Estimating Partially Conditional Quantile Treatment Effects*†

Zongwu Cai[a,b], Ying Fang[b,c,†], Ming Lin[b,c], Shengfang Tang[c]

[a]Department of Economics, University of Kansas, Lawrence, KS 66045, USA. E-mail: caiz@ku.edu.

[b]Wang Yanan Institute for Studies in Economics (WISE) and Fujian Key Laboratory of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China.

E-mails: yifst1@xmu.edu.cn (Fang) and linming50@xmu.edu.cn (Lin).

[c]Department of Statistics, School of Economics, Xiamen University, Xiamen, Fujian 361005, China.

E-mail: tangshengfang103@163.com (Tang).

January 7, 2021

Abstract: This paper proposes a new model, termed as the partially conditional quantile treatment effect model, to characterize the heterogeneity of treatment effect conditional on some predetermined variable(s). We show that this partially conditional quantile treatment effect is identified under the assumption of selection on observables, which leads to a semiparametric estimation procedure in two steps: first, parametric estimation of the propensity score function and then, nonparametric estimation of conditional quantile treatment effects. Under some regularity conditions, the consistency and asymptotic normality of the proposed semiparametric estimator are derived. Furthermore, the finite sample performance of the proposed method is illustrated through Monte Carlo experiments. Finally, we apply our methods to estimate the quantile treatment effects of a first-time mother's smoking during the pregnancy on the baby's weight as a function of the mother's age, and our empirical results show substantial heterogeneity across different mother's ages with a significant negative effect of smoking on infant birth weight across all mother's ages and quantiles considered.

Keywords: Conditional quantile treatment effect; Heterogeneity; Propensity score; Semiparametric estimation; Treatment effect on treated

JEL classification: C21; C13; C14; C54.

*The authors acknowledge the financial supports partially from the National Science Fund of China (NSFC) for Distinguished Scholars (71625001) and the NSFC key projects with grant numbers 71631004 and 72033005.

†*Corresponding author:* Y. Fang (E-mail: yifst1@xmu.edu.cn)

1 Introduction

Understanding the causal effect of a treatment or policy or intervention, such as participating into a training program, is a basic goal of many empirical studies in economics and many other applied fields. This interest has led to a surge in theoretical and applied work focusing on estimating average treatment effects (ATE) or average treatment effects on the treated (ATT) group under various environments. Influential surveys include, but not limited to, the papers by Angrist and Krueger (1999), Heckman, Lalonde and Smith (1999), Blundell and Dias (2002), and among others. Moreover, Imbens (2004) and Imbens and Wooldridge (2009) provided comprehensive reviews on the recent developments in the treatment effect literature.

The average treatment effect, although vital, sometimes reveals only a partial picture for the outcome distribution of interest. For example, the mean effect can not measure how the dispersion of the outcome distribution has altered after a treatment, and furthermore, it is usually uninformative on whether the effects are stronger in some quantiles than in others. However, such distributional information can be important in many applications, particularly from policy-making of views. Here, there are some examples, evaluating the effect of the unionization on wage inequality as in Freeman (1980) and Card (1996), the effects of government training programs on lower quantiles of earning distributions studied by LaLonde (1995) and Abadie, Angrist and Imbens (2002), the effect of the government-subsidized saving program on lower tails of savings distributions, and among many others applications. From a policy perspective, a policy treatment that helps to raise the lower tail of an income distribution is often more appreciated than one that shifts the median, even though the average treatment effects of both are identical. To characterize the distributional effects of policy variables, quantile treatment effects (QTE), as addressed in the papers by Lehmann (1975) and Doksum (1974), can be an effective way which has emerged as an important concept for measuring distributional impacts in the literature. Recent studies on QTE include, but not limited to, the papers by Abadie et al. (2002), Chernozhukov and Hansen (2005), Donald and Hsu (2014), Firpo (2007), Frölich and Melly (2013), and the references therein.

Another challenge in the policy evaluation literature is how to characterize the heterogeneity of treatment effects across different individuals as in Heckman and Robb (1985) and Heckman, Smith and Clements (1997). Researchers are of interest to estimate the effect

of a treatment or a policy on outcomes in various sub-populations defined by some characterizations of components of pre-treatment variables X. For example, when estimating the effect of maternal smoking during pregnancy on the birth weight, it is interesting to catch heterogenous effects across mothers with different ages. To this end, Abrevaya, Hsu and Lieli (2015) and Lee, Okui and Whang (2017) developed the concept of partially conditional average treatment effect (PCATE) to measure the heterogeneity in mean effects across sub-populations. To be more detailed, Abrevaya et al. (2015) proposed using a nonparametric method to estimate the PCATE, whereas Lee et al. (2017) suggested a doubly robust estimation approach.

In this paper, our attempt is to capture heterogeneities for both across-distribution and across-individuals simultaneously. To this end, we propose a partially conditional quantile treatment effect (PCQTE) to characterize the heterogeneity along the outcome distribution conditional on some continuous covariate Z, which is only a strict subset of covariates X, under the condition that the unconfoundedness assumption holds (see Assumption 2.1(i) later). It is worth noting that the unconfoundedness assumption does not hold in general when only conditioning on the sub-vector Z, so new techniques are needed to identify the PCQTE parameter. We show that the PCQTE is nonparametrically identified and a semiparametric estimation is provided. Furthermore, under some regularity conditions, the proposed semiparametric estimator is shown to be consistent and asymptotically normal, which allows us to make point-wise statistical inference about PCQTE as a function of Z.

Our motivation of this paper comes actually from exploring an empirical study for estimating treatment effects of first-time mothers' smoking status during pregnancy on birth weight conditional on their ages. Indeed, Abrevaya et al. (2015) and Lee et al. (2017) considered the case by investigating the ATE of maternal smoking during pregnancy on infant birth weights conditional on mothers' ages, whereas Abrevaya et al. (2015) proposed nonparametric and semiparametric estimators of the conditional average treatment effect conditional on some continuous covariates. A semiparametric estimator was proposed if the propensity score function is estimated parametrically at the first stage, and also, a fully nonparametric estimator is provided when the propensity score function is estimated nonparametrically. To avoid the curse of dimensionality for nonparametric estimation, Lee et al. (2017) instead proposed a doubly robust estimator based on parametric regression in the sense that the estimator is consistent when either the regression model or the propensity score model is

correctly specified. However, the aforementioned papers do not address the heterogeneity issue. In other words, it is interesting to note that the distributions of infant birth weights for both whites and blacks are actually asymmetric and fat-tailed in the left side; see Figure 1 and Table 2 in Section 3 for details. Therefore, in this paper, we re-analyze this real example by using the proposed PCQTE model and its modeling approach, which might be more suitable for analyzing this dataset. Also, we investigate the conditional quantile treatment effect on treated group conditional on the mother's age, termed as PCQTT. As a result, our findings look very interesting, appealing, and novel in the literature, and further, different interpretations are provided to this real application. In a sum, our empirical results show substantial heterogeneity across different mothers' ages and there is a significant negative effect of smoking on infant birth weight across all mothers' ages and quantiles considered. More specifically, the smoking quantile effects become stronger, more negative on birth weights, at higher ages, and particularly, for whites, the estimated values at lower quantiles are bigger than those at the median or higher quantiles, conditional on mothers' ages. The detailed analysis of this real example is presented in Section 4.

The rest of this paper is organized as follows. Section 2 introduces the partially conditional quantile treatment effect model and discusses its identification conditions as well as estimation procedures, together with the presentation of the asymptotic properties of the proposed estimator. Also, Section 2 extends the proposed method to estimate the analogous parameter for the partially conditional quantile treatment effects on the treated group (PCQTT). Monte Carlo simulations are conducted in Section 3 to illustrate the finite sample performance of the proposed estimator, and Section 4 is devoted to them aforementioned empirical example to investigate how the distributional effect of maternal smoking on birth weights varies across different groups of mothers. Section 5 concludes. The proofs of the main results are delegated to Appendix.

2 Partially Conditional Quantile Treatment Effect Model

2.1 Model Setup

Let D_i be the binary treatment variable of individual i, where $D_i = 1$ if individual i receives the treatment of interest and otherwise, $D_i = 0$. Using the potential outcome framework initialized by Rubin (1974), let $Y_i(0)$ and $Y_i(1)$ be the potential outcomes of individual i if it is in the control group or in the treated group, respectively. Note that

for each individual i, we can only observe $Y_i(D_i)$ but $Y_i(1-D_i)$ is missing. The observed outcome variable Y_i can be written as

$$Y_i = (1 - D_i) \cdot Y_i(0) + D_i \cdot Y_i(1).$$

In addition, we observe a L-dimensional vector of pre-treatment variables, denoted by X_i. Throughout this paper, it is assumed that $\big(Y_i(0), Y_i(1), X_i, D_i\big)$, $i = 1, \cdots, n$, are independent and identically distributed. Since only one of $Y_i(0)$ and $Y_i(1)$ is observable for each individual i, the following assumptions are needed to identify the treatment effect.

Assumption 2.1. *(i) (Unconfounded Treatment Assignment) Given pre-treatment variables X_i, the potential outcomes are jointly independent from the treatment variable D_i, namely,*

$$\big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid X_i,$$

where $\perp\!\!\!\perp$ indicates statistical independence.
(ii) (Common Support) For almost all x in the support of X_i,

$$0 < \underline{p} \leq p(x) = P(D_i = 1 | X_i = x) \leq \overline{p} < 1,$$

for some $0 < \underline{p} < \overline{p} < 1$, where $p(x)$ is called propensity score function.

Assumption 2.1(i) is also known as the (strongly) "ignorable treatment assignment", or "conditional independence assumption" or "selection on observables" in the econometrics and/or statistics literature; see Rosenbaum and Rubin (1983) and Lechner (1999, 2002). It rules out the existence of unobserved factors that affect the treatment choice and are also correlated with the potential outcomes. Assumption 2.1(ii) states that in the population for almost all values of X_i, both treatment assignment levels have a positive probability of occurrence. However, lack of common support is one of main concerns in practice. A common approach to address this problem is to drop observations with the propensity score close to zero or one, and focus on the treatment effect in the subpopulation with propensity score bounded away from zero and one. These two assumptions have been widely used in literature on treatment effect evaluation, such as Heckman, Ichimura, Smith and Todd (1998), Dehejia and Wahba (1999), Hirano, Imbens and Ridder (2003), Firpo (2007), and among others.

In this paper, our purpose is on the quantile treatment effect conditional on a subset of

the pre-treatment variables. Specifically, let Z_i be a k-dimensional sub-vector of X_i, where $1 \leq k \ll L$, and then, the τ-th partially conditional quantile treatment effect is defined as

$$\Delta_\tau(z) = q_{1,\tau}(z) - q_{0,\tau}(z), \tag{2.1}$$

where for $j = 0, 1$ and $\tau \in (0, 1)$, $q_{j,\tau}(z)$ is the τ-th conditional quantile of $Y_i(j)$ conditional on $Z_i = z$. Note that the unconfounded treatment assignment assumption may not hold if one only controls the sub-vector Z_i instead of X_i. Also, note that if there is no Z_i in (2.1), it becomes to the unconditional quantile treatment effect model in Firpo (2007).

2.2 Estimation Procedures

Since the potential outcomes $Y_i(0)$ and $Y_i(1)$ are not observable for each individual, $Y_1(j), \cdots, Y_n(j)$ can not be used directly to estimate $q_{j,\tau}(z)$ in (2.1) for $j = 0$ and 1. Now, by defining $W_0(X_i, D_i) = (1 - D_i)/[1 - p(X_i)]$ and $W_1(X_i, D_i) = D_i/p(X_i)$, it is easy to show by Assumption 2.1 that

$$E\big[W_j(X_i, D_i)\, g(Y_i)\big|Z_i\big] = E\big[W_j(X_i, D_i)\, g(Y_i(j))\big|Z_i\big] = E\big[g(Y_i(j))\big|Z_i\big]$$

for $j = 0$ and 1 and any function $g(\cdot)$ with finite expectation, which implies that $q_{j,\tau}(z)$, $j = 0$ and 1, can be easily expressed as

$$q_{j,\tau}(z) = \arg\min_q E\Big(\rho_\tau(Y_i(j); q)\Big|Z_i = z\Big) = \arg\min_q E\Big(W_j(X_i, D_i)\rho_\tau(Y_i; q)\Big|Z_i = z\Big), \tag{2.2}$$

where $\rho_\tau(Y; q) = (Y - q)\{\tau - I(Y \leq q)\}$ is the check function as in Koenker and Bassett (1978) and Koenker (2005). Here, $I\{\cdot\}$ is the indicator function. When the propensity score function $p(x)$ is known, observations (Y_i, X_i, D_i), $i = 1, \cdots, n$, can be used directly to estimate $q_{j,\tau}(z)$ for $j = 0$ and 1 by running a weighted quantile regression model as in Koenker and Bassett (1978) and Koenker (2005).

Because $p(x)$ is unknown, in view of (2.2), a two-step estimation procedure is needed for estimating $\Delta_\tau(z)$ at any given grid point z. Firstly, one needs to obtain the estimated propensity score function $\widehat{p}_n(x)$ using (X_i, D_i), $i = 1, \cdots, n$, and then, at the second stage, the kernel-based locally weighted method is used to estimate $q_{j,\tau}(z)$ for $j = 0$ and 1 and thus, $\Delta_\tau(z)$. Specifically,

$$\widehat{\Delta}_\tau(z) = \widehat{q}_{1,\tau}(z) - \widehat{q}_{0,\tau}(z), \tag{2.3}$$

where for $j = 0$ and 1,

$$\widehat{q}_{j,\tau}(z) = \arg\min_{q} \frac{1}{n}\sum_{i=1}^{n} K_h\big(Z_i - z\big)\widehat{W}_{n,j}(X_i, D_i)\rho_\tau(Y_i; q) \tag{2.4}$$

with $\widehat{W}_{n,0}(X_i, D_i) = (1-D_i)/[1-\widehat{p}_n(X_i)]$, $\widehat{W}_{n,1}(X_i, D_i) = D_i/\widehat{p}_n(X_i)$, and $K_h(u) = K(u/h)/h$. Here, $K(\cdot)$ is a kernel function, h is the bandwidth parameter, and $\widehat{p}_n(x)$ is a consistent estimate of $p(x)$. Of course, the estimation procedure in (2.4) can be extended to the local linear estimation scheme as follows

$$(\widehat{q}^{ll}_{j,\tau}, \widehat{q}^{\prime ll}_{j,\tau}) = \arg\min_{q_0,\, q_1} \frac{1}{n}\sum_{i=1}^{n} K_h\big(Z_i - z\big)\widehat{W}_{n,j}(X_i, D_i)\rho_\tau\big(Y_i; q_0 + q_1(Z_i - z)\big),\ j = 0, 1,$$

which gives the local linear estimate of $\Delta_\tau(z)$, $\widehat{\Delta}_{\tau,ll}(z) = \widehat{q}^{ll}_{1,\tau} - \widehat{q}^{ll}_{0,\tau}$. The theoretical derivations for $\widehat{\Delta}_{\tau,ll}(z)$ should be the same as those for $\widehat{\Delta}_\tau(z)$ so that for simplicity, the asymptotic theory for $\widehat{\Delta}_\tau(z)$ is only provided below. Indeed, the asymptotic properties for $\widehat{\Delta}_{\tau,ll}(z)$ are available upon request.

Now, the question is how to obtain a consistent estimate of $p(x)$. It is well documented in the literature that there are two common approaches used for estimating the propensity score function $p(x)$. The first approach is to assume a parametric model as $p(x) = p(x;\theta)$, for example, a logit model or a probit model so that the parameter θ can be easily estimated through the maximum likelihood method. The second one is nonparametric. For a nonparametric method, one can use the so-called series logit estimator as in Hirano et al. (2003) and Firpo (2007) or other suitable consistent estimators of $p(x)$ are also possible. For example, Ichimura and Linton (2005) used local polynomial regression and Abrevaya et al. (2015) used higher order kernel regression to estimate $p(x)$. The first one, parametric form of $p(x)$, is used in this paper so that the estimation in (2.3) is called a semiparametric estimator. Indeed, Tang (2020) considered using a nonparametric method and derived the first-order asymptotic results displayed in the following section, which do not depend critically on the choice of $\widehat{p}_n(x)$ as long as the bandwidth used for estimating $p(x)$ is under-smoothed. Therefore, the similar conclusions as in Theorem 2.1 can be obtained under some regularity conditions for a nonparametric estimation of $p(x)$; see Tang (2020) for details.

2.3 Asymptotic Theory

This subsection is devoted to investigating the asymptotic properties for the semiparametric estimator for $\widehat{\Delta}_\tau(z)$ in (2.3), in the sense that the propensity score function $p(x)$ is estimated parametrically, and $\Delta_\tau(z)$ is estimated nonparametrically using equations (2.3) and (2.4). Although the asymptotic theory for $\widehat{\Delta}_\tau(z)$ can be obtained for any k-dimensional Z_i with $k < L$, the result is presented only for $k = 1$ to save notation throughout the rest of this paper. As pointed out by Abrevaya et al. (2015), the case for $k = 1$ is the most relevant case in practice, since $\widehat{\Delta}_\tau(z)$ can easily be displayed in a two-dimensional graph when Z_i is a scalar. Before studying the asymptotic properties of the proposed estimators, the following technical assumptions are needed, list below.

Assumption 2.2. (Distributions of X_i and Z_i) There exists a constant $c > 0$ such that the density function of X_i, $f_X(x)$ satisfies $\inf_{x\in\mathcal{X}} f_X(x) \geq c$, where $\mathcal{X}$ is the support of X_i. Furthermore, the density function of Z_i, $f_Z(z)$ is twice continuously differentiable on the support of Z_i.

Assumption 2.3. (i) The conditional density function $f_{Y(j)|X}(y|x)$ is continuous and bounded on the support of $Y_i(j)$ and X_i for $j = 0, 1$. (ii) The conditional density function $f_{Y(j)|Z}(y|z)$ is continuous and uniformly bounded away from zero in a neighborhood of $q_{j,\tau}(z)$ for $j = 0$ and 1. It is twice differentiable with respect to z, and its first derivative with respect to y is continuous and bounded on the support of $Y_i(j)$ and Z_i.

Assumption 2.4. (Kernel and bandwidth) (i) The kernel function $K(u)$ is a symmetric density function with compact support. It is also continuously differentiable on its support. (ii) $h \to 0$, $nh^{1+\varepsilon} \to \infty$ for some $\varepsilon > 0$ and nh^5 is bounded as $n \to \infty$.

Assumption 2.5. (Parametric propensity score function) Suppose the propensity score function has a parametric form $p(x) = p(x; \theta_0)$ with a fixed dimensional parameter θ_0. Also assume that the estimated propensity score function $\widehat{p}_n(x) = p(x; \widehat{\theta}_n)$ satisfies $\sup_{x\in\mathcal{X}} \left|p(x; \widehat{\theta}_n) - p(x; \theta_0)\right| = O_p(n^{-1/2})$.

The restriction imposed on the distribution of X_i in Assumption 2.2 is commonly used in the literature on treatment effect evaluation, see Hirano et al. (2003), Abadie and Imbens (2006, 2016), Firpo (2007), Abrevaya et al. (2015), and among others. Assumption 2.3 guarantees the solution of (2.2) is unique and the smoothness conditions imposed are easily

satisfied in practice. Assumption 2.4 on kernel function and bandwidth is frequently assumed in the literature on nonparametric estimation. Many commonly used kernel functions, such as the Epanechnikov kernel, satisfy the requirements. Assumption 2.5 typically holds for standard parametric estimation methods under reasonably mild regularity conditions.

Next, we establish the asymptotic properties of $\widehat{\Delta}_\tau(z)$, which are stated in the following theorem with the detailed proof given in Appendix. For easy presentation, define some notations as follows. First, define $F_j(y|z) = F_{Y(j)|Z}(y|z)$ to be the conditional CDF of $Y(j)$ given $Z = z$ for $j = 0$ and 1, and its m-th order derivative $F_j^{(m)}(y|z) = \partial^m F_j(y\,|z)/\partial z^m$ for $m \geq 0$. Also, let $\psi_j(Y_i, X_i, D_i; z) = W_j(X_i, D_i)\,(I\{Y_i \leq q_{j,\tau}(z)\} - \tau)$ and $\delta_\tau(z) = \delta_{1,\tau}(z) - \delta_{0,\tau}(z)$, where for $j = 0$ and 1,

$$\delta_{j,\tau}(z) = \frac{2f_Z'(z)F_j^{(1)}(q_{j,\tau}(z)|z)}{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)} + \frac{F_j^{(2)}(q_{j,\tau}(z)|z))}{f_{Y(j)|Z}(q_{j,\tau}(z)|z)}, \tag{2.5}$$

which is in the asymptotic bias term in $\widehat{\Delta}_\tau(z)$, given in the following theorem.

Theorem 2.1. *Suppose that Assumptions 2.1-2.5 hold. Then, for each z in the support of Z_i, one has*

$$\begin{aligned}
&\sqrt{nh}\left[\widehat{\Delta}_\tau(z) - \Delta_\tau(z) + \frac{h^2}{2}\mu_2(K)\delta_\tau(z) + o_p(h^2)\right] \\
= \;& -\frac{1}{\sqrt{nh}}\frac{1}{f_Z(z)}\sum_{i=1}^n \left\{ \frac{hK_h(Z_i - z)\psi_1(Y_i, X_i, D_i, z) - E\Big(hK_h(Z_i - z)\psi_1(Y_i, X_i, D_i, z)\Big)}{f_{Y(1)|Z}(q_{1,\tau}(z)|z)} \right. \\
& \left. - \frac{hK_h(Z_i - z)\psi_0(Y_i, X_i, D_i, z) - E\Big(hK_h(Z_i - z)\psi_0(Y_i, X_i, D_i, z)\Big)}{f_{Y(0)|Z}(q_{0,\tau}(z)|z)} \right\} + o_p(1) \qquad (2.6) \\
\xrightarrow{D} \;& \mathcal{N}\Big(0, \nu_0(K)\sigma_\tau^2(z)/f_Z(z)\Big), \qquad (2.7)
\end{aligned}$$

where $\delta_\tau(z)$ is defined in (2.5), $\mu_2(K) = \int u^2K(u)du$, $\nu_0(K) = \int K^2(u)du$, and

$$\sigma_\tau^2(z) = E\left\{ \left(\frac{\psi_1(Y_i, X_i, D_i, z)}{f_{Y(1)|Z}(q_{1,\tau}(z)|z)} - \frac{\psi_0(Y_i, X_i, D_i, z)}{f_{Y(0)|Z}(q_{0,\tau}(z)|z)} \right)^2 \Big| Z_i = z \right\},$$

which is in the asymptotic variance term of $\widehat{\Delta}_\tau(z)$.

It can be seen from Theorem 2.1 that the first term in (2.6) is the first-order approximation for $\widehat{\Delta}_\tau(z)$, which is the so-called local Bahadur representation; see Cai and Xu (2008), which makes the asymptotic analysis in (2.7) much easier. Another consequence of Theorem

2.1 is to provide a formulation for constructing a confidence interval for making a statistical inference. To construct a pointwise confidence interval for $\Delta_\tau(z)$ for each given z, by ignoring the asymptotic bias term, one needs to obtain a consistent estimate for both $f_Z(z)$ and $\sigma_\tau^2(z)$. Clearly, the density function of Z_i can be estimated by the kernel density estimator as $\widehat{f}_Z(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i - z)$. However, it is much more involved to estimating $\sigma_\tau^2(z)$ because it includes the unknown conditional density function $f_{Y(j)|Z}(q_{j,\tau}(z)|z)$ for $j = 0$ and 1. As pointed out by Koenker and Xiao (2004), Koenker (2005), and Cai and Xu (2008), it might not be easy to estimate consistently the conditional density function $f_{Y(j)|Z}(q_{j,\tau}(z)|z)$. Following Koenker (2005), we propose in this paper using the following estimator to estimate the conditional density function of $Y_i(j)$ conditional on $Z_i = z$ for $j = 0$ and 1,

$$\widehat{f}_{Y(j)|Z}(q_{j,\tau}(z)|z) = \frac{2h^*}{\widehat{q}_{j,\tau+h^*}(z) - \widehat{q}_{j,\tau-h^*}(z)},$$

where h^* is a bandwidth parameter. Indeed, Koenker (2005) showed that $\widehat{f}_{Y(j)|Z}(q_{j,\tau}(z)|z)$ converges to $f_{Y(j)|Z}(q_{j,\tau}(z)|z)$ in probability if $h^* \to 0$ and $h^*\sqrt{nh} \to \infty$. Then, a consistent estimate of $\sigma_\tau^2(z)$ can be given by

$$\widehat{\sigma}_\tau^2(z) = \sum_{i=1}^{n} K_h(Z_i - z)\left(\frac{\widehat{\psi}_1(Y_i, X_i, D_i; z)}{\widehat{f}_{Y(1)|Z}(q_{1,\tau}(z)|z)} - \frac{\widehat{\psi}_0(Y_i, X_i, D_i; z)}{\widehat{f}_{Y(0)|Z}(q_{0,\tau}(z)|z)}\right)^2 \Big/ \sum_{i=1}^{n} K_h(Z_i - z),$$

where $\widehat{\psi}_j(Y_i, X_i, D_i; z) = \widehat{W}_{n,j}(X_i, D_i)\big(I\{Y_i \le \widehat{q}_{j,\tau}(z)\} - \tau\big)$ for $j = 0$ and 1. Therefore, one can compute a pointwise confidence interval for $\Delta_\tau(z)$ by ignoring the asymptotic bias term.

2.4 Estimation of PCQTT

In general, policy-makers may be interested not only in the treatment effect for the whole population, but also the treatment effect for the treated group as in Heckman and Robb (1985) and Heckman et al. (1999). Of course, individual treatment effect might be heterogeneous within the treated subpopulations as well. Therefore, this section is devoted to the estimation of the PCQTT; that is,

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

where $q_{j,\tau|D=1}(z) = \inf\big\{y : P(Y_i(j) \le y \mid Z_i = z, D_i = 1) \ge \tau\big\}$ is the τ-th quantile of $Y_i(j)$ conditional on $Z_i = z$ and $D_i = 1$ for $j = 0$ and 1.

To derive the nonparametric estimation of $\Delta_{\tau|D=1}(z)$, define $V_1(X_i, D_i) = D_i$ and $V_0(X_i, D_i)$

$= (1-D_i)p(X_i)/\big[1-p(X_i)\big]$. Then, by the law of iterated expectations and the unconfounded treatment assignment assumption, it is easy to verify that

$$E\Big(V_j(X_i,D_i)g(Y_i)\Big|Z_i=z\Big)=E\Big(V_j(X_i,D_i)g(Y_i(j))\Big|Z_i=z\Big)=E\Big(D_i\cdot g(Y_i(j))\Big|Z_i=z\Big)$$

for $j=0,1$ and for any function $g(\cdot)$ with finite expectation. Consequently, similar to (2.2), the conditional quantile function $q_{j,\tau|D=1}(z)$ can be identified by

$$q_{j,\tau|D=1}(z)=\arg\min_q E\Big(V_j(X_i,D_i)\rho_\tau(Y_i;q)\big|Z_i=z\Big),$$

from which, the following semiparametric estimator for the PCQTT is proposed

$$\widehat{\Delta}_{\tau|D=1}(z)=\widehat{q}_{1,\tau|D=1}(z)-\widehat{q}_{0,\tau|D=1}(z),$$

where

$$\widehat{q}_{j,\tau|D=1}(z)=\arg\min_q\frac{1}{n}\sum_{i=1}^n K_h\big(Z_i-z\big)\widehat{V}_{n,j}(X_i,D_i)\rho_\tau(Y_i;q),$$

with $\widehat{V}_{n,0}(X_i,D_i)=(1-D_i)\widehat{p}_n(X_i)/\big(1-\widehat{p}_n(X_i)\big)$, $\widehat{V}_{n,1}(X_i,D_i)=D_i$, and $\widehat{p}_n(x)$ being a parametric estimator of $p(x)$. To establish the large sample properties for the proposed PCQTT estimator $\widehat{\Delta}_{\tau|D=1}(z)$, Assumption 2.3 is needed a modification, given below.

Assumption 2.3*. (i) The conditional density function $f_{Y(j)|X,D=1}(y|x)$ is continuous and bounded on the support of $Y_i(j)$ and X_i for $j=0$ and 1. (ii) The conditional density function $f_{Y(j)|Z,D=1}(y|z)$ is continuous and uniformly bounded away from zero in a neighborhood of $q_{j,\tau|D=1}(z)$ for $j=0$ and 1. It is twice differentiable with respect to z, and its first derivative with respect to y is continuous and bounded on the support of $Y_i(j)$ and Z_i. (iii) Finally, $p_Z(z)=P(D_i=1|Z_i=z)$ is twice continuously differentiable.

For simplicity of exposition, let $\varphi_1(Y_i,X_i,D_i;z)=V_1(X_i,D_i)\big(I\{Y_i\le q_{1,\tau|D=1}-\tau\}\big)$ and $\varphi_0(Y_i,X_i,D_i;z)=V_0(X_i,D_i)\big(I\{Y_i\le q_{1,\tau|D=1}-\tau\}\big)$. For $j=0$ and 1, define $F_{j|D=1}(y|z)=F_{Y(j)|Z,D=1}(y|z)$ to be the conditional CDF of $Y(j)$ conditional on $Z=z$ and $D=1$, and its m-th order derivative $F^{(m)}_{j|D=1}(y|z)=\partial^m F_{j|D=1}(y|z)/\partial z^m$ for $m\ge 0$. Also, denote $\zeta_\tau(z)=\zeta_{1,\tau|D=1}-\zeta_{0,\tau|D=1}$, where for $j=0$ and 1,

$$\begin{aligned}\zeta_{j,\tau|D=1} &= \frac{1}{p_Z(z)f_Z(z)f_{Y(j)|Z,D=1}(q_{j,\tau|D=1}(z)|z)}\Big\{2p_Z(z)f'_Z(z)F^{(1)}_{j|D=1}(q_{j,\tau|D=1}(z)|z)\\ &\quad +p_Z(z)f_Z(z)F^{(2)}_{j|D=1}(q_{j,\tau|D=1}(z)|z)+2p'_Z(z)f_Z(z)F^{(1)}_{j|D=1}(q_{j,\tau|D=1}(z)|z)\Big\}.\end{aligned}$$

The following theorem similar to Theorem 2.1 summarizes the asymptotic properties of the estimator $\widehat{\Delta}_{\tau|D=1}$ with the proof similar to that for Theorem 2.1 and omitted.

Theorem 2.2. *Suppose Assumptions 2.1, 2.2, 2.3*, 2.4 and 2.5 hold. Then, for each z in the support of Z, we have*

$$\sqrt{nh}\left[\widehat{\Delta}_{\tau|D=1}(z) - \Delta_{\tau|D=1}(z) + \frac{h^2}{2}\mu_2(K)\zeta_\tau(z) + o_p(h^2)\right] \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{\tau,T}\right),$$

where $\zeta_\tau(z)$ is defined above, $\sigma^2_{\tau,T} = \nu_0(K)\sigma^2_{\tau|D=1}(z)/\left[p^2_Z(z)f_Z(z)\right]$ with

$$\sigma^2_{\tau|D=1}(z) = E\left\{\left(\frac{\varphi_1(Y_i, X_i, D_i; z)}{f_{Y(1)|Z,D=1}(q_{1,\tau|D=1}(z)|z)} - \frac{\varphi_0(Y_i, X_i, D_i; z)}{f_{Y(0)|Z,D=1}(q_{0,\tau|D=1}(z)|z)}\right)^2 \Big| Z_i = z\right\},$$

and other notations are the same as those in Theorem 2.1.

3 Monte Carlo Studies

In this section, Monte Carlo experiments are conducted to examine the finite sample performance of the proposed estimation procedure. The goal is to assess the finite sample accuracy in various scenarios.

Example 1. We consider a Skorohod representation[1] for the potential outcomes $Y(0)$ and $Y(1)$. Specifically, the data generating process is given by

$$Y(0) = \lambda_0 X_1 + \gamma_0\sqrt{U_0}X_2 \quad \text{and} \quad Y(1) = \lambda_1 X_1 + \gamma_1\sqrt{U_1}X_2,$$

where $\lambda_0 = 3.0$, $\gamma_0 = 0.4$, $\lambda_1 = 4.0$, $\gamma_1 = 1.6$, U_0 and U_1 independently follow the uniform $U[0,1]$ distribution, X_1 and X_2 are independent with $X_1 \sim U[0,1]$ and $X_2 \sim \text{Beta}(3,1)$, and the propensity score function is

$$P(D = 1|X_1, X_2) = \frac{\exp\{-0.5 + X_1 + X_2\}}{1 + \exp\{-0.5 + X_1 + X_2\}}.$$

Finally, the conditional variable Z is taken to be X_1. Under this setting, the conditional quantile function for $Y(j)$ for $j = 0$ and 1, conditional on $Z = z$, is given by

$$q_{j,\tau}(z) = \lambda_j z + \gamma_j a_\tau, \tag{3.1}$$

where a_τ is the unique solution of equation $-2a^3 + 3a^2 - \tau = 0$ within the interval $(0,1)$.

[1]For the definition of the Skorohod representation, the reader is referred to Durrett (1996).

Therefore, the PCQTE is

$$\Delta_\tau(z) = (\lambda_1 - \lambda_0)z + (\gamma_1 - \gamma_0)a_\tau. \tag{3.2}$$

To assess the finite sample performance of the estimator $\widehat{\Delta}_\tau(z)$, the mean absolute deviation error (MADE) criterion is used, which is defined as

$$\text{MADE}\big(\widehat{\Delta}_\tau(\cdot)\big) = \frac{1}{m}\sum_{j=1}^{m}\big|\widehat{\Delta}_\tau(z_j) - \Delta_\tau(z_j)\big|,$$

where $\{z_j\}_{j=1}^m$ are the grid points taken from the support of z with equal increments. The semiparametric estimator (2.3) with Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u| \leq 1)$ is used to compute $\widehat{\Delta}_\tau(z)$. It is well known that the choice of bandwidth in kernel-based estimation is important. By Assumption 2.4, the bandwidth h is set to be $h = c \cdot n^{-1/5}$ for $c \in \{0.25, 0.5, 1.0\}$ to illustrate how the choice of h affects the performance of the estimator. In the treatment effect literature, the estimated propensity score is often trimmed to prevent it from getting too close to 0 or 1. Therefore, following the convention in the literature, the estimated propensity score $\widehat{p}_n(x)$ is truncated to be between $[0.005, 0.995]$ in the following simulation studies.

The simulation is replicated $1,000$ times to compute the median and standard deviation (in parentheses) of the $1,000$ MADE values for each setting. Table 1 reports the simulation results for the proposed semiparametric estimator. As seen in Table 1, the semiparametric estimator performs well in terms of MADE and the choice of the bandwidth h should be around $0.5\,n^{-1/5}$ based on the MADEs and their standard deviations. As expected, due to the sparsity of sample observations in tail regions, the estimator performs better around median regions than in tail regions. Finally, from the results presented in Table 1, one can see clearly that there is a sharply decrease in MADEs and their standard deviations as sample size goes from $n = 500$ to $n = 2,000$ in all cases, which is in line with the asymptotic theory.

Table 1: Median and standard deviation (in parentheses) of 1000 MADE values for $\widehat{\Delta}_\tau(\cdot)$.

τ	$h = 0.25n^{-1/5}$			$h = 0.5n^{-1/5}$			$h = 1.0n^{-1/5}$		
	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
	MADE	MADE	MADE	MADE	MADE	MADE	MADE	MADE	MADE
0.1	0.090 (0.022)	0.068 (0.014)	0.060 (0.010)	0.089 (0.020)	0.072 (0.015)	0.058 (0.011)	0.101 (0.027)	0.093 (0.020)	0.084 (0.014)
0.25	0.084 (0.020)	0.064 (0.013)	0.047 (0.009)	0.080 (0.017)	0.062 (0.013)	0.048 (0.010)	0.094 (0.025)	0.083 (0.017)	0.080 (0.012)
0.5	0.080 (0.019)	0.058 (0.012)	0.044 (0.008)	0.066 (0.015)	0.049 (0.011)	0.036 (0.008)	0.080 (0.023)	0.059 (0.015)	0.045 (0.011)
0.75	0.082 (0.021)	0.062 (0.013)	0.046 (0.009)	0.078 (0.018)	0.060 (0.012)	0.045 (0.009)	0.093 (0.025)	0.085 (0.018)	0.075 (0.013)
0.9	0.091 (0.022)	0.069 (0.013)	0.061 (0.010)	0.090 (0.020)	0.075 (0.015)	0.060 (0.011)	0.098 (0.028)	0.092 (0.021)	0.086 (0.015)

Example 2. In this example, we investigate the finite sample performance of the proposed semiparametric estimator $\widehat{\Delta}_{\tau|D=1}(z)$ with the same setting as that in Example 1. Again, the conditional variable Z is taken to be X_1 and the simulation is replicated $1,000$ times to compute the median and standard deviation (in parentheses) of the $1,000$ MADE values for each setting. Table 2 below displays the simulation results for the proposed semiparametric estimator $\widehat{\Delta}_{\tau|D=1}(z)$.

Table 2: Median and standard deviation (in parentheses) of 1000 MADE values for $\widehat{\Delta}_{\tau|D=1}(\cdot)$.

τ	$h = 0.25n^{-1/5}$			$h = 0.5n^{-1/5}$			$h = 1.0n^{-1/5}$		
	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
	MADE	MADE	MADE	MADE	MADE	MADE	MADE	MADE	MADE
0.1	0.087 (0.023)	0.063 (0.015)	0.055 (0.010)	0.088 (0.020)	0.069 (0.014)	0.054 (0.009)	0.099 (0.027)	0.091 (0.020)	0.083 (0.014)
0.25	0.082 (0.021)	0.061 (0.013)	0.046 (0.008)	0.076 (0.017)	0.057 (0.012)	0.042 (0.008)	0.092 (0.023)	0.076 (0.017)	0.067 (0.012)
0.5	0.080 (0.019)	0.060 (0.011)	0.045 (0.008)	0.068 (0.015)	0.049 (0.010)	0.037 (0.007)	0.077 (0.021)	0.058 (0.016)	0.045 (0.011)
0.75	0.083 (0.021)	0.062 (0.013)	0.047 (0.010)	0.087 (0.018)	0.065 (0.012)	0.056 (0.009)	0.101 (0.024)	0.092 (0.017)	0.083 (0.012)
0.9	0.090 (0.023)	0.066 (0.014)	0.051 (0.011)	0.093 (0.020)	0.076 (0.013)	0.064 (0.011)	0.103 (0.028)	0.092 (0.021)	0.086 (0.015)

From Table 2, one can observe similar pattern as in Example 1. Specifically, similar to the semiparametric estimator $\widehat{\Delta}_\tau(z)$, the semiparametric estimator $\widehat{\Delta}_{\tau|D=1}(z)$ also performs well in terms of MADE. Moreover, the choice of the bandwidth h in a reasonable range seems to have little influence on the MADEs and their standard deviations. Again, the estimators perform better around median regions than in tail regions and the MADEs and

their standard deviations sharply decrease as the sample size increases from $n = 500$ to $n = 2000$ in all cases considered.

4 An Empirical Application

Many studies document that low infant birth weight is associated with prolonged negative effects on health, educational and labor market outcomes throughout life, although there has been a debate over its magnitude; see, for example, Abrevaya (2006), Almond, Chay and Lee (2005) and Currie and Almond (2011) and among others. It is well known that there are many risk factors which can cause low birth weight, and it is generally recognized that maternal smoking is considered to be the most important preventable negative cause of low birth weight; see Kramer (1987) for more discussions. Over the last decades, there have been many studies that attempt to estimate the effect of maternal smoking on low birth weight using various procedures. Recently, program evaluation approach is employed to estimate this effect; see, for example, Almond et al. (2005), Abrevaya (2006), da Veiga and Wilder (2008), Abrevaya and Dahl (2008) and Abrevaya et al. (2015) and the references therein. In this paper, our interest is to see how the effect of maternal smoking changes across different age groups of mothers along with the infant birth weight distribution. To capture this heterogeneity, the proposed procedure is used to estimate the quantile effect of maternal smoking on infant birth weight conditional on different mothers' ages, which is different from the studies by Abrevaya et al. (2015) and Lee et al. (2017) by considering the average effect of maternal smoking on infant birth weight conditional on different mothers' ages in their application. Because a large number of covariates is needed to make the unconfoundedness assumption plausible in this example, our focus is on the parametric estimator for the propensity score function $p(x)$ as in Abrevaya et al. (2015) and Lee et al. (2017).

To this end, the same data as Abrevaya et al. (2015) is used, which is based on the records between 1988 and 2002 by the North Carolina State Center Health Services, accessible through the Odum Institute at the University of North Carolina. As in Abrevaya et al. (2015), our sample is limited to first-time mothers and as routine in the literature, the total sample contains whites which consist of a sample of $433,558$ observations and blacks which consist of a sample of $157,989$ observations as separate samples throughout. Note that some aforementioned papers considered both samples for whites and blacks; say, da Veiga and

Wilder (2008) and Abrevaya et al. (2015), but some only investigated the sample for whites; say, Lee et al. (2017). Following Abrevaya et al. (2015), in our analysis below, we explore both samples separately.

In this empirical example, the outcome of interest Y is the infant birth weight measured in grams and the treatment variable D is a binary variable which takes value 1 if the mother smokes and 0 otherwise. $Y(0)$ denotes birth weights for the untreated (no-smoking) group and $Y(1)$ for the treated (smoking) group. Since our interest is to see how the quantile effect of smoking varies across different values of the mother's ages, hence the conditional variable Z is the mother's age in this application. The kernel density estimations of the infant birth weights are displayed in Figures 1 for whites (the left panel) and blacks (the right panel), respectively. For both whites and blacks, skewness and kurtosis of infant birth weights and the results of the symmetry test for the distributions of $Y(0)$ and $Y(1)$ are all reported in Table 3. Based on these results, one can observe that the distributions of infant birth weights for both whites and blacks are fat-tailed in the left side. Therefore, this motivates us to consider the distributional effect of maternal smoking on infant birth weight instead of mean effect.

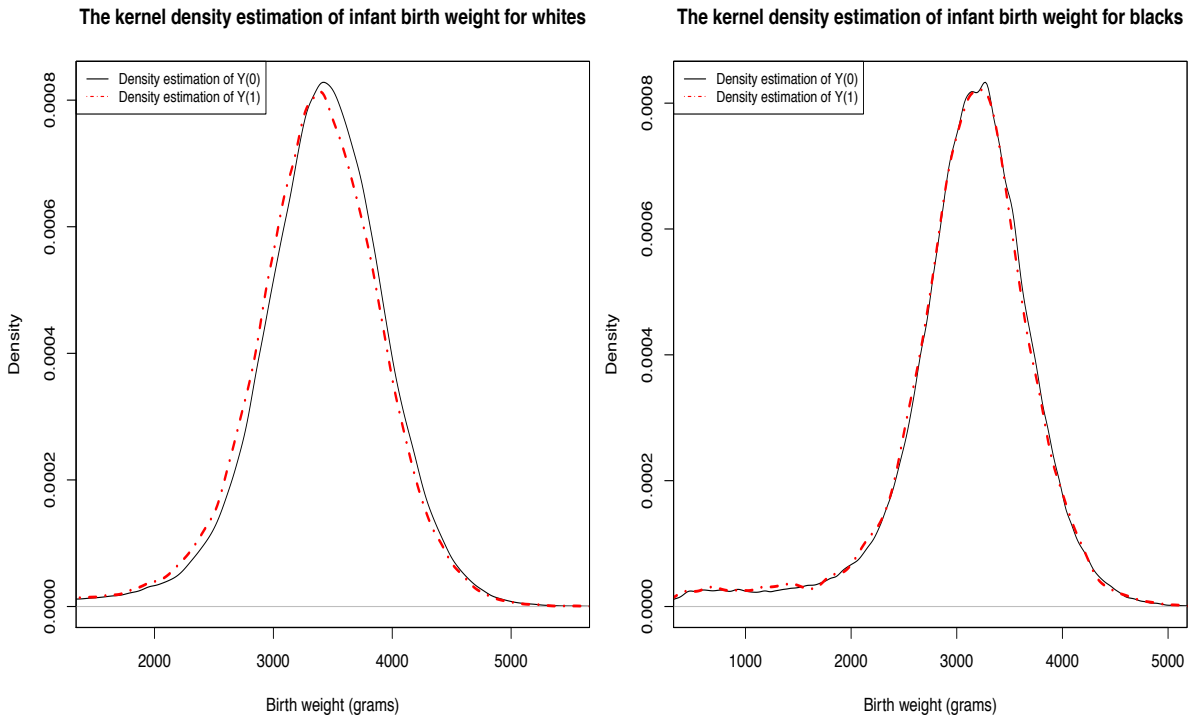

Figure 1: The kernel density estimation of birth weight for whites (the left panel) and blacks (the right panel).

Table 3: Descriptive statistics and results for testing symmetry and kurtosis.

	Whites		Blacks	
Variable	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Mean	3398.681	3346.848	3103.722	3082.726
Skewness	-0.846	-0.840	-1.181	-1.204
Kurtosis	5.931	5.734	6.245	6.164
Symmetry test (p-value)	0.000	0.000	0.000	0.000
Number of observations	359172	74386	146399	11590

To estimate the PCQTE function $\Delta_\tau(z)$, the same set of covariates X is used as in Abrevaya et al. (2015). Specifically, the set of covariates X includes the mother's age, education, month of first prenatal visit, number of prenatal visits, and indicators for the baby's gender, the mother's marital status, whether or not the father's age is missing, gestational diabetes, hypertension, amniocentesis, taking ultra sound exams, previous (terminated) pregnancies, and alcohol use; see Abrevaya et al. (2015) for the detailed discussion. A logit model is used to estimate the propensity score function $p(x)$ with the explanatory variables consisting of all the elements of X, the square of the mother's age, and the interaction terms between the mother's age and all other elements of X. As in Crump et al. (2008) and Abrevaya et al. (2015), the estimated propensity score $\widehat{p}_n(x)$ is truncated to be between $[0.01, 0.99]$ (about 0.136% of the observations dropped for whites and 0.378% for blacks)[2], which is different from that in Abrevaya et al. (2015). The PCQTE function is estimated for mothers aged between 20 and 30 for both whites and blacks.

First, Figure 2 presents the estimated curves of the conditional CDFs for infant birth weights conditional on mother's age ($z = 26$) for whites. Also, the estimated conditional CDFs of infant birth weights under different mother's ages can be obtained but the patterns are quite similar. It can be seen from Figure 2 that the estimated conditional CDF curve for $Y(1)$ is all on the left of $Y(0)$, which implies that the partially conditional quantile treatment effects should be negative across all quantile levels.

[2]Note that in Abrevaya et al. (2015), the estimated propensity score $\widehat{p}_n(x)$ is truncated to be between $[0.03, 0.97]$ for blacks (about 20% of the observations dropped) and $[0.08, 0.92]$ for whites (about 33% of observations dropped). We believe that such truncation rate is slightly high, especially for whites. Indeed, we used the same truncation as in Abrevaya et al. (2015) and the estimated PCQTEs for both whites and blacks are similar to those for our truncation, which are available upon request.

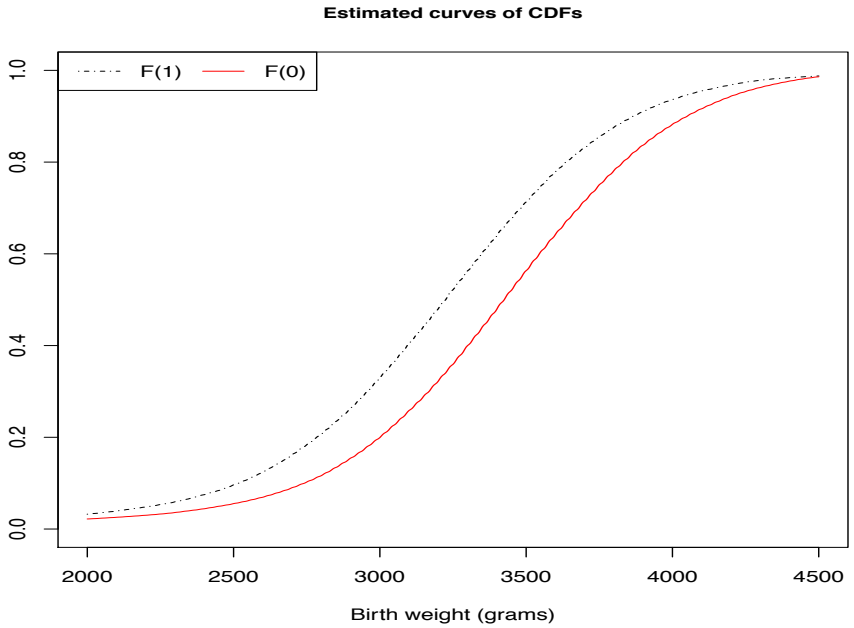

Figure 2: Estimated curves of conditional CDFs for infant birth weight conditional on mother's age ($z = 26$) for whites. $F(0)$ and $F(1)$ are for the non-smoking group and the smoking group, respectively.

Second, Figure 3 plots the estimated PCQTE curves across mothers' ages for three quantile levels $\tau = 0.10$ (the dashed-dotted line), 0.25 (the long dashed line) and $\tau = 0.50$ (the short dashed line) for whites. For comparison, Figure 3 also depicts the estimated PCATE curve by the the solid line, considered in Abrevaya et al. (2015) and Lee et al. (2017), across mothers' ages, and the estimated unconditional ATE as well (the dotted line). From Figure 3, first, one can see that $\widehat{\Delta}_\tau(z)$ for three τ values seem to change over age linearly and in particular, $\widehat{\Delta}_{0.5}(z)$ and the estimated PCATE curve as in Lee et al. (2017) are similar but they are not exactly same. Indeed, $\widehat{\Delta}_{0.5}(z)$ is slightly larger than its PCATE. More importantly, one can observe that there is a significant negative effect of smoking on infant birth weight across all ages and quantile levels considered. These results are in line with the findings displayed by Figure 2. Moreover, the estimated results displayed in Figure 3 show substantially heterogeneity across different ages. Overall, the estimated quantile effects become stronger (more negative) at higher ages. On the other hand, the estimated values at lower quantiles are bigger than those at the median, conditional on the same mother's age.

Furthermore, Figure 4 displays the estimated PCQTEs for blacks for three quantile levels $\tau = 0.10$ (the dashed-dotted line), 0.50 (the short dashed line) and 0.80 (the long dashed line), respectively. The unconditional QTE for $\tau = 0.5$ (the solid line) is also reported with the value at -141.75, together with its 95% confidence interval indicated by the the dotted

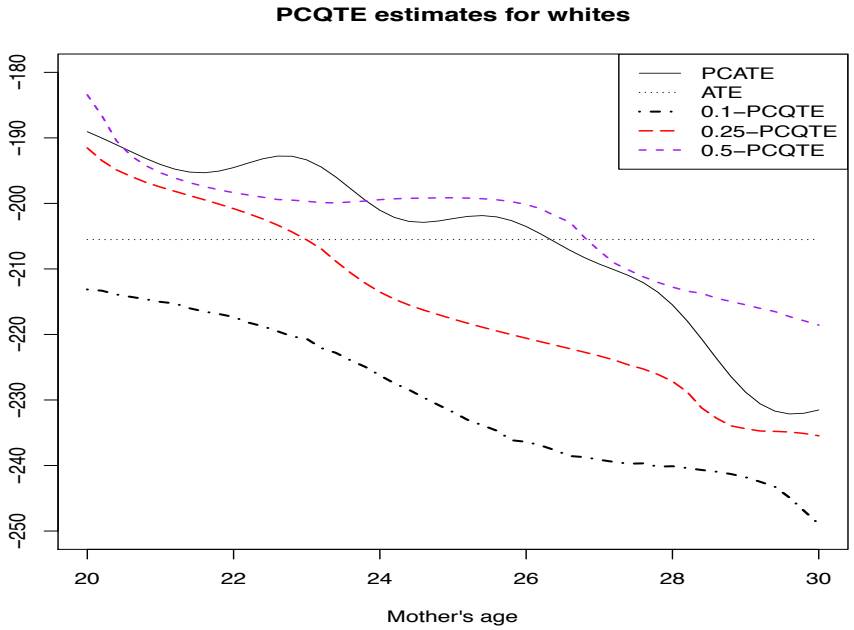

Figure 3: Estimation results for PCQTE for whites for three quantile levels 0.1, 0.25 and 0.50, together with PCATE and the unconditional ATE.

lines, which is computed using the method in Firpo (2007). Clearly, it can be seen from Figure 4 that the estimated PCQTEs for blacks all decrease slightly (however, statistically insignificant) as z increases but three curves are almost same and lying in the 95%-confidence interval of the unconditional QTE for $\tau = 0.5$. In other words, $\widehat{\Delta}_\tau(z)$ may not depend statistically on z for all quantiles considered.

Finally, in addition to estimating the PCQTE above, we also investigate the partially conditional quantile treatment effect conditional on the mother's age and the estimation results of the PCQTT curves are displayed in Figure 5 for whites (the left panel) and blacks (the right panel), respectively, with the estimated PCQTT curves across mothers' ages for two quantile levels $\tau = 0.25$ (the long dashed line) and $\tau = 0.5$ (the short dashed line). For an easy comparison, in Figure 5, we plot the estimated PCATT[3] (partially conditional average treatment effect on the treated) curve by the solid line, together with its 95% confidence interval indicated by the the dotted lines, and the estimated unconditional average treatment effect on the treated group (ATT) as well (the dashed-dotted line). One can observe from Figure 5 that there exists substantially heterogeneity across different mothers' ages for two estimated PCQTT curves considered. To be specific, the numerical values of the estimated $\widehat{\Delta}_{\tau|D=1}(z)$ for two τ values considered increase as mothers' ages increase; that is, the esti-

[3]The estimation of PCATT is computed based on the method in Abrevaya et al. (2015).

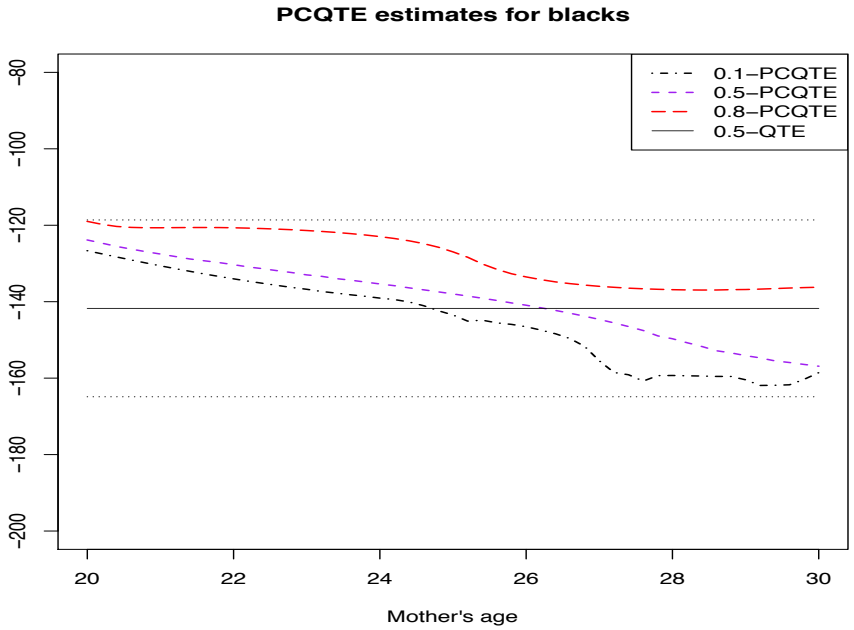

Figure 4: Estimation results for PCQTE for blacks for three quantile levels $\tau = 0.10$ (the dashed-dotted line), $\tau = 0.50$ (the short dashed line) and $\tau = 0.80$ (the long dashed line), together with unconditional QTE for $\tau = 0.5$ (solid line) and its 95% confidence interval (dotted lines).

mated quantile effects become stronger (more negative) at higher mothers' ages. Also, for a given mothers' age, the numerical values of the PCQTT point estimates at lower quantiles are bigger than that at the median. More specifically, the estimated results displayed in Figure 6 also show that the estimated quantile effects become stronger (more negative) at higher mothers' ages and for a given mothers' age, the numerical values of the PCQTT point estimates at lower quantiles are bigger than that at the median.

5 Conclusion

In this paper, we consider estimation for the partially conditional quantile treatment effect, a functional parameter designed to capture the variation in the quantile treatment effect conditional on some covariate(s). We propose a new estimation method and establish the asymptotic theory for the proposed semiparametric estimator. Using the proposed semiparametric estimator, we estimate the quantile effect of the first-time mother's smoking on her baby's birth weight conditional on the mother's age. We find that smoking has a more negative impact at higher ages or at lower quantile levels for whites. Meanwhile, we also find that the partially conditional quantile treatment effects for whites change over mothers' ages but not for blacks for some quantile levels considered.

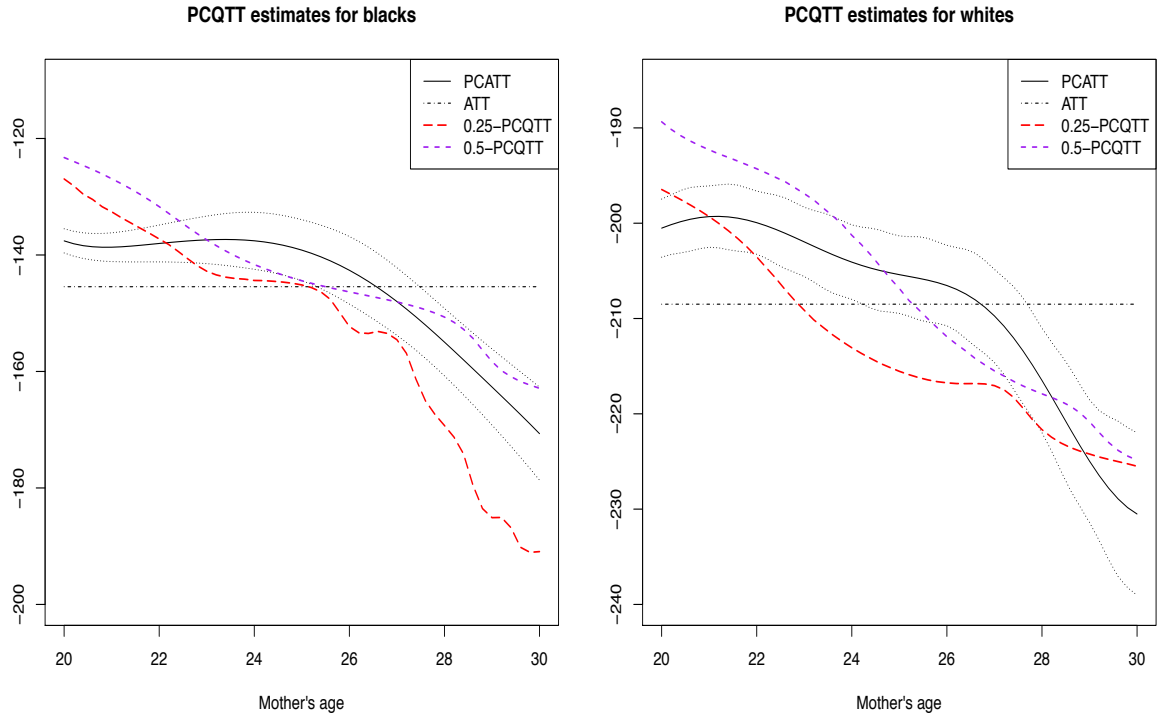

Figure 5: Estimation results for PCQTT for whites (the left panel) and blacks (the right panel) for two quantile levels $\tau = 0.25$ (the long dashed line) and $\tau = 0.50$ (the short dashed line), together with the unconditional ATT (the dashed-dotted line) and PCATT (the solid line) and its 95% confidence interval (dotted lines).

Therefor, it needs to investigate whether there exists heterogeneity in quantile treatment effects for covariate Z. To this end, one might consider the following hypothesis testing problem:

$$H_0 : \Delta_\tau(z) = \Delta_\tau \quad \text{for all } z \in \mathcal{Z} \text{ versus} \quad H_1 : \Delta_\tau(z) \neq \Delta_\tau \quad \text{for some } z \in \mathcal{Z},$$

where Δ_τ is the τ-th unconditional quantile treatment effect and $\mathcal{Z}$ is the domain of Z. Under the null hypothesis, the conditional quantile effect of the treatment equals to the unconditional QTE for all z, whereas, under the alternative, there is at least one value of z under which the conditional quantile treatment effect $\Delta_\tau(z)$ differs from Δ_τ. We leave this as a future research topic.

References

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abadie, A. and Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, 84(2):781–807.

Abrevaya, J. (2006). Estimating the effect of smoking on birth outcomes using a matched panel data approach. *Journal of Applied Econometrics*, 21(4):489–519.

Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics*, 26(4):379–397.

Abrevaya, J., Hsu, Y.-C., and Lieli, R. P. (2015). Estimating conditional average treatment effects. *Journal of Business & Economic Statistics*, 33(4):485–505.

Almond, D., Chay, K. Y., and Lee, D. S. (2005). The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1083.

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. *Handbook of Labor Economics*, 3:1277–1366.

Blundell, R. and Dias, M. C. (2002). Alternative approaches to evaluation in empirical microeconomics. *Portuguese Economic Journal*, 1(2):91–115.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 101(485):1595–1608.

Card, D. (1996). The effect of unions on the structure of wages: A longitudinal analysis. *Econometrica*, 64(4):957–979.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2008). Nonparametric tests for treatment effect heterogeneity. *Review of Economics and Statistics*, 90(3):389–405.

Currie, J. and Almond, D. (2011). Human capital development before age five. *Handbook of Labor Economics*, 4:1315–1486.

da Veiga, P. V. and Wilder, R. P. (2008). Maternal smoking during pregnancy and birthweight: A propensity score matching approach. *Maternal and Child Health Journal*, 12(2):194–203.

Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Annals of Statistics*, 2(2):267–277.

Donald, S. G. and Hsu, Y.-C. (2014). Estimation and inference for distribution functions and quantile functions in treatment effect models. *Journal of Econometrics*, 178(3):383–397.

Durrett, R. (1996). *Probability: Theory and Examples*. Second Edition. Duxbury Press, Belmont, CA.

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

Freeman, R. B. (1980). Unionism and the dispersion of wages. *ILR Review*, 34(1):3–23.

Frölich, M. and Melly, B. (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*, 31(3):346–357.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017–1098.

Heckman, J. J., Lalonde, R. J., and Smith, J. A. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics*, 3:1865–2097.

Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1-2):239–267.

Heckman, J. J., Smith, J., and Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies*, 64(4):487–535.

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

Ichimura, H. and Linton, O. (2005). Asymptotic expansions for some semiparametric program evaluation estimators. *In Andrews, D. W. K. and Stock, J. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, 149–170. Cambridge University Press.*

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge, UK.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Koenker, R. and Xiao, Z. (2004). Unit root quantile autoregression inference. *Journal of the American Statistical Association*, 99(467):775–787.

Kramer, M. S. (1987). Intrauterine growth and gestational duration determinants. *Pediatrics*, 80(4):502–511.

LaLonde, R. J. (1995). The promise of public sector-sponsored training programs. *Journal of Economic Perspectives*, 9(2):149–168.

Lechner, M. (1999). Earnings and employment effects of continuous gff-the-job training in east germany after unification. *Journal of Business & Economic Statistics*, 17(1):74–90.

Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics*, 84(2):205–220.

Lee, S., Okui, R., and Whang, Y.-J. (2017). Doubly robust uniform confidence band for the conditional average treatment effect function. *Journal of Applied Econometrics*, 32(7):1207–1225.

Lehmann, E. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

Tang, S. (2020). *Inferences for Partially Conditional Quantile Treatment Effects*. Ph.D. Dissertation:Department of Statistics, Xiamen University.

Appendix: Mathematical Proofs

Recall that $W_0(X_i, D_i) = \frac{1-D_i}{1-p(X_i)}$, $W_1(X_i, D_i) = \frac{D_i}{p(X_i)}$ and $\widehat{W}_{n,0}(X_i, D_i) = \frac{1-D_i}{1-\widehat{p}_n(X_i)}$, $\widehat{W}_{n,1}(X_i, D_i) = \frac{D_i}{\widehat{p}_n(X_i)}$, where $\widehat{p}_n(x) = p(x; \widehat{\theta}_n)$ is the parametric estimator of the propensity score function using (X_i, D_i), $i = 1, \cdots, n$. To prove Theorem 2.1, we need the following lemma.

Lemma 1. *For $j = 0$ and 1, consider random functions*

$$\Gamma_{n,j}(q, z) = \sum_{i=1}^{n} hK_{h,i}(z)\widehat{W}_{n,j}(X_i, D_i)\Big[\rho_\tau(Y_i; q) - \rho_\tau\big(Y_i; q_{j,\tau}(z)\big)\Big]$$

and

$$\begin{aligned}\widetilde{\Gamma}_{n,j}(q, z) \;=\; & \sum_{i=1}^{n} hK_{h,i}(z)W_j(X_i, D_i)\varphi_\tau\big(Y_i; q_{j,\tau}(z)\big)\big(q - q_{j,\tau}(z)\big) \\ & + \frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2} \cdot nh\big(q - q_{j,\tau}(z)\big)^2,\end{aligned}$$

where $K_{h,i}(z) = K\big((Z_i - z)/h\big)/h$ and $\varphi_\tau(y; q) = I(y \le q) - \tau$. Under Assumptions 2.1-2.5, one has

$$\sup_{|q - q_{j,\tau}(z)| \le \varepsilon/\sqrt{nh}} \Big|\Gamma_{n,j}(q, z) - \widetilde{\Gamma}_{n,j}(q, z)\Big| = o_p(1)$$

for any $z \in \mathcal{Z}$ and any $\varepsilon > 0$.

Proof of Lemma 1: By the definition of $\rho_\tau(y; q)$ and $\varphi_\tau(y; q)$, we can write

$$\begin{aligned}\Gamma_{n,j}(q, z) \;=\; & \sum_{i=1}^{n} hK_{h,i}(z)\widehat{W}_{n,j}(X_i, D_i)\Big[\varphi_\tau\big(Y_i; q_{j,\tau}(z)\big)\big(q - q_{j,\tau}(z)\big) \\ & + (Y_i - q)\big(I\{Y_i \le q_{j,\tau}(z)\} - I\{Y_i \le q\}\big)\Big].\end{aligned}$$

Therefore,

$$\begin{aligned}& \sup_{|q - q_{j,\tau}(z)| \le \varepsilon/\sqrt{nh}} \Big|\Gamma_{n,j}(q, z) - \widetilde{\Gamma}_{n,j}(q, z)\Big| \\ \le\; & \sup_{|q - q_{j,\tau}(z)| \le \varepsilon/\sqrt{nh}} \left\{\big|q - q_{j,\tau}(z)\big| \cdot \sum_{i=1}^{n} hK_{h,i}(z) \cdot \Big|\widehat{W}_{n,j}(X_i, D_i) - W_j(X_i, D_i)\Big| \cdot \Big|\varphi_\tau\big(Y_i; q_{j,\tau}(z)\big)\Big|\right\}\end{aligned}$$

$$
\begin{aligned}
&+ \sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}} \Bigg| \sum_{i=1}^{n} hK_{h,i}(z)\widehat{W}_{n,j}(X_i,D_i)(Y_i-q)\big(I\{Y_i\le q_{j,\tau}(z)\}-I\{Y_i\le q\}\big) \\
&-\frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2}\cdot nh\big(q-q_{j,\tau}(z)\big)^2\Bigg| \\
:= \;& \mathcal{A}_1+\mathcal{A}_2. \qquad (A.1)
\end{aligned}
$$

First, we consider $\mathcal{A}_1$. Note that $\sup_{x\in\mathcal{X}}\left|\widehat{W}_{n,j}(x,D_i)-W_j(x,D_i)\right| = O_p(n^{-1/2})$ and $\left|\varphi_\tau\big(Y_i;q_{j,\tau}(z)\big)\right|$ is bounded, it is easy to show

$$
\begin{aligned}
\mathcal{A}_1 &= \sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}} \left\{\big|q-q_{j,\tau}(z)\big|\cdot\sum_{i=1}^{n} hK_{h,i}(z)\cdot\left|\widehat{W}_{n,j}(X_i,D_i)-W_j(X_i,D_i)\right|\cdot\left|\varphi_\tau\big(Y_i;q_{j,\tau}(z)\big)\right|\right\} \\
&\le \frac{\varepsilon}{\sqrt{nh}}\cdot\sum_{i=1}^{n} hK_{h,i}(z)\cdot O_p(n^{-1/2})\cdot O(1) = O_p(h^{1/2})\cdot\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z).
\end{aligned}
$$

Since $\frac{1}{n}\sum_{i=1}^{n}K_{h,i}=O_p(1)$, it is easy to show that

$$
\mathcal{A}_1 \;=\; O_p(h^{1/2})\cdot O_p(1) = o_p(1). \qquad (A.2)
$$

Now, we move to $\mathcal{A}_2$. Define $\Psi(y;q_1,q_2)=(y-q_1)\big(I\{y\le q_2\}-I\{y\le q_1\}\big)$. Then,

$$
\begin{aligned}
&\mathcal{A}_2 \;= \\
&\sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}}\left|\sum_{i=1}^{n} hK_{h,i}(z)\widehat{W}_{n,j}(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))-\frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2}\cdot nh\big(q-q_{j,\tau}(z)\big)^2\right| \\
&\le \sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}}\left|\sum_{i=1}^{n} hK_{h,i}(z)\Big[\widehat{W}_{n,j}(X_i,D_i)-W_j(X_i,D_i)\Big]\Psi(Y_i;q,q_{j,\tau}(z))\right| \\
&+ \sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}}\left|\sum_{i=1}^{n} hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))-\frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2}\cdot nh\big(q-q_{j,\tau}(z)\big)^2\right| \\
&:= \mathcal{A}_{21}+\mathcal{A}_{22}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}}\left|\Psi(Y_i;q,q_{j,\tau}(z)\right| \;&=\; \sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}}\left|(Y_i-q)\big(I\{Y_i\le q_{j,\tau}(z)\}-I\{Y_i\le q\}\big)\right| \\
&\le \sup_{|q-q_{j,\tau}(z)|\le\varepsilon/\sqrt{nh}}\left|q-q_{j,\tau}(z)\right| = \varepsilon/\sqrt{nh}.
\end{aligned}
$$

By the similar argument to show $\mathcal{A}_1=o_p(1)$, we also have $\mathcal{A}_{21}=o_p(1)$. Next, we focus on

the term $\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))$ in $\mathcal{A}_{22}$. Indeed,

$$\begin{aligned}
& E\Big[\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi\big(Y_i;q,q_{j,\tau}(z)\big)\Big] \\
= \;& nE\Big[hK_{h,i}(z)\Psi(Y_i(j);q,q_{j,\tau}(z))\Big] \\
= \;& nh\cdot E\Big\{K_{h,i}(z)E\Big[(Y_i(j)-q)\big(I\{Y_i(j)\le q_{j,\tau}(z)\}-I\{Y_i(j)\le q\}\big)\Big|Z_i\Big]\Big\} \\
= \;& nh\cdot E\Big\{K_{h,i}(z)\int_q^{q_{j,\tau}(z)}(y-q)\,f_{Y(j)|Z}(y|Z_i)\,dy\Big\} \\
= \;& nh\cdot E\Big\{K_{h,i}(z)\int_q^{q_{j,\tau}(z)}(y-q)\,\big[f_{Y(j)|Z}(q_{j,\tau}(z)|Z_i)+O(|q_{j,\tau}(z)-q|)\big]\,dy\Big\} \\
= \;& nh\cdot\frac{(q_{j,\tau}(z)-q)^2}{2}\cdot E\Big\{K_{h,i}(z)\big[f_{Y(j)|Z}(q_{j,\tau}(z)|Z_i)+O(|q_{j,\tau}(z)-q|)\big]\Big\} \\
= \;& nh\cdot\frac{(q_{j,\tau}(z)-q)^2}{2}\cdot\big[f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)+O(|q_{j,\tau}(z)-q|)+o(1)\big].
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{Var}\Big[\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big] \\
= \;& n\mathrm{Var}\Big[hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big] \\
\le \;& n\cdot E\Big[hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big]^2 \\
= \;& nh^2\cdot E\Big[K_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big]^2 \\
= \;& nh\cdot O(1)\cdot E\Big\{h\,K_{h,i}^2(z)E\Big[(Y_i(j)-q)^2\big(I\{Y_i(j)\le q_{j,\tau}(z)\}-I\{Y_i(j)\le q\}\big)^2\Big|Z_i\Big]\Big\} \\
= \;& nh\cdot O(1)\cdot E\Big\{h\,K_{h,i}^2(z)\cdot\Big|\int_q^{q_{j,\tau}(z)}(y-q)^2\,f_{Y(j)|Z}(y|Z_i)\,dy\Big|\Big\} \\
= \;& nh\cdot O(1)\cdot O(|q_{j,\tau}(z)-q|^3).
\end{aligned}$$

Therefore, one can conclude that

$$\begin{aligned}
& \sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z)) \\
= \;& E\Big[\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big] \\
& + O_p\Big(\mathrm{Var}\Big[\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big]\Big)^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&= nh \cdot \frac{(q_{j,\tau}(z)-q)^2}{2} \cdot \big[f_Z(z) f_{Y(j)|Z}(q_{j,\tau}(z)|z) + O(|q_{j,\tau}(z)-q|) + o(1)\big] \\
&\quad + O_p\Big(nh \cdot |q_{j,\tau}(z)-q|^3\Big)^{1/2},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{A}_{22} &= \sup_{|q-q_{j,\tau}(z)|\le \varepsilon/\sqrt{nh}} \Big| \sum_{i=1}^{n} hK_{h,i}(z) W_j(X_i, D_i) \Psi(Y_i; q, q_{j,\tau}(z)) \\
&\qquad\qquad - \frac{f_Z(z) f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2} \cdot nh \cdot \big(q - q_{j,\tau}(z)\big)^2 \Big| \\
&= \sup_{|q-q_{j,\tau}(z)|\le \varepsilon/\sqrt{nh}} \Big| nh \cdot \frac{(q_{j,\tau}(z)-q)^2}{2} \cdot \big[O(|q_{j,\tau}(z)-q|) + o(1)\big] + O_p\Big(nh \cdot |q_{j,\tau}(z)-q|^3\Big)^{1/2} \Big| \\
&= o_p(1).
\end{aligned}
$$

Thus, one has the following result:

$$
\mathcal{A}_2 = \mathcal{A}_{21} + \mathcal{A}_{22} = o_p(1). \tag{A.3}
$$

It follows from (A.1), (A.2) and (A.3) that

$$
\sup_{|q-q_{j,\tau}(z)|\le \varepsilon/\sqrt{nh}} \Big| \Gamma_{n,j}(q,z) - \widetilde{\Gamma}_{n,j}(q,z) \Big| = o_p(1). \qquad \square
$$

Proof of Theorem 2.1: We first consider

$$
\begin{aligned}
\widetilde{q}_{j,\tau}(z) &= \arg\min_q \big\{ \widetilde{\Gamma}_{n,j}(q,z) \big\} \\
&= q_{j,\tau}(z) - \frac{1}{f_Z(z) f_{Y(j)|Z}(q_{j,\tau}(z)|z)} \cdot \frac{1}{n} \sum_{i=1}^{n} K_{h,i}(z) W_j(X_i, D_i) \varphi_\tau\big(Y_i; q_{j,\tau}(z)\big) \\
&= q_{j,\tau}(z) - \frac{1}{f_Z(z) f_{Y(j)|Z}(q_{j,\tau}(z)|z)} \cdot \frac{1}{n} \sum_{i=1}^{n} K_{h,i}(z) \psi_j(Y_i, X_i, D_i; z)
\end{aligned}
$$

for $j = 0$ and 1. By some calculations, one obtains

$$
E\big(\widetilde{q}_{j,\tau}(z)\big) = q_{j,\tau}(z) - \frac{h^2}{2} \mu_2(K) \delta_{j,\tau}(z) + o(h^2),
$$

where $\mu_2(K) = \int u^2 K(u)du$ and

$$\delta_{j,\tau}(z) = \frac{2f_Z'(z)\frac{\partial F_{Y(j)|Z}(q_{j,\tau}(z)|u)}{\partial u}\big|_{u=z} + f_Z(z)\frac{\partial^2 F_{Y(j)|Z}(q_{j,\tau}(z)|u)}{\partial u^2}\big|_{u=z}}{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)},$$

which leads to

$$\begin{aligned}
&\sqrt{nh}\Big(\widetilde{q}_{j,\tau}(z) - E\big(\widetilde{q}_{j,\tau}(z)\big)\Big) = \sqrt{nh}\Big(\widetilde{q}_{j,\tau}(z) - q_{j,\tau} + \frac{h^2}{2}\mu_2(K)\delta_{j,\tau}(z) + o(h^2)\Big) \\
= \; & -\frac{1}{\sqrt{nh}}\frac{1}{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)} \\
& \times \sum_{i=1}^{n}\Big[hK_{h,i}(z)\psi_j(Y_i, X_i, D_i; z) - E\big(hK_{h,i}(z)\psi_j(Y_i, X_i, D_i; z)\big)\Big]. \qquad \text{(A.4)}
\end{aligned}$$

Next, we consider the difference between $\widetilde{q}_{j,\tau}(z)$ and $\widehat{q}_{j,\tau}(z)$, where

$$\begin{aligned}
\widehat{q}_{j,\tau}(z) &= \arg\min_q \sum_{i=1}^{n} hK_{h,i}(z)\widehat{W}_{n,j}(X_i, D_i)\rho_\tau(Y_i; q) \\
&= \arg\min_q \sum_{i=1}^{n} hK_{h,i}(z)\widehat{W}_{n,j}(X_i, D_i)\Big[\rho_\tau(Y_i; q) - \rho_\tau\big(Y_i; q_{j,\tau}(z)\big)\Big] \\
&= \arg\min_q \big\{\Gamma_{n,j}(q, z)\big\}.
\end{aligned}$$

Since $\Gamma_{n,j}(q, z)$ is convex in q, it is easy to show that

$$\Big(1 - \frac{\epsilon/\sqrt{nh}}{|q - \widetilde{q}_{j,\tau}(z)|}\Big)\Gamma_{n,j}\big(\widetilde{q}_{j,\tau}(z), z\big) + \frac{\epsilon/\sqrt{nh}}{|q - \widetilde{q}_{j,\tau}(z)|}\,\Gamma_{n,j}(q, z) \geq \Gamma_{n,j}\Big(\widetilde{q}_{j,\tau}(z) + \frac{q - \widetilde{q}_{j,\tau}(z)}{|q - \widetilde{q}_{j,\tau}(z)|}\,\frac{\epsilon}{\sqrt{nh}}, z\Big)$$

for any $\epsilon > 0$ and $|q - \widetilde{q}_{j,\tau}(z)| > \epsilon/\sqrt{nh}$. Hence,

$$\begin{aligned}
& \frac{\epsilon/\sqrt{nh}}{|q - \widetilde{q}_{j,\tau}(z)|}\Big[\Gamma_{n,j}(q, z) - \Gamma_{n,j}\big(\widetilde{q}_{j,\tau}(z), z\big)\Big] \\
\geq \; & \Gamma_{n,j}\Big(\widetilde{q}_{j,\tau}(z) + \frac{q - \widetilde{q}_{j,\tau}(z)}{|q - \widetilde{q}_{j,\tau}(z)|}\,\frac{\epsilon}{\sqrt{nh}}, z\Big) - \Gamma_{n,j}(\widetilde{q}_{j,\tau}(z), z) \\
\geq \; & \widetilde{\Gamma}_{n,j}\Big(\widetilde{q}_{j,\tau}(z) + \frac{q - \widetilde{q}_{j,\tau}(z)}{|q - \widetilde{q}_{j,\tau}(z)|}\,\frac{\epsilon}{\sqrt{nh}}, z\Big) - \widetilde{\Gamma}_{n,j}(\widetilde{q}_{j,\tau}(z), z) \\
& -2\sup_{|u - \widetilde{q}_{j,\tau}(z)| \leq \epsilon/\sqrt{nh}}\Big|\Gamma_{n,j}(u, z) - \widetilde{\Gamma}_{n,j}(u, z)\Big|
\end{aligned}$$

for all $|q - \widetilde{q}_{j,\tau}(z)| > \epsilon/\sqrt{nh}$. Note that $\widetilde{\Gamma}_{n,j}(q, z)$ is a quadratic function of q and $\widetilde{q}_{j,\tau}(z) =$

$\arg\min_q \{\widetilde{\Gamma}_{n,j}(q,z)\}$. Then,

$$\begin{aligned}
&\frac{\epsilon/\sqrt{nh}}{|q-\widetilde{q}_{j,\tau}(z)|}\Big[\Gamma_{n,j}(q,z)-\Gamma_{n,j}\big(\widetilde{q}_{j,\tau}(z),z\big)\Big] \\
\geq\ & \frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2}\cdot\epsilon^2-2\sup_{|u-\widetilde{q}_{j,\tau}(z)|\leq\epsilon/\sqrt{nh}}\Big|\Gamma_{n,j}(u,z)-\widetilde{\Gamma}_{n,j}(u,z)\Big| \\
\geq\ & \frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2}\cdot\epsilon^2-2\sup_{|u-q_{j,\tau}(z)|\leq\epsilon/\sqrt{nh}+|q_{j,\tau}(z)-\widetilde{q}_{j,\tau}(z)|}\Big|\Gamma_{n,j}(u,z)-\widetilde{\Gamma}_{n,j}(u,z)\Big|
\end{aligned}$$

for all $|q-\widetilde{q}_{j,\tau}(z)|>\epsilon/\sqrt{nh}$. Since $|q_{j,\tau}(z)-\widetilde{q}_{j,\tau}(z)|=O_p(1/\sqrt{nh})$ from (A.4) and Assumption 2.4, together with Lemma 1,

$$\frac{\epsilon/\sqrt{nh}}{|q-\widetilde{q}_{j,\tau}(z)|}\Big[\Gamma_{n,j}(q,z)-\Gamma_{n,j}\big(\widetilde{q}_{j,\tau}(z),z\big)\Big]\geq\frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2}\cdot\epsilon^2+o_p(1)$$

for all $|q-\widetilde{q}_{j,\tau}(z)|>\epsilon/\sqrt{nh}$. Since $\Gamma_{n,j}\big(\widehat{q}_{j,\tau}(z),z\big)-\Gamma_{n,j}\big(\widetilde{q}_{j,\tau}(z),z\big)\leq 0$ by the definition that $\widehat{q}_{j,\tau}(z)=\arg\min_q\{\Gamma_{n,j}\big(q,z\big)\}$, one can show that

$$\begin{aligned}
&P\Big(\sqrt{nh}|\widehat{q}_{j,\tau}(z)-\widetilde{q}_{j,\tau}(z)|>\epsilon\Big) \\
\leq\ & P\Big(\inf_{|q-\widetilde{q}_{j,\tau}(z)|>\epsilon/\sqrt{nh}}\Big\{\Gamma_{n,j}(q,z)-\Gamma_{n,j}\big(\widetilde{q}_{j,\tau}(z),z\big)\Big\}\leq 0\Big) \\
\leq\ & P\Big(\frac{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2}\cdot\epsilon^2+o_p(1)\leq 0\Big)\to 0,
\end{aligned}$$

which implies $\widehat{q}_{j,\tau}(z)=\widetilde{q}_{j,\tau}(z)+o_p(1/\sqrt{nh})$. It follows by combining (A.4) and $\widehat{q}_{j,\tau}(z)=\widetilde{q}_{j,\tau}(z)+o_p(1/\sqrt{nh})$ that

$$\begin{aligned}
&\sqrt{nh}\left[\widehat{\Delta}_\tau(z)-\Delta_\tau(z)+\frac{h^2}{2}\mu_2(K)\delta_\tau(z)+o_p(h^2)\right] \\
&\sqrt{nh}\left[\widetilde{\Delta}_\tau(z)-\Delta_\tau(z)+\frac{h^2}{2}\mu_2(K)\delta_\tau(z)+o_p(h^2)+\widehat{\Delta}_\tau(z)-\widetilde{\Delta}_\tau(z)\right] \\
=\ & -\frac{1}{\sqrt{nh}}\frac{1}{f_Z(z)}\sum_{i=1}^n\left\{\frac{hK_{h,i}(z)\psi_1(Y_i,X_i,D_i,z)-E\Big(hK_{h,i}(z)\psi_1(Y_i,X_i,D_i,z)\Big)}{f_{Y(1)|Z}(q_{1,\tau}(z)|z)}\right. \\
&\qquad\left.-\frac{hK_{h,i}(z)\psi_0(Y_i,X_i,D_i,z)-E\Big(hK_{h,i}(z)\psi_0(Y_i,X_i,D_i,z)\Big)}{f_{Y(0)|Z}(q_{0,\tau}(z)|z)}\right\}+o_p(1),
\end{aligned}$$

where $\widetilde{\Delta}_\tau(z)=\widetilde{q}_{1,\tau}(z)-\widetilde{q}_{0,\tau}(z)$. By the fact that

$$E\Big[hK_{h,i}(z)\psi_j(Y_i,X_i,D_i;z)-E\big(hK_{h,i}(z)\psi_j(Y_i,X_i,D_i;z)\big)\Big]=0,$$

and the Lyapunov's central limit theorem, we can easily show that

$$\sqrt{nh}\left[\widehat{\Delta}_\tau(z) - \Delta_\tau(z) + \frac{h^2}{2}\mu_2(K)\delta_\tau(z) + o_p(h^2)\right] \xrightarrow{D} \mathcal{N}\Big(0, ||K||_2^2\sigma_\tau^2(z)/f_Z(z)\Big).$$

This completes the proof. □