Inferences for Partially Conditional Quantile Treatment Effect Model[*][†]

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.

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

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

January 16, 2020

Abstract: In this paper, a new model, termed as the partially conditional quantile treatment effect (PCQTE) model, is proposed to characterize the heterogeneity of treatment effect conditional on some predetermined variable(s). We show that the 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 effect. Under some regularity conditions, the consistency and asymptotic normality of the proposed semiparametric estimator are derived. In addition, a specification test is seminally proposed in quantile regression literature, to test whether there exits heterogeneity for PCQTE across sub-populations, a consistent test, based on the Cramér-von Mises type criterion. The asymptotic properties of the proposed test statistic are investigated, including consistency and asymptotic normality. Finally, the performance of the proposed methods is illustrated through Monte-Carlo experiments and an empirical application on estimating the effect of the first-time mother's smoking during pregnancy on the baby's birth weight conditional on mother's age and testing whether the partially conditional quantile treatment effect varies across different mother's age.

Keywords: Conditional quantile treatment effect; Heterogeneity; Specification test; Propensity score; Semiparametric estimation.

JEL classification: C12; C13; C14; C23

[*]The authors acknowledge the financial supports from the National Science Fund of China (NSFC) for Distinguished Scholars (71625001) and the NSFC key projects with grant numbers 71631004 and 71131008.

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

1 Introduction

Understanding the causal effect of a treatment or a policy, 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) 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 cannot 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 method 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 the 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. Moreover, it is also important for policy makers to understand how the heterogenous effects of the participation into 401(k) programs on financial assets for families with different incomes and/or 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. More detailedly, Abrevaya, Hsu and Lieli (2015) proposed using a nonparametric method to estimate the PCATE, whereas Lee, Okui and Whang (2017) suggested a doubly roust 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) for characterizing 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). We show that the PCQTE is nonparametrically identified and a semiparametric estimation is provided. Furthermore, a specification test is conducted for testing whether there exists heterogeneity in quantile effects across sub-populations defined by Z. To be specific, a test statistic is proposed based on the Cramér-von Mises type criterion to test whether the PCQTE conditional on Z is equal to the corresponding unconditional quantile treatment effect. To the best of our knowledge, it is believed that this test is novel in the quantile regression literature, although there are some papers in the literature of testing treatment effect heterogeneity under the framework of average treatment effect but not quantile treatment effect. For example, Crump, Hotz, Imbens and Mitnik (2008) developed two nonparametric tests based on series approach, in which the first is to test whether a treatment has a zero average effect for all sub-populations defined by covariates, and the second is to test whether the average treatment effect conditional on the covariates is identical for all sub-populations, in other words, whether there is heterogeneity in average treatment effects by covariates.

Our motivation comes actually from exploring an empirical example for estimating the quantile treatment effect of first-time mothers' smoking status during pregnancy on birth weight conditional on their ages. Abrevaya et al. (2015) and Lee et al. (2017) considered the case of investigating the average treatment effect 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 continuous covariates. A semiparametric estimator was proposed if the propensity score function is estimated parametrically in the first stage, and the authors also provided a fully nonparametric estimator 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. But the aforementioned papers do not address the heterogeneity issue. Therefore, in this paper, we re-analyze this real example by using the proposed methods for the PCQTE model. As a result, our findings look very interesting and novel in the literature, and further, they provide different interpretations to this application. First, the smoking quantile effects become stronger, more negative on birth weights, at higher ages, and for whites, the estimated values at lower quantiles are bigger than those at the median or higher quantiles, conditional on mothers' ages. Secondly, when the proposed test is applied to test whether the quantile effects varies across ages, it turns out that the quantile treatment effects of whites changes over ages but not for blacks for the quantile levels considered.

The rest of the 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. Section 3 develops a specification test for testing whether there exists heterogeneity by some covariates. Monte Carlo simulations are conducted in Section 4 to illustrate the finite sample performances of the proposed estimators and test statistic, and Section 5 is devoted to an empirical study to investigate how the distributional effect of maternal smoking on birth weights varies across different groups of mothers. Section 6 concludes. The proofs of the main results are delegated to mathematical appendices.

2 Partially Conditional Quantile Treatment Effect Model

2.1 Model Setup

We consider the effect of a treatment on a continuous outcome variable. 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. Hence, the observed outcome variable Y_i can be written as

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

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", "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, 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 $1 \leq k < L$, and then, the τ-th partially conditional quantile treatment effect (PCQTE), where $\tau \in (0, 1)$, is defined as

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

where for $j = 0$ and 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.

2.2 Estimation Procedure

Since the potential outcomes $Y_i(0)$ and $Y_i(1)$ are not observable for every individual, $Y_1(j), \cdots, Y_n(j)$ can not directly used 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 leads to that $q_{j,\tau}(z)$, $j = 0$ and 1, can be easily expressed as

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

where $\rho_\tau(Y, q) = (Y - q)\{\tau - I(Y \le q)\}$ 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, from (2.2), a two-step estimation procedure is needed for estimating $\Delta_\tau(z)$. Firstly, one needs to obtain the estimated propensity score function $\widehat{p}(x)$ using (X_i, D_i), $i = 1, \cdots, n$, and then, at the second stage, the kernel-based local average method is used to estimate $q_{j,\tau}(z)$ and $\Delta_\tau(z)$. Specifically,

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

where for $j = 0$ and 1,

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

with $\widehat{W}_{n,0}(X_i, D_i) = (1 - D_i)/[1 - \widehat{p}(X_i)]$, $\widehat{W}_{n,1}(X_i, D_i) = D_i/\widehat{p}(X_i)$, and $K_h(u) = K(u/h)/h$. Here, $K(\cdot)$ is a kernel function, h is the bandwidth parameter, and $\widehat{p}(x)$ is a consistent estimate of $p(x)$. 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 for $p(x) = p(x; \theta)$, for example, a logit model or a probit model. The parameter θ can be easily estimated through the maximum likelihood method. The second approach is nonparametric[1].

[1] For a nonparametric method, one can use the so-called series logit estimator as in Hirano et al. (2003)

The first one is used in this paper so that the estimation in (2.3) is called a semiparametric estimator.

2.3 Asymptotic Theory

This subsection is devoted to investigating the asymptotic properties for the semiparametric estimator for PCQTE 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) The support of X_i, denoted by $\mathcal{X}$, is a Cartesian product of compact intervals, that is, $\mathcal{X} = \prod_{l=1}^{L} [\underline{x}_l, \overline{x}_l]$, and there exists a constant $c > 0$, such that the density function of X_i, $f_X(x)$ satisfies $\inf_{x \in \mathcal{X}} f_X(x) \geq c$. 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 $\widehat{p}(x) = p(x; \widehat{\theta}_n)$ 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 2.1 is commonly used in the literature on treatment effect evaluation; see Hirano et al. (2003), Abadie and Imbens

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 $p(x)$. Of course, we conjecture that the first-order asymptotic results displayed in the following section do not depend critically on the choice of $\widehat{p}(x)$. That is, the similar conclusions as in Theorem 2.1 could be obtained under similar conditions for a nonparametric estimation of $p(x)$.

(2006, 2016), Firpo (2007), Abrevaya et al. (2015), and among others. Assumption 2.2 guarantees the solution of (2.2) is unique and the smoothness conditions imposed in Assumption 2.3 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 the Appendix. For easy presentation, define some notation 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 ith order derivative $F_j^{(i)}(y|z) = \partial^i F_j(y\,|z)/\partial z^i$ for $i \geq 0$. 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,

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

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

$$\begin{aligned}
&\sqrt{nh}\left[\widehat{\Delta}_\tau(z) - \Delta_\tau(z) + \frac{h^2}{2}\mu_2(K)\delta_\tau(z) + o_p(h^2)\right] \\
= \quad & -\frac{1}{\sqrt{nh}}\frac{1}{f_Z(z)}\sum_{i=1}^n\left\{\frac{K_i\psi_1(Y_i,X_i,D_i,z) - E\Big(K_i\psi_1(Y_i,X_i,D_i,z)\Big)}{f_{Y(1)|Z}(q_{1,\tau}(z)|z)}\right. \\
&\left. - \frac{K_i\psi_0(Y_i,X_i,D_i,z) - E\Big(K_i\psi_0(Y_i,X_i,D_i,z)\Big)}{f_{Y(0)|Z}(q_{0,\tau}(z)|z)}\right\} + o_p(1) \qquad (2.6) \\
\xrightarrow{D} \quad & \mathcal{N}\Big(0, ||K||_2^2\sigma_\psi^2(z)/f_Z(z)\Big), \qquad (2.7)
\end{aligned}$$

where $K_i = K\big((Z_i - z)/h\big)$, $\mu_2(K) = \int u^2K(u)du$, $||K||_2^2 = \int K^2(u)du$, and

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

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 pointwise confidence interval for making an easy statistical inference. To construct a pointwise confidence interval for $\Delta_\tau(z)$ for each given z, by ignoring the asymptotic bias term, $h^2\mu_2(K)\delta_\tau(z)/2$, one needs some consistent estimations for $f_Z(z)$ and $\sigma^2_\psi(z)$, respectively. Clearly, the density function of Z_i can be estimated by the kernel density estimator as $\widehat{f}_Z(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i - z)$. But, it is much more involved to estimate $\sigma^2_\psi(z)$ 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 $f_{Y(j)|Z}(q_{j,\tau}(z)|z)$. Indeed, there are two methods available in the literature. For example, DiNardo, Fortin and Lemieux (1996) and Firpo (2007) used the re-weighted kernel method to estimate the density function of $Y_i(j)$ conditional on $Z_i = z$, whereas Koenker (2005) proposed using the following estimator 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 $h^*\sqrt{nh} \to \infty$. Therefore, a consistent estimate of $\sigma^2_\psi(z)$ can be given by

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

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, it is easy to compute a pointwise confidence interval for $\Delta_\tau(z)$ by ignoring the asymptotic bias and this is implemented in Section 5 for an empirical study; see Figure 6(b) in Section 5 for detail.

3 Specification Test

It is of interest to investigate whether there exists heterogeneity in quantile treatment effects by covariates Z. To this end, we consider the following hypothesis testing problem:

$$H_0 : \Delta_\tau(z) = \Delta_\tau, \text{ for all } z \in \mathcal{Z} \quad \text{versus} \quad H_1 : \Delta_\tau(z) \neq \Delta_\tau, \text{ for some } z \in \mathcal{Z}, \tag{3.1}$$

where $\Delta_\tau = q_{1,\tau} - q_{0,\tau}$ with $q_{j,\tau}$ being the τ-th unconditional quantile of $Y_i(j)$, and $\mathcal{Z}$ is the support of Z_i. Under the null hypothesis, the conditional quantile effect of the treatment equals to the unconditional QTE for all z, whereas, under the alternative, there are at least some values of z under which the conditional quantile treatment effect $\Delta_\tau(z)$ differs from the unconditional QTE.

Next, we propose a test statistic for (3.1) based on the Cramér-von Mises criterion. To this end, let

$$J = \int \Big(\Delta_\tau(z) - \Delta_\tau\Big)^2 \omega(z) dz \geq 0, \tag{3.2}$$

where $\omega(z)$ is a pre-specified continuous and strictly positive weighting function. Note that $J = 0$ only and only if the null hypothesis in (3.1) is true. Hence, a test statistic using the sample analogue of J is defined by

$$J_n = \int \Big(\widehat{\Delta}_\tau(z) - \widehat{\Delta}_\tau\Big)^2 \omega(z) dz, \tag{3.3}$$

where $\widehat{\Delta}_\tau$ is a $\sqrt{n}$-consistent estimator for Δ_τ, such as the estimator proposed in Firpo (2007), and $\widehat{\Delta}_\tau(z)$ is the semiparametric estimator of $\Delta_\tau(z)$ in (2.3).

Remark 3.1. Except for the testing issue displayed in (3.1), one may be interested in testing whether the partially conditional quantile treatment effect model is correctly specified; that is, the more general interest than testing (3.1) is to consider the hypothesis testing problem

$$H_0 : \Delta_\tau(z) = \Delta_{\tau,0}(z; \theta_\tau) \text{ ; for all } \; z \in \mathcal{Z}$$

where $\Delta_{\tau,0}(\cdot)$ is a known function with unknown parameter θ_τ. In particular, one might have an interest in testing

$$H_0 : \Delta_\tau(z) \leq 0 \;\text{ or }\; \geq 0 \text{ for all } \; z \in \mathcal{Z},$$

which leads to studying the stochastic dominance such as $Y(1) \leq Y(0)$ or $Y(1) \geq Y(0)$ for all Z. These extensions are beyond the scope of this paper but certainly worth pursuing in future research.

The following theorem describes the asymptotic properties of the proposed test statistic J_n with their proofs given in the Appendix. To this end, now, define some notation. Let $\mu_0(z; u) = E\Big[(I\{Y(0) \leq q_{0,\tau}(u)\} - \tau)^2/[1 - p(X)]\big|Z = z\Big]$ and $\mu_1(z; u) = E\Big[(I\{Y(1) \leq q_{1,\tau}(u)\} - \tau)^2/p(X)\big|Z = z\Big]$.

Theorem 3.1. *Suppose Assumptions 2.1-2.5 are satisfied. We further assume that* $q_{j,\tau}(z)$, $j = 0, 1$, *is differentiable on* $z \in \mathcal{Z}$ *with bounded second order derivatives and* $nh^{9/2} \to 0$. *Then, under the null hypothesis* H_0, *one has*

$$n\sqrt{h}\big(J_n - \mu_J\big) \xrightarrow{D} \mathcal{N}(0, \sigma_J^2)$$

and under the alternative hypothesis H_1,

$$n\sqrt{h}\big(J_n - \mu_J\big) \xrightarrow{p} +\infty,$$

where

$$\mu_J = \frac{1}{nh}\int K^2(s)ds \int \left\{ \frac{\mu_1(z;z)}{f^2_{Y(1)|Z}(q_{1,\tau}(z)|z)} + \frac{\mu_0(z;z)}{f^2_{Y(0)|Z}(q_{0,\tau}(z)|z)} \right\} \frac{\omega(z)}{f_Z(z)} dz$$

and

$$\sigma^2_J = 2\int \Big(\int K(t)K(t+s)dt\Big)^2 ds \int \left\{ \frac{\mu_1(u;u)}{f^2_{Y(1)|Z}(q_{1,\tau}(u)|u)} + \frac{\mu_0(u;u)}{f^2_{Y(0)|Z}(q_{0,\tau}(u)|u)} \right\}^2 \frac{\omega^2(u)}{f^2_Z(u)} du.$$

Although the asymptotic distribution of J_n is established in the above theorem, the test based on the asymptotic distribution might be sensitive to the choice of bandwidth h and the consistent estimation of σ^2_J in small samples. In particular, it is well known in the literature that the consistent estimation of the conditional density of $Y(j)$ given Z is not an easy task; see, for example, Koenker and Xiao (2004) and Cai and Xu (2008). To overcome this difficulty, a Bootstrap method adopted by Chen, Linton and Van Keilegom (2003), Firpo, Galvao and Song (2017) and Lehrer, Pohl and Song (2018), is suggested to determine p-value for J_n. Specifically, first, generate B Bootstrapping samples by drawing samples from the original sample $\big\{(Y_i, X_i, D_i)\big\}_{i=1}^n$ with replacement, each denoted by $\big\{(Y_i^b, X_i^b, D_i^b)\big\}_{i=1}^n$, $b = 1, \cdots, B$, and then, compute $J_n^b := \int \big[(\widehat{\Delta}^b_\tau(z) - \widehat{\Delta}_\tau(z))\big]^2 dz$, $b = 1, \cdots, B$, where $\widehat{\Delta}^b_\tau(z)$ is the estimated PCQTE using the Bootstrapping sample b and $\widehat{\Delta}_\tau(z)$ is the estimated PCQTE using the original dataset, and finally, compute the Bootstrap p-value for J_n by $B^{-1} \sum\limits_{i=1}^{B} I\{J_n^b \geq J_n\}$.

4 Monte Carlo Studies

In this section, Monte Carlo experiments are conducted to examine the finite sample performance of the proposed estimation procedures and the proposed test J_n. The goal is to assess the finite sample accuracy in various scenarios. The scenarios examined in estimation differ in the choices of the bandwidth parameter h. It has been well known that the choice of the smoothing parameter in kernel-based or local polynomial estimation is important. To assess the performance of the finite sample sizes, for each settings, the mean absolute deviation error (MADE) criterion is used and which is defined as

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

where $\{z_j\}_{j=1}^J$ are the grid points taken from the support of Z with an equal increments. At last, the median and standard deviation of the MADE values for each setting are reported.

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

$$Y(0) = \alpha_0 X_1 + \beta_0 \exp(U_0 + c_0) X_2, \quad \text{and} \quad Y(1) = \alpha_1 X_1 + \beta_1 \exp(U_1 + c_1) X_2,$$

and the propensity score function is $p(X, \theta) = P(D = 1|X) = \exp(\theta' X)/[1 + \exp(\theta' X)]$, where $X = (1, X_1, X_2)'$, $\theta = (-0.5, 1.5, 0.5)'$, $\alpha_0 = -1.5$, $\beta_0 = 0.4$, $c_0 = -0.5$, $\alpha_1 = -0.8$, $\beta_1 = 0.5$, $c_1 = 0.2$, $X_1 \sim iid\ 2Beta(1, 2) - 1$, $X_2 \sim iid\ unif[0, 1]$ and $U_j \sim iid\ unif[0, 1]$ for $j = 0$ and 1.

Motivated by Assumption 2.4, $h = c \cdot n^{-1/5}$ is set for $c = \{0.25, 0.5, 1.0\}$ to illustrate how they affect the quality of the estimator. As presented in (2.3), the estimate $\widehat{\Delta}_\tau(z)$ is computed. Following the treatment effect literature, the estimated propensity score is often trimmed to prevent it from getting too close to the boundaries of the $[0, 1]$ interval. Therefore, in this paper, the estimated propensity score $\widehat{p}(x)$ is truncated to be between $[\varepsilon, 1 - \varepsilon]$ with $\varepsilon = 0.005$; that is, if the fitted value $\widehat{p}(X_i)$ is strictly less than the threshold ε, $\widehat{p}(X_i)$ is reset to be ε. Similarly, if $\widehat{p}(X_i)$ is strictly greater than $1 - \varepsilon$, $\widehat{p}(X_i)$ is set to be $1 - \varepsilon$. As comparison, the estimation results for the partially conditional average treatment effect (PCATE) are also considered as in Abrevaya et al. (2015). Finally, the conditional variable Z is taken to be X_1 in this example and the number of replications are 1,000 for all cases and the normal kernel function is used.

Table 1: Median and standard deviation (in parentheses).

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	0.115 (0.025)	0.110 (0.020)	0.079 (0.015)	0.104 (0.029)	0.091 (0.022)	0.073 (0.018)	0.103 (0.030)	0.098 (0.019)	0.084 (0.015)
0.25	0.097 (0.018)	0.083 (0.016)	0.067 (0.010)	0.094 (0.019)	0.081 (0.012)	0.065 (0.009)	0.099 (0.024)	0.085 (0.017)	0.078 (0.014)
0.5	0.092 (0.016)	0.078 (0.013)	0.061 (0.008)	0.081 (0.015)	0.068 (0.010)	0.059 (0.008)	0.094 (0.017)	0.078 (0.015)	0.067 (0.010)
0.75	0.107 (0.022)	0.091 (0.016)	0.072 (0.011)	0.102 (0.020)	0.088 (0.015)	0.071 (0.011)	0.101 (0.025)	0.094 (0.017)	0.081 (0.015)
0.9	0.132 (0.030)	0.128 (0.025)	0.111 (0.020)	0.127 (0.031)	0.119 (0.024)	0.107 (0.018)	0.125 (0.029)	0.111 (0.021)	0.104 (0.017)
CATE	0.110 (0.024)	0.091 (0.020)	0.072 (0.012)	0.098 (0.025)	0.085 (0.017)	0.068 (0.010)	0.101 (0.025)	0.097 (0.018)	0.084 (0.015)

Table 1 reports the simulation results for the semiparametric estimator. As seen in Table

1, the semiparametric estimator performs well in terms of MADE and the choice of the bandwidth h in a reasonable range seems to have little influence on the MADEs and its 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 its standard errors as sample size goes from $n = 500$ to $n = 2000$ in all cases, which is in line with the asymptotic theory.

Example 2. In this example, we investigate the finite-sample performance of the proposed test J_n. Also, the Skorohod representation is used to generate the potential outcomes $Y(1)$ and $Y(0)$. To be specific, the data generating process is given by

$$Y(1) = (\lambda + \lambda_0)X_1 + \gamma_1 \exp(U_1 + \ell_1)X_2, \quad Y(0) = \lambda_0 X_1 + \gamma_0 \exp(U_0 + \ell_0)X_2,$$

and the propensity score function is

$$p(X_1, X_2) = P(D = 1|X_1, X_2) = \exp(0.5 + X_1 - X_2)/[1 + \exp(0.5 + X_1 - X_2)],$$

where U_1 and U_0 are generated from *iid* $U(0, 1)$, X_1 and X_2 are generated from *iid* $U(-1, 1)$ and $U(0, 1)$ respectively. The constants above are set by $\lambda_0 = 1.0$, $\gamma_1 = 1.0$, $\gamma_0 = 0.5$, $\ell_0 = 0.2$ and $\ell_1 = -0.5$. The constant λ is varied in experiments.

In this example, the conditional variable Z is taken to be X_1 again. By a simple calculation, the exact conditional quantile functions for $Y(1)$ and $Y(0)$ conditional on $Z = z$, are given by, if $0 < \tau \leq 1 - e^{-1}$,

$$q_{1,\tau}(z) = \frac{\tau\gamma_1 e^{\ell_1}}{1 - e^{-1}} + (\lambda + \lambda_0)z, \quad q_{0,\tau}(z) = \frac{\tau\gamma_0 e^{\ell_0}}{1 - e^{-1}} + \lambda_0 z,$$

and if $1 > \tau > 1 - e^{-1}$,

$$q_{1,\tau}(z) = \gamma_1 e^{\ell_1}\omega + (\lambda + \lambda_0)z, \quad q_{0,\tau}(z) = \gamma_0 e^{\ell_0}\omega + \lambda_0 z,$$

where $\tau \in (0, 1)$ is the quantile level and ω is the solution of equation $\ln w - we^{-1} - (\tau - 1) = 0$. The power function is indexed by λ and it is easy to see that only when $\lambda = 0$, the partially conditional quantile treatment effect $\Delta_\tau(z)$ is a constant, which corresponds to the null hypothesis. The Bootstrap procedure outlined in Section 3 is used to determine the critical value. The number of Bootstrap replications is set as $B = 999$. To examine the the size and power properties of the test statistic J_n, three different sample sizes $n = 200$, $n = 400$ and $n = 800$ are considered. To check the sensitivity of the test with respect to different values of the bandwidth h, motivated by the asymptotic theory, $h = a \cdot n^{-1/4}$ is considered with

$a = 0.5$, 1.0 and 2.0. Finally, three quantiles of the distribution, namely, $\tau = 0.2$, 0.5 and $\tau = 0.8$, are computed. The estimated sizes and power of the test J_n are computed for $1,000$ simulations under the nominal size $\alpha = 5\%$, respectively.

Table 2: Estimated sizes of J_n (nominal size $\alpha = 5\%$)

		Empirical rejection probability of J_n with		
		$a = 0.5$		
λ	n	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$
0	200	0.024	0.030	0.023
	400	0.036	0.038	0.032
	800	0.045	0.048	0.042
		$a = 1.0$		
λ	n	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$
0	200	0.032	0.038	0.030
	400	0.036	0.043	0.037
	800	0.047	0.049	0.044
		$a = 2.0$		
λ	n	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$
0	200	0.032	0.036	0.030
	400	0.060	0.046	0.056
	800	0.046	0.052	0.042

The empirical sizes of the J_n test based on Bootstrap critical value are reported in Table 2 which presents the empirical rejection probabilities of the test J_n when the different values of bandwidth h are considered. It can be seen from Table 2 that the empirical sizes converge to their nominal sizes as the sample sizes n increases. Particularly, when the sample size increases to 800, the test J_n performs well in all cases considered. Also, one can observe that the choice of the bandwidth h seems to have little influence on empirical sizes.

Next, Figures 1-3 display the estimated power curves with nominal size $\alpha = 5\%$ of the J_n test for different quantiles and different choices of the bandwidth. In general, the test J_n performs reasonably powerful in detecting the deviation from the null in all cases considered. It is not surprising that the powers increase quickly with the sample size n or the value of λ increasing. It is also noticed from these figures that the bandwidth h in a certain range seems to have little impact on the power of the test J_n.

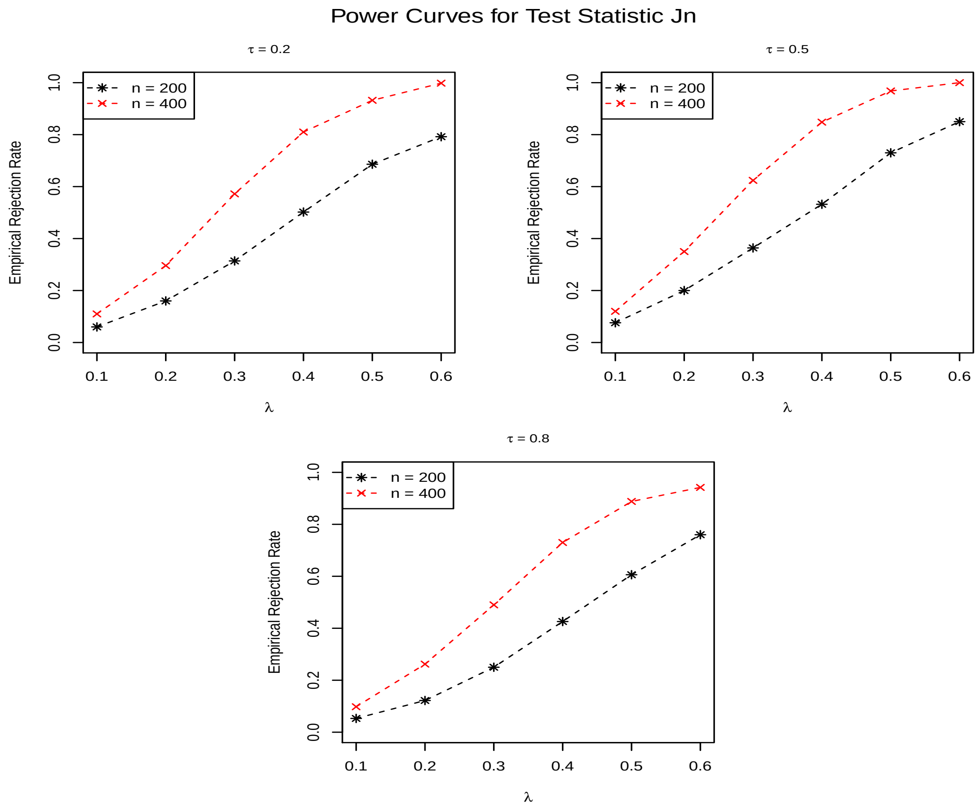

Figure 1: Power curves for test statistic J_n with nominal size $\alpha = 5\%$ and $a = 0.5$.

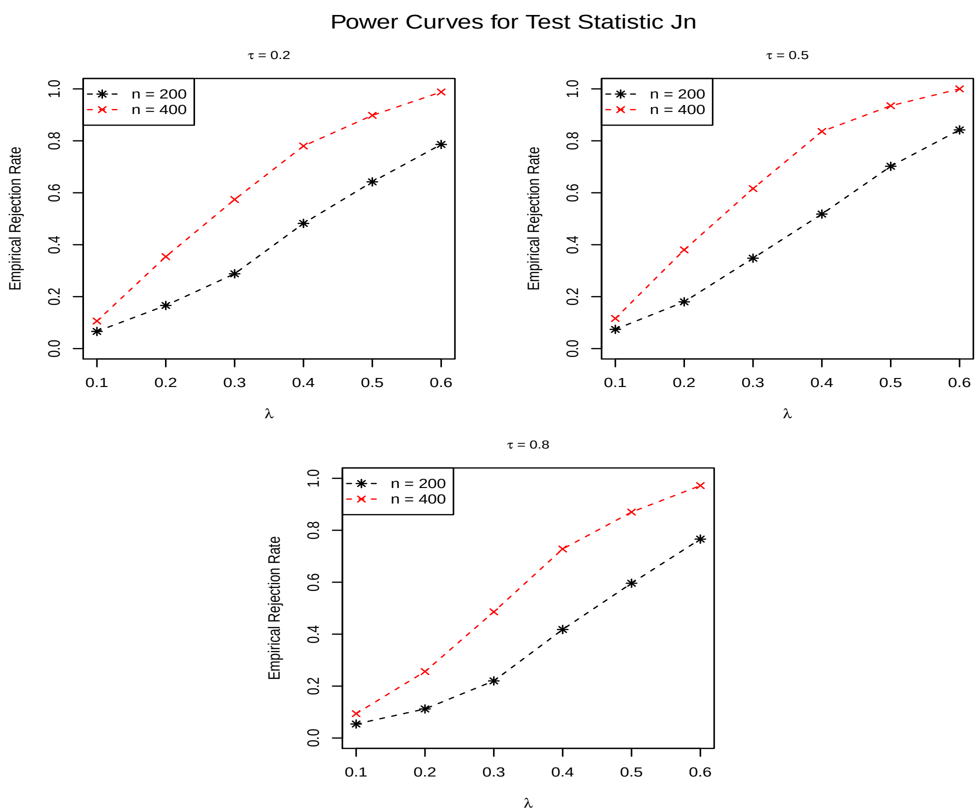

Figure 2: Power curves for test statistic J_n with nominal size $\alpha = 5\%$ and $a = 1.0$.

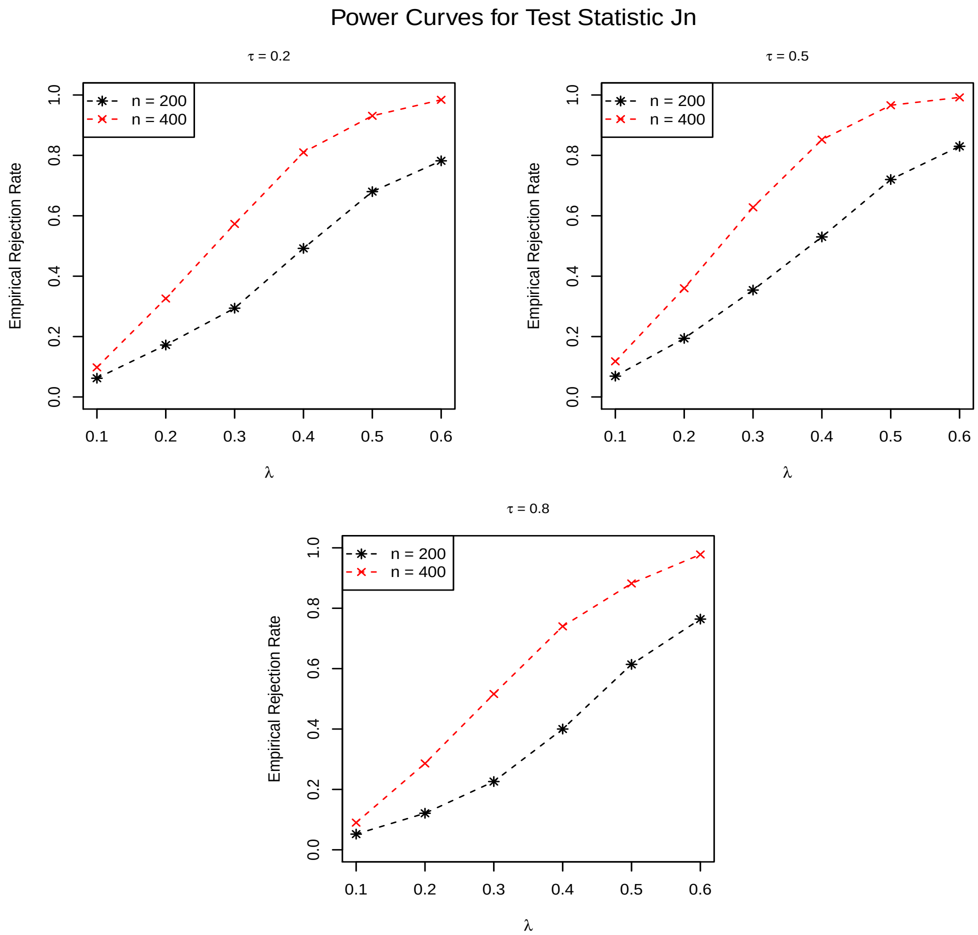

Figure 3: Power curves for test statistic J_n with nominal size $\alpha_0 = 5\%$ and $a = 2.0$.

5 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), da Veiga and Wilder (2008), Abrevaya (2006), 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 above 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 blacks which consist of a sample of $157,989$ observations and whites which consist of a sample of $433,558$ observations as separate populations throughout.

In this 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 4(a) for blacks and 4(b) for whites, respectively. For blacks and whites, 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 both the distributions of infant birth weights for blacks and whites are fat-tailed in the left side. Therefore, this motivates us to consider the distributional effect of maternal smoking on infant birth weight.

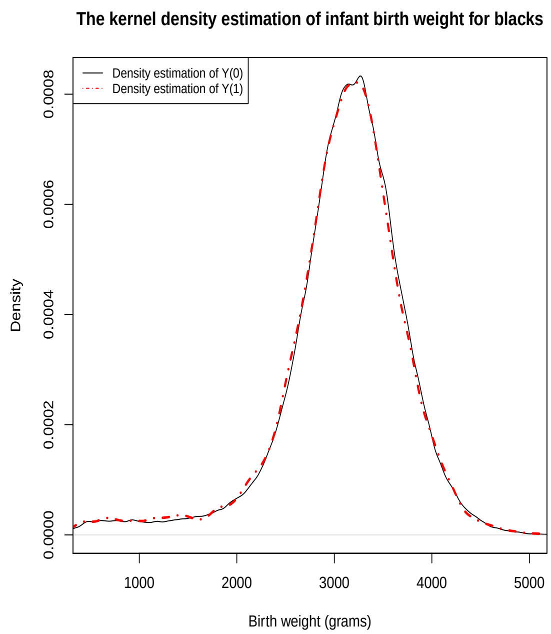

Figure 4(a): The kernel density estimation of birth weight for blacks

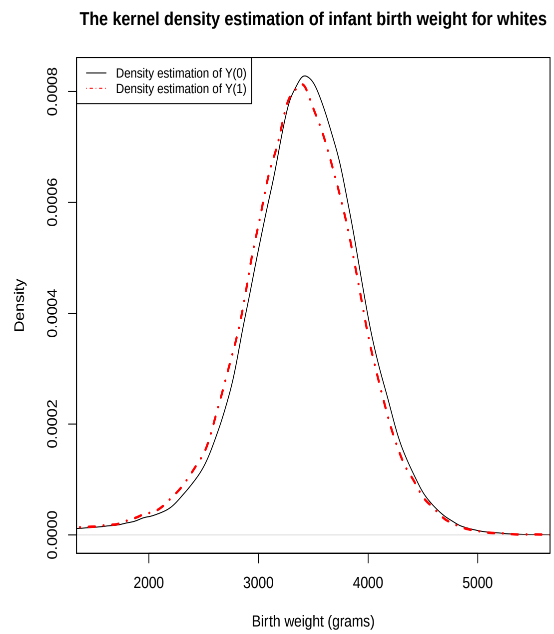

Figure 4(b): The kernel density estimation of birth weight for whites

Table 3: Descriptive statistics and some test results.

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

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 which is equal to 10 if prenatal care is foregone, 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, 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 $p(x)$. The explanatory variables used in the logit model consist 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. The PCQTE function is estimated for mothers aged between 20 and 30.

Figure 5 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 5 that the estimated conditional CDF curves for $Y(1)$ are all on the right of $Y(0)$, which implies that the partially conditional quantile treatment effects should be negative across all quantile levels.

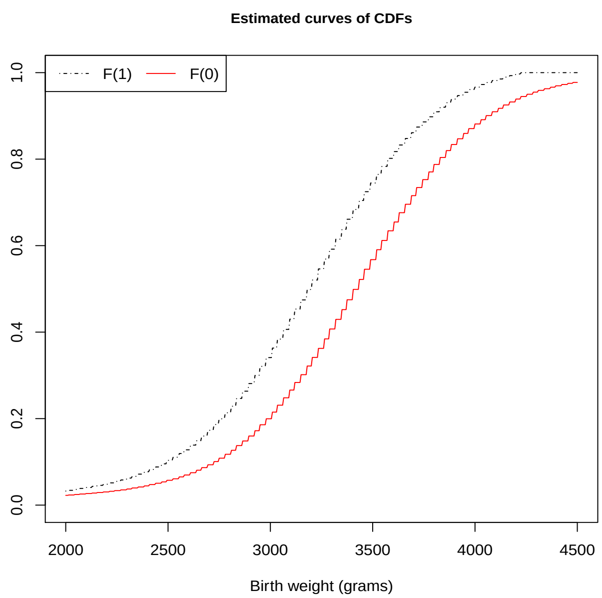

Figure 5: Estimated curves of conditional CDFs for infant birth weight conditional on mother's age ($z = 26$) for whites ($F(1)$ is for treated and $F(0)$ is for untreated)

Figure 6(a) plots the estimated PCQTE curves across mothers' ages for three quantile levels $\tau = 0.10$ (the solid line), 0.25 (the long dashed line) and $\tau = 0.50$ (the short dashed line) for whites. For a comparison, Figure 6(a) also depicts the estimated PCATE curve by the long dashed-dotted line, considered in Lee et al. (2017), across mothers' ages with its 95%-confidence interval denoted by the thin dotted lines, which is described in detail in Lee et al. (2017), and the estimated unconditional ATE as well (the short dashed-dotted line). From Figure 6(a), first, one can see that $\widehat{\Delta}_\tau(z)$ for three τ's and the estimated PCATE curve seem to change over age linearly and in particular, $\widehat{\Delta}_{0.5}(z)$ is similar to its PCATE curve although they are not exactly same. 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. Moreover, the estimated results displayed in Figure 6(a) show substantially heterogeneity across different ages. Overall speaking, 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 or higher quantiles, conditional on the same mother's age. Therefore, these results are in line with the findings displayed by Figure 5. Finally, for making an easy statistical inference, Figure 6(b) demonstrates the estimated PCQTE curve, $\widehat{\Delta}_\tau(z)$ and the unconditional QTE, $\widehat{\Delta}_\tau$ for $\tau = 0.10$ and $\tau = 0.50$, respectively, together with their 95% confidence intervals. Note that the 95% confidence interval for $\Delta_\tau(z)$ is pointwise and by ignoring the asymptotic bias for $\widehat{\Delta}_\tau(z)$. Clearly, one can see that $\widehat{\Delta}_\tau$ is not complete within the 95% pointwise confidence interval of $\Delta_\tau(z)$. Therefore, this means that $\Delta_\tau(z)$ is not a constant.

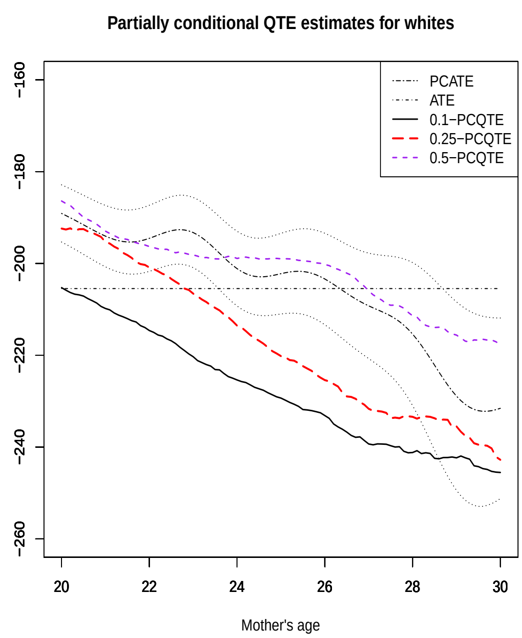

Figure 6(a): Estimation results for PCQTE for whites for three quantile levels, together PCATE and its 95% confidence interval, as well as the unconditional ATE.

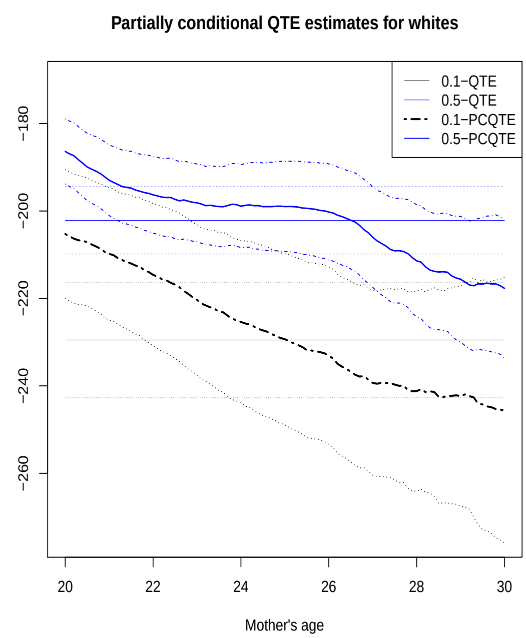

Figure 6(b): Estimation results of the PCQTE and the unconditional QTE for $\tau = 0.10$ and $\tau = 0.50$ for whites with their 95% confidence intervals.

Figure 7 displays the estimated PCQTEs for blacks under two quantile levels $\tau = 0.10$ (the dashed-dotted line) and 0.50 (the solid line), respectively. Their corresponding unconditional QTEs for $\tau = 0.1$ (the dashed line) and 0.5 (the thin solid line), respectively, are also reported, together with their 95% confidence intervals indicated by the dashed dotted lines for $\tau = 0.10$ and the dotted lines for $\tau = 0.50$, which are given by Firpo (2007). The estimated PCQTEs all lie in the 95%-confidence intervals of the corresponding unconditional QTE for all two quantile levels considered. In other words, $\widehat{\Delta}_\tau(z)$ does not depend on z for all quantiles. Therefore, motivated by the estimation results above, it is interesting to test whether or not the partially conditional quantile treatment effects, for whites and blacks, change over mothers' ages, and the testing results are presented in Table 4.

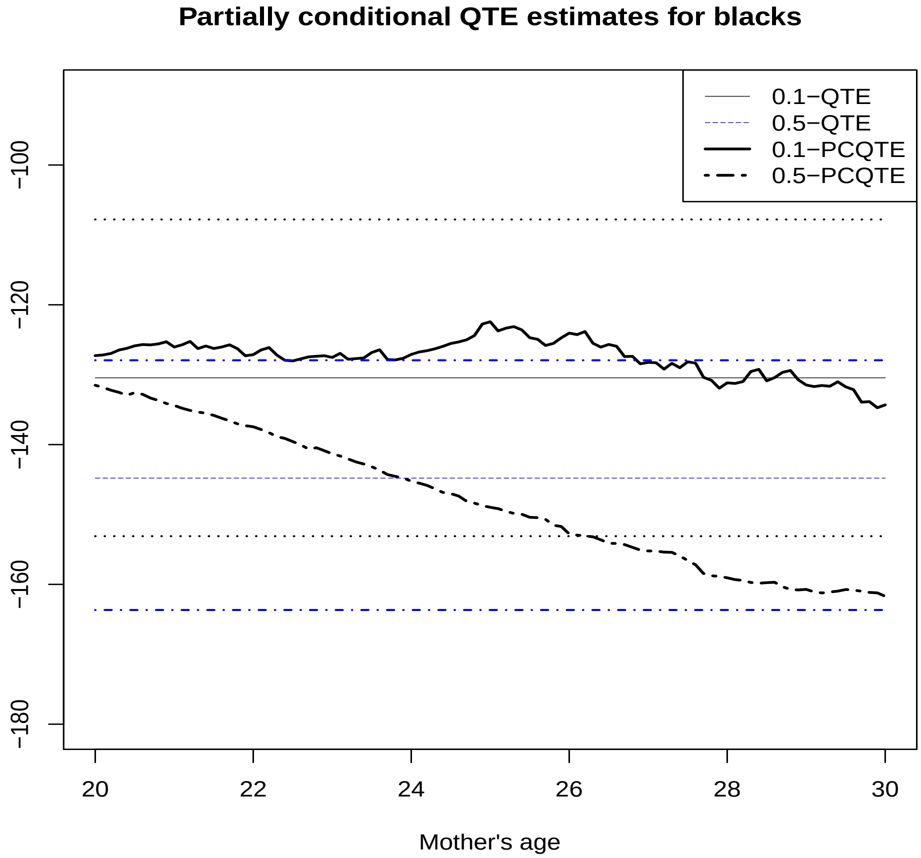

Figure 7: Estimation results for PCQTE for blacks for two quantile levels $\tau = 0.10$ (the dashed-dotted line) and $\tau = 0.50$ (the solid line), together with their unconditional QTEs and their 95% confidence intervals.

Table 4: Testing results for PCQTE over z

Quantile level	Test statistic J_n (Bootstrap p-value)	
τ	Blacks	Whites
0.10	0.633	0.028
0.25	0.218	0.002
0.50	0.201	0.003
0.75	0.479	0.005
0.90	0.785	0.035

It can be seen from Table 4 that there is a strong evidence to support the homogeneity of

PCQTE over z for blacks for all quantile levels considered, especially for $\tau = 0.1$ and $\tau = 0.9$ with p-value being 0.633 and 0.785, respectively. These testing results are in line with the findings presented by Figures 6(a) and 6(b). For whites, one can observe from Table 4 that one should reject the null hypothesis that there is a constant effect over z for all quantile levels considered at the significance level $\alpha = 5\%$.

6 Conclusion

In this paper, we consider the estimation of 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. Furthermore, we propose a new procedure to test the homogeneity for the partially conditional quantile treatment effects over the conditional variable and derive the asymptotic normality for the proposed test statistic. Using the proposed semiparametric estimator, we estimate the quantile effect of a first time mother's smoking on her baby's birth weight conditional on the mother's age. Moreover, using the proposed testing procedure, we test whether the partially conditional quantile treatment effect for both whites and blacks changes over mothers' ages. 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.

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(1):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.

Chen, X., Linton, O., and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5):1591–1608.

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. *The Review of Economics and Statistics*, 90(3):389–405.

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

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.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044.

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

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

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

Firpo, S., Galvao, A. F., and Song, S. (2017). Measurement errors in quantile regression models. *Journal of Econometrics*, 198(1):146–164.

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.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, 14(1):1–16.

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). Chapter 31 – the economics and econometrics of active labor market programs. *Handbook of Labor Economics*, 3, part a(1):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. *The 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. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 149–170.

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, UK: Cambridge University Press.

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

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

Kong, E., Linton, O., and Xia, Y. (2010). Uniform bahadur representation for local polynomial estimates of m-regression and its application to the additive model. *Econometric Theory*, 26(5):1529–1564.

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

Lehrer, S. F., Pohl, R. V., and Song, K. (2018). Multiple testing and the distributional effects of accountability incentives in education. *Available at SSRN 3266793*.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *The Annals of Statistics*, 15(2):780–799.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag.

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.

Appendix

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 $\widehat{W}_{n,0}(X_i, D_i) = \frac{1-D_i}{1-\widehat{p}(X_i)}$, $\widehat{W}_{n,1}(X_i, D_i) = \frac{D_i}{\widehat{p}(X_i)}$, where $\widehat{p}(x)$ 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 $j = 0, 1$, 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 $K_{h,i}(z) = \frac{1}{h}K\big((Z_i - z)/h\big)$ 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,

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

$$- \frac{f_Z(z) f_{Y(j)|Z}(q_{j,\tau}(z)|z)}{2} \cdot nh\big(q - q_{j,\tau}(z)\big)^2 \Bigg|$$

$$:= \quad \mathcal{A}_1 + \mathcal{A}_2.$$

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

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

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

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

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

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

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,

$$E\Big[\sum_{i=1}^n hK_{h,i}(z) W_j(X_i, D_i) \Psi\big(Y_i; q, q_{j,\tau}(z)\big) \Big]$$

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

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

$$
\begin{aligned}
&\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z)) \\
&= E\Big[\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big] \\
&\qquad\qquad + O_p\Big(\mathrm{Var}\Big[\sum_{i=1}^n hK_{h,i}(z)W_j(X_i,D_i)\Psi(Y_i;q,q_{j,\tau}(z))\Big]\Big)^{1/2} \\
&= nh\cdot\frac{(q_{j,\tau}(z)-q)^2}{2}\cdot\big[f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)+O(|q_{j,\tau}(z)-q|)+o(1)\big]+\ O_p\Big(nh\cdot|q_{j,\tau}(z)-q|^3\Big)^{1/2},
\end{aligned}
$$

and

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

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 $j=0,1$. 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 $\mu_2(K)=\int u^2K(u)du$ 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

$$\begin{aligned}\sqrt{nh}\Big(\widetilde{q}_{j,\tau}(z)-E\big(\widetilde{q}_{j,\tau}(z)\big)\Big) &= \sqrt{nh}\Big(\widetilde{q}_{j,\tau}(z)-q_{j,\tau}+\frac{h^2}{2}\mu_2(K)\delta_{j,\tau}(z)+o(h^2)\Big) \qquad (A.4)\\ &= -\frac{1}{\sqrt{nh}}\frac{1}{f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z)}\end{aligned}$$

$$\times \sum_{i=1}^{n} \Big[hK_{h,i}(z)\psi_j(Y_i, X_i, D_i; z) - E\big(hK_{h,i}(z)\psi_j(Y_i, X_i, D_i; z)\big)\Big].$$

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 $\widetilde{q}_{j,\tau}(z) = \arg\min_q \big\{\widetilde{\Gamma}_{n,j}(q, z)\big\}$. 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 $\Gamma_{n,j}\big(\widehat{q}_{j,\tau}(z), z\big) \leq \Gamma_{n,j}\big(\widetilde{q}_{j,\tau}(z), z\big)$ by definition of $\widehat{q}_{j,\tau}(z)$, one can show that

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

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

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

where $\widetilde{\Delta}_\tau(z) = \widetilde{q}_{1,\tau}(z) - \widetilde{q}_{0,\tau}(z)$. Note that $E\Big[hK_{h,i}(z)\psi_j(Y_i, X_i, D_i; z) - E\big(hK_{h,i}(z)\psi_j(Y_i, X_i, D_i; z)\big)\Big] = 0$, by applying the Lyapunov's central limit theorem, we can easily show that

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

This completes the proof. □

Now, it turns to the proof of Theorem 3.1. To this end, first, one needs to show the following lemmas.

Lemma 2. *Suppose that Assumptions 2.2-2.4 hold, then*

$$
\sup_{z\in\mathcal{Z}}\Big|S_{n,j}(z) - f_{Y(j)|Z}(q_{j,\tau}(z)|z) f_Z(z)\Big| = O\Big(h + (nh/\ln n)^{-1/2}\Big),
$$

where

$$
S_{n,j}(z) = \int K(u) f_{Y(j)|Z}\big(q_{j,\tau}(z+hu)|z+hu\big) f_Z(z+hu)du, \quad j = 0, 1.
$$

Proof of Lemma 2: The proof is given by Lemma 8 in Kong et al. (2010). □

Lemma 3. *For any $\alpha, \beta \in \mathbb{R}$ and $j = 0, 1$, define*

$$\begin{aligned} & \Omega_{n,i,j,\tau}(z; \alpha, \beta) \qquad \text{(A.5)} \\ = \; & W_j(X_i, D_i) h K_{h,i}(z) \Big\{ \rho_\tau(Y_i; \alpha + \beta + q_{j,\tau}(z)) - \rho_\tau(Y_i; \beta + q_{j,\tau}(z)) - \alpha \varphi_\tau(Y_i; q_{j,\tau}(z)) \Big\} \\ = \; & W_j(X_i, D_i) h K_{h,i}(z) \Big\{ \rho_\tau(Y_i(j); \alpha + \beta + q_{j,\tau}(z)) - \rho_\tau(Y_i(j); \beta + q_{j,\tau}(z)) - \alpha \varphi_\tau(Y_i(j); q_{j,\tau}(z)) \Big\} \\ = \; & W_j(X_i, D_i) h K_{h,i}(z) \int_\beta^{\alpha+\beta} \Big\{ \varphi_\tau(Y_i(j); q_{j,\tau}(z) + t) - \varphi_\tau(Y_i(j); q_{j,\tau}(z)) \Big\} dt \\ = \; & W_j(X_i, D_i) h K_{h,i}(z) \int_\beta^{\alpha+\beta} \Big(I\big\{ Y_i(j) < q_{j,\tau}(z) + t \big\} - I\big\{ Y_i(j) < q_{j,\tau}(z) \big\} \Big) dt, \end{aligned}$$

and

$$G_{n,i,j,\tau}(z; \alpha, \beta) = \Omega_{n,i,j,\tau}(z; \alpha, \beta) - E\big[\Omega_{n,i,j,\tau}(z; \alpha, \beta)\big], \qquad \text{(A.6)}$$

where $\varphi_\tau(y; q) = I(y \le q) - \tau$. Then, under Assumptions 2.1-2.4, it is clear that for all large $M > 0$,

$$\sup_{z \in \mathcal{Z}} \sup_{\substack{|\alpha| \le M d_n^{(1)} \\ |\beta| \le M^{1/4} d_n^{(2)}}} \Big| \sum_{i=1}^n G_{n,i,j,\tau}(z; \alpha, \beta) \Big| \le M^{3/2} d_n \quad \textit{almost surely},$$

where $d_n^{(1)} = (nh)^{-3/4} (\ln n)^{3/4}$, $d_n^{(2)} = (nh)^{-1/2} (\ln n)^{1/2}$ and $d_n = (nh)^{-1/2} (\ln n)^{3/2}$.

Proof of Lemma 3: This result can be proved using similar arguments as in the proof of Lemma 1 in Kong et al. (2010). □

Lemma 4. *Suppose that Assumptions 2.1-2.4 hold. Then there is a constant $C > 0$ such that, for each $M > 0$ and all large n,*

$$\sup_{z \in \mathcal{Z}} \sup_{\substack{|\alpha| \le M \, d_n^{(1)} \\ |\beta| \le M^{1/4} d_n^{(2)}}} \Big| \sum_{i=1}^n E\big[\Omega_{n,i,j,\tau}(z; \alpha, \beta)\big] - \frac{nh}{2} \alpha(\alpha + 2\beta) S_{n,j}(z) \Big| \le C M^{3/2} d_{n1},$$

where $d_{n1} = (nh)^{-3/4} (\ln n)^{7/4}$.

Proof of Lemma 4: The result can be proved following the proof of Lemma 9 in Kong et al. (2010). □

Lemma 5. *Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent random variables such that $E(\xi_i) = 0$, $|\xi_i| \le \varsigma$*

for all i and $\sum_{i=1}^{n} E\xi_i^2 \le A$. *Then, for any* $\eta > 0$,

$$P\Big(\Big|\sum_{i=1}^{n} \xi_i\Big| \ge \eta\Big) \le 2\exp\Big\{ -\lambda\,\eta + \lambda^2 A\Big\}$$

for all $\lambda \le 1/2\varsigma$.

Proof of Lemma 5: Since $|\xi_i| \le \varsigma$ for all $i = 1, \cdots, n$, we have $\lambda|\xi_i| \le 1/2$. Using the inequality $\exp(x) \le 1 + x + x^2$ for all $|x| \le 1/2$, we can obtain

$$\exp\Big(\pm\lambda_n\xi_i\Big) \le 1 \pm \lambda_n\xi_i + \lambda_n^2\xi_i^2.$$

Hence, we have

$$E\Big[\exp\big(\pm\lambda\xi_i\big)\Big] \le 1 \pm \lambda E(\xi_i) + \lambda^2 E\big(\xi_i^2\big) \le \exp\Big[E\big(\lambda^2\xi_i^2\big)\Big],$$

where we used $E\xi_i = 0$ and $1 + u \le \exp(u)$ for the second inequality. Therefore, by the Markov inequality and independence of $\xi_1, \xi_2, \cdots, \xi_n$, we have

$$\begin{aligned}
P\Big(\Big|\sum_{i=1}^{n}\xi_i\Big| \ge \eta\Big) &\le P\Big(\lambda\sum_{i=1}^{n}\xi_i \ge \lambda\eta\Big) + P\Big(-\lambda\sum_{i=1}^{n}\xi_i \ge \lambda\eta\Big) \\
&\le \Big\{E\Big[\exp(\lambda\sum_{i=1}^{n}\xi_i)\Big] + E\Big[\exp(-\lambda\sum_{i=1}^{n}\xi_i)\Big]\Big\}\Big/\exp(\lambda\eta) \\
&= \exp(-\lambda\eta)\Big\{\prod_{i=1}^{n} E\Big[\exp(\lambda\xi_i)\Big] + \prod_{i=1}^{n} E\Big[\exp(-\lambda\xi_i)\Big]\Big\} \\
&\le \exp(-\lambda\eta)\Big\{\prod_{i=1}^{n}\exp\Big[E\big(\lambda^2\xi_i^2\big)\Big] + \prod_{i=1}^{n}\exp\Big[E\big(\lambda^2\xi_i^2\big)\Big]\Big\} \\
&= 2\exp(-\lambda\eta)\Big[\exp\Big(\lambda^2\sum_{i=1}^{n}E\xi_i^2\Big)\Big] \\
&\le 2\exp(-\lambda\eta)\Big[\exp(\lambda^2 A)\Big] = 2\exp\Big\{-\lambda\eta + \lambda^2 A\Big\}.
\end{aligned}$$

The proof is completed. □

Lemma 6. *Suppose that Assumptions 2.1-2.4 are satisfied, we then have*

$$\sup_{z\in\mathcal{Z}}\Big|\frac{1}{n}S_{n,j}^{-1}(z)\sum_{i=1}^{n}W_j(X_i, D_i)K_h\big(Z_i - z\big)\varphi_\tau(Y_i; q_{j,\tau}(Z_i))\Big| = O\Big\{\Big(\frac{\ln n}{nh}\Big)^{1/2}\Big\}$$

almost surely for $j = 0, 1$.

Proof of Lemma 6: Note that $S_{n,j}(z)$ is bounded away from zero by Assumption 2.2-2.3 and Lemma 2, we only need to show

$$\sup_{z\in\mathcal{Z}}\Big|V_{n,j}(z)\Big| = O\Big\{\Big(\frac{\ln n}{nh}\Big)^{1/2}\Big\} \quad \text{almost surely,}$$

where $V_{n,j}(z) = \sum_{i=1}^{n} \xi_{n,i,j}(z)$ and

$$\xi_{n,i,j}(z) = n^{-1}W_j(X_i, D_i)K_h\big(Z_i - z\big)\varphi_\tau(Y_i; q_{j,\tau}(Z_i)).$$

Since the support $\mathcal{Z}$ of Z_i is compact, it can be covered by a finite number $\mathcal{L}_n = \big(\frac{n}{h^3 \ln n}\big)^{1/2}$ of intervals $\mathcal{Z}_{n,k}$ with length $\ell_n = O(\mathcal{L}_n^{-1}) = O\left(\frac{h^3 \ln n}{n}\right)^{1/2}$ and centers $z_{n,k}$, $k = 1, \cdots, L_n$. Then

$$\begin{aligned}
\sup_{z\in\mathcal{Z}}|V_{n,j}(z)| &= \max_{1\le k\le\mathcal{L}_n}\sup_{z\in\mathcal{Z}\cap\mathcal{Z}_k}|V_{n,j}(z)| \\
&\le \max_{1\le k\le\mathcal{L}_n}\sup_{z\in\mathcal{Z}\cap\mathcal{Z}_k}|V_{n,j}(z) - V_{n,j}(z_{n,k})| + \max_{1\le k\le\mathcal{L}_n}|V_{n,j}(z_{n,k})| \\
&:= Q_1 + Q_2. \qquad\qquad (A.7)
\end{aligned}$$

We first consider Q_1. It is easy to see by Assumption 2.4 that

$$\begin{aligned}
|V_{n,j}(z) - V_{n,j}(z_k)| &= \left|\frac{1}{n}\sum_{i=1}^{n} W_j(X_i, D_i)\varphi_\tau(Y_i; q_{j,\tau}(Z_i))\Big[K_h\big(Z_i - z\big) - K_h\big(Z_i - z_k\big)\Big]\right| \\
&\le \frac{1}{n}\sum_{i=1}^{n} W_j(X_i, D_i)\big|\varphi_\tau(Y_i; q_{j,\tau}(Z_i))\big|\Big|K_h\big(Z_i - z\big) - K_h\big(Z_i - z_k\big)\Big| \\
&\le \frac{C}{h^2}|z - z_k|,
\end{aligned}$$

so that

$$Q_1 \le \frac{C\ell_n}{h^2} = O\Big\{\Big(\frac{\ln n}{nh}\Big)^{1/2}\Big\} \quad \text{almost surely.} \qquad (A.8)$$

Now, we consider Q_2. For any $\eta > 0$,

$$P(Q_2 \ge \eta) = P\Big(\max_{1\le k\le\mathcal{L}_n}\Big|V_{n,j}(z_{n,k})\Big| \ge \eta\Big) \le \sum_{k=1}^{\mathcal{L}_n} P\Big(\Big|V_{n,j}(z_{n,k})\Big| \ge \eta\Big). \qquad (A.9)$$

It is easy to show that

$$E[\xi_{n,i,j}(z)] = E\big[n^{-1}W_j(X_i, D_i)K_h\big(Z_i - z\big)\varphi_\tau(Y_i; q_{j,\tau}(Z_i))\big]$$

$$
\begin{aligned}
&= E\big[n^{-1}K_h\big(Z_i - z\big)\varphi_\tau(Y_i(j); q_{j,\tau}(Z_i))\big] \\
&= E\Big\{n^{-1}K_h\big(Z_i - z\big)\,E\big[\varphi_\tau(Y_i(j); q_{j,\tau}(Z_i))\big|Z_i\big]\Big\} = 0,
\end{aligned}
$$

$|\xi_{n,i,j}(z)| \leq \frac{C_1}{nh}$ and $\sum_{i=1}^n E[\xi_{n,i,j}(z)]^2 \leq \frac{C_2}{nh}$, where C_1 and C_2 do not depend on z. Note that $(nh\ln n)^{1/2} < \frac{1}{2|\xi_{n,i,j}(z)|}$ for large n, we can apply Lemma 5 to $V_{n,j}(z) = \sum_{i=1}^{n}\xi_{n,i,j}(z)$ with $\eta_n = C_3\left(\frac{\ln n}{nh}\right)^{1/2}$ and $\lambda_n = (nh\ln n)^{1/2}$. Then, we obtain

$$
\begin{aligned}
P\Big(\big|V_{n,j}(z)\big| \geq \eta_n\Big) &\leq 2\exp\left\{-\lambda_n\eta_n + \lambda_n^2\frac{C_2}{nh}\right\} \\
&= 2\exp\{-C_3\ln n + C_2\ln n\} = n^{C_2 - C_3} \quad \text{for all } z.
\end{aligned}
$$

It is clear from (A.9) to see that

$$
\begin{aligned}
P(Q_2 \geq \eta_n) &\leq \sum_{k=1}^{\mathcal{L}_n} P\Big(\big|V_{n,j}(z_{n,k})\big| \geq \eta\Big) \\
&\leq \mathcal{L}_n n^{C_2 - C_3} = \left(\frac{n}{h^3\ln n}\right)^{1/2} n^{C_2 - C_3}.
\end{aligned}
$$

Choosing a large C_3, we can ensure that $\sum_{n=1}^{\infty} P(Q_2 \geq \eta_n)$ is finite. An application of the Borel-Cantelli Lemma implies that

$$
Q_2 = O(\eta_n) = O\left\{\left(\frac{\ln n}{nh}\right)^{1/2}\right\} \quad \text{almost surely.} \tag{A.10}
$$

Combining (A.7), (A.8) and (A.10), we obtain

$$
\sup_{z\in\mathcal{Z}}\Big|V_{n,j}(z)\Big| = O\left\{\left(\frac{\ln n}{nh}\right)^{1/2}\right\} \quad \text{almost surely.}
$$

This completes the proof. □

Lemma 7. *Under Assumptions 2.1-2.4,*

$$
\beta_{n,j}(z) - E\beta_{n,j}(z) - \frac{1}{n}S_{n,j}^{-1}(z)\sum_{i=1}^{n} W_j(X_i, D_i)K_h\big(Z_i - z\big)\varphi_\tau(Y_i; q_{j,\tau}(Z_i)) = O\left\{\left(\frac{\ln n}{nh}\right)^{1/2}\right\}
$$

uniformly in $z \in \mathcal{Z}$ with probability 1, where

$$
\begin{aligned}
\beta_{n,j}(z) &= \frac{1}{n}\sum_{i=1}^{n}\varrho_{n,j}(Y_i, X_i, D_i; z) \qquad\qquad \text{(A.11)} \\
&= -\frac{1}{n}\sum_{i=1}^{n} S_{n,j}^{-1}(z)W_j(X_i, D_i)K_h\big(Z_i - z\big)\varphi_\tau(Y_i; q_{j,\tau}(z)).
\end{aligned}
$$

Proof of Lemma 7: The result can be proved by following the proof of Corollary 1 in Kong et al. (2010). □

Lemma 8. *Suppose Assumptions 2.1-2.4 hold. Then,*

$$\sup_{z\in\mathcal{Z}}\Big|\bar{q}_{j,\tau}(z)-q_{j,\tau}(z)-\frac{1}{n}\sum_{i=1}^{n}\varrho_{n,j}(Y_i,X_i,D_i;z)\Big|=O\Big[\Big(\frac{\ln n}{nh}\Big)^{3/4}\Big] \quad \textit{almost surely} \tag{A.12}$$

for $j=0,1,$ *where*

$$\bar{q}_{j,\tau}(z)=\arg\min_{q}\sum_{i=1}^{n}W_j(X_i,D_i)hK_{h,i}(z)\rho_\tau(Y_i;q)$$

and

$$\varrho_{n,j}(Y_i,X_i,D_i;z)=-S_{n,j}^{-1}(z)W_j(X_i,D_i)K_h\big(Z_i-z\big)\varphi_\tau(Y_i;q_{j,\tau}(z)).$$

Proof of Lemma 8: Based on Lemmas 3 and 4, one can see that with probability 1, there exists some $C_1>1$, such that for all large $M>0$,

$$\begin{aligned}&\sup_{z\in\mathcal{Z}}\sup_{\substack{|\alpha|\le Md_n^{(1)}\\|\beta|\le M^{1/4}d_n^{(2)}}}\Big|\sum_{i=1}^{n}\Omega_{n,i,j,\tau}(z;\alpha,\beta)-\frac{nh}{2}S_{n,j}(z)\alpha(\alpha+2\beta)\Big|\\ \le\ & C_1M^{3/2}(d_{n1}+d_n)\le 2C_1M^{3/2}d_n\end{aligned}$$

for large n, where $d_n^{(1)}=(nh)^{-3/4}(\ln n)^{3/4}$, $d_n^{(2)}=(nh)^{-1/2}(\ln n)^{1/2}$, $d_n=(nh)^{-1/2}(\ln n)^{3/2}$ and $d_{n1}=(nh)^{-3/4}(\ln n)^{7/4}$. Hence,

$$\inf_{z\in\mathcal{Z}}\inf_{\substack{|\alpha|\le M\,d_n^{(1)}\\|\beta|\le M^{1/4}d_n^{(2)}}}\Big\{\sum_{i=1}^{n}\Omega_{n,i,j,\tau}(z;\alpha,\beta)-\frac{nh}{2}S_{n,j}(z)\alpha(\alpha+2\beta)\Big\}\ge -2C_1M^{3/2}(nh)^{-1/2}(\ln n)^{3/2}.$$

Note that based on the definition in (A.11),

$$\alpha\sum_{i=1}^{n}W_j(X_i,D_i)hK_h\big(Z_i-z\big)\varphi_\tau(Y_i;q_{j,\tau}(z))=-nhS_{n,j}(z)\beta_{n,j}(z)\alpha.$$

Then, by the definition of $\Omega_{n,i,j,\tau}(z;\alpha,\beta)$ in (A.5), it is easy to show that

$$\begin{aligned}
&\inf_{z\in\mathcal{Z}} \inf_{\substack{|\alpha|\le M\, d_n^{(1)}\\ |\beta|\le M^{1/4} d_n^{(2)}}} \Big\{\sum_{i=1}^n \Omega_{n,i,j,\tau}(z;\alpha,\beta) - \frac{nh}{2} S_{n,j}(z)\alpha(\alpha+2\beta)\Big\}\\
= &\inf_{z\in\mathcal{Z}} \inf_{\substack{|\alpha|\le M\, d_n^{(1)}\\ |\beta|\le M^{1/4} d_n^{(2)}}} \Big\{\sum_{i=1}^n W_j(X_i,D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i;\alpha+\beta+q_{j,\tau}(z)) - \rho_\tau(Y_i;\beta+q_{j,\tau}(z))\big]\\
&-\alpha\sum_{i=1}^n W_j(X_i,D_i)hK_{h,i}(z)\varphi_\tau(Y_i;q_{j,\tau}(z)) - \frac{nh}{2}S_{n,j}(z)\alpha(\alpha+2\beta)\Big\}\\
= &\inf_{z\in\mathcal{Z}} \inf_{\substack{|\alpha|\le M\, d_n^{(1)}\\ |\beta|\le M^{1/4} d_n^{(2)}}} \Big\{\sum_{i=1}^n W_j(X_i,D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i;\alpha+\beta+q_{j,\tau}(z)) - \rho_\tau(Y_i;\beta+q_{j,\tau}(z))\big]\\
&+nh\cdot S_{n,j}(z)\alpha\beta_{n,j}(z) - nh\cdot S_{n,j}(z)\alpha\beta - \frac{nh}{2}S_{n,j}(z)\alpha^2\Big\}\\
\ge &-2C_1M^{3/2}(nh)^{-1/2}(\ln n)^{3/2}.
\end{aligned}$$

Moving the term $-\frac{nh}{2}S_{n,j}(z)\alpha^2$ to the right-hand side, we obtain

$$\begin{aligned}
&\inf_{z\in\mathcal{Z}} \inf_{\substack{|\alpha|\le M\, d_n^{(1)}\\ |\beta|\le M^{1/4} d_n^{(2)}}} \Big\{\sum_{i=1}^n W_j(X_i,D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i;\alpha+\beta+q_{j,\tau}(z)) - \rho_\tau(Y_i;\beta+q_{j,\tau}(z))\big]\\
&\qquad\qquad +nh\cdot S_{n,j}\alpha\big(\beta_{n,j}(z)-\beta\big)\Big\}\\
\ge &\sup_{z\in\mathcal{Z}}\sup_{|\alpha|\le M d_n^{(1)}} \Big\{\frac{nh}{2}S_{n,j}(z)\alpha^2\Big\} - 2C_1M^{3/2}(nh)^{-1/2}(\ln n)^{3/2}
\end{aligned}$$

for all large enough M. Let $B_{n,k} = \{\alpha : kNd_n^{(1)} < |\alpha| \le (k+1)Nd_n^{(1)}\}$, $k = 1, 2, \cdots$, and $M = (K+1)N$. Then,

$$\begin{aligned}
&\inf_{z\in\mathcal{Z}} \inf_{\substack{\alpha\in B_{n,k},\\ |\beta|\le [(k+1)N]^{1/4} d_n^{(2)}}} \Big\{\sum_{i=1}^n W_j(X_i,D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i;\alpha+\beta+q_{j,\tau}(z)) - \rho_\tau(Y_i;\beta+q_{j,\tau}(z))\big]\\
&\qquad\qquad +nh\cdot S_{n,j}(z)\alpha\big(\beta_{n,j}(z)-\beta\big)\Big\}\\
\ge &\sup_{z\in\mathcal{Z}}\sup_{\alpha\in B_{n,k}} \Big\{\frac{nh}{2}S_{n,j}(z)\alpha^2\Big\} - 2C_1[(k+1)N]^{3/2}(nh)^{-1/2}(\ln n)^{3/2}\\
\ge &\sup_{z\in\mathcal{Z}} \Big\{\frac{nh}{2}S_{n,j}(z)k^2N^2(nh)^{-3/2}(\ln n)^{3/2}\Big\} - 2C_1[(k+1)N]^{3/2}(nh)^{-1/2}(\ln n)^{3/2}\\
= &\sup_{z\in\mathcal{Z}} \Big\{\big[N^{1/2} - 4C_1S_{n,j}^{-1}(z)k^{-2}(k+1)^{3/2}\big]N^{3/2}(nh)^{-1/2}(\ln n)^{3/2}\Big\}.
\end{aligned}$$

Since $C_1 S_{n,j}^{-1}(z)k^{-2}(k+1)^{3/2}$ is uniformly bounded for all $z \in \mathcal{Z}$ and $k = 1, 2, \cdots$, we can find a large enough N such that

$$\inf_{z\in\mathcal{Z}} \inf_{\substack{\alpha\in B_{n,k}, \\ |\beta|\le[(k+1)N]^{1/4}d_n^{(2)}}} \Big\{ \sum_{i=1}^{n} W_j(X_i, D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i; \alpha+\beta+q_{j,\tau}(z)) - \rho_\tau(Y_i;\beta+q_{j,\tau}(z))\big] + nh\cdot S_{n,j}(z)\alpha\big(\beta_{n,j}(z)-\beta\big)\Big\} > 0$$

for all $k = 1, 2, \cdots$ almost surely. Furthermore, we can obtain that

$$\begin{aligned} E\beta_{n,j}(z) &= E\big[\varrho_{n,j}(Y_i, X_i, D_i; z)\big] \\ &= -\frac{1}{S_{n,j}(z)} f_Z'(z) \frac{\partial F_{Y(j)|Z}(q_{j,\tau}(z)|u)}{\partial u}\Big|_{u=z} h^2 \int s^2 K(s)ds + o(h^2) = O(h^2) \end{aligned} \tag{A.13}$$

uniformly in z. Since nh^5 is bounded, it is clear to see that $E\beta_{n,j}(z) = O(h^2) = O\left\{\left(\frac{\ln n}{nh}\right)^{1/2}\right\}$, which together with Lemmas 6 and implies that $|\beta_{n,j}(z)| = O\left(d_n^{(2)}\right)$ uniformly in $z \in \mathcal{Z}$ almost surely. Hence, $|\beta_{n,j}(z)| \le [(k+1)N]^{1/4} d_n^{(2)}$ for large N and for all $k = 1, 2, \cdots$ with probability 1. Then,

$$\begin{aligned} 0 &< \inf_{z\in\mathcal{Z}} \inf_{\substack{\alpha\in B_{n,k}, \\ |\beta|\le[(k+1)N]^{1/4}d_n^{(2)}}} \Big\{ \sum_{i=1}^{n} W_j(X_i, D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i; \alpha+\beta+q_{j,\tau}(z)) - \rho_\tau(Y_i;\beta+q_{j,\tau}(z))\big] \\ &\qquad + nh\cdot S_{n,j}(z)\alpha\big(\beta_{n,j}(z)-\beta\big)\Big\} \\ &\le \inf_{z\in\mathcal{Z}} \inf_{\substack{\alpha\in B_{n,k}, \\ \beta=\beta_{n,j}(z)}} \Big\{ \sum_{i=1}^{n} W_j(X_i, D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i; \alpha+\beta+q_{j,\tau}(z)) - \rho_\tau(Y_i;\beta+q_{j,\tau}(z))\big] \\ &\qquad + nh\cdot S_{n,j}(z)\alpha\big(\beta_{n,j}(z)-\beta\big)\Big\} \qquad \text{(A.14)} \\ &= \inf_{z\in\mathcal{Z}} \inf_{\alpha\in B_{n,k}} \Big\{ \sum_{i=1}^{n} W_j(X_i, D_i)hK_{h,i}(z)\big[\rho_\tau(Y_i; \alpha+\beta_{n,j}(z)+q_{j,\tau}(z)) \\ &\qquad - \rho_\tau(Y_i;\beta_{n,j}(z)+q_{j,\tau}(z))\big]\Big\}. \end{aligned}$$

Note that

$$\bar{q}_{j,\tau}(z) = \arg\min_q \sum_{i=1}^{n} W_j(X_i, D_i)hK_{h,i}(z)\rho_\tau(Y_i; q),$$

Inequality (A.14) implies clearly that $\bar{q}_{j,\tau}(z) \ne \alpha + \beta_{n,j}(z) + q_{j,\tau}(z)$ for any $\alpha \in B_{n,k} = \{\alpha : kNd_n^{(1)} < |\alpha| \le (k+1)Nd_n^{(1)}\}$ for any $k = 1, 2, \cdots$. This concludes that

$$|\bar{q}_{j,\tau}(z) - q_{j,\tau}(z) - \beta_{n,j}(z)| < Nd_n^{(1)} = O\Big[\Big(\frac{\ln n}{nh}\Big)^{3/4}\Big]$$

for some $N > 0$ almost surely, where N does not depend on z. This completes the proof. $\square$

Lemma 9. *Suppose that Assumptions 2.1-2.5 are satisfied, then,*

$$\sup_{z\in\mathcal{Z}}\left|\widehat{q}_{j,\tau}(z)-\bar{q}_{j,\tau}(z)\right| = O_p\left\{\max\left(\frac{\ln n}{\sqrt{n}},\left(\frac{\ln n}{nh}\right)^{3/4}\right)\right\},\quad j=0,1.$$

Proof of Lemma 9: For $j=0,1$, define cumulative distribution functions

$$\bar{F}_{n,j}(y\,|\,z)=\frac{\sum_{i=1}^{n}K_{h,i}(z)W_j(X_i,D_i)I\{Y_i\le y\}}{\sum_{i=1}^{n}K_{h,i}(z)W_j(X_i,D_i)}$$

and

$$\widehat{F}_{n,j}(y\,|\,z)=\frac{\sum_{i=1}^{n}K_{h,i}(z)\widehat{W}_{n,j}(X_i,D_i)I\{Y_i\le y\}}{\sum_{i=1}^{n}K_{h,i}(z)\widehat{W}_{n,j}(X_i,D_i)}.$$

By the definitions of $\bar{q}_{j,\tau}(z)$ and $\widehat{q}_{j,\tau}(z)$, and the properties of the check function $\rho(y;q)$, it follows that $\bar{q}_{j,\tau^*}(z)=\inf\{y:\bar{F}_{n,j}(y\,|\,z)\ge\tau^*\}$ and $\widehat{q}_{j,\tau^*}(z)=\inf\{y:\widehat{F}_{n,j}(y\,|\,z)\ge\tau^*\}$ for $0<\tau^*<1$. We also have $\bar{F}_{n,j}\big(\bar{q}_{j,\tau^*}(z)\,|\,z\big)=\tau^*+O_p(1/nh)$. For $j=0$ and 1, define

$$\bar{f}_{n,j}(z)=n^{-1}\sum_{i=1}^{n}K_{h,i}(z)W_j(X_i,D_i)$$

and

$$\widehat{f}_{n,j}(z)=n^{-1}\sum_{i=1}^{n}K_{h,i}(z)\widehat{W}_{n,j}(X_i,D_i).$$

Then, by using $\sup_{x\in\mathcal{X}}\left|\widehat{W}_{n,j}(x,D_i)-W_j(x,D_i)\right|=O_p(n^{-1/2})$ and $\sup_{z\in\mathcal{Z}}\frac{1}{n}\sum\limits_{i=1}^{n}K_{h,i}(z)=O_p(1)$, it is easy to show that

$$\begin{aligned}\sup_{z\in\mathcal{Z}}\left|\bar{f}_{n,j}(z)-\widehat{f}_{n,j}(z)\right| &\le \sup_{z\in\mathcal{Z}}\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)\left|W_j(X_i,D_i)-\widehat{W}_{n,j}(X_i,D_i)\right|\\ &\le O_p(n^{-1/2})\cdot\sup_{z\in\mathcal{Z}}\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)=O_p(n^{-1/2}).\end{aligned}$$

In addition, by the proof of Lemma 4.4 in Donald and Hsu (2014), we know that

$$\sup_{z\in\mathcal{Z}}\left|\widehat{f}_{n,j}(z)-f_Z(z)\right|=o_p(1),$$

which implies that $1\Big/\big(\inf_{z\in\mathcal{Z}}\widehat{f}_{n,j}(z)\big)=O_p(1)$ by Assumption 2.2. Therefore, for $j=0$ and

1,

$$
\begin{aligned}
& \sup_{y\in\mathcal{Y}_j}\sup_{z\in\mathcal{Z}}\left|\widehat{F}_{n,j}(y\,|\,z)-\bar{F}_{n,j}(y\,|\,z)\right| \\
\leq\ & \frac{1}{\inf_{z\in\mathcal{Z}}\bar{f}_{n,j}(z)}\sup_{y\in\mathcal{Y}_j}\sup_{z\in\mathcal{Z}}\left|\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)I\{Y_i\leq y\}\big(W_j(X_i,D_i)-\widehat{W}_{n,j}(X_i,D_i)\big)\right| \\
& +\frac{1}{\inf_{z\in\mathcal{Z}}\bar{f}_{n,j}(z)}\frac{1}{\inf_{z\in\mathcal{Z}}\widehat{f}_{n,j}(z)}\sup_{y\in\mathcal{Y}_j}\sup_{z\in\mathcal{Z}}\left|\bar{f}_{n,j}(z)-\widehat{f}_{n,j}(z)\right|\left|\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)\widehat{W}_{n,j}(X_i,D_i)I\{Y_i\leq y\}\right| \\
\leq\ & O_p(1)\cdot O_p(n^{-1/2})\cdot\sup_{z\in\mathcal{Z}}\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)+O_p(1)\cdot O_p(n^{-1/2})\cdot\sup_{z\in\mathcal{Z}}\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)\widehat{W}_{n,j}(X_i,D_i) \\
=\ & O_p(n^{-1/2}),
\end{aligned}
$$

where $\mathcal{Y}_j$ is the support of $Y(j)$. Now, let $c_n=\max\left\{\frac{\ln n}{\sqrt{n}},\frac{(\ln n)^2}{nh}\right\}$. Then,

$$
\begin{aligned}
\widehat{F}_{n,j}\big(\bar{q}_{j,\tau+c_n}(z)\,|\,z\big) &= \bar{F}_{n,j}\big(\bar{q}_{j,\tau+c_n}(z)\,|\,z\big)+O_p(1/\sqrt{n}) \\
&= \tau+c_n+O_p(1/nh)+O_p(1/\sqrt{n})>\tau
\end{aligned}
$$

in probability as $n\to\infty$. Similarly,

$$
\begin{aligned}
\widehat{F}_{n,j}\big(\bar{q}_{j,\tau-c_n}(z)\,|\,z\big) &= \bar{F}_{n,j}\big(\bar{q}_{j,\tau-c_n}(z)\,|\,z\big)+O_p(1/\sqrt{n}) \\
&= \tau-c_n+O_p(1/nh)+O_p(1/\sqrt{n})<\tau
\end{aligned}
$$

in probability as $n\to\infty$. Since $\widehat{q}_{j,\tau}(z)=\inf\{y:\widehat{F}_{n,j}(y\,|\,z)\geq\tau\}$, then, we have

$$
\bar{q}_{j,\tau-c_n}(z)\leq\widehat{q}_{j,\tau}(z)\leq\bar{q}_{j,\tau+c_n}(z).
$$

Obviously, one also can see that

$$
\bar{q}_{j,\tau-c_n}(z)\leq\bar{q}_{j,\tau}(z)\leq\bar{q}_{j,\tau+c_n}(z).
$$

Therefore, by using the Bahadur representation of $\bar{q}_{j,\tau^*}(z)$ provided by Lemma 8, it follows that

$$
\begin{aligned}
\sup_{z\in\mathcal{Z}}\left|\widehat{q}_{j,\tau}(z)-\bar{q}_{j,\tau}(z)\right| &\leq \sup_{z\in\mathcal{Z}}\left|\bar{q}_{j,\tau+c_n}(z)-\bar{q}_{j,\tau-c_n}(z)\right| \\
&\leq \sup_{z\in\mathcal{Z}}\left|q_{j,\tau+c_n}(z)-q_{j,\tau-c_n}(z)\right|+\sup_{z\in\mathcal{Z}}\left|S_{n,j}^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)W_j(X_i,D_i)\right. \\
&\quad \left.\times\big(I\{Y_i\leq q_{j,\tau-c_n}(z)\}-I\{Y_i\leq q_{j,\tau+c_n}(z)\}+2c_n\big)\right|+O_p\left\{\left(\frac{\ln n}{nh}\right)^{3/4}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{z\in\mathcal{Z}}\big|q_{j,\tau+c_n}(z)-q_{j,\tau-c_n}(z)\big| + \sup_{z\in\mathcal{Z}}\Big|S_{n,j}^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)W_j(X_i,D_i)\\
&\quad \times\big(I\{Y_i\leq q_{j,\tau-c_n}(z)\}-I\{Y_i\leq q_{j,\tau+c_n}(z)\}\big)\Big|\\
&\quad +2c_n\cdot\sup_{z\in\mathcal{Z}}\Big|S_{n,j}^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)W_j(X_i,D_i)\Big| + O_p\Big\{\Big(\frac{\ln n}{nh}\Big)^{3/4}\Big\}\\
&:= \mathcal{M}_1+\mathcal{M}_2+\mathcal{M}_3+O_p\Big\{\Big(\frac{\ln n}{nh}\Big)^{3/4}\Big\}.
\end{aligned}
$$

For $\mathcal{M}_1$, first note that $F_{Y(j)|Z}(q_{j,\tau+c_n}(z)|z)=\tau+c_n$ and $F_{Y(j)|Z}(q_{j,\tau-c_n}(z)|z)=\tau-c_n$. Thus we can obtain

$$
2c_n = F_{Y(j)|Z}(q_{j,\tau+c_n}(z)|z)-F_{Y(j)|Z}(q_{j,\tau-c_n}(z)|z) = f_{Y(j)|Z}(q_n^*|z)\big(q_{j,\tau+c_n}(z)-q_{j,\tau-c_n}(z)\big),
$$

where q_n^* lies in between $q_{j,\tau-c_n}(z)$ and $q_{j,\tau+c_n}(z)$, which implies that

$$
\sup_{z\in\mathcal{Z}}\big|q_{j,\tau+c_n}(z)-q_{j,\tau-c_n}(z)\big| = O(c_n)
$$

by Assumption 2.3.

For $\mathcal{M}_2$, since $q_{j,\tau-c_n}(z)\leq q_{j,\tau}(z)\leq q_{j,\tau+c_n}(z)$ and $\sup_{z\in\mathcal{Z}}\big|q_{j,\tau+c_n}(z)-q_{j,\tau-c_n}(z)\big|=O(c_n)$, we know that there exists a constant A which does not rely on z, such that

$$
\begin{aligned}
\mathcal{M}_2 &\leq \sup_{z\in\mathcal{Z}}\Big|S_{n,j}^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}K_{h,i}(z)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\leq Y_i\leq q_{j,\tau}(z)+Ac_n\}\Big|\\
&\leq \sup_{z\in\mathcal{Z}}\Big|S_{n,j}^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\Big[K_{h,i}(z)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\leq Y_i\leq q_{j,\tau}(z)+Ac_n\}\\
&\quad -E\big(K_{h,i}(z)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\leq Y_i\leq q_{j,\tau}(z)+Ac_n\}\big)\Big]\Big|\\
&\quad +\sup_{z\in\mathcal{Z}}\Big|S_{n,j}^{-1}(z)E\big(K_{h,i}(z)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\leq Y_i\leq q_{j,\tau}(z)+Ac_n\}\big)\Big|\\
&:= \mathcal{M}_{2,1}+\mathcal{M}_{2,2}.
\end{aligned}
$$

Next, we show $\mathcal{M}_{2,1}=o_p(c_n)$. To this aim, first note that the following classes of functions (i) $\{K\big((Z_i-z)/h\big):z\in\mathcal{Z}\}$, (ii) $\big\{I\{Y_i\leq q_{j,\tau}(z)-Ac_n\}:z\in\mathcal{Z}\big\}$ and (iii) $\big\{I\{Y_i\leq q_{j,\tau}(z)+Ac_n\}:z\in\mathcal{Z}\big\}$ are all Euclidean for a constant envelope (Lemma 18 and Lemma 22 of Nolan and Pollard (1987)). The closer properties of Euclidean classes further dictates that $\big\{K\big((Z_i-z)/h\big)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\leq Y_i\leq q_{j,\tau}(z)+Ac_n\}:z\in\mathcal{Z}\big\}$ is also Euclidean, which together with $\frac{\ln n}{nhc_n}=o(1)$ implies the conditions required by Theorem II.37 of Pollard

(1984) are met. In addition, by straightforward calculations, we know that

$$\begin{aligned}
& E\Big[\Big(K\big((Z_i-z)/h\big)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\le Y_i\le q_{j,\tau}(z)+Ac_n\}\Big)^2\Big] \\
= \ & E\Big\{K^2\big((Z_i-z)/h\big)p(X_i)^{-j}(1-p(X_i))^{j-1}E\Big[I\{q_{j,\tau}(z)-Ac_n\le Y_i(j)\le q_{j,\tau}(z)+Ac_n\}\Big|X_i\Big]\Big\} \\
= \ & O(hc_n)
\end{aligned}$$

by Assumption 2.3. Hence, by Theorem II.37 of Pollard (1984),

$$\begin{aligned}
\sup_{z\in\mathcal{Z}}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big[& K\big((Z_i-z)/h\big)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\le Y_i\le q_{j,\tau}(z)+Ac_n\} \\
& -E\big(K\big((Z_i-z)/h\big)W_j(X_i,D_i)I\{q_{j,\tau}(z)-Ac_n\le Y_i\le q_{j,\tau}(z)+Ac_n\}\big)\Big]\Big| = o_p(hc_n).
\end{aligned}$$

Since $S_{n,j}(z)$ is bounded away from zero by Assumption 2.2-2.3 and Lemma 2, we know that $\mathcal{M}_{2,1}=o_p(c_n)$. Also, it is easy to show that $\mathcal{M}_{2,2}=O(c_n)$. This together with $\mathcal{M}_{2,1}=o_p(c_n)$ leads to $\mathcal{M}_2=O_p(c_n)$. Finally, it is easy to see that

$$\mathcal{M}_3=O_p(c_n)$$

holds. Therefore,

$$\sup_{z\in\mathcal{Z}}\big|\widehat{q}_{j,\tau}(z)-\bar{q}_{j,\tau}(z)\big|=O_p\Bigg\{\max\bigg(\frac{\ln n}{\sqrt{n}},\Big(\frac{\ln n}{nh}\Big)^{3/4}\bigg)\Bigg\}.$$

This completes the proof. □

Lemma 10. *Under Assumptions 2.1-2.5, we have*

$$\sup_{z\in\mathcal{Z}}\Big|\widehat{q}_{j,\tau}(z)-q_{j,\tau}(z)-\frac{1}{n}\sum_{i=1}^{n}\varrho_{n,j}(Y_i,X_i,D_i;z)\Big|=O_p\Bigg\{\max\bigg(\frac{\ln n}{\sqrt{n}},\Big(\frac{\ln n}{nh}\Big)^{3/4}\bigg)\Bigg\},$$

for $j=0,1$.

Proof of Lemma 10: The result comes from Lemma 8 and Lemma 9. □

Lemma 11. *Let $R_1,R_2,\cdots$ be an i.i.d. sequence. Suppose that the U-statistic $U_n=\sum_{1\le i<j\le n}H_n(R_i,R_j)$ with symmetric variable function H_n is centered (i.e., $E[H_n(R_1,R_2)]=0$) and degenerated (i.e., $E[H_n(R_1,R_2)|R_1=z_1]=0$ almost surely for all z_1). Let*

$$\sigma_n^2=E\big[H_n^2(R_1,R_2)\big],\quad \widetilde{\Pi}_n(z_1,z_2)=E\big[H_n(R_i,z_1)H_n(R_i,z_2)\big].$$

Then if

$$\lim_{n\to\infty} \frac{E\big[\widetilde{\Pi}_n^2(R_1,R_2)\big] + n^{-1}E\big[H_n^4(R_1,R_2)\big]}{\big(E[H_n^2(R_1,R_2)]\big)^2} = 0,$$

we have that as $n \to \infty$,

$$\frac{2^{1/2}}{n\sigma_n} U_n \xrightarrow{D} \mathcal{N}(0,1).$$

Proof of Lemma 11: The result is Theorem 1 given in Hall (1984).

Lemma 12. *Suppose the conditions required by Theorem 3.1 are satisfied. Then*

$$n\sqrt{h}\left\{\int \Big[\frac{1}{n}\sum_{i=1}^{n}\big(\varrho_{n,1}(Y_i,X_i,D_i;z) - \varrho_{n,0}(Y_i,X_i,D_i;z)\big)\Big]^2 \omega(z)dz - \mu_J\right\} \xrightarrow{D} \mathcal{N}(0,\sigma_J^2),$$

where

$$\mu_J = \frac{1}{nh}\int K^2(s)ds \int \left\{\frac{\mu_1(z;z)}{f^2_{Y(1)|Z}(q_{1,\tau}(z)|z)} + \frac{\mu_0(z;z)}{f^2_{Y(0)|Z}(q_{0,\tau}(z)|z)}\right\}\frac{\omega(z)}{f_Z(z)}dz,$$

and

$$\sigma_J^2 = 2\int\Big(\int K(t)K(t+s)dt\Big)^2 ds \int \left\{\frac{\mu_1(u;u)}{f^2_{Y(1)|Z}(q_{1,\tau}(u)|u)} + \frac{\mu_0(u;u)}{f^2_{Y(0)|Z}(q_{0,\tau}(u)|u)}\right\}^2 \frac{\omega^2(u)}{f_Z^2(u)}du,$$

with

$$\mu_0(z;u) = E\left[\frac{1}{1-p(X_i)}\big(I\{Y_i(0)\le q_{0,\tau}(u)\} - \tau\big)^2 \big| Z_i = z\right],$$

and

$$\mu_1(z;u) = E\left[\frac{1}{p(X_i)}\big(I\{Y_i(1)\le q_{1,\tau}(u)\} - \tau\big)^2 \big| Z_i = z\right].$$

Proof of Lemma 12: Let $\gamma_n(Y_i,X_i,D_i;z) = \varrho_{n,1}(Y_i,X_i,D_i;z) - \varrho_{n,0}(Y_i,X_i,D_i;z)$ and $\widetilde{\gamma}_n(Y_i,X_i,D_i;z) = \gamma_n(Y_i,X_i,D_i;z) - E\big[\gamma_n(Y_i,X_i,D_i;z)\big]$, then,

$$\begin{aligned}
&\int \Big(\frac{1}{n}\sum_{i=1}^{n}\gamma_n(Y_i,X_i,D_i;z)\Big)^2\omega(z)dz \qquad \text{(A.15)}\\
=\;& n^{-2}\sum_{i,k=1}^{n}\int \gamma_n(Y_i,X_i,D_i;z)\gamma_n(Y_k,X_k,D_k;z)\omega(z)dz\\
=\;& 2n^{-2}\sum_{1\le i<k\le n}\int \widetilde{\gamma}_n(Y_i,X_i,D_i;z)\widetilde{\gamma}_n(Y_k,X_k,D_k;z)\omega(z)dz + n^{-2}\sum_{i=1}^{n}\int \gamma_n^2(Y_i,X_i,D_i;z)\omega(z)dz
\end{aligned}$$

$$
\begin{aligned}
&+2n^{-2}(n-1)\sum_{i=1}^{n}\int \widetilde{\gamma}_n(Y_i, X_i, D_i; z)\cdot E[\gamma_n(Y_1, X_1, D_1; z)]\omega(z)dz \\
&+n^{-1}(n-1)\int E^2\big[\gamma_n(Y_1, X_1, D_1; z)\big]\omega(z)dz \\
:=\;& I_{n,1}+I_{n,2}+I_{n,3}+I_{n,4}.
\end{aligned}
$$

We first consider the term $I_{n,1}$. Let $R_i=(Y_i, X_i, D_i)$ and

$$
H_n(R_i, R_k)=\frac{2}{n^2}\int \widetilde{\gamma}_n(R_i; z)\widetilde{\gamma}_n(R_k; z)\omega(z)dz.
$$

Then, $I_{n,1}=\sum_{1\le i<k\le n} H_n(R_i, R_k)$ is a centered and degenerated U-statistic. Thus,

$$
\begin{aligned}
&E\big[H_n(R_i, R_k)^2\big] \qquad\qquad (A.16)\\
=\;& E\left[\frac{4}{n^4}\int\int \widetilde{\gamma}_n(R_i; u)\widetilde{\gamma}_n(R_k; u)\omega(u)\,\widetilde{\gamma}_n(R_i; v)\widetilde{\gamma}_n(R_k; v)\omega(v)\,dudv\right] \\
=\;& \frac{4}{n^4}\int\int E\left[\widetilde{\gamma}_n(R_i; u)\widetilde{\gamma}_n(R_i; v)\widetilde{\gamma}_n(R_k; u)\widetilde{\gamma}_n(R_k; v)\right]\omega(u)\omega(v)\,dudv \\
=\;& \frac{4}{n^4}\int\int E^2\big[\widetilde{\gamma}_n(R_i; u)\widetilde{\gamma}_n(R_i; v)\big]\omega(u)\omega(v)\,dudv
\end{aligned}
$$

By noting that $E\big[\varrho_{n,j}(Y_i, X_i, D_i; z)\big]=O(h^2)$ for $j=0$ and 1, as in (A.13), then, we have

$$
\begin{aligned}
&E\Big[\widetilde{\gamma}_n(R_i; u)\widetilde{\gamma}_n(R_i; v)\Big] \qquad\qquad (A.17)\\
=\;& S_{n,1}^{-1}(u)S_{n,1}^{-1}(v)E\Big[K_h\big(Z_i-u\big)K_h\big(Z_i-v\big)\frac{D_i}{p^2(X_i)}\varphi_\tau(Y_i; q_{1,\tau}(u))\varphi_\tau(Y_i; q_{1,\tau}(v))\Big] \\
&+S_{n,0}^{-1}(u)S_{n,0}^{-1}(v)E\Big[K_h\big(Z_i-u\big)K_h\big(Z_i-v\big)\frac{1-D_i}{(1-p(X_i))^2}\varphi_\tau(Y_i; q_{0,\tau}(u))\varphi_\tau(Y_i; q_{0,\tau}(v))\Big]+O(h^4) \\
=\;& S_{n,1}^{-1}(u)S_{n,1}^{-1}(v)E\Big[K_h\big(Z_i-u\big)K_h\big(Z_i-v\big)\frac{1}{p(X_i)}\varphi_\tau(Y_i; q_{1,\tau}(u))\varphi_\tau(Y_i; q_{1,\tau}(v))\Big] \\
&+S_{n,0}^{-1}(u)S_{n,0}^{-1}(v)E\Big[K_h\big(Z_i-u\big)K_h\big(Z_i-v\big)\frac{1}{1-p(X_i)}\varphi_\tau(Y_i; q_{0,\tau}(u))\varphi_\tau(Y_i; q_{0,\tau}(v))\Big]+O(h^4) \\
=\;& S_{n,1}^{-1}(u)S_{n,1}^{-1}(v)E\Big[K_h\big(Z_i-u\big)K_h\big(Z_i-v\big)\kappa_1(Z_i; u, v)\Big] \\
&+S_{n,0}^{-1}(u)S_{n,0}^{-1}(v)E\Big[K_h\big(Z_i-u\big)K_h\big(Z_i-v\big)\kappa_0(Z_i; u, v)\Big]+O(h^4) \\
=\;& \frac{1}{h}S_{n,1}^{-1}(u)S_{n,1}^{-1}(v)\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_1(u+ht; u, v)f_Z(u+ht)dt \\
&+\frac{1}{h}S_{n,0}^{-1}(u)S_{n,0}^{-1}(v)\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_0(u+ht; u, v)f_Z(u+ht)dt+O(h^4),
\end{aligned}
$$

where

$$\kappa_1(z;u,v) = E\Big[\frac{1}{p(X_i)}\varphi_\tau(Y_i(1);q_{1,\tau}(u))\varphi_\tau(Y_i(1);q_{1,\tau}(v))\big|Z_i = z\Big],$$

and

$$\kappa_0(z;u,v) = E\Big[\frac{1}{1-p(X)}\varphi_\tau(Y_i(0);q_{0,\tau}(u))\varphi_\tau(Y_i(0);q_{0,\tau}(v))\big|Z_i = z\Big],$$

with $\varphi_\tau(y;q) = I(y \le q) - \tau$. Thus,

$$\begin{aligned}
& E^2\Big[\widetilde{\gamma}_n(R_i;u)\widetilde{\gamma}_n(R_i;v)\Big] \\
= \;& \frac{1}{h^2}S_{n,1}^{-2}(u)S_{n,1}^{-2}(v)\Big(\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_1(u+ht;u,v)f_Z(u+ht)dt\Big)^2 \\
& +\frac{1}{h^2}S_{n,0}^{-2}(u)S_{n,0}^{-2}(v)\Big(\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_0(u+ht;u,v)f_Z(u+ht)dt\Big)^2 \\
& +\frac{2}{h^2}S_{n,1}^{-1}(u)S_{n,1}^{-1}(v)S_{n,0}^{-1}(u)S_{n,0}^{-1}(v)\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_1(u+ht;u,v)f_Z(u+ht)dt \\
& \times\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_0(u+ht;u,v)f_Z(u+ht)dt \\
& +2O(h^3)S_{n,1}^{-1}(u)S_{n,1}^{-1}(v)\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_1(u+ht;u,v)f_Z(u+ht)dt \\
& +2O(h^3)S_{n,0}^{-1}(u)S_{n,0}^{-1}(v)\int K(t)K\Big(t+\frac{u-v}{h}\Big)\kappa_0(u+ht;u,v)f_Z(u+ht)dt + O(h^8).
\end{aligned}$$

An application of (A.16) and some straightforward calculations implies that

$$\begin{aligned}
E\big[H_n(R_i,R_k)^2\big] &= \frac{4}{n^4}\int\int E^2\big[\widetilde{\gamma}_n(R_i;u)\widetilde{\gamma}_n(R_i;v)\big]\omega(u)\omega(v)\,dudv \\
&= \frac{4}{n^4h}\Big\{\int\Big(\int K(t)K(t+s)dt\Big)^2 ds\cdot\Big(\int S_{n,1}^{-4}(u)\kappa_1^2(u;u,u)f_Z^2(u)\omega^2(u)du \\
&\quad +\int S_{n,0}^{-4}(u)\kappa_0^2(u;u,u)f_Z^2(u)\omega^2(u)du \\
&\quad +2\int S_{n,1}^{-2}(u)S_{n,0}^{-2}(u)\kappa_1(u;u,u)\kappa_0(u;u,u)f_Z^2(u)\omega^2(u)du\Big) + o(1)\Big\}.
\end{aligned}$$

This, coupled with $S_{n,j}(z) = f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z) + O(h^2)$ for $j = 0,1$, yields

$$\begin{aligned}
E\big[H_n(R_1,R_2)^2\big] &= \frac{4}{n^4h}\Big(\int\Big(\int K(t)K(t+s)dt\Big)^2 ds \\
&\quad\times\int\Big\{\frac{\kappa_1(u;u,u)}{f_{Y(1)|Z}^2(q_{1,\tau}(u)|u)}+\frac{\kappa_0(u;u,u)}{f_{Y(0)|Z}^2(q_{0,\tau}(u)|u)}\Big\}^2\frac{\omega^2(u)}{f_Z^2(u)}du + o(1)\Big) \\
&= O\Big(\frac{1}{n^4h}\Big).
\end{aligned}$$

Similarly, by straightforward calculations, we can obtain

$$E\big[\widetilde{\Pi}_n(R_1,R_2)^2\big] = O\Big(\Big(\frac{1}{n^2h^2}\Big)^4 h^7\Big),$$

and

$$E\big[H_n(R_1,R_2)^4\big] = O\Big(\Big(\frac{1}{n^2h^2}\Big)^4 h^5\Big).$$

Thus, the condition

$$\lim_{n\to\infty} \frac{E\big[\widetilde{\Pi}_n(R_1,R_2)^2\big] + n^{-1}E\big[H_n(R_1,R_2)^4\big]}{\Big(E\big[H_n(R_1,R_2)^2\big]\Big)^2} = 0$$

in Lemma 11 is satisfied, so that

$$\frac{\sqrt{2}}{nE^{1/2}\big[H_n(R_1,R_2)^2\big]}\, I_{n,1} \xrightarrow{D} \mathcal{N}(0,1),$$

or equivalently,

$$n\sqrt{h}\, I_{n,1} \xrightarrow{D} \mathcal{N}(0,\sigma_J^2). \tag{A.18}$$

Next, we move to the term $I_{n,2} = n^{-2}\sum_{i=1}^{n}\int \gamma_n^2(Y_i,X_i,D_i;z)\omega(z)dz$. Note that

$$\begin{aligned}
& E\big[\gamma_n^2(Y_i,X_i,D_i;z)\big] \\
= \;& S_{n,1}^{-2}(z)E\Big[\frac{1}{h^2}K^2\Big(\frac{Z_i-z}{h}\Big)\frac{D_i}{p^2(X_i)}(I\{Y_i\le q_{1,\tau}(z)\}-\tau)^2\Big] \\
& \quad + S_{n,0}^{-2}(z)E\Big[\frac{1}{h^2}K^2\Big(\frac{Z_i-z}{h}\Big)\frac{1-D_i}{(1-p(X_i))^2}(I\{Y_i\le q_{0,\tau}(z)\}-\tau)^2\Big] \\
= \;& S_{n,1}^{-2}(z)E\Big[\frac{1}{h^2}K^2\Big(\frac{Z_i-z}{h}\Big)\frac{1}{p(X_i)}(I\{Y_i(1)\le q_{1,\tau}(z)\}-\tau)^2\Big] \\
& \quad + S_{n,0}^{-2}(z)E\Big[\frac{1}{h^2}K^2\Big(\frac{Z_i-z}{h}\Big)\frac{1}{1-p(X_i)}(I\{Y_i(0)\le q_{0,\tau}(z)\}-\tau)^2\Big] \\
= \;& S_{n,1}^{-2}(z)E\Big[\frac{1}{h^2}K^2\Big(\frac{Z_i-z}{h}\Big)\mu_1(Z_i;z)\Big] + S_{n,0}^{-2}(z)E\Big[\frac{1}{h^2}K^2\Big(\frac{Z_i-z}{h}\Big)\mu_0(Z_i;z)\Big] \\
= \;& S_{n,1}^{-2}(z)\Big[\frac{1}{h}\Big(\mu_1(z;z)f_Z(z)\int K^2(s)ds + O(h)\Big)\Big] + S_{n,0}^{-2}(z)\Big[\frac{1}{h}\Big(\mu_0(z;z)f_Z(z)\int K^2(s)ds + O(h)\Big)\Big] \\
= \;& \frac{1}{h}\Big\{\int K^2(s)ds\cdot\Big(S_{n,1}^{-2}(z)\mu_1(z;z) + S_{n,0}^{-2}(z)\mu_0(z;z)\Big)f_Z(z) + O(h)\Big\},
\end{aligned}$$

coupled with $S_{n,j}(z) = f_Z(z)f_{Y(j)|Z}(q_{j,\tau}(z)|z) + O(h^2)$ for $j=0$ and 1, we have

$$E\big(I_{n,2}\big) = \frac{1}{n}\int E\big[\gamma_n^2(Y_i,X_i,D_i;z)\big]\omega(z)dz \tag{A.19}$$

$$
\begin{aligned}
&= \frac{1}{nh}\Big\{\int K^2(s)ds\cdot\int\Big(S_{n,1}^{-2}(z)\mu_1(z;z)+S_{n,0}^{-2}(z)\mu_0(z;z)\Big)f_Z(z)\omega(z)\Big\}+O\Big(\frac{1}{n}\Big)\\
&= \frac{1}{nh}\int K^2(s)ds\cdot\int\left\{\frac{\lambda_1(z;z)}{f_{Y(1)|Z}^2(q_{1,\tau}(z)|z)}+\frac{\lambda_0(z;z)}{f_{Y(0)|Z}^2(q_{0,\tau}(z)|z)}\right\}\frac{\omega(z)}{f(z)}dz+O\Big(\frac{1}{n}\Big)\\
&= \mu_J+O\Big(\frac{1}{n}\Big).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\mathrm{Var}\big(n\sqrt{h}I_{n,2}\big) &= E\Big\{n\sqrt{h}\big[I_{n,2}-E(I_{n,2})\big]\Big\}^2\\
&= n^{-1}h\Big\{E\Big[\Big(\int\gamma_n^2(Y_i,X_i,D_i;z)\omega(z)dz\Big)^2\Big]-E^2\Big[\int\gamma_n^2(Y_i,X_i,D_i;z)\omega(z)dz\Big]\Big\}\\
&= n^{-1}h\Big\{E\Big[\Big(\int\gamma_n^2(Y_i,X_i,D_i;z)\omega(z)dz\Big)^2\Big]-O(h^{-2})\Big\}\\
&= n^{-1}h\Big\{\int\int E\Big[\gamma_n^2(Y_i,X_i,D_i;u)\gamma_n^2(Y_i,X_i,D_i;v)\Big]\omega(u)\omega(v)dudv-O(h^{-2})\Big\}\\
&= n^{-1}h\cdot O(h^{-2})\to 0,
\end{aligned}
$$

together with (A.19), we have

$$
n\sqrt{h}\big[I_{n,2}-\mu_J\big] \;=\; n\sqrt{h}\big[I_{n,2}-E(I_{n,2})\big]+o_p(1)=o_p(1). \tag{A.20}
$$

We now consider $I_{n,3}$ and $I_{n,4}$. By noting that $E\Big[\gamma_n(Y_i,X_i,D_i;z)\Big]=O(h^2)$ from (A.13) and

$$
\begin{aligned}
& E\Big(\int\widetilde{\gamma}_n(Y_i,X_i,D_i;z)\omega(z)\,dz\Big)^2\\
=\; & \int\int E\Big[\widetilde{\gamma}_n(Y_i,X_i,D_i;u)\widetilde{\gamma}_n(Y_i,X_i,D_i;v)\Big]\omega(u)\omega(v)dudv=O(1)
\end{aligned}
$$

from (A.17), we obtain

$$
\begin{aligned}
E\big[I_{n,3}\big]^2 &= \mathrm{Var}\big[I_{n,3}\big]\\
&= 4n^{-4}(n-1)^2\sum_{i=1}^{n}\mathrm{Var}\Big(\int\widetilde{\gamma}_n(Y_i,X_i,D_i;z)\cdot E[\gamma_n(Y_1,X_1,D_1;z)]\omega(z)dz\Big)\\
&= 4n^{-4}(n-1)^2\cdot n\,E\Big(\int\widetilde{\gamma}_n(Y_i,X_i,D_i;z)\cdot E[\gamma_n(Y_1,X_1,D_1;z)]\omega(z)dz\Big)^2\\
&= 4n^{-3}(n-1)^2E\Big(\int\widetilde{\gamma}_n(Y_i,X_i,D_i;z)\cdot O(h^2)\omega(z)dz\Big)^2\\
&= O(n^{-1}h^4).
\end{aligned}
$$

Hence,

$$n\sqrt{h}\ I_{n,3} = n\sqrt{h} \cdot O(n^{-1/2}h^2) = O(\sqrt{nh^5}). \tag{A.21}$$

Furthermore, we have

$$n\sqrt{h}\ I_{n,4} = n\sqrt{h}\ n^{-1}(n-1)\int E^2\big[\gamma_n(Y_1, X_1, D_1; z)\big]\omega(z)dz = O(nh^{9/2}). \tag{A.22}$$

It follows by combining (A.15), (A.18), (A.20), (A.21) and (A.22) that

$$n\sqrt{h}\left[\int \Big(\frac{1}{n}\sum_{i=1}^{n}\gamma_n(Y_i, X_i, D_i; z)\Big)^2\omega(z)dz - \mu_J\right] \xrightarrow{D} \mathcal{N}(0, \sigma_J^2)$$

under the assumption that $nh^{9/2} \to 0$. □

Proof of Theorem 3.1: Let $\Delta_\tau(z)$ and Δ_τ be the partially conditional quantile treatment effect on $Z_i = z$ and the unconditional quantile treatment effect, respectively. Then

$$\begin{aligned} J_n &= \int \Big(\widehat{\Delta}_\tau(z) - \widehat{\Delta}_\tau\Big)^2\omega(z)dz \\ &= \int \Big[\Big(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\Big) + \Big(\Delta_\tau - \widehat{\Delta}_\tau\Big) + \Big(\Delta_\tau(z) - \Delta_\tau\Big)\Big]^2\omega(z)dz \end{aligned}$$

where $\widehat{\Delta}_\tau$ is a $\sqrt{n}$-consistent estimate of Δ_τ. Recall that $\gamma_n(Y_i, X_i, D_i; z) = \varrho_{n,1}(Y_i, X_i, D_i; z) - \varrho_{n,0}(Y_i, X_i, D_i; z)$. Under the null hypothesis H_0, $\Delta_\tau(z) - \Delta_\tau \equiv 0$, thus,

$$\begin{aligned} J_n &= \int \Big(\frac{1}{n}\sum_{i=1}^{n}\gamma_n(Y_i, X_i, D_i; z) + e_n\Big)^2\omega(z)dz \\ &= \int \Big(\frac{1}{n}\sum_{i=1}^{n}\gamma_n(Y_i, X_i, D_i; z)\Big)^2\omega(z)dz + e_n^2 + 2e_n\int \frac{1}{n}\sum_{i=1}^{n}\gamma_n(Y_i, X_i, D_i; z)\,\omega(z)dz \\ &:= J_{n,1} + J_{n,2} + J_{n,3}, \end{aligned}$$

where $e_n = O_p\left(\max\left\{\frac{\ln n}{\sqrt{n}}, \left(\frac{\ln n}{nh}\right)^{3/4}\right\}\right)$ by Lemma 10. Now, it is easy to verify that $n\sqrt{h}\,J_{n,2} = o_p(1)$. Also, by noting that $E\Big[\gamma_n(Y_i, X_i, D_i; z)\Big] = O(h^2)$ from (A.13) and $E\bigg(\int \widetilde{\gamma}_n(Y_i, X_i, D_i; z)\omega(z)\,dz\bigg)^2 = O(1)$ from (A.17), then, we have

$$e_n \int \frac{1}{n}\sum_{i=1}^{n}\gamma_n(Y_i, X_i, D_i; z)\,\omega(z)dz$$

$$
\begin{aligned}
&= e_n \int \frac{1}{n} \sum_{i=1}^{n} \widetilde{\gamma}_n(Y_i, X_i, D_i; z)\, \omega(z) dz + \int E\Big[\gamma_n(Y_i, X_i, D_i; z)\Big]\, \omega(z) dz \\
&= O_p\big(n^{-1/2} e_n\big) + O_p(h^2 e_n).
\end{aligned}
$$

Hence, $n\sqrt{h}\, J_{n,3} = o_p(1)$ under the assumption $nh^{9/2} \to 0$. Finally, an application of Lemma 12 leads to

$$
n\sqrt{h}\big(J_n - \mu_J\big) = n\sqrt{h}\big(J_{n,1} - \mu_J + J_{n,2} + J_{n,3}\big) \xrightarrow{D} \mathcal{N}(0, \sigma_J^2).
$$

Now, we consider the case under the alternative hypothesis H_1. Under H_1, it is easy to show that $J_n - \mu_J = \int \Big(\Delta_\tau(z) - \Delta_\tau\Big)^2 \omega(z) dz + o_p(1)$. Since $\int \Big(\Delta_\tau(z) - \Delta_\tau\Big)^2 \omega(z) dz$ is a positive constant under H_1, so that

$$
n\sqrt{h}\big(J_n - \mu_J\big) \xrightarrow{p} +\infty.
$$

This completes the proof. □