

Estimation and Inference for Distribution Functions and Quantile Functions in Endogenous Treatment Effect Models

Yu-Chin Hsu*

Robert P. Lieli†

Tsung-Chih Lai‡

Abstract

We propose a new monotonizing method to obtain estimators for the distribution functions of potential outcomes among the group of compliers in an endogenous treatment effect model that are monotonically increasing and bounded between zero and one. Corresponding quantile function estimators are obtained by applying the inverse map to the CDF estimators. We show that both these estimators converge weakly to zero mean Gaussian processes. A simulation method is proposed to approximate the limiting processes for uniform inference. A Monte Carlo simulation and an application addressing the effect of fertility on family income illustrate the usefulness of the results.

JEL classification: C21, C26

Keywords: distribution function, quantile function, treatment effects, instrumental variables, inverse probability weighted estimator.

* Corresponding author. Institute of Economics, Academia Sinica. E-mail: ychsu@econ.sinica.edu.tw. 128 Academia Road, Section 2, Nankang, Taipei, 115 Taiwan. Tel: 886-2-27822791 ext. 322. Fax: 886-2-2785-3946.

† Department of Economics Central European University, Budapest and the National Bank of Hungary. Email: lieli@ceu.hu

‡ Department of Economics, National Taiwan University. Email: d00323011@ntu.edu.tw

1 Introduction

Estimating the causal effect of a policy or a binary treatment has drawn a lot of attention in the last two decades, especially when the treatment is endogenous. Pioneered by Imbens and Angrist (1994), a nonparametric instrumental variable (IV) treatment effect model is now routinely used to evaluate the local average treatment effect (LATE), i.e., the average treatment effect among the group of compliers (to be defined later). The original framework of Imbens and Angrist (1994) has been generalized by Abadie (2003), Frölich (2007), Hong and Nekipelov (2010), and Donald, Hsu and Lieli (2014a, DHL hereafter) to incorporate covariates in the model. In addition to the LATE, Frölich and Lechner (2010), Hong and Nekipelov (2010), and DHL also propose to examine the local average treatment effect for the treated (LATT). Frölich and Melly (2013) go beyond the means and explore the local quantile treatment effects (LQTE) that measures the difference between certain quantiles of potential outcomes among the complier group.

In this paper, we are interested in the distributions of potential outcomes among the group of compliers and their corresponding quantile functions. We propose a two-step estimation procedure for the CDFs. In the first step, we use the inverse probability weighted (IPW) estimators similar to DHL. However, the resulting estimators are not necessarily monotonically increasing nor bounded between 0 and 1 since some of the weights could be strictly negative. In the second step, we monotonize and rescale the preliminaries to obtain legitimate CDF estimators. We then apply the inverse mapping to the distribution functions to obtain the quantile processes. Both CDF and quantile function estimators are shown to converge weakly to Gaussian processes at the parametric rate. Next, we propose a simulation method to approximate the limiting processes that takes the estimation errors into account. These results are useful for conducting inference that involves the whole distribution functions or a continuum of quantile indices. Our results on quantile function estimators generalize the pointwise results of Frölich and Melly (2013). Moreover, we propose estimators for the CDFs and quantile functions of potential outcomes among the *treated* compliers, and discuss the simulation method for conducting inference as well. Our results regarding the group of treated compliers allow us to introduce the local quantile treatment effect for the treated (LQTT), which is not considered in Frölich and Melly (2013).

We illustrate the usefulness of the theoretical results by a Monte Carlo simulation and an empirical application. In the simulation study, the finite sample performance of the estimators is examined by various global measures, including one-sided and two-sided uniform coverage rates obtained from the simulation method. The empirical coverage rates are shown to be close to the nominal ones with relatively small samples. In the empirical study, we employ U.S. census data to investigate the effect of fertility on family income using twin births as an instrument. Our findings suggest that the presence of more than one child does not cause any change in family income in the 1990s and it only affects family income significantly above the 85th quantile in 2000. If we focus on the *positive* effects instead, the result from one-sided confidence band indicates that the positive effects only exist above the 83th quantile. Nevertheless, we cannot find any evidence supporting that either the LQTE or LQTT is heterogeneous over the entire range of quantiles considered. This is in sharp contrast to the findings of Frölich and Melly (2013) that are based on the pointwise asymptotics.

This paper is related to numerous studies in the treatment effects literature. While here we use an IV framework, most of the literature considers the case in which the treatment assignment is unconfounded. The unconfoundedness assumption introduced by Rosenbaum and Rubin (1983) requires that treatment assignment is independent of the potential outcomes conditional on observable covariates. It allows us to identify the treatment effects of the whole population. Under this assumption, Rosenbaum and Rubin (1983, 1985), Heckman, Ichimura, and Todd (1997, 1998), Heckman, Ichimura, Smith and Todd (1998), Hahn (1998), Hirano, Imbens, and Ridder (2003, HIR hereafter) consider the average treatment effect (ATE) and the average treatment effect for the treated (ATT). Firpo (2007) examines the quantile treatment effects (QTE) and the quantile treatment effects for the treated (QTT), and Firpo and Pinto (2015) propose the inequality treatment effects and the inequality treatment effects for the treated. Donald and Hsu (2014, DH hereafter) propose to estimate the CDFs of potential outcomes and the corresponding quantile functions, and apply the simulation method to approximate limiting processes similar to ours. Their results can be used to construct tests for stochastic dominance relationship between the potential outcome distributions. Although our methods are similar to DH's, new techniques are needed because the monotonic mapping considered here is not Hadamard differentiable, and as a consequence, the functional delta method can not be applied. The monotonic method we propose is an alternative to Cher-

nozhlukov, Fernández-Val and Galichon’s (2009, 2010) rearrangement approaches, but as we will argue later, our method is easier to implement.

The remainder of the paper is organized as follows. In Section 2, we discuss the IV treatment effect model as well as the identification and estimation of the CDF and quantile function estimators. Section 3 presents the regularity conditions and asymptotic properties. We introduce the simulation method in Section 4 and extend our results to the group of treated compliers in Section 5. The Monte Carlo simulation and empirical application can be found respectively in Section 6 and 7. Finally, Section 8 concludes. All proofs are contained in the Appendix.

2 Identification and Estimation

2.1 Model framework

We consider the following IV framework augmented with covariates which is now standard in the treatment effects literature. For each population unit (individual), one can observe the value of a binary instrument $Z \in \{0, 1\}$ and a vector of covariates $X \in \mathbb{R}^r$. For $Z = z$, the random variable $D(z) \in \{0, 1\}$ specifies individual’s potential treatment status with $D(z) = 1$ corresponding to treatment and $D(z) = 0$ to no treatment. The actually observed treatment status is then given by $D \equiv D(Z) = D(1)Z + D(0)(1 - Z)$. Similarly, the random variable $Y(d)$ denotes the potential outcomes in the population that would obtain if one were to set $D = d$ exogenously. The observed outcome is $Y \equiv Y(D) = Y(1)D + Y(0)(1 - D)$. Let $W \equiv \{Y, D, Z, X\}$ and $\{W_i\}_{i=1}^n$ denote a random sample of observations on W .

The following assumptions, taken from Abadie (2003) or Frölich (2007) with some modifications, describe the relationships between variables defined above and justify Z being referred to as an instrument:¹

Assumption 2.1 *Suppose that*

(i) *(Independence of the instrument):* $(Y(0), Y(1), D(0), D(1)) \perp Z | X$.

(ii) *(First stage):* $P(D(1) = 1) > P(D(0) = 1)$ and $0 < P(Z = 1 | X) < 1$.

¹In the standard framework, one would define $Y(z, d)$ as the potential outcomes in the population that would obtain if one were to set $Z = z$ and $D = d$ exogenously and impose the exclusion of the instrumental assumption: $P(Y(1, d) = Y(0, d)) = 1$ for $d \in \{0, 1\}$. Then it is equivalent to our approach where we define $Y(d)$ directly.

(iii) (Monotonicity): $P(D(1) \geq D(0)) = 1$.

Assumption 2.1(i) is the conditional independence assumption in the sense that Z is “as good as randomly assigned” once we condition on X . Assumption 2.1(ii) guarantees that Z and D are correlated and it also implies that the group of compliers defined below is of strictly positive measure. In addition, Assumption 2.1(ii) indicates that the distributions $X|Z = 0$ and $X|Z = 1$ have common support. The monotonicity of $D(z)$ in z , required in the last part, rules out the existence of defiers $[D(0) = 1, D(1) = 0]$ and allows for other three different types of population units with non-zero mass: compliers $[D(0) = 0, D(1) = 1]$, always-takers $[D(0) = 1, D(1) = 1]$ and never-takers $[D(0) = 0, D(1) = 0]$.

2.2 Identification of Distribution Functions

In the IV framework, we can only identify the treatment effect among the group of compliers without further assumptions. Therefore, the objects of our interest are the distribution functions and quantile functions of $Y(0)$ and $Y(1)$ among the group of compliers. We first discuss the identification and estimation of the distribution functions.

For $y \in \mathbb{R}$, let $F_{Y^0|\mathcal{C}}(y) = E[1(Y(0) \leq y)|\mathcal{C}]$ and $F_{Y^1|\mathcal{C}}(y) = E[1(Y(1) \leq y)|\mathcal{C}]$ denote the CDFs of $Y(0)$ and $Y(1)$ among the group of compliers, where \mathcal{C} denotes the group of compliers and $1(\cdot)$ is the indicator function. Let $q(x) = P(Z = 1|X = x)$ be the instrument propensity score which denotes the probability of $Z = 1$ given $X = x$.² The following lemma summarizing the identification results of $F_{Y^0|\mathcal{C}}$ and $F_{Y^1|\mathcal{C}}$ is from Theorem 3.1 of Abadie (2003). We present it here for the completeness of the paper.

Lemma 2.1 (*Identification of $F_{Y^0|\mathcal{C}}$ and $F_{Y^1|\mathcal{C}}$*): Under Assumption 2.1, $F_{Y^0|\mathcal{C}}$ and $F_{Y^1|\mathcal{C}}$ are

²As in DHL, we called $q(x)$ as the instrument propensity score to distinguish from the conventional use of the term propensity score (the conditional probability of being treated).

identified by

$$\begin{aligned}
F_{Y^0|C}(y) &= \frac{1}{\Gamma_0} E \left[\frac{(1-Z)(1-D) \cdot 1(Y \leq y)}{1-q(X)} - \frac{Z(1-D) \cdot 1(Y \leq y)}{q(X)} \right], \\
&= \frac{1}{\Gamma_0} E \left[\frac{Z(D-1) \cdot 1(Y \leq y)}{q(X)} - \frac{(1-Z)(D-1) \cdot 1(Y \leq y)}{1-q(X)} \right], \\
F_{Y^1|C}(y) &= \frac{1}{\Gamma_1} E \left[\frac{ZD \cdot 1(Y \leq y)}{q(X)} - \frac{(1-Z)D \cdot 1(Y \leq y)}{1-q(X)} \right], \\
\Gamma_0 &= E \left[\frac{(1-Z)(1-D)}{1-q(X)} - \frac{Z(1-D)}{q(X)} \right] = E \left[\frac{Z(D-1)}{q(X)} - \frac{(1-Z)(D-1)}{1-q(X)} \right], \\
\Gamma_1 &= E \left[\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)} \right],
\end{aligned} \tag{2.1}$$

and $\Gamma_0 = \Gamma_1 = E[D(1) - D(0)] > 0$.

If we treat Z as the treatment indicator and $D \cdot 1(Y \leq y)$ as the outcome variable, then $F_{Y^1|C}$ is identified by the ratio of the ATE of Z on $D \cdot 1(Y \leq y)$ over the ATE of Z on D . Similarly, $F_{Y^0|C}$ is identified by the ratio of the ATE of Z on $(D-1) \cdot 1(Y \leq y)$ over the ATE of Z on $D-1$. Furthermore, define

$$\begin{aligned}
\kappa_0 &= \frac{1}{\Gamma_0} \left[\frac{Z(D-1)}{q(X)} - \frac{(1-Z)(D-1)}{1-q(X)} \right], \\
\kappa_1 &= \frac{1}{\Gamma_1} \left[\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)} \right],
\end{aligned} \tag{2.2}$$

and we have $E[\kappa_0] = E[\kappa_1] = 1$. Then we can see that $F_{Y^0|C}(y)$ and $F_{Y^1|C}(y)$ are weighted averages of $1(Y \leq y)$ with respectively weights κ_0 and κ_1 , i.e., $F_{Y^0|C}(y) = E[\kappa_0 \cdot 1(Y \leq y)]$ and $F_{Y^1|C}(y) = E[\kappa_1 \cdot 1(Y \leq y)]$. However, in some cases the weights κ_0 and κ_1 can be strictly negative, e.g., $\kappa_0 < 0$ if $D = 0$ and $Z = 1$, and $\kappa_1 < 0$ if $D = 1$ and $Z = 0$.

2.3 Estimation of $F_{Y^0|C}$ and $F_{Y^1|C}$

Based on (2.1), the IPW estimators for $F_{Y^0|C}$, $F_{Y^1|C}$, Γ_0 and Γ_1 are

$$\begin{aligned}
\tilde{F}_{Y^0|C}(y) &= \frac{1}{\hat{\Gamma}_0} \frac{1}{n} \sum_{i=1}^n \frac{Z_i(D_i-1) \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} - \frac{(1-Z_i)(D_i-1) \cdot 1(Y_i \leq y)}{1-\hat{q}(X_i)}, \\
\tilde{F}_{Y^1|C}(y) &= \frac{1}{\hat{\Gamma}_1} \frac{1}{n} \sum_{i=1}^n \frac{Z_i D_i \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} - \frac{(1-Z_i) D_i \cdot 1(Y_i \leq y)}{1-\hat{q}(X_i)}, \\
\hat{\Gamma}_0 &= \frac{1}{n} \sum_{i=1}^n \frac{Z_i(D_i-1)}{\hat{q}(X_i)} - \frac{(1-Z_i)(D_i-1)}{1-\hat{q}(X_i)}, \\
\hat{\Gamma}_1 &= \frac{1}{n} \sum_{i=1}^n \frac{Z_i D_i}{\hat{q}(X_i)} - \frac{(1-Z_i) D_i}{1-\hat{q}(X_i)},
\end{aligned} \tag{2.3}$$

where $\hat{q}(x)$ is a nonparametric estimator for $q(x)$. As in HIR, we use the series logit estimator (SLE) to estimate $q(x)$ based on the power series. Let $\lambda = (\lambda_1, \dots, \lambda_r)' \in \mathbb{Z}_+^r$ be a r -dimensional vector of non-negative integers and define the norm for λ as $|\lambda| = \sum_{j=1}^r \lambda_j$. Let $\{\lambda(k)\}_{k=1}^\infty$ be a sequence including all distinct $\lambda \in \mathbb{Z}_+^r$ such that $|\lambda(k)|$ is non-decreasing in k and let $x^\lambda = \prod_{j=1}^r x_j^{\lambda_j}$. For any integer K , define $R^K(x) = (x^{\lambda(1)}, \dots, x^{\lambda(K)})'$ as a vector of power functions. Let $\mathcal{L}(a) = \exp(a)/(1 + \exp(a))$ be the logistic CDF. The SLE for $q(x)$ is defined as $\hat{q}(x) = \mathcal{L}(R^K(x)' \hat{\pi}_K)$ where

$$\hat{\pi}_K = \arg \max_{\pi_K} \frac{1}{n} \sum_{i=1}^n \left(Z_i \cdot \ln \mathcal{L}(R^K(X_i)' \pi_K) + (1 - Z_i) \cdot \ln (1 - \mathcal{L}(R^K(X_i)' \pi_K)) \right).$$

The asymptotic properties of $\hat{q}(x)$ are discussed in Appendix A of HIR.

One can use other nonparametric estimators for $q(x)$ such as the local polynomial estimator as in Ichimura and Linton (2005) and Donald, Hsu and Lieli (2014b), or the higher order kernel estimator as in Abrevaya, Hsu and Lieli (2014). However, a drawback of using these estimators is that the estimated instrument propensity score is not necessarily bounded away from 0 and 1 in finite sample, and proper trimming is required. In contrast, SLE by construction avoids this problem and therefore the trimming is not necessary.³ One can also use the imputation estimators for $\tilde{F}_{Y_0|C}(y)$ and $\tilde{F}_{Y_1|C}(y)$, e.g., Frölich (2007), and Hong and Nekipelov (2010). Under suitable regularity assumptions, we expect that all the results discussed below still hold for the imputation estimators.

As shown below, the estimators $\tilde{F}_{Y_0|C}(y)$ and $\tilde{F}_{Y_1|C}(y)$ will converge weakly to a two-dimensional mean zero Gaussian process at the parametric rate. However, $\tilde{F}_{Y_0|C}(y)$ and $\tilde{F}_{Y_1|C}(y)$ are not guaranteed to be monotonically increasing nor bounded in the unit interval. To see this, let

$$\begin{aligned} \hat{\kappa}_{0,i} &= \frac{1}{\hat{\Gamma}_0} \left[\frac{Z_i(D_i - 1)}{\hat{q}(X)} - \frac{(1 - Z_i)(D_i - 1)}{1 - \hat{q}(X_i)} \right], \\ \hat{\kappa}_{1,i} &= \frac{1}{\hat{\Gamma}_1} \left[\frac{Z_i D_i}{\hat{q}(X)} - \frac{(1 - Z_i) D_i}{1 - \hat{q}(X_i)} \right], \end{aligned} \tag{2.4}$$

be the estimated weights of $\kappa_{0,i}$ and $\kappa_{1,i}$. It is obvious that $n^{-1} \sum_{i=1}^n \hat{\kappa}_{0,i} = n^{-1} \sum_{i=1}^n \hat{\kappa}_{1,i} = 1$. Hence, the estimators can be represented as the weighted averages of $1(Y_i \leq y)$, i.e., $\tilde{F}_{Y_0|C}(y) = n^{-1} \sum_{i=1}^n \hat{\kappa}_{0,i} 1(Y_i \leq y)$ and $\tilde{F}_{Y_1|C}(y) = n^{-1} \sum_{i=1}^n \hat{\kappa}_{1,i} 1(Y_i \leq y)$. Since $\hat{\kappa}_{0,i} < 0$ if $D_i = 0$

³SLE with trimming can nevertheless improve estimation results. See DHL for more details.

and $Z_i = 1$, and $\hat{\kappa}_{1,i} < 0$ if $D_i = 1$ and $Z_i = 0$, $\tilde{F}_{Y^0|C}(y)$ and $\tilde{F}_{Y^1|C}(y)$ are not necessarily monotonically increasing, nor are they bounded between 0 and 1.⁴ This is not a desirable property of CDF estimators. To fix this, we propose a method to restore the monotonicity without changing first-order asymptotics of our estimators. Define ϕ_1 , ϕ_2 and ϕ be functionals such that for any f with $\sup_{a \in \mathcal{Y}} f(a) > 0$,

$$\phi_1(f)(y) = \sup_{a \leq y} f(a), \quad \phi_2(f)(y) = \frac{f(y)}{\sup_{a \in \mathcal{Y}} f(a)}, \quad \phi = \phi_2 \circ \phi_1.⁵ \quad (2.5)$$

Next, define that

$$\hat{F}_{Y^0|C}(y) = \phi(\tilde{F}_{Y^0|C})(y), \quad \hat{F}_{Y^1|C}(y) = \phi(\tilde{F}_{Y^1|C})(y). \quad (2.6)$$

It is easy to see that $\hat{F}_{Y^0|C}(y)$ and $\hat{F}_{Y^1|C}(y)$ are bounded by 0 and 1, and non-decreasing in y .

An advantage of the proposed monotonicizing method is that $\tilde{F}_{Y^0|C}(y)$ and $\tilde{F}_{Y^1|C}(y)$ are step functions with jumps at Y_i 's. Therefore, estimators in (2.6) can be easily implemented by the following procedure:

1. WLOG assume Y is bounded by 0 and \bar{y} , and there are no ties between Y_i 's.
2. Let $Y_{(0)} = 0$ and $Y_{(n+1)} = \bar{y}$. For $i = 1, \dots, n$, let $Y_{(i)}$ denote the i -th smallest element among the Y_i 's so that $0 = Y_{(0)} < Y_{(1)} < \dots < Y_{(n)} < Y_{(n+1)} = \bar{y}$.
3. Let $\hat{B}_0 = \sup_{y \in \mathcal{Y}} \tilde{F}_{Y^0|C}(y)$ and $\hat{B}_1 = \sup_{y \in \mathcal{Y}} \tilde{F}_{Y^1|C}(y)$. Note that $\hat{B}_0 \geq 1$ and $\hat{B}_1 \geq 1$ because $\tilde{F}_{Y^0|C}(\bar{y}) = \tilde{F}_{Y^1|C}(\bar{y}) = 1$.
4. Define $\hat{F}_{Y^0|C}(y)$ by induction. Let $\hat{F}_{Y^0|C}(0) = 0$ for $Y_{(0)} \leq y < Y_{(1)}$ and $\hat{F}_{Y^0|C}(\bar{y}) = 1$. Suppose $\hat{F}_{Y^0|C}(y)$ is already defined for $Y_{(0)} \leq y < Y_{(i)}$, we then define $\hat{F}_{Y^0|C}(y)$ for $Y_{(i)} \leq y < Y_{(i+1)}$ as

$$\begin{aligned} \hat{F}_{Y^0|C}(y) = & \hat{F}_{Y^0|C}(Y_{(i-1)}) \cdot 1 \left(\frac{\tilde{F}_{Y^0|C}(Y_{(i)})}{\hat{B}_0} \leq \hat{F}_{Y^0|C}(Y_{(i-1)}) \right) \\ & + \frac{\tilde{F}_{Y^0|C}(Y_{(i)})}{\hat{B}_0} \cdot 1 \left(\frac{\tilde{F}_{Y^0|C}(Y_{(i)})}{\hat{B}_0} > \hat{F}_{Y^0|C}(Y_{(i-1)}) \right). \end{aligned}$$

5. $\hat{F}_{Y^1|C}(y)$ can be constructed similarly.

⁴The sign of $\hat{\kappa}_{0,i}$ also depends on the sign of $\hat{\Gamma}_0$, which converges in probability to $\Gamma_0 > 0$. For simplicity we assume that $\hat{\Gamma}_0$ is strictly positive. Similarly, we assume that $\hat{\Gamma}_1 > 0$.

⁵Since $\phi_2 \circ \phi_1(f)$ and $\phi_1 \circ \phi_2(f)$ are identical for any f with $\sup_{a \in \mathcal{Y}} f(a) > 0$, it does not matter whether we define $\phi = \phi_2 \circ \phi_1$ or $\phi = \phi_1 \circ \phi_2$.

2.4 Estimation of Quantile Functions

The quantile function $Q(t)$ of the corresponding distribution function $F(z)$ is defined as

$$Q(t) = \inf\{z : F(z) \geq t\}$$

for $t \in [0, 1]$. If $F(z)$ is strictly increasing, we have $Q(F(z)) = z$ and the quantile function is denoted by $Q(t) \equiv F^{-1}(t)$. The quantile functions of $Y(0)$ and $Y(1)$ among the compliers are

$$Q_{Y^0|C}(t) = \inf\{y : F_{Y^0|C}(y) \geq t\}, \quad Q_{Y^1|C}(t) = \inf\{z : F_{Y^1|C}(y) \geq t\},$$

for $t \in [0, 1]$, which can be estimated by

$$\widehat{Q}_{Y^0|C}(t) = \inf\{y : \widehat{F}_{Y^0|C}(y) \geq t\}, \quad \widehat{Q}_{Y^1|C}(t) = \inf\{y : \widehat{F}_{Y^1|C}(y) \geq t\}.$$

3 Asymptotic Properties

3.1 Assumptions

In addition to Assumption 2.1, we make the following assumptions similar to those in HIR.

Assumption 3.1 (*Distributions of $Y(0)$ and $Y(1)$ for the Compliers*):

(i) *Conditional on the group of compliers, $Y(0)$ and $Y(1)$ have convex and compact supports $[y_{0\ell}, y_{0u}]$ and $[y_{1\ell}, y_{1u}]$. Let $\mathcal{Y} = [\min\{y_{0\ell}, y_{1\ell}\}, \max\{y_{0u}, y_{1u}\}]$ and WLOG assume that $\mathcal{Y} = [0, \bar{y}]$ with $\bar{y} < \infty$.*

(ii) *$F_{Y^0|C}(y)$ and $F_{Y^1|C}(y)$ are continuous on \mathcal{Y} with $F_{Y^0|C}(0) = F_{Y^1|C}(0) = 0$.*

Assumption 3.2 (*Distribution of X*):

(i) *The support of the r -dimensional covariate X is a Cartesian product of compact intervals, $\mathcal{X} = \prod_{j=1}^r [x_{\ell j}, x_{u j}]$.*

(ii) *The density of X is bounded away from 0 on \mathcal{X} .*

We define the following conditional moments: $m_z^0(y, x) = E[(1 - D) \cdot 1(Y \leq y) | Z = z, X = x]$, $m_z^1(y, x) = E[D \cdot 1(Y \leq y) | Z = z, X = x]$, and $\mu_z(x) = E[D | X = x, Z = z]$ for $z \in \{0, 1\}$. Note that $m_z^d(y, x)$ are monotonically increasing in y for all $x \in \mathcal{X}$ for $d, z \in \{0, 1\}$ and under Assumption 2.1(i), we have $\mu_z(x) = E[D(z) | X = x]$.⁶

Assumption 3.3 (*Conditional moments*):

- (i) For any $x \in \mathcal{X}$, $m_z^d(y, x)$ is continuous in $y \in \mathcal{Y}$ for $d, z \in \{0, 1\}$.
- (ii) $\mu_0(x)$ and $\mu_1(x)$ are continuously differentiable in $x \in \mathcal{X}$.
- (iii) For any $y \in \mathcal{Y}$, $m_z^d(y, x)$ is continuously differentiable in $x \in \mathcal{X}$ for $d, z \in \{0, 1\}$.
- (iv) For some constant $0 < C < \infty$, $|m_z^d(y_1, x) - m_z^d(y_2, x)| \leq C \cdot |F_{Y^d|C}(y_1) - F_{Y^d|C}(y_2)|$ for all $y_1, y_2 \in \mathcal{Y}$, for $d, z \in \{0, 1\}$ and for all $x \in \mathcal{X}$.

Assumption 3.4 (*Instrument Propensity Score*):

- (i) $q(x)$ is continuously differentiable of order $s \geq 7r$ for all $x \in \mathcal{X}$.
- (ii) $q(x)$ is bounded away from 0 and 1, i.e., $0 < \underline{q} \leq q(x) \leq \bar{q} < 1$ on \mathcal{X} .

Assumption 3.5 (*SLE Estimator*): The SLE of $q(x)$ uses a power series with $K = a \cdot n^\nu$ for some ν satisfying $r/4(s - r) < \nu < 1/9$ and some $a > 0$.

Some remarks are made on the assumptions:

1. Assumption 3.1 requires that conditional on the group of compliers, $Y(0)$ and $Y(1)$ have compact supports with the probabilities of $Y(0)$ and $Y(1)$ at $y = 0$ being equal to 0. For the CDF estimators, it can be weakened to cases that $F_{Y^0|C}$ and $F_{Y^1|C}$ have support on the whole real line or $F_{Y^d|C}(0) > 0$ for $d \in \{0, 1\}$. However, to estimate the whole quantile functions at the parametric rate, it is necessary that the corresponding density functions of $F_{Y^0|C}$ and $F_{Y^1|C}$ are bounded away from 0 on their support, which automatically

⁶Note that we define $m_z^0(y, x) = E[(1 - D) \cdot 1(Y \leq y) | Z = z, X = x]$ instead of $E[(D - 1) \cdot 1(Y \leq y) | Z = z, X = x]$ so that $m_z^0(y, x) \geq 0$ and is increasing in y .

rules out the unbounded support case.⁷ Also, $F_{Y^0|\mathcal{C}}(0) = F_{Y^1|\mathcal{C}}(0) = 0$ are needed for estimating the whole quantile functions.

2. Assumption 3.2 requires that all of the covariates are continuous. However, at the expense of additional notation, we can deal with the case where X has both continuous and discrete components. For example and for simplicity, let X_1 be a binary variable taking value on $\{0, 1\}$. Let the power series in the SLE be $R^K(x) = (1(x_1 = 0)\tilde{R}^K(x_{-1})', 1(x_1 = 1)\tilde{R}^K(x_{-1})')'$, where $\tilde{R}^K(x_{-1})$ is a vector of power functions in $x_{-1} \equiv (x_2, \dots, x_r)$. Note that this method is equivalent to sample splitting and we can generalize this approach to deal with discrete covariates in the model. Please see Comments 6–8 in Section 3.2 of DHL for more details.

3. Assumption 3.3(iv) requires that $m_z^d(y, x)$ is uniformly continuous on \mathcal{Y} w.r.t. the norm $d(y_1, y_2) = |F_{Y^d|\mathcal{C}}(y_1) - F_{Y^d|\mathcal{C}}(y_2)|$ for $d, z \in \{0, 1\}$. This is not restrictive in the following sense. For example, when $d = 0$ and if we define $\tilde{Y}_i = F_{Y^0|\mathcal{C}}(Y_i)$ and $\tilde{Y}_i(0) = F_{Y^0|\mathcal{C}}(Y_i(0))$. Then it is true that conditional on the group of compliers, $\tilde{Y}_i(0)$ will be a uniform distribution over $[0, 1]$, or equivalently, $P(\tilde{Y}_i(0) \leq a|\mathcal{C}) = E[1(\tilde{Y}_i(0) \leq a)|\mathcal{C}] = a$ for $a \in [0, 1]$. Let $\tilde{m}_z^0(a, x) = E[(1 - D) \cdot 1(\tilde{Y} \leq a)|Z = z, X = x]$, then Assumption 3.3(iv) is equivalent to assuming that $\tilde{m}_z^0(a, x)$ is uniformly continuous on $[0, 1]$ w.r.t. the Euclidean norm uniformly over $x \in \mathcal{X}$ for $z \in \{0, 1\}$. Similar argument applies to $d = 1$ case.

As we will show in the appendix, this condition (with other regularity conditions) implies that for $d \in \{0, 1\}$, $\sqrt{n}(\tilde{F}_{Y^d|\mathcal{C}}(y) - F_{Y^d|\mathcal{C}}(y))$ is stochastically equicontinuous w.r.t. the pseudo-metric $\nu_d(y_1, y_2) = |F_{Y^d|\mathcal{C}}(y_1) - F_{Y^d|\mathcal{C}}(y_2)|^{1/2}$. This result plays an important role in the proof where we show that $\hat{F}_{Y^d|\mathcal{C}}(y)$ and $\tilde{F}_{Y^d|\mathcal{C}}(y)$ are asymptotically equivalent.

4. As in HIR, Assumption 3.4 ensures the existence of a ν satisfying the conditions in Assumption 3.5. Note that our theory holds for any choices of a and ν such that $K = a \cdot n^\nu$ under the regulation of Assumption 3.5.

5. Also, as noted by Khan and Tamer (2010), the assumption that the instrument propensity

⁷However, we can still estimate the quantile functions $Q_{Y^0|\mathcal{C}}(t)$ and $Q_{Y^1|\mathcal{C}}(t)$ for $t \in [\epsilon, 1 - \epsilon]$ for any $0 < \epsilon < 1/2$ at the parametric rate given that the density functions are bounded away from zero on $y \in [Q_{Y^0|\mathcal{C}}(\epsilon), Q_{Y^0|\mathcal{C}}(1 - \epsilon)]$ and $y \in [Q_{Y^1|\mathcal{C}}(\epsilon), Q_{Y^1|\mathcal{C}}(1 - \epsilon)]$, respectively.

score is bounded away from zero and one plays an important role in determining the convergence rate of inverse probability weighted estimators.

3.2 Asymptotic Properties of $\widehat{F}_{Y^0|C}$ and $\widehat{F}_{Y^1|C}$

Define $\widehat{\mathbf{F}} = (\widehat{F}_{Y^0|C}, \widehat{F}_{Y^1|C})'$, $\mathbf{F} = (F_{Y^0|C}, F_{Y^1|C})'$ and $\eta = (y_1, y_2)' \in \mathcal{Y} \times \mathcal{Y}$.

Theorem 3.1 *Suppose Assumptions 2.1 and 3.1–3.5 hold. Then*

$$\sqrt{n}(\widehat{\mathbf{F}}(\cdot) - \mathbf{F}(\cdot)) \Rightarrow \Psi(\cdot),$$

where \Rightarrow denotes weak convergence, $\Psi(\cdot)$ is a two dimensional mean zero Gaussian process with covariance functions $\Omega(\eta_1, \eta_2) = E[\psi(W, \eta_1)\psi(W, \eta_2)']$ and $\psi(W, \eta) = (\psi_0(W, y_1), \psi_1(W, y_2))'$ with

$$\begin{aligned} \psi_0(W, y) = \frac{1}{\Gamma_0} & \left\{ \frac{Z[(D-1) \cdot 1(Y \leq y) + m_1^0(y, X) - F_{Y^0|C}(y)(D - \mu_1(X))]}{q(X)} \right. \\ & - \frac{(1-Z)[(D-1) \cdot 1(Y \leq y) + m_0^0(y, X) - F_{Y^0|C}(y)(D - \mu_0(X))]}{1-q(X)} \\ & \left. - m_1^0(y, X) + m_0^0(y, X) - F_{Y^0|C}(y)(\mu_1(X) - \mu_0(X)) \right\}, \\ \psi_1(W, y) = \frac{1}{\Gamma_1} & \left\{ \frac{Z[D \cdot 1(Y \leq y) - m_1^1(y, X) - F_{Y^1|C}(y)(D - \mu_1(X))]}{q(X)} \right. \\ & - \frac{(1-Z)[D \cdot 1(Y \leq y) - m_0^1(y, X) - F_{Y^1|C}(y)(D - \mu_0(X))]}{1-q(X)} \\ & \left. + m_1^1(y, X) - m_0^1(y, X) - F_{Y^1|C}(y)(\mu_1(X) - \mu_0(X)) \right\}, \end{aligned} \quad (3.1)$$

and the convergence is in $\ell^\infty(\mathcal{Y}) \times \ell^\infty(\mathcal{Y})$.^{8,9}

To show Theorem 3.1, we first show that $\sqrt{n}(\widetilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y))$ and $\sqrt{n}(\widetilde{F}_{Y^1|C}(y) - F_{Y^1|C}(y))$ are asymptotically equivalent to the following linear expressions respectively:

$$\begin{aligned} \sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widetilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_0(W_i, y) \right| &= o_p(1), \\ \sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widetilde{F}_{Y^1|C}(y) - F_{Y^1|C}(y)) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_1(W_i, y) \right| &= o_p(1). \end{aligned}$$

⁸The weak convergence is in the sense of Definition 1.3.3 of van der Vaart and Wellner (1996).

⁹ $\ell^\infty(\mathcal{Y})$ denotes the set of all uniformly bounded real functions on \mathcal{Y} .

Next, we show that $\mathcal{K}_0 = \{\psi_0(W, y) | y \in \mathcal{Y}\}$ and $\mathcal{K}_1 = \{\psi_1(W, y) | y \in \mathcal{Y}\}$ are Donsker classes. Note that the Cartesian product of two Donsker classes of functions is still a Donsker class as in van der Vaart (2000). By Donsker's theorem or functional central limit theorem, Theorem 3.1 holds for $\tilde{\mathbf{F}} = (\tilde{F}_{Y^0|C}, \tilde{F}_{Y^1|C})'$ in place of $\hat{\mathbf{F}}$. However, since ϕ is not Hadamard differentiable, we can not apply the functional delta method to $\phi(\tilde{F}_{Y^0|C})$ and $\phi(\tilde{F}_{Y^1|C})$ directly. To obtain the asymptotics of $\hat{F}_{Y^0|C}(y)$ and $\hat{F}_{Y^1|C}(y)$, we first prove that Theorem 3.1 holds for $\phi_1(\tilde{F}_{Y^0|C})$ and $\phi_1(\tilde{F}_{Y^1|C})$ in place of $\hat{\mathbf{F}}$ by applying Lemma 1.5.4 and Theorem 1.5.7 of van der Vaart and Wellner (1996). We then show that $\sqrt{n}(\hat{B}_0 - 1) = o_p(1)$ and $\sqrt{n}(\hat{B}_1 - 1) = o_p(1)$, meaning that $\phi_1(\tilde{F}_{Y^0|C})$ and $\phi_1(\tilde{F}_{Y^1|C})$ are first order equivalent to $\hat{F}_{Y^0|C}(y)$ and $\hat{F}_{Y^1|C}(y)$. These steps complete the proof for Theorem 3.1.

3.3 Asymptotic Properties of $\hat{Q}_{Y^0|C}$ and $\hat{Q}_{Y^1|C}$

We introduce additional conditions for the asymptotic properties of $\hat{Q}_{Y^0|C}$ and $\hat{Q}_{Y^1|C}$. Let $f_{Y^0|C}$ and $f_{Y^1|C}$ be the PDFs of $F_{Y^0|C}$ and $F_{Y^1|C}$.

Assumption 3.6 (*Densities of $Y(0)$ and $Y(1)$ for the Compliers*):

- (i) $f_{Y^0|C}$ and $f_{Y^1|C}$ are continuously differentiable of order 2.
- (ii) There exists $\delta > 0$ such that $f_{Y^0|C}(y) \geq \delta$ for $y \in [y_{0\ell}, y_{0u}]$ and $f_{Y^1|C}(y) \geq \delta$ for all $y \in [y_{1\ell}, y_{1u}]$.

Assumption 3.6 implies that $F_{Y^0|C}$ and $F_{Y^1|C}$ are strictly increasing on $[y_{0\ell}, y_{0u}]$ and $[y_{1\ell}, y_{1u}]$ respectively. Therefore, the quantiles functions, $Q_{Y^0|C}$ and $Q_{Y^1|C}$, are well-defined on $[0, 1]$. Define $\hat{\mathbf{Q}} = (\hat{Q}_{Y^0|C}, \hat{Q}_{Y^1|C})'$, $\mathbf{Q} = (Q_{Y^0|C}, Q_{Y^1|C})'$ and $\tau = (t_1, t_2)' \in [0, 1] \times [0, 1]$.

Theorem 3.2 *Suppose Assumptions 2.1, 3.1–3.6 hold. Then*

$$\sqrt{n}(\hat{\mathbf{Q}}(\cdot) - \mathbf{Q}(\cdot)) \Rightarrow \mathcal{Q}(\cdot),$$

where $\mathcal{Q}(\tau) = (\mathcal{Q}_0(t_1), \mathcal{Q}_1(t_2))$ is a two dimensional mean zero Gaussian process such that

$$\mathcal{Q}_0(t_1) \equiv -\frac{\Psi_0(Q_{Y^0|C}(t_1))}{f_{Y^0|C}(Q_{Y^0|C}(t_1))}, \quad \mathcal{Q}_1(t_2) \equiv -\frac{\Psi_1(Q_{Y^1|C}(t_2))}{f_{Y^1|C}(Q_{Y^1|C}(t_2))}.$$

The convergence is in $\ell^\infty([0, 1]) \times \ell^\infty([0, 1])$.

Given that the quantile map is Hadamard differentiable, Theorem 3.2 follows from the functional delta method. Theorem 3.2 extends Theorem 1 of Frölich and Melly (2013) which focuses on the asymptotics of pointwise quantiles to the asymptotics of whole quantile functions.

3.4 Our Method v.s. Rearrangement Method

The rearrangement approaches have been proposed by Chernozhukov, Fernández-Val and Galichon (2009, 2010) to address the problem of quantile crossing, i.e., the lack of monotonicity in the distribution or quantile estimators. Chernozhukov, Fernández-Val and Galichon’s (2010) rearrangement approach can be used to obtain CDF estimators that are monotonically increasing. However, it requires the pre-arranged quantile function estimators weakly converging to a stochastic process at the parametric rate, so that the CDF estimators will converge at the parametric rate accordingly. See Assumption 2 of Chernozhukov, Fernández-Val and Galichon (2010). In our framework, this assumption is equivalent to assuming that the densities $f_{Y^0|C}(y)$ and $f_{Y^1|C}(y)$ are bounded away from zero as in Assumption 3.6. This assumption, as mentioned above, is not necessary if we are interested in the asymptotics of CDF estimators only.¹⁰ Another advantage of our method over theirs is that our estimator is easier to implement. To be fair, Chernozhukov, Fernández-Val and Galichon’s (2010) rearrangement approach can be applied to cases where the underlying quantile functions are not monotonically increasing, but ours cannot. Similar to Chernozhukov, Fernández-Val and Galichon (2009), our method can be extended to construct monotonically increasing estimators for univariate and multivariate target functions that are monotonically increasing. The generalization is straightforward, but the derivation of these results is beyond the scope of this paper and we leave it for future studies.

4 Simulating the Limiting Processes

4.1 Estimating Conditional Quantities

Before introducing the simulation-based method, we construct uniformly consistent estimators for $m_z^d(y, x)$ that is monotonically increasing in y for any given x for $d, z \in \{0, 1\}$. The uniformly

¹⁰Of course, if we are interested in the asymptotics of whole quantile functions as in Section 3.3, this assumption is needed.

consistent estimators for $\mu_0(x)$ and $\mu_1(x)$ are constructed in the same way. These estimators play important roles in the simulation-based method.

We define the following series estimators:

$$\begin{aligned}
\tilde{m}_0^0(y, x) &= \left[\frac{1}{n} \sum_{i=1}^n \frac{(1 - Z_i)(1 - D_i) \cdot 1(Y_i \leq y)}{1 - \hat{q}(X_i)} R^K(X_i) \right]' \hat{\Xi}(x), \\
\tilde{m}_1^0(y, x) &= \left[\frac{1}{n} \sum_{i=1}^n \frac{Z_i(1 - D_i) \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} R^K(X_i) \right]' \hat{\Xi}(x), \\
\tilde{m}_0^1(y, x) &= \left[\frac{1}{n} \sum_{i=1}^n \frac{(1 - Z_i)D_i \cdot 1(Y_i \leq y)}{1 - \hat{q}(X_i)} R^K(X_i) \right]' \hat{\Xi}(x), \\
\tilde{m}_1^1(y, x) &= \left[\frac{1}{n} \sum_{i=1}^n \frac{Z_i D_i \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} R^K(X_i) \right]' \hat{\Xi}(x), \\
\hat{\mu}_0(x) &= \hat{m}_0^1(\bar{y}, x), \quad \hat{\mu}_1(x) = \hat{m}_1^1(\bar{y}, x), \\
\hat{\Xi}(x) &= \left[\frac{1}{n} \sum_{i=1}^n R^K(X_i) R^K(X_i)' \right]^{-1} R^K(x),
\end{aligned} \tag{4.1}$$

where $R^K(x)$ is the same power series used in SLE estimator. Note that \tilde{m} 's are not necessarily non-decreasing. To restore the monotonicity, we apply ϕ_1 defined in (2.5) on them and define

$$\hat{m}_z^d(y, x) = \phi_1(\tilde{m}_z^d(\cdot, x))(y), \quad \text{for } d, z \in \{0, 1\}. \tag{4.2}$$

It is obvious to see that $\hat{m}_z^d(y, x)$'s are non-decreasing in y . We summarize the asymptotic properties of these estimators in the following lemma.

Lemma 4.1 *Suppose Assumptions 2.1 and 3.1–3.5 hold. Then*

$$\sup_{y \in \mathcal{Y}, x \in \mathcal{X}} \left\{ \sum_{d, z \in \{0, 1\}} \left| \hat{m}_z^d(y, x) - m_z^d(y, x) \right| + \sum_{z \in \{0, 1\}} \left| \hat{\mu}_z(x) - \mu_z(x) \right| \right\} = o_p(1). \tag{4.3}$$

Lemma 4.1 follows from $\sup_{y \in \mathcal{Y}, x \in \mathcal{X}} |\tilde{m}_z^d(y, x) - m_z^d(y, x)| = o_p(1)$ and $\sup_{z \in \mathcal{Z}} |\hat{m}_z^d(y, x) - m_z^d(y, x)| \leq \sup_{z \in \mathcal{Z}} |\tilde{m}_z^d(y, x) - m_z^d(y, x)|$ for all $x \in \mathcal{X}$. Note that the compactness of \mathcal{X} is needed to obtain the uniform result.

4.2 Simulating $\Psi(\zeta)$

Let $\{U_1, U_2, \dots\}$ be i.i.d. random variables with mean 0 and variance 1 which are independent of the sequence $\mathcal{W} \equiv W^\infty = \{W_1, W_2, \dots\}$. For all $\eta = (y_1, y_2)' \in \mathcal{Y} \times \mathcal{Y}$, define the simulated

stochastic processes as $\widehat{\Psi}^u(\eta) = (\widehat{\Psi}_0^u(y_1), \widehat{\Psi}_1^u(y_2))'$ and for all $y \in \mathcal{Y}$,

$$\begin{aligned} \widehat{\Psi}_0^u(y) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \cdot \widehat{\psi}_0(W_i, y), & \widehat{\Psi}_1^u(y) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \cdot \widehat{\psi}_1(W_i, y), \\ \widehat{\psi}_0(W_i, y) &= \frac{1}{\widehat{\Gamma}_0} \left\{ \frac{Z_i [(D_i - 1) \cdot 1(Y_i \leq y) + \widehat{m}_1^0(y, X_i) - \widehat{F}_{Y^0|C}(y)(D_i - \widehat{\mu}_1(X_i))]}{\widehat{q}(X_i)} \right. \\ &\quad - \frac{(1 - Z_i) [(D_i - 1) \cdot 1(Y_i \leq y) + \widehat{m}_0^0(y, X_i) - \widehat{F}_{Y^0|C}(y)(D_i - \widehat{\mu}_0(X_i))]}{1 - \widehat{q}(X_i)} \\ &\quad \left. - \widehat{m}_1^0(y, X_i) + \widehat{m}_0^0(y, X_i) - \widehat{F}_{Y^0|C}(y) [\widehat{\mu}_1(X_i) - \widehat{\mu}_0(X_i)] \right\}, \\ \widehat{\psi}_1(W_i, y) &= \frac{1}{\widehat{\Gamma}_1} \left\{ \frac{Z_i [D_i \cdot 1(Y_i \leq y) - \widehat{m}_1^1(y, X_i) - \widehat{F}_{Y^1|C}(y)(D_i - \widehat{\mu}_1(X_i))]}{\widehat{q}(X_i)} \right. \\ &\quad - \frac{(1 - Z_i) [D_i \cdot 1(Y_i \leq y) - \widehat{m}_0^1(y, X_i) - \widehat{F}_{Y^1|C}(y)(D_i - \widehat{\mu}_0(X_i))]}{1 - \widehat{q}(X_i)} \\ &\quad \left. + \widehat{m}_1^1(y, X_i) - \widehat{m}_0^1(y, X_i) - \widehat{F}_{Y^1|C}(y) [\widehat{\mu}_1(X_i) - \widehat{\mu}_0(X_i)] \right\}. \end{aligned} \quad (4.4)$$

Theorem 4.1 *Suppose Assumptions 2.1 and 3.1–3.5 hold. Then $\widehat{\Psi}^u(\cdot) \Rightarrow \Psi(\cdot)$ conditional on sample path \mathcal{W} with probability approaching 1 which is denoted by $\widehat{\Psi}^u(\cdot) \xrightarrow{P} \Psi(\cdot)$.¹¹*

Theorem 4.1 shows that we can approximate $\Psi(\cdot)$ by the simulated stochastic process $\widehat{\Psi}^u(\cdot)$. To show this, we first prove that $n^{-1/2} \sum_{i=1}^n U_i \cdot \widehat{\psi}(W_i, \cdot) \xrightarrow{P} \Psi(\cdot)$ by the conditional multiplier central limit theorem, or Theorem 2.9.6 of van der Vaart and Wellner (1996). Second, we prove that the effect of estimation errors of $\widehat{\psi}(W_i, \eta) = (\widehat{\psi}_0(W_i, y_1), \widehat{\psi}_1(W_i, y_2))'$ will disappear in the limit.

¹¹The conditional weak convergence is in the sense of Section 2.9 of van der Vaart and Wellner (1996) and Chapter 2 of Kosorok (2008). To be more specific, $\widehat{\Psi}^u \xrightarrow{P} \Psi$ in the metric space (\mathbb{D}, d) if and only if $\sup_{f \in BL_1} |E_u f(\widehat{\Psi}^u) - E f(\Psi)| \xrightarrow{P} 0$ and $E_u f(\widehat{\Psi}^u)^* - E_u f(\widehat{\Psi}^u)_* \xrightarrow{P} 0$, where the subscript u in E_u indicates conditional expectation over the weights U_i 's given the remaining data, BL_1 is the space of functions $f: \mathbb{D} \rightarrow \mathbb{R}$ with Lipschitz norm bounded by 1, and $f(\widehat{\Psi}^u)^*$ and $f(\widehat{\Psi}^u)_*$ denote respectively measurable majorants and minorants w.r.t. the joint data including the U_i 's. We use the notation $\widehat{\Psi}^u \xrightarrow{a.s.} \Psi$ to express the same thing except replacing all \xrightarrow{P} 's with $\xrightarrow{a.s.}$'s. Note that by Lemma 1.9.2 (ii) of van der Vaart and Wellner (1996), it is true that $\widehat{\Psi}_n^u \xrightarrow{P} \Psi$ if and only if every subsequence k_n of n has a further subsequence ℓ_n of k_n such that $\widehat{\Psi}_{\ell_n}^u \xrightarrow{a.s.} \Psi$.

4.3 Simulating $\mathcal{Q}(\tau)$

As shown in Theorem 3.2, $f_{Y^0|C}$ and $f_{Y^1|C}$ have to be estimated before proceeding to the simulation method for the quantile case. We construct uniformly consistent IPW kernel estimators for $f_{Y^0|C}$ and $f_{Y^1|C}$ in \mathcal{Y} which play important roles in the proposed simulated process for $\mathcal{Q}(\tau)$.

Let h denote the bandwidth, which depends on sample size n and $K(u)$ the kernel function. Define $\tilde{f}_{Y^0|C}(y)$ and $\tilde{f}_{Y^1|C}(y)$ as

$$\begin{aligned}\tilde{f}_{Y^0|C}(y) &= \frac{1}{nh} \sum_{i=1}^n \hat{\kappa}_{0,i} K\left(\frac{Y_i - y}{h}\right) \quad \text{for } y \in [y_{0\ell} + h, y_{0u} - h], \\ \tilde{f}_{Y^1|C}(y) &= \frac{1}{nh} \sum_{i=1}^n \hat{\kappa}_{1,i} K\left(\frac{Y_i - y}{h}\right) \quad \text{for } y \in [y_{1\ell} + h, y_{1u} - h],\end{aligned}$$

and define

$$\begin{aligned}\tilde{f}_{Y^0|C}(y) &= \begin{cases} \tilde{f}_{Y^0|C}(h) & \text{if } y \in [y_{0\ell}, y_{0\ell} + h), \\ \tilde{f}_{Y^0|C}(\bar{y} - h) & \text{if } y \in (y_{0u} - h, y_{0u}], \end{cases} \\ \tilde{f}_{Y^1|C}(y) &= \begin{cases} \tilde{f}_{Y^1|C}(h) & \text{if } y \in [y_{1\ell}, y_{1\ell} + h), \\ \tilde{f}_{Y^1|C}(\bar{y} - h) & \text{if } y \in (y_{1u} - h, y_{1u}]. \end{cases}\end{aligned}$$

It is well known that the density estimator $\tilde{f}_{Y^d|C}(y)$ is in general inconsistent around the boundary points $y_{d\ell}$ and y_{du} , $d \in \{0, 1\}$. Therefore, we follow Donald, Hsu and Barrett (2012) and accommodate $\tilde{f}_{Y^0|C}(y)$ and $\tilde{f}_{Y^1|C}(y)$ around the boundary points to obtain uniformly consistent estimators for $f_{Y^0|C}(y)$ and $f_{Y^1|C}(y)$. Under regularity conditions introduced below, it can be shown that $\tilde{f}_{Y^0|C}$ and $\tilde{f}_{Y^1|C}$ are uniformly consistent for $f_{Y^0|C}$ and $f_{Y^1|C}$ that are bounded away from 0. However, the same problem arises in that given the weights $\hat{\kappa}_{0,i}$ and $\hat{\kappa}_{1,i}$ could be negative in finite sample, $\tilde{f}_{Y^0|C}$ and $\tilde{f}_{Y^1|C}$ could be negative accordingly. For this reason, we introduce the following trimming method. Let $\{b_n : n \geq 1\}$ be a sequence of positive numbers that converges to zero, we define

$$\hat{f}_{Y^0|C}(y) = \max\{\tilde{f}_{Y^0|C}(y), b_n\}, \quad \hat{f}_{Y^1|C}(y) = \max\{\tilde{f}_{Y^1|C}(y), b_n\}. \quad (4.5)$$

Note that despite the fact that $\hat{f}_{Y^0|C}(y)$ defined in (4.5) is not integrated to 1, it is fine for our use because we only need an uniform consistent estimator for $f_{Y^0|C}(y)$. Same comment applies to $\hat{f}_{Y^1|C}(y)$.

Assumption 4.1 (*IPW Kernel Estimator*):

(i) Suppose $K(u)$ is non-negative and has support $[-1, 1]$. $K(u)$ is symmetric around 0 and is continuously differentiable of order 1. The bandwidth h satisfies that $h \rightarrow 0$, $nh^4 \rightarrow \infty$ and $nh/\log n \rightarrow \infty$ when $n \rightarrow \infty$.

(ii) $\{b_n : n \geq 1\}$ is a non-increasing sequence of positive numbers and $b_n \rightarrow 0$ as $n \rightarrow \infty$.

Lemma 4.2 *Suppose Assumptions 2.1, 3.1–3.6 and 4.1 hold. Then*

$$\sup_{y \in [y_{0\ell}, y_{0u}]} \left| \hat{f}_{Y^0|C}(y) - f_{Y^0|C}(y) \right| + \sup_{y \in [y_{1\ell}, y_{1u}]} \left| \hat{f}_{Y^1|C}(y) - f_{Y^1|C}(y) \right| = o_p(1). \quad (4.6)$$

Note that b_n converges to 0 and $\tilde{f}_{Y^0|C}(y)$ is uniformly bounded away from zero with probability approaching 1, so $\tilde{f}_{Y^0|C}(y)$ and $\hat{f}_{Y^0|C}(y)$ are asymptotically equivalent.

We define the simulated processes as $\mathcal{Q}^u(\tau) = (\mathcal{Q}_0^u(t_1), \mathcal{Q}_1^u(t_2))'$ with

$$\hat{\mathcal{Q}}_0^u(\tau) = -\frac{\hat{\Psi}_0^u(\hat{Q}_{Y^0|C}(t_1))}{\hat{f}_{Y^0|C}(\hat{Q}_{Y^0|C}(t_1))}, \quad \hat{\mathcal{Q}}_1^u(\tau) = -\frac{\hat{\Psi}_1^u(\hat{Q}_{Y^1|C}(t_2))}{\hat{f}_{Y^1|C}(\hat{Q}_{Y^1|C}(t_2))}.$$

Theorem 4.2 *Suppose Assumptions 2.1, 3.1–3.6 and 4.1 hold. Then $\hat{\mathcal{Q}}^u(\cdot) \xrightarrow{P} \mathcal{Q}(\cdot)$.*

Theorem 4.2 shows that $\hat{\mathcal{Q}}^u(\tau)$ can approximate $\mathcal{Q}(\tau)$ well. The results regarding quantile functions allow us to test for the Lorenz dominance relations between potential outcomes, or construct uniform confidence bands for the LQTE over a continuum of quantile indices.

5 Estimation among the Treated Compliers

In general, researchers are interested not only in the group of compliers, but also certain sub-population such as the treated compliers, e.g., Frölich and Lechner (2010), Hong and Nekipelov (2010), and DHL. In this section we extend our estimation results and the simulation method to the treated compliers.

Let $F_{Y^0|C}^t(y) = E[1(Y(0) \leq y)|C, D = 1]$ and $F_{Y^1|C}^t(y) = E[1(Y(1) \leq y)|C, D = 1]$ denote the CDFs of potential outcomes of the treated compliers. The following lemma, which is similar to Lemma 2.1, summarizes the identification results for the group of treated compliers.

Lemma 5.1 (Identification of $F_{Y_0|C}^t$ and $F_{Y_1|C}^t$): Under Assumption 2.1, $F_0^t(y|C)$ and $F_1^t(y|C)$ are identified by

$$\begin{aligned} F_{Y_0|C}^t(y) &= \frac{1}{\Gamma_0^t} E \left[q(X) \left(\frac{Z(D-1) \cdot 1(Y \leq y)}{q(X)} - \frac{(1-Z)(D-1) \cdot 1(Y \leq y)}{1-q(X)} \right) \right], \\ F_{Y_1|C}^t(y) &= \frac{1}{\Gamma_1^t} E \left[q(X) \left(\frac{ZD \cdot 1(Y \leq y)}{q(X)} - \frac{(1-Z)D \cdot 1(Y \leq y)}{1-q(X)} \right) \right], \\ \Gamma_0^t &= E \left[q(X) \left(\frac{Z(D-1)}{q(X)} - \frac{(1-Z)(D-1)}{1-q(X)} \right) \right], \\ \Gamma_1^t &= E \left[q(X) \left(\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)} \right) \right], \end{aligned} \quad (5.1)$$

and $\Gamma_0^t = \Gamma_1^t = E[D(1) - D(0)|D=1] \cdot E[q(X)] > 0$.

Note that under the group of compliers, $D = 1$ iff $Z = 1$. With suitable modification, we can see that $F_{Y_1|C}^t$ is identified by the ratio of the ATT of Z on $D \cdot 1(Y \leq y)$ over the ATT of Z on D .¹² Similarly, $F_{Y_0|C}^t$ is identified by the ratio of the ATT of Z on $(D-1) \cdot 1(Y \leq y)$ over the ATT of Z on $D-1$. Also, define

$$\begin{aligned} \kappa_0^t &= \frac{q(X)}{\Gamma_0^t} \left[\frac{Z(D-1)}{q(X)} - \frac{(1-Z)(D-1)}{1-q(X)} \right], \\ \kappa_1^t &= \frac{q(X)}{\Gamma_1^t} \left[\frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)} \right], \end{aligned} \quad (5.2)$$

we have $F_{Y_0|C}^t(y) = E[\kappa_0^t \cdot 1(Y \leq y)]$ and $F_{Y_1|C}^t(y) = E[\kappa_1^t \cdot 1(Y \leq y)]$ with $E[\kappa_0^t] = E[\kappa_1^t] = 1$.

Based on (5.1), the IPW estimators for $F_{Y_0|C}^t$, $F_{Y_1|C}^t$, Γ_0^t and Γ_1^t are:

$$\begin{aligned} \tilde{F}_{Y_0|C}^t(y) &= \frac{1}{n\hat{\Gamma}_0^t} \sum_{i=1}^n \hat{q}(X_i) \left[\frac{Z_i(D_i-1) \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} - \frac{(1-Z_i)(D_i-1) \cdot 1(Y_i \leq y)}{1-\hat{q}(X_i)} \right], \\ \tilde{F}_{Y_1|C}^t(y) &= \frac{1}{n\hat{\Gamma}_1^t} \sum_{i=1}^n \hat{q}(X_i) \left[\frac{Z_i D_i \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} - \frac{(1-Z_i) D_i \cdot 1(Y_i \leq y)}{1-\hat{q}(X_i)} \right], \\ \hat{\Gamma}_0^t &= \frac{1}{n} \sum_{i=1}^n \hat{q}(X_i) \left[\frac{Z_i(D_i-1)}{\hat{q}(X_i)} - \frac{(1-Z_i)(D_i-1)}{1-\hat{q}(X_i)} \right], \\ \hat{\Gamma}_1^t &= \frac{1}{n} \sum_{i=1}^n \hat{q}(X_i) \left[\frac{Z_i D_i}{\hat{q}(X_i)} - \frac{(1-Z_i) D_i}{1-\hat{q}(X_i)} \right]. \end{aligned} \quad (5.3)$$

¹²Note that Γ_0^t is not equal to the ATT of Z on D , but $\Gamma_0^t/E[q(X)]$ is. Similarly, $E \left[q(X) \left(\frac{ZD \cdot 1(Y \leq y)}{q(X)} - \frac{(1-Z)D \cdot 1(Y \leq y)}{1-q(X)} \right) \right] / E[q(X)]$ is the ATT of Z on $D \cdot 1(Y \leq y)$. Hence, the identification result of $F_{Y_1|C}^t(y)$ in (5.1) can be interpreted as of the ATT of Z on $D \cdot 1(Y \leq y)$ over the ATT of Z on D .

Similar to $\tilde{F}_{Y^0|C}(y)$ and $\tilde{F}_{Y^1|C}(y)$, $\tilde{F}_{Y^0|C}^t(y)$ and $\tilde{F}_{Y^1|C}^t(y)$ are not necessarily monotonically increasing nor bounded by 0 and 1. Hence, we define

$$\hat{F}_{Y^0|C}^t(y) = \phi(\tilde{F}_{Y^0|C}^t(y)), \quad \hat{F}_{Y^1|C}^t(y) = \phi(\tilde{F}_{Y^1|C}^t(y)). \quad (5.4)$$

Also, we estimate the quantile functions for the treated compliers by

$$\hat{Q}_{Y^0|C}^t(t) = \inf\{y : \hat{F}_{Y^0|C}^t(y) \geq t\}, \quad \hat{Q}_{Y^1|C}^t(t) = \inf\{y : \hat{F}_{Y^1|C}^t(y) \geq t\}. \quad (5.5)$$

Let $f_{Y^0|C}^t(y)$ and $f_{Y^1|C}^t(y)$ be the corresponding PDFs of $F_{Y^0|C}^t(y)$ and $F_{Y^1|C}^t(y)$. We modify Assumptions 3.1 and 3.6 for the treated case.

Assumption 5.1 (*Distributions of $Y(0)$ and $Y(1)$ for the Treated Compliers*):

1. *Conditional on the treated compliers, $Y(0)$ and $Y(1)$ have convex and compact supports $[y_{0\ell}, y_{0u}]$ and $[y_{1\ell}, y_{1u}]$. Let $\mathcal{Y} = [\min\{y_{0\ell}, y_{1\ell}\}, \max\{y_{0u}, y_{1u}\}]$ and WLOG assume that $\mathcal{Y} = [0, \bar{y}]$ with $\bar{y} < \infty$.*
2. *$F_{Y^0|C}^t(y)$ and $F_{Y^1|C}^t(y)$ are continuous on \mathcal{Y} with $F_{Y^0|C}^t(0) = F_{Y^1|C}^t(0) = 0$.*
3. *$f_{Y^0|C}^t(y)$ and $f_{Y^1|C}^t(y)$ are continuously differentiable of order 2 and bounded away from 0 on $[y_{0\ell}, y_{0u}]$ and $[y_{1\ell}, y_{1u}]$, respectively.*

Note that the first two parts of Assumption 5.1 are stronger than necessary regarding the asymptotic results of the CDF estimators and it can be relaxed as in Assumption 3.1.

Define $\hat{\mathbf{F}}^t = (\hat{F}_{Y^0|C}^t, \hat{F}_{Y^1|C}^t)'$, $\mathbf{F}^t = (F_{Y^0|C}^t, F_{Y^1|C}^t)'$, $\hat{\mathbf{Q}}^t = (\hat{Q}_{Y^0|C}^t, \hat{Q}_{Y^1|C}^t)'$, and $\mathbf{Q}^t = (Q_{Y^0|C}^t, Q_{Y^1|C}^t)'$.

Theorem 5.1 *Suppose Assumptions 2.1, 3.2–3.5 and 5.1 hold. Then*

$$\sqrt{n}(\hat{\mathbf{F}}^t(\cdot) - \mathbf{F}^t(\cdot)) \Rightarrow \Psi^t(\cdot), \quad \sqrt{n}(\hat{\mathbf{Q}}^t(\cdot) - \mathbf{Q}^t(\cdot)) \Rightarrow \mathcal{Q}^t(\cdot),$$

where $\Psi^t(\cdot)$ is a two dimensional mean zero Gaussian process with covariance functions $\Omega^t(\eta_1, \eta_2) =$

$E[\psi^t(W, \eta_1)\psi^t(W, \eta_2)']$, where $\psi(W, \eta) = (\psi_0^t(W, y_1), \psi_1^t(W, y_2))'$ with

$$\begin{aligned}\psi_0^t(W, y) &= \frac{q(X)}{\Gamma_0^t} \left\{ \frac{Z[(D-1) \cdot 1(Y \leq y) + m_1^0(y, X) - F_{Y^0|C}^t(y)(D - \mu_1(X))]}{q(X)} \right. \\ &\quad - \frac{(1-Z)[(D-1) \cdot 1(Y \leq y) + m_0^0(y, X) - F_{Y^0|C}^t(y)(D - \mu_0(X))]}{1-q(X)} \\ &\quad \left. + \frac{Z[-m_1^0(y, X) + m_0^0(y, X) - F_{Y^0|C}^t(y)(\mu_1(X) - \mu_0(X))]}{q(X)} \right\}, \\ \psi_1^t(W, y) &= \frac{q(X)}{\Gamma_1^t} \left\{ \frac{Z[D \cdot 1(Y \leq y) - m_1^1(y, X) - F_{Y^1|C}^t(y)(D - \mu_1(X))]}{q(X)} \right. \\ &\quad - \frac{(1-Z)[D \cdot 1(Y \leq y) - m_0^1(y, X) - F_{Y^1|C}^t(y)(D - \mu_0(X))]}{1-q(X)} \\ &\quad \left. + \frac{Z[m_1^1(y, X) - m_0^1(y, X) - F_{Y^1|C}^t(y)(\mu_1(X) - \mu_0(X))]}{q(X)} \right\}.\end{aligned}\tag{5.6}$$

Also, $\mathcal{Q}^t(\tau) = (\mathcal{Q}_0^t(t_1), \mathcal{Q}_1^t(t_2))$ is a two dimensional mean zero Gaussian process such that

$$\mathcal{Q}_0^t(t_1) \equiv -\frac{\Psi_0^t(q_{Y^0|C}^t(t_1))}{f_{Y^0|C}^t(q_{Y^0|C}^t(t_1))}, \quad \mathcal{Q}_1^t(t_2) \equiv -\frac{\Psi_1^t(q_{Y^1|C}^t(t_2))}{f_{Y^1|C}^t(q_{Y^1|C}^t(t_2))}.$$

We again apply the simulation method for uniform inference on $\widehat{F}_{Y^0|C}^t$, $\widehat{F}_{Y^1|C}^t$, $\widehat{Q}_{Y^0|C}^t$ and $\widehat{Q}_{Y^1|C}^t$. To start with, we construct the uniform consistent estimators for $f_{Y^0|C}^t(y)$ and $f_{Y^1|C}^t(y)$ as follows. Let

$$\begin{aligned}\hat{\kappa}_{0,i}^t &= \frac{\hat{q}(X_i)}{\widehat{\Gamma}_0^t} \left[\frac{Z_i(D_i - 1)}{\hat{q}(X)} - \frac{(1 - Z_i)(D_i - 1)}{1 - \hat{q}(X_i)} \right], \\ \hat{\kappa}_{1,i}^t &= \frac{\hat{q}(X_i)}{\widehat{\Gamma}_1^t} \left[\frac{Z_i D_i}{\hat{q}(X)} - \frac{(1 - Z_i) D_i}{1 - \hat{q}(X_i)} \right].\end{aligned}\tag{5.7}$$

Define $\tilde{f}_{Y^0|C}^t(y)$ and $\tilde{f}_{Y^1|C}^t(y)$ as

$$\begin{aligned}\tilde{f}_{Y^0|C}^t(y) &= \frac{1}{nh} \sum_{i=1}^n \hat{\kappa}_{0,i}^t K\left(\frac{Y_i - y}{h}\right) \quad \text{for } y \in [y_{0l} + h, y_{0u} - h], \\ \tilde{f}_{Y^1|C}^t(y) &= \frac{1}{nh} \sum_{i=1}^n \hat{\kappa}_{1,i}^t K\left(\frac{Y_i - y}{h}\right) \quad \text{for } y \in [y_{1l} + h, y_{1u} - h],\end{aligned}$$

and define

$$\begin{aligned} \tilde{f}_{Y^0|C}^t(y) &= \begin{cases} \tilde{f}_{Y^0|C}^t(h) & \text{if } y \in [y_{0l}, y_{0l} + h), \\ \tilde{f}_{Y^0|C}^t(\bar{y} - h) & \text{if } y \in (y_{0u} - h, y_{0u}], \end{cases} \\ \tilde{f}_{Y^1|C}^t(y) &= \begin{cases} \tilde{f}_{Y^1|C}^t(h) & \text{if } y \in [y_{1l}, y_{1l} + h), \\ \tilde{f}_{Y^1|C}^t(\bar{y} - h) & \text{if } y \in (y_{1u} - h, y_{1u}]. \end{cases} \end{aligned}$$

Like $\tilde{f}_{Y^0|C}$ and $\tilde{f}_{Y^1|C}$, $\tilde{f}_{Y^0|C}^t$ and $\tilde{f}_{Y^1|C}^t$ are uniformly consistent but possibly negative. We accommodate the problem by defining

$$\hat{f}_{Y^0|C}^t(y) = \max \{ \tilde{f}_{Y^0|C}^t(y), b_n \}, \quad \hat{f}_{Y^1|C}^t(y) = \max \{ \tilde{f}_{Y^1|C}^t(y), b_n \}. \quad (5.8)$$

The estimated influence functions are

$$\begin{aligned} &\hat{\psi}_0^t(W_i, y) \\ &= \frac{\hat{q}(X_i)}{\hat{\Gamma}_0^t} \left\{ \frac{Z_i [(D_i - 1) \cdot 1(Y_i \leq y) + \hat{m}_1^0(y, X_i) - \hat{F}_{Y^0|C}^t(y)(D_i - \hat{\mu}_1(X_i))]}{\hat{q}(X_i)} \right. \\ &\quad - \frac{(1 - Z_i) [(D_i - 1) \cdot 1(Y_i \leq y) + \hat{m}_0^0(y, X_i) - \hat{F}_{Y^0|C}^t(y)(D_i - \hat{\mu}_0(X_i))]}{1 - \hat{q}(X_i)} \\ &\quad \left. + \frac{Z_i [-\hat{m}_1^0(y, X_i) + \hat{m}_0^0(y, X_i) - \hat{F}_{Y^0|C}^t(y)(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))]}{\hat{q}(X_i)} \right\}, \end{aligned} \quad (5.9)$$

$$\begin{aligned} &\hat{\psi}_1^t(W_i, y) \\ &= \frac{\hat{q}(X_i)}{\hat{\Gamma}_1^t} \left\{ \frac{Z_i [D_i \cdot 1(Y_i \leq y) - \hat{m}_1^1(y, X_i) - \hat{F}_{Y^1|C}^t(y)(D_i - \hat{\mu}_1(X_i))]}{\hat{q}(X_i)} \right. \\ &\quad - \frac{(1 - Z_i) [D_i \cdot 1(Y_i \leq y) - \hat{m}_0^1(y, X_i) - \hat{F}_{Y^1|C}^t(y)(D_i - \hat{\mu}_0(X_i))]}{1 - \hat{q}(X_i)} \\ &\quad \left. + \frac{Z_i [\hat{m}_1^1(y, X_i) - \hat{m}_0^1(y, X_i) - \hat{F}_{Y^1|C}^t(y)(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))]}{\hat{q}(X_i)} \right\}. \end{aligned}$$

Accordingly, the simulated stochastic processes are defined as

$$\begin{aligned} \hat{\Psi}_0^{t,u}(y) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \cdot \hat{\psi}_0^t(W_i, y), & \hat{\Psi}_1^{t,u}(y) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \cdot \hat{\psi}_1^t(W_i, y), \\ \hat{\mathcal{Q}}_0^{t,u}(t) &= -\frac{\hat{\Psi}_0^{t,u}(\hat{q}_{Y^0|C}^t(t))}{\hat{f}_{Y^0|C}^t(\hat{q}_{Y^0|C}^t(t))}, & \hat{\mathcal{Q}}_1^{t,u}(t) &= -\frac{\hat{\Psi}_1^{t,u}(\hat{q}_{Y^1|C}^t(t))}{\hat{f}_{Y^1|C}^t(\hat{q}_{Y^1|C}^t(t))}. \end{aligned} \quad (5.10)$$

Similar to Theorem 4.1 and Theorem 4.2, we can show that $\Psi^t(\cdot)$ and $\mathcal{Q}^t(\cdot)$ can be approximated by $\hat{\Psi}^{t,u}(\cdot) = (\hat{\Psi}_0^{t,u}(\cdot), \hat{\Psi}_1^{t,u}(\cdot))'$ and $\hat{\mathcal{Q}}^{t,u}(\cdot) = (\hat{\mathcal{Q}}_0^{t,u}(\cdot), \hat{\mathcal{Q}}_1^{t,u}(\cdot))'$ well and we omit the formal statement.

Note that our results allow us to define the LQTT which is not considered in Frölich and Melly (2013). To be more specific, for a quantile index $\tau \in [0, 1]$, LQTT(τ) is defined as $Q_{Y^1|C}^t(\tau) - Q_{Y^0|C}^t(\tau)$ which can be estimated by $\widehat{Q}_{Y^1|C}^t(\tau) - \widehat{Q}_{Y^0|C}^t(\tau)$. The asymptotic properties of LQTT estimator can be obtained based on our results above.

6 Monte Carlo Simulation

In this section, we explore the finite sample properties of the proposed estimators by a Monte Carlo simulation. The data generating process (DGP) adapted from DH and DHL is in the following:

$$\begin{aligned}
D(0) &= 0, \\
D(1) &= 1(X \leq b\epsilon + (1-b)U_D), \\
Y(0) &= 1(U_{Y_0} \leq X + \epsilon) \frac{U_{Y_0}^2}{X + \epsilon} + 1(U_{Y_0} > X + \epsilon)U_{Y_0}, \\
Y(1) &= 1(U_{Y_1} \leq 1 - (X + \epsilon)) \frac{U_{Y_1}^2}{1 - (X + \epsilon)} + 1(U_{Y_1} > 1 - (X + \epsilon))U_{Y_1}, \\
Z &= 1(q(X) > U_Z), \\
D &= Z \cdot D(1) + (1 - Z) \cdot D(0), \\
Y &= D \cdot Y(1) + (1 - D) \cdot Y(0),
\end{aligned}$$

where X , ϵ , U_D , U_{Y_0} , U_{Y_1} and U_Z are mutually independent uniform distributions over $[0, 1]$. As in DHL, $b \in [0, 1]$ governs whether the unconfoundedness assumption holds. When $b = 0$, D is independent of $(Y(0), Y(1))$ conditional on X . When $b > 0$, a larger b indicates that the unobserved confounding between D and $(Y(0), Y(1))$ is more severe. However, please note that the instrument Z is valid conditional on X for any b .

We consider three specifications of $q(x)$ similar to DHL:

$$(i) \ q(x) = 0.4, \quad (ii) \ q(x) = \frac{1}{1 + \exp(1 - x)}, \quad (iii) \ q(x) = \frac{1}{1 + \exp\left(1 - \frac{1}{1+x}\right)}.$$

The first benchmark case implies that the instrument Z is completely randomly assigned. The second linear specification acts as if it can be “correctly specified” by SLE using a power series of order 1. In contrast, the last specification is intended to be the “nonparametric” one in that

given any fixed order of power series, the SLE can only be seen as an approximation of the true specification.

The DGP is designed in a way to satisfy all assumptions mentioned above. For example, Assumption 2.1 is clearly satisfied since $D(0) = 0$. Assumptions 3.1, 3.2 and 3.6 are also true given that the uniform distribution has compact support and the density is bounded away from zero. Finally, by construction $q(x)$ is bounded away from 0 and 1 in all cases, and this is exactly what Assumption 3.4 required for.

We consider various settings in the simulation study. To begin with, the observations are set to 100, 400, 900 and 1600 so that the rate of convergence can be easily verified. The number of Monte Carlo repetitions is 1000 and the uniform coverage rates are calculated based on 1000 simulations. We let $b = 0.5$ and focus on quantiles $t \in [0.2, 0.8]$ with nominal significance level 10%. Next, we use a quadratic power series (i.e., the basis functions are 1, X and X^2) to estimate the instrument propensity score as well as other conditional quantities. A cubic power series is also utilized as a robustness check. In estimating $f_{Y^0|C}$ and $f_{Y^1|C}$, we use the Gaussian kernel along with Silverman's rule-of-thumb bandwidth. The estimated instrument propensity score is trimmed to lie in the interval $[0.005, 0.995]$ and the estimated PDFs are trimmed to be larger than 0.005.¹³ Finally, the multipliers U_i 's are drawn from the standard normal distribution.

Table 1 presents the simulation results regarding to the finite sample performance of $\widehat{F}_{Y^d|C}$, $\widehat{Q}_{Y^d|C}$ and the estimators for LQTE and LQTT. We report various measures for global accuracy such as the integrated bias (IBIAS), the root integrated mean squared error (RIMSE), the integrated mean absolute error (IMAE), the one-sided uniform coverage rate (1UCR) and the two-sided uniform coverage rate (2UCR). From Table 1, we see that the finite sample biases for all estimators decrease gradually with sample sizes and so do the RIMSEs and IMAEs. In addition, the magnitudes of RIMSEs reduce by a half when n increases from 100 to 400, and become one-third and one-fourth when n varies from 100 to 900 and 1600 respectively. This result suggests that our estimators indeed converge at the parametric rate of \sqrt{n} . Next, we validate the simulation method by comparing the 1UCR and 2UCR with nominal coverage rate 90%. Table 1 shows that almost all empirical coverage rates, regardless one- or two-sided, are

¹³The estimates are in fact not affected by the trimming in all cases considered.

close to the nominal one with moderate sample sizes.¹⁴ For these reasons, our methods should be able to provide reliable inference even when the sample size is relatively small.

7 Empirical Application

In the empirical study, we aim to investigate the distributional effects of fertility on family income, using twin births as an instrument to control for the endogeneity of fertility. The twins-based IV strategy first introduced by Rosenzweig and Wolpin (1980) has been adopted by, for example, Angrist and Evans (1998), Vere (2011) and Frölich and Melly (2013). Among these papers, this study is closest to Frölich and Melly (2013) in that they propose to assess the unconditional QTEs for the compliers, i.e., those who had planned to have only one child but ended up with several because of a twin birth. Despite sharing the same goal of distributional effects, we emphasize again that our methods can not only provide pointwise results but also conduct uniform inference over the whole family income distribution. In addition, we also consider the LQTT which is not present in their paper. Empirically speaking, our results allow us to assess the effects of more than one child on parents' income, both for the whole complier group and for the compliers who actually have two or more children.

The data we use come from the 1% and 5% Public Use Microdata Sample (PUMS). We consider the 1990 and 2000 censuses similar to Angrist and Evans (1998) and Frölich and Melly (2013), respectively. Suggested by previous research, the sample is limited to 21–35 years old married women who have at least one child. The outcome variable Y is the sum of the mother's and father's annual labor incomes. The treatment indicator D is equal to one if two or more children were born and zero otherwise. The instrument Z is equal to one if the first birth is a twin birth and zero otherwise. We consider covariates X such as mother's and father's age, race and educational level. All descriptive statistics are summarized in Table 3. Since the monotonicity assumption is automatically satisfied by the nature of twin birth, we can identify the treatment effects for the group of compliers up to 35% of the whole sample in 1990, and 40% in 2000, as shown in Table 3.

¹⁴Specification (ii) with size of 100 may be the only exception for a stringent standard that some deviations are over $\pm 7\%$. However, the result is as expected since in this case we overfit the true linear specification $q(x)$ with a quadratic power series. The deduction can be further confirmed by Table 2, where the overfitting becomes more severe if a cubic power series is adopted.

We describe the estimation details in the following. Similar to the Monte Carlo exercise, a quadratic power series is adopted to estimate the instrument propensity score and other conditional moments. For the IPW kernel estimation we again use the Gaussian kernel with Silverman’s bandwidth selection. The PDF estimates are trimmed to be larger than one over the square of the sample size.¹⁵ We restrict our attention to the quantiles $t \in [0.10, 0.95]$ at the 5% level of significance.¹⁶ The uniform confidence bands are calculated by 10000 simulations.

Figure 4 shows the LQTE, the pointwise confidence intervals and one- and two-sided confidence bands in 2000. According to the figure, we claim that the causal effect of fertility on family income does not exist for compliers below the 85th quantile. More importantly, in contrast to the findings of Frölich and Melly (2013), the treatment is not heterogeneous given that we cannot reject the null hypothesis of a constant treatment effect slightly below 5000 for all quantiles. The same argument applies to the LQTT in Figure 5. On the other hand, it is well known that the one-sided confidence bands, comparing to the two-sided ones, are more suitable for those who are interested in examining the positive effects. Thus, we present the one- and two-sided confidence bands for LQTE and LQTT respectively in Figures 6 and 7. For concreteness we concentrate on the case of $t \in [0.8, 0.95]$. As can be seen, we now conclude that having two or more children will increase family income for compliers above the 83th quantile, and 81th quantile for treated compliers. As a robustness check we also use the 1990 data. It turns out that the whole family income distribution is not affected by the presence of two or more children, as shown in Figures 8 and 9.

8 Conclusion

This paper proposes two-step estimators for the distribution functions of potential outcomes among the group of compliers in the IV treatment effect model. We first adopt the IPW estimators for the CDFs, and then modify the first-stage estimators for monotonicity and boundedness. Our monotonizing method is different from the existing ones and is easier to use. We then apply the inverse mapping to the distribution estimators to obtain the corresponding quantile functions, and show that the proposed estimators weakly converge to mean zero Gaus-

¹⁵No trimming is required for the instrument propensity score since it is bounded away from 0 and 1 per se.

¹⁶The family income is reported as zero for 4.6% in 1990 sample and 4.0% in 2000 sample.

sian processes. A simulation-based method taking into account the estimation effects is also introduced. These results allow us to conduct inference regarding the whole distribution functions or a continuum of quantile indices. We also extend the results to the treated compliers. Finally, the finite sample properties are demonstrated by a Monte Carlo simulation. We then apply the methods to the empirical study using U.S. census data. Our findings suggest that fertility will increase family income for high quantiles in 2000, whereas this effect disappears in 1990. Furthermore, there is no significant evidence that the treatment effect is heterogeneous for both 1990 and 2000.

APPENDIX

A Proof of Theorems

Proof of Theorem 3.1: We define additional notations. Let $M_0(y) = E[(D-1) \cdot 1(Y \leq y) | Z = 1] - E[(D-1) \cdot 1(Y \leq y) | Z = 0]$, $M_1(y) = E[D \cdot 1(Y \leq y) | Z = 1] - E[D \cdot 1(Y \leq y) | Z = 0]$. It is true that $F_{Y^0|C}(y) = M_0(y)/\Gamma_0$ and $F_{Y^1|C}(y) = M_1(y)/\Gamma_1$. $M_0(y)$ and $M_1(y)$ are estimated by

$$\begin{aligned}\widehat{M}_0(y) &= \frac{1}{n} \sum_{i=1}^n \frac{Z_i(D_i - 1) \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} - \frac{(1 - Z_i)(D_i - 1) \cdot 1(Y_i \leq y)}{1 - \hat{q}(X_i)} \\ \widehat{M}_1(y) &= \frac{1}{n} \sum_{i=1}^n \frac{Z_i D_i \cdot 1(Y_i \leq y)}{\hat{q}(X_i)} - \frac{(1 - Z_i) D_i \cdot 1(Y_i \leq y)}{1 - \hat{q}(X_i)}.\end{aligned}$$

Replacing Y with $(D-1) \cdot 1(Y \leq y)$ and $D \cdot 1(Y \leq y)$ respectively in DHL, and by the same argument of Lemma A1 of DH, we have

$$\begin{aligned}\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widehat{M}_0(y) - M_0(y)) - \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_{M_0}(W_i, y) - M_0(y)) \right| &= o_p(1), \\ \sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widehat{M}_1(y) - M_1(y)) - \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_{M_1}(W_i, y) - M_1(y)) \right| &= o_p(1),\end{aligned}\tag{A.1}$$

with

$$\begin{aligned}\psi_{M_0}(W, y) &= \frac{Z(D-1) \cdot 1(Y \leq y)}{q(X)} - \frac{(1-Z)(D-1) \cdot 1(Y \leq y)}{1-q(X)} + \left(\frac{m_1^0(y, X)}{q(X)} + \frac{m_0^0(y, X)}{1-q(X)} \right) (Z - q(X)), \\ \psi_{M_1}(W, y) &= \frac{ZD \cdot 1(Y \leq y)}{q(X)} - \frac{(1-Z)D \cdot 1(Y \leq y)}{1-q(X)} - \left(\frac{m_1^1(y, X)}{q(X)} + \frac{m_0^1(y, X)}{1-q(X)} \right) (Z - q(X)).\end{aligned}$$

By the same argument of Lemma A2 of DH, we have $\mathcal{K}_{M_0} = \{\psi_{M_0}(W, y) | y \in \mathcal{Y}\}$ and $\mathcal{K}_{M_1} = \{\psi_{M_1}(W, y) | y \in \mathcal{Y}\}$ are both \mathcal{P} -Donsker. This implies that

$$\begin{aligned}\sup_{y \in \mathcal{Y}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_{M_0}(W_i, y) - M_0(y)) \right| &= O_p(1), \\ \sup_{y \in \mathcal{Y}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_{M_1}(W_i, y) - M_1(y)) \right| &= O_p(1).\end{aligned}\tag{A.2}$$

(A.1) and (A.2) together imply that

$$\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widehat{M}_0(y) - M_0(y)) \right| = O_p(1), \quad \sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widehat{M}_1(y) - M_1(y)) \right| = O_p(1).\tag{A.3}$$

Similarly, we have

$$\begin{aligned}\sqrt{n}(\widehat{\Gamma}_0 - \Gamma_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_{\Gamma_0}(W_i) - \Gamma_0) + o_p(1), \\ \sqrt{n}(\widehat{\Gamma}_1 - \Gamma_1) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_{\Gamma_1}(W_i) - \Gamma_1) + o_p(1),\end{aligned}\tag{A.4}$$

with

$$\psi_{\Gamma_0}(W) = \psi_{\Gamma_1}(W) = \frac{ZD}{q(X)} - \frac{(1-Z)D}{1-q(X)} - \left(\frac{\mu_1(X)}{q(X)} + \frac{\mu_0(X)}{1-q(X)} \right) (Z - q(X)).$$

Recall that $\widetilde{F}_{Y^0|C}(y) = \widehat{M}_0(y)/\widehat{\Gamma}_0$ and $\widetilde{F}_{Y^1|C}(y) = \widehat{M}_1(y)/\widehat{\Gamma}_1$. Therefore, uniformly over $y \in \mathcal{Y}$,

$$\begin{aligned}& \sqrt{n}(\widetilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \\ &= \sqrt{n} \left(\frac{\widehat{M}_0(y)}{\widehat{\Gamma}_0} - \frac{M_0(y)}{\Gamma_0} \right) \\ &= \frac{1}{\Gamma_0} \left(\sqrt{n}(\widehat{M}_0(y) - M_0(y)) - \frac{M_0(y)}{\Gamma_0} \sqrt{n}(\widehat{\Gamma}_0 - \Gamma_0) \right) + o_p(1) \\ &= \frac{1}{\Gamma_0} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{M_0}(W_i, y) - M_0(y) - F_{Y^0|C}(y) (\psi_{\Gamma_0}(W_i) - \Gamma_0) + o_p(1) \right) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_0(W_i, y) + o_p(1),\end{aligned}$$

or equivalently

$$\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widetilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_0(W_i, y) \right| = o_p(1).\tag{A.5}$$

The argument for $\widetilde{F}_{Y^1|C}(y)$ is similar and we omit it for brevity. Note that $\mathcal{K}_0 = \{\psi_0(W, y) | y \in \mathcal{Y}\}$ is \mathcal{P} -Donsker. This result and (A.5) imply that (1) $\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\widetilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \right| = O_p(1)$ and (2) $\sqrt{n}(\widetilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y))$ is stochastically equicontinuous w.r.t. the pseudo-metric $\rho_0(y_1, y_2) = \left\{ E[\psi_0(W, y_1) - \psi_0(W, y_2)]^2 \right\}^{1/2}$.

Next, we claim that $\rho_0(y_1, y_2)$ is dominated by $\nu_0(y_1, y_2) = |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)|^{1/2}$ in that for all $y_1, y_2 \in \mathcal{Y}$, $\rho_0(y_1, y_2) \leq M \cdot \nu_0(y_1, y_2)$ for some $M > 0$. That $\rho_0(y_1, y_2)$ is dominated by $\nu_0(y_1, y_2)$ implies the stochastic equicontinuity of $\sqrt{n}(\widetilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y))$ w.r.t. the pseudo-metric $\nu_0(y_1, y_2)$. Note that

$$\begin{aligned}\rho_0(y_1, y_2) &= \left\{ E[\psi_0(W, y_1) - \psi_0(W, y_2)]^2 \right\}^{1/2} \\ &= \left\{ E \left[(\psi_{M_0}(W, y_1) - \psi_{M_0}(W, y_2)) - (F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)) \psi_{\Gamma_0}(W) \right]^2 \right\}^{1/2} \\ &\leq \left\{ 2E[\psi_{M_0}(W, y_1) - \psi_{M_0}(W, y_2)]^2 + 2E[(F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)) \psi_{\Gamma_0}(W)]^2 \right\}^{1/2} \\ &\leq \sqrt{2}(A_1(y_1, y_2) + A_2(y_1, y_2))\end{aligned}\tag{A.6}$$

where $A_1(y_1, y_2) = \{E[\psi_{M_0}(W, y_1) - \psi_{M_0}(W, y_2)]^2\}^{1/2}$ and $A_2(y_1, y_2) = \{E[(F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2))\psi_{\Gamma_0}(W)]^2\}^{1/2}$. The first inequality holds because $(a - b)^2 \leq 2(a^2 + b^2)$ and the last inequality holds because $(x + y) \leq (\sqrt{x} + \sqrt{y})^2$ for $x, y \geq 0$. To proceed, we first show that

$$\begin{aligned}
A_2(y_1, y_2) &= \left\{ E \left[(F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)) \psi_{\Gamma_0}(W) \right]^2 \right\}^{1/2} \\
&= |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)| \cdot \left\{ E[\psi_{\Gamma_0}(W)]^2 \right\}^{1/2} \\
&= C_2 \cdot |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)| \\
&\leq C_2 \cdot |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)|^{1/2} = C_2 \cdot \nu_0(y_1, y_2)
\end{aligned} \tag{A.7}$$

where the last equality holds because $C_2 = \{E[\psi_{\Gamma_0}(W)]^2\}^{1/2} < \infty$ by Assumption 3.4 and the last inequality holds because $a \leq \sqrt{a}$ for $0 \leq a \leq 1$. Next, WLOG we assume that $y_1 \geq y_2$. Note that

$$\begin{aligned}
&A_1(y_1, y_2) \\
&= \left\{ E[\psi_{M_0}(W, y_1) - \psi_{M_0}(W, y_2)]^2 \right\}^{1/2} \\
&= \left\{ E \left[\frac{Z(D-1) \cdot 1(y_2 < Y \leq y_1)}{q(X)} - \frac{(1-Z)(D-1) \cdot 1(y_2 < Y \leq y_1)}{1-q(X)} \right. \right. \\
&\quad \left. \left. + \left[\frac{m_1^0(y_1, X) - m_1^0(y_2, X)}{q(X)} + \frac{m_0^0(y_1, X) - m_0^0(y_2, X)}{1-q(X)} \right] (Z - q(X)) \right]^2 \right\} \\
&\leq \left\{ 4E \left[\frac{Z(D-1) \cdot 1(y_2 < Y \leq y_1)}{q(X)} \right]^2 + 4E \left[\frac{(1-Z)(D-1) \cdot 1(y_2 < Y \leq y_1)}{1-q(X)} \right]^2 \right. \\
&\quad \left. + 4E \left[\frac{m_1^0(y_1, X) - m_1^0(y_2, X)}{q(X)} (Z - q(X)) \right]^2 + 4E \left[\frac{m_0^0(y_1, X) - m_0^0(y_2, X)}{1-q(X)} (Z - q(X)) \right]^2 \right\}^{1/2} \\
&\leq 2 \left\{ \left\{ E \left[\frac{Z(D-1) \cdot 1(y_2 < Y \leq y_1)}{q(X)} \right]^2 \right\}^{1/2} + \left\{ E \left[\frac{(1-Z)(D-1) \cdot 1(y_2 < Y \leq y_1)}{1-q(X)} \right]^2 \right\}^{1/2} \right. \\
&\quad \left. + \left\{ E \left[\frac{m_1^0(y_1, X) - m_1^0(y_2, X)}{q(X)} (Z - q(X)) \right]^2 \right\}^{1/2} + \left\{ E \left[\frac{m_0^0(y_1, X) - m_0^0(y_2, X)}{1-q(X)} (Z - q(X)) \right]^2 \right\}^{1/2} \right\} \\
&\equiv 2(A_{11}(y_1, y_2) + A_{12}(y_1, y_2) + A_{13}(y_1, y_2) + A_{14}(y_1, y_2)),
\end{aligned} \tag{A.8}$$

where $A_{1j}(y_1, y_2)$ is defined term by term in the last inequality. We first claim that $A_{11}(y_1, y_2) \leq$

$C_{11}\nu_0(y_1, y_2)$ for some $C_{11} > 0$. Note that

$$\begin{aligned}
A_{11}(y_1, y_2) &= \left\{ E \left[\frac{Z(D-1) \cdot 1(y_2 < Y \leq y_1)}{q(X)} \right]^2 \right\}^{1/2} \\
&\leq \underline{q}^{-1} \cdot \left\{ E \left[\frac{Z(D-1) \cdot 1(y_2 < Y \leq y_1)}{q(X)} \right] \right\}^{1/2} \\
&= \underline{q}^{-1} \cdot \{ E[m_1^0(y_1, X) - m_1^0(y_2, X)] \}^{1/2} \\
&\leq \underline{q}^{-1} C \cdot |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)|^{1/2} \\
&= C_{11} \cdot \nu_0(y_1, y_2),
\end{aligned} \tag{A.9}$$

where the second line holds because $q^{-1}(X) \leq \underline{q}^{-1}$ by Assumption 3.4 and $[Z(D-1) \cdot 1(y_2 < Y \leq y_1)]^2 = Z(D-1) \cdot 1(y_2 < Y \leq y_1)$, the third line holds by the law of iterated expectations, and the fourth line holds because $|m_1^0(y_1, X) - m_1^0(y_2, X)| \leq C \cdot |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)|$ by Assumption 3.3(iv). Similar argument applies to $A_{12}(y_1, y_2)$. Next,

$$\begin{aligned}
A_{13}(y_1, y_2) &= \left\{ E \left[\frac{m_1^0(y_1, X) - m_1^0(y_2, X)}{q(X)} (Z - q(X)) \right]^2 \right\}^{1/2} \\
&\leq C \cdot |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)| \\
&\leq C \cdot |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)|^{1/2} \\
&= C_{13} \cdot \nu_0(y_1, y_2),
\end{aligned} \tag{A.10}$$

where the first inequality holds again by Assumption 3.3 and the second inequality holds because $|F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)| \leq 1$. Same argument applies to $A_{14}(y_1, y_2)$. (A.9) and (A.10) imply that $A_1(y_1, y_2) \leq C_1 \cdot \nu_0(y_1, y_2)$ for some $C_1 > 0$. This result and (A.7) imply that $\rho_0(y_1, y_2) \leq M \cdot \nu_0(y_1, y_2)$ for some $M > 0$.

Define $\ddot{F}_{Y^0|C}(y) = \phi_1(\tilde{F}_{Y^0|C}(y)) = \sup_{y' \leq y} \tilde{F}_{Y^0|C}(y')$. Let $\Psi_0(y)$ denote the first element of $\Psi(\eta)$ and $\Psi_0(y)$ is a zero mean Gaussian process with covariance kernel $\Omega_0(y_1, y_2) = E[\psi_0(W, y_1)\psi_0(W, y_2)]$. We first show that $\sqrt{n}(\ddot{F}_{Y^0|C}(\cdot) - F_{Y^0|C}(\cdot)) \Rightarrow \Psi_0(\cdot)$. By Lemma 1.5.4 and Theorem 1.5.7 of van der Vaart and Wellner (1996), it is sufficient to show that (a) the corresponding marginals of $\sqrt{n}(\ddot{F}_{Y^0|C}(\cdot) - F_{Y^0|C}(\cdot))$ and $\Psi_0(\cdot)$ are equal in law, (b) \mathcal{Y} is totally bounded w.r.t. $\nu_0(y_1, y_2)$ and (c) the stochastic equicontinuity of $\sqrt{n}(\ddot{F}_{Y^0|C}(\cdot) - F_{Y^0|C}(\cdot))$ with respect to $\nu_0(y_1, y_2)$. For (b), it is obvious that $\mathcal{Y} = [0, \bar{y}]$ is totally bounded w.r.t. $\nu_0^2(y_1, y_2) = |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)|$ and this implies the total boundedness of \mathcal{Y} w.r.t. $\nu_0(y_1, y_2)$. To check (c), we need to show that for all $\epsilon_1, \epsilon_2 > 0$, there exist N_1 large enough and $\delta_1 > 0$ such that for all $n \geq N_1$,

$$P \left(\sup_{\nu_0(y_1, y_2) \leq \delta_1} \left| \sqrt{n}(\ddot{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1)) - \sqrt{n}(\ddot{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq \epsilon_1 \right) \geq 1 - \epsilon_2. \tag{A.11}$$

Note that $\sup_{y \in \mathcal{Y}} |\sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2))| = O_p(1)$ which implies that for any $\epsilon_2 > 0$, there exist

$M > 0$ and N_m large enough such that for all $n \geq N_m$

$$P \left(\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq M \right) \geq 1 - \frac{\epsilon_2}{2}. \quad (\text{A.12})$$

Also, the stochastic equicontinuity of $\sqrt{n}(\tilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y))$ w.r.t. the pseudo-metric $\nu_0(y_1, y_2)$ means that for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$, there exists $\tilde{\delta} > 0$ small enough and $\tilde{N} > 0$ large enough such that for all $n > \tilde{N}$,

$$P \left(\sup_{\rho_0(y_1, y_2) \leq \tilde{\delta}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1)) - \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq \frac{\epsilon_1}{3} \right) \geq 1 - \frac{\epsilon_2}{2}. \quad (\text{A.13})$$

We pick N_m large enough such that

$$2M/\sqrt{N_m} < \tilde{\delta}^2. \quad (\text{A.14})$$

Therefore, for $y_1 \geq y_2$ with $\nu_0(y_1, y_2) > \tilde{\delta}$ and for $n > \sqrt{N_m}$, if $\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq M$, then

$$\begin{aligned} & \tilde{F}_{Y^0|C}(y_2) - \tilde{F}_{Y^0|C}(y_1) \\ &= (\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) - (\tilde{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1)) - (F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)) \\ &\leq 2M/\sqrt{n} - \tilde{\delta}^2 < 2M/\sqrt{N_m} - \tilde{\delta}^2 < 0, \end{aligned}$$

where the inequality holds because $\sqrt{n}(\tilde{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1)) \geq -M$, $\sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \leq M$ and $F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2) > \tilde{\delta}^2$. This implies that for all $y \in \mathcal{Y}$ and for $n > \sqrt{N_m}$, if $\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq M$,

$$\ddot{F}_{Y^0|C}(y) = \sup_{y' \leq y} \tilde{F}_{Y^0|C}(y') = \sup_{\{y': y' \leq y, \nu_0(y', y) \leq \tilde{\delta}\}} \tilde{F}_{Y^0|C}(y'). \quad (\text{A.15})$$

Also, for all $y \in \mathcal{Y}$ and for $n > \sqrt{N_m}$ with $\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq M$, we have

$$\begin{aligned} & 0 \leq \sqrt{n}(\ddot{F}_{Y^0|C}(y) - \tilde{F}_{Y^0|C}(y)) \\ &= \sqrt{n} \left((\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) - (\tilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \right) \\ &= \sqrt{n} \left(\sup_{\{y': y' \leq y, \nu_0(y', y) \leq \tilde{\delta}\}} \left((\tilde{F}_{Y^0|C}(y') - F_{Y^0|C}(y)) - (\tilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \right) \right) \\ &\leq \sup_{\{y': y' \leq y, \nu_0(y', y) \leq \tilde{\delta}\}} \sqrt{n} \left((\tilde{F}_{Y^0|C}(y') - F_{Y^0|C}(y')) - (\tilde{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \right) \\ &\leq \sup_{\nu_0(y_1, y_2) \leq \tilde{\delta}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1)) - \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right|. \end{aligned} \quad (\text{A.16})$$

where the second inequality holds because for $y' \leq y$, $F_{Y^0|C}(y') \leq F_{Y^0|C}(y)$. Therefore, when we pick

$N_1 = \max\{N_m, \tilde{N}\}$, $\epsilon_1, \epsilon_2, M, \delta_1 = \tilde{\delta}$ such that (A.12), (A.13) and (A.14) hold, then

$$P\left(\sup_{y \in \mathcal{Y}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq M \text{ and} \right. \\ \left. \sup_{\nu_0(y_1, y_2) \leq \tilde{\delta}} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1)) - \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2)) \right| \leq \frac{\epsilon_1}{3} \right) \geq 1 - \epsilon_2, \quad (\text{A.17})$$

where the last inequality holds from (A.12), (A.13) and the fact that for any two events A and B , $P(A \cap B) \geq 1 - P(A^c) - P(B^c)$. Let Ω_1 denote the event in (A.17) and it is true that $P(\Omega_1) \geq 1 - \epsilon_2$. Therefore, for each $\omega \in \Omega_1$, we have

$$\begin{aligned} & \sup_{\nu_0(y_1, y_2) \leq \delta_1} \left| \sqrt{n}(\ddot{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1))(\omega) - \sqrt{n}(\ddot{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2))(\omega) \right| \\ = & \sup_{\nu_0(y_1, y_2) \leq \delta_1} \left| \sqrt{n}(\ddot{F}_{Y^0|C}(y_1) - \tilde{F}_{Y^0|C}(y_1))(\omega) \right. \\ & \quad \left. + \left[\sqrt{n}(\tilde{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1))(\omega) - \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2))(\omega) \right] \right. \\ & \quad \left. - \sqrt{n}(\ddot{F}_{Y^0|C}(y_2) - \tilde{F}_{Y^0|C}(y_2))(\omega) \right| \quad (\text{A.18}) \\ \leq & \sup_{\nu_0(y_1, y_2) \leq \delta_1} \left| \sqrt{n}(\ddot{F}_{Y^0|C}(y_1) - \tilde{F}_{Y^0|C}(y_1))(\omega) \right| + \sup_{\nu_0(y_1, y_2) \leq \delta_1} \left| \sqrt{n}(\ddot{F}_{Y^0|C}(y_2) - \tilde{F}_{Y^0|C}(y_2))(\omega) \right| \\ & \quad + \sup_{\nu_0(y_1, y_2) \leq \delta_1} \left| \sqrt{n}(\tilde{F}_{Y^0|C}(y_1) - F_{Y^0|C}(y_1))(\omega) - \sqrt{n}(\tilde{F}_{Y^0|C}(y_2) - F_{Y^0|C}(y_2))(\omega) \right| \\ \leq & \frac{\epsilon_1}{3} + \frac{\epsilon_1}{3} + \frac{\epsilon_1}{3} = \epsilon_1, \end{aligned}$$

where the last line follows by (A.16) and (A.17). Therefore, (A.17) and (A.18) are sufficient to show (A.11). We last show (a) the corresponding marginals of $\sqrt{n}(\tilde{F}_{Y^0|C}(\cdot) - F_{Y^0|C}(\cdot))$ and $\Psi_0(\cdot)$ are equal in law. It is sufficient for us to show that for every $y \in \mathcal{Y}$, $\sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \xrightarrow{d} \Psi_0(y)$ because the argument for the joint convergence of for every finite subset $\{y_1, \dots, y_k\} \subset \mathcal{Y}$ can be done similarly. We apply the same argument in the proof of Proposition 1 of Barrett and Donald (2003) to show this. For any $y \in \mathcal{Y}$, let $\mathcal{Y}^* = \{a : a \leq y \text{ and } F_{Y^0|C}(y) = F_{Y^0|C}(a)\}$ where we suppress the dependence of \mathcal{Y}^* on y for notational simplicity. It is obvious that \mathcal{Y}^* is non-empty. Note that

$$\begin{aligned} \sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) &= \sup_{a \in [0, y]} \sqrt{n}(\tilde{F}_{Y^0|C}(a) - F_{Y^0|C}(y)) \\ &\geq \sup_{a \in \mathcal{Y}^*} \sqrt{n}(\tilde{F}_{Y^0|C}(a) - F_{Y^0|C}(y)) \\ &= \sup_{a \in \mathcal{Y}^*} \sqrt{n}(\tilde{F}_{Y^0|C}(a) - F_{Y^0|C}(a)) \\ &\xrightarrow{d} \sup_{a \in \mathcal{Y}^*} \Psi_0(a), \end{aligned} \quad (\text{A.19})$$

where the first line follows by definition of $\ddot{F}_{Y^0|C}(y)$, the second line follows by that $\mathcal{Y}^* \subseteq [0, y]$, then third line follows by that $F_{Y^0|C}(a) = F_{Y^0|C}(y)$ for all $a \in \mathcal{Y}^*$ and the last line follows by continuous

mapping theorem. Therefore, (A.19) implies that for any $c \in \mathbb{R}$,

$$\limsup P \left(\sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \leq c \right) \leq P \left(\sup_{a \in \mathcal{Y}^*} \Psi_0(a) \leq c \right) \quad (\text{A.20})$$

Next, following the same argument in their proof, we can show that for any $\epsilon > 0$ and for any $c \in \mathbb{R}$,

$$\liminf P \left(\sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \leq c \right) \geq P \left(\sup_{a \in \mathcal{Y}^*} \Psi_0(a) \leq c \right) - \epsilon. \quad (\text{A.21})$$

Then by (A.21) and by the fact that ϵ is arbitrary, we have for any $c \in \mathbb{R}$,

$$\liminf P \left(\sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \leq c \right) \geq P \left(\sup_{a \in \mathcal{Y}^*} \Psi_0(a) \leq c \right). \quad (\text{A.22})$$

(A.20) and (A.22) together imply that

$$P \left(\sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \leq c \right) = P \left(\sup_{a \in \mathcal{Y}^*} \Psi_0(a) \leq c \right). \quad (\text{A.23})$$

or equivalently $\sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \xrightarrow{d} \sup_{a \in \mathcal{Y}^*} \Psi_0(a)$. Then this part of proof can be done when we claim that $\sup_{a \in \mathcal{Y}^*} \Psi(a) = \Psi(y) + o_p(1)$. Recall that $\sqrt{n}(\tilde{F}_{Y^0|C}(a) - F_{Y^0|C}(y))$ is stochastically equicontinuous w.r.t. $\rho_0(y_1, y_2) = \{E[\psi_0(W, y_1) - \psi_0(W, y_2)]^2\}^{1/2}$ that is dominated by $\nu_0(y_1, y_2) = |F_{Y^0|C}(y_1) - F_{Y^0|C}(y_2)|^{1/2}$. For $a \in \mathcal{Y}^*$, $|F_{Y^0|C}(y) - F_{Y^0|C}(a)| = 0$ which implies that $\nu_0(y, a) = 0$ and $\rho_0(y, a) = 0$. $\rho_0(y, a) = 0$ iff $E[\psi_0(W, y) - \psi_0(W, a)]^2 = 0$ which further implies that $\text{Var}(\Psi_0(y) - \Psi_0(a)) = 0$ for all $a \in \mathcal{Y}^*$. By continuous mapping theorem, we have $\sup_{a \in \mathcal{Y}^*} \Psi_0(y) - \Psi_0(a) = o_p(1)$ and this is sufficient to show that $\sup_{a \in \mathcal{Y}^*} \Psi_0(a) \stackrel{d}{\sim} \Psi_0(y)$ in that $\sup_{a \in \mathcal{Y}^*} \Psi_0(a) - \Psi_0(y) = o_p(1)$. By Lemma 1.5.4 and Theorem 1.5.7 of van der Vaart and Wellner (1996), we have shown that $\sqrt{n}(\ddot{F}_{Y^0|C}(\cdot) - F_{Y^0|C}(\cdot)) \Rightarrow \Psi_0(\cdot)$.

We claim that $\sup_{y \in \mathcal{Y}} \sqrt{n}(\ddot{F}_{Y^0|C}(y) - 1) = o_p(1)$. Note that $\sup_{y \in \mathcal{Y}} \sqrt{n}(\ddot{F}_{Y^0|C}(y) - 1)$ is numerically identical to $\sup_{y \in \mathcal{Y}} \sqrt{n}(\tilde{F}_{Y^0|C}(y) - 1)$. By previous argument, we have $\sup_{y \in \mathcal{Y}} \sqrt{n}(\tilde{F}_{Y^0|C}(y) - 1) \stackrel{d}{\sim} \sqrt{n}(\tilde{F}_{Y^0|C}(\bar{y}) - 1)$. By definition of $\tilde{F}_{Y^0|C}(\bar{y})$, we have $\tilde{F}_{Y^0|C}(\bar{y}) = 1$. Therefore, $\sqrt{n}(\tilde{F}_{Y^0|C}(\bar{y}) - 1) = 0$ and $\sup_{y \in \mathcal{Y}} \sqrt{n}(\tilde{F}_{Y^0|C}(y) - 1) = o_p(1)$. Last, $\hat{F}_{Y^0|C}(y) = \tilde{F}_{Y^0|C}(y) / \hat{B}_0$ where $\hat{B}_0 = \sup_{y \in \mathcal{Y}} \tilde{F}_{Y^0|C}(y)$. Then uniformly over $y \in \mathcal{Y}$,

$$\begin{aligned} & \sqrt{n}(\hat{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) \\ &= \sqrt{n} \left(\frac{\ddot{F}_{Y^0|C}(y)}{\hat{M}_0} - F_{Y^0|C}(y) \right) \\ &= \sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) - F_{Y^0|C}(y) \sqrt{n}(\hat{M}_0 - 1) + o_p(1) \\ &= \sqrt{n}(\ddot{F}_{Y^0|C}(y) - F_{Y^0|C}(y)) + o_p(1), \end{aligned} \quad (\text{A.24})$$

where the second line follows from a mean-valued expansion and the fact that $\hat{M}_0 \xrightarrow{P} 1$ and the last lines follows that $0 \leq F_{Y^0|C}(y) \leq 1$ and $\sqrt{n}(\hat{M}_0 - 1) = \sup_{y \in \mathcal{Y}} \sqrt{n}(\tilde{F}_{Y^0|C}(y) - 1) + o_p(1)$. Therefore, (A.24)

and the fact that $\sqrt{n}(\ddot{F}_{Y^0|\mathcal{C}}(\cdot) - F_{Y^0|\mathcal{C}}(\cdot)) \Rightarrow \Psi_0(\cdot)$ imply that $\sqrt{n}(\widehat{F}_{Y^0|\mathcal{C}}(\cdot) - F_{Y^0|\mathcal{C}}(\cdot)) \Rightarrow \Psi_0(\cdot)$. The proof for $\widehat{F}_{Y^1|\mathcal{C}}(\cdot)$ is similar and we omit it for brevity. These complete the proof of Theorem 3.1. ■

Proof of Theorem 3.2: The result follows when we apply the functional delta method to the quantile map, e.g., Section 3.9 of van der Vaart and Wellner (1996). ■

Proof of Theorem 4.1: The proofs are similar to those for Lemma 4.2 of DH, so we omit them for brevity. ■

Proof of Theorem 4.2: The proofs are similar to those for Theorem 4.5 of DH, so we omit them for brevity. ■

Proof of Theorem 5.1: The proofs are similar to those for Theorem 3.1 and Theorem 3.2, so we omit them for brevity. ■

B Proof of Lemmas

Proof of Lemma 2.1: The proof is an application of Theorem 3.1 of Abadie (2003) and can be done by defining $g(Y, D, X) = 1(Y \leq y)$ for each $y \in \mathcal{Y}$. ■

Proof of Lemma 4.1: The proofs are similar to those for Lemma 4.1 of DH, so we omit them for brevity. ■

Proof of Lemma 4.2: By the same arguments for Lemma 4.4 of DH, we can show that

$$\sup_{y \in [y_{0\ell}, y_{0u}]} \left| \tilde{f}_{Y^0|\mathcal{C}}(y) - f_{Y^0|\mathcal{C}}(y) \right| = o_p(1). \quad (\text{B.1})$$

Given that b_n converges to 0, $\inf_{y \in [y_{0\ell}, y_{0u}]} f_{Y^0|\mathcal{C}}(y) \geq \delta$ by Assumption 3.6(ii) and (B.1), we have $\inf_{y \in [y_{0\ell}, y_{0u}]} \tilde{f}_{Y^0|\mathcal{C}}(y) \geq b_n$ with probability one. This implies that

$$\sup_{y \in [y_{0\ell}, y_{0u}]} |\tilde{f}_{Y^0|\mathcal{C}}(y) - \hat{f}_{Y^0|\mathcal{C}}(y)| = o_p(1). \quad (\text{B.2})$$

(B.1), (B.2) and the triangular inequality show Lemma 4.2. ■

Proof of Lemma 5.1: We replace Y with $(D - 1) \cdot 1(Y \leq y)$ in (2) of DHL. To be more specific, we let $W(z) \equiv D(z)(D(z) - 1)1(Y(1) \leq y) + (1 - D(z))(D(z) - 1)1(Y(0) \leq y)$ and it is true that $W(z) \equiv (D(z) - 1)1(Y(0) \leq y)$ because $D(z)(D(z) - 1) = 0$ and $(1 - D(z))(D(z) - 1) = -(1 - D(z))^2 = -(1 - D(z))$. Let $W \equiv W(Z) = ZW(1) + (1 - Z)W(0)$ and it is easy to see that $W = (D - 1) \cdot 1(Y(0) \leq y) = (D - 1) \cdot 1(Y \leq y)$. Then we have

$$\begin{aligned} E[W(1) - W(0)|Z = 1] &= E\left[\left((D(1) - 1) - (D(0) - 1)\right) \cdot 1(Y(0) \leq y)|Z = 1\right] \\ &= E\left[(D(1) - D(0)) \cdot 1(Y(0) \leq y)|Z = 1\right] \\ &= E[1(Y(0) \leq y)|\mathcal{C}, Z = 1] \cdot E[D(1) - D(0)|Z = 1]. \end{aligned}$$

Note that among the group of compliers, $D = 1$ iff $Z = 1$, so $F_{Y_0|C}^t(y) = E[1(Y(0) \leq y)|C, D = 1] = E[1(Y(0) \leq y)|C, Z = 1]$. Therefore, we have

$$E[1(Y(0) \leq y)|C, D = 1] = E[W(1) - W(0)|Z = 1]/E[D(1) - D(0)|Z = 1].$$

Therefore, if we treat Z as the primary treatment, and W and D as the outcomes of interest, then $E[W(1) - W(0)|Z = 1]$ is the ATT of Z on W , and $E[D(1) - D(0)|Z = 1]$ is the ATT of Z on D . Hence,

$$E[W(1) - W(0)|Z = 1] = E \left[q(X) \left(\frac{Z(D-1)1(Y \leq y)}{q(X)} - \frac{Z(D-1)1(Y \leq y)}{1-q(X)} \right) \right] / P(Z = 1),$$

$$E[D(1) - D(0)|Z = 1] = E \left[q(X) \left(\frac{Z(D-1)}{q(X)} - \frac{Z(D-1)}{1-q(X)} \right) \right] / P(Z = 1) = \Gamma_0^t / P(Z = 1),$$

and the identification of $E[1(Y(0) \leq y)|C, D = 1]$ follows. To show the identification of $E[1(Y(1) \leq y)|C, D = 1]$, we just need to apply the same argument to $D \cdot 1(Y \leq y)$ and we omit the proof for brevity of the paper. ■

Acknowledgements: We thank Blaise Melly for providing us with the data for the application. All the errors and omissions are our own responsibility. Yu-Chin Hsu gratefully acknowledges the research support from National Science Council Grant (101-2410-H-001-109-MY2) and Junior Research Investigators Award, Academia Sinica.

References

- Abadie, A. (2003): “Semiparametric Instrument Variable Estimation of Treatment Response Models,” *Journal of Econometrics*, **113**, 231–263.
- Abrevaya, J., Y.-C. Hsu and R. P. Lieli (2014): “Estimating Conditional Average Treatment Effects,” *Journal of Business and Economic Statistics*, forthcoming.
- Angrist, J. D. and W. N. Evans (1998): “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size,” *American Economic Review*, **88**, 450–477.
- Chernozhukov, V., I. Fernández-Val and A. Galichon (2009): “Improving Point and Interval Estimators of Monotone Functions by Rearrangement,” *Biometrika*, **96**, 559–575.
- Chernozhukov, V., I. Fernández-Val and A. Galichon (2010): “Quantile and Probability Curves without Crossing,” *Econometrica*, **78**, 1093–1125.
- Donald, S. G., Y.-C. Hsu, and G. F. Barrett (2012): “Incorporating Covariates in the Measurement of Welfare and Inequality: Methods and Applications,” *The Econometrics Journal*, **15**, C1–C30.
- Donald, S. G., and Y.-C. Hsu (2014): “Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models,” *Journal of Econometrics*, **178**, 383–397.
- Donald, S. G., Y.-C. Hsu, and R. P. Lieli (2014a): “Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT,” *Journal of Business and Economic Statistics*, **32**, 395–415.
- Donald, S. G. Y.-C. Hsu, and R. P. Lieli (2014b): “Inverse Probability Weighted Estimation of Local Average Treatment Effects: A Higher Order MSE Expansion,” *Statistics and Probability Letters*, **95**, 132–138.
- Firpo, S. (2007): “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, **75**, 259–276.
- Firpo, S. and C. Pinto (2015): “Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures,” *Journal of Applied Econometrics*, forthcoming.
- Frölich, M. (2007): “Nonparametric IV Estimation of Local Average Treatment Effects with Covariates,” *Journal of Econometrics*, **139**, 35–75.
- Frölich, M. and M. Lechner (2010): “Exploiting Regional Treatment Intensity for the Evaluation of Labor Market Policies,” *Journal of the American Statistical Association*, **105**, 1014–1029.

- Frölich, M. and B. Melly (2013): “Unconditional Quantile Treatment Effects under Endogeneity,” *Journal of Business and Economic Statistics*, **31**, 346–357.
- Hahn, J. (1998): “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects,” *Econometrica*, **66**, 315–331.
- Heckman, J., H. Ichimura, J. Smith and P. Todd (1998): “Characterizing Selection Bias Using Experimental Data,” *Econometrica*, **66**, 1017–1098.
- Heckman, J., H. Ichimura and P. Todd (1997): “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program,” *Review of Economic Studies*, **64**, 605–654.
- Heckman, J., H. Ichimura and P. Todd (1998): “Matching as an Econometric Evaluations Estimator,” *Review of Economic Studies*, **65**, 261–294.
- Hirano, K., G. Imbens and G. Ridder (2003): “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, **71**, 1161–1189.
- Hong, H. and D. Nekipelov (2010): “Semiparametric Efficiency in Nonlinear LATE Models,” *Quantitative Economics*, **1**, 279–304.
- Ichimura, H. and O. Linton (2005): “Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators,” in D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models*, 149–170. Cambridge University Press.
- Imbens, G. and J. Angrist (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, **62**, 467–475.
- Khan, S. and E. Tamer (2010): “Irregular Identification, Support Conditions, and Inverse Weight Estimation,” *Econometrica*, **6**, 2021–2042.
- Kosorok, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science and Business Media.
- Rosenbaum, P. and D. Rubin (1983): “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, **70**, 41–55.
- Rosenbaum, P. and D. Rubin (1985): “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score,” *Journal of American Statistical Association*, **79**, 516–524.
- Rosenzweig, M. R. and K. I. Wolpin (1980): “Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment,” *Econometrica*, **48** 227–240.
- van der Vaart, A. (2000): *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Application to Statistics*. New York: Springer-Verlag.

Vere, J. P. (2011): "Fertility and Parents' Labour Supply: New Evidence from US Census Data," *Oxford Economic Papers*, **63**, 211–231.

Table 1: Simulation Results – Quadratic Case

Object	$n = 100$					$n = 400$					$n = 900$					$n = 1600$				
	IBIAS	RIMSE	IMAE	1UCR	2UCR	IBIAS	RIMSE	IMAE	1UCR	2UCR	IBIAS	RIMSE	IMAE	1UCR	2UCR	IBIAS	RIMSE	IMAE	1UCR	2UCR
	Specification 1: $q(x) = 0.4$																			
$F_{Y^0 C}$	0.014	0.135	0.106	0.933	0.965	0.002	0.071	0.056	0.919	0.935	0.004	0.048	0.038	0.901	0.919	0.001	0.037	0.029	0.893	0.906
$F_{Y^1 C}$	-0.001	0.107	0.086	0.852	0.846	-0.001	0.054	0.043	0.897	0.888	0.001	0.035	0.028	0.885	0.904	0.001	0.027	0.021	0.896	0.888
$F_{Y^0 C}^t$	0.016	0.139	0.109	0.942	0.966	0.002	0.072	0.057	0.926	0.935	0.005	0.048	0.038	0.900	0.923	-0.001	0.037	0.029	0.902	0.905
$F_{Y^1 C}^t$	-0.004	0.106	0.085	0.863	0.860	-0.004	0.054	0.043	0.913	0.891	0.003	0.035	0.028	0.875	0.895	-0.001	0.027	0.021	0.905	0.889
$Q_{Y^0 C}$	-0.007	0.143	0.111	0.939	0.878	0.000	0.080	0.062	0.928	0.941	-0.005	0.055	0.043	0.932	0.941	0.000	0.042	0.032	0.905	0.924
$Q_{Y^1 C}$	0.001	0.111	0.089	0.868	0.856	0.001	0.058	0.046	0.892	0.917	-0.001	0.038	0.030	0.909	0.907	-0.001	0.029	0.023	0.881	0.887
$Q_{Y^0 C}^t$	-0.010	0.146	0.114	0.933	0.874	0.000	0.080	0.062	0.928	0.937	-0.005	0.055	0.043	0.937	0.941	0.002	0.042	0.033	0.898	0.921
$Q_{Y^1 C}^t$	0.003	0.110	0.088	0.872	0.869	0.004	0.058	0.046	0.890	0.920	-0.003	0.038	0.030	0.918	0.910	0.001	0.029	0.023	0.868	0.882
LQTE	0.008	0.183	0.145	0.897	0.917	0.001	0.098	0.078	0.926	0.942	0.004	0.067	0.053	0.915	0.931	-0.001	0.051	0.041	0.903	0.909
LQTT	0.013	0.186	0.147	0.897	0.917	0.004	0.099	0.078	0.922	0.939	0.001	0.067	0.053	0.914	0.931	-0.001	0.052	0.041	0.907	0.910
Specification 2: $q(x) = [1 + \exp(1 - x)]^{-1}$																				
$F_{Y^0 C}$	0.013	0.133	0.105	0.941	0.954	0.002	0.071	0.056	0.923	0.931	0.004	0.047	0.037	0.898	0.930	0.001	0.036	0.028	0.902	0.902
$F_{Y^1 C}$	-0.001	0.124	0.099	0.841	0.816	-0.002	0.060	0.048	0.874	0.880	0.001	0.039	0.031	0.899	0.902	0.001	0.030	0.024	0.883	0.881
$F_{Y^0 C}^t$	0.027	0.159	0.125	0.956	0.972	0.007	0.087	0.068	0.921	0.944	0.005	0.056	0.044	0.904	0.928	-0.004	0.043	0.034	0.917	0.904
$F_{Y^1 C}^t$	-0.004	0.118	0.094	0.865	0.847	-0.004	0.059	0.047	0.893	0.881	0.004	0.038	0.031	0.875	0.898	0.000	0.030	0.023	0.883	0.878
$Q_{Y^0 C}$	-0.007	0.142	0.110	0.935	0.873	0.000	0.079	0.061	0.924	0.932	-0.004	0.054	0.041	0.934	0.931	0.000	0.041	0.032	0.919	0.917
$Q_{Y^1 C}$	0.001	0.127	0.101	0.836	0.823	0.002	0.064	0.051	0.884	0.909	-0.001	0.041	0.033	0.912	0.917	-0.001	0.032	0.025	0.898	0.897
$Q_{Y^0 C}^t$	-0.019	0.162	0.127	0.926	0.842	-0.004	0.094	0.073	0.944	0.926	-0.005	0.064	0.049	0.944	0.938	0.004	0.049	0.038	0.911	0.933
$Q_{Y^1 C}^t$	0.003	0.120	0.096	0.866	0.840	0.004	0.063	0.050	0.880	0.916	-0.005	0.041	0.033	0.924	0.910	-0.001	0.032	0.025	0.888	0.889
LQTE	0.008	0.189	0.149	0.896	0.910	0.002	0.102	0.080	0.921	0.920	0.002	0.067	0.053	0.904	0.935	-0.001	0.053	0.041	0.891	0.894
LQTT	0.021	0.203	0.160	0.879	0.894	0.008	0.114	0.089	0.914	0.914	0.000	0.075	0.059	0.921	0.936	-0.005	0.059	0.046	0.931	0.920
Specification 3: $q(x) = [1 + \exp(1 - 1/(1 + x))]^{-1}$																				
$F_{Y^0 C}$	0.014	0.137	0.108	0.943	0.960	0.002	0.072	0.057	0.916	0.929	0.004	0.049	0.038	0.900	0.921	0.000	0.037	0.030	0.898	0.901
$F_{Y^1 C}$	0.000	0.100	0.080	0.852	0.851	-0.001	0.051	0.041	0.888	0.893	0.001	0.033	0.026	0.889	0.900	0.001	0.025	0.020	0.894	0.880
$F_{Y^0 C}^t$	0.014	0.132	0.105	0.943	0.954	0.000	0.069	0.055	0.925	0.931	0.004	0.046	0.036	0.896	0.918	-0.001	0.035	0.028	0.910	0.903
$F_{Y^1 C}^t$	-0.002	0.100	0.080	0.861	0.864	-0.002	0.051	0.041	0.896	0.893	0.002	0.033	0.026	0.882	0.894	-0.001	0.025	0.020	0.905	0.887
$Q_{Y^0 C}$	-0.007	0.145	0.113	0.936	0.875	0.000	0.081	0.063	0.935	0.946	-0.004	0.056	0.043	0.936	0.939	0.000	0.043	0.033	0.904	0.908
$Q_{Y^1 C}$	0.000	0.105	0.084	0.877	0.875	0.000	0.054	0.044	0.903	0.910	-0.001	0.036	0.028	0.904	0.907	-0.001	0.027	0.021	0.891	0.884
$Q_{Y^0 C}^t$	-0.009	0.141	0.110	0.936	0.888	0.003	0.077	0.060	0.927	0.936	-0.005	0.053	0.041	0.944	0.936	0.001	0.041	0.031	0.894	0.913
$Q_{Y^1 C}^t$	0.002	0.104	0.084	0.879	0.873	0.002	0.055	0.044	0.905	0.913	-0.002	0.036	0.028	0.919	0.904	0.002	0.027	0.021	0.874	0.885
LQTE	0.008	0.180	0.142	0.893	0.904	0.000	0.098	0.077	0.929	0.941	0.003	0.066	0.052	0.922	0.927	-0.001	0.051	0.040	0.908	0.911
LQTT	0.011	0.177	0.139	0.895	0.912	0.000	0.094	0.075	0.933	0.944	0.003	0.064	0.050	0.908	0.927	0.000	0.049	0.038	0.909	0.902

IBIAS = Integrated bias. RIMSE = Root integrated mean squared error. IMAE = Integrated mean absolute error. 1UCR = One-sided uniform coverage rate. 2UCR = Two-sided uniform coverage rate.

Table 2: Simulation Results – Cubic Case

Object	$n = 100$					$n = 400$					$n = 900$					$n = 1600$				
	IBIAS	RIMSE	IMAE	1UCR	2UCR	IBIAS	RIMSE	IMAE	1UCR	2UCR	IBIAS	RIMSE	IMAE	1UCR	2UCR	IBIAS	RIMSE	IMAE	1UCR	2UCR
	Specification 1: $q(x) = 0.4$																			
$F_{Y^0 C}$	0.014	0.136	0.107	0.930	0.960	0.001	0.071	0.056	0.918	0.935	0.004	0.048	0.038	0.904	0.921	0.001	0.037	0.029	0.890	0.905
$F_{Y^1 C}$	-0.001	0.108	0.086	0.854	0.846	-0.002	0.054	0.044	0.895	0.882	0.001	0.035	0.028	0.885	0.905	0.001	0.027	0.021	0.893	0.887
$F_{Y^0 C}^t$	0.017	0.142	0.111	0.937	0.963	0.003	0.072	0.057	0.926	0.936	0.005	0.049	0.038	0.896	0.921	-0.001	0.037	0.029	0.902	0.902
$F_{Y^1 C}^t$	-0.004	0.106	0.085	0.863	0.860	-0.004	0.054	0.043	0.913	0.891	0.003	0.035	0.028	0.875	0.895	-0.001	0.027	0.021	0.905	0.889
$Q_{Y^0 C}$	-0.008	0.144	0.112	0.938	0.878	0.001	0.080	0.062	0.929	0.937	-0.005	0.055	0.043	0.933	0.944	0.000	0.042	0.032	0.905	0.920
$Q_{Y^1 C}$	0.002	0.112	0.089	0.870	0.851	0.001	0.058	0.046	0.892	0.910	-0.001	0.038	0.030	0.911	0.909	-0.001	0.029	0.023	0.877	0.887
$Q_{Y^0 C}^t$	-0.011	0.149	0.115	0.930	0.872	0.000	0.081	0.063	0.934	0.935	-0.005	0.055	0.043	0.944	0.944	0.002	0.042	0.033	0.899	0.920
$Q_{Y^1 C}^t$	0.003	0.110	0.088	0.872	0.869	0.004	0.058	0.046	0.890	0.920	-0.003	0.038	0.030	0.918	0.910	0.001	0.029	0.023	0.868	0.882
LQTE	0.009	0.184	0.146	0.891	0.918	0.001	0.099	0.078	0.922	0.939	0.004	0.067	0.053	0.914	0.932	-0.001	0.051	0.041	0.904	0.906
LQTT	0.014	0.188	0.148	0.892	0.911	0.004	0.099	0.078	0.920	0.941	0.001	0.067	0.053	0.916	0.929	-0.001	0.052	0.041	0.904	0.909
Specification 2: $q(x) = [1 + \exp(1 - x)]^{-1}$																				
$F_{Y^0 C}$	0.013	0.134	0.106	0.939	0.953	0.002	0.071	0.056	0.920	0.926	0.004	0.047	0.037	0.898	0.931	0.001	0.036	0.028	0.902	0.900
$F_{Y^1 C}$	-0.002	0.125	0.100	0.840	0.801	-0.002	0.060	0.048	0.870	0.874	0.001	0.039	0.031	0.901	0.903	0.001	0.030	0.024	0.877	0.878
$F_{Y^0 C}^t$	0.028	0.162	0.128	0.947	0.978	0.007	0.087	0.068	0.919	0.939	0.005	0.056	0.044	0.905	0.930	-0.004	0.043	0.034	0.919	0.902
$F_{Y^1 C}^t$	-0.004	0.118	0.094	0.865	0.847	-0.004	0.059	0.047	0.893	0.881	0.004	0.038	0.031	0.875	0.898	0.000	0.030	0.023	0.883	0.878
$Q_{Y^0 C}$	-0.007	0.143	0.111	0.935	0.874	0.000	0.079	0.061	0.923	0.929	-0.004	0.054	0.042	0.936	0.930	0.000	0.041	0.032	0.916	0.918
$Q_{Y^1 C}$	0.003	0.128	0.102	0.828	0.810	0.002	0.064	0.051	0.882	0.900	-0.001	0.041	0.033	0.912	0.919	-0.001	0.032	0.025	0.892	0.901
$Q_{Y^0 C}^t$	-0.019	0.165	0.129	0.924	0.829	-0.004	0.095	0.073	0.947	0.924	-0.005	0.064	0.049	0.945	0.938	0.004	0.049	0.038	0.912	0.930
$Q_{Y^1 C}^t$	0.003	0.120	0.096	0.866	0.840	0.004	0.063	0.050	0.880	0.916	-0.005	0.041	0.033	0.924	0.910	-0.001	0.032	0.025	0.888	0.889
LQTE	0.010	0.191	0.150	0.886	0.906	0.002	0.102	0.080	0.915	0.917	0.002	0.067	0.053	0.904	0.935	-0.001	0.053	0.041	0.889	0.903
LQTT	0.021	0.205	0.161	0.878	0.888	0.008	0.114	0.089	0.910	0.915	0.000	0.075	0.059	0.921	0.936	-0.005	0.059	0.046	0.931	0.918
Specification 3: $q(x) = [1 + \exp(1 - 1/(1 + x))]^{-1}$																				
$F_{Y^0 C}$	0.015	0.138	0.109	0.936	0.955	0.002	0.072	0.057	0.917	0.927	0.004	0.049	0.038	0.898	0.918	0.001	0.038	0.030	0.897	0.904
$F_{Y^1 C}$	0.000	0.101	0.081	0.846	0.844	-0.001	0.051	0.041	0.883	0.890	0.001	0.033	0.026	0.892	0.897	0.001	0.025	0.020	0.895	0.884
$F_{Y^0 C}^t$	0.015	0.134	0.106	0.936	0.959	0.000	0.069	0.055	0.923	0.923	0.004	0.046	0.036	0.902	0.919	-0.001	0.035	0.028	0.912	0.897
$F_{Y^1 C}^t$	-0.002	0.100	0.080	0.861	0.864	-0.002	0.051	0.041	0.896	0.893	0.002	0.033	0.026	0.882	0.894	-0.001	0.025	0.020	0.905	0.887
$Q_{Y^0 C}$	-0.008	0.146	0.114	0.936	0.866	0.000	0.081	0.063	0.931	0.944	-0.004	0.056	0.043	0.931	0.940	0.000	0.043	0.033	0.901	0.910
$Q_{Y^1 C}$	0.001	0.105	0.084	0.875	0.868	0.000	0.055	0.044	0.901	0.911	-0.001	0.036	0.028	0.903	0.905	-0.001	0.027	0.021	0.893	0.885
$Q_{Y^0 C}^t$	-0.010	0.143	0.112	0.934	0.878	0.003	0.078	0.060	0.926	0.936	-0.005	0.053	0.041	0.942	0.939	0.001	0.041	0.031	0.893	0.914
$Q_{Y^1 C}^t$	0.002	0.104	0.084	0.879	0.873	0.002	0.055	0.044	0.905	0.913	-0.002	0.036	0.028	0.919	0.904	0.002	0.027	0.021	0.874	0.885
LQTE	0.009	0.181	0.142	0.883	0.904	0.000	0.098	0.077	0.931	0.938	0.003	0.066	0.052	0.920	0.923	-0.001	0.051	0.040	0.909	0.907
LQTT	0.012	0.178	0.140	0.892	0.911	0.000	0.095	0.075	0.926	0.940	0.003	0.064	0.050	0.906	0.924	0.000	0.049	0.038	0.907	0.899

IBIAS = Integrated bias. RIMSE = Root integrated mean squared error. IMAE = Integrated mean absolute error. 1UCR = One-sided uniform coverage rate. 2UCR = Two-sided uniform coverage rate.

Table 3: Descriptive Statistics

	1990			2000		
	All	Twin Birth	No Twin Birth	All	Twin Birth	No Twin Birth
Observations	602,767	6,524	596,243	573,437	8,569	564,868
Number of children	1.93	2.53	1.92	1.88	2.51	1.87
At least two children	64.92%	100.00%	64.52%	61.73%	100.00%	61.14%
Mother						
Age in years	29.79	29.99	29.79	30.06	30.45	30.05
Years of education	12.90	12.94	12.90	12.93	13.07	12.93
Black	6.37%	7.54%	6.36%	7.33%	7.99%	7.32%
Asian	3.23%	2.76%	3.23%	4.48%	3.38%	4.50%
Currently at work	57.74%	53.64%	57.79%	56.37%	51.69%	56.44%
Usual hours per week	24.55	23.01	24.56	25.16	23.10	25.19
Week worked	28.78	26.38	28.81	29.79	27.24	29.82
Wage or salary income last year	8,958	8,203	8,967	14,210	13,495	14,222
Father						
Age in years	32.54	32.81	32.54	32.93	33.33	32.92
Years of education	13.08	13.18	13.08	12.89	13.02	12.89
Black	6.70%	7.82%	6.68%	8.03%	9.32%	8.01%
Asian	2.89%	2.40%	2.90%	4.04%	3.36%	4.05%
Currently at work	89.73%	89.03%	89.74%	85.11%	85.38%	85.11%
Usual hours per week	43.97	44.06	43.97	43.86	44.00	43.86
Week worked	47.33	47.01	47.33	47.05	47.24	47.05
Wage or salary income last year	26,977	28,012	26,965	37,989	41,559	37,934
Parents						
Wage or salary income last year	35,935	36,215	35,932	52,200	55,054	52,156

Data from the 1% and 5% PUMS in 1990 and 2000. The sample consists of married mothers between 21 and 35 years of age with at least one child. Own calculations using the PUMS sample weights.

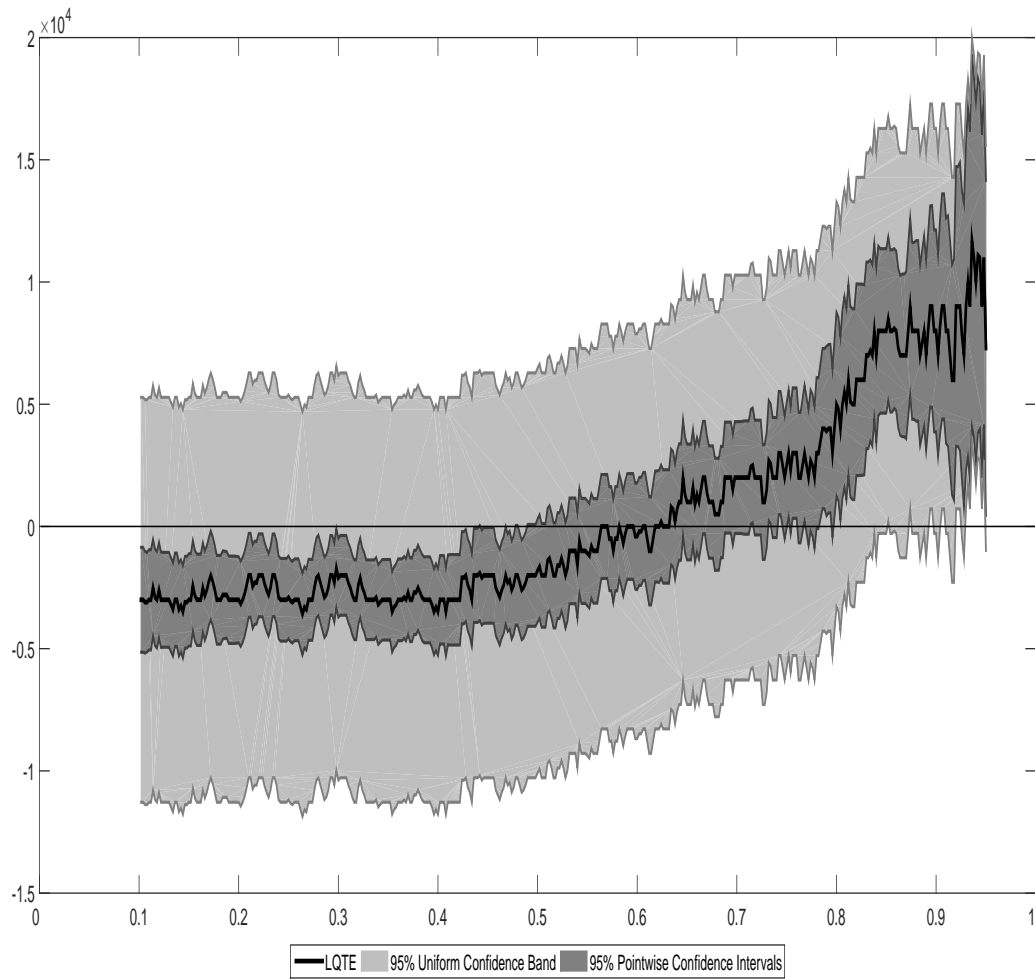


Figure 4: LQTE of Having More than One Child on Family Income in 2000

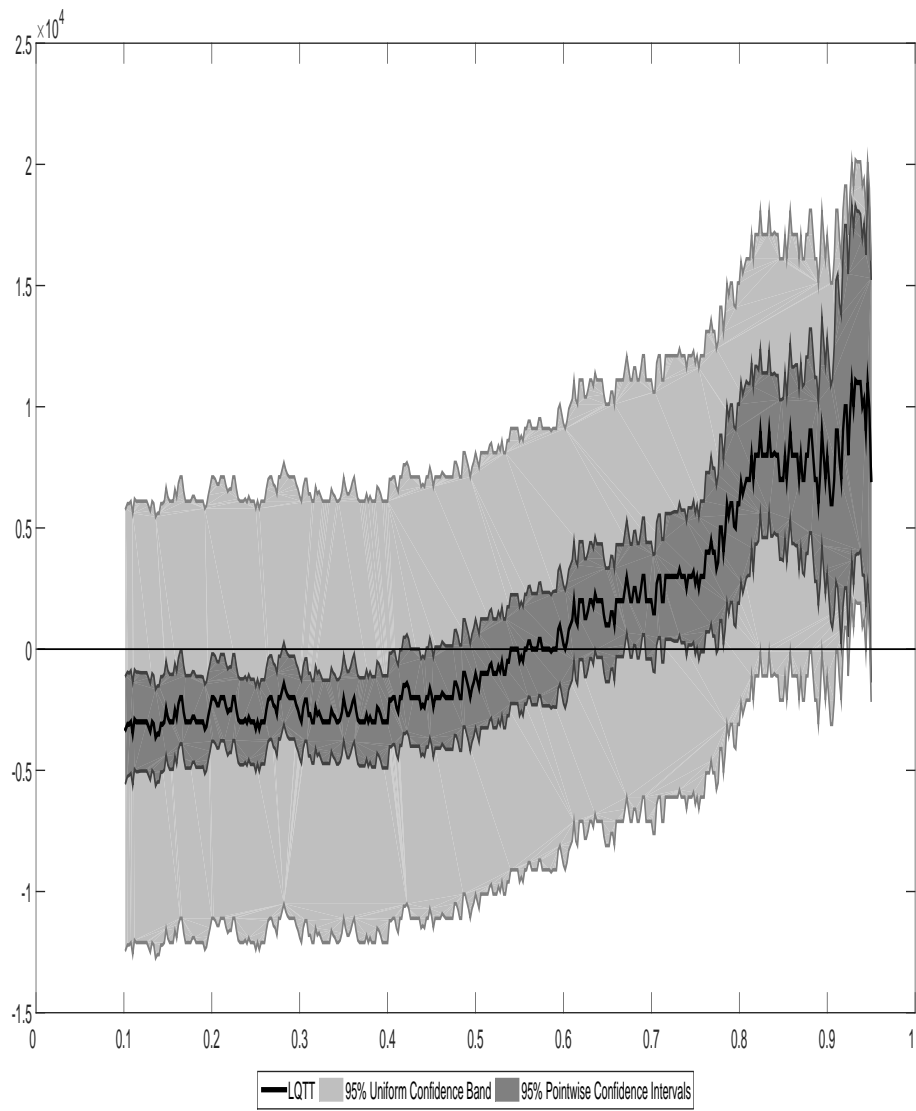


Figure 5: LQTT of Having More than One Child on Family Income in 2000

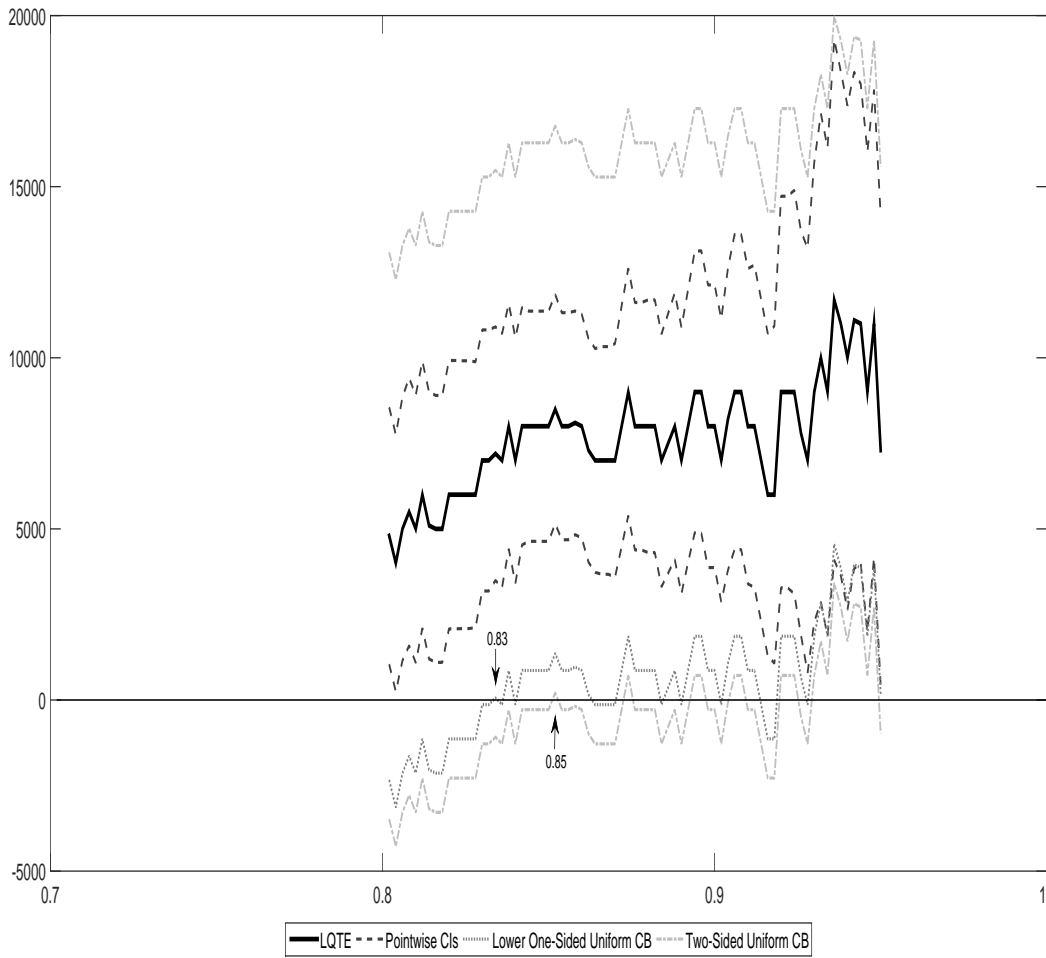


Figure 6: LQTE in 2000, Above the 8th Decile

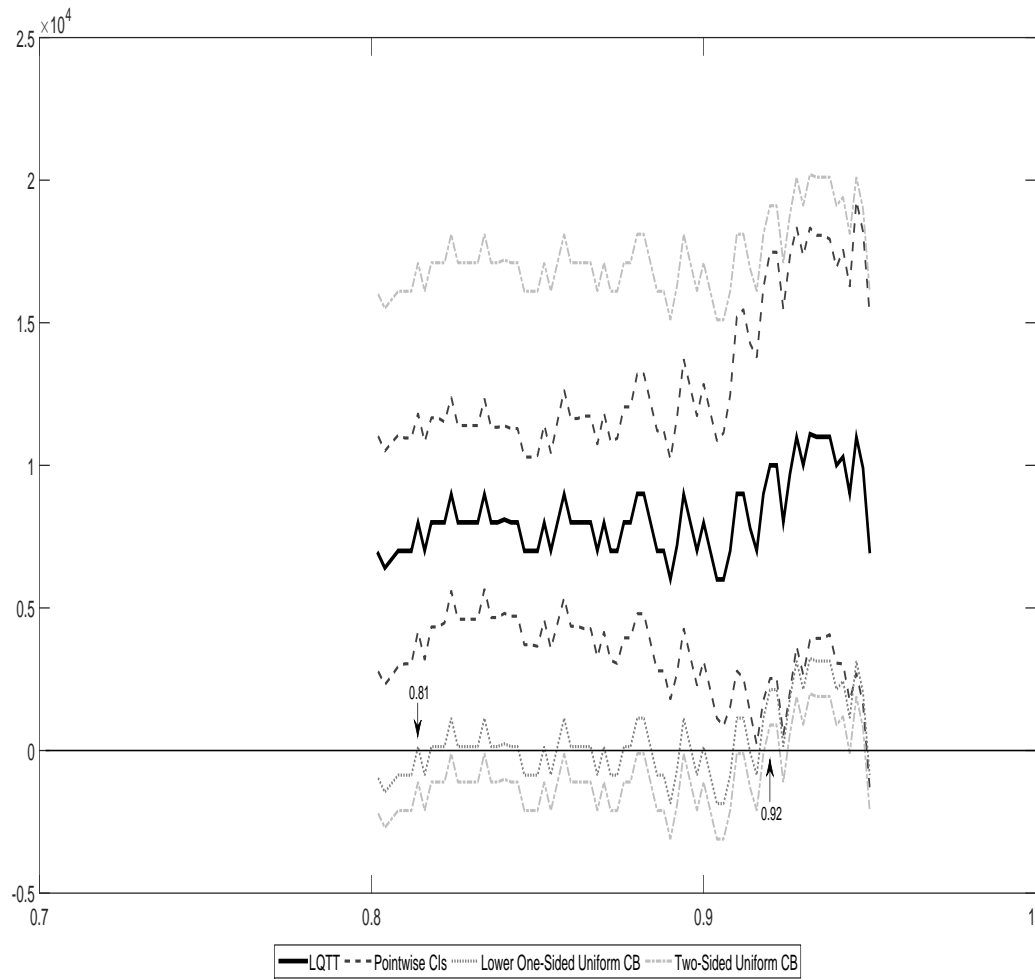


Figure 7: LQTT in 2000, Above the 8th Decile

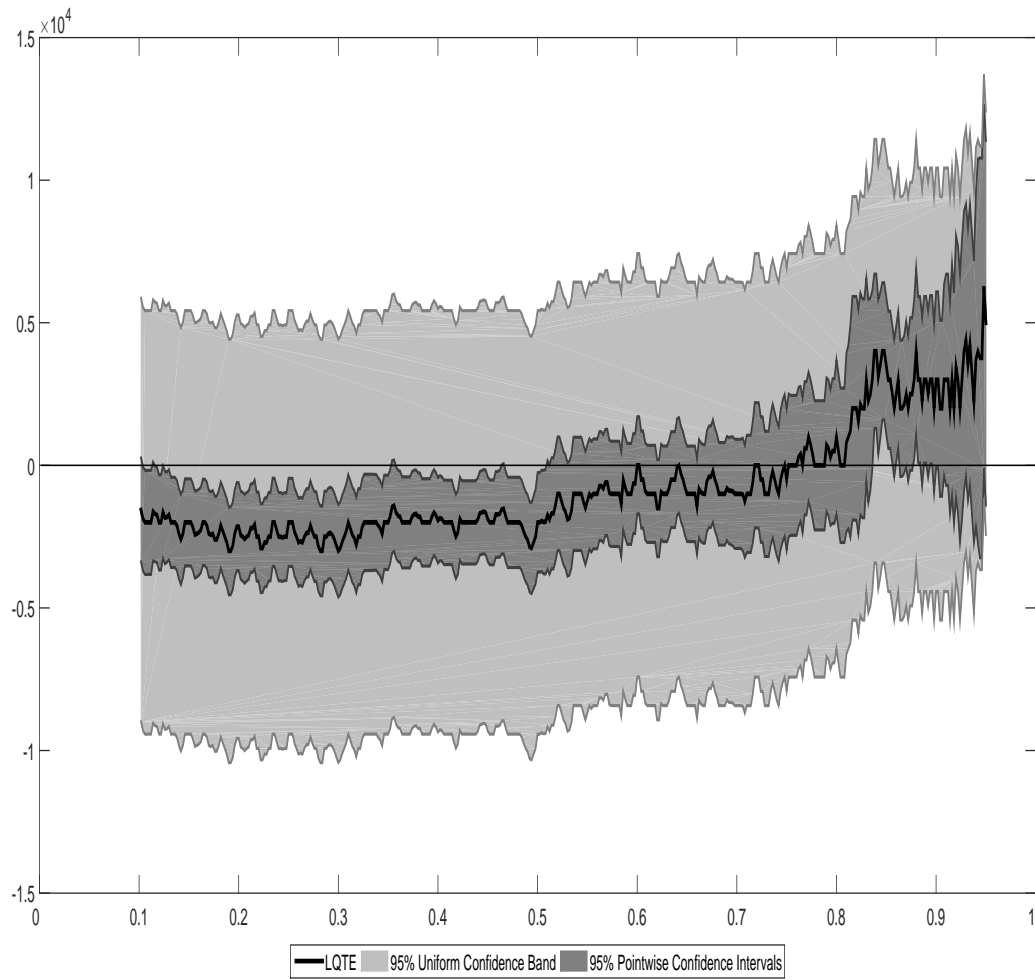


Figure 8: LQTE of Having More than One Child on Family Income in 1990

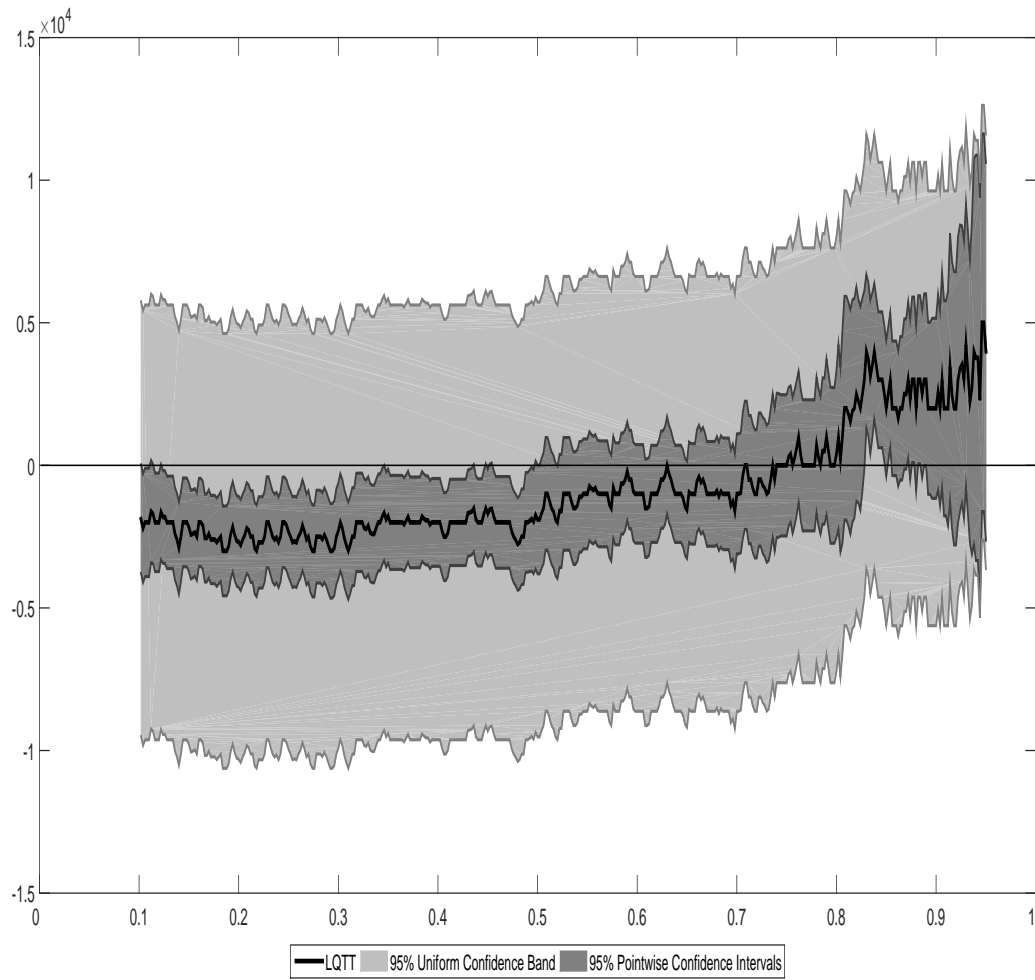


Figure 9: LQTT of Having More than One Child on Family Income in 1990