}	-0.665	-0.079	-0.205	-0.509^{**}	-0.320	-0.357^{**}	-0.441^{***}

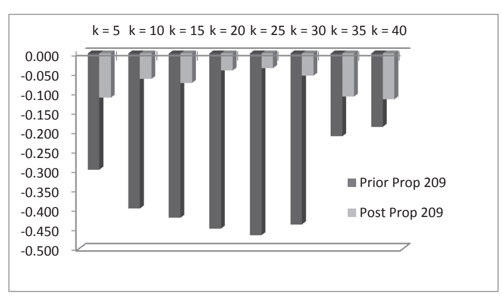
The sample size (subsample size) for students with a science major and campus from Berkeley to Riverside are 3906 (700), 4159 (700), 3861 (700), 4319 (700), 4361 (700), 2594 (600), 1596 (350), 2180 (375). For students with non-science major, they are 4695 (750), 6029 (775), 4024 (650), 5418 (750), 4432 (650), 6108 (800), 4537 (650), 4529 (500). $*$, $**$, and $***$ indicate 90%, 95%, and 99% significance level, respectively. I use standard bootstrap CI for the inference of ATE and BN-CI for extreme QTE.

Table 7: Index gap across campus and initial major

D.3 Prior and post-Prop 209 comparison



(a) Berkeley Science



(b) San Diego Science

Figure 12: Minority gaps pre- and post-Prop 209

Science	ATE	k=5	k=10	k=15	k=20	k=25	k=30	k=35	k=40
Berkeley	-0.213^{***}	0.292	0.535	0.356	0.574	0.611	0.291	0.165	0.148
San Diego	-0.376^{***}	-0.170^{*}	-0.322^{***}	-0.344^{***}	-0.408^{***}	-0.433^{***}	-0.381^{***}	-0.106	-0.077

Table 8: Difference of the racial gaps

E Theoretical proofs

E.1 Proof of Theorem 3.1

Before starting the proof, I first state a maximal inequality which is derived in Chernozhukov, Chetverikov, and Kato (2014). See Corollary 5.1 in their paper. Let $(X_1, \cdots, X_n)$ be a sequence of i.i.d random variables taking values in a measurable space $(S, \mathcal{S})$ with common distribution P. $\mathcal{F}$ is a generic class of measurable function $S \to \mathbb{R}$ with an envelope function F. Let $\sigma^2 > 0$ be any positive constant such that

$$\sup_{f \in \mathcal{F}} Pf^2 \leq \sigma^2 \leq ||F||_{P,2}^2 \quad \text{and} \quad M = \max_{1 \leq i \leq n} F(X_i).$$

Lemma E.1. *If $F \in L^2(P)$ and suppose that there exist constants $a \geq e$ and $v \geq 1$ such that the following uniform entropy condition holds:*

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

then

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

Throughout the appendix, for simplicity of notation, I call a term $U_n(k) = o_p^*(r_n)$ $(O_p^*(r_n))$ if

$$\sup_{k \in [\kappa_1, \kappa_2]} \left|\frac{U_n(k)}{r_n}\right| = o_p(1)(O_p(1))$$

for some fixed positive constants κ_1 and κ_2.

Now I return to the proof of Theorem 3.1. Let $\hat{\Delta}_{1,n}(k) = \lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n))$ be the maximizer of the rescaled objective function, that is,

$$\hat{\Delta}_{1,n}(k) = \arg\min_{\Delta \in \mathbb{R}} -\hat{W}_n(k)\Delta(k) + \hat{G}_n(\Delta, k) \tag{E.1}$$

where

$$\hat{W}_n(k) = \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} (k\tau_n - \mathbb{1}\{Y_i \le q_1(k\tau_n)\}),$$

$$\hat{G}_n(\Delta, k) = \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} \int_0^{\Delta} \left(\mathbb{1}\left\{ Y_i \le q_1(k\tau_n) + \frac{s}{\lambda_{1,n}(k)} \right\} - \mathbb{1}\left\{ Y_i \le q_1(k\tau_n) \right\} \right) ds.$$

The proof of the first part of the theorem is divided into three steps. In the first step, by defining

$$R_n(\Delta, k) = \hat{G}_n(\Delta, k) - \frac{\Delta^2}{2},$$

I show that

$$\sup_{|\Delta| \le M, k \in [\kappa_1, \kappa_2]} |R_n(\Delta, k)| = o_p(1). \tag{E.2}$$

In the second step, I show that

$$\hat{W}_n(k) = W_n(k) + o_p^*(1)$$

where

$$W_n(k) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{i,1,n}(k)$$

and

$$\phi_{i,1,n}(k) = \frac{1}{\sqrt{k\tau_n}} \left[\frac{D_i}{P(X_i)} T_{i,1,n}(k) - \frac{\mathbb{E}(T_{i,1,n}(k)|X_i)}{P(X_i)} (D_i - P(X_i)) \right].$$

In the third step, I show that $\{W_n(k) : k \in [\kappa_1, \kappa_2]\}$ is tight. This implies that $\{\hat{W}_n(k) : k \in [\kappa_1, \kappa_2]\}$ is tight too. Given the tightness of $\{\hat{W}_n(k) : k \in [\kappa_1, \kappa_2]\}$ and (E.2), I can apply a generalized version of the Convexity lemma in Pollard (1991) proved in Lemma 2 of Chernozhukov (2000), I can conclude that

$$\hat{\Delta}_{n,1}(k) = \hat{W}_n(k) + o_p^*(1) = W_n(k) + o_p^*(1)$$

and $\{\hat{\Delta}_{n,1}(k) : k \in [\kappa_1, \kappa_2]\}$ is tight. Similarly, I can show that

$$\lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{0,i,n}(k) + o_p^*(1)$$

where

$$\phi_{0,i,n}(k) = \frac{1}{\sqrt{k\tau_n}} \left[\frac{1 - D_i}{1 - P(X_i)} T_{i,0,n}(k) + \frac{\mathbb{E}(T_{i,0,n}(k)|X_i)}{1 - P(X_i)} (D_i - P(X_i)) \right]$$

and that the stochastic process $\{\phi_{0,i,n}(k) : k \in [\kappa_1, \kappa_2]\}$ is tight. This concludes the first half of the results in Theorem 3.1.

Step 1.

Define

$$G_n(\Delta,k)=\frac{1}{\sqrt{nk\tau_n}}\sum_{i=1}^{n}\frac{D_i}{P(X_i)}\int_0^{\Delta}\left(\mathbb{1}\left\{Y_i\le q_1(k\tau_n)+\frac{s}{\lambda_{1,n}(k)}\right\}-\mathbb{1}\left\{Y_i\le q_1(k\tau_n)\right\}\right)ds.$$

By Lemma 1 in Hirano et al. (2003), $\sup_x|\hat{P}(x)-P(x)|=o_p(1)$. In addition, $\inf_x P(x)$ is bounded away from zero. Therefore,

$$\sup_x|\frac{1}{\hat{P}(x)}-\frac{1}{P(x)}|=o_p(1).$$

Then, uniformly over $|\Delta|\le M$,

$$\begin{aligned}|\hat{G}_n(\Delta,k)-G_n(\Delta,k)|\le& o_p(1)\Bigg[\frac{M}{\sqrt{nk\tau_n}}\sum_{i=1}^{n}\Bigg(\mathbb{1}\left\{Y_i\le q_1(k\tau_n)+\frac{M}{\lambda_{1,n}(k)}\right\}-\mathbb{1}\left\{Y_i\le q_1(k\tau_n)\right\}\\&+\mathbb{1}\left\{Y_i\le q_1(k\tau_n)\right\}-\mathbb{1}\left\{Y_i\le q_1(k\tau_n)-\frac{M}{\lambda_{1,n}(k)}\right\}\Bigg)\Bigg]\\\lesssim& o_p(1)|\sqrt{n}\mathcal{P}_nf||_{\mathcal{F}_{1,n}}\lesssim o_p(1)(||\sqrt{n}(\mathcal{P}_n-P)||_{\mathcal{F}_{1,n}}+\sqrt{n}|Pf|_{\mathcal{F}_{1,n}})\end{aligned}\tag{E.3}$$

where

$$\begin{aligned}\mathcal{F}_{1,n}=\Bigg\{\frac{1}{\sqrt{\tau_n}}\Bigg(&\mathbb{1}\left\{Y_i\le q_1(k\tau_n)+\frac{M}{\lambda_{1,n}(k)}\right\}-\mathbb{1}\left\{Y_i\le q_1(k\tau_n)\right\}\\&+\mathbb{1}\left\{Y_i\le q_1(k\tau_n)\right\}-\mathbb{1}\left\{Y_i\le q_1(k\tau_n)-\frac{M}{\lambda_{1,n}(k)}\right\}\Bigg),k\in[\kappa_1,\kappa_2]\Bigg\},\end{aligned}$$

with an envelope function

$$\begin{aligned}F_{1,n}=\frac{1}{\sqrt{\tau_n}}\Bigg(&\mathbb{1}\left\{Y_i\le q_1(\kappa_2\tau_n)+\frac{M}{\underline{\lambda}_{1,n}}\right\}-\mathbb{1}\left\{Y_i\le q_1(\kappa_1\tau_n)\right\}\\&+\mathbb{1}\left\{Y_i\le q_1(\kappa_2\tau_n)\right\}-\mathbb{1}\left\{Y_i\le q_1(\kappa_1\tau_n)-\frac{M}{\underline{\lambda}_{1,n}}\right\}\Bigg).\end{aligned}$$

Note that $f(q_1(k\tau_n))$ is monotone in k for n large enough and $k\in[\kappa_1,\kappa_2]$. Hence $\lambda_{1,n}(k)\ge\overline{\lambda}_{1,n}:=\frac{\sqrt{n}}{\sqrt{\kappa_2\tau_n}}f_1(q_1(\underline{k}\tau_n))$ where $\underline{k}=\kappa_1$ or κ_2 depends on whether f_1 is monotone decreasing or increasing at the tail. Then I have

$$||F_{1,n}||_{P,2}\le C<\infty,\quad M_{1,n}=\max_{1\le i\le n}F_{1,n}\le\frac{2}{\sqrt{\tau_n}}.$$

Furthermore, $q_1(\cdot\tau_n)$ and $\lambda_{1,n}(\cdot)$ are monotone. So by repeatedly using Lemma 2.6.18 (iv), (v), and (viii) of Van der Vaart and Wellner (1996), I have

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

By Lemma E.1 with $\sigma = ||F_{1,n}||_{P,2}$, I have

$$\mathbb{E}||\sqrt{n}(\mathcal{P}_n - P)||_{\mathcal{F}_{1,n}} \lesssim ||F_{1,n}||_{P,2} + \frac{1}{\sqrt{\tau_n n}} = O(1)$$

and thus

$$||\sqrt{n}(\mathcal{P}_n - P)||_{\mathcal{F}_{1,n}} = O_p(1). \tag{E.4}$$

I next want to show$\sqrt{n}|Pf|_{\mathcal{F}_{1,n}} = O(1)$. In fact, I have

$$\sqrt{n}|Pf|_{\mathcal{F}_{1,n}} \lesssim \sup_{k\in[\kappa_1,\kappa_2]} \left(\frac{f_1\left(q_1(k\tau_n) + \frac{\tilde{M}}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))} + \frac{f_1\left(q_1(k\tau_n) - \frac{\tilde{M}}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))} \right)$$

where $\tilde{M}$ is between zero and M. Since $\tau_n n \to \infty$, for any constant $l > 1$ independent of k, there exists $N_0 > 0$ independent of k such that for $n > N_0$,

$$\frac{\tilde{M}}{\lambda_{1,n}(k)} = \frac{\tilde{M}(q_1(lk\tau_n) - q_1(k\tau_n))}{\sqrt{n\tau_n}\int_k^{lk} \frac{f(q_1(k\tau_n))}{f(q_1(t\tau_n))}dt} \leq (q_1(lk\tau_n) - q_1(k\tau_n)). \tag{E.5}$$

Therefore, if f_1 is monotone increasing at its tail,

$$\sup_{k\in[\kappa_1,\kappa_2]} \frac{f_1\left(q_1(k\tau_n) + \frac{\tilde{M}}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))} + \frac{f_1\left(q_1(k\tau_n) - \frac{\tilde{M}}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))} \leq \frac{f_1(q_1(l\kappa_2\tau_n))}{f_1(q_1(\kappa_1\tau_n))} + 1 = O(1).$$

Similar argument shows $\sup_{k\in[\kappa_1,\kappa_2]} \frac{f_1\left(q_1(k\tau_n) + \frac{\tilde{M}}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))} + \frac{f_1\left(q_1(k\tau_n) - \frac{\tilde{M}}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))} = O(1)$ when f_1 is monotone decreasing at its tail. So I obtain the desired result that

$$\sqrt{n}|Pf|_{\mathcal{F}_{1,n}} = O_p(1). \tag{E.6}$$

Combining (E.1), (E.4), and (E.6), I have

$$\sup_{\Delta,k} |\hat{G}_n(\Delta, k) - G_n(\Delta, k)| = o_p(1). \tag{E.7}$$

Next, I want to show $G_n(\Delta, k) \to \frac{\Delta^2}{2}$ uniformly in $|\Delta| \leq M$ and $k \in [\kappa_1, \kappa_2]$. It suffices to show

$$\sup_{|\Delta|\leq M, k\in[\kappa_1,\kappa_2]} \left| \mathbb{E}G_n(\Delta, k) - \frac{\Delta^2}{2} \right| = o(1) \tag{E.8}$$

and

$$\sup_{|\Delta|\leq M, k\in[\kappa_1,\kappa_2]} |G_n(\Delta, k) - \mathbb{E}G_n(\Delta, k)| = o_p(1). \tag{E.9}$$

For (E.8), I have

$$\mathbb{E}G_n(\Delta,k)=\frac{n}{\sqrt{nk\tau_n}}\int_0^\Delta\left(F_1\left(q_1(k\tau_n)+\frac{s}{\lambda_{1,n}(k)}\right)-F_1(q_1(k\tau_n))\right)ds=\frac{\Delta^2}{2}\frac{f_1\left(q_1(k\tau_n)+\frac{\tilde{s}(k,\Delta)}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))}.$$

By (E.5), for any $l>1$, there exists $N_0>1$ independent of k such that for $n>N_0$, if f_1 is monotone increasing at its lower tail,

$$\frac{f_1\left(q_1(k\tau_n)+\frac{\tilde{s}(k,\Delta)}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))}\in\left(\frac{f_1\left(q_1(\frac{k}{l}\tau_n)\right)}{f_1(q_1(k\tau_n))},\frac{f_1(q_1(lk\tau_n))}{f_1(q_1(k\tau_n))}\right),$$

and if f_1 is monotone decreasing in its lower tail,

$$\frac{f_1\left(q_1(k\tau_n)+\frac{\tilde{s}(k,\Delta)}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))}\in\left(\frac{f_1(q_1(lk\tau_n))}{f_1(q_1(k\tau_n))},\frac{f_1\left(q_1(\frac{k}{l}\tau_n)\right)}{f_1(q_1(k\tau_n))}\right).$$

By first Letting $n\to\infty$ and then $l\to 1$, both the upper and lower bound converge to 1 uniformly over $k\in[\kappa_1,\kappa_2]$. This implies $\frac{f_1\left(q_1(k\tau_n)+\frac{\tilde{s}(k,\Delta)}{\lambda_{1,n}(k)}\right)}{f_1(q_1(k\tau_n))}\to 1$ uniformly in k. Therefore, $\mathbb{E}G_n(\Delta,k)\to\frac{\Delta^2}{2}$ uniformly in Δ and k.

For (E.9), I have

$$G_n(\Delta,k)-\mathbb{E}G_n(\Delta,k)=\sqrt{n}(\mathcal{P}_n-\mathcal{P})f \text{ for } f\in\mathcal{F}_{2,n}$$

where

$$\mathcal{F}_{2,n}=\left\{\frac{1}{\sqrt{\tau_n}}\frac{D_i}{P(X_i)}\int_0^\Delta\left(\mathbb{1}\left\{Y_i\le q_1(k\tau_n)+\frac{s}{\lambda_{1,n}(k)}\right\}-\mathbb{1}\left\{Y_i\le q_1(k\tau_n)\right\}\right)ds,\right.$$
$$\left.|\Delta|<M,k\in[\kappa_1,\kappa_2]\right\}$$

with an envelope function $F_{2,n}=\frac{D_i}{P(X_i)}F_{1,n}$. I note that $||F_{2,n}||_{P,2}\le C<\infty$,

$$M_{2,n}=\max_{1\le i\le n}F_{2,n}\le\frac{C}{\sqrt{\tau_n}}.$$

Since $\mathbb{E}G_n^2(\Delta,k)=O(\frac{1}{\sqrt{n\tau_n}})=o(1)$, $\sqrt{n}(\mathcal{P}_n-P)f\rightsquigarrow 0$ on any subset of $\mathcal{F}_{2,n}$ with finite number of elements. In addition, the empirical process indexed by $f\in\mathcal{F}_{2,n}$ is stochastically equicontinuous. To see this, consider $\mathcal{F}_{2,n}^\delta=\{f-g,f,g\in\mathcal{F}_{2,n},||f-g||_{P,2}\le\delta\}$ with an envelope $F_{2,n}^\delta=2F_{2,n}$ and $M_{2,n}^\delta=\frac{C}{\sqrt{\tau_n}}$. By applying Lemma E.1 on $\mathcal{F}_{2,n}^\delta$ with $\sigma:=\delta$, the Markov inequality, and the fact that $\tau_n n \to \infty$, I obtain that for any $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_n P\left(||\sqrt{n}(\mathcal{P}_n - P)||_{\mathcal{F}_{2,n}^\delta} \geq \varepsilon \right)$$
$$\leq \lim_{\delta \downarrow 0} \limsup_n C\varepsilon^{-1} \left(\sqrt{v\delta^2 \log\left(\frac{2a||F_{2,n}^\delta||_{P,2}}{\delta}\right)} + \frac{v}{\sqrt{n\tau_n}} \log\left(\frac{2a||F_{2,n}^\delta||_{P,2}}{\delta}\right) \right) = 0.$$

This implies $\sup_{|\Delta| \leq M, k \in [\kappa_1, \kappa_2]} |G_n(\Delta, k) - \mathbb{E}G_n(\Delta, k)| = \sqrt{n}||\mathcal{P}_n - P||_{\mathcal{F}_{2,n}} = o_p(1)$.

Combining (E.8) and (E.9), I obtain that

$$G_n(\Delta, k) := \sqrt{n}\mathcal{P}_n f \xrightarrow{p} \frac{\Delta^2}{2} \tag{E.10}$$

uniformly in Δ and k. Then, combining (E.7) and (E.10), I obtain (E.2). This concludes step 1.

Step 2.
Next I consider $\hat{W}_n$ in (E.1):

$$\hat{W}_n(k) = J_{n,1}(k) - J_{n,2}(k) + J_{n,3}(k)$$

where

$$J_{n,1}(k) := \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^n \frac{D_i}{P(X_i)} T_{i,1,n}(k),$$
$$J_{n,2}(k) := \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^n \frac{D_i(\hat{P}(X_i) - P(X_i))}{P(X_i)^2} T_{i,1,n}(k),$$
$$J_{n,3}(k) := \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^n \frac{D_i(\hat{P}(X_i) - P(X_i))^2}{P(X_i)^2 \hat{P}(X_i)} T_{i,1,n}(k),$$

and $T_{i,1,n}(k) = k\tau_n - \mathbb{1}\{Y_{i,1} \leq q_1(k\tau_n)\}$. Note that $T_{i,1,n}(k)$ has an envelope

$$\sup_k |T_{i,1,n}(k)| \leq \overline{T}_{i,1,n} := \kappa_2 \tau_n + \mathbb{1}\{Y_{i,1} \leq q_1(\kappa_2 \tau_n)\}.$$

In the following, I will bound $(J_{n,1}(k), J_{n,2}(k), J_{n,3}(k))$ uniformly over $k \in [\kappa_1, \kappa_2]$.

For $J_{n,3}(k)$, I have

$$\sup_k |J_{n,3}(k)| \lesssim \frac{1}{\sqrt{n\kappa_1 \tau_n}} \sum_{i=1}^n |\overline{T}_{i,1,n}| o_p(\frac{1}{\sqrt{n}}) = o_p(1). \tag{E.11}$$

This is based on two observations: (1) $\mathbb{E}\sup_k \sum_{i=1}^n |T_{i,1,n}| \leq n\mathbb{E}\overline{T}_{i,1,n} = Cn\tau_n$, so $\sum_{i=1}^n |T_{i,1,n}| = O_p^*(n\tau_n)$; (2) under Assumption 3, Lemma 1 of Hirano et al. (2003) shows that $\sup_x |\hat{P}(x) - P(x)| = o_p(n^{-1/4})$.

For $J_{n,2}(k)$, I have $J_{n,2}(k) = J_{n,4}(k) + J_{n,5}(k)$ where

$$J_{n,4}(k) := \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{1}{P(x)} (\hat{P}(x) - P(x))(\mathbb{E}(T_{i,1,n}(k)|x))dF_X(x)$$

and

$$J_{n,5}(k) := \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^{n} \Big[\frac{D_i}{P(X_i)^2} \Big(\hat{P}(X_i) - P(X_i) \Big) T_{i,1,n}(k) - \int_{\text{Supp}(X)} \frac{1}{P(x)} \Big(\hat{P}(x) - P(x) \Big) \Big(\mathbb{E}(T_{i,1,n}(k)|x) \Big) dF_X(x) \Big].$$

Next, I show $J_{n,5}(k) = o_p^*(1)$. Denote $P_h(x) = L(H_h(x)'\pi_h)$ where

$$\pi_h = \arg\min_{\pi \in \mathbb{R}^h} \mathbb{E}(P(X)\log(L(H_h(X)\pi)) + (1 - P(X))\log(1 - L(H_h(X)'\pi))),$$

$H_h(X)$ is the series bases used for approximation such as polynomials or B-splines, and h is the number of terms of the series. I have $J_{n,5}(k) = J_{n,6}(k) + J_{n,7}(k)$ where

$$J_{n,6}(k) := \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^{n} \Big(\frac{D_i T_{i,1,n}(k)}{P(X_i)^2} (\hat{P}(X_i) - P_h(X_i)) - \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} (\hat{P}(x) - P_h(x))dF_X(x) \Big)$$

and

$$J_{n,7}(k) := \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^{n} \Big(\frac{D_i T_{i,1,n}(k)}{P(X_i)^2} (P_h(X_i) - P(X_i)) - \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} (P_h(x) - P(x))dF_X(x) \Big).$$

By Lemma 1 of Hirano et al. (2003), $\sup_x |P_h(x) - P(x)| \lesssim \zeta(h)h^{-\frac{s}{2r}}$ where $\zeta(h) = \sup_x ||H_h(x)||$ and $||A|| = \sqrt{tr(A^T A)}$. For polynomial bases, $\zeta(h) \le Ch$. All the rates restriction in Assumption 3 are stated under this circumstance.

Next, I first compute the order of magnitude of $J_{n,7}(k)$.

$$J_{n,7}(k) = \sqrt{n}(\mathcal{P}_n - \mathcal{P})f, f \in \mathcal{F}_{3,n}$$

where

$$\mathcal{F}_{3,n} = \Big\{ \frac{1}{\sqrt{\tau_n}} \Big(\frac{D_i T_{i,1,n}(k)}{P(X_i)^2} (P_h(X_i) - P(X_i)) - \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} (P_h(x) - P(x))\mathbb{E}(T_{i,1,n}(k)|x)dF_X(x) \Big), k \in [\kappa_1, \kappa_2] \Big\}$$

with an envelope function $F_{3,n} = \frac{C}{\sqrt{\tau_n}}(\overline{T}_{i,1,n} + \mathbb{E}(\overline{T}_{i,1,n}|X))$. Since

$$\mathbb{E}J^2_{n,7}(k) \lesssim \frac{1}{\tau_n}\mathbb{E}\left(\frac{D_i T_{i,1,n}(k)}{P(X_i)^2}(P_h(X_i) - P(X_i))\right)^2 \lesssim \zeta(h)^2 h^{-\frac{s}{r}}\frac{\mathbb{E}T^2_{i,1,n}(k)}{\tau_n} = o(1),$$

$J_{n,7}(k) \rightsquigarrow 0$ on any subsets of $[\kappa_1, \kappa_2]$ with finite elements. I next show that $\sqrt{n}(\mathcal{P}_n - \mathcal{P})f, f \in \mathcal{F}_{3,n}$ is stochastically equicontinuous.

I note that $||F_{3,n}||_{P,2} \le C < \infty$ and $M_{3,n} = \max_{1\le i\le n} F_{3,n} \le \frac{C}{\sqrt{\tau_n}}$. Therefore,

$$\mathcal{F}^{\delta}_{3,n} = \{f - g, f, g \in \mathcal{F}_{3,n}, ||f - g||_{P,2} \le \delta\}$$

with an envelope $2F_{3,n}$ and $M^{\delta}_{3,n} = \frac{C}{\sqrt{\tau_n}}$. In addition, $\{T_{i,1,n}(k) : k \in [\kappa_1, \kappa_2]\}$ satisfies the uniform entropy condition because it is a VC-class, and the class of functions $\{\mathbb{E}(T_{i,1,n}(k)|X) : k \in [\kappa_1, \kappa_2]\}$ is generated by taking the conditional expectation which implies that it also satisfies the uniform entropy condition. Therefore, $\mathcal{F}^{\delta}_{3,n}$ satisfies the uniform entropy condition, that is,

$$\sup_Q N(\varepsilon||F^{\delta}_{3,n}||_{Q,2}, \mathcal{F}^{\delta}_{3,n}, ||\cdot||_{Q,2}) \le \left(\frac{a}{\varepsilon}\right)^v, \quad \forall\varepsilon \in (0, 1].$$

By applying Lemma E.1 on $\mathcal{F}^{\delta}_{3,n}$ with $\sigma := \delta$ and the Markov inequality, I have

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\limsup_n P\bigg(||\sqrt{n}(\mathcal{P}_n - P)||_{\mathcal{F}^{\delta}_{3,n}} \ge \varepsilon\bigg) \\
\le &\lim_{\delta\downarrow 0}\limsup_n C\varepsilon^{-1}\left(\sqrt{v\delta^2 \log\left(\frac{2a||F^{\delta}_{3,n}||_{P,2}}{\delta}\right)} + \frac{v}{\sqrt{n\tau_n}}\log\left(\frac{2a||F^{\delta}_{3,n}||_{P,2}}{\delta}\right)\right) = 0.
\end{aligned}$$

This verifies that $\sqrt{n}(\mathcal{P}_n - \mathcal{P})f, f \in \mathbb{F}_{3,n}$ is stochastically equicontinuous. Combining this with the finite-dimensional convergence, I obtain that $J_{n,7}(k) = o^*_p(1)$.

For $J_{n,6}(k)$, by the Taylor expansion, I have $J_{n,6}(k) = (W_{h,1}(k) + W_{2,h}(k) - W_{3,h}(k))(\hat{\pi}_h - \pi_h)$, in which

$$\begin{aligned}
W_{1,h}(k) := \frac{1}{\sqrt{nk\tau_n}}\sum_{i=1}^n \bigg[&\frac{D_i T_{i,1,n}(k)}{P(X_i)^2} L'(H^T_h(X_i)\pi_h)H^T_h(X_i) \\
&- \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} L'(H^T_h(x)\pi_h)H^T_h(x)dF_X(x)\bigg],
\end{aligned}$$

$$W_{2,h}(k) := \frac{1}{\sqrt{nk\tau_n}}\sum_{i=1}^n \frac{D_i T_{i,1,n}(k)}{P(X_i)^2} L''(H^T_h(X_i)\overline{\pi}_h)H_h(X_i)H^T_h(X_i)(\tilde{\pi}_h - \pi_h),$$

and

$$W_{3,h}(k) := \sqrt{\frac{n}{k\tau_n}}\int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} L''(H^T_h(x)\overline{\pi}_h)H_h(x)H^T_h(x)(\tilde{\pi}_h - \pi_h).$$

For an arbitrary deterministic sequence $l_n \to \infty$ and $f \in \mathcal{F}_{4,n}$,

$$\frac{W_{1,h}(k)}{\zeta(h)l_n} = \sqrt{n}(\mathcal{P}_n - \mathcal{P})f$$

where

$$\mathcal{F}_{4,n} = \Bigg\{ \frac{1}{\sqrt{k\tau_n}\zeta(h)l_n}\Bigg[\frac{D_iT_{i,1,n}(k)}{P(X_i)^2}L'(H_h^T(X_i)\pi_h)H_h^T(X_i) \\ - \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)}L'(H_h^T(x)\pi_h)H_h^T(x)dF_X(x)\Bigg], k \in [\kappa_1, \kappa_2]\Bigg\}$$

with an envelope function

$$F_{4,n} = \frac{C}{\sqrt{\tau_n}\zeta(h)l_n}(H_h^T(X_i)\overline{T}_{i,1,n} + \int H_h^T(x)\mathbb{E}(\overline{T}_{i,1,n}|X=x)dF_X(x)).$$

Since

$$\mathbb{E}||W_{1,h}(k)||^2 \lesssim \left(\frac{\mathbb{E}T_{i,1,n}^2(k)}{\tau_n}\right)\zeta^2(h) = O(\zeta^2(h)),$$

$\{\sqrt{n}(\mathcal{P}_n - \mathcal{P})f : f \in \mathcal{F}_{4,n}\} \rightsquigarrow 0$ in finite dimension. In addition, $||F_{4,n}||_{P,2} \le C < \infty$ and $M_{4,n} = \max_{1\le i\le n} F_{4,n}(X_i) \le \frac{C}{\sqrt{\tau_n}l_n}$. Therefore, for

$$\mathcal{F}_{4,n}^\delta = \{f-g, f, g \in \mathcal{F}_{4,n}, ||f-g||_{P,2} \le \delta\}$$

with an envelope $2F_{4,n}$, I have $||F_{4,n}^\delta||_{P,2} \le C$, $M_{4,n}^\delta = \frac{C}{\sqrt{\tau_n}l_n}$, and

$$\sup_Q N(\varepsilon||F_{4,n}^\delta||_{Q,2}, \mathcal{F}_{4,n}^\delta, ||||_{Q,2}) \le \left(\frac{a}{\varepsilon}\right)^v, \quad \forall\varepsilon \in (0,1].$$

By applying Lemma E.1 on $\mathcal{F}_{4,n}^\delta$ with $\sigma := \delta$ and the Markov inequality, I have

$$\lim_{\delta\downarrow 0}\limsup_n P\left(||\sqrt{n}(\mathcal{P}_n - P)||_{\mathcal{F}_{4,n}^\delta} \ge \varepsilon\right) \\ \le \lim_{\delta\downarrow 0}\limsup_n C\varepsilon^{-1}\left(\sqrt{v\delta^2\log\left(\frac{2a||F_{4,n}^\delta||_{P,2}}{\delta}\right)} + \frac{v}{\sqrt{n\tau_n}l_n}\log\left(\frac{2a||F_{4,n}^\delta||_{P,2}}{\delta}\right)\right) = 0.$$

Therefore, $W_{1,h}(k) = o_p^*(\zeta(h)l_n)$ for any sequence of l_n such that $l_n \to \infty$.

For $W_{2,h}(k)$,

$$\mathbb{E}\sup_k ||W_{2,h}(k)|| \lesssim \mathbb{E}\left|\frac{D_i\overline{T}_{i,1,n}}{P(X_i)^2}L''\right| ||H_h(X_i)||^2||\tilde{\pi}_h - \pi_h||\frac{n}{\sqrt{n\tau_n}} = O(\zeta(h)^2\sqrt{h})$$

So $W_{2,h}(k) = O_p^*(\zeta(h)^2\sqrt{h})$. Similarly,

$$\mathbb{E}\sup_k ||W_{3,h}(k)|| \lesssim \sqrt{\frac{n}{\tau_n}} \int_{\text{Supp}(X)} \left| \frac{\mathbb{E}(\overline{T}_{i,1,n}|x)}{P(x)} L'' \right| ||H_h(x)||^2 dF_X(x) ||\tilde{\pi}_h - \pi_h|| = O(\zeta(h)^2\sqrt{h})$$

So $W_{3,h}(k) = O_p^*(\zeta(h)^2\sqrt{h})$. Combining all the results, $J_{n,7}(k) = O_p^*(\zeta(h)^2\sqrt{h}\sqrt{\frac{h}{n}}) = o_p^*(1)$ and thus $J_{n,5}(k) = o_p^*(1)$.

I next decompose $J_{n,4}$: $J_{n,4}(k) = J_{n,8}(k) + J_{n,9}(k)$ where

$$J_{n,8}(k) := \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}(k)|x)}{P(x)} (\hat{P}(x) - P_h(x)) dF_X(x),$$
$$J_{n,9}(k) := \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}(k)|x)}{P(x)} (P_h(x) - P(x)) dF_X(x).$$

For $J_{n,9}(k)$ I have,

$$J_{n,9}(k) \leq \sqrt{\frac{n}{\kappa_1 \tau_n}} \int_{\text{Supp}(X)} \left| \frac{\mathbb{E}(\overline{T}_{i,n,i}|x)}{P(x)} \right| dF_X(x)\zeta(h)h^{-\frac{s}{2r}} = O_p^*(\sqrt{n\tau_n}\zeta(h)h^{-\frac{s}{2r}}) = o_p^*(1).$$

For $J_{n,8}$, by the Taylor expansion,

$$J_{n,8}(k) = \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}(k)|x)}{P(x)} L'(H_h(x)^T\tilde{\pi}_h) H_h(x)^T dF_X(x)(\hat{\pi}_h - \pi_h).$$

Since $\hat{\pi}_h$ solves the first order condition, $\hat{\pi}_h - \pi_h = \frac{1}{n}\sum_{i=1}^n (\tilde{\Sigma}_h)^{-1}(D_i - P_h(X_i))H_h(X_i)$, in which

$$\tilde{\Sigma}_h = \frac{1}{n}\sum_{i=1}^n L'(H_h(X_i)^T\tilde{\pi}_h)H_h(X_i)H_h(X_i)^T.$$

Hence, I have

$$\begin{aligned}
J_{n,8}(k) &= \frac{1}{n}\sum_{i=1}^n \sqrt{\frac{n}{\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}|x)}{P(x)} L'(H_h(x)^T\tilde{\pi}_h)H_h(x)^T dF_X(x)(\tilde{\Sigma}_h)^{-1}(D_i - P_h(X_i))H_h(X_i) \\
&= \tilde{\Psi}_h^T(k)(\tilde{\Sigma}_h)^{-1}V_h \\
&= \Psi_h^T(k)\Sigma_h^{-1}V_h + (\tilde{\Psi}_h^T(k) - \Psi_h^T(k))\tilde{\Sigma}_h^{-1}V_h + \Psi_h^T(k)(\tilde{\Sigma}_h^{-1} - \Sigma_h^{-1})V_h \\
&:= \Psi_h^T(k)\Sigma_h^{-1}V_h + J_{n,10}(k) + J_{n,11}(k)
\end{aligned}$$

where

$$\tilde{\Psi}_h(k) := \frac{1}{\sqrt{\tau_n}}\int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}|x)}{P(x)} L'(H_h(x)^T\tilde{\pi}_h)H_h(x)dF_X(x),$$
$$\Psi_h(k) := \frac{1}{\sqrt{\tau_n}}\int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}|x)}{P(x)} L'(H_h(x)^T\pi_h)H_h(x)dF_X(x),$$
$$\Sigma_h := \mathbb{E}(H_h(x)H_h(x)^T L'(H_h(x)^T\pi_h)),$$
$$V_h := \frac{1}{\sqrt{n}}\sum_{i=1}^n H_h(X_i)(D_i - P_h(X_i)).$$

Since $\lambda_{\min}(\tilde{\Sigma}_h) \geq \varepsilon > 0$[17], $V_h = O_p(\zeta(h))$, and

$$\begin{aligned}
&||(\tilde{\Psi}_h(k) - \Psi_h(k))|| \\
\lesssim&\frac{1}{\sqrt{\kappa_1\tau_n}}\int_{\text{Supp}(X)} |\frac{\mathbb{E}(\overline{T}_{i,1,n}|x)}{P(x)} L''(H_h(x)^T\overline{\pi}_h)|||H_h(x)||^2 dF_X(x)||\tilde{\pi}_h - \pi_h|| \\
=&O_p^*(\sqrt{\tau_n}\zeta(h)^2\sqrt{\frac{h}{n}}),
\end{aligned}$$

I have $J_{n,10}(k) = O_p^*(\sqrt{\tau_n}\zeta(h)^3\sqrt{\frac{h}{n}}) = o_p^*(1)$.

For $J_{n,11}(k)$, I first denote

$$\hat{\Sigma}_h = \frac{1}{n}\sum_{i=1}^n L'(H_h(X_i)^T\pi_h)H_h(X_i)H_h(X_i)^T.$$

By noticing that $\mathbb{E}||V_h||^2 = O(\zeta(h)^2)$, I have

$$\begin{aligned}
&||(\tilde{\Sigma}_h - \Sigma_h)\Sigma_h^{-1}V_h|| \\
\lesssim&||(\tilde{\Sigma}_h - \hat{\Sigma}_k)\Sigma_h^{-1}V_h|| + ||(\hat{\Sigma}_k - \Sigma_h)\Sigma_h^{-1}V_h|| \\
\lesssim&\frac{1}{n}\sum_{i=1}^n ||H_h(X_i)^T(\tilde{\pi}_h - \pi_h)L''(H_h(X_i)^T\overline{\pi}_h)H_h(X_i)H_h(X_i)^T\Sigma_h^{-1}V_h|| \\
&+\frac{1}{n}||\sum_i^n[L'(H_h(X_i)^T\pi_h)H_h(X_i)H_h(X_i)^T - \mathbb{E}L'(H_h(X_i)^T\pi_h)H_h(X_i)H_h(X_i)^T]\Sigma_h^{-1}V_h|| \\
\lesssim&O_p(\zeta(h)^4\sqrt{\frac{h}{n}}) + O_p((\frac{1}{\sqrt{n}}\mathbb{E}||L'(H_h(X_i)^T\pi_h)H_h(X_i)H_h(X_i)^T||^2)^{1/2}||\zeta(h)||) \\
=&O_p(\zeta(h)^4\sqrt{\frac{h}{n}} + \frac{\zeta(h)^3}{\sqrt{n}}).
\end{aligned}$$

Furthermore, $||\Psi_h(k)|| \lesssim O_p(\frac{\zeta(h)}{\sqrt{\tau_n}}\mathbb{E}(\mathbb{E}(\overline{T}_{i,1,n}|x))) = O_p^*(\sqrt{\tau_n}\zeta(h))$. This implies

$$J_{n,11}(k) = O_p^*(\sqrt{\tau_n}(\zeta(h)^5\sqrt{\frac{h}{n}} + \frac{\zeta(h)^4}{\sqrt{n}}))$$

[17] $\lambda_{\min}(A)$ is the minimal eigenvalue of a positive definite matrix A.

and

$$J_{n,8}(k) = \Psi_h^T(k)\Sigma_h^{-1}V_h + O_p^*(\sqrt{\tau_n}(\zeta(h)^5\sqrt{\frac{h}{n}})) = \Psi_h(k)^T\Sigma_h^{-1}V_h + o_p^*(1).$$

Next, I compute the leading term of $J_{n,8}(k)$: $\Psi_h^T(k)\Sigma_h^{-1}V_h$. Define

$$\delta_0(x,k) := \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{\sqrt{k\tau_n}P(x)}\sqrt{P(x)(1-P(x))},$$
$$\delta_h(x,k) := \Psi_h^T(k)\Sigma_h^{-1}\sqrt{P_h(x)(1-P_h(x))}H_h(x).$$

Then

$$\Psi_h^T(k)\Sigma_h^{-1}V_h = \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_h(X_i,k)\frac{D_i - P_h(X_i)}{\sqrt{P_h(X_i)(1-P_h(X_i))}}.$$

I want to compute the difference

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\delta_h(X_i,k)\frac{D_i - P_h(X_i)}{\sqrt{P_h(X_i)(1-P_h(X_i))}} - \delta_0(X_i)\frac{D_i - P(X_i)}{\sqrt{P(X_i)(1-P(X_i))}}\right] := J_{n,12}(k) + J_{n,13}(k)$$

where

$$J_{n,12}(k) := \frac{1}{\sqrt{n}}\sum_{i=1}^n[(\delta_h(X_i,k) - \delta_0(X))\frac{D - P(X_i)}{P(X_i)(1-P(X_i))}],$$
$$J_{n,13}(k) := \frac{1}{\sqrt{n}}\sum_{i=1}^n[\delta_h(X_i,k)(\frac{D_i - P_h(X_i)}{\sqrt{P_h(X_i)(1-P_h(X_i))}} - \frac{(D_i - P(X_i))}{\sqrt{P(X_i)(1-P(X_i))}})].$$

For $J_{n,12}(k)$, notice that $\sqrt{\tau_n}\delta_h(x,k)$ is the projection of $\sqrt{\tau_n}\delta_0(x,k)$ on $\sqrt{L'(H_h(x)^T\pi_h)}H_h(x)$. By Assumption 3, $\mathbb{E}(T_{i,1,n}(k)|x)$ and $P(x)$ are t times differentiable with their derivatives being bounded by M_n on $\text{Supp}(X)$ uniformly over the quantile index (and thus k). Hence,

$$\sup_{(x,k)\in\text{Supp}(X)\times[\kappa_1,\kappa_2]}||\delta_0(x,k) - \delta_h(x,k)|| \lesssim M_n h^{-\frac{t}{2r}}/\sqrt{\tau_n}$$

and

$$J_{n,12}(k) = O_p^*(\sqrt{\frac{nM_n}{\tau_n}}h^{-\frac{t}{2r}}) = o_p^*(1).$$

For $J_{n,13}(k)$, I have

$$||J_{n,13}(k)|| \le \sqrt{n}\sup_{k,x}||\delta_h(x,k)||\zeta(h)h^{-\frac{s}{2r}} = O_p^*(\sqrt{n\tau_n}\zeta^3(h)h^{-\frac{s}{2r}}) = o_p^*(1).$$

Combining the bounds on $(J_{n,10}(k), \cdots, J_{n,13}(k))$, I obtain that

$$\begin{aligned} J_{n,8}(k) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_h(X_i, k) \frac{D_i - P_h(X_i)}{\sqrt{P_h(X_i)(1 - P_h(X_i))}} + o_p^*(1) \\ &= \frac{1}{\sqrt{nk\tau_n}} \sum_{i=1}^{n} \frac{\mathbb{E}(T_{i,1,n}(k)|X_i)}{P(X_i)} (D_i - P(X_i)) + o_p^*(1). \end{aligned}$$

Then by combining $J_{n,1}(k)$-$J_{n,8}(k)$, I have,

$$\hat{W}_n(k) = W_n(k) + o_p^*(1).$$

This concludes Step 2.

Step 3.
Note that

$$W_n(k) = \sqrt{n}(\mathcal{P}_n - \mathcal{P})f$$

for $f \in \mathcal{F}_{5,n}$, in which $\mathcal{F}_{5,n} = \{\phi_{i,1,n}(k), k \in [\kappa_1, \kappa_2]\}$ and

$$\phi_{i,1,n}(k) = \frac{1}{\sqrt{k\tau_n}} \left[\frac{D_i}{P(X_i)} T_{i,1,n}(k) - \frac{\mathbb{E}(T_{i,1,n}(k)|X_i)}{P(X_i)} (D_i - P(X_i)) \right].$$

Then,

$$F_{5,n} = \frac{C}{\sqrt{\tau_n}} (\overline{T}_{i,1,n} + \mathbb{E}(\overline{T}_{i,1,n}|X_i))$$

is an envelope for $\mathcal{F}_{5,n}$. We have $||F_{5,n}||_{P,2} \leq C < \infty$. $M_{5,n} := \max_{1 \leq i \leq n} F_{5,n}(Y_i, X_i) \leq \frac{C}{\sqrt{\tau_n}}$.

First notice that, for $f \in \mathcal{F}_{5,n}$, $\mathcal{P}f = 0$, $Pf^2 \lesssim \frac{1}{\tau_n} \mathbb{E}T_{i,1,n}^2(k) = O(1)$. So the empirical process $\sqrt{n}(\mathcal{P}_n - \mathcal{P})f$ indexed by $f \in \mathcal{F}_{5,n}$ is bounded in probability in any subsets of $\mathcal{F}_{5.n}$ with finite number of elements.

Next, I want to show the empirical process is stochastically equicontinuous. Let

$$\mathcal{F}_{5,n}^{\delta} = \{f - g, f, g \in \mathcal{F}_{5,n}, ||f - g||_{P,2} \leq \delta\}$$

with envelope $2F_{5,n}$. Then similar to $\mathcal{F}_{3,n}^{\delta}$, there exists $v > 0$ and $a > e$ such that

$$\sup_Q N(\varepsilon ||F_{5,n}^{\delta}||_{Q,2}, \mathcal{F}_{5,n}^{\delta}, ||||_{Q,2}) \leq \left(\frac{a}{\varepsilon}\right)^v, \quad \forall \varepsilon \in (0, 1].$$

By applying Lemma E.1 on $\mathcal{F}_{5,n}^{\delta}$ with $\sigma := \delta$ and the Markov inequality, I have

$$\lim_{\delta \downarrow 0} \limsup_{n} P\left(||\sqrt{n}(\mathcal{P}_n - P)||_{\mathcal{F}_{5,n}^{\delta}} \geq \varepsilon\right)$$
$$\leq \lim_{\delta \downarrow 0} \limsup_{n} C\varepsilon^{-1}\left(\sqrt{v\delta^2 \log\left(\frac{2a||F_{5,n}^{\delta}||_{P,2}}{\delta}\right)} + \frac{v}{\sqrt{n\tau_n}} \log\left(\frac{2a||F_{5,n}^{\delta}||_{P,2}}{\delta}\right)\right) = 0.$$

Therefore, the empirical process $\sqrt{n}(\mathcal{P}_n - P)$ indexed by $f \in \mathcal{F}_{5,n}$ is stochastically equicontinuous and the stochastic process $\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{i,1,n}(k) : k \in [\kappa_1, \kappa_2]\right\}$ is tight. It further implies that the stochastic process $\{\hat{W}_n(k) : k \in [\kappa_1, \kappa_2]\}$ is tight. This concludes Step 3 as well as the proof of the first part of Theorem 3.1.

I next turn to the proof of the second part of Theorem 3.1. By the additional assumption in the theorem, the covariance kernel satisfies that

$$\mathbb{E}(\phi_{i,1,n}(k_1), \phi_{i,0,n}(k_2))(\phi_{i,1,n}(k_1), \phi_{i,0,n}(k_2))' \rightarrow \mathcal{H}(k_1, k_2).$$

This is sufficient for the finite-dimensional convergence of

$$(\lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)), \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n))).$$

Combining the finite-dimensional convergence with the stochastic equicontinuity of

$$\left\{(\lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)), \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n))), k \in [\kappa_1, \kappa_2]\right\},$$

I can conclude the proof for the second part of Theorem 3.1.

E.2 Proof of Theorem 3.3

Hereafter, all bootstrap counterparts are starred. Let $\{I_{n,j}\}_{j\geq 1}$ denote an i.i.d. sequence distributed as multinomial with parameter 1 and probability $(\frac{1}{n}, \cdots, \frac{1}{n})$, so that the bootstrap weight for individual i, $w_{n,i}$, satisfies $w_{n,i} = \sum_{j=1}^{n} \mathbb{1}\{I_{n,j} = i\}$. Also, let $\hat{\Delta}_{1,n}^* = \lambda_{1,n}(\hat{q}_1^*(\tau_n) - q(\tau_n))$ where $\lambda_{1,n}$ is defined in (3.3). Similar to the proof of Theorem 3.1,

$$\hat{\Delta}_{1,n}^* = \arg\min_{\Delta \in \mathbb{R}} -\hat{W}_n^* \Delta + \hat{G}_n^*(\Delta) \tag{E.12}$$

where

$$\hat{W}_n^* = \frac{1}{\sqrt{n\tau_n}} \sum_{i=1}^{n} \frac{w_{n,i} D_i}{\hat{P}(X_i)}(\tau_n - \mathbb{1}\{Y_i \leq q_1(\tau_n)\}),$$
$$\hat{G}_n^*(\Delta) = \frac{1}{\sqrt{n\tau_n}} \sum_{i=1}^{n} \frac{w_{n,i} D_i}{\hat{P}(X_i)} \int_0^{\Delta}\left(\mathbb{1}\left\{Y_i \leq q_1(\tau_n) + \frac{s}{\lambda_{1,n}}\right\} - \mathbb{1}\left\{Y_i \leq q_1(\tau_n)\right\}\right) ds.$$

Since $Ew_{n,i} = 1$, same as in the proof of Theorem 3.1,

$$\hat{G}_n^*(\Delta) = \frac{\Delta^2}{2} + o_p(1). \tag{E.13}$$

Next, let $w_{N_n,i} = \sum_{j=1}^{N_n} \mathbb{1}\{I_{n,j} = i\}$, so that $\{w_{N_n,i}\}_{i=1}^n$ are i.i.d. Poisson random variable with unit mean. Let

$$\tilde{W}_n^* = \frac{1}{\sqrt{n\tau_n}} \sum_{i=1}^n \frac{w_{N_n,i} D_i}{\hat{P}(X_i)} (\tau_n - \mathbb{1}\{Y_i \leq q_1(\tau_n)\}).$$

I aim to show that

$$\hat{W}_n^* - \tilde{W}_n^* = o_p(1).$$

Fix $\eta > 0$ and let $\mathcal{I}_j = \{i : |w_{N_n,i} - w_{n,i}| \geq j\}$ and $n_j = \#\mathcal{I}_j$. Then, for n large enough and with a probability greater than $1 - \eta$ (see (Van der Vaart and Wellner, 1996), p.348),

$$\hat{W}_n^* - \tilde{W}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n (w_{N_n,i} - w_{n,i}) M_{n,i}(\tau_n) = \operatorname{sign}(N_n - n) \sum_{j=1}^2 \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}_j} M_{n,i}(\tau_n) \tag{E.14}$$

with $M_{n,i}(\tau_n) = \frac{1}{\sqrt{\tau_n}} \frac{D_i}{\hat{P}(X_i)} (\tau_n - \mathbb{1}\{Y_i \leq q_1(\tau_n)\})$ and the convention that $\sum_{i \in \mathcal{I}_j} M_{n,i}(\tau_n) = 0$ when $n_j = 0$. I now show that $\sum_{i \in \mathcal{I}_j} M_{n,i}(\tau_n)/\sqrt{n} = o_p(1)$. Note that

$$M_{n,i}(\tau_n) = M_{n,i}^*(\tau_n) + R_{n,i}$$

where

$$M_{n,i}^*(\tau_n) = \frac{1}{\sqrt{\tau_n}} \frac{D_i}{\hat{P}(X_i)} (\tau_n - \mathbb{1}\{Y_i \leq q_1(\tau_n)\})$$

and

$$R_{n,i} = \frac{1}{\sqrt{\tau_n}} \frac{D_i(P(X_i) - \hat{P}(X_i))}{\hat{P}(X_i) P(X_i)} (\tau_n - \mathbb{1}\{Y_i \leq q_1(\tau_n)\}).$$

I first show

$$\sum_{i \in \mathcal{I}_j} R_{n,i}/\sqrt{n} = o_p(1). \tag{E.15}$$

Note that

$$\begin{aligned} |\sum_{i \in \mathcal{I}_j} R_{n,i}/\sqrt{n}| &\lesssim \frac{1}{\sqrt{n\tau_n}} \sum_{i \in \mathcal{I}_j} |\tau_n - \mathbb{1}\{Y_{i,1} \leq q_1(\tau_n)\}| \sup_{x \in \operatorname{Supp}(X)} |\hat{P}(x) - P(x)| \\ &\lesssim \frac{1}{\sqrt{n\tau_n}} \sum_{i \in \mathcal{I}_j} |\tau_n - \mathbb{1}\{Y_{i,1} \leq q_1(\tau_n)\}| o_p(1). \end{aligned}$$

In addition,

$$\frac{1}{n\tau_n}E\left[\left(\sum_{i\in\mathcal{I}_j}|\tau_n - \mathbb{1}\{Y_{i,1}\leq q_1(\tau_n)\}|\right)^2 |(I_{n,j})_{j\geq 1}, N_n\right] \lesssim \left(\frac{n_j}{\sqrt{n}}\right)^2 \lesssim \left(\frac{N_n - n}{\sqrt{n}}\right)^2 = O_p(1).$$

Thus (E.15) holds. Next, since $E(M^*_{n,i}(\tau_n)|(I_{n,j})_{j\geq 1}, N_n) = 0$ and

$$\frac{1}{n}\text{Var}\left[\left(\sum_{i\in\mathcal{I}_j} M^*_{n,i}(\tau_n)\right)|(I_{n,j})_{j\geq 1}, N_n\right] \leq \frac{n_j}{n} \leq \frac{|N_n - n|}{n} = o_p(1),$$

I have

$$\sum_{i\in\mathcal{I}_j} M^*_{n,i}(\tau_n)/\sqrt{n} = o_p(1). \tag{E.16}$$

Combining (E.15) and (E.16), I have shown that $\sum_{i\in\mathcal{I}_j} M_{n,i}(\tau_n)/\sqrt{n} = o_p(1)$ and thus

$$\hat{W}^*_n - \tilde{W}^*_n = o_p(1). \tag{E.17}$$

In addition, by the same argument in the proof of Theorem 3.1, I have

$$\tilde{W}^*_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n w_{N_n,i}\phi_{i,1,n}(1) + o_p(1). \tag{E.18}$$

Combining (E.12), (E.17), and (E.18), I obtain that

$$-\hat{W}^*_n\Delta + \hat{G}^*_n(\Delta) = -\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n w_{N_n,i}\phi_{i,1,n}(1)\right)\Delta + \frac{\Delta^2}{2}.$$

By the Convexity lemma in Pollard (1991), I have

$$\hat{\Delta}^*_{1,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n w_{N_n,i}\phi_{i,1,n}(1) + o_p(1).$$

Recall that, from the proof of Theorem 3.1, I have

$$\hat{\Delta}_{1,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_{i,1,n}(1) + o_p(1).$$

Thus

$$\lambda_{n,1}(\hat{q}^*_1(\tau_n) - \hat{q}_1(\tau_n)) = \hat{\Delta}^*_{1,n} - \hat{\Delta}_{1,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (w_{N_n,i} - 1)\phi_{i,1,n}(1) + o_p(1). \tag{E.19}$$

Similarly,

$$\lambda_{n,0}(\hat{q}_0^*(\tau_n)-\hat{q}_0(\tau_n))=\hat{\Delta}_{0,n}^*-\hat{\Delta}_{0,n}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(w_{N_n,i}-1)\phi_{i,0,n}(1)+o_p(1). \tag{E.20}$$

Also note that, with $C_1(\rho,m)$, $C_0(\rho,m)$, $\hat{\lambda}_n$, and Σ_n defined in Theorem 3.2, I have

$$\Sigma_n^{-1/2}\hat{\lambda}_n(\hat{q}(\tau_n)-q(\tau_n))=\Sigma_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(C_1(\rho,m)\phi_{i,1,n}(1)-C_0(\rho,m)\phi_{i,0,n}(1))+o_p(1)\rightsquigarrow\mathcal{N}(0,1). \tag{E.21}$$

Then combining (E.19), (E.20), and (E.21) with the continuous mapping theorem, I obtain that

$$\begin{aligned}&\Sigma_n^{-1/2}\hat{\lambda}_n(\hat{q}^*(\tau_n)-\hat{q}(\tau_n))\\=&\Sigma_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(w_{N_n,i}-1)\left(C_1(\rho,m)\phi_{i,1,n}(1)-C_0(\rho,m)\phi_{i,0,n}(1)\right)+o_p(1)\rightsquigarrow\mathcal{N}(0,1).\end{aligned}$$

Here the variance Σ_n is the same in (E.21) because $w_{N_n,i}$ is independent of data and has unit mean and variance. This concludes the proof.

E.3 Proof of Theorem 3.4

Note that

$$\frac{\hat{q}_j(ml^r\tau_n)-\hat{q}_j(l^r\tau_n)}{\hat{q}_j(ml^{r-1}\tau_n)-\hat{q}_j(l^{r-1}\tau_n)}\sim(1+O_p(\frac{1}{\sqrt{\tau_n n}}))\frac{q_j(ml^r\tau_n)-q_j(l^r\tau_n)}{q_j(ml^{r-1}\tau_n)-q_j(l^{r-1}\tau_n)}\sim(1+O_p(\frac{1}{\sqrt{\tau_n n}}))l^{-\xi_j}.$$

This implies (1) by the continuous mapping theorem. (2) follows from the delta-method and a triangular array CLT in such as Theorem 3.4.5 in Durrett (2010).

E.4 Proof of Theorem 4.1

Note that

$$\begin{aligned}\hat{Z}_{1,n}(k)=&\arg\min_z\frac{1}{\alpha_{1,n}}\Big[-\sum_{i=1}^{n}\frac{D_i}{\hat{P}(X_i)}(\tau_n z-(z-\alpha_{1,n}(U_{i,1}-\beta_{1,n}))\mathbb{1}\{\alpha_{1,n}(U_{i,1}-\beta_{1,n})\le z\})\\&+\sum_{i=1}^{n}\frac{D_i}{\hat{P}(X_i)}\tau_n\alpha_{1,n}(U_{i,1}-\beta_{1,n})\Big].\end{aligned}$$

Multiplying the LHS by $\alpha_{1,n}$ and subtracting

$$\sum_{i=1}^{n}\frac{D_i}{\hat{P}(X_i)}(\tau_n\alpha_{1,n}(U_{i,1}-\beta_{1,n})+(-\delta-\alpha_{1,n}(U_{i,1}-\beta_{1,n}))\mathbb{1}\{\alpha_{1,n}(U_{i,1}-\beta_{1,n})\le-\delta\}),$$

I obtain

$$\hat{Z}_{1,n}(k) = \arg\min_{z_1} -\sum_{i=1}^{n} W_1(D_i, \hat{P}(X_i))\tau_n z_1 + \sum_{i=1}^{n} W_1(D_i, \hat{P}(X_i)) l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1).$$

Similarly,

$$\hat{Z}_{0,n}(k) = \arg\min_{z_0} -\sum_{i=1}^{n} W_0(D_i, \hat{P}(X_i))\tau_n z_0 + \sum_{i=1}^{n} W_0(D_i, \hat{P}(X_i)) l_\delta(\alpha_{0,n}(U_{i,0} - \beta_{0,n}), z_0).$$

So overall,

$$(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k)) := \arg\min_{z_1, z_0} \sum_{j=0,1} \mathcal{Q}_{j,n}(z_j),$$

where

$$\mathcal{Q}_{j,n}(z_j, k) = -\sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i))\tau_n z_j + \sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i)) l_\delta(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), z_j).$$

In the following, I divide the proof into five steps. In the first step, I show the marginal convergence, that is, for $j = 0, 1$ and fixed z_j,

$$\mathcal{Q}_{j,n}(z_j, k) \rightsquigarrow \mathcal{Q}_{j,\infty}(z_j, k),$$

in which

$$\mathcal{Q}_{j,\infty}(z_j, k) = -k z_j + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j})) l_\delta(\mathcal{J}_{i,j}, z_j).$$

In the second step, I show that for any (z_1, z_0), $\mathcal{Q}_{1,n}(z_1, k)$ and $\mathcal{Q}_{0,n}(z_0, k)$ are asymptotically independent. Hence, the marginal convergence is sufficient for the joint convergence of $(\mathcal{Q}_{1,n}(z_1, k), \mathcal{Q}_{0,n}(z_0, k))$ to $(\mathcal{Q}_{1,\infty}(z_1, k), \mathcal{Q}_{0,\infty}(z_0, k))$. Then by the continuous mapping theorem,

$$\mathcal{Q}_{1,n}(z_1, k) + \mathcal{Q}_{0,n}(z_0, k) \rightsquigarrow \mathcal{Q}_{1,\infty}(z_1, k) + \mathcal{Q}_{0,\infty}(z_0, k).$$

In the third step, I apply the convexity lemma to show the weak convergence of the sample minimizers $(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k))$ to their population counterparts $(Z_{1,\infty}(k), Z_{0,\infty}(k))$ when k satisfies Assumption 9.

In the fourth step, I enhance the result to the finite-dimensional convergence, that is, for $(k_l)_{l=1}^L$ satisfying Assumption 9,

$$\begin{aligned}&(\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l))_{l=1}^L \rightsquigarrow (Z_{1,\infty}(k_l), Z_{0,\infty}(k_l))_{l=1}^L \\ := &\arg\min_{(z_{1,l}, z_{0,l})_{l=1}^L} \sum_{j=0,1} \sum_{l=1}^{L} \left\{ -k_l z_{j,l} + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j})) l_\delta(\mathcal{J}_{i,j}, z_{j,l}) \right\}.\end{aligned}$$

In the last step, I show $(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k))$ as a two-dimensional stochastic process indexed by k in $\mathcal{D}^2([\kappa_1, \kappa_2])$ weakly converges to a two-dimensional stochastic process $(Z_{1,\infty}(k), Z_{0,\infty}(k))$.

Before showing the five steps, I first present four technical statements. Their proofs can be found at the end of this section.

Lemma E.2. *Under the assumptions in Theorem 4.1, for* $j = 0, 1$,
(1) $\frac{1}{n}\sum_{i=1}^n W_j(D_i, P(X_i)) \to 1$ *a.s.*
(2) Let

$$\begin{aligned}
&\textit{for type 1 tails } (\xi_1 = 0)\textit{:} && E_j = E^1 = [-\infty, +\infty) \times \{0,1\} \times \mathit{Supp}(\mathcal{X}),\\
&\textit{for type 2 tails } (\xi_1 > 0)\textit{:} && E_j = E^2 = [-\infty, 0) \times \{0,1\} \times \mathit{Supp}(\mathcal{X}),\\
&\textit{for type 3 tails } (\xi_1 < 0)\textit{:} && E_j = E^3 = [0, +\infty) \times \{0,1\} \times \mathit{Supp}(\mathcal{X}).
\end{aligned}$$

Then $\hat{N}_j := \sum_{i=1}^n \mathbb{1}\{\alpha_{j,n}(U_{i,j} - \beta_{j,n}), D_i, X_i\}$ *as a point process on state space* E_j *weakly converges to* $N_j = \sum_{i=1}^\infty \mathbb{1}\{\mathcal{J}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}\}$.
(3) Let

$$g_1(u,x) = \frac{1}{P(x)} l_\delta(u, x, z_1), \quad g_0(u,x) = \frac{1}{1-P(x)} l_\delta(u, x, z_0),$$

and

$$\Psi_{j,n} = \sum_{i=1}^n (jD_i + (1-j)(1-D_i)) g_j(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), X_i).$$

Then for a pair of constants (t_1, t_0), *and* $\tilde{i}$ *representing the imaginary number,*

$$\mathbb{E}\exp\left(\tilde{i}t_1\Psi_{1,n} + \tilde{i}t_0\Psi_{0,n}\right) \to \mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 d g_1 dN_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0(1-d) g_0 dN_0\right),$$

in which N_j *is defined in (2).*
(4) The distances between two closest discontinuities of the sample paths of the two marginal stochastic processes $\hat{Z}_{1,n}(k)$ *and* $\hat{Z}_{0,n}(k))$ *indexed by* k *are both greater than 1.*

Step 1:
I focus on the case for $j = 1$ because the case for $j = 0$ can be proved in a similar manner. First note that for fixed z_1, by Lemma E.2, $-\sum_{i=1}^n \frac{D_i}{\hat{P}(X_i)}\tau_n z_1 = -kz_1 + o_p(1)$. In order to compute the second piece of the objective function, I first define

$$\begin{aligned}
\theta_{n,1}(z_1) &:= \sum_{i=1}^n \frac{D_i}{P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1),\\
\theta_{n,2}(z_1) &:= \sum_{i=1}^n \frac{D_i}{P(X_i)} \Big| l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1)\Big|,\\
\theta_{n,3}(z_1) &:= \sum_{i=1}^n \frac{D_i(\hat{P}(X_i) - P(X_i))}{\hat{P}(X_i)P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1).
\end{aligned}$$

Then $\sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) = \theta_{n,1}(z_1) + \theta_{n,3}(z_1)$ and $|\theta_{n,3}(z_1)| \lesssim \theta_{n,2}(z) \sup_x |\hat{P}(x) - P(x)|$. Also notice that $\theta_{n,1}(z_1)$ and $\theta_{n,2}(z_1)$ can be rewritten as

$$\theta_{n,1}(z_1) = \int_E \frac{d}{P(x)} l_\delta(u, z_1) d\hat{N}_1,$$
$$\theta_{n,2}(z_1) = \int_E \frac{d}{P(x)} |l_\delta(u, z_1)| d\hat{N}_1,$$

in which $\hat{N}_1$ is defined in Lemma E.2. Following part 2 of the proof of Theorem 4.1 in Chernozhukov (2005), for type 1 and 3 tails, $\frac{d}{P(x)} l_\delta(u, z_1) \in C_K(E)$ for any fixed z, and for type 2 tails, $\frac{d}{P(x)} l_\delta(u, z_1) \in C_K(E)$ for $z_1 < 0$. Also, by Lemma E.2(2), $\hat{N}_1 \rightsquigarrow N_1$. Therefore, for any z for type 1 and 3 tails and negative z for type 2 tails,

$$\theta_{n,1}(z) \rightsquigarrow \theta_{\infty,1}(z_1) = \int_E \frac{d}{P(x)} l_\delta(u, z_1) dN_1$$
$$\theta_{n,2}(z) \rightsquigarrow \theta_{\infty,2}(z_1) = \int_E \frac{d}{P(x)} |l_\delta(u, z_1)| dN_1.$$

This implies that, for the aforementioned region of z_1, $\theta_{\infty,2}(z_1) = O_p(1)$, $\theta_{n,3}(z_1) = O_p(\theta_{n,2}(z_1) \sup_x |\hat{P}(x) - P(x)|) = o_p(1)$, and thus

$$\sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) \to \theta_{\infty,1}(z_1).$$

The last thing to check is $\sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) \to +\infty$ for type 2 tails when $z_1 > 0$. Again, following Chernozhukov (2005), if $z_1 > 0$, $\alpha_{1,n} \to 0$, $\beta_{1,n} = 0$, $l_\delta(u, z_1) \geq \mathbb{1}\{-\delta \leq u \leq 0\} z_1$ if $u > -\delta$, and $l_\delta(u, z) = z + \delta$ if $u \leq -\delta$. Because $P(\alpha_n U_{i,1} > -\delta) \to 1$, I have,

$$\sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_n U_{i,1}, z_1) \mathbb{1}\{\alpha_n U_{i,1} \leq -\delta\} \lesssim \sum_{i=1}^{n} \mathbb{1}\{\alpha_n U_{i,1} \leq -\delta\} = O_p(1),$$

and

$$\sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_n U_{i,1}, z_1) \mathbb{1}\{\alpha_n U_{i,1} > -\delta\} \gtrsim \sum_{i=1}^{n} z_1 \mathbb{1}\{\alpha_n U_{i,1} > -\delta\} = +\infty,$$

which lead to the desired result that

$$\sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) \to +\infty.$$

Noting that $\mathcal{Q}_{1,\infty}(z_1, k) = -kz_1 + \int_E \frac{d}{P(x)} l_\delta(u_1, z_1) dN_1$, I have shown that, for all types fo tails,

$$\mathcal{Q}_{1,n}(z_1, k) \rightsquigarrow \mathcal{Q}_{1,\infty}(z_1, k).$$

Similarly, by denoting $\mathcal{Q}_{0,\infty}(z_0,k) = -kz_0 + \int_E \frac{1-d}{1-P(x)} l_\delta(u_0, z_0) dN_0$, I can show that

$$\mathcal{Q}_{0,n}(z_0) \rightsquigarrow \mathcal{Q}_{0,\infty}(z_0).$$

Step 2:

From the proof of step 1, it is sufficient to show the asymptotic independence of

$$\Psi_{1,n} := \sum_{i=1}^{n} W_1(D_i, P(X_i)) l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1)$$

and

$$\Psi_{0,n} := \sum_{i=1}^{n} W_0(D_i, P(X_i)) l_\delta(\alpha_{0,n}(U_{i,0} - \beta_{0,n}), z_0)$$

for any (z_1, z_0). Also I have already shown in step 1 that

$$\Psi_{1,n} \rightsquigarrow \int_{E_1} d g_1(j,d,x) dN_1(j,d,x)$$

and

$$\Psi_{0,n} \rightsquigarrow \int_{E_0} (1-d) g_0(j,d,x) dN_0(j,d,x).$$

Therefore, I only have to show that, for any pair of constants (t_1, t_0),

$$\mathbb{E} \exp\left(\tilde{i} t_1 \Psi_{1,n} + \tilde{i} t_0 \Psi_{0,n}\right) \to \mathbb{E} \exp\left(\tilde{i} \int_{E_1} t_1 d g_1 dN_1\right) \mathbb{E} \exp\left(\tilde{i} \int_{E_0} t_0 (1-d) g_0 dN_0\right).$$

This is done by Lemma E.2(3).

Step 3:

From the results in step 1 and 2, I obtain the joint convergence:

$$(\mathcal{Q}_{1,n}(z_1,k), \mathcal{Q}_{0,n}(z_0,k)) \rightsquigarrow (\mathcal{Q}_{1,\infty}(z_1,k), \mathcal{Q}_{0,\infty}(z_0,k)) \quad \text{and} \quad \mathcal{Q}_{1,\infty}(z_1,k) \perp\!\!\!\perp \mathcal{Q}_{0,\infty}(z_0,k).$$

By the continuous mapping theorem,

$$\mathcal{Q}_{1,n}(z_1,k) + \mathcal{Q}_{0,n}(z_0,k) \rightsquigarrow \mathcal{Q}_{1,\infty}(z_1,k) + \mathcal{Q}_{0,\infty}(z_0,k).$$

This result can be easily improved to hold over finite pairs of (z_1, z_0). For fixed k as the limiting of $\tau_n n$ who satisfies Assumption 9, recall that

$$\mathcal{Q}_{j,n}(z_{j,l},k) = -\sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i)) \tau_n z_{j,l} + \sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i)) l_\delta(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), z_{j,l}),$$

and

$$\mathcal{Q}_{j,\infty}(z_{j,l},k) = -kz_{j,l} + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j}))l_\delta(\mathcal{J}_{i,j}, z_{j,l}).$$

Then

$$\sum_{l=1}^{L}[\mathcal{Q}_{1,n}(z_{1,l},k) + \mathcal{Q}_{0,n}(z_{0,l},k)] \rightsquigarrow \sum_{l=1}^{L}[\mathcal{Q}_{1,\infty}(z_{1,l},k) + \mathcal{Q}_{0,\infty}(z_{0,l},k)].$$

This is the finite-dimensional convergence of the objective function. Also notice that $\mathcal{Q}_{1,\infty}(z_1) + \mathcal{Q}_{0,\infty}(z_0)$ is convex in (z_1, z_0). Therefore, in order to apply the convexity lemma as in Chernozhukov (2005), I only have to further verify two statements: (1) $\mathcal{Q}_{j,\infty}(z_j)$ is finite over a non-empty open set of (z_j) and (2) $Z_{j,\infty}(k)$, $j = 0, 1$ is a unique pair of random variables who minimizes $\sum_{j=0,1} \mathcal{Q}_{j,\infty}(z_j)$. In fact, (1) can be proved similar to the proof of Theorem 4.1 Part 2(II) in Chernozhukov (2005). (2) holds by the fact that k satisfies Assumption 9. One sufficient condition for Assumption 9 is $k \in [\kappa_1, \kappa_2]/(\mathcal{L}_1 \cup \mathcal{L}_2)$[18], in which

$$\mathcal{L}_j = \Bigg\{k \in [\kappa_1,\kappa_2] : P\left(\sum_{i\in\mu}\frac{1}{P(\mathcal{X}_{i,j})} = k\right) > 0 \text{ or } P\left(\sum_{i\in\mu}\frac{1}{P(\mathcal{X}_{i,j})} + \frac{1}{P(\mathcal{X}_{j,h})} = k\right) > 0$$
$$\text{for some } h \text{ and } \mu \in \mathcal{M}(l), l \le h-1\Bigg\}.$$

Then, the convexity lemma implies that

$$(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k)) \rightsquigarrow (Z_{1,\infty}(k), Z_{0,\infty}(k)) := \underset{(z_1,z_0)\in\mathbb{R}^2}{\arg\min} \sum_{j=0,1}\left[-kz_j + \sum_{i=1}^{n} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j}))l_\delta(\mathcal{J}_{i,j}, z_j)\right].$$

Step 4:
Recall that

$$\mathcal{Q}_{j,n}(z_j,k) = -\sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i))\tau_n z_j + \sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i))l_\delta(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), z_j)$$

and

$$\mathcal{Q}_{j,\infty}(z_j,k) = \left(-kz_j + \sum_{i=1}^{\infty} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j}))l_\delta(\mathcal{J}_{i,j}, z_j)\right).$$

Then I have

$$(\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l))_{l=1}^{L} = \underset{(z_{1,l},z_{0,l})_{l=1}^{L}\in\mathbb{R}^{2L}}{\arg\min} \sum_{l=1}^{L}\sum_{j=0,1} \mathcal{Q}_{j,n}(z_{j,l},k_l).$$

[18] Lemma E.5 and E.6 show that when $k \in [\kappa_1, \kappa_2]/(\mathcal{L}_1 \cup \mathcal{L}_2)$, uniqueness and tightness of $Z_{j,\infty}(k)$, $j = 0, 1$ hold. This sufficient condition will be used later in the proof.

When k_l satisfies Assumption 9 for $l = 1, 2, \cdots, L$, by repeating Step 1–3, I can establish that

$$\sum_{l=1}^{L}\sum_{j=0,1}\mathcal{Q}_{j,n}(z_{j,l},k_l) \rightsquigarrow \sum_{l=1}^{L}\sum_{j=0,1}\mathcal{Q}_{j,\infty}(z_{j,l},k_l).$$

By the same Convexity Lemma used in Step 3, I have

$$\begin{aligned}(\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l))_{l=1}^{L} \rightsquigarrow &(Z_{1,\infty}(k_l), Z_{0,\infty}(k_l))_{l=1}^{L} \\ &:= \operatorname*{arg\,min}_{(z_{1,l},z_{0,l})_{l=1}^{L}\in\mathbb{R}^{2(L+1)}} \sum_{j=0,1}\sum_{l=1}^{L}\left[-k_l z_{j,l} + \sum_{i=1}^{n} W_j(\mathcal{D}_{i,j}, P(\mathcal{X}_{i,j}))l_\delta(\mathcal{J}_{i,j}, z_{j,l})\right].\end{aligned}$$

Step 5:
I aim to prove the result by applying Theorem 13.1 of Billingsley (1999) with $T_p = [\kappa_1,\kappa_2]/(\mathcal{L}_1\cup\mathcal{L}_0)$ because as mentioned above, all the discontinuities of the $Z_{j,\infty}(k)$ occurs in $\mathcal{L}_j$. In fact, with $(\kappa_1,\kappa_2) \notin \mathcal{L}_1\cup\mathcal{L}_0$, I only need to show $(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k))$ indexed by $k\in[\kappa_1,\kappa_2]$ is tight. Then based on Theorem 13.3 of Billingsley (1999), it suffices to show that (1) T_p's complement in $[\kappa_1,\kappa_2]$ is at most countable, (2) for $j = 0, 1$ and every ε,

$$\lim_{\delta\to 0}\Big[P\left(|Z_{j,\infty}(\kappa_2) - Z_{j,\infty}(\kappa_2-\delta)| \geq \varepsilon\right) + P\left(|Z_{j,\infty}(\kappa_1) - Z_{j,\infty}(\kappa_1+\delta)| \geq \varepsilon\right)\Big] = 0, \tag{E.22}$$

and (3) for $j = 0, 1$, any positive ε, and any η, there exists constants δ and n_0 such that

$$P\left(|\omega''_{j,n}(\delta)| \geq \varepsilon\right) \leq \eta \tag{E.23}$$

in which

$$\omega''_{j,n}(\delta) := \sup_{k_1\leq k_2\leq k_3, k_3-k_1\leq\delta}\Big\{|\hat{Z}_{j,n}(k_2) - \hat{Z}_{j,n}(k_1)| \wedge |\hat{Z}_{j,n}(k_3) - \hat{Z}_{j,n}(k_2)|\Big\}.$$

(E.22) holds by Assumption 9. For (E.23), I focus on the case for $j = 1$. The case for $j = 0$ can be handled similarly. First, by convention, I define $Z_{1,\infty}(k)$ as the left limit of the sample path, that is, $Z_{1,\infty}(k) = \lim_{k'\downarrow k} Z_{1,\infty}(k')$. Notice that $Z_{1,\infty}(k)$ is piece-wise constant and the jumps only occur when $k - \frac{1}{P(X_h)} = \sum_{i\neq h}\frac{T_i}{P(X_i)}\mathbb{1}\{\mathcal{J}_i < \mathcal{J}_h\}$ or $k = \sum_{i\neq h}\frac{T_i}{P(X_i)}\mathbb{1}\{\mathcal{J}_i < \mathcal{J}_h\}$ for some h such that $T_h = 1$. By Lemma E.2(4), for $k_1 < k_2 < k_3$, such that $k_3 - k_1 < 1$, either $\hat{Z}_{j,n}(k_2) = \hat{Z}_{j,n}(k_1)$ or $\hat{Z}_{j,n}(k_2) = \hat{Z}_{j,n}(k_3)$. This implies that (E.23) holds whenever $\delta < 1$. Last, for $k\in\mathcal{L}_1$, k can be written as $\sum_{i=1}^{I_1} N_i\frac{1}{P(x_i)}$ where $\{x_i\}_{i=1}^{I_1}$ are the point mass of the CDF of $\mathcal{X}_{i,1}$, $\{N_i\}_{i=1}^{I_1}$ are a sequence of nonnegative integers, and I_1 is the total number of point mass, which is finite by Assumption 8. Since $\frac{1}{P(x_i)} > 1$, $\sum_{i=1}^{I_1} N_i \leq \kappa_2$ which implies that the cardinality of $\mathcal{L}_1$ is at most finite. Similarly, the cardinality of $\mathcal{L}_0$ is also finite. This implies that T_p's complement in $[\kappa_1,\kappa_2]$ is finite. Hence by Theorem 13.3 of Billingsley (1999), the marginal processes $\hat{Z}_{1,n}(k)$ and $\hat{Z}_{2,n}(k)$ indexed by k in $\mathcal{D}[\kappa_1,\kappa_2]$ are tight and $(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k))$ converges to $(\hat{Z}_{1,\infty}(k), \hat{Z}_{0,\infty}(k))$ under

Skorohod metric.

E.5 Proof of Theorem 4.2

First note that

$$\frac{\alpha_{1,n}}{\alpha_{0,n}} = \frac{\alpha_{1,n}(q_1(\frac{mk_{l'}}{n}) - q_1(\frac{k_{l'}}{n}))}{\alpha_{0,n}(q_(\frac{mk_{l'}}{n}) - q_0(\frac{k_{l'}}{n}))} \frac{q_0(\frac{mk_{l'}}{n}) - q_0(\frac{k_{l'}}{n})}{q_1(\frac{mk_{l'}}{n}) - q_1(\frac{k_{l'}}{n})} \to \ k_{l'}^{\xi_0-\xi_1} \frac{m^{-\xi_1} - 1}{\rho(m^{-\xi_0} - 1)} := \tilde{\rho}.$$

Hence,

$$\begin{aligned}\frac{\hat{\alpha}_n}{\alpha_{1,n}} &\sim \frac{\sqrt{k_{l'}}}{\max(\hat{Z}_{1,n}(mk_{l'}) - \hat{Z}_{1,n}(k_{l'}), \frac{\alpha_{1,n}}{\alpha_{0,n}}(\hat{Z}_{0,n}(mk_{l'}) - \hat{Z}_{0,n}(k_{l'})))} \\ &\sim \frac{\sqrt{k_{l'}}}{\max(Z_{1,\infty}(mk_{l'}) - Z_{1,\infty}(k_{l'}), \tilde{\rho}(Z_{0,\infty}(mk_{l'}) - Z_{0,\infty}(k_{l'})))}.\end{aligned}$$

Similarly, $\frac{\hat{\alpha}_n}{\alpha_{0,n}} \sim \frac{\sqrt{k_{l'}}\tilde{\rho}}{\max(Z_{1,\infty}(mk_{l'}) - Z_{1,\infty}(k_{l'}), \tilde{\rho}(Z_{0,\infty}(mk_{l'}) - Z_{0,\infty}(k_{l'})))}$. By combining the above results with Theorem 4.1, I obtain that

$$\hat{Z}_n(k) = \hat{\alpha}_n(\hat{q}(\tau_n) - q(\tau_n)) = \frac{\hat{\alpha}_n}{\alpha_{1,n}} \hat{Z}^c_{1,n}(k) - \frac{\hat{\alpha}_n}{\alpha_{0,n}} \hat{Z}^c_{0,n}(k) \rightsquigarrow Z^c_\infty(k).$$

Note that the limiting distribution is non-degenerate even when $\rho = 0$ or ∞.

E.6 Proof of Proposition 4.2

$$\begin{aligned}&\hat{\alpha}_n(\sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \sum_{l=1}^{L} r_l q(\tau_{n,l})) \\ =&\hat{\alpha}_n(\sum_{l=1}^{L}(\hat{r}_l - r_l)q(\tau_{n,l})) + \hat{\alpha}_n(\sum_{l=1}^{L} \hat{r}_l(\hat{q}(\tau_{n,l}) - q(\tau_{n,l}))) \\ =&\hat{\alpha}_n(\sum_{l=1}^{L}(\hat{r}_l - r_l)(q(\tau_{n,l}) - q(0))) + \hat{\alpha}_n(\sum_{l=1}^{L} \hat{r}_l(\hat{q}(\tau_{n,l}) - q(\tau_{n,l}))).\end{aligned}$$

Since $\alpha_{j,n}(q_j(\tau_n) - q_j(0)) \to \eta_j(k)$, $\frac{\hat{\alpha}_n}{\alpha_{j,n}} = O_p(1)$ for $j = 0, 1$, and $\hat{\gamma}_l \to \gamma_l$, the first term is $o_p(1)$. The second term converges to $\sum_{l=1}^{L} \gamma_l Z^c_\infty(k_l)$. This concludes the proof.

E.7 Proof of Theorem 5.1

The proof follows the five steps in the proof of Theorem 4.1 which I will not repeat. The key ingredient, Lemma E.2, is replaced by the following Lemma.

Lemma E.3. *Let $P_{n,i} = \sum_{l=1}^{n} \mathbb{1}\{I_l = i\}$. Under the conditions of Theorem 5.1, for $j = 0, 1$,*
(1) $\frac{1}{n}\sum_{i=1}^{n} P_{n,i} W_j(D_i, P(X_i)) \to 1$ a.s.

(2) For $\hat{N}_j^* := \sum_{i=1}^n P_{n,i}\mathbb{1}\{\alpha_{j,n}(U_{i,j} - \beta_{j,n}), D_i, X_i\}$,

$$\hat{N}_j^* \rightsquigarrow N_j^* := \sum_{i=1}^{\infty} \Gamma_{i,j}\mathbb{1}\{\mathcal{J}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}\}.$$

(3) Let

$$g_1(u,x) = \frac{1}{P(x)} l_\delta(u,x,z_1), \quad g_0(u,x) = \frac{1}{1-P(x)} l_\delta(u,x,z_0),$$

and

$$\Psi_{j,n} = \sum_{i=1}^n (jD_i + (1-j)(1-D_i))P_{n,i}g_j(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), X_i).$$

Then for a pair of constants (t_1, t_0),

$$\mathbb{E}\exp(\tilde{i}t_1\Psi_{1,n} + \tilde{i}t_0\Psi_{0,n}) \to \mathbb{E}\exp(\tilde{i}\int_{E_1} t_1 d g_1 dN_1^*)\mathbb{E}\exp(\tilde{i}\int_{E_0} t_0(1-d)g_0 dN_0^*),$$

in which N_j *is defined in (2).*
(4) The distances between the two closest discontinuities of the marginal sample paths of the two-dimensional stochastic process $(\hat{Z}_{1,n}^*(k), \hat{Z}_{0,n}^*(k))$ *indexed by* k *are both greater than 1.*

E.8 Proof of Theorem 5.2

The proof is divided into three steps. For $j = 0, 1$, denote $Z_{j,n}^*(k) = \alpha_{j,b}(\hat{q}_j^*(\tau_b) - q_j(0))$[19] where $\alpha_{j,b}$ is the infeasible convergence rate defined after Assumption 7. In the first step, I want to show that $(Z_{1,n}^*(k), Z_{0,n}^*(k))$ as a two-dimensional stochastic process indexed by k in $\mathcal{D}([\kappa_1, \kappa_2])$ converges weakly to $(Z_{1,\infty}(k), Z_{0,\infty}(k))$ defined in Theorem 4.1 under Skorohod metric. In the second step, I want to show that $\hat{\alpha}_b^*(\hat{q}^*(\tau_b) - q(\tau_b))$ as a stochastic process indexed by k in $\mathcal{D}([\kappa_1, \kappa_2])$ converges weakly to $Z_\infty^c(k)$ defined in Theorem 4.2 under the Skorohod metric. Last, I want to show that $\hat{\alpha}_b^*(\hat{q}(\tau_b) - q(\tau_b))$ as a stochastic process indexed by k in $\mathcal{D}([\kappa_1, \kappa_2])$ converges weakly to 0 under the uniform metric. Combining the results from the last two steps, I can establish the desired result that

$$\hat{\alpha}_b^*(\hat{q}^*(\tau_b) - \hat{q}(\tau_b)) = \hat{\alpha}_b^*(\hat{q}^*(\tau_b) - q(\tau_b)) - \hat{\alpha}_b^*(\hat{q}(\tau_b) - q(\tau_b)) \rightsquigarrow Z_\infty^c(k).$$

Step 1.

$$(\hat{Z}_{1,b}^*(k), \hat{Z}_{0,b}^*(k)) = \arg\min_{(z_1,z_2)} \sum_{j=0,1}\Big\{-\sum_{i=1}^n P_{n,i}W_j(D_i, \hat{P}(X_i))\tau_b z_j + \sum_{i=1}^n P_{n,i}W_j(D_i, \hat{P}(X_i)) l_\delta(\alpha_{j,b}(U_{i,j} - q_j(0)), z_j)\Big\}.$$

If the replacement is allowed, $P_{n,i} = \sum_{l=1}^b \mathbb{1}\{I_l = i\}$, $(I_{n,1}, I_{n,2}, \cdots, I_{n,b})$ is a multinomial vector

[19]It is different from $\hat{Z}_n^*(k) = \hat{\alpha}_b^*(\hat{q}^*(\tau_b) - \hat{q}(\tau_b))$. $\hat{q}_j^*(\tau_b)$ is defined before Theorem 5.2.

with parameter b and probabilities $(\frac{1}{n}, \cdots, \frac{1}{n})$. If replacement is not allowed, $\{P_{n,i}\}_{i=1}^n$ has b 1's and $n-b$ 0's and each combination of $\{P_{n,i}\}_{i=1}^n$ has equal probability $\frac{1}{C_n^b}$. The proof of this step follows the five steps in the proof of Theorem 4.1 which I will not repeat. The key ingredient, Lemma E.2, is replaced by the following Lemma.

Lemma E.4.
(1) $\frac{1}{n}\sum_{i=1}^n P_{n,i}W_j(D_i, P(X_i)) \to 1$ *a.s.*
(2) For $\hat{N}_j^* := \sum_{i=1}^n P_{n,i}\mathbb{1}\{\alpha_{j,b}(U_{i,j}-\beta_{j,b}), D_i, X_i\}$,

$$\hat{N}_j^* \rightsquigarrow N_j := \sum_{i=1}^{\infty} \mathbb{1}\{\mathcal{J}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}\}.$$

(3) Let

$$g_1(u,x) = \frac{1}{P(x)} l_\delta(u,x,z_1), \quad g_0(u,x) = \frac{1}{1-P(x)} l_\delta(u,x,z_0),$$

and

$$\Psi_{j,n} = \sum_{i=1}^n (jD_i + (1-j)(1-D_i))P_{n,i}g_j(\alpha_{j,b}(U_{i,j}-\beta_{j,b}), X_i).$$

Then for a pair of constants (t_1, t_0),

$$\mathbb{E}\exp(\tilde{i}t_1\Psi_{1,n} + \tilde{i}t_0\Psi_{0,n}) \to \mathbb{E}\exp(\tilde{i}\int_{E_1} t_1 d g_1 dN_1)\mathbb{E}\exp(\tilde{i}\int_{E_0} t_0(1-d)g_0 dN_0),$$

in which N_j *is defined in (2).*
(4) The distances between the two closest discontinuities of the marginal sample paths of the two-dimensional stochastic process $(\hat{Z}_{1,n}^*(k), \hat{Z}_{0,n}^*(k))$ *indexed by* k *are both greater than 1.*

Step 2.
First, I note that

$$\hat{\alpha}_b^*(\hat{q}^*(\tau_b) - q(0)) = \frac{\alpha_b^*}{\alpha_{1,b}}\hat{Z}_{1,b}^*(k) - \frac{\alpha_b^*}{\alpha_{0,b}}\hat{Z}_{0,b}^*(k),$$

$$\begin{aligned}&\alpha_{1,b}\max(\hat{q}_1^*(m\tau_{b,l'}) - \hat{q}_1^*(\tau_{b,l'}), \hat{q}_0^*(m\tau_{b,l'}) - \hat{q}_0^*(\tau_{b,l'}))\\ \to &\max(Z_{1,\infty}(mk_{l'}) - Z_{1,\infty}(k_{l'}), \tilde{\rho}(Z_{0,\infty}(mk_{l'}) - Z_{0,\infty}(k_{l'}))),\end{aligned}$$

and similarly,

$$\begin{aligned}&\alpha_{0,b}\max(\hat{q}_1^*(m\tau_{b,l'}) - \hat{q}_1^*(\tau_{b,l'}), \hat{q}_0^*(m\tau_{b,l'}) - \hat{q}_0^*(\tau_{b,l'}))\\ \to &\max(\frac{1}{\tilde{\rho}}(Z_{1,\infty}(mk_{l'}) - Z_{1,\infty}(k_{l'})), Z_{0,\infty}(mk_{l'}) - Z_{0,\infty}(k_{l'})).\end{aligned}$$

By step 1, I have

$$(\hat{Z}_{1,b}^*(k), \hat{Z}_{0,b}^*(k)) \rightsquigarrow (Z_{1,\infty}(k), Z_{0,\infty}(k)).$$

Therefore

$$\hat{\alpha}_b^*(\hat{q}^*(\tau_b) - q(\tau_b)) \rightsquigarrow \frac{\sqrt{k_{l'}}(Z_{1,\infty}(k) - \tilde{\rho}Z_{0,\infty}(k))}{\max(Z_{1,\infty}(mk_{l'}) - Z_{1,\infty}(k_{l'}), \tilde{\rho}(Z_{0,\infty}(mk_{l'}) - Z_{0,\infty}(k_{l'})))}.$$

Last, I have that $\alpha_{j,b}(q_j(\tau_b) - q_j(0)) \to \eta_j(k)$ uniformly in $k \in [\kappa_1, \kappa_2]$. Combining this with the above result, I obtain that

$$\hat{\alpha}_b^*(\hat{q}^*(\tau_b) - q(\tau_b)) \rightsquigarrow Z_\infty^c(k) := \frac{\sqrt{k_{l'}}(Z_{1,\infty}^c(k) - \tilde{\rho} Z_{0,\infty}^c(k))}{\max(Z_{1,\infty}(mk_{l'}) - Z_{1,\infty}(k_{l'}), \tilde{\rho}(Z_{0,\infty}(mk_{l'}) - Z_{0,\infty}(k_{l'})))}.$$

This concludes step 2.

Step 3.
By construction, $\tau_b n = \tau_n n \frac{n}{b} \to \infty$. By Theorem 3.1, $\lambda_{j,n}(k)(\hat{q}_j(\tau_b) - q_j(\tau_b))$ as a stochastic process indexed by k is tight. I only need to show $\frac{\hat{\alpha}_b^*}{\lambda_{j,n}(k)} \to 0$. To see this, I note that, by step 1, $\hat{\alpha}_b^* = O_p(\min(\alpha_{1,b}, \alpha_{0,b}))$. Furthermore, since $k \in [\kappa_1, \kappa_2]$, I have

$$\frac{\hat{\alpha}_b^*}{\lambda_{j,n}(k)} \lesssim_p \frac{\alpha_{j,b}}{\lambda_{j,n}(k)} \lesssim_p \sqrt{\frac{b}{n\kappa_1}} = o(1).$$

This concludes the proof.

E.9 Proof of Corollary 5.1

By Assumption 13 and Theorem 5.2, I have

$$\hat{Z}_n^{c*}(k)/S_n(k) \rightsquigarrow Z_\infty^c(k)/\sigma(k) \text{ in } \mathcal{D}[\kappa_1, \kappa_2].$$

Let ρ be the Skorohod metric on $\mathcal{D}([\kappa_1, \kappa_2])$. Since 0 is a constant function, the map $\rho(s, 0) = \sup_{k \in [\kappa_1, \kappa_2]} |s|$ is continuous in $s \in \mathcal{D}([\kappa_1, \kappa_2])$. Therefore,

$$\sup_{k \in [\kappa_1, \kappa_2]} |\hat{Z}_n^{c*}(k)/S_n(k)| \rightsquigarrow \sup_{k \in [\kappa_1, \kappa_2]} |Z_\infty^c(k)/\sigma(k)|.$$

Next, I note that $\sup_{k \in [\kappa_1, \kappa_2]} |Z_\infty^c(k)/\sigma(k)|$ is continuously distributed by Lemma E.8. Thus,

$$\widehat{C}_{1-a} \xrightarrow{p} C_{1-a}$$

in which $\widehat{C}_{1-a}$ and C_{1-a} are the $(1-a)$-th quantiles of

$$\sup_{k \in [\kappa_1, \kappa_2]} |\hat{Z}_n^{c*}(k)/S_n(k)| \quad \text{and} \quad \sup_{k \in [\kappa_1, \kappa_2]} |Z_\infty^c(k)/\sigma(k)|, \quad \text{respectively.}$$

This implies that the $(1-a)$-th uniform confidence band is consistent, that is,

$$\lim_{n \to \infty} P\left(q(\frac{k}{n}) \in \left[\hat{q}(\frac{k}{n}) - S_n(k)\widehat{C}_{1-a}/\hat{\alpha}_n, \hat{q}(\frac{k}{n}) + S_n(k)\widehat{C}_{1-a}/\hat{\alpha}_n\right] : k \in [\kappa_1, \kappa_2]\right) = 1 - \alpha.$$

E.10 Proof of Theorem 5.3

If $\{\tau_n\}_{n\geq 1} \in \Gamma_{ex}$ and $\tau_n \leq \tau_{n,1}$ for n large enough,

$$\widetilde{C}_a^h(\tau_n) = \widetilde{C}_a^{bn}(\tau_n).$$

By Theorem 5.1,

$$P\left(q(\tau_n) \in \left(\hat{q}(\tau_n) - \widetilde{C}^h_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^h_{\frac{a}{2}}(\tau_n)\right)\right) = 1 - a.$$

If $\{\tau_n\}_{n\geq 1} \in \Gamma_{ex}$ and for n large enough, $\tau_n > \tau_{n,1}$,

$$\widetilde{C}_a^h(\tau_n) = \widetilde{C}_a^{lf}(\tau_n)$$

and thus

$$\begin{aligned}
&P\left(q(\tau_n) \in \left(\hat{q}(\tau_n) - \widetilde{C}^h_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^h_{\frac{a}{2}}(\tau_n)\right)\right) \\
\geq &P\left(q(\tau_n) \in \left(\hat{q}(\tau_n) - \widetilde{C}^{bn}_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^{bn}_{\frac{a}{2}}(\tau_n)\right)\right) = 1 - a.
\end{aligned}$$

These two situations exhaust all sequences in Γ_{ex}.

If $\{\tau_n\}_{n\geq 1} \in \Gamma_{int}$, for n large enough, I have $\tau_n \geq \tau_{n,1}$. This implies that

$$\begin{aligned}
&P\left(q(\tau_n) \in \left(\hat{q}(\tau_n) - \widetilde{C}^h_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^h_{\frac{a}{2}}(\tau_n)\right)\right) \\
\geq &P\left(q(\tau_n) \in \left(\hat{q}(\tau_n) - \widetilde{C}^{nn}_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^{nn}_{\frac{a}{2}}(\tau_n)\right)\right) = 1 - a,
\end{aligned}$$

where the last equality is by Theorem 3.3.

If $\{\tau_n\}_{n\geq 1} \in \Gamma_{reg}$, for n large enough, I have $\tau_n \geq \tau_{n,2}$. This implies that

$$\widetilde{C}_a^h(\tau_n) = \widetilde{C}_a^{nn}(\tau_n),$$

and thus by the assumption in the theorem,

$$P\left(q(\tau_n) \in \left(\hat{q}(\tau_n) - \widetilde{C}^h_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \widetilde{C}^h_{\frac{a}{2}}(\tau_n)\right)\right) = 1 - a.$$

E.11 Proof of Proposition 5.2

It suffices to show that $\hat{\alpha}_n(\hat{q}(0) - q(0)) \rightsquigarrow \sum_{l=1}^{L} \gamma_l Z^c_\infty(k_l)$. Then Proposition 5.1 shows that $\widehat{C}_a$ is consistent for the a-th quantile of $\sum_{l=1}^{L} \gamma_l Z^c_\infty(k_l)$.

First, by Theorem 3.4, $\hat{\xi}_j \xrightarrow{p} \xi_j$ for $j = 0, 1$. This implies that $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3) \xrightarrow{p} (\gamma_1, \gamma_2, \gamma_3)$ where $(\gamma_1, \gamma_2, \gamma_3)$ is the unique solution to the follow system of equations:

$$\sum_{l=1}^{3} r_l = 1, \quad \sum_{l=1}^{3} r_l k_l^{-\xi_1} = 0, \quad \sum_{l=1}^{3} r_l k_l^{-\xi_0} = 0. \tag{E.24}$$

In addition,

$$\hat{\alpha}_n(\hat{q}(0) - q(0)) = \hat{\alpha}_n\left\{\sum_{l=1}^{3}[\hat{r}_l\hat{q}(\tau_{n,l}) - r_l q(\tau_{n,l})]\right\} + \hat{\alpha}_n\left\{\sum_{l=1}^{3} r_l[q(\tau_{n,l}) - q(0)]\right\}.$$

Since $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3) \xrightarrow{p} (\gamma_1, \gamma_2, \gamma_3)$, by Proposition 4.2, the first term converges weakly to $\sum_{l=1}^{L}\gamma_l Z_\infty^c(k_l)$. For the second term, since $\alpha_{j,n}(q_j(\tau_{n,l}) - q_j(0)) \to \eta_j(k_l) = k_l^{-\xi_j}$ and $\frac{\hat{\alpha}_n}{\alpha_{j,n}} = O_p(1)$, by (E.24), I have

$$\hat{\alpha}_n\left\{\sum_{l=1}^{3} r_l[q(\tau_{n,l}) - q(0)]\right\} = \left(\frac{\hat{\alpha}_n}{\alpha_{1,n}} + \frac{\hat{\alpha}_n}{\alpha_{0,n}}\right)o(1) = o_p(1).$$

This concludes the proof.

E.12 Proof of Theorem 5.4

Let $\xi_1^{(1)}$, $\xi_0^{(1)}$, $\xi_1^{(2)}$, and $\xi_0^{(2)}$ be the EV index for $Y_1^{(1)}$, $Y_0^{(1)}$, $Y_1^{(2)}$, and $Y_0^{(2)}$, respectively. Denote $c(s,t) = k_{l'}^{s-t}\frac{m^{-s}-1}{m^{-t}-1}$. Then following the proof of Theorem 4.2, I have

$$\frac{\alpha_{0,n_1}^{(1)}}{\alpha_{1,n_1}^{(1)}} \to c(\xi_1^{(1)}, \xi_0^{(1)})\rho_1, \quad \frac{\alpha_{0,n_2}^{(2)}}{\alpha_{1,n_2}^{(2)}} \to c(\xi_1^{(2)}, \xi_0^{(2)})\rho_2 \quad \frac{\alpha_{1,n_2}^{(2)}}{\alpha_{1,n_1}^{(1)}} \to \upsilon^{-\xi_1^{(2)}}c(\xi_1^{(1)}, \xi_1^{(2)})\rho_3,$$
$$\frac{\alpha_{0,n_2}^{(2)}}{\alpha_{1,n_1}^{(1)}} \to \upsilon^{-\xi_0^{(2)}}c(\xi_1^{(1)}, \xi_0^{(2)})\rho_4, \quad \frac{\alpha_{1,n_2}^{(2)}}{\alpha_{0,n_1}^{(1)}} \to \upsilon^{-\xi_1^{(2)}}c(\xi_0^{(1)}, \xi_1^{(2)})\rho_5, \quad \text{and} \quad \frac{\alpha_{0,n_2}^{(2)}}{\alpha_{0,n_1}^{(1)}} \to \upsilon^{-\xi_0^{(2)}}c(\xi_0^{(1)}, \xi_0^{(2)})\rho_0. \tag{E.25}$$

In addition,

$$\begin{aligned}&\hat{\alpha}_n\left(\hat{q}^{(1)}(\frac{k}{n_1}) - \hat{q}^{(2)}(\frac{k}{n_2})\right)\\ =&\hat{\alpha}_n(\hat{q}_1^{(1)}(\frac{k}{n_1}) - q_1^{(1)}(\frac{k}{n_1})) - \hat{\alpha}_n(\hat{q}_0^{(1)}(\frac{k}{n_1}) - q_0^{(1)}(\frac{k}{n_1})) - \hat{\alpha}_n(\hat{q}_1^{(2)}(\frac{k}{n_2}) - q_1^{(2)}(\frac{k}{n_2})) + \hat{\alpha}_n(\hat{q}_0^{(2)}(\frac{k}{n_2}) - q_0^{(2)}(\frac{k}{n_2})).\end{aligned}$$

Following (E.25),

$$\begin{aligned}&\hat{\alpha}_n(\hat{q}_1^{(1)}(\frac{k}{n_1}) - q_1^{(1)}(\frac{k}{n_1}))\\ =&\min\left\{V_1^{(1)}, \left(\frac{\alpha_{0,n_1}^{(1)}}{\alpha_{1,n_1}^{(1)}}\right)V_0^{(1)}, \left(\frac{\alpha_{1,n_2}^{(2)}}{\alpha_{1,n_1}^{(1)}}\right)V_1^{(2)}, \left(\frac{\alpha_{0,n_2}^{(2)}}{\alpha_{1,n_1}^{(1)}}\right)V_0^{(2)}\right\}Z_{1,\infty}^{c,(1)}(k) + o_p(1)\\ =&\min\left\{V_1^{(1)}, \left(c(\xi_1^{(1)}, \xi_0^{(1)})\rho_1\right)V_0^{(1)}, \left(\frac{c(\xi_1^{(1)}, \xi_1^{(2)})\rho_3}{\upsilon^{\xi_1^{(2)}}}\right)V_1^{(2)}, \left(\frac{c(\xi_1^{(1)}, \xi_0^{(2)})\rho_4}{\upsilon^{\xi_0^{(2)}}}\right)V_0^{(2)}\right\}Z_{1,\infty}^{c,(1)}(k) + o_p(1),\end{aligned}$$

in which

$$V_j^{(s)} = \frac{\sqrt{k_{l'}}}{Z_{j,\infty}^{(s)}(mk_{l'}) - Z_{j,\infty}^{(s)}(k_{l'})}, \quad j = 0, 1, \quad s = 1, 2.$$

Similarly,

$$
\begin{aligned}
&\hat{\alpha}_n(\hat{q}_0^{(1)}(\frac{k}{n_1}) - q_0^{(1)}(\frac{k}{n_1})) \\
=&\min\Big\{\Big(\frac{1}{c(\xi_1^{(1)},\xi_0^{(1)})\rho_1}\Big)V_1^{(1)}, V_0^{(1)}, \Big(\frac{c(\xi_0^{(1)},\xi_1^{(2)})\rho_5}{\upsilon^{\xi_1^{(2)}}}\Big)V_1^{(2)}, \Big(\frac{c(\xi_0^{(1)},\xi_0^{(2)})\rho_0}{\upsilon^{\xi_0^{(2)}}}\Big)V_0^{(2)}\Big\}Z_{0,\infty}^{c,(1)}(k) + o_p(1), \\
&\hat{\alpha}_n(\hat{q}_1^{(2)}(\frac{k}{n_2}) - q_1^{(2)}(\frac{k}{n_2})) \\
=&\min\Big\{\Big(\frac{\upsilon^{\xi_1^{(2)}}}{c(\xi_1^{(1)},\xi_1^{(2)})\rho_3}\Big)V_1^{(1)}, \Big(\frac{\upsilon^{\xi_1^{(2)}}}{c(\xi_0^{(1)},\xi_1^{(2)})\rho_5}\Big)V_0^{(1)}, V_1^{(2)}, \Big(c(\xi_1^{(2)},\xi_0^{(2)})\rho_2\Big)V_0^{(2)}\Big\}Z_{1,\infty}^{c,(2)}(k) + o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\hat{\alpha}_n(\hat{q}_0^{(2)}(\frac{k}{n_2}) - q_0^{(2)}(\frac{k}{n_2})) \\
=&\min\Big\{\Big(\frac{\upsilon^{\xi_0^{(2)}}}{c(\xi_1^{(1)},\xi_0^{(2)})\rho_4}\Big)V_1^{(1)}, \Big(\frac{\upsilon^{\xi_0^{(2)}}}{c(\xi_0^{(1)},\xi_0^{(2)})\rho_0}\Big)V_0^{(1)}, \Big(\frac{1}{c(\xi_1^{(2)},\xi_0^{(2)})\rho_2}\Big)V_1^{(2)}, V_0^{(2)}\Big\}Z_{1,\infty}^{c,(2)}(k) + o_p(1).
\end{aligned}
$$

Since the four $\min\{\cdot\}$ terms are all $O_p(1)$ and at least one of them is non-degenerate, there exists a non-degenerate random variable $Z^{\text{TS}}(k)$ such that

$$
\hat{\alpha}_n\Big(\hat{q}^{(1)}(\frac{k}{n_1}) - \hat{q}^{(2)}(\frac{k}{n_2})\Big) \rightsquigarrow Z_\infty^{\text{TS}}(k).
$$

In addition, since the $\min\{\cdot\}$ terms are independent of k and by Lemma E.7, $Z_{j,\infty}^{c,(s)}(k)$ are all continuously distributed for $j = 0, 1,\ s = 1, 2$, $Z_\infty^{\text{TS}}(k)$ is also continuous. Following the similar argument in the proof of Theorem 5.2, I can also show that

$$
\hat{\alpha}_b^*\Big[\Big(\hat{q}^{(1)*}(\frac{k}{b_1}) - \hat{q}^{(1)}(\frac{k}{b_1})\Big) - \Big(\hat{q}^{(2)*}(\frac{k}{b_2}) - \hat{q}^{(2)}(\frac{k}{b_2})\Big)\Big] \rightsquigarrow Z_\infty^{\text{TS}}(k).
$$

The detail is omitted for brevity. This concludes the proof.

E.13 Proof of Corollary A.1

I only have to show the weak convergence of $P(X \in .|Y_1 = y)$ to $\sum_t P(x_t)p_t\mathbb{1}\{x_t \in .\}$, that is, for any $F \in \text{Supp}(X)$ with $\partial F \cap \{x_1, x_2, \cdots, x_T\} = \emptyset$, $\lim_{y\to q_1(0)} P(X \in F|Y_1 = y) = \sum_{t=1}^T p_t\mathbb{1}\{x_t \in F\}$. I first claim that for an arbitrarily small constant γ, there exist a small constant η, such that for any $t = 1, \cdots, T$, if $|y - q_1(0)| < \eta$, $S_{y,t} \subset \{x : |x - x_t| \le \gamma\}$.

Suppose not, since T is finite, as $y \downarrow q_1(0)$, there exists a t and a sequence $x_{y,t} \in S_{y,t}$, such that $|x_{y,t} - x_t| > \gamma_0$. Also because $x_{y,t} \in S_y$, there exists a corresponding $\varepsilon_{y,t}$ such that $g(x_{y,t}, \varepsilon_{y,t}) \le y$. Since $\text{Supp}(X) \times [0, 1]$ is compact, there is a convergent subsequence $\{x_{y',t}, \varepsilon_{y',t}\}$ of $\{x_{y,t}, \varepsilon_{y,t}\}$ with limiting point $(x_{t'}, \varepsilon_{t'})$. Since $g(x_{y',t}, \varepsilon_{y',t}) \le y'$ and g is lower semi-continuous, as $y' \to q_1(0)$,

$g(x_{t'}, \varepsilon'_h) \leq \liminf_{y' \to q_1(0)} g(x_{y',t}, \varepsilon_{y',t}) \leq q_1(0)$. So $g(x_{t'}, \varepsilon'_h) = q_1(0)$. This means $x_{t'} \in S_0$. But $|x_{t'} - x_t| \geq \gamma_0$. In addition, $S_{y,t}$ is monotone decreasing in y by construction so $\{x_{y',t}\} \subset S_{y_0,t}$. This implies $d(x_{t'}, S_{y_0,t}) = 0$ for some $t' \neq t$, which contradicts with the construction of $S_{y_0,t}$.

Let $\delta_0 = \min_{(x_t,x_{t'}) \in S_0 \times S_0} ||x_t - x_{t'}||$ and $\mathcal{B}(x, d)$ be a ball with radius d and center x. Then when y is small enough, $S_{y,t} = S_y \cap \mathcal{B}(x_t, \delta_0/2)$, which is defined independent of the initial partition $\{S_{y_0,t}\}_{t=1}^T$. This implies p_t is well defined independent of $S_{y_0,t}$. Furthermore, for any F such that $\partial F \cap \{x_1, x_2, \cdots, x_T\} = \emptyset$, either $d(x_t, F) > 0$ or $d(x_t, F^c) > 0$ for all $t = 1, 2, \cdots, T$. Whenever y is small enough, either $s_{y,t} \subset F$ if $d(x_t, F^c) > 0$ or $S_{y,t} \cap F = \emptyset$ if $d(x_t, F) > 0$. Therefore, for some arbitrarily small γ, there always exists a y small enough such that

$$
\begin{aligned}
&|P(X \in F|Y_1 = y) - \sum_t p_t \mathbb{1}\{x_t \in F\}| \\
=&\left|\sum_{t=1}^T \left[\frac{\mathbb{E}\mathbb{1}\{X \in S_{y,t} \cap F\}\frac{\partial \lambda(X,y)}{\partial y}}{\mathbb{E}\mathbb{1}\{X \in S_y\}\frac{\partial \lambda(X,y)}{\partial y}} - p_t \mathbb{1}\{x_t \in F\}\right]\right| \\
\leq&\sum_{t=1}^T |p_{y,t} - p_t| \mathbb{1}\{x_t \in F\} \\
\leq&M\gamma
\end{aligned}
$$

This implies that $P(X \in .|Y_1 = y)$ weakly converges to $\sum_t p_t \mathbb{1}\{x_t \in .\}$.

E.14 Proof of Corollary A.3

In the proof of corollary A.1, I have shown that for any $\gamma > 0$, $S^d_{y,r} \subset \mathcal{B}(x_r, \gamma)$. I next show that it is also true for $S^c_{y,t}$, that is, $S^c_{y,t} \subset (S_{0,t})^\gamma$.

Suppose not, there exists $\gamma_0 > 0$ and a sequence $x_{y,t} \in S^c_{y,t}$ such that $d(x_{y,t}, S_{0,t}) > \gamma_0$. $x_{y,t} \in S^c_{y,t}$ implies that there exists a corresponding sequence $\{e_{y,t}\}$ such that $g(x_{y,t}, e_{y,t}) \leq y$. Then there exists a convergent subsequence $(x_{y',t}, e_{y',t})$ with limit (x', e') such that $g(x', e') \leq \liminf_{y \to q_1(0)} g(x_{y,t}, e_{y,t}) \leq q_1(0)$. This implies $x' \in S_0$. But $d(x', S_{0,t}) > \gamma_0$, so $x' \in S_{0,t'}$ for $t' \neq t$ or $x' = x_r$, for some $r = 1, 2, \cdots, R^d$. But $S^c_{y,t}$ is decreasing so I have $d(x', S^c_{y_0,t}) = 0$, which contradicts with the way I construct $\{S^c_{y_0,t}\}_{t=1}^T$ and $\{S^d_{y_0,t}\}_{r=1}^{R^d}$.

The above claim implies that whenever y is small enough, $S^d_{y,r} = \mathcal{B}(x_r, \delta_0/2) \cap S_y$ and $S^c_{y,t} = (S_{0,t})^{\delta_0/2} \cap S_y$. Then $\{S^d_{y,r}\}_{r=1}^{R^d}$ and $\{S^c_{y,t}\}_{t=1}^T$ are defined independent of $\{S^c_{y_0,t}\}_{t=1}^T$ and $\{S^d_{y_0,r}\}_{r=1}^{R^d}$ and they are disjoint. This implies that $p^d_{y,r}$ and $p^c_{y,t}$ are well defined independent of $\{S^c_{y_0,t}\}_{t=1}^T$ and $\{S^d_{y_0,r}\}_{r=1}^{R^d}$. Furthermore, $S_{0,t}$ is compact because for a convergent sequence $\{x_n\}_{n=1}^\infty$ with limit x, there exists a corresponding sequence $\{\varepsilon_n\}_{n=1}^\infty \subset [0, 1]$ such that it has a convergent subsequence $\{\varepsilon'_n\}$ with limit ε. Then $g(x, \varepsilon) \leq \liminf_{n'} g(x_{n'}, \varepsilon_{n'}) \leq q_1(0)$, which implies $x \in S_0$. Since all $S_{0,t'}$, $t' = 1, 2, \cdots, T$ are separate, it implies $x \in S_{0,t}$. Therefore, $F \cap S^c_{y,t} \to F \cap S_{0,t}$.

The potential discontinuity $\mathcal{S}$ of the limiting distribution is $\{x_r\}_{r=1}^{R^d} \cup \left(\mathcal{S}_X \cap (\cup_{r=1}^{R^c}(\partial S_{0,r}))\right)$ where

$\mathcal{S}_X$ is the discontinuity of X. Let $\mathcal{F}$ be a collection of all open and relatively compact set such that $\partial F \cap \mathcal{S} = \emptyset$. Then, in order to show the weak convergence, it suffices to show that

$$\lim_{y \to q_1(0)} P(X \in F | Y_1 = y) = \sum_r \mathbb{1}\{X_r \in F\} p_r^d + \sum_{t=1}^{T_c} p_t^c \int_{S_{0,t} \cap F} \frac{\sigma_t(x)^{1/\xi_t} dF_X(x)}{\int_{S_{0,t}} \sigma_t(x)^{1/\xi_t} dF_X(x)},$$

for all F with $\partial F \cap \mathcal{S} = \emptyset$.

Notice that $\frac{f_U(y-q_1^*|X)}{f_{\varepsilon_t}(y-q_1^*)} \to \sigma_t(X)^{-1/\xi_t}$ locally uniformly and $F \cap S_{y,t}^c \to F \cap S_{0,t}$. Then, by the dominated convergence theorem, as $y \to q_1(0)$, I have

$$\begin{aligned} \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{y,t}^c\} \frac{\partial \lambda(X,y)}{\partial y}}{\mathbb{E}\mathbb{1}\{X \in S_{y,t}^c\} \frac{\partial \lambda(X,y)}{\partial y}} &= \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{y,t}^c\} \frac{f_U(y-q_1^*|X)}{f_{\varepsilon_t}(y-q_1^*)}}{\mathbb{E}\mathbb{1}\{X \in S_{y,t}^c\} \frac{f_U(y-q_1^*|X)}{f_{\varepsilon_t}(y-q_1^*)}} \\ &\to \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{0,t}\} \sigma_t(X)^{-1/\xi_t}}{\mathbb{E}\mathbb{1}\{X \in S_{0,t}\} \sigma_t(X)^{-1/\xi_t}}. \end{aligned}$$

Therefore, for any fixed F such that $\partial F \cap \mathcal{S} = \emptyset$, as $y \to q_1(0)$,

$$\begin{aligned} &P(X \in F | Y = y) \\ =& \frac{\mathbb{E}\mathbb{1}\{X \in F\} \frac{\partial \lambda(X,y)}{\partial y}}{\mathbb{E}\mathbb{1}\{X \in S_y\} \frac{\partial \lambda(X,y)}{\partial y}} \\ =& \sum_{r=1}^{R^d} \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{y,r}^d\} \frac{\partial \lambda(X,y)}{\partial y}}{\mathbb{E}\mathbb{1}\{X \in S_y\} \frac{\partial \lambda(X,y)}{\partial y}} + \sum_{t=1}^{T} \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{y,t}^c\} \frac{\partial \lambda(X,y)}{\partial y}}{\mathbb{E}\mathbb{1}\{X \in S_y\} \frac{\partial \lambda(X,y)}{\partial y}} \\ =& \sum_{r=1}^{R^d} p_{y,r}^d \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{y,r}^d\} \frac{\partial \lambda(X,y)}{\partial y}}{\mathbb{E}\mathbb{1}\{X \in S_{y,r}^d\} \frac{\partial \lambda(X,y)}{\partial y}} + \sum_{t=1}^{T} p_{y,t}^c \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{y,t}^c\} \frac{\partial \lambda(X,y)}{\partial y}}{\mathbb{E}\mathbb{1}\{X \in S_{y,t}^c\} \frac{\partial \lambda(X,y)}{\partial y}} \\ \to& \sum_{r=1}^{R^d} p_r^d \mathbb{1}\{x_r \in F\} + \sum_{t=1}^{T} p_t^c \frac{\mathbb{E}\mathbb{1}\{X \in F \cap S_{0,t}\} \sigma_t(X)^{-1/\xi_t}}{\mathbb{E}\mathbb{1}\{X \in S_{0,t}\} \sigma_t(X)^{-1/\xi_t}}. \end{aligned}$$

This concludes the proof.

E.15 Proof of Lemma E.2

(1) is trivial.

For (2), it is known that a Poisson random measure (PRM) with the Lebesgue mean measure can be written as $\sum_{i=1}^{\infty} \mathbb{1}\{\sum_{l=1}^{i} \mathcal{E}_i \in .\}$ where $\mathcal{E}_i$ is independent and identically standard exponentially distributed. Then by Proposition 3.7 and 3.8 in Resnick (1987), I can transform and augment the baseline point process and show that PRM$(\mu_j) = N_j(.) := \sum_{i=1}^{\infty} \mathbb{1}\{(\mathcal{J}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}) \in .\}$ for $j = 0, 1$, in which for $j = 0, 1$,

$$\mu_j((a,b) \times \{d\} \times F) = \int_F (dP(x) + (1-d)(1-P(x))) P_j^+(dx | Y_j = q_j(0))(h_j(b) - h_j(a)).$$

I focus on $j = 1$. Since $P_1^+(X \in .|Y_1 = q_1(0))$ is a bounded measure, its discontinuities are at most countable. So there exists $\mathcal{F}_1$, a basis of relatively compact open sets of $\mathbb{R}^d$ such that $\mathcal{F}_1$ is closed under finite unions and intersections and for any $F \in \mathcal{F}_1$, $P_1^+(X \in \partial F|Y_1 = q_1(0)) = 0$. Then by Lemma 9.3 and 9.4 in Chernozhukov (2005), I only have to verify that, for any $F \in \mathcal{F}_1$ and any interval (a, b), $\mathbb{E}\hat{N}_1((a, b) \times \{d\} \times F) \to \mu_1((a, b) \times \{d\} \times F)$. Notice that $l/\alpha_{1,n} + \beta_{1,n} \downarrow F_{u_1}^{-1}(0) = 0$ or $-\infty$ for any $l \in (-\infty, +\infty)$ for type 1 tails, any $l \in (-\infty, 0)$ for type 2 tails, and any $l \in [0, +\infty)$ for type 3 tails. Let $S_n = (q_1(0) + \beta_{1,n} + a/\alpha_{1,n}, q_1(0) + \beta_{1,n} + b/\alpha_{1,n})$ and $f_1(y)$ be the density of Y_1. By the continuous mapping theorem, I obtain that

$$
\begin{aligned}
&\mathbb{E}\hat{N}_1((a, b) \times \{d\} \times F) \\
=&P(D = d, X \in F|\alpha_{1,n}(U_1 - \beta_{1,n}) \in (a, b))nP(\alpha_{1,n}(U_1 - \beta_{1,n}) \in (a, b)) \\
=&(1 + o(1))\frac{\int_{S_n} P(D = d, X \in F|Y_1 = y)f_1(y)dy}{\int_{S_n} f_1(y)dy}(h_1(b) - h_1(a)) \\
=&(1 + o(1))\frac{\int_{S_n \times F}(dP(x) + (1 - d)(1 - P(x)))P(dx|Y_1 = y)f_1(y)dy}{\int_{S_n} f_1(y)dy}(h_1(b) - h_1(a)) \\
\to& \int_F (dP(x) + (1 - d)(1 - P(x)))P_1^+(dx|Y_1 = q_1(0))(h_1(b) - h_1(a)).
\end{aligned}
$$

This is the desired result for the marginal convergence.

For (3), let $(U'_{i,j}, X'_{i,j})_{j=0,1}$ be an i.i.d. sequence such that $(U'_{i,1}, X'_{i,1}) \perp\!\!\!\perp (U'_{i,0}, X'_{i,0})$ and that $(U'_{i,j}, X'_{i,j})$ is distributed as $(U_{i,j}, X_i)|D_i = j$. Let $p = P(D_i = 1)$. Then

$$
\begin{aligned}
&\mathbb{E}\exp(\tilde{i}t_1\Psi_{1,n} + \tilde{i}t_0\Psi_{0,n})\mathbb{1}\{D_1 = 1, \cdots, D_s = 1, D_{s+1} = 0, \cdots, D_n = 0\} \\
=&\mathbb{E}\exp\left(\tilde{i}t_1(\sum_{i=1}^{s} g_1(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), X_i)) + \tilde{i}t_0(\sum_{i=s+1}^{n} g_0(\alpha_{0,n}(U_{i,0} - \beta_{0,n}), X_i))\right) \\
&\times \mathbb{1}\{\{D_i = 1\}_{i=1}^s, \{D_i = 0\}_{i=s+1}^n\} \\
=&p^s(1 - p)^{n-s}\mathbb{E}\exp\left(\tilde{i}t_1(\sum_{i=1}^{s} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}))\right) \\
&\times \mathbb{E}\exp\left(\tilde{i}t_0(\sum_{i=s+1}^{n} g_0(\alpha_{0,n}(U'_{i,0} - \beta_{0,n}), X'_{i,0}))\right).
\end{aligned}
$$

Therefore, by symmetry,

$$
\begin{aligned}
&\mathbb{E}\exp(\tilde{i}t_1\Psi_{1,n} + \tilde{i}t_0\Psi_{0,n}) \\
=&\sum_{s=0}^{n}\Bigg\{C_n^s p^s(1 - p)^{n-s}\mathbb{E}\exp\left(\tilde{i}t_1(\sum_{i=1}^{s} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}))\right) \\
&\times \mathbb{E}\exp\left(\tilde{i}t_0(\sum_{i=s+1}^{n} g_0(\alpha_{0,n}(U'_{i,0} - \beta_{0,n}), X'_{i,0}))\right)\Bigg\}.
\end{aligned}
$$

Define E'_j for $j = 0, 1$ as follows:

$$\begin{aligned} &\text{for type 1 tails } (\xi_j = 0): && E'_j = [-\infty, +\infty) \times \mathrm{Supp}(\mathcal{X}), \\ &\text{for type 2 tails } (\xi_j > 0): && E'_j = [-\infty, 0) \times \mathrm{Supp}(\mathcal{X}), \\ &\text{for type 3 tails } (\xi_j < 0): && E'_j = [0, +\infty) \times \mathrm{Supp}(\mathcal{X}). \end{aligned}$$

Let N'_j be PRM(μ'_j) on E'_j with

$$\mu'_j([a,b] \times F) = \int_F (jP(x) + (1-j)(1-P(x)))P_j^+(dx|Y_j = q_j(0))(h_j(b) - h_j(a))$$

and

$$\widehat{N}'_j(.) := \sum_{i=1}^{js+(n-s)(1-j)} \mathbb{1}\left\{(\alpha_{j,n}(U'_{i,j} - \beta_{j,n}), X'_{i,j}) \in .\right\}.$$

Let $r_n = \sqrt{2n\log(\log(n))}$, $S_n = \{s \in \mathcal{Z}, |s - np| \le r_n\}$. Then,

$$\begin{aligned} &\left|\mathbb{E}\exp(\tilde{i}t_1\Psi_{1,n} + \tilde{i}t_0\Psi_{0,n}) - \mathbb{E}\exp(\tilde{i}\int_{E_1} t_1 g_1 dN'_1)\mathbb{E}\exp(\tilde{i}\int_{E_0} t_0 g_0 dN'_0)\right| \\ &\le \sum_{s\in S_n} C_n^s p^s(1-p)^{n-s}\left|\mathbb{E}\exp\left(\tilde{i}t_1\int_{E'_1} g_1 d\widehat{N}'_1\right)\mathbb{E}\exp\left(\tilde{i}t_0\int_{E'_0} g_0 d\widehat{N}'_0\right)\right. \\ &\quad \left.- \mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 g_1 dN'_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0 g_0 dN'_0\right)\right| \\ &+ \sum_{s\in S_n^c} C_n^s p^s(1-p)^{n-s}\left|\mathbb{E}\exp\left(\tilde{i}t_1\int_{E'_1} g_1 d\widehat{N}'_1\right)\mathbb{E}\exp\left(\tilde{i}t_0\int_{E'_0} g_0 d\widehat{N}'_0\right)\right. \\ &\quad \left.- \mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 g_1 dN'_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0 g_0 dN'_0\right)\right| \\ &\le \sum_{s\in S_n} C_n^s p^s(1-p)^{n-s}\left|\mathbb{E}\exp\left(\tilde{i}t_1\int_{E'_1} g_1 d\widehat{N}'_1\right)\mathbb{E}\exp\left(\tilde{i}t_0\int_{E'_0} g_0 d\widehat{N}'_0\right)\right. \\ &\quad \left.- \mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 g_1 dN'_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0 g_0 dN'_0\right)\right| + const \times \left(\sum_{s\in S_n^c} C_n^s p^s(1-p)^{n-s}\right). \end{aligned} \tag{E.26}$$

By the law of iterated logarithm, $\sum_{s\in S_n^c} C_n^s p^s(1-p)^{n-s} = o(1)$ as $n \to \infty$. Therefore, the second term is asymptotically negligible. For the first term, if $s \geq [np]$,

$$
\begin{aligned}
&\left|\mathbb{E}\exp(\tilde{i}t_1\int_{E_1'} g_1 d\widehat{N}_1') - \mathbb{E}\exp(\tilde{i}t_1\int_{E_1'} g_1 dN_1')\right| \\
\leq&\left|\mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1'} g_1 d\widehat{N}_1'\right) - \mathbb{E}\exp\left(\tilde{i}t_1\sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')\right)\right| \\
+&\left|\mathbb{E}\exp\left(\tilde{i}t_1\sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')\right) - \mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1'} g_1 dN_1'\right)\right| \\
\leq&\left|\mathbb{E}\exp\left(\tilde{i}t_1\sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')\right)\left[\exp\left(\tilde{i}t_1\sum_{i=[np]}^{s} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')\right) - 1\right]\right| \\
+&\left|\mathbb{E}\exp\left(\tilde{i}t_1\sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')\right) - \mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1'} g_1 dN_1'\right)\right| \\
\leq&\mathbb{E}\left(2 - 2\cos(t_1\sum_{i=[np]+1}^{s} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}'))\right)^{1/2} \\
+&\left|\mathbb{E}\exp\left(\tilde{i}t_1\sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')\right) - \mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1'} g_1 dN_1'\right)\right|,
\end{aligned}
\tag{E.27}
$$

in which the last inequality is by the fact that $|\exp(\tilde{i}t) - 1|^2 \leq 2 - 2\cos(t)$.

Similar to the proof in step 1,

$$
\begin{aligned}
&[np]P(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}) \in [a,b], X_{i,1}' \in F) \\
=&\frac{[np]}{p}P(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}) \in [a,b], X_{i,1}' \in F, D_i = 1) \\
=&\frac{[np]}{p}\int_a^b\int_F P(x)P(dx|\alpha_{1,n}(U_{i,1} - \beta_{1,n}) = u)dP(\alpha_{1,n}(U_{i,1} - \beta_{1,n}) \leq u) \\
\to&\int_F P(x)P_1^+(dx|Y_1 = q_1(0))(h_1(b) - h_1(a)) \\
=&\mu_1'([a,b] \times F).
\end{aligned}
$$

Then by the continuous mapping theorem and the fact that $g_1(u,x) \in C_k(E_1')$, I have

$$
\sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}') \rightsquigarrow \int_{E_1} g_1 dN_1'.
$$

Similarly, because $\frac{r_n}{n} \to 0$, I have that

$$\sum_{i=[np]+1}^{s} |g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1})| \leq \sum_{i=[np]+1}^{[np+r_n]+1} |g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1})| = o_p(1).$$

Therefore, for the first term on the RHS of (E.27), I have

$$\begin{aligned}
&\sup_{s\in S_n, s\geq [np]} \left(2 - 2\cos(t_1 \sum_{i=[np]+1}^{s} g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1}))\right) \\
\leq & 2\left(1-\cos(|t_1|\sum_{i=[np]+1}^{[np+r_n]+1} |g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1}))\right) \mathbb{1}\left\{\sum_{i=[np]+1}^{[np+r_n]+1} |g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1})| \leq \frac{\pi}{|t_1|}\right\} \\
& + 2\mathbb{1}\left\{\sum_{i=[np]+1}^{[np+r_n]+1} |g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1})| \geq \frac{\pi}{|t_1|}\right\} \\
= & o_p(1).
\end{aligned}$$

Therefore, by the dominated convergence theorem, I have

$$\sup_{s\in S_n, s\geq [np]} \mathbb{E}\left(2-2\cos(t_1 \sum_{i=[np]+1}^{s} g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1}))\right) \to 0.$$

For the second term of (E.27), I have, by the dominated convergence theorem, that

$$\left|\mathbb{E}\exp\left(\tilde{i}t_1 \sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1})\right) - \mathbb{E}\exp\left(\tilde{i}t_1 \int_{E'_1} g_1 dN'_1\right)\right| \to 0.$$

Combining the two terms, I obtain that

$$\sup_{s\in S_n, s\geq [np]} \left|\mathbb{E}\exp\left(\tilde{i}t_1 \int_{E'_1} g_1 d\widehat{N}'_1\right) - \mathbb{E}\exp\left(\tilde{i}t_1 \int_{E'_1} g_1 dN'_1\right)\right| \to 0.$$

If $s < [np]$, then $\sum_{i=s}^{[np]-1} \left|g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1})\right| \leq \sum_{i=[np-r_n]}^{[np]-1} \left|g_1(\alpha_{1,n}(U'_{i,1}-\beta_{1,n}), X'_{i,1})\right| = o_p(1)$. By the same argument, I have

$$\sup_{s\in S_n, s< [np]} \left|\mathbb{E}\exp\left(\tilde{i}t_1 \int_{E'_1} g_1 d\widehat{N}'_1\right) - \mathbb{E}\exp\left(\tilde{i}t_1 \int_{E'_1} g_1 dN'_1\right)\right| \to 0.$$

To sum up, I have $\sup_{s\in S_n} \left|\mathbb{E}\exp\left(\tilde{i}t_1 \int_{E'_1} g_1 d\widehat{N}'_1\right) - \mathbb{E}\exp\left(\tilde{i}t_1 \int_{E'_1} g_1 dN'_1\right)\right| \to 0$. Similarly, I can show that

$$\sup_{s\in S_n} \left|\mathbb{E}\exp\left(\tilde{i}t_0 \int_{E'_0} g_0 d\widehat{N}'_0\right) - \mathbb{E}\exp\left(\tilde{i}t_0 \int_{E'_0} g_0 dN'_0\right)\right| \to 0.$$

This implies

$$\begin{aligned}
&\sum_{s\in S_n} C_n^s p^s(1-p)^{n-s}\Bigg|\mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1'} g_1 d\widehat{N}_1'\right)\mathbb{E}\exp\left(\tilde{i}t_0\int_{E_0'} g_0 d\widehat{N}_0'\right)\\
&-\mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 d g_1 dN_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0(1-d) g_0 dN_0\right)\Bigg|\\
&\leq \sup_{s\in S_n}\left|\mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1'} g_1 d\widehat{N}_1'\right)-\mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1'} g_1 dN_1'\right)\right|\\
&\quad+\sup_{s\in S_n}\left|\mathbb{E}\exp\left(\tilde{i}t_0\int_{E_0'} g_0 d\widehat{N}_0'\right)-\mathbb{E}\exp\left(\tilde{i}t_0\int_{E_0'} g_0 dN_0'\right)\right|\\
&\to 0
\end{aligned} \tag{E.28}$$

Combining (E.26) and (E.28),

$$\left|\mathbb{E}\exp\left(\tilde{i}t_1\Psi_{1,n}+\tilde{i}t_0\Psi_{0,n}\right)-\mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 g_1 dN_1'\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0 g_0 dN_0'\right)\right|\to 0.$$

Last, notice that the random variable $\int_{E_j'} g_j dN_j'$ is uniquely determined by its characteristic function

$$\mathbb{E}\left(\exp(\tilde{i}t\int_{E_j'} g_j dN_j')\right)=\exp\left(-\int_{E_j'}(1-\exp(-\tilde{i}t g_j))d\mu_j'\right).$$[20]

Similarly, the random variable $\int_{E_j}(dj+(1-d)(1-j))g_j dN_j$ is uniquely determined by its characteristic function

$$\mathbb{E}\exp\left(\tilde{i}t\int_{E_j}(dj+(1-d)(1-j))g_j dN_j\right)=\exp\left(-\int_{E_j'}\left(1-\exp(-\tilde{i}t(jd+(1-j)(1-d))g_j)\right)d\mu_j\right).$$

In addition, I have

$$\begin{aligned}
&\int_{E_j}\left(1-\exp(-\tilde{i}t(jd+(1-j)(1-d))g_j)\right)d\mu_j\\
=&\int_{E_j}(jd+(1-j)(1-d))(1-\exp(-\tilde{i}tg_j))d\mu_j\\
=&\int_{E_j'} jP(x)(1-\exp(-\tilde{i}tg_j))d\mu_j(u,1,x)+\int_{E_j'}(1-j)(1-P(x))(1-\exp(-\tilde{i}tg_j))d\mu_j(u,0,x)\\
=&\int_{E_j'}(1-\exp(-\tilde{i}tg_j))d\mu_j'(u,x),
\end{aligned}$$

[20] See the definition of Laplace functional of PRM(μ) in section 3.2 of Resnick (1987).

that is, the two characteristic functions are the same. This implies

$$\int_{E'_j} g_j dN'_j = \int_{E_j} (dj + (1-d)(1-j)) g_j dN_j.$$

Therefore

$$\mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 g_1 dN'_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0 g_0 dN'_0\right) = \mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 d g_1 dN_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0(1-d) g_0 dN_0\right)$$

and

$$\left|\mathbb{E}\exp\left(\tilde{i}t_1\Psi_{1,n} + \tilde{i}t_0\Psi_{0,n}\right) - \mathbb{E}\exp\left(\tilde{i}\int_{E_1} t_1 d g_1 dN_1\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0} t_0(1-d) g_0 dN_0\right)\right| \to 0.$$

For part (4), it is easy to see that $(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k))$ are piece-wise constant because for instance, when $j = 1$ and $k - \frac{1}{\hat{P}(X_h)} < \sum_{i\neq h}\frac{T_i}{\hat{P}(X_i)}\mathbb{1}\{\alpha_n U_{i,1} < \alpha_n U_{h,1}\} < k$ for some h such that $T_h = 1$, then $\hat{Z}_{1,n}(k) = \alpha_n U_{h,1}$. The discontinuity for the sample path only occurs at $k - \frac{1}{\hat{P}(X_h)} = \sum_{i\neq h}\frac{T_i}{\hat{P}(X_i)}\mathbb{1}\{\alpha_n U_{i,1} < \alpha_n U_{h,1}\}$ or $k = \sum_{i\neq h}\frac{T_i}{\hat{P}(X_i)}\mathbb{1}\{\alpha_n U_{i,1} < \alpha_n U_{h,1}\} = k$. W.l.o.g., I assume $0 < \hat{P}(X_i) < 1$ for all i. This implies that the distances between the two closest discontinuities for the sample paths are $\min_{1\leq i\leq n}\frac{1}{\hat{P}(X_i)} \geq 1$.

E.16 Proof of Lemma E.3

For (1), I compute its characteristic function conditioning on data Φ_n. Let $\tilde{i}$ be the imaginary number. I have

$$\begin{aligned}
&\mathbb{E}\left\{\exp\left[\tilde{i}t\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{n}\mathbb{1}\{I_l = i\}W_j(D_i, P(X_i))\right)\right]|\Phi_n\right\}\\
=&\left\{\mathbb{E}\left[\exp\left(\tilde{i}t(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{I_1 = i\}W_j(D_i, P(X_i)))\right)|\Phi_n\right]\right\}^n\\
=&\left\{1 - \frac{1}{n}\left[\sum_{l=1}^{n} 1 - \exp\left(\tilde{i}t\left(\frac{1}{n}W_j(D_l, P(X_l))\right)\right)\right]\right\}^n.
\end{aligned}$$

By the Taylor expansion, $\sum_{l=1}^{n} 1 - \exp\Big(\tilde{i}t\left(\frac{1}{n}W_j(D_l, P(X_l))\right)\Big) - \tilde{i}t\frac{1}{n}\sum_{l=1}^{n} W_j(D_l, P(X_l)) \to 0$ a.s. By SLLN,

$$\frac{1}{n}\sum_{l=1}^{n} W_j(D_l, P(X_l)) \to EW_j(D_l, P(X_l)) = 1 \ a.s.$$

So $\mathbb{E}\Big\{\exp\Big[\tilde{i}t\Big(\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{n}\mathbb{1}\{I_l = i\}W_j(D_i, P(X_i))\Big)\Big]|\Phi_n\Big\} \to \exp(\tilde{i}t)$ a.s, which implies the desired result.

For (2), I first note that $\sum_{i=1}^{n}\mathbb{1}\{\alpha_{j,n}(U_{i,j} - \beta_{j,n}), D_i, X_i\} \rightsquigarrow \sum_{i=1}^{n}\mathbb{1}\{\mathcal{J}_{i,j}, \mathcal{D}_{i,j}, \mathcal{X}_{i,j}\}$ by Lemma E.2(2). Then (2) follows by Proposition 6.3 of Resnick (2007).

For (3),

$$
\begin{aligned}
&\mathbb{E}\exp\left(\tilde{i}(t_1\Psi_{1,n}+t_0\Psi_{0,n})\right)\\
=&\mathbb{E}\exp\left(\sum_{l=1}^{n}\sum_{i=1}^{n}\mathbb{1}\{I_l=i\}\tilde{i}\left(t_1 g_{1,n}(\alpha_{1,n}U_{1.n},X_i,z_1)+t_0 g_{0,n}(\alpha_{0,n}U_{0.n},X_i,z_0)\right)\right)\\
=&\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\exp\Big(\tilde{i}(t_1 D_i g_{1,n}(\alpha_{1,n}U_{1.n},X_i,z_1)+t_0(1-D_i)g_{0,n}(\alpha_{0,n}U_{0.n},X_i,z_0))\Big)\right]^n\\
=&\mathbb{E}\left[1-\frac{1}{n}\sum_{i=1}^{n}\Big(1-\exp(\tilde{i}(t_1 D_i g_{1,n}(\alpha_{1,n}U_{1.n},X_i,z_1)+t_0(1-D_i)g_{0,n}(\alpha_{0,n}U_{0.n},X_i,z_0)))\Big)\right]^n.
\end{aligned}
$$

Conditioning on $D_1=\cdots=D_s=1$ and $D_{s+1}=\cdots=D_n=0$, I have

$$
\begin{aligned}
&\sum_{i=1}^{n}(1-\exp(\tilde{i}(t_1 D_i g_{1,n}(\alpha_{1,n}U_{1.n},X_i,z_1)+t_0(1-D_i)g_{0,n}(\alpha_{0,n}U_{0.n},X_i,z_0))))\\
=&\sum_{i=1}^{s}(1-\exp(\tilde{i}t_1 g_{1,n}(\alpha_{1,n}U'_{1,n},X'_i,z_1)))+\sum_{i=s+1}^{n}(1-\exp(\tilde{i}t_0 g_{0,n}(\alpha_{0,n}U'_{0,n},X'_i,z_0)))\\
=&J_{1,s,n}+J_{0,s,n},
\end{aligned}
\tag{E.29}
$$

in which $(U'_{i,j},X'_{i,j})$ is defined in the proof of Lemma E.2 and $p=P(D=1)$. Then $J_{1,s,n}\perp\!\!\!\perp J_{0,s,n}$ and

$$
\mathbb{E}\exp\left(\tilde{i}(t_1\Psi_{1,n}+t_0\Psi_{0,n})\right)=\sum_{s=0}^{n}C_n^s p^s(1-p)^{n-s}\mathbb{E}\left[1-\frac{1}{n}(J_{1,s,n}+J_{0,s,n})\right]^n.
$$

Similar to the proof of Lemma E.2, it can be shown that $J_{j,s,n}-\int(1-\exp(\tilde{i}t_j g_j))dN'_j=o_p(1)$ uniformly over $|s-np|\le r_n$. Therefore,

$$
\begin{aligned}
&\sum_{s=0}^{n}C_n^s p^s(1-p)^{n-s}\mathbb{E}\left[1-\frac{1}{n}(J_{1,s,n}+J_{0,s,n})\right]^n\\
=&\sum_{|s-np|\le r_n}C_n^s p^s(1-p)^{n-s}\mathbb{E}\left[1-\frac{1}{n}(J_{1,s,n}+J_{0,s,n})\right]^n+o(1)\\
\to&\mathbb{E}\exp\left(-\int(1-\exp(\tilde{i}t_1 g_1))dN'_1-\int(1-\exp(\tilde{i}t_0 g_0))dN'_0\right)\\
=&\mathbb{E}\exp\left(-\int(1-\exp(\tilde{i}t_1 g_1))dN'_1\right)\mathbb{E}\exp\left(-\int(1-\exp(\tilde{i}t_0 g_0))dN'_0\right)\\
=&\mathbb{E}\exp\left(-\int(1-\exp(\tilde{i}t_1 d g_1))dN_1\right)\mathbb{E}\exp\left(-\int(1-\exp(\tilde{i}t_0(1-d)g_0))dN_0\right)\\
=&\mathbb{E}\exp\left(\tilde{i}\int_{E_1}t_1 d g_1 dN_1^*\right)\mathbb{E}\exp\left(\tilde{i}\int_{E_0}t_0(1-d)g_0 dN_0^*\right).
\end{aligned}
$$

In the above derivation, the first line is by the law of iterated logarithm. The second line is by the fact that $J_{j,s,n}-\int(1-\exp(\tilde{i}t_j g_j))dN'_j=o_p(1)$ uniformly over $|s-np|\le r_n$ and then the dominated convergence theorem because $|[1-\frac{1}{n}(J_{1,s,n}+J_{0,s,n})]^n|\leq 1$. The third line is because $J_{1,s,n} \perp\!\!\!\perp J_{0,s,n}$ and thus so are their limits. The fourth line is because, for any $f\in C_K(E_j')$, $\int_{E_j'} fdN_j' = \int_{E_j}(dj+(1-d)(1-j))fdN_j$. The last line is because, for example, for $j=1$ and any $f\in C_K(E_1)$,

$$\begin{aligned}\mathbb{E}\exp\left(\int_{E_1} fdN_1^*\right) =&\mathbb{E}\exp\left(\sum_{i=1}^{\infty}\Gamma_{i,1}f(\mathcal{J}_{i,1},\mathcal{D}_{i,1},\mathcal{X}_{i,1})\right)\\ =&\mathbb{E}\Pi_{i=1}^{\infty}\mathbb{E}\exp(\Gamma_{i,1}f(\mathcal{J}_{i,1},\mathcal{D}_{i,1},\mathcal{X}_{i,1})|\{\mathcal{J}_{i,1},\mathcal{D}_{i,1},\mathcal{X}_{i,1}\}_{i\geq 0})\\ =&\mathbb{E}\Pi_{i=1}^{\infty}\exp(-(1-\exp(f(\mathcal{J}_{i,1},\mathcal{D}_{i,1},\mathcal{X}_{i,1}))))\\ =&\mathbb{E}\exp\left(-\int_{E_1}(1-\exp(f))dN_1\right).\end{aligned} \tag{E.30}$$

For (4), I note that $\hat{Z}_{1,n}^*(k)$ and $\hat{Z}_{0,n}^*(k))$ are also piece-wise constant as $(Z_{1,\infty}(k), Z_{0,\infty}(k))$, that is, when $k-\frac{1}{\hat{P}(X_h^*)} < \sum_{i\neq h}\frac{D_i^*}{\hat{P}(X_i^*)}\mathbb{1}\{\alpha_n U_{i,1}^* < \alpha_n U_{h,1}^*\} < k$ for some h such that $D_h^*=1$, then $\hat{Z}_{1,n}^*(k)=\alpha_n U_{h,1}^*$. And the discontinuity for the sample path occurs at

$$k-\frac{1}{\hat{P}(X_h^*)} = \sum_{i\neq h}\frac{D_i^*}{\hat{P}(X_i^*)}\mathbb{1}\{\alpha_n U_{i,1}^* < \alpha_n U_{h,1}^*\}$$

or

$$k = \sum_{i\neq h}\frac{D_i^*}{\hat{P}(X_i^*)}\mathbb{1}\{\alpha_n U_{i,1}^* < \alpha_n U_{h,1}^*\} = k.$$

W.l.o.g., I assume $\hat{P}(X_i)<1$ for all i. This implies the distances between the two closest discontinuities for the sample paths are $\min_{1\leq i\leq n}\frac{1}{\hat{P}(X_i)}\geq 1$.

E.17 Proof of Lemma E.4

For (1), Let $\tilde{i}$ be the imaginary number. When replacement is allowed,

$$\begin{aligned}&\mathbb{E}\left(\exp\left(\tilde{i}t\left(\frac{1}{b}\sum_{i=1}^{n}\sum_{l=1}^{b}\mathbb{1}\{I_l=i\}W_j(D_i,P(X_i))\right)\right)|\Phi_n\right)\\ =&\left[\mathbb{E}\left(\exp\left(\tilde{i}t\left(\frac{1}{b}\sum_{i=1}^{n}\mathbb{1}\{I_1=i\}W_j(D_i,P(X_i))\right)\right)|\Phi_n\right)\right]^b\\ =&\left[\left(\mathbb{E}\exp\left(\tilde{i}t(\frac{1}{b}\sum_{i=1}^{n}\mathbb{1}\{I_1=i\}W_j(D_i,P(X_i)))\right)|\Phi_n\right)\right]^b\\ =&\left[1-\frac{1}{b}\frac{b}{n}\left(\sum_{l=1}^{n}1-\exp\left(\tilde{i}t(\frac{1}{b}W_j(D_l,P(X_l)))\right)\right)\right]^b.\end{aligned}$$

Because $\frac{b}{n}\left\{\sum_{l=1}^{n}\left[1-\exp\left(\tilde{i}t(\frac{1}{b}W_j(D_l,P(X_l)))\right)\right]\right\}\to\tilde{i}t$ as $b,n\to\infty$ a.s., the characteristic function converges to $\exp(\tilde{i}t)$. This implies that $\frac{1}{b}\sum_{i=1}^{n}\sum_{l=1}^{m}\mathbb{1}\{I_l=i\}W_j(D_i,P(X_i))\to 1$ a.s.

When replacement is not allowed,

$$\frac{1}{b}\sum_{i=1}^{n} P_{n,i}W_j(D_i,P(X_i)) = \frac{1}{b}\sum_{i=1}^{n}(P_{n,i}-\frac{b}{n})W_j(D_i,P(X_i)) + \frac{1}{n}\sum_{i=1}^{n}W_j(D_i,P(X_i)). \tag{E.31}$$

The second term of (E.31) converges to 1 almost surely by SLLN. For the first term of (E.31), W_j is bounded and $\mathbb{E}(\frac{1}{b}\sum_{i=1}^{n}(P_{n,i}-\frac{b}{n})W_j(D_i,P(X_i)))^2 \lesssim \frac{1}{b}+\frac{1}{n} \to 0$. This concludes part (1).

For part (2), $\mathbb{E}P_{n,i} = \frac{b}{n}$ and $\hat{N}_j := \sum_{i=1}^{n}\mathbb{1}\{\alpha_{j,n}(U_{i,j}-\beta_{j,n}), X_i, D_i\} \rightsquigarrow N_j$. By Proposition 6.2 of Resnick (2007), for $\hat{N}_j^*$ and N_j as random element in the space of point measure,

$$P(\hat{N}_j^* \in .|\{\alpha_{j,n}(U_{i,j}-\beta_{j,n}), X_i, D_i\}_{i=1}^{n}) \xrightarrow{p} P(N_j \in .).$$

Taking expectation on both sides, I obtain $\hat{N}_j^* \rightsquigarrow N_j$.

For part (3), I first denote $(U'_{i,j}, X'_{i,j})$ as is defined in the proof of Lemma E.2 and $p = P(D=1)$. When replacement is allowed,

$$\begin{aligned}
&\mathbb{E}\exp(\tilde{i}(t_1\Psi_{1,n}+t_0\Psi_{0,n}))\\
=&\mathbb{E}\exp\left(\tilde{i}\left(\sum_{l=1}^{b}\sum_{i=1}^{n}\mathbb{1}\{I_l=i\}(t_1D_ig_1(\alpha_{1,b}(U_{i,1}-\beta_{1,b}),X_i)+t_0(1-D_i)g_0(\alpha_{0,b}(U_{i,0}-\beta_{0,b}),X_i))\right)\right)\\
=&\mathbb{E}\left[1-\frac{1}{b}\left(\frac{b}{n}\sum_{i=1}^{n}\left(1-\exp(\tilde{i}(t_1D_ig_1(\alpha_{1,b}(U_{i,1}-\beta_{1,b}),X_i)+t_0(1-D_i)g_0(\alpha_{0,b}(U_{i,0}-\beta_{0,b}),X_i)))\right)\right)\right]^b\\
=&\sum_{s=0}^{n}C_n^sp^s(1-p)^{n-s}\mathbb{E}\Bigg\{1-\frac{1}{b}\Bigg[\frac{b}{n}\sum_{i=1}^{s}\left(1-\exp(\tilde{i}t_1g_1(\alpha_{1,b}(U'_{i,1}-\beta_{1,b}),X'_i))\right)\\
&+\frac{b}{n}\sum_{i=s+1}^{n}(1-\exp(\tilde{i}t_0g_0(\alpha_{0,b}(U'_{i,0}-\beta_{0,b}),X'_i)))\Bigg]\Bigg\}^b.
\end{aligned} \tag{E.32}$$

For $s=[np]$, $\mathbb{E}\frac{b}{n}\sum_{i=1}^{[np]}\mathbb{1}\{(\alpha_{1,b}(U'_{i,1}-\beta_{1,b}),X'_i)\in .\}\to\mu'_1(.)$ and $\mathbb{E}\frac{b}{n}\sum_{i=[np]+1}^{n}\mathbb{1}\{(\alpha_{0,b}(U'_{i,0}-\beta_{0,b}),X'_i)\in .\}\to\mu'_0(.)$, where μ'_j is defined as the mean measure of N'_j and N'_j is defined in Lemma E.2. Then by Theorem 5.3 of Resnick (2007), $\frac{b}{n}\sum_{i=1}^{[np]}\mathbb{1}\{(\alpha_{1,b}(U'_{i,1}-\beta_{1,b}),X'_i)\}\rightsquigarrow N'_j$ as $\frac{b}{n}\to 0$. By the same argument in the proof of (3) of Lemma E.2, I can show that this convergence is uniform over $|s-np|\le r_n$. Therefore, uniformly over $|s-np|\le r_n$,

$$\frac{b}{n}\sum_{i=1}^{s}\left(1-\exp(\tilde{i}(t_1g_1(\alpha_{1,b}(U'_{i,1}-\beta_{1,b}),X'_i)))\right)\xrightarrow{p}\int_{E'_1}\left[1-\exp(\tilde{i}(t_1g_1(u,x)))\right]d\mu'_1,$$

and

$$\frac{b}{n}\sum_{i=s+1}^{n}\left(1-\exp(\tilde{i}(t_0g_0(\alpha_{0,b}(U'_{i,0}-\beta_{0,b}),X'_i)))\right)\xrightarrow{p}\int_{E'_0}\left[1-\exp(\tilde{i}(t_0g_0(u,x)))\right]d\mu'_0.$$

Since the term inside the expectation of the RHS of (E.32) is bounded by 1, by the dominated convergence theorem, the RHS of (E.32) converges to

$$
\begin{aligned}
&\exp\left\{\int_{E_1'}\left[1-\exp(\tilde{i}(t_1 g_1(u,x)))\right]d\mu_1' + \int_{E_0'}\left[1-\exp(\tilde{i}(t_0 g_0(u,x)))\right]d\mu_0\right\} \\
=&\exp\left\{\int_{E_1}\left[1-\exp(\tilde{i}(t_1 d g_1(u,x)))\right]d\mu_1 + \int_{E_0}\left[1-\exp(\tilde{i}(t_0(1-d) g_0(u,x)))\right]d\mu_0\right\} \\
=&\mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1} dg_1 dN_1\right)\mathbb{E}\exp\left(\tilde{i}t_0\int_{E_0} dg_0 dN_0\right),
\end{aligned}
$$

in which the first equality is by the relation between μ_j and μ_j' and the second equality is by the definition of Laplace functional of Poisson random measure with mean measure μ_j.

If replacement is not allowed, then by the exchangeablity of the weights $P_{n,i}$,

$$
\begin{aligned}
&\mathbb{E}\exp(\tilde{i}(t_1\Psi_{1,n}+t_0\Psi_{0,n})) \\
=&\mathbb{E}\exp\left(\tilde{i}(t_1\sum_{i=1}^{b} D_i g_1(\alpha_{1,b}(U_{i,1}-\beta_{1,b}),X_i)+t_0\sum_{i=1}^{b}(1-D_i)g_0(\alpha_{0,b}(U_{i,0}-\beta_{0,b}),X_i))\right) \\
=&\mathbb{E}\exp\left(\tilde{i}t_1\int_{E_1} dg_1 dN_1\right)\mathbb{E}\exp\left(\tilde{i}t_0\int_{E_0} dg_0 dN_0\right),
\end{aligned}
$$

in which the second equality is by the same argument in the proof of (3) in Lemma E.2 with n is replaced by b.

(4) holds for the same reason as in the proof of (4) in Lemma E.3.

E.18 Tightness, uniqueness and continuity

Lemma E.5. *$Z_{j,\infty}(k)$, $j=0,1$ are tight.*

Proof. Here I focus on the case for $j=1$. The proof follows the proof of Lemma 9.7 in Chernozhukov (2005). The difference is that $l_\delta(u,v)$ is reweighted by the inverse propensity score $\frac{d}{P(x)}$.

First, note that the limiting objective function is $\mathcal{Q}_{1,\infty}(z_1,k)=-kz_1+\int_{\mathbb{E}}\frac{d}{P(x)}(z_1-j)^+dN_1(j,d,x)$ when $j>-\delta$. I can choose z^f such that $-kz^f+\int_{\mathbb{E}}\frac{d}{P(x)}(z^f-j)^+dN_1(j,d,x)=O_p(1)$. Let $z^*=z^f+Mv$, where $v=\pm 1$. Then by the convexity of objective function in z and the argument between Equation (9.74) and (9.75) of Chernozhukov (2005), I only need to show that, for any K and $\varepsilon>0$, there is an M large enough such that

$$
P(\min_{v=\pm 1}\mathcal{Q}_{1,\infty}(z^*)>K)\geq 1-\varepsilon. \tag{E.33}
$$

The claim holds trivially when $v=-1$. For $v=1$, first note that $P(x)\leq 1-c$. When Y_1 has the type 1 or 3 tail,

$$\begin{aligned}
&\int_{\mathbb{E}} \frac{d}{P(x)}(z^f + M - j)^+ dN_1(j,d,x) \\
\geq &\int_{[0,\kappa]\times\{1\}\times \text{Supp}(\mathcal{X})} \frac{d}{P(x)}(z^f + M - j)^+ dN_1(j,d,x) \\
\geq &N([0,\kappa]\times\{1\}\times \text{Supp}(\mathcal{X}))\frac{(z^f + M - \kappa)^+}{1-c}.
\end{aligned}$$

Because $N([0,\kappa]\times\{1\}\times \text{Supp}(\mathcal{X}))$ is a Poisson random variable with mean $\int P(x)P_1^+(dx|Y = q_1(0))h(\kappa) \to \infty$ as $\kappa \to \infty$. For $\kappa \to \infty$, $N([0,\kappa]\times\{1\}\times \text{Supp}(X)) > (k+1)(1-c)$ with probability greater of equal to $1-\varepsilon$.

When Y_1 has type 2 tail, I have, for any $\kappa < 0$,

$$\begin{aligned}
&\int_{\mathbb{E}} \frac{d}{P(x)}(z^f + M - j)^+ dN_1(j,d,x) \\
\geq &\int_{[-\infty,\kappa]\times\{1\}\times \text{Supp}(\mathcal{X})} \frac{d}{P(x)}(z^f + M - j)^+ dN_1(j,d,x) \\
\geq &N([-\infty,\kappa]\times\{1\}\times \text{Supp}(\mathcal{X}))\frac{(z^f + M - \kappa)^+}{1-c}.
\end{aligned}$$

Then similarly, $N([-\infty,\kappa]\times\{1\}\times \text{Supp}(\mathcal{X}))$ is a Poisson random variable with mean $\int P(x)P_1^+(dx|Y = q_1(0))h(\kappa) \to \infty$ as $\kappa \to 0$. For $\kappa \to 0$, $N([-\infty,\kappa]\times\{1\}\times \text{Supp}(X)) > (k+1)(1-c)$ with probability greater of equal to $1-\varepsilon$

So by letting M be large enough, with probability greater or equal to $1-\varepsilon$, I have

$$\begin{aligned}
\mathcal{Q}_{1,\infty}(z^*, k) &= -kz^f - kM + \int_{\mathbb{E}} \frac{d}{P(x)}(z^f + M - j)^+ dN_1(j,d,x) \\
&\geq -kz^f - kM + (z^f + M - \kappa)^+(k+1) > K.
\end{aligned}$$

This verifies (E.33). ☐

Lemma E.6. *Let $\mathcal{M}(l)$ be the set of l-element subsets of $\mathcal{N} = \{1, 2, \cdots\}$. For $j = 0, 1$, the sequence $(\mathcal{D}_i, \mathcal{X}_{i,j})$ are i.i.d such that $\mathcal{D}_i$ is Bernoulli distributed with success probability $P(\mathcal{X}_{i,j})$ and $\mathcal{X}_{i,j}$ has law $P_j^+(X \in .|Y_j = q_j(0))$. If $P(\sum_{i\in\mu}\frac{1}{P(\mathcal{X}_{i,1})} = k) = 0$, $P(\sum_{i\in\mu}\frac{1}{P(\mathcal{X}_{i,1})} + \frac{1}{P(\mathcal{X}_{h,1})} = k) = 0$, $P(\sum_{i\in\mu}\frac{1}{1-P(\mathcal{X}_{i,0})} = k) = 0$, and $P(\sum_{i\in\mu}\frac{1}{1-P(\mathcal{X}_{i,0})} + \frac{1}{1-P(\mathcal{X}_{h,0})} = k) = 0$, for any h and $\mu \in \mathcal{M}(l)$, $l \leq h-1$, then both $Z_{1,\infty}(k)$ and $Z_{0,\infty}(k)$ are unique minimizers a.s.*

Proof. Here I focus on the case for $j = 1$. Following the notation in Theorem 4.1, $\mathcal{J}_i = h_1^{-1}(\sum_{l=1}^{i} E_l)$. By Proposition 6.1 of Koenker (2005) and Lemma E.5, $Z_{1,\infty}(k) = \mathcal{J}_h$ for some h such that $T_h = 1$. Then by taking directional derivative of the objective function,

$$k - \frac{1}{P(\mathcal{X}_h)} \leq \sum_{i\neq h} \frac{\mathcal{D}_i}{P(\mathcal{X}_i)} \mathbb{1}\{\mathcal{J}_i < \mathcal{J}_h\} \leq k. \tag{E.34}$$

Since $\mathcal{J}_i$ is monotone increasing,

$$\begin{aligned} & P\left(\sum_{i\neq h}\frac{\mathcal{D}_i}{P(\mathcal{X}_i)}\mathbb{1}\{\mathcal{J}_i<\mathcal{J}_h\}=k\right) \\ \leq & \sum_{l\leq h-1,\mu\in\mathcal{M}(l),h} P\left(\sum_{i\in\mu}\frac{1}{P(\mathcal{X}_i)}=k\right) \\ =&0. \end{aligned} \tag{E.35}$$

Similarly, $P(\sum_{i\neq h}\frac{\mathcal{D}_i}{P(\mathcal{X}_i)}\mathbb{1}\{J_i<J_h\}+\frac{1}{P(\mathcal{X}_h)}=k)=0$. Therefore, the inequality (E.34) holds strictly. This implies $Z_{1,\infty}(k)$ is the unique minimizer. □

Lemma E.7. *$Z_{j,\infty}(k)$ is continuous for any k and $j=0,1$. If $k'(m-1)>\frac{1}{\inf_{x\in Supp(\mathcal{X})}P(x)}$ and $k'(m-1)>\frac{1}{\inf_{x\in Supp(\mathcal{X})}(1-P(x))}$, then*

$$\sqrt{k'}\frac{Z_{j,\infty}(k)+c}{\max(Z_{1,\infty}(mk')-Z_{1,\infty}(k'),\tilde{\rho}(Z_{0,\infty}(mk')-Z_{0,\infty}(k')))}$$

is also continuous for $j=0,1$.

Proof. $Z_{1,\infty}(k)=J_h$ for some h with $T_h=1$. Because J_h is continuous, $P(Z^*_{1,\infty}(k)=z)=\sum_h P(J_h=z)=0$. Therefore, $Z_{1,\infty}(k)$ is continuous. Similarly, $Z_{0,\infty}(k)$ is also continuous. Assume h_1 and h_2 solve the following two first order conditions:

$$k'-\frac{1}{P(\mathcal{X}_{h_1})}\leq\sum_{i\neq h_1}\frac{D_i}{P(\mathcal{X}_i)}\mathbb{1}\{\mathcal{J}_i<\mathcal{J}_{h_1}\}\leq k',$$

$$mk'-\frac{1}{P(\mathcal{X}_{h_2})}\leq\sum_{i\neq h_2}\frac{D_i}{P(\mathcal{X}_i)}\mathbb{1}\{\mathcal{J}_i<\mathcal{J}_{h_2}\}\leq mk'.$$

Then $h_1=h_2=h$ implies $(m-1)k'\leq\frac{1}{P(\mathcal{X}_h)}$ for some $\mathcal{X}_h\in\text{Supp}(\mathcal{X})$. However, the imposed condition rules out this situation. Thus $h_1\neq h_2$ and $Z^*_{j,\infty}(mk')\neq Z^*_{j,\infty}(k')$. In fact, following the same argument in step 3 of proof of Lemma E.1 in Chernozhukov and Fernández-Val (2011), I can prove that $Z_{j,\infty}(mk')-Z_{j,\infty}(k')>0$, $j=0,1$. Last, noting that function $1/\max(u,v)$ is continuous on $(u,v)\in\mathbb{R}^+\times\mathbb{R}^+$, I have proved the stated result. □

Next, I aim to show $\sup_{k\in[\kappa_1,\kappa_2]}|Z^c_\infty(k)/\sigma(k)|$ is continuous. Recall the definition of $J_{1,i}$ and $J_{0,i}$ in Theorem 4.1. I rely on the next technical assumption to derive the result.

Assumption 19. *If $\tilde{\rho}\in(0,\infty)$, for any pair of positive integers (h_0,h_1), $\left|\frac{\mathcal{J}_{h_1,1}-\tilde{\rho}\mathcal{J}_{h_0,0}-(\eta_1(k)-\tilde{\rho}\eta_0(k))}{\sigma(k)}\right|$ has at most L local extremum which are denoted as $\{k^*_l(\mathcal{J}_{h_1,1}-\tilde{\rho}\mathcal{J}_{h_0,0})\}_{l=1}^L$ for some finite integer L. Furthermore, the following two conditions holds:*

1. *$k^*_l(\mathcal{J}_{h_1,1}-\tilde{\rho}\mathcal{J}_{h_0,0})$ is continuously distributed for $l=1,\cdots,L$.*

2. For any positive integers (h_0, h_1), any z, and any $l = 1, \cdots, L$,

$$P\left(\left|\frac{\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho}\eta_0(k))}{\sigma(k)}\right| = z | k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0}) = k\right) = 0$$

for almost all $k \in [\kappa_1, \kappa_2]$.

If $\tilde{\rho} = 0$, for any pair of positive integers (h_0, h_1), $|\frac{\mathcal{J}_{h_1,1} - \eta_1(k)}{\sigma(k)}|$ has at most L local extremum which are denoted as $\{k_l^*(\mathcal{J}_{h_1,1})\}_{l=1}^L$ for some finite integer L. Furthermore, the following two conditions holds:

1. $k_l^*(\mathcal{J}_{h_1,1})$ is continuously distributed for $l = 1, \cdots, L$.

2. For any positive integers (h_0, h_1), any z, and any $l = 1, \cdots, L$,

$$P\left(\left|\frac{\mathcal{J}_{h_1,1} - \eta_1(k)}{\sigma(k)}\right| = z | k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0}) = k\right) = 0$$

for almost all $k \in [\kappa_1, \kappa_2]$.

If $\tilde{\rho} = \infty$, for any pair of positive integers (h_0, h_1), $|\frac{\mathcal{J}_{h_0,0} - \eta_0(k)}{\sigma(k)}|$ has at most L local extremum which are denoted as $\{k_l^*(\mathcal{J}_{h_0,0})\}_{l=1}^L$ for some finite integer L. Furthermore, the following two conditions holds:

1. $k_l^*(\mathcal{J}_{h_0,0})$ is continuously distributed for $l = 1, \cdots, L$.

2. For any positive integers (h_0, h_1), any z, and any $l = 1, \cdots, L$,

$$P\left(\left|\frac{\mathcal{J}_{h_0,0} - \eta_0(k)}{\sigma(k)}\right| = z | k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0}) = k\right) = 0$$

for almost all $k \in [\kappa_1, \kappa_2]$.

This assumption is mild. For example, if $\sigma(k) := 1$, the assumption holds automatically. To see this, note that $\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0}$ is continuously distributed and $k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0})$ does not depends on $\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0}$, that is, it is deterministic.

Lemma E.8. *κ_1 and κ_2 are not in the discontinuity of either $Z_{1,\infty}(k)$ and $Z_{0,\infty}(k)$, and Assumption 19 holds. If $\tilde{\rho} \in (0, \infty)$, then*

$$\sup_{k \in [\kappa_1, \kappa_2]} |(Z_{1,\infty}^c(k) - \tilde{\rho}Z_{0,\infty}^c(k))/\sigma(k)|$$

is continuous.

If $\tilde{\rho} = 0$, then

$$\sup_{k \in [\kappa_1, \kappa_2]} |Z_{1,\infty}^c(k)/\sigma(k)|$$

is continuous.

If $\tilde{\rho} \in (0, \infty)$, *then*

$$\sup_{k \in [\kappa_1, \kappa_2]} |Z^c_{0,\infty}(k)/\sigma(k)|$$

is continuous.

If $k'(m-1) > \frac{1}{\inf_{x \in Supp(\mathcal{X})} P(x)}$ *and* $k'(m-1) > \frac{1}{\inf_{x \in Supp(\mathcal{X})}(1-P(x))}$, *then*

$$\sup_{k \in [\kappa_1, \kappa_2]} |Z^c_{\infty}(k)/\sigma(k)| = \sup_{k \in [\kappa_1, \kappa_2]} \left| \frac{\sqrt{k'}}{\sigma(k)} \frac{Z^c_{1,\infty}(k) - \tilde{\rho} Z^c_{0,\infty}(k)}{\max(Z_{1,\infty}(mk') - Z_{1,\infty}(k'), \tilde{\rho}(Z_{0,\infty}(mk') - Z^*_{0,\infty}(k')))} \right|$$

is also continuous.

Proof. I only consider the case for $\tilde{\rho} \in (0, \infty)$. The other two cases can be proved similarly. Let $\mathcal{L}_{h,1} = \{k : D_h = 1, k = \sum_{i<h} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1})$ or $k = \sum_{i \leq h} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1})\}$ and $\mathcal{L}_{h,0} = \{k : 1 - D_h = 1, k = \sum_{i<h} W_0(\mathcal{D}_{i,0}, \mathcal{X}_{i,0})$ or $k = \sum_{i \leq h} W_0(\mathcal{D}_{i,0}, \mathcal{X}_{i,0})\}$. Then the discontinuities for the sample path of $Z_{j,\infty}(k)$ is $\cup_{h \geq 1} \mathcal{L}_{j,h}$. Since the closest distance between two distinct discontinuities of $Z_{j,\infty}(k)$ is at least 1, there are at most finite number of discontinuities of either $Z_{1,\infty}(k)$ or $Z_{0,\infty}(k)$. This implies the closest distance between two distinct discontinuities of $Z^c_{\infty}(k)$ is strictly positive. For a fixed event ω, if $\sup_{k \in [\kappa_1, \kappa_2]} |Z^c_{\infty}(k)(\omega)| = z$, then there exists a convergent sequence $\hat{k}_m(\omega)$[21] with limit $\hat{k}(\omega)$ such that $|Z^c_{\infty}(\hat{k}_m(\omega))(\omega)| \to z$. Since $Z_{j,\infty}(k)$ is piece-wise constant, κ_1 and κ_2 are not in $\cup_{j=0,1} \cup_{h \geq 1} \mathcal{L}_{j,h}$, there exist $M(\omega)$ large enough such that for $m > M(\omega)$,

$$\begin{aligned} z = & \sup_{k \in [\kappa_1, \kappa_2]} |Z^c_{\infty}(k)(\omega)/\sigma(k)| \\ = & |(Z_{1,\infty}(\hat{k}_m) - \tilde{\rho} Z_{0,\infty}(\hat{k}_m) - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| \\ = & |(\mathcal{J}_{\hat{h}_1,1} - \tilde{\rho}\mathcal{J}_{\hat{h}_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})|, \end{aligned}$$

in which $\hat{k}_m - \frac{1}{P(\mathcal{X}_{\hat{h}_1,1})} < \sum_{i<\hat{h}_1} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1}) < \hat{k}_m$, $\hat{k}_m - \frac{1}{1-P(\mathcal{X}_{\hat{h}_0,0})} < \sum_{i<\hat{h}_0} W_0(\mathcal{D}_{i,0}, \mathcal{X}_{i,0}) < \hat{k}_m$, and $\hat{k} \in \mathcal{L}(\hat{h}_1, \hat{h}_0) := \mathcal{L}_{\hat{h}_1,1} \cup \mathcal{L}_{\hat{h}_0,0} \cup \{k^*_l(\mathcal{J}_{\hat{h}_1,1} - \tilde{\rho}\mathcal{J}_{\hat{h}_0,0})\}_{l=1}^{L} \cup \{\kappa_1\} \cup \{\kappa_2\}$. Furthermore, let $\mathcal{A}_h = \{\sum_{i \leq h} \mathcal{D}_{i,1} > \kappa_2, \sum_{i \leq h}(1 - \mathcal{D}_{i,0}) > \kappa_2\}$. Then on $\mathcal{A}_h$, $\hat{h}_j \leq h$ for $j = 0, 1$. Therefore,

$$\begin{aligned} & P\bigg(\sup_{k \in [\kappa_1, \kappa_2]} |Z^c_{\infty}(k)(\omega)/\sigma(k)| = z \bigg) \\ \leq & \sum_{h > \kappa_2} P\bigg(|(\mathcal{J}_{\hat{h}_1,1} - \tilde{\rho}\mathcal{J}_{\hat{h}_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} \in \mathcal{L}(\hat{h}_1, \hat{h}_0), \mathcal{A}_h \bigg) \\ \leq & \sum_{h > \kappa_2} \sum_{h_1 \leq h, h_0 \leq h} P\bigg(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} \in \mathcal{L}(h_1, h_0) \bigg) \end{aligned} \tag{E.36}$$

[21] $\hat{k}_m(\omega)$ depends on the sample path and thus is random.

In order to bound the last equation, I note that $\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho}\eta_0(k))$ is continuously distributed, $\mathcal{J}_{i,j}$ is independent of $(\mathcal{D}_{i,j}, \mathcal{X}_{i,j})$ for any realization (h_1, h_0) of $(\hat{h}_1, \hat{h}_0)$, and $(\mathcal{J}_{h_1,1}, \mathcal{J}_{h_0,0}) \perp\!\!\!\perp \mathcal{L}(h_1, h_0)$. Hence, if $\hat{k} \in \mathcal{L}_{h_1,1}$ and for instance, $\hat{k} = \sum_{i<h_1} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1})$, I have

$$\begin{aligned}
&P\left(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} = \sum_{i<h_1} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1})\right) \\
\leq &\int P\left(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho}\eta_0(k)))/\sigma(k)| = z | \sum_{i<h_1} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1}) = k\right) \\
&\times dP\left(\sum_{i<h_1} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1}) \leq k\right) \\
= &\int P\left(|\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho}\eta_0(k))| = z\right) dP\left(\sum_{i<h_1} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1}) \leq k\right) \\
= &0.
\end{aligned}$$

Similarly, if $\hat{k} \in \mathcal{L}_{h_0,0}$ and $\hat{k} = \sum_{i \leq h_1} W_1(\mathcal{D}_{i,1}, \mathcal{X}_{i,1})$,

$$P\left(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} = \sum_{i \leq h_1} W_0(\mathcal{D}_{i,0}, \mathcal{X}_{i,0})\right) = 0.$$

If $\hat{k} \in \{k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0})\}_{l=1}^L$,

$$\begin{aligned}
&P\left(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} \in \{k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0})\}_{l=1}^L\right) \\
\leq &\sum_{l=1}^L \int_{\kappa_1}^{\kappa_2} P\left(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho}\eta_0(k)))/\sigma(k)| = z | k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0}) = k\right) \\
&\times dP\left(k_l^*(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0}) \leq k\right) = 0.
\end{aligned}$$

Last, if $\hat{k} = \kappa_1$ or κ_2,

$$P\left(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} = \kappa_1 \text{ or } \kappa_2\right) = 0.$$

To sum up, I have

$$P\left(|(\mathcal{J}_{h_1,1} - \tilde{\rho}\mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} \in \mathcal{L}(\hat{h}_1, \hat{h}_0)\right) = 0.$$

Then by (E.36), I have, for any $z \in \Re$,

$$P\left(\sup_{k \in [\kappa_1, \kappa_2]} |Z_\infty^c(k)(\omega)/\sigma(k)| = z\right) = 0.$$

This means $\sup_{k\in[\kappa_1,\kappa_2]}|Z_\infty^c(k)(\omega)/\sigma(k)|$ is continuously distributed. The second result can be proved in a same manner as in Lemma E.7. $\square$

F Additional simulation results

F.1 Simulation results with $n = 300$

Tables 9 and 10 report the coverage of BN-CI and NN-CI as well as their corresponding median lengths. I am interested in the QTE at quantile order $k = (5, 10, 20, 40)$. In this case, the corresponding quantile indices are $\tau_n = (0.017, 0.033, 0.067, 0.133)$. Y_1 and Y_0 have four different conditional boundary structures: (1) single minimizer, (2) multiple minimizers, (3) continuum minimizers, and (4) mixture minimizers. When reading the table, the row indicates the potential outcome Y_1 while the column indicates the potential outcome Y_0. The detail of each model can be found in Appendix B. The subsample size used to compute Table 9 and Figure 13 is 120. Figure 13 shows the evolution of the BN-CI coverage over $k \in [5, 40]$. In all cases, the coverage before the cutoff line $k = \min(40, \frac{0.2b}{m})$ is close to the nominal rate. Figure 14 shows that the evolution of BN-CI's coverage against subsample size b is stable.

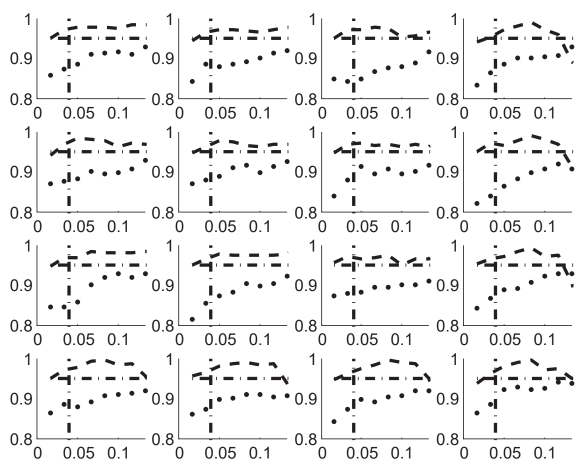
τ_n = 0.001, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.002, k = 10	(1)	(2)	(3)	(4)
(1)	0.949	0.942	0.948	0.939	(1)	0.971	0.964	0.972	0.952
	(0.176)	(0.167)	(0.152)	(0.169)		(0.186)	(0.174)	(0.160)	(0.187)
(2)	0.940	0.947	0.947	0.948	(2)	0.967	0.961	0.969	0.972
	(0.155)	(0.140)	(0.116)	(0.166)		(0.162)	(0.147)	(0.126)	(0.184)
(3)	0.946	0.950	0.955	0.952	(3)	0.967	0.964	0.970	0.964
	(0.135)	(0.122)	(0.061)	(0.106)		(0.138)	(0.127)	(0.069)	(0.118)
(4)	0.950	0.954	0.947	0.937	(4)	0.970	0.966	0.962	0.961
	(0.185)	(0.177)	(0.171)	(0.165)		(0.205)	(0.200)	(0.191)	(0.186)
τ_n = 0.004, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.008, k = 40	(1)	(2)	(3)	(4)
(1)	0.978	0.971	0.976	0.981	(1)	0.983	0.978	0.965	0.891
	(0.229)	(0.223)	(0.208)	(0.281)		(0.193)	(0.185)	(0.166)	(0.328)
(2)	0.980	0.974	0.964	0.976	(2)	0.968	0.968	0.963	0.912
	(0.202)	(0.185)	(0.165)	(0.282)		(0.164)	(0.163)	(0.137)	(0.327)
(3)	0.982	0.975	0.967	0.982	(3)	0.983	0.978	0.966	0.903
	(0.173)	(0.166)	(0.098)	(0.198)		(0.156)	(0.145)	(0.089)	(0.249)
(4)	0.992	0.987	0.984	0.989	(4)	0.955	0.938	0.948	0.949
	(0.362)	(0.354)	(0.347)	(0.348)		(0.401)	(0.399)	(0.389)	(0.274)

Table 9: Coverage of 95% b out of n bootstrap CI, sample size = 300

τ_n = 0.001, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.002, k = 10	(1)	(2)	(3)	(4)
(1)	0.858	0.841	0.847	0.833	(1)	0.872	0.884	0.842	0.864
	(0.119)	(0.112)	(0.096)	(0.118)		(0.137)	(0.129)	(0.117)	(0.140)
(2)	0.868	0.868	0.837	0.820	(2)	0.874	0.877	0.878	0.840
	(0.104)	(0.096)	(0.076)	(0.107)		(0.115)	(0.110)	(0.092)	(0.130)
(3)	0.846	0.814	0.871	0.842	(3)	0.844	0.855	0.879	0.866
	(0.085)	(0.077)	(0.041)	(0.072)		(0.097)	(0.088)	(0.051)	(0.089)
(4)	0.864	0.861	0.841	0.863	(4)	0.884	0.872	0.871	0.886
	(0.118)	(0.109)	(0.108)	(0.117)		(0.142)	(0.137)	(0.136)	(0.147)

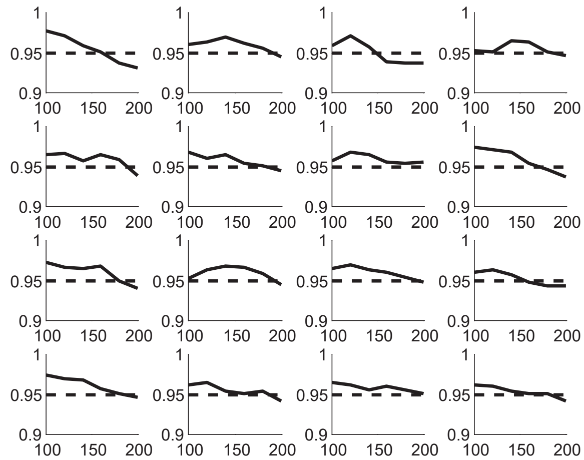
τ_n = 0.004, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.008, k = 40	(1)	(2)	(3)	(4)
(1)	0.908	0.885	0.867	0.901	(1)	0.929	0.919	0.915	0.927
	(0.159)	(0.152)	(0.139)	(0.169)		(0.187)	(0.180)	(0.168)	(0.218)
(2)	0.901	0.908	0.894	0.881	(2)	0.928	0.924	0.916	0.907
	(0.131)	(0.128)	(0.112)	(0.162)		(0.156)	(0.155)	(0.140)	(0.214)
(3)	0.901	0.881	0.893	0.892	(3)	0.927	0.921	0.909	0.927
	(0.110)	(0.101)	(0.066)	(0.113)		(0.129)	(0.124)	(0.088)	(0.159)
(4)	0.892	0.901	0.892	0.928	(4)	0.917	0.905	0.919	0.938
	(0.186)	(0.185)	(0.185)	(0.200)		(0.274)	(0.277)	(0.280)	(0.305)

Table 10: Coverage of 95% n out of n bootstrap CI, sample size = 300



Each (i, j)-th subplot represents the (i, j)-th model. The dashed line is the coverage of BN-CI with $b = 120$ and $n = 300$ for quantile index $\tau \in [1.67\%, 16.67\%]$. The dotted line is the coverage of NN-CI. The horizontal dotted dashed line is the 95% nominal coverage rate, and the vertical dotted dashed line is $\tau = \min(\frac{40}{n}, \frac{0.2b}{mn})$.

Figure 13: Coverage across quantiles



Each (i, j)-th subplot represents the (i, j)-th model. The solid line is the coverage for b out of n bootstrap CI at $k = 10$ in which $b \in [100, 200]$.

Figure 14: Coverage across subsample size

$\tau_n = 0.017$, k = 5	(1)	(2)	(3)	(4)	$\tau_n = 0.033$, k = 10	(1)	(2)	(3)	(4)
(1)	0.372	0.132	0.376	-0.138	(1)	0.195	0.113	0.288	-0.199
(2)	0.315	0.154	0.049	-0.058	(2)	0.149	0.120	0.122	-0.091
(3)	-0.109	0.139	0.011	-0.127	(3)	-0.129	-0.017	0.033	-0.177
(4)	0.198	0.011	0.100	0.086	(4)	0.028	0.177	0.060	-0.006
$\tau_n = 0.067$, k = 20	(1)	(2)	(3)	(4)	$\tau_n = 0.133$, k = 40	(1)	(2)	(3)	(4)
(1)	-0.024	0.024	0.139	-0.075	(1)	-0.168	-0.096	0.136	-0.102
(2)	0.079	0.182	0.113	-0.089	(2)	-0.125	0.081	0.136	-0.127
(3)	-0.154	-0.087	0.038	-0.149	(3)	-0.021	-0.150	-0.013	-0.098
(4)	-0.163	0.030	0.055	-0.033	(4)	-0.460	-0.154	-0.005	1.487

Table 11: Bias of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

τ_n = 0.017, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.033, k = 10	(1)	(2)	(3)	(4)
(1)	2.996	2.992	2.588	3.047	(1)	3.194	3.218	2.868	3.462
(2)	2.608	2.421	1.911	2.964	(2)	2.748	2.770	2.351	3.226
(3)	2.260	1.998	0.995	1.899	(3)	2.397	2.141	1.222	2.212
(4)	2.754	2.672	2.695	2.890	(4)	3.253	3.291	3.301	3.598
τ_n = 0.067, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.133, k = 40	(1)	(2)	(3)	(4)
(1)	3.678	3.691	3.426	4.163	(1)	4.057	4.037	4.035	5.375
(2)	3.120	3.062	2.725	3.974	(2)	3.494	3.450	3.238	5.299
(3)	2.714	2.433	1.527	2.682	(3)	3.002	2.795	2.010	3.935
(4)	4.053	4.298	4.156	4.511	(4)	5.819	5.951	6.107	8.230

Table 12: root-MSE of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

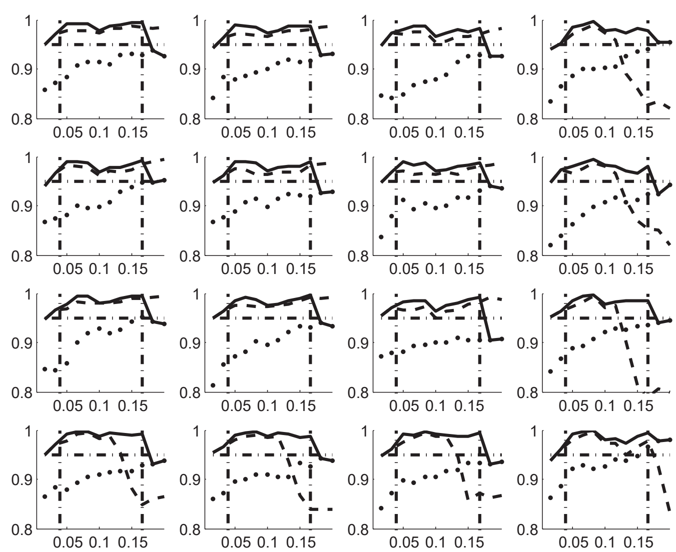
τ_n = 0.017, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.033, k = 10	(1)	(2)	(3)	(4)
(1)	0.262	-0.005	0.217	-0.044	(1)	0.138	-0.060	0.155	0.009
(2)	0.322	0.235	-0.040	0.156	(2)	0.155	0.021	0.055	0.103
(3)	-0.012	0.248	-0.010	-0.079	(3)	-0.079	0.045	0.016	-0.124
(4)	0.082	-0.013	0.010	0.028	(4)	-0.136	-0.036	-0.224	-0.160
τ_n = 0.067, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.133, k = 40	(1)	(2)	(3)	(4)
(1)	-0.192	-0.072	-0.096	0.239	(1)	-0.164	-0.271	0.098	0.228
(2)	-0.061	0.144	-0.057	-0.037	(2)	-0.092	-0.076	0.066	0.271
(3)	-0.144	-0.089	-0.007	-0.082	(3)	-0.041	-0.044	-0.010	0.010
(4)	-0.397	-0.229	-0.122	-0.231	(4)	-0.553	-0.510	-0.196	1.542

Table 13: median-bias of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

τ_n = 0.017, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.033, k = 10	(1)	(2)	(3)	(4)
(1)	1.763	1.883	1.512	2.030	(1)	2.021	2.218	1.817	2.368
(2)	1.687	1.633	1.213	1.980	(2)	1.766	1.838	1.520	2.097
(3)	1.502	1.395	0.617	1.318	(3)	1.594	1.430	0.784	1.487
(4)	1.765	1.701	1.588	1.861	(4)	2.128	2.167	2.040	2.192
τ_n = 0.067, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.133, k = 40	(1)	(2)	(3)	(4)
(1)	2.354	2.510	2.320	2.766	(1)	2.836	2.754	2.662	3.488
(2)	2.057	2.116	1.817	2.537	(2)	2.430	2.511	2.195	3.598
(3)	1.841	1.576	0.994	1.880	(3)	1.868	1.887	1.363	2.549
(4)	2.776	2.805	2.871	3.117	(4)	3.999	3.936	4.106	5.272

Table 14: MAE of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

To compute the robust CI, $\tau_1 := \min(\frac{40}{n}, \frac{0.2b}{mn})$ where the spacing parameter m here is 2. To compute the feasible normalizing factor $\hat{\alpha}_n$ for τ, when $k := \tau n \leq 25$, the spacing parameter is 2 and $k'_l = 10$ while $m = 1.2$ and $k'_l = 20$ when $k > 25$.



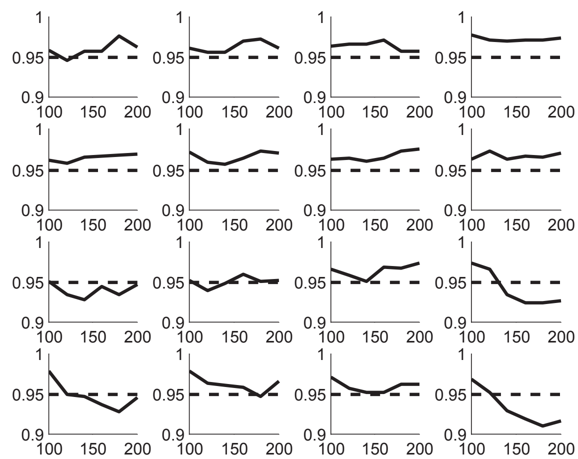
The dashed line is the coverage for BN-CI. The dotted line is the coverage for NN-CI. The solid line is the coverage for the robust CI. When $b = 120$, $n = 300$, and $\tau \in [6.67\%, 20\%]$. The horizontal dotted dashed line is the 95% nominal coverage rate. $\tau_1 = 4\%$ and $\tau_2 = 16.75\%$.

Figure 15: Coverage across quantiles

For the lower boundary, I use $\tau_n = (0.02, .0.04, 0.06)$ for $n = 300$ to compute the EV index. The subsample size used is the same as in Table 9.

	(1)	(2)	(3)	(4)
(1)	0.946	0.956	0.967	0.972
	(0.605)	(0.551)	(0.431)	(0.497)
(2)	0.958	0.960	0.964	0.973
	(0.481)	(0.456)	(0.329)	(0.428)
(3)	0.935	0.940	0.959	0.966
	(0.392)	(0.352)	(0.153)	(0.226)
(4)	0.950	0.964	0.958	0.953
	(0.570)	(0.514)	(0.438)	(0.303)

Table 15: Coverage of 95% CI, sample size = 300.



The solid line is the coverage for b out of n bootstrap CI at $k = 0$ in which $b \in [100, 200]$.

Figure 16: Coverage across subsample size

	(1)	(2)	(3)	(4)
(1)	-1.639	-3.178	0.201	-0.408
(2)	-1.635	-0.927	-0.834	-0.145
(3)	-1.097	-0.436	-0.559	-1.122
(4)	-3.313	-2.065	-1.116	-1.909

Table 16: Bias of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	19.809	47.957	27.608	10.520
(2)	16.819	19.857	19.598	9.221
(3)	10.612	11.724	4.066	4.378
(4)	16.769	11.378	8.510	6.018

Table 17: root-MSE of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	-1.623	-0.504	0.966	-1.335
(2)	-1.881	-1.243	-0.233	-1.130
(3)	-1.920	-1.365	-0.461	-1.431
(4)	-2.591	-2.069	-0.925	-1.965

Table 18: median-bias of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	7.336	7.378	5.705	6.080
(2)	7.048	7.063	4.015	4.738
(3)	5.900	4.937	1.997	2.933
(4)	6.387	5.198	4.241	3.679

Table 19: MAE of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.

F.2 Simulation results with $n = 1,000$

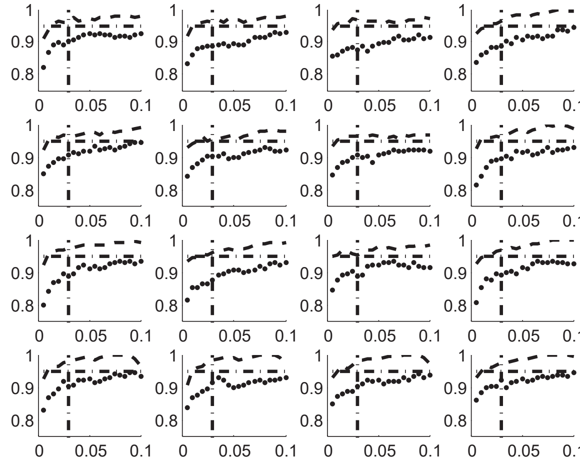
Next I consider the QTE estimator with a moderate size sample: 1,000. I am still interested in $k = (5, 10, 20, 40)$ and the corresponding quantile indices become $\tau_n = (0.005, 0.01, 0.015, 0.02)$. The subsample size used in Table 20 and Figure 17 is 300. For Figure 18, the subsample size ranges from 150 to 500.

τ_n = 0.5%, k = 5	(1)	(2)	(3)	(4)	τ_n = 1%, k = 10	(1)	(2)	(3)	(4)
(1)	0.915	0.918	0.927	0.924	(1)	0.939	0.946	0.957	0.934
	(0.065)	(0.063)	(0.051)	(0.060)		(0.073)	(0.070)	(0.059)	(0.068)
(2)	0.918	0.930	0.942	0.931	(2)	0.950	0.945	0.955	0.953
	(0.061)	(0.054)	(0.038)	(0.058)		(0.066)	(0.061)	(0.045)	(0.068)
(3)	0.926	0.933	0.949	0.921	(3)	0.950	0.949	0.954	0.948
	(0.055)	(0.048)	(0.019)	(0.036)		(0.060)	(0.053)	(0.023)	(0.044)
(4)	0.917	0.902	0.935	0.931	(4)	0.958	0.949	0.957	0.955
	(0.059)	(0.052)	(0.047)	(0.052)		(0.072)	(0.065)	(0.060)	(0.066)
τ_n = 2%, k = 20	(1)	(2)	(3)	(4)	τ_n = 4%, k = 40	(1)	(2)	(3)	(4)
(1)	0.966	0.958	0.959	0.957	(1)	0.969	0.968	0.964	0.981
	(0.077)	(0.073)	(0.063)	(0.079)		(0.097)	(0.097)	(0.082)	(0.111)
(2)	0.961	0.952	0.965	0.958	(2)	0.966	0.966	0.967	0.981
	(0.070)	(0.064)	(0.049)	(0.083)		(0.090)	(0.081)	(0.066)	(0.116)
(3)	0.966	0.954	0.953	0.954	(3)	0.978	0.971	0.972	0.977
	(0.064)	(0.058)	(0.025)	(0.051)		(0.081)	(0.073)	(0.039)	(0.078)
(4)	0.957	0.963	0.957	0.966	(4)	0.990	0.993	0.990	0.989
	(0.080)	(0.070)	(0.066)	(0.073)		(0.142)	(0.127)	(0.122)	(0.132)

Table 20: Coverage of 95% b out of n bootstrap CI, sample size $= 1,000$

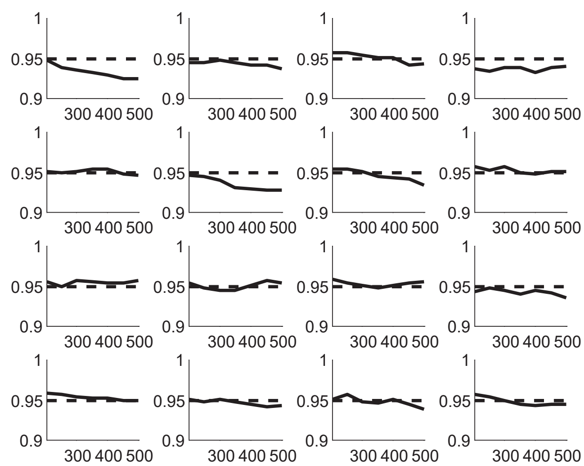
τ_n = 0.5%, k = 5	(1)	(2)	(3)	(4)	τ_n = 1%, k = 10	(1)	(2)	(3)	(4)
(1)	0.822	0.832	0.857	0.836	(1)	0.868	0.861	0.859	0.861
	(0.057)	(0.051)	(0.041)	(0.050)		(0.061)	(0.056)	(0.047)	(0.058)
(2)	0.850	0.841	0.844	0.812	(2)	0.873	0.868	0.867	0.841
	(0.049)	(0.044)	(0.031)	(0.045)		(0.053)	(0.049)	(0.037)	(0.054)
(3)	0.797	0.814	0.846	0.804	(3)	0.839	0.852	0.876	0.852
	(0.042)	(0.034)	(0.015)	(0.028)		(0.045)	(0.039)	(0.018)	(0.035)
(4)	0.827	0.835	0.849	0.858	(4)	0.866	0.866	0.873	0.884
	(0.049)	(0.043)	(0.038)	(0.042)		(0.055)	(0.050)	(0.046)	(0.052)
τ_n = 2%, k = 20	(1)	(2)	(3)	(4)	τ_n = 4%, k = 40	(1)	(2)	(3)	(4)
(1)	0.899	0.883	0.879	0.883	(1)	0.914	0.893	0.871	0.896
	(0.066)	(0.063)	(0.055)	(0.066)		(0.074)	(0.072)	(0.065)	(0.077)
(2)	0.895	0.892	0.885	0.885	(2)	0.910	0.911	0.901	0.894
	(0.057)	(0.054)	(0.043)	(0.062)		(0.063)	(0.062)	(0.051)	(0.072)
(3)	0.872	0.864	0.896	0.875	(3)	0.912	0.893	0.917	0.903
	(0.047)	(0.044)	(0.022)	(0.040)		(0.052)	(0.049)	(0.029)	(0.051)
(4)	0.894	0.887	0.886	0.902	(4)	0.922	0.922	0.921	0.914
	(0.064)	(0.061)	(0.058)	(0.064)		(0.079)	(0.076)	(0.075)	(0.084)

Table 21: Coverage of 95% n out of n bootstrap CI, sample size $= 1,000$



Each (i,j)-th subplot represents the (i,j)-th model. The dashed line is the coverage of BN-CI with $b = 300$ and $n = 1,000$ for quantile index $\tau \in [0.5\%, 10\%]$. The dotted line is the coverage of NN-CI. The horizontal dotted dashed line is the 95% nominal coverage rate, and the vertical dotted dashed line is $\tau = \min(\frac{40}{n}, \frac{0.2b}{mn})$.

Figure 17: Coverage across quantiles



Each (i,j)-th subplot represents the (i,j)-th model. The solid line is the coverage for b out of n bootstrap CI at $k = 10$ in which $b \in [150, 500]$.

Figure 18: Coverage across subsample size

τ_n = 0.005, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.010, k = 10	(1)	(2)	(3)	(4)
(1)	-0.323	0.077	1.079	-0.441	(1)	-0.349	0.021	0.940	-0.107
(2)	-0.963	0.386	0.206	-1.186	(2)	-0.964	-0.450	0.432	-1.487
(3)	-0.561	-0.070	-0.154	-0.647	(3)	-0.459	-0.272	0.067	-0.767
(4)	-0.044	-0.368	0.229	-0.527	(4)	0.230	-0.455	1.002	-0.665
τ_n = 0.020, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.040, k = 40	(1)	(2)	(3)	(4)
(1)	0.297	0.674	1.324	-0.213	(1)	0.956	0.123	0.981	-0.933
(2)	-0.715	-0.185	0.006	-0.652	(2)	0.273	0.450	0.229	-1.097
(3)	-0.482	-0.261	0.066	-0.328	(3)	-0.523	0.437	-0.010	-0.322
(4)	0.270	-0.360	0.977	-0.178	(4)	0.328	0.391	1.286	-0.636

Table 22: Bias of the median-unbiased estimator, sample size = $1,000$. All values are inflated by 1,000.

τ_n = 0.005, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.010, k = 10	(1)	(2)	(3)	(4)
(1)	1.348	1.251	1.121	1.273	(1)	1.409	1.338	1.168	1.402
(2)	1.242	1.083	0.790	1.252	(2)	1.322	1.211	0.924	1.403
(3)	1.111	0.958	0.384	0.748	(3)	1.130	0.993	0.456	0.889
(4)	1.194	1.030	0.933	1.072	(4)	1.319	1.209	1.156	1.199
τ_n = 0.020, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.040, k = 40	(1)	(2)	(3)	(4)
(1)	1.583	1.547	1.363	1.640	(1)	1.803	1.736	1.594	1.941
(2)	1.389	1.377	1.079	1.611	(2)	1.538	1.483	1.262	1.827
(3)	1.177	1.103	0.555	1.010	(3)	1.307	1.239	0.666	1.202
(4)	1.492	1.441	1.438	1.480	(4)	1.866	1.832	1.830	1.935

Table 23: root-MSE of the median-unbiased estimator, sample size = $1,000$. All values are inflated by 100.

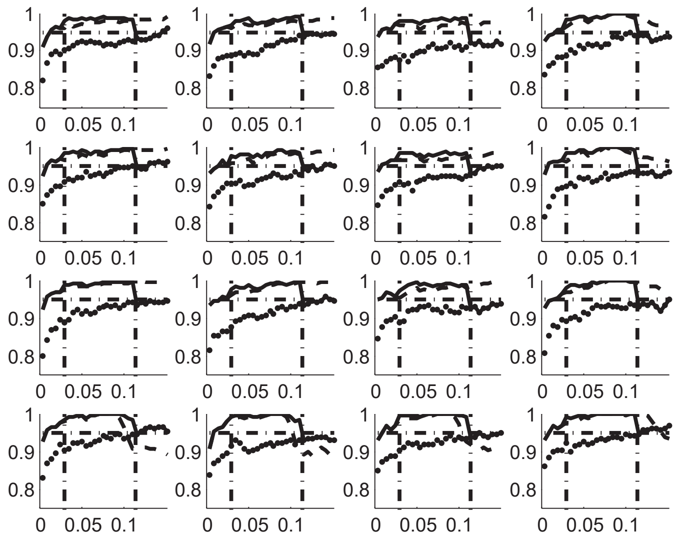
τ_n = 0.005, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.010, k = 10	(1)	(2)	(3)	(4)
(1)	-0.237	-0.347	-0.205	0.759	(1)	-0.483	-0.414	0.430	0.080
(2)	-0.892	0.703	0.359	-0.511	(2)	-1.326	-0.187	0.559	-0.854
(3)	-0.101	0.459	-0.077	-0.041	(3)	-0.380	0.156	0.220	-0.447
(4)	-0.368	-0.188	0.302	-0.309	(4)	0.364	-0.403	0.919	-0.926
τ_n = 0.020, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.040, k = 40	(1)	(2)	(3)	(4)
(1)	-0.089	0.041	0.639	1.274	(1)	0.158	-0.312	0.725	0.338
(2)	-1.243	-0.342	-0.315	0.261	(2)	0.617	0.511	-0.131	0.157
(3)	-0.181	0.585	0.011	0.129	(3)	-0.229	0.316	0.037	-0.391
(4)	0.119	-0.297	0.592	-0.285	(4)	-0.029	-0.542	0.178	-0.868

Table 24: median-bias of the median-unbiased estimator, sample size = $1,000$. All values are inflated by 1,000.

τ_n = 0.005, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.010, k = 10	(1)	(2)	(3)	(4)
(1)	0.867	0.798	0.724	0.895	(1)	0.902	0.901	0.765	0.916
(2)	0.842	0.733	0.525	0.807	(2)	0.887	0.814	0.602	0.893
(3)	0.729	0.660	0.232	0.509	(3)	0.786	0.688	0.289	0.626
(4)	0.832	0.686	0.588	0.697	(4)	0.894	0.819	0.789	0.802
τ_n = 0.020, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.040, k = 40	(1)	(2)	(3)	(4)
(1)	1.050	1.017	0.907	1.140	(1)	1.226	1.180	1.080	1.316
(2)	0.942	0.966	0.726	1.084	(2)	1.006	0.982	0.794	1.216
(3)	0.780	0.741	0.373	0.688	(3)	0.895	0.811	0.448	0.840
(4)	1.020	0.988	0.928	0.968	(4)	1.255	1.241	1.225	1.252

Table 25: MAE of the median-unbiased estimator, sample size = $1,000$. All values are inflated by 100.

To compute the robust CI, $\tau_1 := \min(\frac{40}{n}, \frac{0.2b}{mn})$ where the spacing parameter m here is 2 and $\tau_2 = \frac{b}{n\sqrt{\log(n)}}$. To compute the feasible normalizing factor $\hat{\alpha}_n$ for τ, when $k := \tau n \leq 50$, the spacing parameter is 2 and $k_l' = 10$ while $m = 1.2$ and $k_l' = 20$ when $k > 50$.



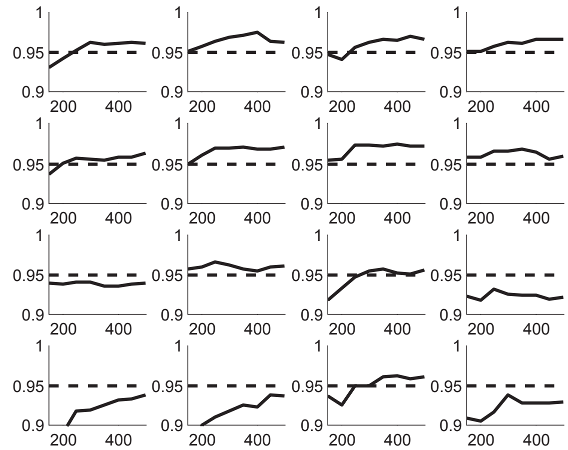
The dashed line is the coverage for BN-CI. The dotted line is the coverage for NN-CI. The solid line is the coverage for the robust CI. When $b = 300$, $n = 1,000$, and $\tau \in [0.5\%, 15\%]$. The horizontal dotted dashed line is the 95% nominal coverage rate. $\tau_1 = 4\%$ and $\tau_2 = 11.41\%$.

Figure 19: Coverage across quantiles

For the lower boundary: I use $\tau_n = (0.02, 0.04, \cdots, 0.1)$ for $n = 1,000$ to compute the EV index. The subsample size used is the same as in Table 20.

	(1)	(2)	(3)	(4)
(1)	0.963	0.969	0.963	0.963
	(0.200)	(0.172)	(0.119)	(0.145)
(2)	0.956	0.969	0.973	0.966
	(0.168)	(0.155)	(0.093)	(0.129)
(3)	0.941	0.963	0.955	0.926
	(0.140)	(0.112)	(0.037)	(0.063)
(4)	0.920	0.918	0.950	0.938
	(0.139)	(0.116)	(0.099)	(0.085)

Table 26: Coverage of 95% CI, sample size = 1,000.



The solid line is the coverage for b out of n bootstrap CI at $k = 0$ in which $b \in [150, 500]$.

Figure 20: Coverage across subsample size

	(1)	(2)	(3)	(4)
(1)	-0.475	-0.693	0.092	-0.023
(2)	-0.586	-0.301	-0.046	-0.287
(3)	-0.813	-0.754	-0.294	-0.675
(4)	-0.831	-0.924	-0.637	-0.870

Table 27: Bias of the median-unbiased 0-QTE estimator, sample size = $1,000$. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	6.162	5.751	3.711	4.096
(2)	5.502	4.651	2.809	3.450
(3)	4.042	3.140	1.152	1.859
(4)	4.473	3.836	3.049	2.862

Table 28: root-MSE of the median-unbiased 0-QTE estimator, sample size = $1,000$. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	-0.634	-0.861	0.348	-0.322
(2)	-0.623	-0.304	-0.094	-0.670
(3)	-1.258	-0.946	-0.305	-0.705
(4)	-1.135	-0.931	-0.728	-0.937

Table 29: median-bias of the median-unbiased 0-QTE estimator, sample size = $1,000$. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	4.046	3.736	2.413	2.616
(2)	3.546	3.029	1.785	2.277
(3)	2.889	2.189	0.711	1.257
(4)	3.048	2.429	1.910	1.918

Table 30: MAE of the median-unbiased 0-QTE estimator, sample size = $1,000$. All values are inflated by 100.

F.3 Simulation results with $n = 5,000$

τ_n = 0.001, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.002, k = 10	(1)	(2)	(3)	(4)
(1)	-0.043	0.179	0.021	-0.483	(1)	0.447	0.317	-0.030	-0.843
(2)	-0.247	-0.833	0.129	-0.589	(2)	-0.407	-0.798	0.077	-0.916
(3)	-0.194	-0.123	-0.087	-0.421	(3)	-0.398	-0.322	-0.056	-0.235
(4)	-0.662	-0.962	-0.106	-0.386	(4)	-0.691	-0.608	-0.034	-0.411
τ_n = 0.004, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.008, k = 40	(1)	(2)	(3)	(4)
(1)	0.156	0.498	0.098	-0.586	(1)	0.392	0.358	-0.265	-0.245
(2)	-0.172	-0.727	0.123	-0.757	(2)	-0.209	-0.629	-0.009	-0.090
(3)	-0.390	-0.026	-0.063	-0.417	(3)	-0.278	-0.056	-0.144	-0.257
(4)	-0.877	-0.180	0.326	-0.501	(4)	-0.280	-0.361	0.150	-0.192

Table 31: Bias of the median-unbiased estimator, sample size = $5,000$. All values are inflated by 1,000.

τ_n = 0.001, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.002, k = 10	(1)	(2)	(3)	(4)
(1)	5.806	4.715	3.553	4.202	(1)	5.924	5.190	3.816	4.733
(2)	5.259	4.284	2.571	4.057	(2)	5.447	4.410	2.858	4.608
(3)	5.112	3.530	1.151	2.489	(3)	5.211	3.897	1.291	2.869
(4)	4.574	3.676	2.830	3.113	(4)	4.908	3.828	3.296	3.707
τ_n = 0.004, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.008, k = 40	(1)	(2)	(3)	(4)
(1)	6.310	5.429	4.406	5.364	(1)	6.820	5.904	4.971	6.109
(2)	5.659	4.675	3.310	5.033	(2)	5.804	5.231	3.802	5.674
(3)	5.052	3.962	1.565	3.130	(3)	5.070	4.110	1.848	3.638
(4)	5.344	4.289	3.890	4.381	(4)	5.399	5.050	4.567	5.029

Table 32: root-MSE of the median-unbiased estimator, sample size = $5,000$. All values are inflated by 1,000.

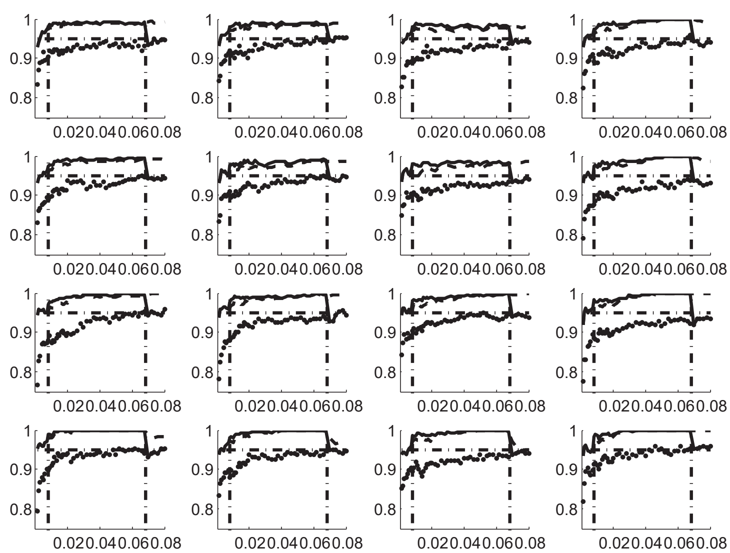
τ_n = 0.001, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.002, k = 10	(1)	(2)	(3)	(4)
(1)	-0.247	-0.013	-0.337	-0.190	(1)	0.247	0.243	-0.332	-0.745
(2)	-0.039	-0.684	0.023	-0.261	(2)	-0.354	-0.724	-0.030	-0.745
(3)	0.110	0.010	-0.017	-0.246	(3)	-0.192	-0.203	-0.019	-0.138
(4)	-0.622	-0.758	0.066	-0.174	(4)	-0.575	-0.303	0.158	-0.111
τ_n = 0.004, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.008, k = 40	(1)	(2)	(3)	(4)
(1)	0.186	0.244	-0.175	-0.352	(1)	0.342	0.375	-0.503	-0.291
(2)	0.132	-0.610	0.039	-0.411	(2)	-0.228	-0.661	0.152	0.105
(3)	0.073	-0.031	-0.008	-0.404	(3)	-0.315	-0.069	-0.154	-0.208
(4)	-1.020	-0.002	0.315	-0.296	(4)	-0.179	-0.391	0.085	-0.080

Table 33: median-bias of the median-unbiased estimator, sample size = $5,000$. All values are inflated by 1,000.

τ_n = 0.001, k = 5	(1)	(2)	(3)	(4)	τ_n = 0.002, k = 10	(1)	(2)	(3)	(4)
(1)	3.916	3.073	2.209	2.783	(1)	4.091	3.678	2.453	2.979
(2)	3.658	2.881	1.644	2.715	(2)	3.542	2.976	1.994	3.099
(3)	3.440	2.349	0.743	1.649	(3)	3.507	2.654	0.863	1.875
(4)	3.098	2.380	1.889	2.169	(4)	3.397	2.492	2.131	2.506
τ_n = 0.004, k = 20	(1)	(2)	(3)	(4)	τ_n = 0.008, k = 40	(1)	(2)	(3)	(4)
(1)	4.407	3.901	2.925	3.599	(1)	4.634	4.084	3.393	4.172
(2)	3.679	3.158	2.242	3.247	(2)	3.806	3.814	2.534	3.813
(3)	3.635	2.673	1.027	2.169	(3)	3.636	2.798	1.219	2.425
(4)	3.837	2.860	2.608	3.035	(4)	3.724	3.478	3.119	3.437

Table 34: MAE of the median-unbiased estimator, sample size $= 5,000$. All values are inflated by 1,000.

To compute the robust CI, $\tau_1 := \min(\frac{40}{n}, \frac{0.2b}{mn})$ where the spacing parameter m here is 2 and $\tau_2 = \frac{b}{n\sqrt{\log(n)}}$. To compute the feasible normalizing factor $\hat{\alpha}_n$ for τ, when $k := \tau n \leq 100$, the spacing parameter is 2 and $k_l' = 10$ while $m = 1.2$ and $k_l' = 20$ when $k > 100$.



The dashed line is the coverage for BN-CI. The dotted line is the coverage for NN-CI. The solid line is the coverage for the robust CI. When $b = 1,000$, $n = 5,000$, and $\tau \in [0.1\%, 8\%]$. The horizontal dotted dashed line is the 95% nominal coverage rate. $\tau_1 = 0.8\%$ and $\tau_2 = 6.85\%$.

Figure 21: Coverage across quantiles

Next are the finite sample performance of the median-unbiased point estimator.

	(1)	(2)	(3)	(4)
(1)	-0.148	-0.282	0.015	0.105
(2)	0.006	-0.120	-0.021	0.063
(3)	0.188	-0.082	-0.056	0.027
(4)	-0.100	-0.284	-0.058	-0.086

Table 35: Bias of the median-unbiased 0-QTE estimator, sample size = $5,000$. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	4.388	2.944	1.861	2.549
(2)	3.416	2.453	1.360	2.149
(3)	2.457	1.646	0.470	0.959
(4)	2.512	1.687	1.205	1.282

Table 36: root-MSE of the median-unbiased 0-QTE estimator, sample size = $5,000$. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	-0.246	-0.251	0.093	0.041
(2)	-0.134	-0.136	-0.005	-0.099
(3)	-0.189	-0.291	-0.082	-0.059
(4)	-0.382	-0.374	-0.109	-0.177

Table 37: median-bias of the median-unbiased 0-QTE estimator, sample size = $5,000$. All values are inflated by 100.

	(1)	(2)	(3)	(4)
(1)	2.287	1.687	1.076	1.498
(2)	1.998	1.478	0.849	1.404
(3)	1.548	1.062	0.300	0.604
(4)	1.637	1.024	0.750	0.830

Table 38: MAE of the median-unbiased 0-QTE estimator, sample size = $5,000$. All values are inflated by 100.