Econometric Theory, **41**, 2025, 660–687.
doi:10.1017/S0266466624000045

A NONPARAMETRIC TEST OF HETEROGENEITY IN CONDITIONAL QUANTILE TREATMENT EFFECTS

Zongwu Cai
University of Kansas

Ying Fang
Xiamen University

Ming Lin
Xiamen University

Shengfang Tang
Jiangxi University of Finance and Economics

This paper proposes a nonparametric test to assess whether there exist heterogeneous quantile treatment effects (QTEs) of an intervention on the outcome of interest across different sub-populations defined by covariates of interest. Specifically, a consistent test statistic based on the Cramér–von Mises type criterion is developed to test if the treatment has a constant quantile effect for all sub-populations defined by covariates of interest. Under some regularity conditions, the asymptotic behaviors of the proposed test statistic are investigated under both the null and alternative hypotheses. Furthermore, a nonparametric Bootstrap procedure is suggested to approximate the finite-sample null distribution of the proposed test; then, the asymptotic validity of the proposed Bootstrap test is theoretically justified. Through Monte Carlo simulations, we demonstrate the power properties of the test in finite samples. Finally, the proposed testing approach is applied to investigate whether there exists heterogeneity for the QTE of maternal smoking during pregnancy on infant birth weight across different age groups of mothers.

1. INTRODUCTION

In program evaluation studies, it is important to learn about the heterogeneous impacts of policy variables on different quantiles of the outcome distribution of interest. Examples include, but are not limited to, evaluating the effects of government training programs on lower quantiles of earning distributions studied

The authors are grateful to the editor (Peter C. B. Phillips), the co-editor, and the two anonymous referees for their insightful comments which substantially improved the paper. This research is supported by the National Science Foundation of China (Grant Nos. 72033008, 72301119, 72133002, and 71988101). Address correspondence to Ming Lin, Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian, China, linming50@xmu.edu.cn; Shengfang Tang, School of Finance, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China, tangshengfang@jxufe.edu.cn.

by LaLonde (1995) and Abadie, Angrist, and Imbens (2002), the effects of government-subsidized saving programs on lower tails of savings distributions, and among many others. To characterize the heterogeneous effects along with the outcome distribution, quantile treatment effect (QTE), originally suggested by Doksum (1974) and Lehmann (1975) and defined as the difference between the quantiles of the marginal potential distributions of the treatment and control responses, provides a powerful tool to document such heterogeneity. In the last few decades, QTE has gained increasing popularity in economics, political science, and many other social, behavioral, and statistical sciences. Recent studies on QTE include Abadie et al. (2002), Chernozhukov and Hansen (2005), Firpo (2007), Frölich and Melly (2013), Donald and Hsu (2014), and the references therein.

The aforementioned papers mainly focus on identification and estimation of the QTE for the overall population or the treated group under various assumptions. It is generally believed in the program evaluation literature that the effect of a treatment can be heterogeneous across different individuals, as in Heckman and Robb (1985) and Heckman, Smith, and Clements (1997). In many cases, researchers may be more interested in studying the effects of programs across different individuals instead of the effects for the overall population or the sub-population of treated individuals. For example, it may be of substantive interest to investigate the heterogeneous effects of maternal smoking during pregnancy on infant birth weight across mothers of different ages. How to characterize the heterogeneity of treatment effects across different individuals is a challenge in the treatment effect literature and has been extensively considered in recent studies. For instance, to characterize the heterogeneous effects across different sub-populations defined by some covariates of interest, Abrevaya, Hsu, and Lieli (2015) and Lee, Okui, and Whang (2017) considered the partially conditional average treatment effect (ATE). Unlike Abrevaya et al. (2015) and Lee et al. (2017), to simultaneously capture heterogeneities across both distributions and individuals, Cai et al. (2021)[1] and Zhou, Guo, and Zhu (2022) proposed a partially conditional quantile treatment effect (PCQTE) model, whereas Tang et al. (2021) considered a parametric model.

In this paper, we investigate whether there exists heterogeneity in quantile effects across sub-populations defined by the covariate of interest, Z, which is a subset of covariates X, under the condition that the treatment assignment is independent of the potential outcomes conditional on X. The study is motivated by the empirical estimation results in Cai et al. (2021) and Tang et al. (2021), which investigate the QTE of maternal smoking during pregnancy on infant birth weight across different age groups of mothers. The main findings in Cai et al. (2021) are that there is a significant negative effect of smoking on infant birth weight across all mothers' ages and quantiles for both whites and blacks and there is substantial heterogeneity across different mothers' ages for whites but not for blacks. Based on

[1]Please note that the working paper version of this paper in English with a modification can be downloaded at https://econpapers.repec.org/paper/kanwpaper/202005.htm.

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

these estimation results, from statistical and empirical perspectives, it is interesting to test whether or not the conditional QTEs, conditional on mothers' ages, for both whites and blacks, change over mothers' ages, in other words, whether there exists heterogeneity for the QTEs of maternal smoking on infant birth weight across different age groups of mothers for both whites and blacks. To this end, we propose a novel test to assess whether there exist heterogeneously distributional effects for an intervention on an outcome of interest across different sub-populations defined by covariates of interest. Specifically, a nonparametric test is developed for testing the null hypothesis that the treatment has a constant QTE for all sub-populations defined by covariates of interest. In other words, the null hypothesis is that there is no heterogeneity in QTEs by covariates of interest. To this end, a consistent test statistic is constructed based on the Cramér–von Mises type criterion.

Under some regularity conditions, we establish the asymptotic distribution of the proposed test statistic under both the null and alternative hypotheses and investigate the power of our test against a sequence of local alternatives. However, to calculate the critical value of the proposed test statistic under the null hypothesis, one needs to consistently estimate the conditional density of the potential outcomes conditional on covariates of interest involved in the asymptotic bias and asymptotic variance. This is not an easy task, as recognized in the literature. To overcome this problem, a nonparametric Bootstrap procedure is proposed to approximate the finite-sample null distribution of the proposed test. Furthermore, the asymptotic validity of the proposed Bootstrap test is justified. Through Monte Carlo simulations, we demonstrate the power properties of the test in finite samples. As an empirical illustration, the proposed testing approach is applied to investigating whether there exists heterogeneity for the QTE of maternal smoking during pregnancy on infant birth weight across different age groups of mothers. The results show that the QTEs of maternal smoking on infant birth weight for whites change statistically significantly over mother's age for all quantile levels. By contrast, for blacks, the effects vary slightly with age for all quantile levels, but the results are not statistically significant. This lends support to the findings in Cai et al. (2021).

The present paper relates to several earlier lines of research. For example, Crump et al. (2008) developed two nonparametric tests based on a 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 ATE conditional on the covariates is identical for all sub-populations, in other words, whether heterogeneity by covariates exists in ATE. Further, Lee and Whang (2009) tested whether the conditional QTE is significant, conditional on the whole set of covariates. By contrast, our focus is on testing whether the partially conditional QTE is a constant, in which the constant needs to be estimated. More importantly, one may be interested in studying the heterogeneous effect on some particular covariates instead of the whole set of covariates, for example, the effect of maternal smoking during pregnancy on infant birth weight across mothers of different ages. Moreover, Escanciano and Goh (2014) considered a nonparametric test of the specification of a linear conditional quantile function over a continuum of quantile levels; they showed that the use of an orthogonal projection on the tangent space of nuisance parameters at each quantile index can improve power and facilitate the simulation of critical values via the application of a simple multiplier Bootstrap procedure. Finally, Dong, Li, and Feng (2019) introduced a new approach to assess the lack of fit for quantile regression models. They first transformed the lack-of-fit tests for parametric quantile regression models into checking the equality of two conditional distributions of covariates. Then, by applying some successful two-sample test statistics in the literature, they constructed two tests to check the lack of fit for low- and high-dimensional quantile regression models. Finally, to calculate the p-values or critical values, they suggested adopting the wild Bootstrap procedure.

The remainder of this paper is organized as follows: Section 2 introduces the proposed test statistic and presents its asymptotic properties under the null hypothesis. A Bootstrap procedure is suggested to approximate the finite-sample null distribution of the proposed test statistic and the asymptotic validity of the Bootstrap test is theoretically justified. Moreover, an extension to testing heterogeneity in conditional QTEs for a continuum of quantile levels is considered. In Sections 3 and 4, the finite sample properties of our test are investigated through Monte Carlo simulations; then, an empirical application is considered. Section 5 concludes the paper. Finally, the key steps for proving the theorems can be found in the Appendix, together with some auxiliary lemmas with their detailed proofs given in the Supplementary Material.

2. TESTING HETEROGENEITY FOR CONDITIONAL QTE

2.1. Test Statistic

Let us first introduce the model framework considered in this paper. To this end, let D_i be the binary treatment variable of individual i in the population, where $D_i = 1$ if individual i receives the treatment of interest, and $D_i = 0$ otherwise. Using the potential outcome framework initialized by Rubin (1974), define $Y_i(0)$ and $Y_i(1)$ as the potential outcomes of individual i if that individual is in the control group or in the treated group, respectively. Also, assume that D_i and Y_i are observed, where Y_i is the realized outcome, and $Y_i = (1 - D_i) \cdot Y_i(0) + D_i \cdot Y_i(1)$. In addition, suppose that X_i, an m-dimensional vector of pre-treatment variables for individual i, is observed too. Throughout the paper, it is assumed that $\left\{Y_i(0), Y_i(1), X_i, D_i\right\}$, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.).

Let Z_i be a d-dimensional sub-vector of X_i, where $1 \leq d \leq m$. In particular, d is much smaller than m in many applications. To simultaneously capture heterogeneities across both distributions and individuals, Cai et al. (2021) and Zhou et al. (2022) considered a PCQTE model, which is defined as

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

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

where $\tau \in (0,1)$ is the quantile level and for $j = 0$ and 1, $q_{j,\tau}(z)$ is the τth conditional quantile function of $Y_i(j)$ conditional on $Z_i = z$. It is important to note that, for each individual in the population, only one of $Y_i(0)$ and $Y_i(1)$ is observable, so that due to the missing variable, the PCQTE parameter $\Delta_\tau(z)$ in (1) can not be identified without further restrictions on the data-generating distribution. To identify the functionals in (1), it is common in the treatment effect literature to assume that assignment to treatment is unconfounded and that the probability of assignment is bounded away from 0 and 1. Formally, the following assumption is imposed throughout the paper.

Assumption 1. Assume:

(i) Unconfoundedness. Conditional on pre-treatment variables X_i, the potential outcomes are jointly independent of 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) Overlap. For all $x \in \mathcal{X}$, where $\mathcal{X}$ is the support of X_i, there exists some $\varepsilon > 0$ so that $\varepsilon < p(x) := P(D_i = 1 | X_i = x) < 1 - \varepsilon$, where $p(x)$ is called the propensity score function.

Part (i) of Assumption 1 is often referred to as the (strongly) *ignorable treatment assignment*, *conditional independence assumption* or *selection on observables* in the econometrics and/or statistics literature. It requires that conditional on the observed individual characteristics X_i, the treatment assignment D_i is independent of the potential outcomes $Y_i(0)$ and $Y_i(1)$. Although it is a strong assumption, it has been extensively employed in many fields to study the effect of treatments or programs, see, for example, Heckman et al. (1998), Dehejia and Wahba (1999), Hirano, Imbens and Ridder (2003), Abadie and Imbens (2006, 2016), and Firpo (2007). Part (ii) of Assumption 1 states that, for all values of X_i in the population, both treatment assignment levels have a positive probability of occurrence. In practice, however, there are often concerns about possible lack of common support. A common approach to address this problem is to drop observations with a propensity score close to zero or one, and focus on the treatment effect in the sub-population with a propensity score bounded away from zero and one (see Crump et al., 2009 for more details).

Under Assumption 1, Cai et al. (2021) showed that the PCQTE function $\Delta_\tau(z)$ is nonparametrically identified and further proposed a semiparametric estimation procedure to estimate $\Delta_\tau(z)$. The reader is referred to Cai et al. (2021) for more details. As discussed in the introduction, our interest is to investigate whether there exists heterogeneity in QTEs across different sub-populations defined by covariates of interest. To this end, the following hypothesis testing problem is investigated:

$$H_0 : \Delta_\tau(z) = \delta_\tau \quad \text{for all } z \in \mathcal{Z} \quad \text{versus} \quad H_1 : \Delta_\tau(z) \neq \delta_\tau \text{ for some } z \in \mathcal{Z} \qquad \textbf{(2)}$$

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

for some constant δ_τ, where $\mathcal{Z}$ is the support of Z_i. Under the null hypothesis, the partially conditional quantile effect of the treatment is a constant and under the alternative, the PCQTE varies across different sub-populations defined by Z_i.

In order to test whether the hypothesis testing problem formulated in (2) holds, a test statistic is constructed based on the Cramér–von Mises criterion as follows. Let

$$J = \int \left(\Delta_\tau(z) - \delta_\tau\right)^2 \omega(z) dz \geq 0,$$

where $\omega(z)$ is a pre-specified strictly positive and integrable weighting function, and the integral is taken over $\mathcal{Z}$. Note that $J = 0$ if and only if the null hypothesis in (2) is true. It is easy to observe that in order to construct a feasible test statistic, one should first estimate the unknown parameters $\Delta_\tau(z)$ and δ_τ.

To this end, we first consider using the Series Logit Estimator (SLE) to estimate $p(x)$ as in Hirano, Imbens, and Ridder (2003). More specifically, let $\kappa = (\kappa_1, \ldots, \kappa_m)' \in \mathbb{N}_0^m$ be an m-dimensional vector of nonnegative integers and define $|\kappa| = \sum_{i=1}^{m} \kappa_i$. Let $\{\kappa(\ell)\}_{\ell=1}^{\infty}$ be a sequence that includes all distinct vectors in $\mathbb{N}_0^m$ and satisfies the condition that $|\kappa(\ell)|$ is nondecreasing in ℓ. For $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$, let x^κ denote the power function $\prod_{j=1}^{m} x_j^{\kappa_j}$. We further define $R^L(x) = (x^{\kappa(1)}, \ldots, x^{\kappa(L)})'$ as an L-vector of power functions for $L > 0$. Then, the SLE for $p(x)$ is

$$\widehat{p}_n(x) = g\left(R^L(x)'\widehat{\pi}_L\right)$$

with $g(u) = \exp(u)/(1+\exp(u))$ and

$$\widehat{\pi}_L = \arg\max_{\pi} \sum_{i=1}^{n} \left\{D_i \ln g\left(R^L(X_i)'\pi\right) + (1-D_i)\ln\left[1 - g\left(R^L(X_i)'\pi\right)\right]\right\}.$$

Consequently, the proposed estimate for $\Delta_\tau(z)$ is

$$\widehat{\Delta}_\tau(z) = \widehat{q}_{1,\tau}(z) - \widehat{q}_{0,\tau}(z), \tag{3}$$

where, for $j = 0$ and 1,

$$\widehat{q}_{j,\tau}(z) = \inf\left\{y : \widehat{F}_{n,j}(y|z) \geq \tau\right\}$$

with $\widehat{F}_{n,j}(y|z) = \sum_{i=1}^{n} K_h(Z_i - z)\widehat{W}_{n,j}(X_i, D_i) I\{Y_i \leq y\} \Big/ \sum_{i=1}^{n} K_h(Z_i - z)\widehat{W}_{n,j}(X_i, D_i)$, $\widehat{W}_{n,0}(X_i, D_i) = (1 - D_i)/[1 - \widehat{p}_n(X_i)]$, $\widehat{W}_{n,1}(X_i, D_i) = D_i/\widehat{p}_n(X_i)$, and $K_h(z) = K(z/h)/h^d$. Here, $K(\cdot)$ is a kernel function and h is the bandwidth parameter. Furthermore, the proposed estimator for δ_τ is given by

$$\widehat{\delta}_\tau = \frac{1}{n} \sum_{i=1}^{n} \widehat{\Delta}_\tau(Z_i). \tag{4}$$

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

Finally, a test statistic using the sample analog of J is defined by

$$J_n = \int \left(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau\right)^2 \omega(z)dz.$$

It is well-known that the accuracy of the kernel estimator $\widehat{q}_{j,\tau}(z)$ might suffer from boundary effects. To overcome this difficulty, one could test $H_0: \Delta_\tau(z) = \delta_\tau$ over a trimmed subset of $\mathcal{Z}$. In this case, we can use a nonnegative weighting function $\omega(z)$ that is strictly positive only for the trimmed subset of $\mathcal{Z}$.

Remark 1. If Z is taken to be X in (2), then the hypothesis testing problem in (2) collapses into testing whether the conditional QTE is a constant for all values of the covariates. Unlike our setting, Crump et al. (2008) tested whether the conditional ATE is a constant or zero for all values of the covariates. Note that even if the conditional ATE is equal to a constant, the conditional QTE may not be a constant. Consequently, our paper complements and extends Crump et al. (2008) by testing whether there exists treatment effect heterogeneity across different subpopulations defined by a subset of the covariates.

Remark 2. Besides the testing issues displayed in (2), one may be interested in testing

$$H_0^*: \Delta_\tau(z) - \delta_\tau \leq 0 \;\; (\text{or } \geq 0) \;\; \text{for all } \; z \in \mathcal{Z}.$$

When $\delta_\tau \equiv 0$ and H_0^* holds for all τ's, it leads to stochastic dominance between $Y(0)$ and $Y(1)$ conditional on $Z = z$ for all $z \in \mathcal{Z}$. Recently, Lee, Song, and Whang (2018) developed a general method for testing inequality restrictions on nonparametric functions using a one-sided version of L_p functionals of kernel-type estimators, which can be extended to test H_0^* for a given δ_τ. Note that testing $\Delta_\tau(z) - \delta_\tau \leq 0$ and $\Delta_\tau(z) - \delta_\tau \geq 0$ for all $z \in \mathcal{Z}$ at the same time is equivalent to testing (2) when δ_τ is known. However, δ_τ is unknown in our testing problem, and needs to be estimated. In addition, when introducing the inverse probability weighting method to correct the bias introduced by the unobserved potential outcomes in Lee et al.'s (2018) testing method, the influence of the convergence rate of the estimated propensity score needs to be carefully examined. Such a topic is beyond the scope of this paper but certainly worth pursuing in future research.

2.2. Limiting Distribution of Test Statistic J_n

This subsection is devoted to investigating the asymptotic properties of the proposed test statistic J_n. Before studying the asymptotic properties of the proposed test statistic, the following technical assumptions are needed.

Assumption 2. X_i has a compact support $\mathcal{X}$ and the density function of $X_i, f_X(x)$, satisfies $\inf_{x\in\mathcal{X}} f_X(x) \geq c$ for some $c > 0$. Furthermore, the density function of Z_i, $f_Z(z)$ is twice continuously differentiable in $\mathcal{Z}$.

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

Assumption 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 $j=0$ and 1. (ii) The conditional density function $f_{Y(j)|Z}(y|z)$ is continuous and uniformly bounded away from zero for all $z \in \mathcal{Z}$ and $j=0,1$. 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 4. For $j=0$ and 1, the conditional quantile function $q_{j,\tau}(z)$ is continuously differentiable on the support of Z with bounded second-order derivatives.

Assumption 5. (i) The kernel function $K(u)$ is a symmetric, continuously differential probability density function with compact support, say, $[-1,1]$. (ii) $nh^{2d}/(\ln n)^2 \to \infty$ and $nh^{d+4}/\ln n \to 0$ as $n \to 0$.

Assumption 6. Suppose the kernel bandwidth is taken as $h = c \cdot n^{-\eta}$, where c is a positive constant and $1/(d+4) < \eta < 1/(2d)$. The propensity score $p(x)$ is continuously differentiable of order $s > \left[6/(\eta d) - 1\right]m$, and the SLE of $p(x)$ uses a power series with $L = a \cdot n^{\nu}$ for some $a > 0$ and $\nu \in \left[\frac{m}{s+m}, \frac{\eta d}{6}\right)$.

The restriction imposed on the distribution of X_i in Assumption 2 is commonly used in the literature on treatment effect evaluation (see Hirano et al., 2003; Abadie and Imbens, 2006, 2016; Firpo, 2007; Abrevaya et al., 2015, among others). Assumption 3 guarantees that the conditional quantile function $q_{j,\tau}(z)$ is unique and well defined and the smoothness conditions imposed are easily satisfied in practice. The smoothness conditions on the conditional quantile function $q_{j,\tau}(z)$ for $j=0$ and 1 imposed in Assumption 4 are also easily satisfied in practice. Assumption 5 gives conditions for the kernel and its bandwidth. It requires $nh^{d+4}/\ln n \to 0$ to remove higher-order terms in the bias of the asymptotic null distribution established in Theorem 1. Note that Assumption 5(ii) requires that the dimension of the covariate of interest, Z_i, cannot exceed 3. Although this is a stringent condition, it is sufficient in many applications. Indeed, as pointed out by Abrevaya et al. (2015), the case of $d=1$ is the most relevant case in practice. Finally, Assumption 6 requires the propensity score function to be sufficiently smooth, so that the SLE of $p(x)$ converges to the true propensity score function at a fast enough rate. Note that the condition $s > \left[6/(\eta d) - 1\right]m$ in Assumption 6 guarantees that $\nu \in \left[\frac{m}{s+m}, \frac{\eta d}{6}\right)$ is a non-empty set.

Under the assumptions listed above, we now can state our main result on the asymptotic properties of the proposed test statistic J_n and its proof can be found in the Appendix. For easy presentation, first, define some notations as follows. Let $\mu_{0,\tau}(z;u) = E\left\{\left[I\{Y_i(0) \le q_{0,\tau}(u)\} - \tau\right]^2/[1-p(X_i)] \,\middle|\, Z_i = z\right\}$ and $\mu_{1,\tau}(z;u) = E\left\{\left[I\{Y_i(1) \le q_{1,\tau}(u)\} - \tau\right]^2/p(X_i) \,\middle|\, Z_i = z\right\}$. Then, we have the following asymptotic results.

THEOREM 1. *Suppose that Assumptions 1–6 are satisfied. Then, under the null hypothesis H_0 in (2), one has*

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

$$nh^{d/2}\big(J_n-\mu_J\big) \xrightarrow{D} \mathcal{N}(0,\sigma_J^2),$$

where

$$\mu_J=\frac{\nu_0(K)}{nh^d}\int\left\{\frac{\mu_{1,\tau}(z;z)}{f^2_{Y(1)|Z}(q_{1,\tau}(z)|z)}+\frac{\mu_{0,\tau}(z;z)}{f^2_{Y(0)|Z}(q_{0,\tau}(z)|z)}\right\}\frac{\omega(z)}{f_Z(z)}dz$$

with $\nu_0(K)=\int K^2(u)du$ *and*

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

and under the alternative hypothesis H_1*,*

$$nh^{d/2}\big(J_n-\mu_J\big) \xrightarrow{p} +\infty. \quad \textbf{(5)}$$

Following Theorem 1, an asymptotic significance level α_0 test is to reject H_0 if $nh^{d/2}\big(J_n-\mu_J\big)/\sigma_J > C_{\alpha_0}$, where C_{α_0} is the α_0 upper-quantile of the standard normal distribution. Clearly, (5) implies that the proposed test is consistent. To the best of our knowledge, the above asymptotic result for testing nonparametric QTE is new in the literature.

Remark 3. The testing setting in (2) can be generalized to the following testing problem:

$$H_0:\Delta_\tau(z)=\Delta_{\tau,0}(z,\theta_\tau) \quad \text{versus} \quad H_1:\Delta_\tau(z)\neq\Delta_{\tau,0}(z,\theta_\tau), \quad \textbf{(6)}$$

where $\Delta_{\tau,0}(z,\theta_\tau)$ is a known function with unknown parameter θ_τ. The purpose of the test in (6) is to see whether $\Delta_\tau(z)$ has a particular parametric form, say, a linear function as in Tang et al. (2021). We can extend the proposed test statistic J_n to this case by replacing $\widehat{\delta}_\tau$ with $\Delta_{\tau,0}(z,\widehat{\theta}_{n,\tau})$, where $\widehat{\theta}_{n,\tau}$ is an estimator of θ_τ. It is easy to show that J_n has the same asymptotic null distribution if $\sup_{z\in\mathcal{Z}}|\Delta_\tau(z,\widehat{\theta}_{n,\tau})-\Delta_{\tau,0}(z,\theta_\tau)|=o_p\big(n^{-1/2}h^{-d/4}\big)$.

In addition to testing the null hypothesis against fixed alternatives, it is of interest to consider testing power for local departures from the null. Suppose that δ_τ is estimated using (4). We focus on a set of Pitman alternatives represented by

$$H_{1n}:\Delta_\tau(z)=\delta_\tau+\rho_n\cdot\zeta(z), \quad \textbf{(7)}$$

where $\rho_n=n^{-1/2}h^{-d/4}\to 0$ as $n\to\infty$ and the function $\zeta(z)$ satisfies

$$\int\zeta(z)f_Z(z)dz=0 \quad \text{and} \quad 0<\int\zeta^2(z)\omega(z)dz<\infty.$$

Note that the requirement $\int\zeta(z)f_Z(z)dz=0$, or equivalently, $\int\Delta_\tau(z)f_Z(z)\,dz=\delta_\tau$, for the local alternatives depends on the estimator $\hat{\delta}_\tau$ used in the test statistic J_n. We require $\hat{\delta}_\tau$ to converge to δ_τ with a rate faster than ρ_n under the local alternative

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

H_{1n}. When $\widehat{\delta}_\tau = \frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}_\tau(Z_i)$, we have

$$\begin{aligned}
\widehat{\delta}_\tau - \delta_\tau &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}_\tau(Z_i) - \int \Delta_\tau(z) f_Z(z)\,dz \\
&= \left[\frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}_\tau(Z_i) - \frac{1}{n}\sum_{i=1}^{n}\Delta_\tau(Z_i)\right] + \left[\frac{1}{n}\sum_{i=1}^{n}\Delta_\tau(Z_i) - \int \Delta_\tau(z) f_Z(z)\,dz\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta}_\tau(Z_i) - \frac{1}{n}\sum_{i=1}^{n}\Delta_\tau(Z_i)\right] + O_p(1/\sqrt{n})
\end{aligned}$$

under H_{1n}, which is shown to be $o_p(\rho_n)$ in the proof of Theorem 1. The following theorem shows that our test can distinguish alternatives H_{1n} that get closer to H_0 at rate $n^{-1/2}h^{-d/4}$ while maintaining a constant power level.

THEOREM 2. *Under Assumptions 1–6, suppose that the local alternative (7) converges to the null in the sense that $\rho_n = n^{-1/2}h^{-d/4}$. Then,*

$$\frac{nh^{d/2}}{\sigma_J}\left(J_n - \mu_J\right) \xrightarrow{D} \mathcal{N}\left(\sigma_J^{-1}\int \zeta^2(z)\omega(z)dz, 1\right).$$

Clearly, it follows from Theorem 2 that under the local alternative (7),

$$P\left(nh^{d/2}\left(J_n - \mu_J\right)/\sigma_J > C_{\alpha_0}\right) \to 1 - \Phi\left(C_{\alpha_0} - \sigma_J^{-1}\int \zeta^2(z)\omega(z)dz\right),$$

where Φ is the cumulative distribution function of the standard normal distribution. This indicates that the testing power for local alternative (7) converges to a constant greater than the significance level α_0.

From Theorem 2, we can see that the choice of the weight function $\omega(z)$ would affect the testing power. However, the "optimal" weight function to maximize the local power depends on the unknown function $\zeta(z)$. In practice, there could be many possible options for the weight function. A simple choice is to use constant weight for all $z \in \mathcal{Z}$. Another choice is letting $\omega(z) = \left\{\frac{\mu_{1,\tau}(z;z)}{f^2_{Y(1)|Z}(q_{1,\tau}(z)|z)} + \frac{\mu_{0,\tau}(z;z)}{f^2_{Y(0)|Z}(q_{0,\tau}(z)|z)}\right\}^{-1} f_Z(z)$, which is proportional to the inverse of the asymptotic variance of $\hat{\Delta}_\tau(z)$ given in Cai et al. (2021). Intuitively, such an inverse-variance weight function avoids large statistical uncertainty by assigning a small weight to z, where $\Delta_\tau(z)$ is estimated imprecisely. It can be seen from Theorem 1 that if the inverse-variance weight function is used, the asymptotic null distribution of the test statistic J_n is free of nuisance parameters. However, to employ the inverse-variance weight function, one would first need a consistent estimator of the weight function at a uniform rate of $o_p(h^{d/2})$.

2.3. A Nonparametric Bootstrap Test

Theorem 1 provides the asymptotic null distribution of the test J_n. Consequently, one can perform tests for H_0 by comparing the value of J_n to its asymptotic critical value. However, as expected, Theorem 1 can not be used directly for an accurate calculation of critical values. This is because the test based on the asymptotic distribution might be sensitive to the choice of bandwidth h and the consistent estimation of μ_J and σ_J^2 in small samples. In particular, it is well known in the quantile regression literature that the consistent estimation of the conditional density of $Y_i(j)$ given Z_i involved in μ_J and σ_J^2 is not an easy task (see, for example, Koenker and Xiao, 2004; Cai and Xu, 2008). To overcome this difficulty, following Chen, Linton, and Van Keilegom (2003) and Firpo, Galvao, and Song (2017), a nonparametric Bootstrap procedure is proposed to determine the p-value for J_n. Other types of Bootstrap methods such as the multiplier Bootstrap in Escanciano and Goh (2014) and the wild Bootstrap in Dong et al. (2019) can be used, but we make this choice for simplicity. The method that we use involves the following steps:

(1) Generate the Bootstrap sample by drawing samples from the original sample $\{(Y_i, X_i, D_i)\}_{i=1}^n$ with replacement, denoted by $\{(Y_i^*, X_i^*, D_i^*)\}_{i=1}^n$.
(2) Compute the Bootstrap test statistic

$$J_n^* = \int \left(\left(\widehat{\Delta}_\tau^*(z) - \widehat{\delta}_\tau^* \right) - \left(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau \right) \right)^2 \omega(z) dz,$$

where $\widehat{\Delta}_\tau^*(z) = \widehat{q}_{1,\tau}^*(z) - \widehat{q}_{0,\tau}^*(z)$ and $\widehat{\delta}_\tau^* = \sum_{i=1}^n \widehat{\Delta}_\tau^*(Z_i^*)/n$ are estimated using the Bootstrap sample $\{(Y_i^*, X_i^*, D_i^*)\}_{i=1}^n$, and $\widehat{\Delta}_\tau(z)$ and $\widehat{\delta}_\tau = \sum_{i=1}^n \widehat{\Delta}_\tau(Z_i)/n$ are computed based on the original data.
(3) Repeat steps (1) and (2) a large number of times, say, B times, to obtain $\{J_n^{*(j)}\}_{j=1}^B$.
(4) Reject H_0 at significance level α_0 if J_n exceeds the $(1-\alpha_0)$th sample quantile of $\{J_n^{*(j)}\}_{j=1}^B$.

Define $\widetilde{J}_n^* = nh^{d/2}(J_n^* - \mu_J)\big/\sigma_J$. The following theorem justifies the asymptotic validity of the Bootstrap test with its proof given in the Appendix.

THEOREM 3. *Suppose the same conditions as in Theorem 1 are satisfied. Then under H_0 or H_1, we have*

$$\sup_{y\in\mathbb{R}} \left| P\left(\widetilde{J}_n^* \le y \,\middle|\, \{Y_i, X_i, D_i\}_{i=1}^n\right) - \Phi(y) \right| = o_p(1),$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution.

Theorem 3 states that the Bootstrap statistic $\widetilde{J}_n^* = nh^{d/2}(J_n^* - \mu_J)\big/\sigma_J$ converges to $\mathcal{N}(0,1)$ in distribution in probability. It is important to note that Theorem 3 holds

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

true whether or not the null hypothesis is true. Therefore, when the null hypothesis is true, the Bootstrap test procedure leads to an asymptotically correct size, because conditional on the data, the Bootstrap statistic J_n^* has the same asymptotic distribution as J_n. When the null hypothesis is false, the Bootstrap procedure leads to a consistent test. This is because the test statistic $nh^{d/2}\left(J_n - \mu_J\right)/\sigma_J$ diverges to $+\infty$ as the sample size n goes to infinity, as shown in Theorem 1, whereas the Bootstrap critical value is still finite.

Remark 4. It is worth stressing that Theorems 1–3 are established under the case where the propensity score function is estimated using SLE. However, it is generally recognized that in the presence of many covariates, nonparametric estimation of the propensity score is infeasible in practice due to the curse of dimensionality. To overcome the curse of dimensionality, practitioners commonly adopt flexible parametric model for the unknown propensity score by assuming that

$$p(x) = P(D = 1|X = x) = p(x;\vartheta),$$

where ϑ is a finite-dimensional parameter. In general, the unknown parameter ϑ can be estimated using the maximum likelihood estimate (MLE), denoted by $\widehat{\vartheta}_n$. Consequently, by using similar arguments, it can be shown that Theorems 1–3 hold under Assumptions 1–5 and the condition that $\sup_{x\in\mathcal{X}}\left|p(x;\widehat{\vartheta}_n) - p(x)\right| = O_p(n^{-1/2})$. In particular, in the real data example section, a parametric model is used to estimate the propensity score function because many covariates need to be controlled to satisfy the unconfoundedness condition.

2.4. An Extension to Testing Heterogeneity for a Continuum of Quantile Levels

The procedure developed above can be used to test whether heterogeneity exists in QTEs across different sub-populations defined by covariates of interest for a specific quantile level τ. In some applications, it may be of interest to test whether heterogeneity exists in conditional QTEs for a continuum of quantile levels.[2] To this end, this section extends our testing approach to the case of a continuum of quantile levels. The following hypothesis testing problem is investigated:

$$H_0 : \Delta_\tau(z) = \delta_\tau \text{ for all } z \in \mathcal{Z} \text{ and } \tau \in \mathcal{A} \text{ versus } H_1 : \\ \Delta_\tau(z) \neq \delta_\tau \text{ for some } z \in \mathcal{Z} \text{ or } \tau \in \mathcal{A}, \quad \textbf{(8)}$$

where $\mathcal{A}$ is a compact subset of $(0,1)$. Similarly, to test whether or not the hypothesis testing problem formulated in (8) holds, we construct the test statistic based on the Cramér–von Mises criterion, defined by

$$S_n = \int_{\mathcal{A}}\int \left(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau\right)^2 \omega(z,\tau)dzd\tau,$$

[2]The authors thank one of the anonymous referees for raising this interesting topic.

where $\omega(z,\tau)$ is a pre-specified strictly positive and integrable weighting function, and $\widehat{\Delta}_\tau(z)$ and $\widehat{\delta}_\tau$ are the estimators of $\Delta_\tau(z)$ and δ_τ defined in (3) and (4), respectively.

To present the asymptotic properties of the proposed test statistic S_n, some additional notations are needed. Define

$$\lambda_{1,\tau_1,\tau_2}(z;u,v) = E\Big[\frac{1}{p(X_i)}\varphi_{\tau_1}(Y_i(1);q_{1,\tau_1}(u))\varphi_{\tau_2}(Y_i(1);q_{1,\tau_2}(v))\big|Z_i = z\Big],$$

and

$$\lambda_{0,\tau_1,\tau_2}(z;u,v) = E\Big[\frac{1}{1-p(X_i)}\varphi_{\tau_1}(Y_i(0);q_{0,\tau_1}(u))\varphi_{\tau_2}(Y_i(0);q_{0,\tau_2}(v))\big|Z_i = z\Big],$$

where $\varphi_\tau(y;\theta) = I\{y \le \theta\} - \tau$. The following theorem summarizes the asymptotics of the test statistic S_n.

THEOREM 4. *Suppose that Assumptions 1–6 are satisfied. Then, under the null hypothesis H_0 in (8), one has*

$$nh^{d/2}\big(S_n - \mu_S\big) \xrightarrow{D} \mathcal{N}(0,\sigma_S^2),$$

where

$$\mu_S = \frac{1}{nh^d}\int K^2(s)ds \cdot \int_{\mathcal{A}}\int \left\{\frac{\mu_{1,\tau}(z;z)}{f^2_{Y(1)|Z}(q_{1,\tau}(z)|z)} + \frac{\mu_{0,\tau}(z;z)}{f^2_{Y(0)|Z}(q_{0,\tau}(z)|z)}\right\}\frac{\omega(z,\tau)}{f_Z(z)}dzd\tau,$$

and

$$\sigma_S^2 = 2\int\Big(\int K(t)K(t+s)dt\Big)^2 ds\int_{\mathcal{A}}\int_{\mathcal{A}}\int\left\{\frac{\lambda_{1,\tau_1,\tau_2}(u;u,u)}{f_{Y(1)|Z}(q_{1,\tau_1}(u)|u)f_{Y(1)|Z}(q_{1,\tau_2}(u)|u)}\right.$$
$$\left.+\frac{\lambda_{0,\tau_1,\tau_2}(u;u,u)}{f_{Y(0)|Z}(q_{0,\tau_1}(u)|u)f_{Y(0)|Z}(q_{0,\tau_2}(u)|u)}\right\}^2\frac{\omega(u,\tau_1)\omega(u,\tau_2)}{f_Z^2(u)}dud\tau_1 d\tau_2,$$

and under the alternative hypothesis H_1 in (8),

$$nh^{d/2}\big(S_n - \mu_S\big) \xrightarrow{p} +\infty.$$

To use the test statistic S_n to perform tests for H_0 in (8), one needs to calculate its asymptotic critical value. However, it is difficult to accurately calculate the critical values. Similar to the nonparametric Bootstrap procedure developed in Section 2.3, we propose the following nonparametric Bootstrap procedure to determine the p-value for S_n, which involves the following steps.

(1) Generate the Bootstrap sample by drawing samples from the original sample $\big\{(Y_i,X_i,D_i)\big\}_{i=1}^n$ with replacement, denoted by $\big\{(Y_i^*,X_i^*,D_i^*)\big\}_{i=1}^n$.
(2) Compute the Bootstrap test statistic

$$S_n^* = \int_{\mathcal{A}}\int\Big(\big(\widehat{\Delta}_\tau^*(z) - \widehat{\delta}_\tau^*\big) - \big(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau\big)\Big)^2\omega(z,\tau)dzd\tau,$$

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

where $\widehat{\Delta}_\tau^*(z)$ and $\widehat{\delta}_\tau^* = \sum_{i=1}^n \widehat{\Delta}_\tau^*(Z_i^*)/n$ are estimated using the Bootstrap sample $\left\{(Y_i^*, X_i^*, D_i^*)\right\}_{i=1}^n$, and $\widehat{\Delta}_\tau(z)$ and $\widehat{\delta}_\tau = \sum_{i=1}^n \widehat{\Delta}_\tau(Z_i)/n$ are computed based on the original data.

(3) Repeat steps (1) and (2) B times to obtain $\{S_n^{*(j)}\}_{j=1}^B$.

(4) Reject H_0 at significance level α_0 if S_n exceeds the $(1-\alpha_0)$th sample quantile of $\{S_n^{*(j)}\}_{j=1}^B$.

3. MONTE CARLO STUDIES

In this section, we investigate the finite sample performance of the proposed tests J_n and S_n by means of simulation studies. The goal is to assess the size and power of the proposed test for moderate sample sizes in various scenarios. The scenarios examined differ in the order of the power series used in estimating $p(x)$ and the choice of the bandwidth parameter h. In these experiments, we simply use the constant weight function $\omega(z) \equiv 1$ or $\omega(z,\tau) \equiv 1$ in the test statistics J_n or S_n, respectively.

Let the data generating process (DGP) be:

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

where $\rho_n = n^{-1/2} h^{-1/4}$, $\gamma_0 = 1.0$, $\gamma_1 = 1.5$, U_0 and U_1 independently follow the $U[0,1]$ distribution, X_1 and X_2 are independently generated from $U[-1,1]$ and Beta$(3,1)$, respectively, and the propensity score function is

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

Finally, the conditional variable Z is taken to be X_1. Under this setting, by straightforward calculations, the conditional quantile function for $Y(j)$ for $j=0$ and 1, conditional on $Z=z$, is given by $q_{0,\tau}(z) = \gamma_0 a_\tau$ and $q_{1,\tau}(z) = \lambda \rho_n z + \gamma_1 a_\tau$, respectively, where a_τ is the unique solution of equation $-2a^3+3a^2-\tau=0$ within the interval $(0,1)$. Therefore, the PCQTE is

$$\Delta_\tau(z) = \lambda \rho_n z + (\gamma_1 - \gamma_0) a_\tau,$$

where λ in the above equation takes different values in the experiment so that we can investigate empirical sizes and local power curves of the test statistic J_n indexed by λ. It is easy to see that $\Delta_\tau(z)$ is equal to a constant only when $\lambda = 0$, which corresponds to the null hypothesis. The Bootstrap procedure outlined in Sections 2.3 is used to determine the critical values. The number of Bootstrap replications is set as $B = 599$. To examine the size and local power performance of the test statistic J_n, three different sample sizes $n=400$, $n=800$ and $n=1,600$ are considered. To check the sensitivity of the test with respect to different values of the bandwidth h, motivated by Assumption 1, $h = cn^{-1/3}$ is used with $c=0.5$, 1.0, and 2.0. Linear and quadratic power series are used in the SLE to estimate $p(x)$. It should be noted that the model for the propensity score $p(x)$ is correctly specified when either a linear power series or a quadratic power series is used in the SLE, except that the

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

Table 1. Empirical sizes of J_n and S_n (nominal size $\alpha = 5\%$, linear power series specification for estimating $p(x)$).

		$h = 0.5n^{-1/3}$			
$\lambda = 0$	n	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau \in [0.1, 0.9]$
	400	0.030	0.034	0.031	0.028
	800	0.040	0.037	0.045	0.035
	1,600	0.047	0.053	0.044	0.043
		$h = 1.0n^{-1/3}$			
$\lambda = 0$	n	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau \in [0.1, 0.9]$
	400	0.043	0.043	0.045	0.041
	800	0.058	0.052	0.055	0.056
	1,600	0.044	0.053	0.053	0.052
		$h = 2.0n^{-1/3}$			
$\lambda = 0$	n	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau \in [0.1, 0.9]$
	400	0.034	0.036	0.038	0.035
	800	0.056	0.056	0.054	0.042
	1,600	0.053	0.051	0.053	0.054

quadratic power series contains redundant terms. Finally, three quantiles levels, namely, $\tau = 0.25$, 0.5 and $\tau = 0.75$, are considered. The empirical sizes and local powers of the test J_n are computed using 1,000 simulations under the nominal size $\alpha = 5\%$, respectively.

The empirical sizes of the test J_n based on Bootstrap critical values with $p(x)$ estimated using linear power series and quadratic power series are reported in Tables 1 and 2, respectively. It can be seen from Tables 1 and 2 that the empirical sizes converge to their nominal sizes as the sample size n increases. In particular, when the sample size increases to 1,600, the test J_n performs well in all cases considered. Also, one can observe that the choice of the bandwidth h and the power series seems to have little influence on empirical sizes. In addition, the empirical sizes of the test S_n based on Bootstrap critical values for $\tau \in [0.1, 0.9]$ are presented in the last column in Tables 1 and 2, respectively. Similarly, the empirical sizes converge to their nominal sizes as the sample size n increases; the choice of the bandwidth h and the power series seems to have little influence on empirical sizes.

Next, Figures 1–6 display the estimated local power curves with nominal size $\alpha = 5\%$ of the test J_n for different quantile levels and different choices of the bandwidth and the power series. In general, the test J_n is reasonably powerful in detecting deviation from the null hypothesis in all cases considered. Specifically, it can be seen from these figures that the test J_n has power against local alternatives converging to the null at the rate of $\rho_n = n^{-1/2}h^{-1/4}$. Moreover, it is not surprising that the power increases quickly with the value of λ increasing. These figures also

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

TABLE 2. Empirical sizes of J_n and S_n (nominal size $\alpha = 5\%$, quadratic power series specification for estimating $p(x)$).

		$h = 0.5n^{-1/3}$			
$\lambda = 0$	n	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau \in [0.1, 0.9]$
	400	0.030	0.032	0.030	0.028
	800	0.035	0.037	0.035	0.034
	1,600	0.045	0.046	0.042	0.041
		$h = 1.0n^{-1/3}$			
$\lambda = 0$	n	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau \in [0.1, 0.9]$
	400	0.038	0.040	0.035	0.032
	800	0.055	0.056	0.057	0.043
	1,600	0.053	0.052	0.054	0.054
		$h = 2.0n^{-1/3}$			
$\lambda = 0$	n	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau \in [0.1, 0.9]$
	400	0.037	0.038	0.032	0.033
	800	0.057	0.058	0.060	0.055
	1,600	0.052	0.052	0.054	0.053

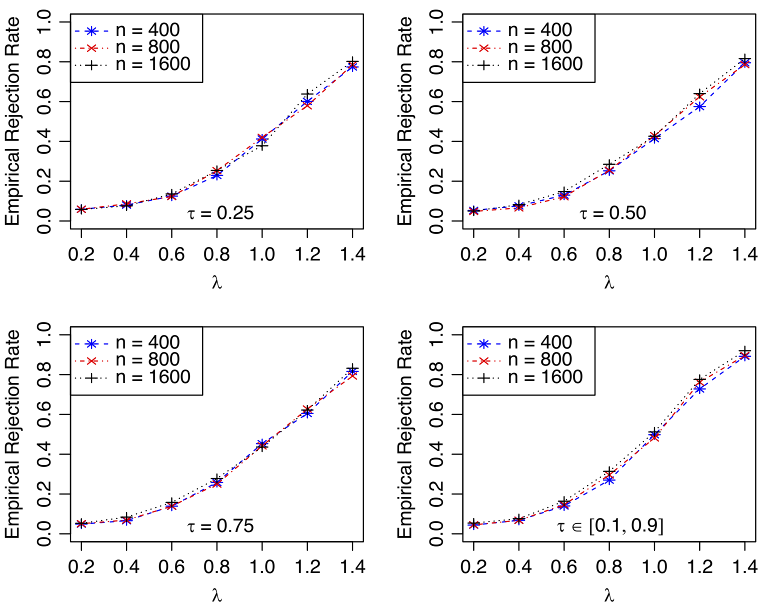

FIGURE 1. Local power curves for test statistic J_n and S_n (lower right panel) with nominal size $\alpha = 5\%$, $\rho_n = n^{-1/2}h^{-1/4}$, bandwidth $h = 0.5n^{-1/3}$ and linear power series specification for estimating $p(x)$.

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

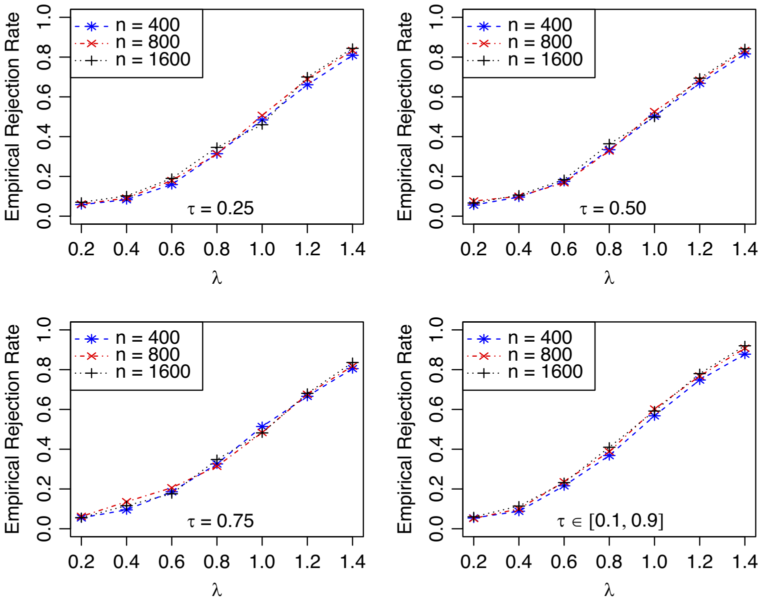

FIGURE 2. Local power curves for test statistic J_n and S_n (lower right panel) with nominal size $\alpha = 5\%$, $\rho_n = n^{-1/2}h^{-1/4}$, bandwidth $h = n^{-1/3}$ and linear power series specification for estimating $p(x)$.

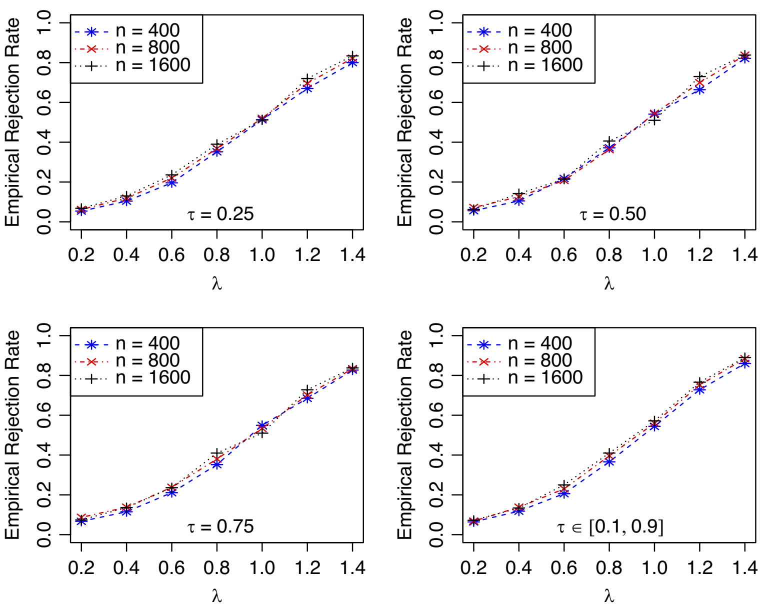

FIGURE 3. Local power curves for test statistic J_n and S_n (lower right panel) with nominal size $\alpha = 5\%$, $\rho_n = n^{-1/2}h^{-1/4}$, bandwidth $h = 2n^{-1/3}$ and linear power series specification for estimating $p(x)$.

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

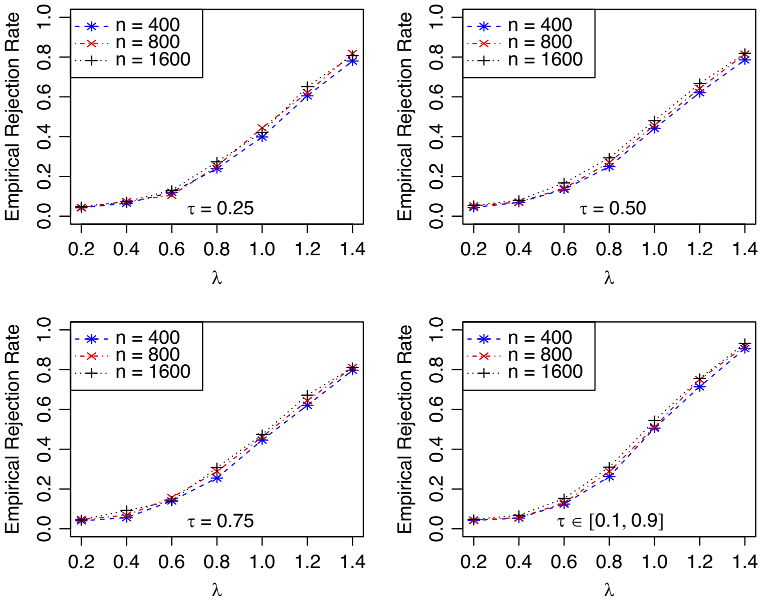

FIGURE 4. Local power curves for test statistic J_n and S_n (lower right panel) with nominal size $\alpha = 5\%$, $\rho_n = n^{-1/2}h^{-1/4}$, bandwidth $h = 0.5n^{-1/3}$ and quadratic power series specification for estimating $p(x)$.

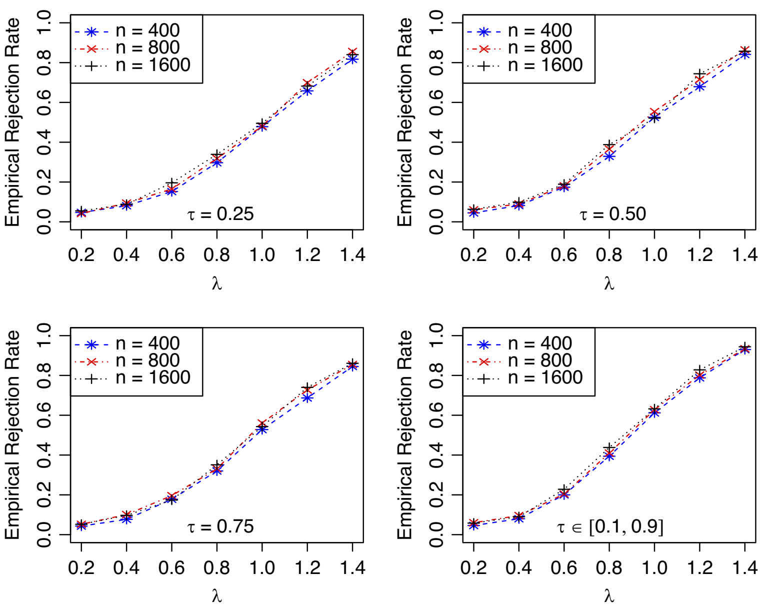

FIGURE 5. Local power curves for test statistic J_n and S_n (lower right panel) with nominal size $\alpha = 5\%$, $\rho_n = n^{-1/2}h^{-1/4}$, bandwidth $h = n^{-1/3}$ and quadratic power series specification for estimating $p(x)$.

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

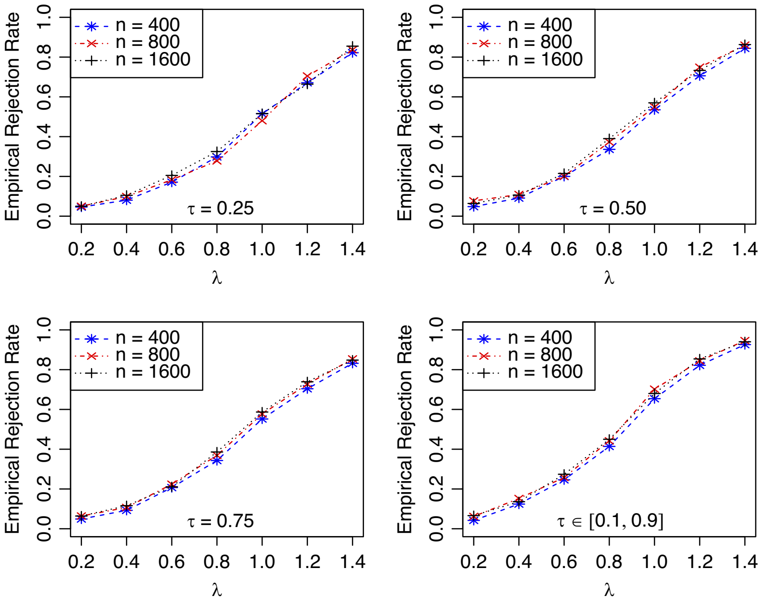

FIGURE 6. Local power curves for test statistic J_n and S_n (lower right panel) with nominal size $\alpha = 5\%$, $\rho_n = n^{-1/2}h^{-1/4}$, bandwidth $h = 2n^{-1/3}$ and quadratic power series specification for estimating $p(x)$.

show that the bandwidth h in a certain range and the power series considered seem to have little impact on the power of the test. Finally, the estimated local power curves of the test S_n for $\tau \in [0.1, 0.9]$ and different choices of the bandwidth and the power series are presented in the lower right panels of Figures 1–6. Similar conclusions can be made from these figures.

4. A REAL DATA EXAMPLE

The proposed testing approach is applied to investigate whether heterogeneity exists for the QTE of maternal smoking during pregnancy on infant birth weight across different age groups of mothers. To this end, we use the same dataset as in Abrevaya et al. (2015), composed of vital statistics collected by the North Carolina State Center Health Services between 1988 and 2002, 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; as is routine in the literature, we treat blacks and whites as separate populations throughout. The number of observations is 157,989 for the black group and 433,558 for the white group.

Low infant birth weight is associated with poor outcomes in health and human capital development throughout life (Black, Devereux, and Salvanes, 2007; Almond and Currie, 2011). Maternal smoking during pregnancy is considered to be the most important preventable cause of low birth weight (see Kramer, 1987

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

for more discussion). Recently, there have been several studies using program evaluation approaches to explore how the effect of maternal smoking during pregnancy on infant birth weight varies across the mothers' ages. In particular, Abrevaya et al. (2015) and Lee et al. (2017) considered the ATE of maternal smoking on infant birth weight conditional on different mothers' ages, and found different degrees of heterogeneity by age. The main qualitative finding in Abrevaya et al. (2015) and Lee et al. (2017) is that smoking has a more severe impact at higher ages. In contrast to Abrevaya et al. (2015) and Lee et al. (2017), Cai et al. (2021) considered the QTE of mothers' smoking status during pregnancy on infant birth weight conditional on different mothers' ages, and found that the QTEs for the quantile levels considered seemed to change significantly over age for whites but not for blacks. Motivated by the estimation results in Cai et al. (2021), it is interesting to test whether or not the partially conditional QTEs, for whites and blacks, change over mothers' ages. Therefore, in this section, we test whether the QTE of maternal smoking on infant birth weight varies across different age groups of mothers, using the testing approach proposed in Section 2.

Here, the conditional variable Z is the mother's age. In addition, the treatment variable D is a binary variable which takes the value 1 if the mother smokes, and 0 otherwise. The outcome variable of interest Y is the infant birth weight measured in grams. Also, in this example, $Y(0)$ denotes the infant birth weight for the untreated (non-smoking) group and $Y(1)$ stands for the infant birth weight for the treated (smoking) group.

To explore the treatment effect heterogeneity of mother's smoking on infant birth weight using the approach in Section 2, one needs to find certain baseline covariates such that the unconfoundedness assumption holds true, that is, the potential infant birth weight outcomes are independent of the smoking decision conditional on the baseline covariates. In this paper, we use the same set of covariates X as in Abrevaya et al. (2015), which includes the mother's age, education, month of first prenatal visit, number of prenatal visits, and indicators for the baby's gender, the mother's marital status, whether 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 a detailed discussion).

Another problem when using our approach is how to estimate the unknown propensity score function $p(x)$. Following Abrevaya et al. (2015) and Cai et al. (2021), we use a logistic model to estimate the propensity score function $p(x)$. The explanatory variables used in the logistic 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. Finally, the partially conditional QTE is estimated for mothers between 20 and 30 years of age, for both whites and blacks. Figure 7 plots the estimated PCQTEs for whites and blacks for three quantile levels $\tau = 0.10$, $\tau = 0.25$, and $\tau = 0.50$, together with the estimated unconditional 0.5-QTEs and their 95% confidence intervals. It should be noted that the 0.5-QTE could be different from $\delta_{0.5}$ in the testing problem (2). It can be observed that, as the mother's

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

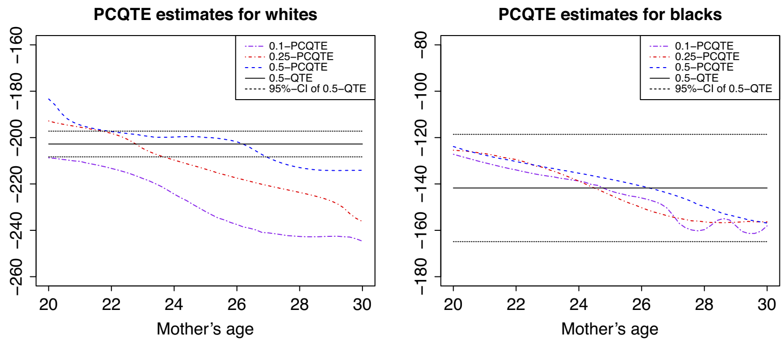

FIGURE 7. Estimated PCQTEs for whites and blacks for three quantile levels $\tau = 0.10$, $\tau = 0.25$, and $\tau = 0.50$, together with the estimated unconditional 0.5-QTEs and their 95% confidence intervals.

TABLE 3. Test results for testing if PCQTE function changes over mother's age.

Quantile level	Test statistic (J_n or S_n) Bootstrap p-values	
τ	Whites	Blacks
0.10	0.022	0.573
0.25	0.002	0.307
0.50	0.033	0.151
0.75	0.035	0.474
0.90	0.042	0.326
[0.1, 0.9]	0.044	0.357

age increases, the PCQTEs of maternal smoking on infant birth weight decrease quickly for whites, while the PCQTEs for blacks decrease slowly, compared to the width of the 95% confidence intervals of the corresponding unconditional 0.5-QTEs.

Table 3 displays the results of testing whether the partially conditional QTE changes over mother's age. Table 3 clearly shows that one should reject the null hypothesis for whites for all quantiles considered at 5%. This means that PCQTEs do change over mother's age for all quantile levels considered at the significance level $\alpha = 5\%$ for whites. However, for blacks, the change in the PCQTE over mother's age is slight and statistically insignificant for all quantile levels. These testing results support the empirical findings in Cai et al. (2021). As pointed out by Yang et al. (2014), this phenomenon may occur through various interconnected mechanisms. Blacks are exposed to an environment with many risk factors for smoking, such as job loss and economic hardship. Finally, we use the proposed test statistic S_n to test whether the partially conditional QTE changes over mother's age for $\tau \in [0.1, 0.9]$. The testing results are displayed in the last row in Table 3. The same conclusion can be made.

5. CONCLUSION

Motivated by the question of whether heterogeneity exists in the QTEs conditional on mothers' ages for both whites and blacks, we propose a nonparametric versus constant test, which is applied to assess whether there exist heterogeneously distributional effects for an intervention on an outcome of interest across different sub-populations defined by covariates of interest. To test whether the null hypotheses of interest hold true, a consistent test statistic is proposed based on the Cramér–von Mises type criterion. To the best of our knowledge, this test is novel in the QTE literature. Under some regularity conditions, we establish the asymptotic distribution of the proposed test statistic under the null hypothesis and its consistency against fixed alternatives. We also study the power of our test against a sequence of local alternatives. Moreover, we propose a Bootstrap procedure to approximate the finite-sample null distribution of the proposed test statistic. Furthermore, the asymptotic validity of the proposed Bootstrap test is established.

Finally, some extensions of our paper can be considered. The first is the case where the number of controlling variables is large. For this case, the estimated propensity score converges at a slower rate, even when a penalized method is used in estimation. The so-called double machine learning method developed by Belloni et al. (2017) may be used to estimate the PCQTE and establish asymptotic properties of the test statistic. Second, one might be interested in extending our results to time series cases, which have potential in a wide range of applications. Such extensions can be warranted as future research.

APPENDIX. MATHEMATICAL PROOFS

Note that this appendix provides some key steps for proving Theorems 1–4. Some auxiliary lemmas with their detailed proofs, as well as some notations, are given in the Supplementary Material.

Proof of Theorem 1. Let $\Delta_\tau(z)$ be the partially conditional QTE conditional on $Z_i = z$. Define $\bar{\delta}_\tau = \int \Delta_\tau(z) f_Z(z)\, dz$. Then,

$$J_n = \int \left(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau\right)^2 \omega(z)dz = \int \left[\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right) + \left(\bar{\delta}_\tau - \widehat{\delta}_\tau\right) + \left(\Delta_\tau(z) - \bar{\delta}_\tau\right)\right]^2 \omega(z)dz,$$

where $\widehat{\delta}_\tau = \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau(Z_i)$. We first claim that $\widehat{\delta}_\tau - \bar{\delta}_\tau = O_p\big(e_n\big)$, where $e_n = \max\left\{\frac{\ln n}{n^{1/2-3\nu/2}}, \frac{\sqrt{\ln n}}{(nh^d)^{3/4}}\right\}$. It is easy to obtain from Lemma 4 that

$$\widehat{\delta}_\tau - \bar{\delta}_\tau$$
$$= \left[\frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau(Z_i) - \frac{1}{n}\sum_{i=1}^n \Delta_\tau(Z_i)\right] + \left[\frac{1}{n}\sum_{i=1}^n \Delta_\tau(Z_i) - \int \Delta_\tau(z) f_Z(z)\, dz\right]$$

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{n}\sum_{j=1}^{n}\gamma_{n,\tau}(Y_j,X_j,D_j;Z_i)\right\} + O_p\big(e_n\big) + O_p(1/\sqrt{n})$$

$$= \frac{1}{n^2}\sum_{1\le i<j\le n}\big[\gamma_{n,\tau}(Y_j,X_j,D_j;Z_i) + \gamma_{n,\tau}(Y_i,X_i,D_i;Z_j)\big]$$

$$+ \frac{1}{n^2}\sum_{i=1}^{n}\gamma_{n,\tau}(Y_i,X_i,D_i;Z_i) + O_p\big(e_n\big) + O_p(1/\sqrt{n})$$

$$= \frac{1}{n^2}\sum_{1\le i<j\le n}\big[\gamma_{n,\tau}(Y_j,X_j,D_j;Z_i) + \gamma_{n,\tau}(Y_i,X_i,D_i;Z_j)\big] + O_p\big(e_n\big),$$

where $\gamma_{n,\tau}(Y_i,X_i,D_i;z) = \varrho_{n,1,\tau}(Y_i,X_i,D_i;z) - E\varrho_{n,1,\tau}(Y_i,X_i,D_i;z) - \varrho_{n,0,\tau}(Y_i,X_i,D_i;z) + E\varrho_{n,0,\tau}(Y_i,X_i,D_i;z)$. Here $\varrho_{n,l,\tau}(Y_i,X_i,D_i;z)$, $l = 0,1$, are defined in Lemma 2. Note that the first term in the above equation is a centered U-statistic. Using Lemma 5, we can verify that

$$\frac{1}{n^2}\sum_{1\le i<j\le n}\big[\gamma_{n,\tau}(Y_j,X_j,D_j;Z_i) + \gamma_{n,\tau}(Y_i,X_i,D_i;Z_j)\big] = O_p(1/\sqrt{n}).$$

Thus,

$$\widehat{\delta}_\tau - \bar{\delta}_\tau = O_p(1/\sqrt{n}) + O_p\big(e_n\big) = O_p\big(e_n\big). \qquad \textbf{(9)}$$

Then, under the null hypothesis H_0, $\Delta_\tau(z) - \bar{\delta}_\tau \equiv 0$. By Lemma 4,

$$J_n = \int\Big(\frac{1}{n}\sum_{i=1}^{n}\gamma_{n,\tau}(Y_i,X_i,D_i;z) + O_p(e_n)\Big)^2\omega(z)dz$$

$$= \int\Big(\frac{1}{n}\sum_{i=1}^{n}\gamma_{n,\tau}(Y_i,X_i,D_i;z)\Big)^2\omega(z)dz + O_p(e_n^2)\int\omega(z)dz$$

$$+ 2O_p(e_n)\int\frac{1}{n}\sum_{i=1}^{n}\gamma_{n,\tau}(Y_i,X_i,D_i;z)\omega(z)dz$$

$$:= J_{n,1} + J_{n,2} + J_{n,3}.$$

Note that under Assumptions 5 and 6, $e_n = o\big(n^{-1/2}h^{-d/4}\big)$. It is easy to verify that $nh^{d/2}J_{n,2} = o_p(1)$. Also, by noting that $E\big(\gamma_{n,\tau}(Y_i,X_i,D_i;z)\big) = 0$, we have

$$nh^{d/2}J_{n,3} = nh^{d/2}\cdot O_p(e_n)\cdot\frac{1}{n}\sum_{i=1}^{n}\int\gamma_{n,\tau}(Y_i,X_i,D_i;z)\omega(z)dz$$

$$= nh^{d/2}\cdot O_p(e_n)\cdot O_p(n^{-1/2}) = o_p(1).$$

Thus, an application of Lemma 6 leads to

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

Now, we consider the case under the alternative hypothesis H_1. Under H_1, it is easy to show that $J_n - \mu_J = \int\big(\Delta_\tau(z) - \bar{\delta}_\tau\big)^2\omega(z)dz + o_p(1)$. Since $\int\big(\Delta_\tau(z) - \bar{\delta}_\tau\big)^2\omega(z)dz$ is a

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

positive constant under H_1, so that

$$nh^{d/2}\left(J_n - \mu_J\right) \xrightarrow{p} +\infty.$$

This completes the proof of Theorem 1. □

Proof of Theorem 2. Under the local alternative $H_{1n} : \Delta_\tau(z) = \delta_\tau + \rho_n \cdot \zeta(z)$ with $\rho_n = n^{-1/2}h^{-d/4}$, it is easy to see that $\bar{\delta} := \int \Delta_\tau(z) f_Z(z)\, dz = \delta_\tau$ since $\int \zeta(z) f_Z(z)\, dz = 0$. According to (9), we have

$$\begin{aligned}
J_n &= \int \left(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau\right)^2 \omega(z)dz \\
&= \int \left[\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right) + \left(\Delta_\tau(z) - \delta_\tau\right) + \left(\delta_\tau - \widehat{\delta}_\tau\right)\right]^2 \omega(z)dz \\
&= \int \left[\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right) + \left(\Delta_\tau(z) - \delta_\tau\right) + \left(\bar{\delta}_\tau - \widehat{\delta}_\tau\right)\right]^2 \omega(z)dz \\
&= \int \left[\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right) + \rho_n \cdot \zeta(z) + O_p(e_n)\right]^2 \omega(z)dz \\
&= \int \left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right)^2 \omega(z)dz + \rho_n^2 \cdot \int \zeta^2(z)\omega(z)dz + O_p\left(e_n^2\right) \\
&\quad + 2\rho_n \cdot \int \zeta(z)\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right)\omega(z)dz + 2O_p\left(e_n\right) \cdot \int \left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right)\omega(z)dz \\
&\quad + 2O_p\left(e_n\right) \cdot \rho_n \cdot \int \zeta(z)\omega(z)dz \\
&:= J_n^{(1)} + J_n^{(2)} + J_n^{(3)} + J_n^{(4)} + J_n^{(5)} + J_n^{(6)}.
\end{aligned}$$

By noting that $\rho_n = n^{-1/2}h^{-d/4}$,

$$\begin{aligned}
\int \left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right)\omega(z)dz &= \int \frac{1}{n}\sum_{i=1}^{n} \gamma_{n,\tau}(Y_i, X_i, D_i; z)\omega(z)dz + O_p\left(e_n\right) \\
&= O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(e_n\right),
\end{aligned}$$

and

$$\int \zeta(z)\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right)\omega(z)dz = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(e_n\right),$$

it is easy to show that $nh^{d/2}J_n^{(k)} = o_p(1)$ for $3 \le k \le 6$. Then, using the result in Theorem 1, we have

$$nh^{d/2}\left(J_n - \mu_J\right) = nh^{d/2}\left(J_n^{(1)} - \mu_J\right) + nh^{d/2}J_n^{(2)} + o_p(1) \xrightarrow{D} \mathcal{N}\left(\int \zeta^2(z)\omega(z)dz, \sigma_J^2\right).$$

This completes the proof of Theorem 2. □

Now, we introduce some notations before considering the proof of Theorem 3. First, let P denote the distribution of $\{(Y_i(0), Y_i(1), X_i, D_i)\}_{i=1}^n$, and let P^* denote the Bootstrap distribution, which is the distribution of $\{(Y_i^*, X_i^*, D_i^*)\}_{i=1}^n$, conditional on $\{(Y_i, X_i, D_i)\}_{i=1}^n$. Also, we use E^* and Var^* to denote the expectation and variance with respect to P^*, respectively. Finally, following Lee et al. (2015), let $S_1, S_2, \ldots$ be a sequence of random

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

variables and $a_1, a_2, \dots$ be a sequence of positive real numbers. Define $S_n = o_{p^*}(a_n)$ if, for any $\varepsilon > 0$ and $\epsilon > 0$, $\lim_{n\to\infty} P\{P^*(|S_n/a_n| > \epsilon) > \varepsilon\} = 0$. Similarly, $S_n = O_{p^*}(a_n)$ means that, for any $\varepsilon > 0$ and $\epsilon > 0$, there exists $M > 0$ such that $\limsup_{n\to\infty} P\{P^*(|S_n/a_n| > M) > \varepsilon\} < \epsilon$.

Proof of Theorem 3. It is easy to see that we have the following decomposition:

$$\begin{aligned}
J_n^* &= \int \Big(\big(\widehat{\Delta}_\tau^*(z) - \widehat{\delta}_\tau^*\big) - \big(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau\big)\Big)^2 \omega(z) dz \\
&= \int \Big(\widehat{\Delta}_\tau^*(z) - \widehat{\Delta}_\tau(z)\Big)^2 \omega(z) dz + \int \big(\widehat{\delta}_\tau^* - \widehat{\delta}_\tau\big)^2 \omega(z) dz \\
&\quad -2\int \big(\widehat{\delta}_\tau^* - \widehat{\delta}_\tau\big)\big(\widehat{\Delta}_\tau^*(z) - \widehat{\Delta}_\tau(z)\big)\omega(z) dz \\
&= \int \Big(\widehat{\Delta}_\tau^*(z) - \widehat{\Delta}_\tau(z)\Big)^2 \omega(z) dz + O_{p^*}(e_n^2) - 2O_{p^*}(e_n) \int \big(\widehat{\Delta}_\tau^*(z) - \widehat{\Delta}_\tau(z)\big)\omega(z) dz \\
&:= \mathcal{Q}_{n,1} + \mathcal{Q}_{n,2} + \mathcal{Q}_{n,3},
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\delta}_\tau^* - \widehat{\delta}_\tau &= \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau^*(Z_i^*) - \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau(Z_i) \\
&= \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau^*(Z_i^*) - \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau(Z_i^*) + \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau(Z_i^*) - \frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_\tau(Z_i) \\
&= O_{p^*}(e_n) + O_{p^*}(1/\sqrt{n}) = O_{p^*}(e_n)
\end{aligned}$$

can be proved using Lemma 9 and similar arguments as in the proof of Theorem 1.

For the term $\mathcal{Q}_{n,1}$, we have

$$\begin{aligned}
\mathcal{Q}_{n,1} &= \int \Big(\widehat{\Delta}_\tau^*(z) - \widehat{\Delta}_\tau(z)\Big)^2 \omega(z) dz \\
&= \int \left(\frac{1}{n}\sum_{i=1}^n \big(\psi_{n,1,\tau}(Y_i^*, X_i^*, D_i^*; z) - \psi_{n,0,\tau}(Y_i^*, X_i^*, D_i^*; z)\big)\right)^2 \omega(z) dz + O_{p^*}(e_n^2) \\
&\quad + 2O_{p^*}(e_n) \cdot \int \left[\frac{1}{n}\sum_{i=1}^n \Big(\psi_{n,1,\tau}(Y_i^*, X_i^*, D_i^*; z) - \psi_{n,0,\tau}(Y_i^*, X_i^*, D_i^*; z)\Big)\right] \omega(z) dz \\
&:= \mathcal{Q}_{n,1}^{(1)} + \mathcal{Q}_{n,1}^{(2)} + \mathcal{Q}_{n,1}^{(3)},
\end{aligned}$$

where $\psi_{n,l,\tau}(Y_i^*, X_i^*, D_i^*; z)$, $l = 0, 1$, are defined in Lemma 10. Recall that $e_n = \max\left\{\frac{\ln n}{n^{1/2-3\nu/2}}, \frac{\sqrt{\ln n}}{(nh^d)^{3/4}}\right\}$, it is easy to see that $nh^{d/2}\mathcal{Q}_{n,1}^{(2)} = o_{p^*}(1)$. Also, according to the proof of Lemma 11, we know that $nh^{d/2}\mathcal{Q}_{n,1}^{(3)} = o_{p^*}(1)$. Therefore, an application of Lemma 10 leads to

$$nh^{d/2}(\mathcal{Q}_{n,1} - \mu_J)/\sigma_J = nh^{d/2}(\mathcal{Q}_{n,1}^{(1)} - \mu_J)/\sigma_J + o_{p^*}(1) \longrightarrow \mathcal{N}(0,1)$$

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

in distribution in probability. For the term $\mathcal{Q}_{n,2}$, it is easy to see that $nh^{d/2}\mathcal{Q}_{n,2} = o_{p^*}(1)$. By using Lemma 11 again, we also know that $nh^{d/2}\mathcal{Q}_{n,3} = o_{p^*}(1)$. Finally, we have

$$nh^{d/2}(J_n^* - \mu_J)/\sigma_J = nh^{d/2}(\mathcal{Q}_{n,1} - \mu_J)/\sigma_J + o_{p^*}(1) \longrightarrow \mathcal{N}(0,1)$$

in distribution in probability. Because the cumulative distribution function of $\mathcal{N}(0,1)$ is continuous, by Polyā's theorem in Bhattacharya and Rao (1986), we obtain Theorem 3. □

Proof of Theorem 4. We have

$$\begin{aligned} S_n &= \int_{\mathcal{A}} \int \left(\widehat{\Delta}_\tau(z) - \widehat{\delta}_\tau\right)^2 \omega(z,\tau)dzd\tau \\ &= \int_{\mathcal{A}} \int \left[\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right) + \left(\bar{\delta}_\tau - \widehat{\delta}_\tau\right) + \left(\Delta_\tau(z) - \bar{\delta}_\tau\right)\right]^2 \omega(z,\tau)dzd\tau, \end{aligned}$$

where $\hat{\delta}_\tau = \frac{1}{n}\sum_{i=1}^n \hat{\Delta}_\tau(Z_i)$ and $\bar{\delta}_\tau = \int \Delta_\tau(z) f_Z(z)\, dz$. According to Lemma 4 and the proof in Theorem 1, we know that $\widehat{\delta}_\tau - \bar{\delta}_\tau = O_p(e_n)$ uniformly for $\tau \in \mathcal{A}$, where $e_n = \max\left\{\frac{\ln n}{n^{1/2-3\nu/2}}, \frac{\sqrt{\ln n}}{(nh^d)^{3/4}}\right\}$. Thus,

$$\begin{aligned} S_n &= \int_{\mathcal{A}} \int \left[\left(\widehat{\Delta}_\tau(z) - \Delta_\tau(z)\right) + \left(\Delta_\tau(z) - \bar{\delta}_\tau\right) + O_p(e_n)\right]^2 \omega(z,\tau)dzd\tau \\ &= \int_{\mathcal{A}} \int \left[\frac{1}{n}\sum_{i=1}^n \gamma_{n,\tau}(R_i;z) + \left(\Delta_\tau(z) - \bar{\delta}_\tau\right) + O_p(e_n)\right]^2 \omega(z,\tau)dzd\tau. \end{aligned}$$

Then, under the null hypothesis $H_0: \Delta_\tau(z) - \bar{\delta}_\tau \equiv 0$, we have

$$\begin{aligned} S_n &= \int_{\mathcal{A}} \int \left(\frac{1}{n}\sum_{i=1}^n \gamma_{n,\tau}(R_i;z) + O_p(e_n)\right)^2 \omega(z,\tau)dzd\tau \\ &= \int_{\mathcal{A}} \int \left(\frac{1}{n}\sum_{i=1}^n \gamma_{n,\tau}(R_i;z)\right)^2 \omega(z,\tau)dzd\tau + O_p(e_n^2) \int_{\mathcal{A}} \int \omega(z,\tau)dzd\tau \\ &\quad + 2O_p(e_n) \int_{\mathcal{A}} \int \frac{1}{n}\sum_{i=1}^n \gamma_{n,\tau}(R_i;z)\omega(z,\tau)dzd\tau \\ &:= S_{n,1} + S_{n,2} + S_{n,3}. \end{aligned}$$

It is easy to verify that $nh^{d/2}S_{n,2} = o_p(1)$ under Assumptions 5 and 6. Also, by noting that $E\big(\gamma_{n,\tau}(Y_i,X_i,D_i;z)\big) = 0$, we have

$$\begin{aligned} nh^{d/2}S_{n,3} &= nh^{d/2} \cdot O_p(e_n) \cdot \frac{1}{n}\sum_{i=1}^n \int_{\mathcal{A}} \int \gamma_{n,\tau}(R_i;z)\omega(z,\tau)dzd\tau \\ &= nh^{d/2} \cdot O_p(e_n) \cdot O_p(n^{-1/2}) = o_p(1). \end{aligned}$$

Thus, an application of Lemma 12 leads to

$$nh^{d/2}\big(S_n - \mu_S\big) = nh^{d/2}\big(S_{n,1} - \mu_S + S_{n,2} + S_{n,3}\big) \xrightarrow{D} \mathcal{N}(0,\sigma_S^2).$$

Under the alternative hypothesis H_1, it is easy to show that $S_n - \mu_S = \int_{\mathcal{A}}\int \big(\Delta_\tau(z) - \bar{\delta}_\tau\big)^2 \omega(z)dzd\tau + o_p(1)$. Because $\int_{\mathcal{A}}\int \big(\Delta_\tau(z) - \bar{\delta}_\tau\big)^2\omega(z)dzd\tau$ is a positive constant under

https://doi.org/10.1017/S0266466624000045 Published online by Cambridge University Press

H_1, then

$$nh^{d/2}(S_n - \mu_S) \xrightarrow{p} +\infty.$$

This completes the proof of Theorem 4. □

SUPPLEMENTARY MATERIAL

Zongwu Cai, Ying Fang, Ming Lin, and Shengfang Tang (2024): Supplement to "A Nonparametric Test of Heterogeneity in Conditional Quantile Treatment Effects," Econometric Theory Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466624000045.

REFERENCES

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

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

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

Almond, D., & Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25(3), 153–172.

Belloni, A., Chernozhukov, V., Fernàndez-Val, I., & Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1), 233–298.

Bhattacharya, R. N., & Rao, R. R. (1986). *Normal approximation and asymptotic expansions*. Wiley.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *Quarterly Journal of Economics*, 122(1), 409–439.

Cai, Z., Fang, Y., Lin, M., & Tang, S. (2021). Estimation of partially conditional quantile treatment effect model. *China Journal of Econometrics*, 1(4), 741–762.

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

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

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

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

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1), 187–199.

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

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

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

Dong, C., Li, G., & Feng, X. (2019). Lack-of-fit tests for quantile regression models. *Journal of the Royal Statistical Society, Series B*, 81(3), 629–648.

Escanciano, J., & Goh, S. (2014). Specification analysis of linear quantile models. *Journal of Econometrics*, 178(2), 495–507.

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

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

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

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

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

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

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

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

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

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

Lee, S., Song, K., & Whang, Y.-J. (2015). Uniform asymptotics for nonparametric quantile regression with an application to testing monotonicity. Preprint, arXiv:1506.05337, 1–30.

Lee, S., Song, K., & Whang, Y.-J. (2018). Testing for a general class of functional inequalities. *Econometric Theory*, 34(5), 1018–1064.

Lee, S., & Whang, Y.-J. (2009). Nonparametric tests of conditional treatment effects. *CeMMAP Working Paper CWP36/09 and Cowles Foundation Discussion Paper 1740*. Yale University, pp. 1–68.

Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks*. Holden-Day.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.

Tang, S., Cai, Z., Fang, Y., & Lin, M. (2021). A new quantile treatment effect model for studying smoking effect on birth weight during mother's pregnancy. *Journal of Management Science and Engineering*, 6(3), 336–343.

Yang, T.-C., Shoff, C., Noah, A. J., Black, N., & Sparks, C. S. (2014). Racial segregation and maternal smoking during pregnancy: A multilevel analysis using the racial segregation interaction index. Social Science & Medicine 107 (April), 26–36.

Zhou, N., Guo, X., & Zhu, L. (2022). The role of propensity score structure in asymptotic efficiency of estimated conditional quantile treatment effect. *Scandinavian Journal of Statistics*, 49(2), 718–743.