



Nonparametric estimation and inference on conditional quantile processes[☆]



Zhongjun Qu^{a,*}, Jungmo Yoon^b

^a Department of Economics, Boston University, United States

^b Robert Day School of Economics and Finance, Claremont McKenna College, United States

ARTICLE INFO

Article history:

Received 5 September 2012

Received in revised form

11 July 2014

Accepted 20 October 2014

Available online 4 November 2014

JEL classification:

C14

C21

Keywords:

Nonparametric quantile regression

Uniform Bahadur representation

Uniform inference

Treatment effect

ABSTRACT

This paper presents estimation methods and asymptotic theory for the analysis of a nonparametrically specified conditional quantile process. Two estimators based on local linear regressions are proposed. The first estimator applies simple inequality constraints while the second uses rearrangement to maintain quantile monotonicity. The bandwidth parameter is allowed to vary across quantiles to adapt to data sparsity. For inference, the paper first establishes a uniform Bahadur representation and then shows that the two estimators converge weakly to the same limiting Gaussian process. As an empirical illustration, the paper considers a dataset from Project STAR and delivers two new findings.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Models for conditional quantiles play an important role in econometrics and statistics. In practice, it is often desirable to consider simultaneously multiple quantiles to obtain a complete analysis of the stochastic relationships between variables. This underlies the consideration of the conditional quantile process. A seminal contribution to this analysis is [Koenker and Portnoy \(1987\)](#), which established a uniform Bahadur representation and serves as the foundation for further developments in this area. More recently, [Koenker and Xiao \(2002\)](#) considered the issue of testing composite hypotheses about quantile regression processes using Khmaladzat (Khmaladze, 1981). [Chernozhukov and Fernandez-Val \(2005\)](#) considered the same issue and suggested re-sampling as an alternative approach. [Angrist et al. \(2006\)](#) established inferential theory in misspecified models. Their results can be used to study a wide range of issues, including but not restricted

to (i) testing alternative model specifications, (ii) testing stochastic dominance, and (iii) detecting treatment effect significance and heterogeneity.

The main focus of the above literature has been on parametric quantile models. However, there are frequent occasions where parametric specifications fail, making more flexible nonparametric methods desirable. This paper aims to achieve two goals. The first is to propose two simple nonparametric estimators for the conditional quantile process. The second is to derive an inferential theory that can be used for constructing uniform confidence bands and testing various hypotheses concerning conditional quantile processes.

The two proposed estimators are both based on local linear regressions ([Fan et al., 1994](#); [Yu and Jones, 1998](#)), but differing in how they ensure the quantile monotonicity. Specifically, the first estimator applies local linear regressions to a grid of quantiles while imposing a set of linear inequality constraints, and then linearly interpolates between adjacent quantiles to obtain an estimate for the quantile process. The second estimator first applies local linear regressions to a grid of quantiles without any constraints and then applies rearrangement ([Chernozhukov et al., 2010](#)) if quantile crossing occurs. They share the following two features. First, the bandwidth parameter is allowed to vary across quantiles to adapt to data sparsity. This is important because data are typically more sparse near the tails of the conditional distribution. Second, the computation is feasible even for large sample sizes. More detailed

[☆] First version: March 24, 2011. We thank the co-editor, the associate editor, two anonymous referees, Adam McCloskey, Denis Tkachenko, seminar participants at University of Iowa, BU, BC-BU mini conference, CUHK, 2011 Econometric Society Summer Meeting and 2011 Asian Meeting of the Econometric Society for useful comments and suggestions.

* Corresponding author.

E-mail addresses: qu@bu.edu (Z. Qu), jyoon@cmc.edu (J. Yoon).

comparisons between the two estimators are provided in Section 3 of the paper.

For inference, three sets of results are established. (1) We derive a uniform Bahadur representation for the unconstrained estimator (i.e., obtained without imposing the monotonicity constraint). This generalizes Theorem 2.1 in Koenker and Portnoy (1987) to the local linear regression setting. While being of independent interest, this representation forms a key step in proving the subsequent results. (2) We show that the first proposed estimator has the same first-order asymptotic distribution as the unconstrained estimator if a certain rate condition on the quantile grid is satisfied. Further, its asymptotic distribution is a continuous Gaussian process, whose critical values can be estimated via simulations by exploiting the fact that it is conditionally pivotal, drawing on the insights of Parzen et al. (1994) and Chernozhukov et al. (2009). (3) We show that the second proposed estimator follows the same asymptotic distribution as the first. This result broadens the application of rearrangement to the local linear regression context.

The inferential theory and methods can be used to analyze a wide range of issues. They include: (1) constructing a uniform confidence band for the conditional quantile process, (2) constructing a uniform confidence band for the difference or other linear functions of multiple such processes, and (3) testing distributional hypotheses such as the homogeneity or equality of quantile treatment effects, as well as first-order and second-order conditional stochastic dominance. They can also be potentially useful for constructing specification tests of parametrically specified conditional quantile processes. Studies considering the latter issue include Escanciano and Velasco (2010) and Mammen et al. (2013).

We evaluate the proposed methods using simulations and briefly summarize the results below. First, the two proposed estimators and the conventional quantile-by-quantile local linear estimator all perform similarly in terms of the integrated mean squared error criterion. This result confirms the finding that they all share the same limiting distribution. Second, the confidence band can have undercoverage because the bias term in the estimator can be difficult to estimate. This is not particular to our problem, but rather is a well known fact in the nonparametric literature. To address this issue, we suggest a simple modification that allows for a more flexible bias correction. The resulting confidence band is asymptotically conservative. Simulation evidence suggests that it has adequate coverage, even with small sample sizes, and that it is only mildly wider than the confidence band that uses the conventional bias correction.

As an empirical illustration, the paper considers a dataset from an experiment known as Project STAR (Student–Teacher Achievement Ratio). Two results emerge. First, the students in the upper quantiles of the test score distribution benefit more from the class size reduction. Second, the effect of the class size reduction is strongest for the classes taught by moderately experienced teachers (i.e., 6–8 years of experience). We also conduct hypotheses tests for treatment significance, homogeneity, equality as well as first order stochastic dominance. The results reconfirm the above two findings.

There are two key differences between this paper and Belloni et al. (2011). The first difference is in the estimation framework. Belloni et al. (2011) consider a series-based framework, where the conditional quantile function is modeled globally with a large number of parameters. The current paper is based on local linear regressions, where the quantile function is modeled locally by a few parameters and the modeling complexity is governed by the bandwidth. Consequently, different techniques are applied to establish the asymptotic properties of the estimators. The second difference is how the quantile monotonicity is achieved. Belloni et al. (2011) apply monotonicity procedures to a preliminary series-based estimator, while in our first estimator the monotonicity enters directly into the estimation through inequality constraints.

When viewed from a methodological perspective, the current paper is related to the following two strands of literature. First, the first estimator is related to the studies on estimating nonparametric regression relationships subject to monotonicity constraints, where the main focus has been the monotonicity with respect to the covariate. For example, Mammen (1991) considered an estimator consisting of a kernel smoothing step and an isotonicization step. Delecroix et al. (1996) studied a procedure that involves unconstrained smoothing followed by a constrained projection. He and Shi (1998) and Koenker and Ng (2005) considered smoothing splines subject to linear inequality constraints. In the current paper, the monotonicity constraint is with regards to the quantiles, giving rise to a different type of estimator than those discussed above, and requiring different techniques for studying its statistical properties. Note that He (1997), Dette and Volgushev (2008), Bondell et al. (2010) and Chernozhukov et al. (2010) considered monotonicity with respect to the quantiles. The connection with their works is discussed later in the paper. Second, there is an active literature that studies uniform confidence bands for nonparametric conditional quantile functions; see Hardle and Song (2010) and Koenker (2010). The former paper considered kernel-based estimators and obtained confidence bands using strong approximations. The latter considered additive quantile models analyzed with total-variation penalties and obtained confidence bands using Hotelling's tube formula. Their results are uniform in covariates but pointwise in quantiles. Therefore, their results and ours complement each other and, when jointly applied, allow one to probe a broad spectrum of topics.

The paper is organized as follows. Section 2 introduces the issue of interest. Section 3 presents the estimators while Section 4 establishes their asymptotic properties. Section 5 discusses the bandwidth selection. Section 6 shows how to construct uniform confidence bands and conduct hypothesis tests on the conditional quantile process. Section 7 reports simulation results. Section 8 contains an empirical application and Section 9 concludes. All proofs are in the two appendices, with Appendix A containing the proofs of the main results and Appendix B some auxiliary lemmas.

The following notation is used. The superscript 0 indicates the true value. $\|z\|$ is the Euclidean norm of a vector z . $1(\cdot)$ is the indicator function. $\text{supp}(f)$ stands for the support of f . The symbols " \Rightarrow " and " \rightarrow^p " denote weak convergence under the Skorohod topology and convergence in probability, and $O_p(\cdot)$ and $o_p(\cdot)$ is the usual notation for the orders of stochastic magnitude.

2. The issues of interest

Let (X, Y) be an \mathbb{R}^{d+1} -valued random vector, where Y is a scalar response variable and X is an \mathbb{R}^d -valued explanatory variable (X does not include a constant). Let $f_{Y|X}(\cdot)$ and $f_X(\cdot)$ be the conditional density of Y and the marginal density of X . Denote the conditional cumulative distribution of Y given $X = x$ by $F_{Y|X}(\cdot|x)$ and its conditional quantile at $\tau \in (0, 1)$ by $Q(\tau|x)$, i.e.,

$$Q(\tau|x) = F_{Y|X}^{-1}(\tau|x) = \inf \{s : F_{Y|X}(s|x) \geq \tau\}.$$

In this paper, $Q(\tau|x)$ is modeled as a general nonlinear function of x and τ . We fix x and treat $Q(\tau|x)$ as a process in τ , where $\tau \in \mathcal{T} = [\lambda_1, \lambda_2]$ with $0 < \lambda_1 \leq \lambda_2 < 1$. Here, \mathcal{T} falls strictly within the unit interval in order to allow the conditional distribution to have an unbounded support.

This paper has two goals. The first is to develop nonparametric estimators for the conditional quantile process. The second is to provide some asymptotic results that can be used for constructing uniform confidence bands and testing various hypotheses concerning $Q(\tau|x)$. Throughout the paper, we assume $\{(x_i, y_i)\}_{i=1}^n$ is a sample of n observations that are *i.i.d.* as (X, Y) . The following examples illustrate the above issues of interest. More discussions will follow in Section 6.

Example 1 (*Quantile Treatment Effect, QTE*). QTE measures the effect of a treatment on the distribution of the potential outcomes. Specifically, let $X = (D, Z)$, where D is a binary policy variable and Z includes the covariates. Let $Q(\tau|d, z)$ denote the τ th conditional quantile of the potential outcome given $D = d$ and $Z = z$. Then, the QTE is defined as¹

$$Q(\tau|1, z) - Q(\tau|0, z).$$

One may be interested in examining: (i) whether the treatment has a significant effect at some quantile, i.e., testing $H_0: Q(\tau|1, z) = Q(\tau|0, z)$ for all $\tau \in \mathcal{T}$ against $H_1: Q(\tau|1, z) \neq Q(\tau|0, z)$ for some $\tau \in \mathcal{T}$; (ii) whether the effects are homogeneous, i.e., testing $H_0: Q(\tau|1, z) - Q(\tau|0, z) = \delta(z)$ for some $\delta(z)$ and for all $\tau \in \mathcal{T}$ against the hypothesis that the preceding difference is quantile dependent; (iii) whether the effects are equal at two different covariate values, i.e., for given z_1 and z_2 , testing $H_0: Q(\tau|1, z_1) - Q(\tau|0, z_1) = Q(\tau|1, z_2) - Q(\tau|0, z_2)$ against $H_1: Q(\tau|1, z_1) - Q(\tau|0, z_1) \neq Q(\tau|1, z_2) - Q(\tau|0, z_2)$ for some $\tau \in \mathcal{T}$. Koenker and Xiao (2002) and Chernozhukov and Fernandez-Val (2005) developed inferential procedures for the first two hypotheses using parametric conditional quantile models. The results obtained here will allow us to analyze these issues under a nonparametric setting. In practice, \mathcal{T} can be flexibly chosen. For example, if the policy target is the lower part of the distribution, then we can choose $\mathcal{T} = [\varepsilon, 0.5]$ with ε being a small positive number.

Example 2 (*Conditional Stochastic Dominance*). Stochastic dominance is an important concept for the study of poverty and income inequality. Although a large part of the literature has focused on unconditional dominance, conditional dominance has also received interest recently, see Koenker and Xiao (2002), Chernozhukov and Fernandez-Val (2005) and Linton et al. (2005). Specifically, let $Q_1(\tau|x)$ and $Q_2(\tau|x)$ denote conditional quantile functions associated with two conditional distributions. Then, Distribution One (weakly) first-order stochastically dominates Distribution Two at x if and only if

$$Q_1(\tau|x) \geq Q_2(\tau|x) \quad \text{for all } \tau \in (0, 1). \tag{1}$$

Distribution One (weakly) second-order stochastically dominates Two at x if and only if

$$\int_0^\tau Q_1(s|x)ds \geq \int_0^\tau Q_2(s|x)ds \quad \text{for all } \tau \in (0, 1). \tag{2}$$

The null hypotheses (1) and (2) can be tested using one sided Kolmogorov–Smirnov tests based on $Q_1(\tau|x) - Q_2(\tau|x)$ and its integral. Note that the above three papers assumed parametric models for the conditional mean or quantiles functions, while the results obtained in this paper will allow us to test these hypotheses under a nonparametric setting.

Example 3 (*Sharp Regression Discontinuity Design, SRD*). The regression discontinuity design provides an alternative framework to randomized experiments for evaluating social programs and policy interventions. Under SRD, the treatment status of an individual, say D_i , is a deterministic function of some scalar variable X_i , i.e., $D_i = 1\{X_i \geq c\}$, where c is some known cut-off. The average effect of the treatment at $X_i = c$ is given by (see Imbens and Lemieux, 2008, for a more comprehensive discussion) $\lim_{x \downarrow c} E[Y_i|x] - \lim_{x \uparrow c} E[Y_i|x]$. Analogously, the quantile treatment effect can be defined as $\delta(\tau|c) = \lim_{x \downarrow c} Q(\tau|x) - \lim_{x \uparrow c} Q(\tau|x)$.

One may be interested in examining (i) whether the treatment has a significant effect at some quantile, i.e., testing $H_0: \delta(\tau|c) = 0$ for all $\tau \in \mathcal{T}$ against $H_1: \delta(\tau|c) \neq 0$ for some $\tau \in \mathcal{T}$, (ii) whether the effect is homogeneous, i.e., testing $H_0: \delta(\tau|c)$ is constant in $\tau \in \mathcal{T}$ against the hypothesis that $\delta(\tau|c)$ is quantile dependent, and (iii) whether the effect is uniformly nonnegative, i.e., testing $H_0: \delta(\tau|c) \geq 0$ for all $\tau \in \mathcal{T}$ against $H_1: \delta(\tau|c) < 0$ for some $\tau \in \mathcal{T}$.

Remark 1. Although one can get glimpses into the hypotheses in the above three examples by considering a few pre-specified quantiles, to address them thoroughly it is necessary to study the entire conditional quantile process. For example, to test whether the treatment effect is always non-negative or whether it is constant across quantiles, we need to examine the entire range of quantiles, rather than merely a few points in the conditional distribution. Concerning stochastic dominance, Abadie (2002) showed that if the outcome distribution of the treatment group first order stochastically dominates that of the control group, then the outcome is socially preferable under any widely acceptable utility function. Thus, the entire distribution, not only a few pre-specified quantiles, matters when designing socially preferable policies. The testing procedures and confidence bands proposed in this paper can provide informative results for such purposes.

3. The estimators

We start by briefly reviewing the idea underlying the local linear regression; more details can be found in Chaudhuri (1991), Fan et al. (1994) and Yu and Jones (1998). For a given $\tau \in (0, 1)$, the method assumes that $Q(\tau|x)$ is a smooth function of x and exploits the following first-order Taylor approximation:

$$Q(\tau|x_i) \approx Q(\tau|x) + (x_i - x)' \frac{\partial Q(\tau|x)}{\partial x}. \tag{3}$$

The local linear estimator of $Q(\tau|x)$, denoted by $\hat{\alpha}(\tau)$, is determined via

$$\begin{aligned} (\hat{\alpha}(\tau), \hat{\beta}(\tau)) = \arg \min_{\alpha(\tau), \beta(\tau)} & \sum_{i=1}^n \rho_\tau(y_i - \alpha(\tau) - (x_i - x)' \beta(\tau)) \\ & \times K\left(\frac{x_i - x}{h_{n,\tau}}\right), \end{aligned} \tag{4}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function, K is a kernel function and $h_{n,\tau}$ is a bandwidth parameter that depends on τ . As demonstrated in Fan et al. (1994), the local linear regression has several advantages over the local constant fit. In particular, (i) the bias of $\hat{\alpha}(\tau)$ is not affected by the value of $f'_x(x)$ and $\partial Q(\tau|x)/\partial x$, (ii) it is of the same order irrespective of whether x is a boundary point, and (iii) plug-in data-driven bandwidth selection does not require estimating the derivatives of the marginal density, therefore is relatively simple to implement. As will be seen later, these three features continue to hold for our estimators. Note that the results in Fan et al. (1994) and Yu and Jones (1998) are pointwise in τ .

3.1. The first estimator

We propose the following two-step procedure to estimate the process $Q(\tau|x)$, $\tau \in \mathcal{T}$.

STEP 1. Partition \mathcal{T} into a grid of equally spaced quantiles $\{\tau_1, \dots, \tau_m\}$. Consider the following constrained optimization problem

$$\begin{aligned} \min_{\{\alpha(\tau_j), \beta(\tau_j)\}_{j=1}^m} & \sum_{j=1}^m \sum_{i=1}^n \rho_\tau(y_i - \alpha(\tau_j) - (x_i - x)' \beta(\tau_j)) \\ & \times K\left(\frac{x_i - x}{h_{n,\tau_j}}\right) \end{aligned} \tag{5}$$

¹ This concept traces back to Lehmann (1975) and Doksum (1974); recent works include Heckman et al. (1997), Abadie et al. (2002), Chernozhukov and Hansen (2005) and Firpo (2007).

subject to

$$\alpha(\tau_j) \leq \alpha(\tau_{j+1}) \quad \text{for all } j = 1, \dots, m - 1. \quad (6)$$

Denote the estimates by $\tilde{\alpha}(\tau_1), \dots, \tilde{\alpha}(\tau_m)$.

STEP 2. Linearly interpolate between $\tilde{\alpha}(\tau_1), \dots, \tilde{\alpha}(\tau_m)$ to obtain an estimate for the quantile process, i.e., for any $\tau \in \mathcal{T}$, compute

$$\tilde{\alpha}^*(\tau) = \gamma(\tau) \tilde{\alpha}(\tau_j) + (1 - \gamma(\tau)) \tilde{\alpha}(\tau_{j+1}) \quad \text{if } \tau \in [\tau_j, \tau_{j+1}], \quad (7)$$

where $\gamma(\tau) = (\tau_{j+1} - \tau) / (\tau_{j+1} - \tau_j)$.

Remark 2. The constraints $\alpha(\tau_j) \leq \alpha(\tau_{j+1})$ ($j = 1, \dots, m - 1$) are necessary and sufficient for ensuring quantile monotonicity. They depend neither on the data (such as their support) nor on the model (such as the number of covariates), but only on the number of quantiles entering the estimation. Consequently, the above procedure can be applied to different models and covariate values without any modification.

Remark 3. The constrained optimization (5) is essentially a special case of Koenker and Ng (2005), who provided algorithms for computing quantile regression estimates under general linear inequality constraints. An efficient implementation can be constructed with little cost.

Remark 4. The linear interpolation step is motivated by Neocleous and Portnoy (2008). It permits obtaining a tractable inferential theory as presented later.

Remark 5. There exist other methods for ensuring quantile monotonicity. He (1997) exploited the structure of a location–scale model. Bondell et al. (2010) considered quantile smoothing splines and showed that non-crossing can be imposed via inequality constraints on the knots. Dette and Volgushev (2008) first estimated the conditional distribution function and then inverted it to obtain the quantiles. However, their results are pointwise in quantiles. It remains an open question whether they can lead to tractable inferential theory for the quantile process.

3.2. The second estimator

Chernozhukov et al. (2010) proposed a generic framework for estimating monotone probability and quantile curves. Their method uses a preliminary estimator that is not necessarily monotonic and applies rearrangement to ensure monotonicity. Adapting their method to the current nonparametric setting leads to the following procedure.

STEP 1. Partition \mathcal{T} into a grid of equally spaced quantiles $\{\tau_1, \dots, \tau_m\}$. Solve the following unconstrained optimization problem

$$\min_{\alpha(\tau_j), \beta(\tau_j)} \sum_{i=1}^n \rho_\tau(y_i - \alpha(\tau_j) - (x_i - x)' \beta(\tau_j)) K\left(\frac{x_i - x}{h_{n,\tau_j}}\right)$$

for all $j = 1, \dots, m$. Denote the estimates by $\hat{\alpha}(\tau_1), \dots, \hat{\alpha}(\tau_m)$.

STEP 2. Linearly interpolate between the estimates:

$$\hat{\alpha}^*(\tau) = \gamma(\tau) \hat{\alpha}(\tau_j) + (1 - \gamma(\tau)) \hat{\alpha}(\tau_{j+1}) \quad \text{if } \tau \in [\tau_j, \tau_{j+1}], \quad (8)$$

where $\gamma(\tau) = (\tau_{j+1} - \tau) / (\tau_{j+1} - \tau_j)$. Apply rearrangement to $\hat{\alpha}^*(\tau)$ (see Displays (2.1) and (2.2) in Chernozhukov et al., 2010) to obtain

$$\tilde{\alpha}^*(\tau) = \inf \left\{ y \in \mathbb{R} : \int_{\mathcal{T}} 1(\hat{\alpha}^*(u) \leq y) du \geq \tau - \lambda_1 \right\}, \quad (9)$$

where λ_1 is the lower limit of \mathcal{T} .

In the second step, the rearrangement is applied after the linear interpolation in order to be consistent with the theoretical

analysis in Chernozhukov et al. (2010). As the paper suggests, to facilitate the implementation, it can also be applied directly to $\hat{\alpha}(\tau_j)$ ($j = 1, \dots, m$) provided that m is sufficiently large. The linear interpolation can then be applied to the monotone estimate to obtain the final estimate.

Although this paper focuses on the nonparametric setting, the above two procedures can also be extended to analyze semiparametric models. Below we provide two illustrations. The discussion focuses on the implementation aspect while the asymptotic analysis is beyond the scope of the current paper. First, consider a partially linear conditional quantile model given by

$$Q(\tau | x_i, z_i) = q(x_i, \tau) + z_i' \delta(\tau),$$

where x_i and z_i are two sets of covariates (excluding a constant), $q(x_i, \tau)$ is a nonlinear function of both arguments and $\delta(\tau)$ a vector of coefficients. Similar models are considered in He and Shi (1996), Chen and Khan (2001), Lee (2003) and Song et al. (2012). Analogously to (3) and (4), a local linear approximation to the τ th conditional quantile at a particular x is given by

$$Q(\tau | x, z_i) \approx \alpha(\tau) + (x_i - x)' \beta(\tau) + z_i' \delta(\tau)$$

and the parameter estimates can be obtained using (see Chen and Khan, 2001)

$$\begin{aligned} (\hat{\alpha}(\tau), \hat{\beta}(\tau), \hat{\delta}(\tau)) = & \operatorname{argmin}_{\alpha(\tau), \beta(\tau), \delta(\tau)} \sum_{i=1}^n \rho_\tau(y_i - \alpha(\tau) \\ & - (x_i - x)' \beta(\tau) - z_i' \delta(\tau)) K\left(\frac{x_i - x}{h_{n,\tau}}\right). \end{aligned}$$

Here, the quantile monotonicity at a particular x and over the support of z_i can be enforced by imposing constraints on $\alpha(\tau)$ and $\delta(\tau)$. This will automatically guarantee the linearity with respect to z_i . Alternatively, we can solve the above unconstrained estimation problem and then apply rearrangement at x and the values of z_i that are of interest. This may not preserve the linearity with respect to z_i when quantile crossing occurs.

Next, consider a varying coefficients quantile model given by (see Honda, 2004; Kim, 2007; Cai and Xiao, 2012)

$$Q(\tau | x_i) = x_i' \beta(\tau, u_i),$$

where u_i is a scalar random variable and $\beta(\tau, u_i)$ is a smooth function of u_i . A local linear approximation to the τ th conditional quantile at a particular u is given by $Q(\tau | u, x_i) \approx x_i' \beta(\tau) + (u_i - u) x_i' \delta(\tau)$ and the parameter estimates can be obtained using

$$\begin{aligned} (\hat{\beta}(\tau), \hat{\delta}(\tau)) = & \operatorname{argmin}_{\beta(\tau), \delta(\tau)} \sum_{i=1}^n \rho_\tau(y_i - x_i' \beta(\tau) - (u_i - u) x_i' \delta(\tau)) \\ & \times K\left(\frac{u_i - u}{h_{n,\tau}}\right). \end{aligned}$$

Similarly to the partially linear model, the quantile monotonicity at a particular u and over the support of x_i can be enforced either by imposing constraints on $\beta(\tau)$, or by repeating the rearrangement operation for this u and all x_i of interest. The comment regarding the linearity with respect to x_i also applies here.

We conclude this section by comparing the two estimators taking into consideration of their further developments. The second estimator is computationally more straightforward if the main goal is to study the conditional quantile process nonparametrically for a fixed x . The first estimator is more flexible if one wishes to impose additional restrictions, say restricting some elements of $\beta(\tau)$ to be non-negative (see the Engel curve example considered in Koenker and Ng, 2005). It can also be attractive when dealing with certain semiparametric models as discussed above. Although the theory in the next section does not consider the estimation of the derivatives and semiparametric models, we conjecture that the techniques developed there can be useful for such further analysis.

4. Asymptotic properties

Let $\delta(x)$ denote some small open neighborhood surrounding x . We make the following assumptions.

Assumption 1. (i) X is an \mathbb{R}^d -valued explanatory variable excluding a constant. (ii) The evaluation point x belongs to the support of X whose density f_X is continuously differentiable at x with $0 < f_X(x) < \infty$. (iii) $\{(x_i, y_i)\}_{i=1}^n$ is a sample of n observations that are *i.i.d.* as (X, Y) .

Assumption 2. (i) $f_{Y|X}(Q(\tau|x)|x)$ is Lipschitz continuous over \mathcal{T} . (ii) There exist finite positive constants f_L, f_U and ϵ such that $f_L \leq f_{Y|X}(Q(\tau|s) + \eta|s) \leq f_U$ for all $|\eta| \leq \epsilon, s \in \delta(x) \cap \text{supp}(f_X)$ and $\tau \in \mathcal{T}$.

Assumption 3. (i) $Q(\tau|x)$ and $\partial Q(\tau|x)/\partial \tau$ are finite and continuous in $\tau \in \mathcal{T}$. (ii) The elements of $\partial^2 Q(\tau|s)/\partial s \partial s'$ are finite and Lipschitz continuous over $\{(s, \tau) : s \in \delta(x) \cap \text{supp}(f_X) \text{ and } \tau \in \mathcal{T}\}$.

Assumption 4. The kernel K is compactly supported, bounded, having finite first-order derivatives and satisfying $K(\cdot) \geq 0, \int K(u)du = 1, \int uK(u)du = 0, \int uu'K(u)du = \mu_2(K)$ with $\mu_2(K)$ being positive definite.

Assumption 5. The bandwidth $h_{n,\tau}$ satisfies $h_{n,\tau} = c(\tau)h_n$, where $h_n = O(n^{-1/(4+d)})$ and $nh_n^d \rightarrow \infty$ and $c(\tau)$ is Lipschitz continuous satisfying $0 < \underline{c} \leq c(\tau) \leq \bar{c} < \infty$ for all $\tau \in \mathcal{T}$.

Assumption 1 is fairly standard. In practice, d is typically a small number. Although the *i.i.d.* assumption is maintained throughout the paper, the asymptotic results derived in Section 4 continue to hold in a Markov time series setting, that is, when $Y = Y_t, X = X_{t-1}$ and Y_t depends on the past only through X_{t-1} . **Assumption 2** imposes restrictions on the conditional density. Part (i) requires it to have finite first-order derivatives with respect to τ over \mathcal{T} . It is used to show that the optimal bandwidth in Display (16) is Lipschitz continuous with respect to τ . This further ensures that allowing the bandwidth parameter to vary across quantiles does not interfere with the stochastic equicontinuity. Part (ii) implies that it is positive and finite in a neighborhood of x and \mathcal{T} . This is needed to ensure that the estimator will have the usual rate of convergence. **Assumption 3** imposes restrictions on the conditional quantile process. Part (i) requires this process and its derivative in τ to vary smoothly with respect to τ . Part (ii) implies that the error in the Taylor expansion (3) is uniformly small. A common theme of **Assumptions 2** and **3** is that they are local. That is, they impose restrictions only on neighborhoods surrounding x and \mathcal{T} . For example, if $\mathcal{T} = [0.5, 0.8]$, then the lower part of the conditional distribution and the extreme quantiles are left unrestricted. There, the conditional densities can be non-smooth or zero.

Assumption 4 permits univariate as well as multivariate kernels, although it rules out those with unbounded support (e.g., the Gaussian kernel) and those with unbounded derivatives. The matrix $\mu_2(K)$ will be diagonal if a product kernel is used. **Assumption 5** imposes restrictions on the bandwidth parameter, in particular requiring it to be a smooth function of τ . This is needed to ensure stochastic equicontinuity. It is not restrictive, and is satisfied by the optimal bandwidth that minimizes the asymptotic MSE; see Section 5. Note that for a given τ , we use the same bandwidth for every coordinate of x . This simplifies the expressions for the bias and variance, especially when x is a boundary point. More general formulas allowing for different bandwidths at different x can be obtained by applying similar arguments as those in Section 5.

The asymptotic properties of the estimator depend on whether x is close to the boundary of the support. As in **Ruppert and Wand (1994)**, define $\mathcal{E}_{x,\tau} = \{v \in \mathbb{R}^d : h_{n,\tau}^{-1}(x - v) \in \text{supp}(K)\}$. Call x an

interior point if $\mathcal{E}_{x,\tau} \subset \text{supp}(f_X)$ for all $\tau \in \mathcal{T}$ and otherwise call it a boundary point. If x is a fixed point in the interior of $\text{supp}(f_X)$, then x is an interior point for all large n . Therefore, to model the boundary point, we consider a sequence of points $x = x^n$ converging to x_∂ on the boundary of $\text{supp}(f_X)$ so that x is a boundary point for all n . Formally, a boundary point x satisfies

$$x = x_\partial + h_{n,\tau}c \quad \text{for some } \tau \in \mathcal{T} \text{ and some fixed } c \in \text{supp}(K). \tag{10}$$

We also define the following set which serves as the domain for integration:

$$\mathcal{D}_{x,\tau} = \{v \in \mathbb{R}^d : (x + h_{n,\tau}v) \in \text{supp}(f_X)\} \cap \text{supp}(K).$$

Note that if x is an interior point, then $\mathcal{D}_{x,\tau} = \text{supp}(K)$. To illustrate the boundary point case, suppose $\text{supp}(K) = [-1, 1], \text{supp}(f_X) = [0, 1]$ and $x = ch_{n,\tau}$ with $c > 0$, then $\mathcal{D}_{x,\tau} = [-c, 1]$. The next assumption rules out degeneracies and is the same as in **Ruppert and Wand (1994)**.

Assumption 6. There is a convex set \mathcal{C} with nonnull interior such that $x_\partial \in \mathcal{C}$ and $\inf_{x \in \mathcal{C}} f_X(x) > 0$.

Below, we establish the asymptotic properties of (7) and (9) in three steps by sequentially analyzing the following three quantities:

- $\hat{\alpha}(\tau)$: the solution of (4) at a given $\tau \in \mathcal{T}$,
- $\hat{\alpha}^*(\tau)$: obtained from the linear interpolation without imposing the constraints, see (8),
- $\tilde{\alpha}^*(\tau)$: the final estimator defined in (7) or (9).

The first step provides a uniform Bahadur representation for $\hat{\alpha}(\tau)$. The second step quantifies the effect of the linear interpolation. The third step analyzes the effect of the inequality constraints or the rearrangement and delivers the limiting distributions of the proposed estimators.

Theorem 1 (Uniform Bahadur Representation). *Let Assumptions 1–5 hold. Then the following results hold uniformly in $\tau \in \mathcal{T}$.*

1. If x is an interior point, then

$$\begin{aligned} & \sqrt{nh_{n,\tau}^d} (\hat{\alpha}(\tau) - Q(\tau|x) - d_\tau h_{n,\tau}^2) \\ &= \frac{(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i^0(\tau) \leq 0)) K_{i,\tau}}{f_X(x) f_{Y|X}(Q(\tau|x)|x)} + o_p(1), \end{aligned}$$

where $d_\tau = \frac{1}{2} \text{tr} \left(\frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \mu_2(K) \right), K_{i,\tau} = K \left(\frac{x_i - x}{h_{n,\tau}} \right), u_i^0(\tau) = y_i - Q(\tau|x_i)$.

2. If x is a boundary point as defined by (10) and **Assumption 6** holds, then

$$\begin{aligned} & \sqrt{nh_{n,\tau}^d} (\hat{\alpha}(\tau) - Q(\tau|x) - d_{b,\tau} h_{n,\tau}^2) \\ &= \frac{\iota_1' N_x(\tau)^{-1} (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i^0(\tau) \leq 0)) z_{i,\tau} K_{i,\tau}}{f_X(x) f_{Y|X}(Q(\tau|x)|x)} \\ &+ o_p(1), \end{aligned}$$

where $\iota_1 = (1, 0_d)'$, $u \in \mathbb{R}^d, \bar{u} = (1, u)'$, $d_{b,\tau} = \frac{1}{2} \iota_1' N_x(\tau)^{-1} \int_{\mathcal{D}_{x,\tau}} u' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} u \bar{u} K(u) du, N_x(\tau) = \int_{\mathcal{D}_{x,\tau}} \bar{u} \bar{u}' K(u) du, z_{i,\tau}' = (1, (x_i - x)' / h_{n,\tau})$.

Note that the bias terms will be zero if $Q(\tau|x)$ is linear in x . **Theorem 1** generalizes **Theorem 2.1** in **Koenker and Portnoy (1987)** to the nonparametric setting. The proof faces two challenges that do not arise in the parametric situation. First, the conditional quantile function is approximated using a first-order Taylor expansion. Second, the bandwidth can vary across quantiles. They both lead to difficulties when establishing the stochastic equicontinuity. Since the strategy of the proof can be of interest, we provide an outline below.

We first introduce some notations that will also be used in the two appendices. Define $u_i(\tau) = y_i - \alpha(\tau) - (x_i - x)' \beta(\tau)$, where $\alpha(\tau) \in \mathbb{R}$ and $\beta(\tau) \in \mathbb{R}^d$ are some candidate parameter values. Let

$$e_i(\tau) = \left[Q(\tau|x) + (x_i - x)' \frac{\partial Q(\tau|x)}{\partial x} \right] - Q(\tau|x_i) \quad \text{and}$$

$$\phi(\tau) = \sqrt{nh_{n,\tau}^d} \left(\begin{matrix} \alpha(\tau) - Q(\tau|x) \\ h_{n,\tau} \left(\beta(\tau) - \frac{\partial Q(\tau|x)}{\partial x} \right) \end{matrix} \right).$$

Then, $u_i(\tau)$ can be decomposed as $u_i(\tau) = u_i^0(\tau) - e_i(\tau) - (nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \phi(\tau)$, where $z'_{i,\tau} = (1, (x_i - x)' / h_{n,\tau})$. This decomposition is useful because it breaks $u_i(\tau)$ into three components: the true residual, the error due to the Taylor approximation and the error caused by replacing the unknown parameter values in the approximation with some estimates. Define

$$V_{n,\tau}(\phi(\tau)) = \sum_{i=1}^n \rho_\tau(u_i^0(\tau) - e_i(\tau) - (nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \phi(\tau)) K_{i,\tau} - \sum_{i=1}^n \rho_\tau(u_i^0(\tau) - e_i(\tau)) K_{i,\tau}, \quad (11)$$

where the first summation is the same as in the objective function in (4) and the second is introduced for recentering the first and has no effects on the estimation. Finally, let

$$S_n(\tau, \phi(\tau), e_i(\tau)) = (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n \{ P(1(u_i^0(\tau) \leq (nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \phi(\tau) + e_i(\tau)) | x_i) - 1(u_i^0(\tau) \leq (nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \phi(\tau) + e_i(\tau)) \} z_{i,\tau} K_{i,\tau}.$$

The proof proceeds in three steps. It is structured similarly to that of **Theorem 2.1** in **Koenker and Portnoy (1987)**. The first step establishes uniform rates of convergence for $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$, i.e., $\sup_{\tau \in \mathcal{T}} \|\hat{\phi}(\tau)\| \leq \log^{1/2}(nh_n^d)$ with probability tending to 1. Specifically, it uses the convexity of $V_{n,\tau}(\phi(\tau))$ and **Knight's (1998)** identity to show that, if the preceding inequality is violated, $V_{n,\tau}(\phi(\tau))$ will be strictly positive with probability tending to 1 for some $\tau \in \mathcal{T}$. This contradicts with $\hat{\phi}(\tau)$ being the minimizer of $V_{n,\tau}(\phi(\tau))$. The second step studies the subgradient over the set

$$\Phi = \{(\tau, \phi(\tau)) : \tau \in \mathcal{T}, \|\phi(\tau)\| \leq \log^{1/2}(nh_n^d)\}. \quad (12)$$

Specifically, the subgradient multiplied by $(nh_{n,\tau}^d)^{-1/2}$ equals $(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n \{ \tau - 1(u_i^0(\tau) \leq e_i(\tau) + (nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \hat{\phi}(\tau)) \} z_{i,\tau} K_{i,\tau}$, which, by **Theorem 2.1** in **Koenker (2005)**, is of order $O_p((nh_n^d)^{-1/2}) = o_p(1)$ uniformly in τ . Adding and subtracting terms, it can be rewritten as

$$\begin{aligned} & \{S_n(\tau, \hat{\phi}(\tau), e_i(\tau)) - S_n(\tau, 0, e_i(\tau))\} + \{S_n(\tau, 0, e_i(\tau)) \\ & - S_n(\tau, 0, 0)\} + S_n(\tau, 0, 0) + (nh_{n,\tau}^d)^{-1/2} \\ & \times \sum_{i=1}^n \left\{ \tau - P(u_i^0(\tau) \leq e_i(\tau) \right. \\ & \left. + (nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \hat{\phi}(\tau) \mid x_i) \right\} z_{i,\tau} K_{i,\tau}. \end{aligned} \quad (13)$$

The first term measures the effect from estimating the parameters, the second term is due to the remainder term in the Taylor expansion, the third term depends only on the true data generating process, while the fourth term involves only smooth functions. To further analyze the first term, the set Φ is partitioned into $C(nh_n^d)^{1/2+\kappa} (\log^{1/2}(nh_n^d)/\delta)^{(d+1)}$ cells with C, κ and δ being some constants such that, within each cell, the differences in τ and $\phi(\tau)$ are at most $(nh_n^d)^{-1/2-\kappa}$ and δ respectively. Then, using a chaining argument dating back to **Bickel (1975)**, it is shown that this term can be bounded using the values of $S_n(\tau, \phi(\tau), e_i(\tau))$ evaluated at certain vertices of the cells. The latter involves only a countable number of points and can be further analyzed using the maximal inequalities developed in **Bai (1996)** and **Oka and Qu (2011)**. It then follows that the first term is $o_p(1)$ uniformly over the set Φ . The second term is analyzed in a similar way. The third term is $O_p(1)$ and appears in **Theorem 1** and the fourth term is analyzed using a first order Taylor expansion. Finally, the third step derives the formulae for d_τ and $d_{b,\tau}$ and verifies the Bahadur representation.

Theorem 1 is uniform in τ but pointwise in x . Establishing uniformity in both dimensions (i.e., showing that remainder term in **Theorem 1** is $o_p(1)$ with respect to both τ and x while maintaining **Assumptions 1–6**) will be a challenging task. We conjecture the techniques developed in the following papers can be useful for such an analysis. **Masry (1996, Theorem 6)** showed that the local polynomial estimator of the conditional mean function converges uniformly and almost surely at the rate $(\log n / (nh_n^d))^{1/2}$. **Kong et al. (2010)** considered a general class of M-regression functions. They established a uniform Bahadur representation with an remainder term of order $(\log n / (nh_n^d))^{3/4}$. In both papers, an important step is to partition the support of the covariates into small cells whose size depends on the bandwidth and the sample size. Subsequently, it is shown that the differences in the relevant quantities are small both within and across the cells; see the proofs of **Theorem 2** in **Masry (1996)** and **Lemma 1** in **Kong et al. (2010)**. **Guerre and Sabbah (2012)** studied uniformity with respect to both τ and x . Combining the techniques in the above papers with the chaining arguments used here can be a potential way to establish the uniformity in both x and τ under the current setting, although substantial work is needed. Assuming such a representation is obtainable, then the techniques developed in **Lee et al. (2009)**, which handle Gaussian processes indexed by multidimensional parameters over an increasing set, can be useful for deriving relevant distributional properties.

Theorem 1 implies that the limit of $\hat{\alpha}(\tau)$ is driven by the leading term in the Bahadur representation. This paves the way for establishing the weak convergence of $\hat{\alpha}^*(\tau)$, stated below.

Theorem 2 (Weak Convergence Under Linear Interpolation). *Let Assumptions 1–5 hold. Suppose $m / (nh_n^d)^{1/4} \rightarrow \infty$ as $n \rightarrow \infty$. Then:*

1. If x is an interior point, then

$$\begin{aligned} & \sqrt{nh_{n,\tau}^d} f_X(x) f_{Y|X}(Q(\tau|x)|x) (\hat{\alpha}^*(\tau) - Q(\tau|x) - d_\tau h_{n,\tau}^2) \\ & \Rightarrow G(\tau), \end{aligned}$$

where $G(\tau)$ is a zero mean continuous Gaussian process defined over \mathcal{T} with covariance

$$\begin{aligned} E(G(r)G(s)) &= (\kappa(r)\kappa(s))^{-d/2} (r \wedge s - rs) \\ & \times \int K\left(\frac{u}{\kappa(r)}\right) K\left(\frac{u}{\kappa(s)}\right) du, \end{aligned} \quad (14)$$

where $\kappa(\tau) = h_{n,\tau} / h_{n,1/2} = c(\tau) / c(1/2)$.

2. If x is a boundary point as defined by (10) and **Assumption 6** holds, then

$$\begin{aligned} & \sqrt{nh_{n,\tau}^d} f_X(x) f_{Y|X}(Q(\tau|x)|x) (\hat{\alpha}^*(\tau) - Q(\tau|x) - d_{b,\tau} h_{n,\tau}^2) \\ & \Rightarrow G_b(\tau), \end{aligned}$$

where $G_b(\tau)$ is a zero mean continuous Gaussian process defined over \mathcal{T} with covariance

$$E(G_b(r)G_b(s)) = (\kappa(r)\kappa(s))^{-d/2} (r \wedge s - rs) \iota_1' N_x(r)^{-1} \times T_x(r, s) N_x(s)^{-1} \iota_1, \tag{15}$$

where

$$T_x(r, s) = \int_{\mathcal{D}_{x, h_{1/2}}} \left[\frac{1}{\kappa(r)} \right] K\left(\frac{u}{\kappa(r)}\right) \begin{bmatrix} 1 & u' \\ \kappa(r) & \kappa(r) \end{bmatrix} \times K\left(\frac{u}{\kappa(s)}\right) du$$

and $\kappa(r)$ and $N_x(r)$ have the same definitions as before.

The Gaussian process $G(\tau)$ depends only on the kernel and the bandwidth. In the extreme case where the bandwidths are equal across quantiles, it is simply the Brownian bridge (truncated to the interval \mathcal{T}) multiplied by $\int K^2(u) du$. In the boundary case, the structure of the process is similar, except now it also depends on the location of x relative to the boundary of the support.

The rate condition $m/(nh_n^d)^{1/4} \rightarrow \infty$ ensures that the linear interpolation induces no loss of efficiency asymptotically, i.e., the limiting process is the same as if all quantiles were being estimated directly. When formulating the result, we have related the bandwidths at τ to the median. This facilitates the computation of the bandwidth selection rule in the next section. The next result presents the asymptotic distributions of the two estimators. They are both denoted by $\tilde{\alpha}^*(\tau)$.

Theorem 3 (Weak Convergence of the Two Proposed Estimators). *Let Assumptions 1–5 hold and assume $m/(nh_n^d)^{1/4} \rightarrow \infty$ as $n \rightarrow \infty$. In addition, assume $m/(nh_n^d)^{1/2} \rightarrow 0$ holds for the first estimator and $(nh_{n,\tau}^{4+d})^{1/2} \rightarrow h(\tau) < \infty$ holds for the second estimator. Then, they both satisfy*

$$\sqrt{nh_{n,\tau}^d f_X(x) f_{Y|X}(Q(\tau|x)|x)} (\tilde{\alpha}^*(\tau) - Q(\tau|x) - d_\tau h_{n,\tau}^2) \Rightarrow G(\tau)$$

if x is an interior point, and

$$\sqrt{nh_{n,\tau}^d f_X(x) f_{Y|X}(Q(\tau|x)|x)} (\tilde{\alpha}^*(\tau) - Q(\tau|x) - d_{b,\tau} h_{n,\tau}^2) \Rightarrow G_b(\tau)$$

if x is a boundary point and Assumption 6 holds.

The first estimator requires an additional rate condition $m/(nh_n^d)^{1/2} \rightarrow 0$. This ensures that the constraints (6) act as a finite sample correction, having no effect on the limiting distribution. We conjecture this condition may not be necessary, but have not been able to relax it. The adequacy of the limiting distributions under different m will be evaluated using simulations. It turns out that the property of $\tilde{\alpha}^*(\tau)$ is rather insensitive to such choices, as long as m is not too small. Motivated by this finding, we suggest the following simple rule $m = \max(10, \sqrt{nh_n^d / \log(nh_n^d)})$, where the first argument safeguards against using too few quantiles when the sample size is relatively small and the second permits choosing a large number if the sample size is large. The second estimator requires $nh_{n,\tau}^{4+d}$ to converge. This is to ensure that $\sqrt{nh_{n,\tau}^d} (\tilde{\alpha}^*(\tau) - Q(\tau|x))$ has a well-defined limiting process. This condition is satisfied by the bandwidth proposed in the next section.

5. Bandwidth selection

Theorem 3 implies that, for any given $\tau \in \mathcal{T}$, $\tilde{\alpha}^*(\tau)$ has the same limiting distribution as the conventional local linear estimator studied by Fan et al. (1994). Therefore, the same bandwidth selection rule can be applied.

Corollary 1 (Optimal Bandwidth for an Interior Point). *Let Assumptions 1–5 and those stated in Theorem 3 hold. Assume $|\text{tr}(\partial^2 Q(\tau|x)/\partial x \partial x')| > 0$. Then, the bandwidth that minimizes the (interior) asymptotic MSE of $\tilde{\alpha}^*(\tau)$ for any $\tau \in \mathcal{T}$ is given by*

$$h_{n,\tau}^* = \left(\frac{\tau(1-\tau) \int K(v)^2 dv}{f_X(x) f_{Y|X}(Q(\tau|x)|x)^2 \left\{ \text{tr} \left(\frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \mu_2(K) \right) \right\}^2} \right)^{1/(4+d)} \times n^{-1/(4+d)}. \tag{16}$$

In Appendix A, we verify that the above bandwidth satisfies Assumption 5. The result implies that the bandwidth selection rule for estimating the conditional quantile process is conceptually no more difficult than in the conventional situation. The result also illustrates how the bandwidths change as τ shifts away from the center of the distribution. It typically widens as changes in $f_{Y|X}(Q(\tau|x)|x)$ often dominate the other terms.

To compute the bandwidth (16), the main challenge is in estimating $\partial^2 Q(\tau|x)/\partial x \partial x'$. Although this quantity can often be estimated reasonably well for τ close to the median, the estimation precision can be low when τ is close to the tail of the conditional distribution. The latter can affect negatively the estimation of the conditional quantile process. In simulations and the empirical application, we have experimented with an approximation due to Yu and Jones (1998), which treats $\partial^2 Q(\tau|x)/\partial x \partial x'$ as constant across quantiles. Specifically, under such an approximation, the optimal bandwidth at τ is related to the median via

$$\left(\frac{h_{n,\tau}^*}{h_{n,1/2}^*} \right)^{4+d} = 4\tau(1-\tau) \left(\frac{f_{Y|X}(Q(\frac{1}{2}|x)|x)}{f_{Y|X}(Q(\tau|x)|x)} \right)^2.$$

Next, applying a Normal reference method (considering $f_{Y|X}$ to be a Gaussian density) as in Yu and Jones (1998), the above relationship simplifies to

$$\left(\frac{h_{n,\tau}^*}{h_{n,1/2}^*} \right)^{4+d} = \frac{2\tau(1-\tau)}{\pi \phi(\Phi^{-1}(\tau))^2}, \tag{17}$$

where ϕ and Φ are the density and the cdf of a standard normal random variable. Finally, the bandwidth $h_{n,1/2}^*$ can be determined using (16) in the above Corollary. This procedure delivers a sequence of bandwidths that automatically satisfies Assumption 5.

6. Uniform confidence bands and hypothesis tests

We first present the uniform band for the conditional quantile process and explain how to estimate the relevant quantities appearing in its expression. Next, we illustrate how to compute confidence band for the difference between two conditional quantile processes. Then, we show how to test a variety of hypotheses on quantile processes using the three examples discussed in Section 2.

6.1. Uniform confidence band for the conditional quantile process

Corollary 2. *Let Assumptions 1–5 and those stated in Theorem 3 hold, then an asymptotic $(1-p)$ confidence band for $Q(\tau|x)$ is given by:*

1. If x is an interior point,

$$C_p = \left[(\tilde{\alpha}^*(\tau) - d_\tau h_{n,\tau}^2) - \sigma_{n,\tau}(x) Z_p, (\tilde{\alpha}^*(\tau) - d_\tau h_{n,\tau}^2) + \sigma_{n,\tau}(x) Z_p \right], \tag{18}$$

where $\sigma_{n,\tau}(x) = \left(\sqrt{nh_{n,\tau}^d f_X(x) f_{Y|X}(Q(\tau|x)|x)} \right)^{-1}$ and Z_p is the $(1-p)$ percentile of $\sup_{\tau \in \mathcal{T}} |G(\tau)|$.

2. If x is a boundary point,

$$C_p = [(\tilde{\alpha}^*(\tau) - d_{b,\tau} h_{n,\tau}^2) - \sigma_{n,\tau}(x) Z_p, (\tilde{\alpha}^*(\tau) - d_{b,\tau} h_{n,\tau}^2) + \sigma_{n,\tau} Z_p],$$

where $\sigma_{n,\tau}(x)$ is the same as above and Z_p is now the $(1 - p)$ percentile of $\sup_{\tau \in \mathcal{T}} |G_b(\tau)|$.

The confidence bands are typically wider near the tails of the conditional distribution, because the change in $f_{Y|X}(Q(\tau|x)|x)$ often dominates the rest. To compute the bands, the terms d_τ and $d_{b,\tau}$ can be estimated using local quadratic or cubic regressions, the conditional density $f_{Y|X}(Q(\tau|x)|x)$ can be estimated from the conditional quantile process and the critical values of $\sup_{\tau \in \mathcal{T}} |G(\tau)/\sqrt{f_X(x)}|$ and $\sup_{\tau \in \mathcal{T}} |G_b(\tau)/\sqrt{f_X(x)}|$ can be estimated via simulations. We now examine the latter simulation in detail. Consider the boundary case. From Theorems 1 and 2,

$$\frac{\ell'_1 N_x(\tau)^{-1} (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i^0(\tau) \leq 0)) z_{i,\tau} K_{i,\tau}}{f_X(x)} \Rightarrow \frac{G_b(\tau)}{\sqrt{f_X(x)}}.$$

The left hand side is conditionally pivotal. Using $(nh_{n,\tau}^d)^{-1} \sum_{i=1}^n z_{i,\tau} z'_{i,\tau} K_{i,\tau} \rightarrow^p f_X(x) N_x(\tau)$ and the Skorohod representation, we have

$$\ell'_1 \left((nh_{n,\tau}^d)^{-1} \sum_{i=1}^n z_{i,\tau} z'_{i,\tau} K_{i,\tau} \right)^{-1} (nh_{n,\tau}^d)^{-1/2} \times \sum_{i=1}^n (\tau - 1(u_i - \tau \leq 0)) z_{i,\tau} K_{i,\tau} \Rightarrow \frac{G_b(\tau)}{\sqrt{f_X(x)}}, \tag{19}$$

where u_i are i.i.d. Uniform(0, 1) random variables and independent of $\{x_i\}_{i=1}^n$. This suggests the following simple procedure: (i) Simulate the left hand side of (19) by drawing i.i.d. Uniform(0, 1) random variables and keeping $K_{i,\tau}$ and $z_{i,\tau}$ fixed; (ii) Repeat the above step for a large number of replications; (iii) Compute their absolute values to estimate the distribution of $\sup_{\tau \in \mathcal{T}} |G_b(\tau)/\sqrt{f_X(x)}|$. The critical values of this distribution can then be combined with estimates of $d_{b,\tau}$ and $f_{Y|X}(Q(\tau|x)|x)$ to form a confidence band. Because (19) automatically accounts for the relative location of the evaluation point from the boundary, it allows us to easily handle situations where the support of f_X has a complicated structure. Note that the resulting critical values are asymptotically valid even if x is an interior point.

Remark 6. The above simulation procedure is inspired by Parzen et al. (1994) and Chernozhukov et al. (2009). The former paper exploited the conditional pivotal property to obtain a resampling method, while the latter used such a property to obtain finite sample confidence bands for quantile regression models. The current paper is the first that applies such a property to the local regression setting and, more importantly, to provide valid confidence bands for both interior and boundary points.

Now consider estimating the bias. From Theorem 1,

$$d_{b,\tau} = \ell'_1 \{f_X(x) N_x(\tau)\}^{-1} \times \left\{ \frac{1}{2} f_X(x) \int_{\mathcal{D}_{x,\tau}} u' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} u \begin{bmatrix} 1 \\ u \end{bmatrix} K(u) du \right\}.$$

The term inside the first curly bracket can be consistently estimated using $(nh_{n,\tau}^d)^{-1} \sum_{i=1}^n z_{i,\tau} z'_{i,\tau} K_{i,\tau}$. The second curly bracket is the limit of $-h_{n,\tau}^{-2} (nh_{n,\tau}^d)^{-1} \sum_{i=1}^n e_i(\tau) z_{i,\tau} K_{i,\tau}$, where

$$e_i(\tau) = -\frac{1}{2} \left(\frac{x_i - x}{h_{n,\tau}} \right)' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \left(\frac{x_i - x}{h_{n,\tau}} \right) h_{n,\tau}^2 + o(h_{n,\tau}^2).$$

Therefore, $d_{b,\tau}$ can be estimated using

$$\ell'_1 \left((nh_{n,\tau}^d)^{-1} \sum_{i=1}^n z_{i,\tau} z'_{i,\tau} K_{i,\tau} \right)^{-1} \times \left(-h_{n,\tau}^{-2} (nh_{n,\tau}^d)^{-1} \sum_{i=1}^n e_i(\tau) z_{i,\tau} K_{i,\tau} \right). \tag{20}$$

In (20), the only unknown is $\partial^2 Q(\tau|x)/\partial x \partial x'$. This can be estimated using a global or local polynomial regression, although caution is needed because the estimate can be imprecise. As before, (20) automatically accounts for the relative location of the evaluation point from the boundary, allowing us to easily handle situations where the support of f_X has a complicated structure.

Remark 7. We use the formulas (19) and (20) throughout in the simulation study and the empirical application when constructing the confidence band. In the simulation section, it will emerge that treating the estimated bias term as the true value can cause the confidence band (18) to have substantial undercoverage (see Panel (b) in Table 2). Some modifications that can reflect the estimation uncertainty are desirable. This leads to the modified confidence band considered below.

We suggest the following simple modified confidence band, where the idea is to allow for, but do not force, a bias adjustment: (Consider the interior point case; the boundary point case can be handled in the same way.)

$$[(\tilde{\alpha}^*(\tau) - d_\tau^+ h_{n,\tau}^2) - \sigma_{n,\tau}(x) Z_p, (\tilde{\alpha}^*(\tau) - d_\tau^- h_{n,\tau}^2) + \sigma_{n,\tau} Z_p], \tag{21}$$

where $d_\tau^+ = \max(d_\tau, 0)$ and $d_\tau^- = \min(d_\tau, 0)$. This modified band has the same or higher coverage relative to not making any bias adjustment (i.e., by setting d_τ to zero, as often done in the literature). This is preferable when the bias is small. It also has the same or higher coverage than the conventional band (18), which is preferable when the bias is large. Here, an important concern is whether the band will be too wide to be informative. Our simulation evidence in Section 7 suggests otherwise. This is because when the curvature of the conditional function (d_τ) is high, the proposed bandwidth selection rule can deliver a small bandwidth, therefore the value of $d_\tau h_{n,\tau}^2$ remains modest. Consequently, (21) is typically only mildly wider than (18).

6.2. Uniform confidence band for the difference between two quantile processes

Suppose there are two groups (say the treatment and control group) of observations with conditional quantile processes $Q_1(\tau|x)$ and $Q_2(\tau|x)$. The issue of interest is to obtain a confidence band for $\delta(\tau|x) = Q_1(\tau|x) - Q_2(\tau|x)$ with $\tau \in \mathcal{T}$.

Without loss of generality, suppose x is an interior point for both groups. Then,

Group 1:

$$\sqrt{n_1 h_{n,\tau,1}^d} f_{Y|X,1}(Q_1(\tau|x)|x) (\tilde{\alpha}_1^*(\tau) - Q_1(\tau|x) - d_{\tau,1} h_{n,\tau,1}^2) \Rightarrow \frac{G^{(1)}(\tau)}{\sqrt{f_{X,1}(x)}}. \tag{22}$$

Group 2:

$$\sqrt{n_2 h_{n,\tau,2}^d} f_{Y|X,2}(Q_2(\tau|x)|x) (\tilde{\alpha}_2^*(\tau) - Q_2(\tau|x) - d_{\tau,2} h_{n,\tau,2}^2) \Rightarrow \frac{G^{(2)}(\tau)}{\sqrt{f_{X,2}(x)}}, \tag{23}$$

where n_1 and n_2 are the number of observations for the two groups and $G^{(1)}(\tau)$ and $G^{(2)}(\tau)$ are independent copies of $G(\tau)$. All the quantities are group dependent as signaled by their sub-indices. Assume the ratio n_1/n_2 converges to a positive constant. Then, by (16), there exists some $r(\tau) > 0$ such that $\sqrt{n_1 h_{n,\tau}^d} / \sqrt{n_2 h_{n,\tau}^d} \rightarrow r(\tau)$. Multiplying (23) by

$$\frac{\sqrt{n_1 h_{n,\tau}^d} f_{Y|X,1}(Q_1(\tau|x)|x)}{\sqrt{n_2 h_{n,\tau}^d} f_{Y|X,2}(Q_2(\tau|x)|x)} \tag{24}$$

and subtracting it from (22), we obtain

$$\begin{aligned} & \sqrt{n_1 h_{n,\tau}^d} f_{Y|X,1}(Q_1(\tau|x)|x) \left| (\tilde{\alpha}_1^*(\tau) - d_{\tau,1} h_{n,\tau}^2 \right. \\ & \quad \left. - \tilde{\alpha}_2^*(\tau) + d_{\tau,2} h_{n,\tau}^2) - \delta(\tau|x) \right| \\ \Rightarrow & \left| \frac{G^{(1)}(\tau)}{\sqrt{f_{X,1}(x)}} - r(\tau) \frac{f_{Y|X,1}(Q_1(\tau|x)|x)}{f_{Y|X,2}(Q_2(\tau|x)|x)} \frac{G^{(2)}(\tau)}{\sqrt{f_{X,2}(x)}} \right|. \end{aligned} \tag{25}$$

Consequently, the following procedure can be used to construct a confidence band for $\delta(\tau|x)$. First, simulate the supremum of (25) for a large number of replications. Compute the $(1 - p)$ percentile of the resulting distribution. Call it $c_p(x)$. Next, compute the confidence band for $\delta(\tau|x)$ as

$$\begin{aligned} & (\tilde{\alpha}_1^*(\tau) - d_{\tau,1} h_{n,\tau}^2 - \tilde{\alpha}_2^*(\tau) + d_{\tau,2} h_{n,\tau}^2) \\ & \pm \frac{c_p(x)}{\sqrt{n_1 h_{n,\tau}^d} f_{Y|X,1}(Q_1(\tau|x)|x)}. \end{aligned} \tag{26}$$

As in Corollary 2, the above band is typically wider near the tails of the conditional distribution. This follows from the multiplication of (23) by (24). Without such a multiplication, the bands will tend to be narrower in the tails because there the bandwidth tends to be greater.

6.3. Hypothesis tests

We revisit the three examples in Section 2. Without loss of generality, assume the evaluation points are interior points except for the third example. To ease the exposition, suppose the bandwidth $h_{n,\tau}$ ($\tau \in \mathcal{T}$) is used to estimate all the conditional quantile processes. To be consistent with the notation used in Section 2, we let $\hat{Q}(\tau|x)$ denote $\tilde{\alpha}^*(\tau)$ when the evaluation point is x .

Example 1 (QTE, Continued). Let $\hat{Q}(\tau|1, z)$ and $\hat{Q}(\tau|0, z)$ be the estimates of $Q(\tau|1, z)$ and $Q(\tau|0, z)$ using the treatment and the control group, respectively. Treatment significance can be tested using a Kolmogorov–Smirnov (KS) type test:

$$\sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| \hat{Q}(\tau|1, z) - \hat{Q}(\tau|0, z) \right|. \tag{27}$$

Let $x_1 = (1, z)$, $x_2 = (0, z)$, and define $G_x^{(j)}(\tau) = G^{(j)}(\tau) / (\sqrt{f_X(x)} f_{Y|X}(Q(\tau|x)|x))$ with $G^{(j)}(\tau)$ ($j = 1, 2$) being independent copies of $G(\tau)$. Assume Assumptions 1–5 hold at x_1 and x_2 and $\partial^2 Q(\tau|x_1) / \partial x \partial x' = \partial^2 Q(\tau|x_2) / \partial x \partial x'$. Then, the test has the following null limiting distribution $\sup_{\tau \in \mathcal{T}} |G_{x_1}^{(1)}(\tau) - G_{x_2}^{(2)}(\tau)|$, whose critical values can be consistently estimated by simulating independently its first and second component using the algorithm described in Section 6.1. The treatment homogeneity hypothesis can be tested using

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| \hat{Q}(\tau|1, z) - \hat{Q}(\tau|0, z) \right| \\ & - \int_{\mathcal{T}} \left\{ \hat{Q}(\tau|1, z) - \hat{Q}(\tau|0, z) \right\} d\tau, \end{aligned} \tag{28}$$

whose null limiting distribution is given by $\sup_{\tau \in \mathcal{T}} |G_{x_1}^{(1)}(\tau) - G_{x_2}^{(2)}(\tau) - \int_{\mathcal{T}} \{G_{x_1}^{(1)}(\tau) - G_{x_2}^{(2)}(\tau)\} d\tau|$, whose critical values can be obtained in a similar manner. The quantile treatment equality hypothesis between z_1 and z_2 can be tested using

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| \hat{Q}(\tau|1, z_1) - \hat{Q}(\tau|0, z_1) \right. \\ & \quad \left. - \left\{ \hat{Q}(\tau|1, z_2) - \hat{Q}(\tau|0, z_2) \right\} \right|. \end{aligned} \tag{29}$$

Its null limiting distribution is given by $\sup_{\tau \in \mathcal{T}} |G_{x_{11}}^{(1)}(\tau) - G_{x_{21}}^{(2)}(\tau) - \{G_{x_{12}}^{(1)}(\tau) - G_{x_{22}}^{(2)}(\tau)\}|$, where $x_{1j} = (1, z_j)$ and $x_{2j} = (0, z_j)$ for $j = 1, 2$ and the distributions involved are mutually independent.

Example 2 (Stochastic Dominance, Continued). Let $\hat{Q}_1(\tau|x)$ and $\hat{Q}_2(\tau|x)$ be the estimates of $Q_1(\tau|x)$ and $Q_2(\tau|x)$ respectively. The first-order conditional stochastic dominance can be tested against the non-dominance alternative using a signed KS test:

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| 1 \left(\hat{Q}_1(\tau|x) - \hat{Q}_2(\tau|x) \leq 0 \right) \right. \\ & \quad \left. \times \left(\hat{Q}_1(\tau|x) - \hat{Q}_2(\tau|x) \right) \right|. \end{aligned} \tag{30}$$

Assume Assumptions 1–5 hold for both samples and $\partial^2 Q_1(\tau|x) / \partial x \partial x' = \partial^2 Q_2(\tau|x) / \partial x \partial x'$. Then, under the least favorable null hypothesis, the test converges to

$$\sup_{\tau \in \mathcal{T}} \left| 1 \left(G_x^{(1)}(\tau) - G_x^{(2)}(\tau) \leq 0 \right) \left(G_x^{(1)}(\tau) - G_x^{(2)}(\tau) \right) \right|,$$

where $G_x^{(j)}$ ($j = 1, 2$) has the same definition as in Example 1, but with $\sqrt{f_X(x)} f_{Y|X}(Q(\tau|x)|x)$ replaced by $\sqrt{f_{X,j}(x)} f_{Y|X,j}(Q_j(\tau|x)|x)$ ($j = 1, 2$) to allow the quantities to be group dependent. The second-order dominance can be tested using

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| 1 \left(\int_{\varepsilon}^{\tau} \left(\hat{Q}_1(s|x) - \hat{Q}_2(s|x) \right) ds \leq 0 \right) \right. \\ & \quad \left. \times \int_{\varepsilon}^{\tau} \left(\hat{Q}_1(s|x) - \hat{Q}_2(s|x) \right) ds \right|. \end{aligned}$$

Then, under the least favorable null hypothesis, the test converges to

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \left| 1 \left(\int_{\varepsilon}^{\tau} \left(G_x^{(1)}(\tau) - G_x^{(2)}(\tau) \right) d\tau \leq 0 \right) \right. \\ & \quad \left. \times \int_{\varepsilon}^{\tau} \left(G_x^{(1)}(\tau) - G_x^{(2)}(\tau) \right) d\tau \right|. \end{aligned}$$

In the above construction, the lower limit of the integral is some positive constant ε instead of zero. This allows the conditional distributions to have unbounded support, at the cost of possibly sacrificing some power if the main differences between two distributions lies in the lower tails. Again, the critical values can be consistently estimated via simulations.

Example 3 (SRD, Continued). Let $\hat{Q}(\tau|c_+)$ and $\hat{Q}(\tau|c_-)$ be the estimates of $\lim_{x \downarrow c} Q(\tau|x)$ and $\lim_{x \uparrow c} Q(\tau|x)$, respectively. The treatment significance hypothesis can be tested using

$$\sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| \hat{Q}(\tau|c_+) - \hat{Q}(\tau|c_-) \right|.$$

The treatment homogeneity hypothesis can be tested using $\sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| \hat{Q}(\tau|c_+) - \hat{Q}(\tau|c_-) - \int_{\mathcal{T}} \{ \hat{Q}(\tau|c_+) - \hat{Q}(\tau|c_-) \} d\tau \right|$. The nonnegative treatment effects hypothesis can be tested using $\sup_{\tau \in \mathcal{T}} \sqrt{nh_{n,\tau}^d} \left| 1 \left(\hat{Q}(\tau|c_+) - \hat{Q}(\tau|c_-) \leq 0 \right) \left(\hat{Q}(\tau|c_+) - \hat{Q}(\tau|c_-) \right) \right|$. Their limiting distributions are analogous to those in Examples 1 and 2, hence the detail is omitted.

Table 1
Root mean integrated squared error (RMISE).

Models	n = 250			n = 500			n = 1000		
	(0.5, 0.5)	(0.75, 0.75)	(0.9, 0.9)	(0.5, 0.5)	(0.75, 0.75)	(0.9, 0.9)	(0.5, 0.5)	(0.75, 0.75)	(0.9, 0.9)
Model 1									
Proposed 1	0.03759	0.07590	0.13775	0.02635	0.05702	0.09151	0.01914	0.04081	0.06612
Proposed 2	0.03759	0.07590	0.13760	0.02635	0.05702	0.09151	0.01914	0.04081	0.06612
Local Linear	0.03759	0.07590	0.13791	0.02635	0.05702	0.09154	0.01914	0.04081	0.06612
Crossing	(0.000)	(0.074)	(0.748)	(0.000)	(0.000)	(0.302)	(0.000)	(0.000)	(0.074)
QR	0.03191	0.05791	0.07686	0.02280	0.04213	0.05245	0.01643	0.02842	0.03605
Model 2									
Proposed 1	0.14561	0.22266	0.29231	0.11908	0.18039	0.22324	0.09612	0.13359	0.16635
Proposed 2	0.14561	0.22265	0.29166	0.11908	0.18039	0.22317	0.09612	0.13359	0.16634
Local Linear	0.14561	0.22267	0.29327	0.11908	0.18040	0.22340	0.09612	0.13359	0.16637
Crossing	(0.078)	(0.118)	(0.740)	(0.010)	(0.066)	(0.618)	(0.000)	(0.010)	(0.358)
QR	0.49038	0.41946	0.21568	0.48823	0.41356	0.18802	0.49100	0.41509	0.18165
Model 3									
Proposed 1	0.14170	0.15985	0.33123	0.10707	0.12180	0.23555	0.08391	0.08776	0.18468
Proposed 2	0.14170	0.15985	0.33103	0.10707	0.12180	0.23541	0.08391	0.08776	0.18466
Local Linear	0.14171	0.15985	0.33203	0.10707	0.12180	0.23572	0.08391	0.08776	0.18471
Crossing	(0.100)	(0.092)	(0.882)	(0.028)	(0.002)	(0.780)	(0.004)	(0.000)	(0.362)
QR	0.52300	0.15640	0.60453	0.52676	0.12593	0.60823	0.52451	0.11409	0.60268

The domain of the conditional quantile process is $\mathcal{T} = [0.2, 0.8]$ and the number of the estimated quantiles is 30. We report the average of the RMISE over 500 simulation replications. “Proposed 1” denotes the first proposed estimator. “Proposed 2” is the second proposed estimator. “Local linear” is the quantile-by-quantile application of the local linear estimator. “QR” is obtained from the conventional linear quantile regression. “Crossing” denotes the fraction of simulations in which “Local linear” has quantile-crossing.

7. Monte Carlo experiments

We focus on two issues: (i) the performance of the proposed estimators relative to some other estimators, and (ii) the property of the uniform confidence band. We consider the following three models whose conditional quantile functions are

Model 1: $Q(\tau|x) = x_1 - x_2 + (0.5x_1 + 0.3x_2)Q_{e_1}(\tau)$.

Model 2: $Q(\tau|x) = (0.5 + 2x_1 + \sin(2\pi x_1 - 0.5)) + x_2Q_{e_1}(\tau)$.

Model 3: $Q(\tau|x) = \log(x_1x_2) + 1/(1 + \exp(-x_1Q_{e_1}(\tau) - x_2Q_{e_2}(\tau))) + x_2Q_{e_1}(\tau)$.

The regressors x_1 and x_2 are mutually independent and are i.i.d. $U(0, 1)$. The error terms e_1 and e_2 are i.i.d. $N(0, 1)$ and $U(0, 1)$ respectively. Model 1 is a linear location–scale model. Model 2 is a similar model but with nonlinearity in the location. Model 3 is a fairly complicated nonlinear model. The conditional quantile functions in Models 2 and 3 exhibit significant curvature in x , which can pose substantial challenges for estimation and inference.

Other aspects of the simulation design are as follows. We consider three evaluation points: $x = (0.5, 0.5)$, $(0.75, 0.75)$ and $(0.9, 0.9)$. The latter can be viewed as a boundary point, since the selected bandwidth at this point is typically greater than 0.1. The sample sizes are: $n = 250, 500$ and 1000 . Given that the experiment involves estimating the quantile process nonparametrically at $x = (0.9, 0.9)$, $n = 250$ should be viewed as a very small sample size. We set $\mathcal{T} = [0.2, 0.8]$ unless stated otherwise. The number of quantiles (m) equals 10, 20 and 30. The kernel function is the product of univariate Epanechnikov kernels. All subsequent results are based on 500 replications.

The bandwidths are estimated in three steps.

STEP 1. Obtain a set of pilot bandwidths for calculating the relevant quantities in (16). This is done by first obtaining a bandwidth for the local median regression using leave-one-out cross validation, and then relating it to the other quantiles using (17) following Yu and Jones (1998).

STEP 2. Estimate $\text{tr}(\partial^2 Q(\tau|x)/\partial x \partial x')$ and $f_{Y|X}(\cdot|x)$. For the former, we apply a local cubic regression with the bandwidth determined via leave-one-out cross validation. For the latter, we use

$$\hat{f}_{Y|X}(z|x) = \int \frac{1}{h_{yx}} K_1\left(\frac{z-y}{h_{yx}}\right) d\hat{F}(y|x), \tag{31}$$

where h_{yx} is a bandwidth parameter and $K_1(\cdot)$ is a univariate Epanechnikov kernel and $\hat{F}(y|x) = \sup\{\tau \in (0, 1) | \hat{Q}(\tau|x) \leq y\}$ with $\hat{Q}(\tau|x)$ equal to our proposed two-step estimator $\hat{\alpha}^*(\tau)$ computed using the bandwidth selected in Step 1. To implement (31), we sample from $\hat{F}(y|x)$ and apply kernel smoothing to the resulting draws. The value of h_{yx} is set to $2\hat{h}_{yx}$, where \hat{h}_{yx} is determined by Silverman’s rule-of-thumb method.

STEP 3. Apply the estimates from Step 2 to compute the optimal bandwidth for the median using (16) and that for the other quantiles using (17).

7.1. The relative performance of different estimators

We contrast the finite sample performance of the proposed estimators (Proposed 1 and 2) with two alternatives. The first alternative is the classical estimator in Koenker and Bassett (1978), labeled as QR. This comparison is used to illustrate the gain or loss from using a nonparametric specification. The second alternative is the standard quantile-by-quantile local linear estimator (labeled Local linear), which is included to illustrate the effect of imposing the monotonicity. The same bandwidth selection rule is used throughout the comparison.

The criterion is the RMISE criterion with $RMISE = \sqrt{m^{-1} \sum_{j=1}^m |\hat{Q}(\tau_j|x) - Q(\tau_j|x)|^2}$, where m is the number of quantile entering the estimation. Table 1 reports the results for $m = 30$. We obtained similar findings with $m = 10$ and 20. They are omitted to save space.

First consider the proposed estimators and QR. When the underlying model is linear (Model 1), the proposed estimators show a moderately higher RMISE. The efficiency loss is not high because the bandwidth selection rule delivers wide bandwidths in this case. When nonlinearity (Models 2 and 3) is present, the nonparametric estimators show substantially superior performance overall.

Next consider the proposed estimators and the standard quantile-by-quantile local linear estimator. For the latter, we also report the fraction of the simulation runs in which the estimates exhibit quantile crossings. The results show that quantile crossings are relatively infrequent at interior points, however they occur frequently at points close to the boundary. Specifically, when $n = 250$, the fractions of crossings are between 0 and 0.100 for

Table 2
Coverage ratio of the 90% confidence band.

Models	x	$m = 10$			$m = 20$			$m = 30$		
		$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
(a) Modified Bias Correction										
1	(0.50, 0.50)	0.892	0.884	0.868	0.884	0.878	0.896	0.878	0.882	0.906
	(0.75, 0.75)	0.906	0.912	0.916	0.894	0.908	0.908	0.894	0.926	0.934
	(0.90, 0.90)	0.908	0.918	0.940	0.888	0.922	0.928	0.882	0.940	0.952
2	(0.50, 0.50)	0.720	0.754	0.790	0.764	0.770	0.774	0.760	0.746	0.784
	(0.75, 0.75)	0.810	0.826	0.880	0.804	0.822	0.874	0.802	0.798	0.910
	(0.90, 0.90)	0.780	0.848	0.916	0.794	0.884	0.898	0.800	0.844	0.914
3	(0.50, 0.50)	0.890	0.934	0.942	0.870	0.928	0.946	0.860	0.914	0.932
	(0.75, 0.75)	0.828	0.888	0.874	0.866	0.872	0.878	0.858	0.854	0.872
	(0.90, 0.90)	0.736	0.806	0.860	0.770	0.810	0.854	0.754	0.848	0.842
(b) Conventional Bias Correction										
1	(0.50, 0.50)	0.490	0.454	0.454	0.414	0.450	0.442	0.402	0.432	0.430
	(0.75, 0.75)	0.712	0.740	0.744	0.756	0.720	0.734	0.752	0.746	0.768
	(0.90, 0.90)	0.392	0.460	0.492	0.382	0.400	0.422	0.388	0.412	0.460
2	(0.50, 0.50)	0.690	0.734	0.776	0.720	0.748	0.766	0.716	0.730	0.772
	(0.75, 0.75)	0.766	0.780	0.842	0.748	0.786	0.836	0.748	0.766	0.876
	(0.90, 0.90)	0.690	0.804	0.848	0.704	0.820	0.826	0.700	0.788	0.852
3	(0.50, 0.50)	0.790	0.826	0.824	0.766	0.798	0.846	0.748	0.784	0.836
	(0.75, 0.75)	0.590	0.586	0.596	0.618	0.590	0.546	0.598	0.544	0.566
	(0.90, 0.90)	0.592	0.714	0.766	0.578	0.722	0.744	0.568	0.742	0.734

Results are based on 500 Replications; m is the number of quantiles entering the estimation; n is the sample size.

$x = (0.5, 0.5)$, 0.074 and 0.118 for $x = (0.75, 0.75)$, and 0.740 and 0.882 for $x = (0.9, 0.9)$. When $n = 1000$, the values are between 0 and 0.004, 0 and 0.010, and 0.074 and 0.362, respectively. Concerning $RMISE$, the proposed estimators typically perform better, but the difference is always small. This is consistent with the result that these estimators are first-order asymptotically equivalent. Therefore, in the current simulations, the monotonicization serves as a way to ensure a coherent estimate, but not a way to substantially improve the precision in finite samples.

Now compare the two proposed estimators. They have very similar $RMISE$, confirming the result in Theorem 3 that these two estimators are first-order asymptotically equivalent. Interestingly, the $RMISE$ of the rearranged estimator tends to be slightly smaller. However, the difference is too small to prefer one estimator to another.

7.2. Properties of the uniform confidence band

We examine the following two issues: whether the modified confidence band (21) shows a significant improvement over the conventional one (18), and whether the improvement comes at the cost of a substantially wider band. The results reported are based on the first proposed estimator.

Table 2 reports the coverage ratios. Panel (a) corresponds to the modified bias adjustment. For Model 1, the modification delivers adequate coverage for all the cases considered. For Models 2 and 3, undercoverage exists in some cases when n is small. However, the coverage becomes adequate when n is increased to 1000, except for Model 2 with $x = (0.5, 0.5)$. To examine the reason for the undercoverage in the latter case, we sequentially replace the quantities appearing in (21) with their true values. First, when plugging in the true conditional density, the coverage does not improve. Second, when plugging in the infeasible optimal bandwidth while keeping the estimated conditional density, the coverage ratio reaches 0.922 when $n = 1000$. Finally, we plug in both the infeasible bandwidth and the true conditional density and obtain 0.938. Therefore, the undercoverage is due to oversmoothing. Panel (b) in Table 2 reports results under conventional bias adjustment. For Model 1, the coverage can be quite low even when $n = 1000$. Because the true bias is zero, the conventional bias correction can only have detrimental effects on the coverage ratio. Further, because the selected

bandwidths are large, such effects can be severe. For the other two models, the improvement from using the modified adjustment is more important when the sample size is small or when the x is close to the boundary of the data support.

The results in Table 2 can be further contrasted with two additional cases: the coverage ratio of the uniform band without bias adjustment and that of the pointwise confidence band with modified bias adjustment. We summarize the results below while omitting the details. Without any bias adjustment, the coverage ratios for Model 1 are indeed close to the nominal rate. However, for Model 2, the ratios are quite low. For model 3, the difference is not as drastic as in Model 2, but the modified bias adjustment still offers improvement. Meanwhile, the pointwise band has severe undercoverage, confirming that it should not be used in problems involving quantile processes.

We now examine the relative width of these confidence bands. Table 3 summarizes the widths of the confidence bands under modified and conventional bias adjustment, measured at two representative quantiles $\tau = 0.5$ and 0.8. The values are based on $n = 500$. The uniform bands with modified bias adjustment are overall only mildly wider. This pattern holds even when the quantile function has a substantial curvature, say in Model 2 with $x = (0.75, 0.75)$. Intuitively, when the curvature is large, the selected bandwidth tends to be small, therefore safeguarding the effect of the bias on the width of the confidence band. We have also compared the uniform band with the pointwise band under modified bias adjustment. The former are overall mildly wider. The ratios are between 1.25 and 1.35 at $\tau = 0.5$ and 1.25 and 1.63 at $\tau = 0.8$. This suggests that such bands can deliver valid inference for the quantile process without losing informativeness about individual quantiles.

8. An illustrative example: effects of class size reduction

We consider a dataset from an experiment known as Project STAR (Student–Teacher Achievement Ratio). In the late 1980s, Tennessee conducted a randomized controlled experiment to measure the effect of class size reduction on student achievements. Students were randomly assigned to a small class with target sizes of 13 to 17, a regular size class of 22 to 25 or a regular size class with a teaching aide. Teachers within a given school were also

Table 3
The length of the 90% uniform confidence band.

Models	x	$m = 10$		$m = 20$		$m = 30$	
		$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.5$	$\tau = 0.8$
(a) Modified Bias Correction							
1	(0.50, 0.50)	0.138 (0.027)	0.186 (0.033)	0.141 (0.027)	0.190 (0.036)	0.142 (0.027)	0.192 (0.033)
	(0.75, 0.75)	0.299 (0.044)	0.395 (0.064)	0.309 (0.045)	0.409 (0.066)	0.317 (0.047)	0.420 (0.071)
	(0.90, 0.90)	0.595 (0.141)	0.814 (0.211)	0.625 (0.132)	0.834 (0.198)	0.624 (0.143)	0.853 (0.200)
2	(0.50, 0.50)	0.372 (0.047)	0.476 (0.088)	0.387 (0.050)	0.503 (0.097)	0.390 (0.051)	0.502 (0.101)
	(0.75, 0.75)	0.668 (0.096)	0.925 (0.222)	0.684 (0.106)	0.953 (0.213)	0.688 (0.107)	0.959 (0.223)
	(0.90, 0.90)	1.070 (0.261)	1.451 (0.482)	1.103 (0.300)	1.504 (0.504)	1.087 (0.263)	1.523 (0.515)
3	(0.50, 0.50)	0.541 (0.071)	0.688 (0.133)	0.562 (0.076)	0.703 (0.128)	0.563 (0.075)	0.703 (0.130)
	(0.75, 0.75)	0.594 (0.113)	0.725 (0.139)	0.617 (0.113)	0.769 (0.176)	0.615 (0.105)	0.776 (0.157)
	(0.90, 0.90)	1.209 (0.306)	1.513 (0.439)	1.231 (0.308)	1.540 (0.498)	1.259 (0.302)	1.589 (0.455)
(b) Conventional Bias Correction							
1	(0.50, 0.50)	0.121 (0.021)	0.161 (0.030)	0.123 (0.021)	0.167 (0.032)	0.124 (0.020)	0.167 (0.029)
	(0.75, 0.75)	0.266 (0.034)	0.359 (0.054)	0.277 (0.036)	0.372 (0.059)	0.283 (0.037)	0.382 (0.059)
	(0.90, 0.90)	0.504 (0.112)	0.679 (0.168)	0.526 (0.111)	0.693 (0.161)	0.523 (0.117)	0.711 (0.176)
2	(0.50, 0.50)	0.319 (0.043)	0.421 (0.087)	0.334 (0.047)	0.448 (0.098)	0.337 (0.048)	0.446 (0.102)
	(0.75, 0.75)	0.570 (0.086)	0.824 (0.224)	0.587 (0.096)	0.852 (0.215)	0.593 (0.097)	0.860 (0.226)
	(0.90, 0.90)	0.971 (0.255)	1.357 (0.478)	1.001 (0.288)	1.408 (0.502)	0.986 (0.257)	1.425 (0.510)
3	(0.50, 0.50)	0.464 (0.066)	0.607 (0.132)	0.485 (0.069)	0.623 (0.127)	0.486 (0.069)	0.625 (0.127)
	(0.75, 0.75)	0.523 (0.090)	0.655 (0.130)	0.544 (0.094)	0.702 (0.172)	0.543 (0.084)	0.707 (0.150)
	(0.90, 0.90)	1.121 (0.301)	1.425 (0.431)	1.143 (0.302)	1.443 (0.495)	1.168 (0.299)	1.500 (0.456)

The sample size is 500 and results are based on 500 replications. We report the average length of the uniform confidence bands. The standard deviations are in parenthesis.

randomly assigned to one of the three types of classes. Over the four-year period from kindergarten to the third grade, more than 11,000 students at 79 schools participated in the program.

We focus on a sample that was previously analyzed many times. It consists of the students who entered the kindergarten when Project STAR started in the fall of 1985. After excluding students in regular sized classes with aide, this sample has 1738 and 2004 observations in the treatment and control group, respectively. Focusing on this sample allows us to compare our results to previous studies, in particular that of Krueger (1999), to show that the new findings are due to the new framework rather than differences in the samples considered.

The existing studies using this dataset mainly considered the average effect. The results broadly agree that class size is an important determinant of student achievement. Our method provides further information by studying the distribution effect in a flexible nonparametric setting. First, we find that students in the upper quantiles of the test score distribution benefit more from the class size reduction. Second, using teachers' experience as a conditioning variable, we find that the effect of the class size reduction is strongest for the classes taught by moderately experienced teachers (i.e., 6–8 years of experience). We also conduct hypotheses tests for treatment significance, homogeneity, equality and first order stochastic dominance. The results reconfirm these two findings.

8.1. Heterogeneous effects of small class sizes

The detailed model specifications are as follows. Student achievement is measured by the sum of math and reading scores of the Stanford Achievement Test after finishing the kindergarten. The total score ranges from 635 to 1253. Because the goal is to estimate distributional effects, we do not use percentile ranks as in Krueger (1999). The conditioning variable is teachers' experience, whose 10th to 90th percentiles are: 2, 4, 6, 8, 9, 11, 12, 14, 17. The process of quantile treatment effects is defined as the difference between the two conditional quantile processes (i.e., for small and regular class sizes) over the range $\tau \in [0.1, 0.9]$. All the results are obtained using the first proposed estimator.

Fig. 1 presents the estimates of the quantile treatment effects (the dashed black lines) along with their 90% uniform confidence bands (the shaded areas). Two interesting patterns emerge. First, the effects are heterogeneous, with the students in the upper quantiles benefiting more than those in the lower quantiles of the conditional distribution. This pattern is particularly pronounced when the classes are taught by less experienced teachers (i.e., Fig. 1(a)–(d)). There, the differences between the 90th and 10th percentiles are 23.2, 23.5, 14.2 and 7.5 points, respectively. Second, the effect of class size reduction depends strongly on teachers' experience. Specifically, the effect is large and significant when the classes are taught by teachers with 6–8 years of

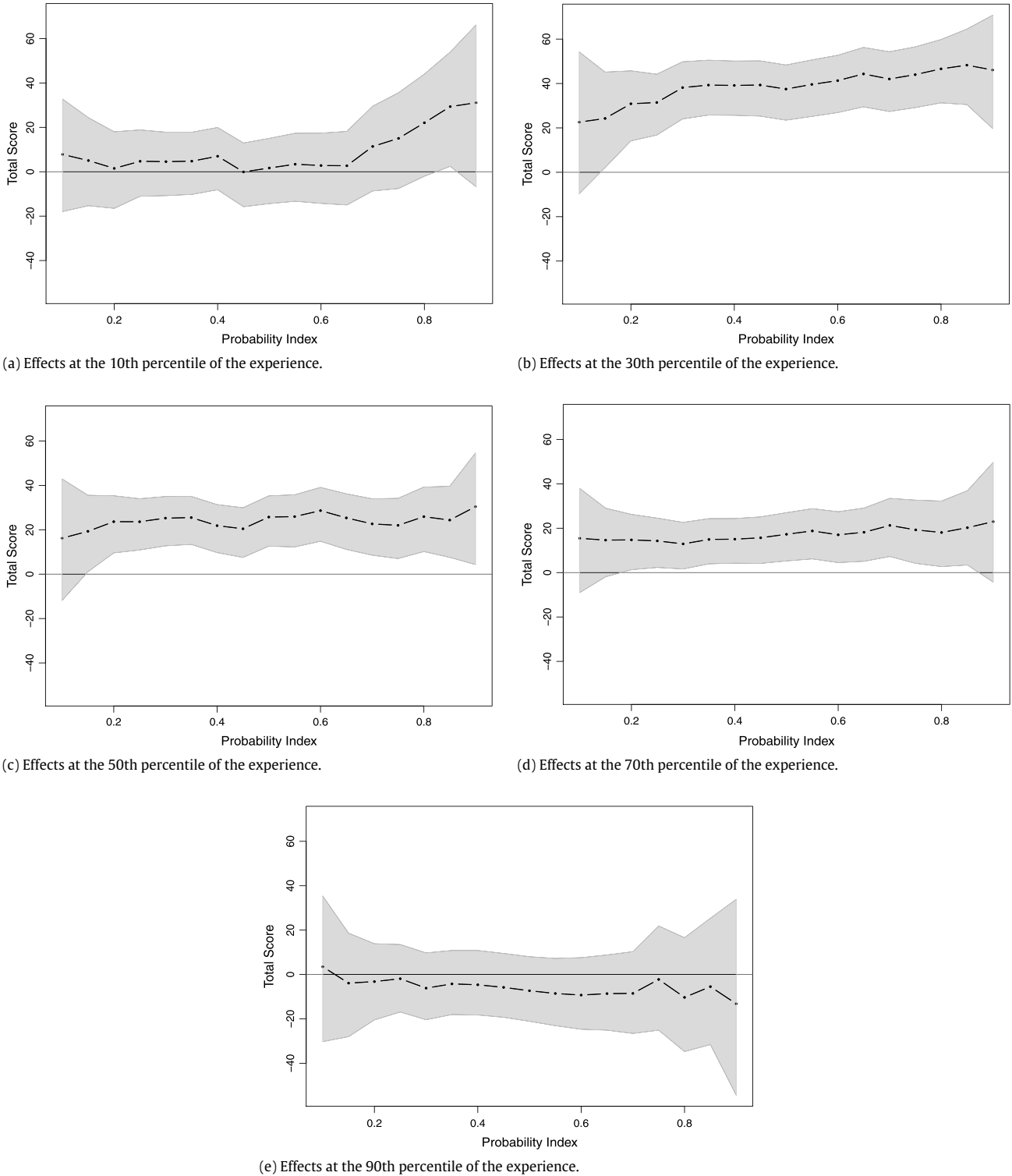


Fig. 1. Processes of quantile treatment effects. Each figure shows quantile dependent treatment effects over the range [0.1, 0.9]. For the values of the covariate (i.e., teacher’s experience) we consider those corresponding to the 10th, 30th, 50th, 70th and 90th of the conditioning distribution, each of which is reported in a separate figure.

experience (i.e., Fig. 1(b)) but small or even insignificant for teachers with little or much experience. The maximum difference across teachers’ experience can be as high as 45.1 points when evaluated at $\tau = 0.5$. This is fairly significant from an economic point of view. The above two findings suggest that the proposed method can be valuable in discovering and quantifying the heterogeneity of the treatment responses in empirically relevant situations.

Note that the second finding is in sharp contrast with that of Krueger (1999), who documented that “the STAR data suggest

that measured teacher characteristics explain relatively little of student achievement on tests”. To examine the reason behind the difference, we evaluate the non-parametrically estimated conditional quantile processes at five selected conditional quantiles {0.1, 0.25, 0.5, 0.75, 0.9} and plot each of them as a function of teachers’ experience (Fig. 2). The solid lines correspond to small classes and the dotted lines the regular sized classes. Two patterns emerge. First, the quantiles of the test scores of small classes are fairly flat in teachers’ experience. There is some gain when

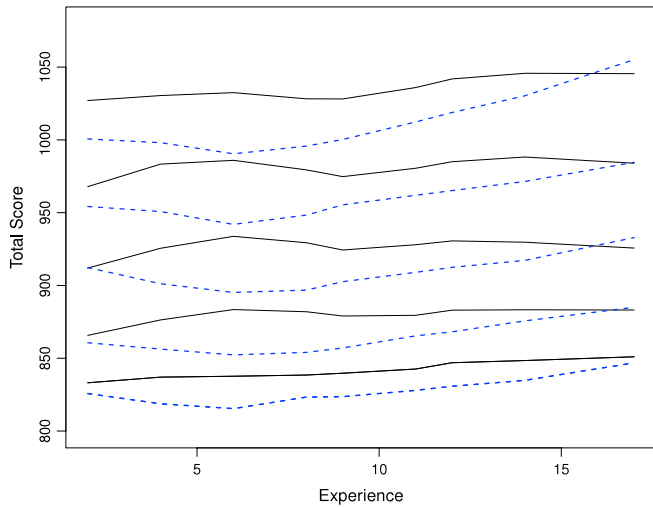


Fig. 2. Conditional quantiles of test scores as functions of teacher’s experience. The solid line corresponds to small classes and the dotted line regular sized classes. We report five quantiles {0.1, 0.25, 0.5, 0.75, 0.9}.

teachers’ experiences increases to 6–8 years, beyond which the curves are basically flat. Second, the quantiles associated with regular sized classes exhibit an initial small dip, after which they show clear upward trends as teachers’ experience increases. The patterns suggest that teachers’ experience is an important asset for managing regular sized classes but less important when class sizes are small. *Krueger (1999)* did not allow the effect of teachers’ experience to be different across class types, and arrived at a different conclusion.

8.2. Testing hypotheses on class size reduction

We now apply the test statistics discussed in Section 6.3 to examine a range of issues concerning quantile treatment effects. We set $\mathcal{T} = [0.1, 0.9]$. The value of the covariate is set to its median (9 years) unless stated otherwise. First, we consider “treatment significance”. The test statistic, given by (27), is 9.80 with the critical values being 4.81 and 5.35 at the 10% and 5% significance levels. So the treatment is strongly significant. Second, we consider “treatment homogeneity”. The test statistic, given by (28), equals 3.69 with the critical values being 3.24 and 3.54 at the 10% and 5% levels, respectively. Therefore the evidence points to treatment heterogeneity. Third, we consider “first order dominance”. The test statistic, given by (30), equals 0.00 with the critical values being 4.14 and 4.74 at the 10% and 5% levels. Therefore, we do not reject the hypothesis that the treatment effects are uniformly positive. Finally, we test “treatment equality”. The test statistic, given by (29) equals 11.20 when the comparison is between teachers with 9 and 2 (i.e., the 10th percentile) years of experience. The critical values at the 10% and 5% levels are 7.29 and 8.27. Thus the evidence supports that at these two experience levels the treatment effects are significantly different. Similarly as above, the test statistics can be carried out at other values of the covariate. We omit such details.

9. Conclusion

We have considered the estimation and inference about a non-parametrically specified conditional quantile process. The two estimation methods are computationally simple to implement and are practically feasible even for relatively large datasets. We obtained a uniform Bahadur representation, a functional central limit theorem and provided practical procedures for constructing uniform

confidence bands and testing hypothesis about the quantile process. When applied to the STAR dataset, the methods produced informative results, suggesting that they can be practically useful for studying distributional effects in a flexible nonparametric setting.

Appendix A. Proof of main results

Proof of Theorem 1. As outlined in the main text, the proof consists of three steps.

Step 1. Let $K_n = \log^{1/2}(nh_n^d)$. Because $V_{n,\tau}(0) = 0, V_{n,\tau}(\hat{\phi}(\tau)) \leq 0$ always holds for each τ and every n . To show $\hat{\phi}(\tau)$ lies within Φ , it is sufficient to show that for any $\epsilon > 0$, there exist some finite constants N_0 and $\eta > 0$ independent of τ , such that if $\|\phi(\tau)\| \geq K_n$ holds for some τ , then $P(V_{n,\tau}(\phi(\tau)) > \eta K_n^2) > 1 - \epsilon$ holds for all $n \geq N_0$. A sufficient condition for the above condition is $P(\inf_{\|\phi\| \geq K_n} \inf_{\tau \in \mathcal{T}} V_{n,\tau}(\phi) > \eta K_n^2) > 1 - \epsilon$ for all $n \geq N_0$. This formulation is useful because ϕ is no longer quantile dependent. Because $V_{n,\tau}(\phi)$ is convex in ϕ , we always have $V_{n,\tau}(\gamma\phi) - V_{n,\tau}(0) \geq \gamma(V_{n,\tau}(\phi) - V_{n,\tau}(0))$ for any $\gamma \geq 1$. Therefore, a further sufficient condition is

$$P\left(\inf_{\|\phi\|=K_n} \inf_{\tau \in \mathcal{T}} V_{n,\tau}(\phi) > \eta K_n^2\right) > 1 - \epsilon \quad \text{for all } n \geq N_0. \quad (\text{A.1})$$

We now establish (A.1). Apply *Knight’s (1998)* decomposition: $V_{n,\tau}(\phi) = W_{n,\tau}(\phi) + Z_{n,\tau}(\phi)$, where

$$W_{n,\tau}(\phi) = -(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n K_{i,\tau} \psi_\tau(u_i^0(\tau) - e_i(\tau)) z'_{i,\tau} \phi$$

with $\psi_\tau(u) = \tau - 1(u < 0)$,

$$Z_{n,\tau}(\phi) = \sum_{i=1}^n K_{i,\tau} \int_0^{(nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \phi} \{1(u_i^0(\tau) - e_i(\tau) \leq s) - 1(u_i^0(\tau) - e_i(\tau) \leq 0)\} ds.$$

Applying this decomposition, we have

$$\inf_{\|\phi\|=K_n} \inf_{\tau \in \mathcal{T}} K_n^{-2} V_{n,\tau}(\phi) \geq \inf_{\|\phi\|=K_n} \inf_{\tau \in \mathcal{T}} K_n^{-2} Z_{n,\tau}(\phi) - \sup_{\|\phi\|=K_n} \sup_{\tau \in \mathcal{T}} |K_n^{-2} W_{n,\tau}(\phi)|. \quad (\text{A.2})$$

We bound the two right hand side terms in (A.2) separately. First consider the second term:

$$\sup_{\|\phi\|=K_n} \sup_{\tau \in \mathcal{T}} |K_n^{-2} W_{n,\tau}(\phi)| \leq K_n^{-1} \sup_{\tau \in \mathcal{T}} \left\| (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n \{ \psi_\tau(u_i^0(\tau) - e_i(\tau)) - \psi_\tau(u_i^0(\tau)) \} z'_{i,\tau} K_{i,\tau} \right\| \quad (\text{L1})$$

$$+ K_n^{-1} \sup_{\tau \in \mathcal{T}} \left\| (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n \psi_\tau(u_i^0(\tau)) z'_{i,\tau} K_{i,\tau} \right\|. \quad (\text{L2})$$

Applying *Lemma B.6*, (L1) = $O_p(K_n^{-1}) = o_p(1)$. For (L2), the quantity inside the norm is $O_p(1)$ for any fixed $\tau \in \mathcal{T}$ by a central limit theorem and it is stochastic equicontinuous in τ by *Lemma B.3*. Therefore, (L2) = $O_p(K_n^{-1}) = o_p(1)$. Thus, $\sup_{\|\phi\|=K_n} \sup_{\tau \in \mathcal{T}} |K_n^{-2} W_{n,\tau}(\phi)| = o_p(1)$. Next, consider the first term in (A.2). We will show that it is strictly positive with probability tending to 1. To this end, note that the integral appearing in

$Z_{n,\tau}(\phi)$ is always nonnegative and satisfies (c.f. Lemma A.1 in Oka and Qu, 2011)

$$\int_0^{(nh_{n,\tau}^d)^{-1/2}z'_{i,\tau}\phi} \{1(u_i^0(\tau) - e_i(\tau) \leq s) - 1(u_i^0(\tau) - e_i(\tau) \leq 0)\} ds \geq (nh_{n,\tau}^d)^{-1/2} \frac{z'_{i,\tau}\phi}{2} \left\{ 1\left(u_i^0(\tau) - e_i(\tau) \leq (nh_{n,\tau}^d)^{-1/2} \frac{z'_{i,\tau}\phi}{2}\right) - 1(u_i^0(\tau) - e_i(\tau) \leq 0) \right\}.$$

Applying this inequality to $Z_{n,\tau}(\phi)$:

$$K_n^{-2}Z_{n,\tau}(\phi) \geq K_n^{-2} \left(\frac{h_n^d}{h_{n,\tau}^d}\right)^{1/2} \left(\frac{\phi}{2}\right)' \{S_n(\tau, \mathbf{0}, e_i(\tau)) - S_n(\tau, \phi/2, e_i(\tau))\} \tag{L3}$$

$$+ K_n^{-2}(nh_{n,\tau}^d)^{-1/2} \left(\frac{\phi}{2}\right)' \sum_{i=1}^n \left\{ P\left(u_i^0(\tau) - e_i(\tau) \leq (nh_{n,\tau}^d)^{-1/2} \frac{z'_{i,\tau}\phi}{2} \middle| x_i\right) - P(u_i^0(\tau) - e_i(\tau) \leq 0 \middle| x_i) \right\} z_{i,\tau}K_{i,\tau}. \tag{L4}$$

We now analyze the two right hand side terms (L3) and (L4) separately. Because of Lemma B.5, $\|\phi\| = K_n$ and $h_{n,\tau}^d/h_n^d = O(1)$, (L3) = $O_p(K_n^{-1}) = o_p(1)$. Applying a mean value theorem, (L4) = $(1/4)K_n^{-2}(nh_{n,\tau}^d)^{-1} \sum_{i=1}^n f_{Y|X}(\tilde{y}_i|x_i) K_{i,\tau} \phi' z_{i,\tau} z'_{i,\tau} \phi$, where \tilde{y}_i lies between $Q(\tau|x_i) + e_i(\tau)$ and $Q(\tau|x_i) + e_i(\tau) + (nh_{n,\tau}^d)^{-1/2}z'_{i,\tau}\phi/2$. Because x_i is in a vanishing neighborhood of x and \tilde{y}_i approaches $Q(\tau|x_i)$ as $n \rightarrow \infty$, Assumption 2 implies that $f_{Y|X}(\tilde{y}_i|x_i) \geq f_L$ for large n . Consequently, (L4) $\geq (1/4)K_n^{-2}f_L\phi'((nh_{n,\tau}^d)^{-1} \sum_{i=1}^n K_{i,\tau}z_{i,\tau}z'_{i,\tau})\phi$ uniformly in τ for large n . Further, $(nh_{n,\tau}^d)^{-1} \sum_{i=1}^n z_{i,\tau}z'_{i,\tau}K_{i,\tau} \rightarrow^p f_X(x)N_x(\tau)$ uniformly in τ , where $N_x(\tau)$ is defined in Theorem 1 and is positive definite for all τ by Assumption 6. Let $\lambda_{\min}(\tau) > 0$ denote the minimum eigenvalue of $f_X(x)N_x(\tau)$, then (L4) $\geq \frac{1}{8}f_L\lambda_{\min}(\tau)$ in probability uniformly in τ . Combining the above results on (L1)–(L4), we see that the first term in (A.2) is strictly positive and dominates the second term with probability tending to 1. This proves (A.1). In fact, using the same arguments as above, we can establish a stronger result: for any $\epsilon > 0$, there exists $K_0 > 0$ and $\eta > 0$, such that if $\|\hat{\phi}(\tau)\| \geq K_0$ holds for some τ , then

$$P(V_{n,\tau}(\hat{\phi}(\tau)) > \eta) > 1 - \epsilon \quad \text{holds for all } n \geq N_0. \tag{A.3}$$

The weaker result suffices for proving Theorem 1 while the stronger one is needed for Theorem 3.

Step 2. We analyze the three terms in (13) separately. Because of Step 1, we can restrict our attention to the set Φ defined in Display (12). The term in the first curly bracket is $o_p(1)$ uniformly over this set by Lemma B.5. The term in the second curly bracket is uniformly $o_p(1)$ implied by the proof of Lemma B.5. Consider the last term, and apply a first-order Taylor expansion. It equals

$$-(nh_n^d)^{-1/2} \sum_{i=1}^n f_{Y|X}(\tilde{y}_i|x_i) e_i(\tau) z_{i,\tau}K_{i,\tau} - (nh_n^d)^{-1/2}(nh_{n,\tau}^d)^{-1/2} \times \left(\sum_{i=1}^n f_{Y|X}(\tilde{y}_i|x_i) K_{i,\tau}z_{i,\tau}z'_{i,\tau}\right) \hat{\phi}(\tau),$$

where \tilde{y}_i lies between $Q(\tau|x_i)$ and $Q(\tau|x_i) + e_i(\tau) + (nh_{n,\tau}^d)^{-1/2}z'_{i,\tau}\hat{\phi}$. These results imply $\hat{\phi}(\tau)$ equals

$$\left(f_{Y|X}(Q(\tau|x))f_X(x)N_x(\tau)\right)^{-1} \left\{ \left(\frac{h_n^d}{h_{n,\tau}^d}\right)^{1/2} S_n(\tau, \mathbf{0}, \mathbf{0}) - (nh_{n,\tau}^d)^{-1/2}f_{Y|X}(Q(\tau|x)) \times \sum_{i=1}^n e_i(\tau) z_{i,\tau}K_{i,\tau} \right\} + o_p(1). \tag{A.4}$$

Step 3. For any x_i satisfying $(x_i - x)/h_{n,\tau} \in \text{supp}(K(\cdot))$, we have

$$e_i(\tau) = -\frac{1}{2} \left(\frac{x_i - x}{h_{n,\tau}}\right)' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \left(\frac{x_i - x}{h_{n,\tau}}\right) h_{n,\tau}^2 + o(h_{n,\tau}^2).$$

Below, we consider two cases separately. Case 1: x is a boundary point. Then,

$$-h_{n,\tau}^{-2} (nh_{n,\tau}^d)^{-1} \sum_{i=1}^n e_i(\tau) z_{i,\tau}K_{i,\tau} \xrightarrow{p} \frac{1}{2}f_X(x) \times \int_{\mathcal{D}_{x,\tau}} u' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} u \begin{bmatrix} 1 \\ u \end{bmatrix} K(u) du. \tag{A.5}$$

Plugging the above expression into (A.4) leads to the second result in Theorem 1. Case 2: x is an interior point. The above formula simplifies because $N_x(\tau)$ is block diagonal with the diagonal blocks being 1 and $\mu_2(K)$, and $h_{n,\tau}^{-2}(nh_{n,\tau}^d)^{-1} \sum_{i=1}^n e_i(\tau) z_{i,\tau}K_{i,\tau} = \frac{1}{2}tr\left(\frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \mu_2(K)\right) + o_p(1)$. Plugging this expression into (A.4) leads to the first result. ■

Proof of Theorem 2. The proof consists of two steps. The first step shows that $\hat{\alpha}(\tau)$, obtained by solving (4) for all $\tau \in \mathcal{T}$, converges weakly to the desired limit. The second step shows that $\hat{\alpha}^*(\tau)$ has the same weak limit as $\hat{\alpha}(\tau)$ over $\tau \in \mathcal{T}$. We focus on the interior point case.

Step 1. It suffices to show that the leading term in the uniform Bahadur representation converges weakly to the desired limit. Its finite dimensional convergence is an immediate consequence of results in Fan et al. (1994). The tightness follows because the denominator is finite and bounded away from 0 by Assumptions 1 and 2 and that the numerator is tight by Lemma B.3.

Step 2. We use similar arguments as in Neocleous and Portnoy (2008). Write $\hat{\alpha}^*(\tau) - \hat{\alpha}(\tau)$ as

$$\gamma(\tau) \left(\hat{\alpha}(\tau_j) - Q(\tau_j|x) - d_{\tau_j}h_{n,\tau_j}^2\right) - \gamma(\tau) \left(\hat{\alpha}(\tau) - Q(\tau|x) - d_\tau h_{n,\tau}^2\right) \tag{L5}$$

$$+ (1 - \gamma(\tau)) \left(\hat{\alpha}(\tau_{j+1}) - Q(\tau_{j+1}|x) - d_{\tau_{j+1}}h_{n,\tau_{j+1}}^2\right) - (1 - \gamma(\tau)) \left(\hat{\alpha}(\tau) - Q(\tau|x) - d_\tau h_{n,\tau}^2\right) \tag{L6}$$

$$+ \gamma(\tau) \left(Q(\tau_j|x) + d_{\tau_j}h_{n,\tau_j}^2\right) + (1 - \gamma(\tau)) \times \left(Q(\tau_{j+1}|x) + d_{\tau_{j+1}}h_{n,\tau_{j+1}}^2\right) - \left(Q(\tau|x) + d_\tau h_{n,\tau}^2\right). \tag{L7}$$

For Term (L5), applying Theorem 1 and Assumption 5, we have

$$\sqrt{nh_{n,\tau}^d} \left(\hat{\alpha}(\tau_j) - Q(\tau_j|x) - d_{\tau_j}h_{n,\tau_j}^2\right) - \sqrt{nh_{n,\tau}^d} \left(\hat{\alpha}(\tau) - Q(\tau|x) - d_\tau h_{n,\tau}^2\right) = (nh_{n,\tau}^d)^{-1/2}$$

$$\begin{aligned} & \times \frac{\sum_{i=1}^n (\tau_j - 1(u_i^0(\tau_j) \leq 0)) K_{i,\tau_j} - \sum_{i=1}^n (\tau - 1(u_i^0(\tau) \leq 0)) K_{i,\tau}}{f_X(x) f_{Y|X}(Q(\tau_j|x)|x)} \\ & + (nh_{n,\tau}^d)^{-1/2} \frac{\sum_{i=1}^n (\tau - 1(u_i^0(\tau) \leq 0)) K_{i,\tau}}{f_X(x)} \\ & \times \left(\frac{1}{f_{Y|X}(Q(\tau_j|x)|x)} - \frac{1}{f_{Y|X}(Q(\tau|x)|x)} \right) + o_p(1). \end{aligned}$$

The first term on the right hand side is $o_p(1)$ by the stochastic equicontinuity of the subgradient process, see Lemma B.3. The second term is also $o_p(1)$ by the tightness of the subgradient process and Assumption 2. Term (L6) can be analyzed similarly. For Term (L7), apply a Taylor expansion: $Q(\tau_j|x) = Q(\tau|x) + \frac{\partial Q(\tau|x)}{\partial \tau}(\tau_j - \tau) + O((\tau_{j+1} - \tau_j)^2)$, where the order of the remainder term is due to the Lipschitz continuity of $\partial Q(\tau|x)/\partial \tau$. This implies $\gamma(\tau)Q(\tau_j|x) + (1 - \gamma(\tau))Q(\tau_{j+1}|x) = Q(\tau|x) + O((\tau_{j+1} - \tau_j)^2)$. Similarly, $d_{\tau_j} h_{n,\tau_j}^2 = d_{\tau} h_{n,\tau}^2 + \frac{\partial d_{\tau} h_{n,\tau}^2}{\partial \tau}(\tau_j - \tau) + O((\tau_{j+1} - \tau_j)^2)$, implying $\gamma(\tau)d_{\tau_j} h_{n,\tau_j}^2 + (1 - \gamma(\tau))d_{\tau_{j+1}} h_{n,\tau_{j+1}}^2 = d_{\tau} h_{n,\tau}^2 + O((\tau_{j+1} - \tau_j)^2)$. Therefore, (L7) = $O((\tau_{j+1} - \tau_j)^2)$. The result then follows because $\tau_{j+1} - \tau_j = o((nh_n^d)^{-1/4})$. ■

Proof of Theorem 3. We first show $\sup_{1 \leq j \leq m} \sqrt{nh_{n,\tau_j}^d} |\tilde{\alpha}(\tau_j) - Q(\tau_j|x)| = O_p(1)$. The proof is by contradiction. Suppose the stochastic order is violated. Then, $\sqrt{nh_{n,\tau_j}^d} |\tilde{\alpha}(\tau_j) - Q(\tau_j|x)| > K_0$ for some $\tau_j \in \mathcal{T}$ and some $n \geq N_0$ with positive probability. We will argue along such a sequence. Without loss of generality, assume $\tilde{\alpha}(\tau_j) > Q(\tau_j|x) + (nh_{n,\tau_j}^d)^{-1/2} K_0$. Then, we consider the following two possibilities separately: (1) the constraint is not binding at τ_j (i.e., $\tilde{\alpha}(\tau_j) > \tilde{\alpha}(\tau_{j-1})$), and (2) the constraint is in fact binding (i.e., $\tilde{\alpha}(\tau_j) = \tilde{\alpha}(\tau_{j-1})$).

In the first case, $\tilde{\alpha}(\tau_j)$ is identical to the unconstrained estimate $\hat{\alpha}(\tau_j)$, therefore $V_{n,\tau_j}(\hat{\phi}(\tau_j)) = V_{n,\tau_j}(\tilde{\phi}(\tau_j)) \leq 0$. However, this contradicts (A.3). Consider the second case. Let τ_l denote the lowest quantile in the grid for which $\tilde{\alpha}(\tau_l) = \tilde{\alpha}(\tau_j)$. By the monotonicity of the quantile function and $\tilde{\alpha}(\tau_j) > Q(\tau_j|x) + (nh_{n,\tau_j}^d)^{-1/2} K_0$, we must have $\tilde{\alpha}(\tau_l) > Q(\tau_l|x) + (nh_{n,\tau_l}^d)^{-1/2} K_0$. Therefore, $P(V_{n,\tau_l}(\tilde{\phi}(\tau_l)) > \eta) > 1 - \epsilon$ holds for all $n \geq N_0$. Now, consider decreasing the value of $\tilde{\alpha}(\tau_l)$ by some small amount. This will decrease the value of the objective function $V_{n,\tau_l}(\tilde{\phi}(\tau_l))$ due to its convexity. This action does not violate any constraint because of the definition of τ_l (i.e., it is the lowest quantile for which $\tilde{\alpha}(\tau_l) = \tilde{\alpha}(\tau_j)$). Therefore, the new value remains admissible and yet returns a smaller objective function value. It contradicts that $\tilde{\alpha}(\tau_l)$ is the minimizer.

Given the above result and the given condition on m , the quantiles will not cross with probability arbitrarily close to 1 in large samples, therefore the large sample distribution is the same as the unconstrained case. This proves the result for the first estimator.

We now verify the conditions in Chernozhukov et al. (2010, Corollary 3) are satisfied for the second proposed estimator. First, our Assumptions 1–3 imply their Assumption 1 with the domain $(0, 1)$ replaced by \mathcal{T} . They also imply the strict monotonicity of $Q(\tau|x)$ in τ as well as its continuous differentiability in both arguments. Their Assumption 2 holds with $a_n = \sqrt{nh_{n,\tau}^d}$ and $\tilde{G}_x(\tau) = G(\tau)/(\sqrt{f_X(x)} f_{Y|X}(Q(\tau|x)|x)) + h(\tau)d_{\tau}$ because of the first result in Theorem 2. ■

Proof of Corollary 1. The mean squared error (MSE) for an interior point x is

$$\begin{aligned} & \frac{1}{4} \text{tr} \left(\mu_2(K) \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \right)^2 h_{n,\tau}^4 + \frac{\tau(1-\tau) \int K(v)^2 dv}{nh_{n,\tau}^d f_X(x) f_{Y|X}(Q(\tau|x)|x)^2} \\ & + o_p(nh_{n,\tau}^d). \end{aligned}$$

Computing the derivatives of the first two terms leads to the desired formula. We now verify that it satisfies Assumption 5. The factor in front of $n^{-1/(4+d)}$ in the Corollary corresponds to $c(\tau)$. It is bounded under Assumptions 1–4. To show that it is Lipschitz continuous, it suffices to verify $f_{Y|X}(Q(\tau|x)|x)^{-2/(4+d)}$, $\text{tr}(\mu_2(K) \partial^2 Q(\tau|x)/\partial x \partial x')^{-2/(4+d)}$ and $(\tau(1-\tau))^{1/(4+d)}$ have bounded first derivatives. Consider the first quantity: $|f_{Y|X}(Q(\tau_1|x)|x)^{-\frac{2}{4+d}} - f_{Y|X}(Q(\tau_2|x)|x)^{-\frac{2}{4+d}}| = |\frac{2}{4+d} f(Q(\tilde{\tau}|x)|x)^{-\frac{6+d}{4+d}} f'_{Y|X}(Q(\tilde{\tau}|x)|x) Q'(\tilde{\tau}|x)|(\tau_2 - \tau_1)|$, where $\tilde{\tau} \in \mathcal{T}$. The terms on the right side are all bounded by Assumptions 1–3. The remaining two quantities can be analyzed similarly. ■

Proof of Corollary 2. We have, for the interior point case,

$$\begin{aligned} & P(Q(\tau|x) \notin C_p \text{ for some } \tau \in \mathcal{T}) \\ & = P \left(\frac{1}{\sigma_{n,\tau}(x)} |Q(\tau|x) - \tilde{\alpha}^*(\tau) + d_{\tau} h_{n,\tau}^2| > Z_p \text{ for some } \tau \in \mathcal{T} \right) \\ & = P \left(\sup_{\tau \in \mathcal{T}} \frac{1}{\sigma_{n,\tau}(x)} |Q(\tau|x) - \tilde{\alpha}^*(\tau) + d_{\tau} h_{n,\tau}^2| > Z_p \right). \end{aligned}$$

The right hand side converges to p by the continuous mapping theorem. The boundary point case can be proved using the same argument. ■

Appendix B. Auxiliary lemmas

The next lemma is needed for analyzing the effect of the quantile dependent bandwidth on the asymptotic properties of the estimators.

Lemma B.1. Let Assumptions 1, 4 and 5 hold. (1) For any $\gamma \geq 1$ there exists a $B > 0$, such that for any $\tau_1, \tau_2 \in \mathcal{T}$ with $\tau_1 \leq \tau_2$, we have $E \|z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}\|^{2\gamma} \leq B h_n^d (\tau_2 - \tau_1)^{2\gamma}$. (2) Let $b_n = (nh_n^d)^{1/2+\kappa}$ with $0 < \kappa < 1/2$ being some arbitrary constant and h_n satisfying $h_n \rightarrow 0$ and $nh_n^d \rightarrow \infty$ as $n \rightarrow \infty$. Let $\delta_n = \{(\tau_1, \tau_2) : \tau_1 \in \mathcal{T}, \tau_2 \in \mathcal{T}, \tau_1 \leq \tau_2 \leq \tau_1 + b_n^{-1}\}$, then $\sup_{(\tau_1, \tau_2) \in \delta_n} (nh_n^d)^{-1/2} \sum_{i=1}^n \|z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}\| = o_p(1)$.

Proof. Without loss of generality, we can assume that the support of $K(\cdot)$ is contained in $D = [-1, 1]^{\otimes d}$. Applying Assumptions 4 and 5, there exists a constant \bar{C} independent of τ such that

$$\|z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}\| \leq \bar{C} * 1 (x_i - x \in \bar{c} D h_n) (\tau_2 - \tau_1). \tag{B.1}$$

Consequently, $E \|z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}\|^{2\gamma} \leq \bar{C}^{2\gamma} (\tau_2 - \tau_1)^{2\gamma} P(x_i - x \in \bar{c} D h_n)$. Because the density of x_i is bounded by Assumption 1, there exists a constant A independent of τ such that $P(x_i - x \in \bar{c} D h_n) \leq A h_n^d$. Letting $B = \bar{C}^{2\gamma} A$ proves the first result. Then, apply (B.1): $(nh_n^d)^{-1/2} \sum_{i=1}^n \|z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}\| \leq \bar{C} (nh_n^d)^{1/2} (\tau_2 - \tau_1) \{ (nh_n^d)^{-1} \sum_{i=1}^n 1(x_i - x \in \bar{c} D h_n) \}$. The quantity inside the curly brackets satisfies a weak law of large numbers, therefore is $O_p(1)$. Finally, $\sup_{(\tau_1, \tau_2) \in \delta_n} \bar{C} (nh_n^d)^{1/2} (\tau_2 - \tau_1) \leq \bar{C} (nh_n^d)^{-\kappa} \rightarrow 0$ because $\kappa > 0$ and $nh_n^d \rightarrow \infty$. ■

The next Lemma is needed to establish the stochastic equicontinuity of the process $S_n(\tau, 0, 0)$.

Lemma B.2. Let $b_n = (nh_n^d)^{1/2+\kappa}$ with $\kappa \in (0, 1/2)$ being some arbitrary constant and $h_n \rightarrow 0$ and $nh_n^d \rightarrow \infty$ as $n \rightarrow \infty$. Then, there exist $\gamma > 1$ and $\bar{C} < \infty$, such that for any $\tau_1, \tau_2 \in \mathcal{T}$ satisfying $|\tau_2 - \tau_1| \geq b_n^{-1}$, we have $E \|S_n(\tau_2, 0, 0) - S_n(\tau_1, 0, 0)\|^{2\gamma} \leq \bar{C} |\tau_2 - \tau_1|^\gamma$.

Proof. It suffices to show $(E \|S_n(\tau_2, 0, 0) - S_n(\tau_1, 0, 0)\|^{2\gamma})^{1/\gamma} \leq \bar{C}^{1/\gamma} (\tau_2 - \tau_1)$ for $\tau_2 \geq \tau_1$. Let

$$A_{1i} = \{(\tau_2 - 1(u_i^0(\tau_2) \leq 0)) - (\tau_1 - 1(u_i^0(\tau_1) \leq 0))\} z_{i,\tau_2} K_{i,\tau_2},$$

$$A_{2i} = (\tau_1 - 1(u_i^0(\tau_1) \leq 0)) (z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}).$$

Let $A_{1i,j}$ and $A_{2i,j}$ denote the j th element of A_{1i} and A_{2i} respectively. By the Minkowski inequality,

$$(E \|S_n(\tau_2, 0, 0) - S_n(\tau_1, 0, 0)\|^{2\gamma})^{1/\gamma} \leq \sum_{j=1}^{d+1} \left\{ (nh_n^d)^{-\gamma} E \left| \sum_{i=1}^n A_{1i,j} + A_{2i,j} \right|^{2\gamma} \right\}^{1/\gamma}. \tag{B.2}$$

Applying the Rosenthal inequality and the Minkowski inequality similarly to Bai (1996, Lemma A1), the term inside the curly brackets is bounded by

$$2^\gamma C (nh_n^d)^{-\gamma} \left(\sum_{i=1}^n E \|A_{1i}\|^2 + E \|A_{2i}\|^2 \right)^\gamma + 2^\gamma (nh_n^d)^{-\gamma} C \times \sum_{i=1}^n ((E \|A_{1i}\|^{2\gamma})^{1/\gamma} + (E \|A_{2i}\|^{2\gamma})^{1/\gamma})^\gamma. \tag{B.3}$$

We now derive bounds for the summands appearing above.

$$E \|A_{1i}\|^{2\gamma} \leq E \left(E \left\{ (\tau_2 - 1(u_i^0(\tau_2) \leq 0)) - \tau_1 + 1(u_i^0(\tau_1) \leq 0) \right\}^2 |x_i \right\} \|z_{i,\tau_2} K_{i,\tau_2}\|^{2\gamma} \right) \leq (\tau_2 - \tau_1) E \|z_{i,\tau_2} K_{i,\tau_2}\|^{2\gamma} \leq Ch_n^d (\tau_2 - \tau_1)$$

where the first inequality is because $|\tau_2 - 1(u_i^0(\tau_2) \leq 0) - \tau_1 + 1(u_i^0(\tau_1) \leq 0)| \leq 1$, the second uses similar arguments as in Lemma B.1 and C is a constant independent of τ . Meanwhile, $E \|A_{2i}\|^{2\gamma} \leq E \left\| (z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}) \right\|^{2\gamma} \leq Bh_n^d (\tau_2 - \tau_1)^{2\gamma}$, where the last inequality uses the first result in Lemma B.1. Similarly, $E \|A_{1i}\|^2 \leq Ch_n^d (\tau_2 - \tau_1)$ and $E \|A_{2i}\|^2 \leq Bh_n^d (\tau_2 - \tau_1)^2$. Applying these bounds, (B.3) is bounded from above by $M (\tau_2 - \tau_1)^\gamma + M (nh_n^d (\tau_2 - \tau_1))^{1-\gamma} (\tau_2 - \tau_1)^\gamma$ for some finite M . The latter is further bounded by $2M (\tau_2 - \tau_1)^\gamma$ using the definition of b_n and $\gamma > 1$. Thus, (B.2) is bounded by $(d + 1) (2M)^{1/\gamma} (\tau_2 - \tau_1)$. Let $\bar{C} = (d + 1) (2M)^{1/\gamma}$, the proof is complete. ■

The next lemma establishes the stochastic equicontinuity of the process $S_n(\tau, 0, 0)$.

Lemma B.3. For any $\varepsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$, such that for large n ,

$$P \left(\sup_{\tau'', \tau' \in \mathcal{T}, |\tau'' - \tau'| \leq \delta} \|S_n(\tau'', 0, 0) - S_n(\tau', 0, 0)\| > \varepsilon \right) < \eta.$$

Proof. Let $z_{i,\tau,j}$ denote the j th component of $z_{i,\tau}$ ($j = 1, \dots, d + 1$). Note that $z_{i,\tau',j} \geq 0$ if and only if $z_{i,\tau'',j} \geq 0$ for any $\tau'', \tau' \in \mathcal{T}$. Without loss of generality, assume that the elements of $z_{i,\tau,j}$ are all nonnegative. Otherwise, we can write $z_{i,\tau,j}$ as $z_{i,\tau,j}^+ - z_{i,\tau,j}^- \equiv z_{i,\tau,j} 1(z_{i,\tau,j} \geq 0) - (-z_{i,\tau,j}) 1(z_{i,\tau,j} < 0)$.

For a given δ , \mathcal{T} contains at most $1/\delta$ intervals of length δ . Therefore, it suffices to show that for any $\varepsilon > 0, \eta > 0$, there exists a $\delta > 0$, such that (see Billingsley, 1968, p. 58, equation 8.12)

$$P \left(\sup_{s \leq \tau \leq \delta+s, \tau \in \mathcal{T}} \|S_n(\tau, 0, 0) - S_n(s, 0, 0)\| > \varepsilon \right) < \delta \eta$$

holds for all $s \in \mathcal{T}$ when n is large. (B.4)

We use a chaining argument to show (B.4). Partition $[s, \delta + s]$ into $b_n = (nh_n^d)^{1/2+\kappa}$ intervals of equal sizes with $0 < \kappa < 1/2$. Let τ_j denote the lower limit of the j th interval; $\tau_1 = s$. Then,

$$\sup_{s \leq \tau \leq \delta+s} \|S_n(\tau, 0, 0) - S_n(s, 0, 0)\| \leq \sup_{1 \leq j \leq b_n} \sup_{\tau \in [\tau_j, \tau_{j+1}]} \|S_n(\tau, 0, 0) - S_n(\tau_j, 0, 0)\| + \sup_{1 \leq j \leq b_n} \|S_n(\tau_j, 0, 0) - S_n(s, 0, 0)\|. \tag{B.5}$$

Consider the first term. For any $\tau \in [\tau_j, \tau_{j+1}]$, we have

$$S_n(\tau, 0, 0) - S_n(\tau_j, 0, 0) \geq S_n(\tau_{j+1}, 0, 0) - S_n(\tau_j, 0, 0) + (nh_n^d)^{-1/2} \times \sum_{i=1}^n (\tau_{j+1} - 1(u_i^0(\tau_{j+1}) < 0)) (z_{i,\tau} K_{i,\tau} - z_{i,\tau_{j+1}} K_{i,\tau_{j+1}}) - (nh_n^d)^{-1/2} \sum_{i=1}^n (\tau_{j+1} - \tau_j) z_{i,\tau} K_{i,\tau}. \tag{B.6}$$

The norm of the first summation is bounded by $(nh_n^d)^{-1/2} \sum_{i=1}^n \|z_{i,\tau} K_{i,\tau} - z_{i,\tau_{j+1}} K_{i,\tau_{j+1}}\| = o_p(1)$ applying Lemma B.1. The second summation is also $o_p(1)$ because $|\tau_{j+1} - \tau_j| \leq (nh_n^d)^{-1/2-\kappa}$. Therefore, $S_n(\tau, 0, 0) - S_n(\tau_j, 0, 0) \geq S_n(\tau_{j+1}, 0, 0) - S_n(\tau_j, 0, 0) - \varepsilon/5$ with probability no less than $1 - \delta \eta$ uniformly in \mathcal{T} for large n . Meanwhile, there is a reversed inequality for (B.6) given by

$$S_n(\tau, 0, 0) - S_n(\tau_j, 0, 0) \leq (nh_n^d)^{-1/2} \sum_{i=1}^n 1(u_i^0(\tau_j) < 0) (z_{i,\tau_j} K_{i,\tau_j} - z_{i,\tau} K_{i,\tau}) + (nh_n^d)^{-1/2} \sum_{i=1}^n (\tau_{j+1} - \tau_j) z_{i,\tau} K_{i,\tau} - (nh_n^d)^{-1/2} \sum_{i=1}^n \tau_j (z_{i,\tau_j} K_{i,\tau_j} - z_{i,\tau} K_{i,\tau}).$$

All the three terms on the right hand side are $o_p(1)$ uniformly in \mathcal{T} following similar arguments as above. Therefore, $S_n(\tau, 0, 0) - S_n(\tau_j, 0, 0) \leq \varepsilon/5$ holds with probability no less than $1 - \delta \eta$ uniformly in \mathcal{T} for large n . Consequently, $\sup_{s \leq \tau \leq \delta+s} \|S_n(\tau, 0, 0) - S_n(s, 0, 0)\| \leq \sup_{1 \leq j \leq b_n} \|S_n(\tau_{j+1}, 0, 0) - S_n(\tau_j, 0, 0)\| + \frac{2\varepsilon}{5} + \sup_{1 \leq j \leq b_n} \|S_n(\tau_j, 0, 0) - S_n(s, 0, 0)\| \leq \sup_{1 \leq j \leq b_n} 3 \|S_n(\tau_j, 0, 0) - S_n(s, 0, 0)\| + \frac{2\varepsilon}{5}$, which holds with probability no less than $1 - \delta \eta$ uniformly in \mathcal{T} for large n . This implies, to establish (B.4) it suffices to show $P(\sup_{1 \leq j \leq b_n} \|S_n(\tau_j, 0, 0) - S_n(\tau_1, 0, 0)\| > \frac{\varepsilon}{5}) < \delta \eta$ for large n . To this end, we follow Bai (1996, p.613) and apply Theorem 12.2 in Billingsley (1968) to bound the above probability by using the bounds on the moments of $S_n(\tau_{j+1}, 0, 0) - S_n(\tau_j, 0, 0)$, established in the previous lemma. Specifically, Billingsley's result states that if there exists $\beta \geq 0, \alpha > 1$ and $u_l \geq 0$ ($l = 1, \dots, b_n$) such that $E(\|S_n(\tau_j, 0, 0) - S_n(\tau_i, 0, 0)\|^\beta) \leq (\sum_{i < l \leq j} u_i)^\alpha, 0 \leq i \leq j \leq b_n$, then $P(\sup_{1 \leq j \leq b_n} \|S_n(\tau_j, 0, 0) - S_n(\tau_1, 0, 0)\| > \varepsilon) \leq \varepsilon^{-\alpha} C_{\beta,\alpha} (u_1 + \dots + u_{b_n})^\alpha$. In our case, set $\beta = 2\gamma$ and $\alpha = \gamma$. Lemma B.2 then implies $E \|S_n(\tau_j, 0, 0) - S_n(\tau_i, 0, 0)\|^\beta \leq \bar{C} (\tau_j - \tau_i)^\alpha$

for $0 \leq i \leq j \leq b_n$. Therefore, $P(\sup_{1 \leq j \leq b_n} \|S_n(\tau_j, \mathbf{0}, \mathbf{0}) - S_n(\tau_1, \mathbf{0}, \mathbf{0})\| > \frac{\varepsilon}{5}) \leq \delta((\frac{\varepsilon}{5})^{-\alpha} C_{\beta, \alpha} \bar{C} \delta^{\alpha-1})$. Choosing δ such that $((\frac{\varepsilon}{5})^{-\alpha} C_{\beta, \alpha} \bar{C} \delta^{\alpha-1}) \leq \eta$ completes the proof. ■

The next two lemmas establish some relationship between $S_n(\tau, \phi, e_i(\tau))$ and $S_n(\tau, \mathbf{0}, e_i(\tau))$. They are needed to quantify the effect of the parameter estimation.

Lemma B.4. Let $b_n = (nh_n^d)^{1/2+\kappa}$ and consider a partition of \mathcal{T} into b_n intervals of equal sizes. Let τ_j denote the lower limit of the j th interval. Then, under Assumptions 1–6, we have

$$\begin{aligned} & \sup_{1 \leq j \leq b_n} \sup_{\tau_{j-1} \leq \tau \leq \tau_j} \left\| (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau} K_{i,\tau} \{1(u_i^0(\tau_j) \leq e_i(\tau)) \right. \\ & \quad \left. - 1(u_i^0(\tau_j) \leq e_i(\tau_j))\right\| = o_p(1), \\ & \sup_{1 \leq j \leq b_n} \sup_{\tau_{j-1} \leq \tau \leq \tau_j} \sup_{\|\phi\| \leq \log^{1/2}(nh_{n,\tau}^d)} \left\| (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau} K_{i,\tau} \right. \\ & \quad \times \left\{ 1(u_i^0(\tau_j) \leq e_i(\tau) + (nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \phi) - (nh_n^d)^{-1/2} \right. \\ & \quad \left. \times \sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} 1(u_i^0(\tau_j) \leq e_i(\tau_j) + (nh_{n,\tau_j}^d)^{-1/2} z'_{i,\tau_j} \phi) \right\} \\ & = o_p(1). \end{aligned}$$

Proof. Without loss of generality, assume $z_{i,\tau}$ is a scalar and nonnegative. Consider the first result. By the second result in Lemma B.1, the term on the left hand side has the same order as

$$\begin{aligned} & \sup_{1 \leq j \leq b_n} \sup_{\tau_{j-1} \leq \tau \leq \tau_j} \left\| (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} \{1(u_i^0(\tau_j) \leq e_i(\tau)) \right. \\ & \quad \left. - 1(u_i^0(\tau_j) \leq e_i(\tau_j))\right\|. \end{aligned} \tag{B.7}$$

Meanwhile, for any $\tau \in [\tau_j, \tau_{j+1}]$, we have $e_i(\tau) - e_i(\tau_j) = o_p((nh_n^d)^{-1/2}) \leq \varepsilon(nh_n^d)^{-1/2}$ in probability uniformly over $\tau \in \mathcal{T}$, where the constant ε can be made arbitrarily small. Thus, (B.7) is bounded by $\sup_{1 \leq j \leq b_n} \|(nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} 1(e_i(\tau_j) - \varepsilon(nh_n^d)^{-1/2} \leq u_i^0(\tau_j) \leq e_i(\tau_j) + \varepsilon(nh_n^d)^{-1/2})\|$. Let

$$\begin{aligned} \xi_{i,\tau_j} &= 1(e_i(\tau_j) - \varepsilon(nh_n^d)^{-1/2} \leq u_i^0(\tau_j) \leq e_i(\tau_j) + \varepsilon(nh_n^d)^{-1/2}) \\ & \quad - E\{1(e_i(\tau_j) - \varepsilon(nh_n^d)^{-1/2} \leq u_i^0(\tau_j) \leq e_i(\tau_j) \\ & \quad + \varepsilon(nh_n^d)^{-1/2}) | x_i\}. \end{aligned}$$

Applying the triangle inequality, (B.7) can thus be further bounded by

$$\begin{aligned} & \sup_{1 \leq j \leq b_n} (nh_n^d)^{-1/2} \left\| \sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} \xi_{i,\tau_j} \right\| \\ & \quad + \sup_{1 \leq j \leq b_n} (nh_n^d)^{-1/2} \sum_{i=1}^n E\{1(e_i(\tau_j) \\ & \quad - \varepsilon(nh_n^d)^{-1/2} \leq u_i^0(\tau_j) \leq e_i(\tau_j) + \varepsilon(nh_n^d)^{-1/2}) | x_i\}. \end{aligned}$$

The second term is $o_p(1)$. The first term satisfies: $P(\sup_{1 \leq j \leq b_n} (nh_n^d)^{-1/2} \|\sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} \xi_{i,\tau_j}\| > \eta) \leq \sum_{j=1}^{b_n} P((nh_n^d)^{-1/2} \|\sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} \xi_{i,\tau_j}\| > \eta)$. Applying the Rosenthal inequality with $\gamma > 1$ to

the j th term on the right side, it is bounded by

$$\begin{aligned} & C\eta^{-2\gamma} (nh_n^d)^{-\gamma/2} \left\{ E \left((nh_n^d)^{-1/2} \sum_{i=1}^n E(\|\xi_{i,\tau_j}\|^2 | x_i) \|z_{i,\tau_j} K_{i,\tau_j}\|^2 \right)^\gamma \right. \\ & \quad \left. + (nh_n^d)^{-\gamma/2} \sum_{i=1}^n E \left(\|z_{i,\tau_j} K_{i,\tau_j}\|^{2\gamma} E(\|\xi_{i,\tau_j}\|^2 | x_i) \right) \right\}. \end{aligned}$$

The term inside the curly brackets is finite because $E(\|\xi_{i,\tau_j}\|^2 | x_i) \leq E(\|\xi_{i,\tau_j}\|^2 | x_i) \leq B(nh_n^d)^{-1/2}$ for some B . Therefore, the preceding display is bounded by $CM\eta^{-2\gamma} (nh_n^d)^{-\gamma/2}$, whose summation over $\sum_{j=1}^{b_n}$ is bounded by $CM\eta^{-2\gamma} (nh_n^d)^{1/2+\kappa-\gamma/2}$, which converges to zero by choosing $\gamma > 1 + 2\kappa$. This proves the first result. The second result can be shown using similar arguments because the difference between $(nh_{n,\tau}^d)^{-1/2} z'_{i,\tau} \phi$ and $(nh_{n,\tau_j}^d)^{-1/2} z'_{i,\tau_j} \phi$ is of lower order than $(nh_n^d)^{-1/2}$ by Assumption 5. The uniformity in ϕ can be shown as in the next lemma. The detail is omitted. ■

Lemma B.5. Under Assumptions 1–6, let $K_n = \log^{1/2}(nh_n^d)$, we have

$$\sup_{\tau \in \mathcal{T}} \sup_{\|\phi\| \leq K_n} \|S_n(\tau, \phi, e_i(\tau)) - S_n(\tau, \mathbf{0}, e_i(\tau))\| = o_p(1).$$

Proof. As in Lemma B.3, assume that the elements of $z_{i,\tau}$ are all nonnegative. Partition \mathcal{T} into $b_n = (nh_n^d)^{1/2+\kappa}$ intervals and let τ_j denote the lower limit of the j th interval. Applying Lemma B.4 and similar arguments as in (B.6), we have, for any $\tau \in [\tau_j, \tau_{j+1}]$ and any $\varepsilon > 0$, $S_n(\tau_{j+1}, \phi, e_i(\tau_{j+1})) - S_n(\tau_j, \mathbf{0}, e_i(\tau_j)) + 2\varepsilon \leq S_n(\tau, \phi, e_i(\tau)) - S_n(\tau, \mathbf{0}, e_i(\tau)) \leq S_n(\tau_j, \phi, e_i(\tau_j)) - S_n(\tau_{j+1}, \mathbf{0}, e_i(\tau_{j+1})) + 2\varepsilon$ uniformly for large n . Adding and subtracting terms:

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \sup_{\|\phi\| \leq K_n} \|S_n(\tau, \phi, e_i(\tau)) - S_n(\tau, \mathbf{0}, e_i(\tau))\| \\ & \leq 2 \sup_{1 \leq j \leq b_n} \|S_n(\tau_{j+1}, \mathbf{0}, \mathbf{0}) - S_n(\tau_j, \mathbf{0}, \mathbf{0})\| \\ & \quad + 4 \sup_{1 \leq j \leq b_n+1} \|S_n(\tau_j, \mathbf{0}, e_i(\tau_j)) - S_n(\tau_j, \mathbf{0}, \mathbf{0})\| \\ & \quad + 2 \sup_{1 \leq j \leq b_n+1} \sup_{\|\phi\| \leq K_n} \|S_n(\tau_j, \phi, e_i(\tau_j)) \\ & \quad - S_n(\tau_j, \mathbf{0}, e_i(\tau_j))\| + 4\varepsilon. \end{aligned} \tag{B.8}$$

Consider the third term on the right hand side. For any $\delta > 0$, the set $\{\phi : \|\phi\| \leq K_n\}$ can be partitioned into $N(\delta)$ spheres such that the diameter of each sphere is less than or equal to δ . Note that $N(\delta) = O((K_n/\delta)^{d+1})$. Denote the spheres by D_h with centers being ϕ_h ($h \in \{1, 2, \dots, N(\delta)\}$). For any $\phi \in D_h$, we have $z'_{i,\tau_j} \phi_h - \|z_{i,\tau_j}\| \delta \leq z'_{i,\tau_j} \phi \leq z'_{i,\tau_j} \phi_h + \|z_{i,\tau_j}\| \delta$. Therefore,

$$\begin{aligned} & \sup_{1 \leq j \leq b_n+1} \sup_{\|\phi\| \leq K_n} \|S_n(\tau_j, \phi, e_i(\tau_j)) - S_n(\tau_j, \mathbf{0}, e_i(\tau_j))\| \\ & \leq \sup_{1 \leq j \leq b_n+1} \sup_{\substack{1 \leq h \leq N(\delta) \\ k=1,2,\phi \in D_h}} \left\| (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} \right. \\ & \quad \times E\{1(u_i^0(\tau_j) \leq e_i(\tau_j) + z'_{i,\tau_j} \phi_h + (-1)^k \|z_{i,\tau_j}\| \delta) | x_i\} \\ & \quad \left. - E\{1(u_i^0(\tau_j) \leq e_i(\tau_j) + (nh_{n,\tau_j}^d)^{-1/2} z'_{i,\tau_j} \phi) | x_i\} \right\| \\ & \quad + \sup_{1 \leq j \leq b_n+1} \max_{1 \leq h \leq N(\delta)} \max_{k=1,2} \|(nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} \xi_{i,j,h,k}\|, \end{aligned} \tag{B.9}$$

where

$$\begin{aligned} \xi_{i,j,h,k} &= E\{1(u_i^0(\tau_j) \leq e_i(\tau_j) + (nh_{n,\tau_j}^d)^{-1/2} z'_{i,\tau_j} \phi_h \\ & \quad + (-1)^k \|z_{i,\tau_j}\| \delta) | x_i\} \end{aligned}$$

$$\begin{aligned}
& -1 \left(u_i^0(\tau_j) \leq e_i(\tau_j) \right) + (nh_{n,\tau_j}^d)^{-1/2} z'_{i,\tau_j} \phi_h \\
& + (-1)^k \|z_{i,\tau_j}\| \delta \\
& - E \left(1 \left(u_i^0(\tau_j) \leq e_i(\tau_j) \right) \middle| x_i \right) + 1 \left(u^0(\tau_j) \leq e_i(\tau_j) \right).
\end{aligned}$$

Applying a Taylor expansion, the first term on the right hand side of (B.9) is of the same order as δ , which can be made arbitrarily small by choosing, say $\delta = K_n^{-\kappa}$. To bound the second norm, it is sufficient to show that for any $\epsilon > 0$, $1 \leq j \leq b_n + 1$, $1 \leq h \leq N(\delta)$ and $k \in \{1, 2\}$, $b_n N(\delta) P(\|(nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau_j} K_{i,\tau_j} \xi_{i,j,h,k}\| > \epsilon) \rightarrow 0$. Applying the Rosenthal inequality, the preceding probability is bounded by $M\epsilon^{-2\gamma} b_n N(\delta) (nh_n^d)^{-\gamma/2} K_n^\gamma$ for large n . Now, using the definition of b_n , $N(\delta)$ and K_n and choosing $\gamma = 1 + 2\kappa + c$ for some $c > 0$, the preceding quantity converges to 0 for $\delta = K_n^{-\kappa}$. Finally, the second term on the right hand side of (B.8) can be analyzed similarly and is $o_p(1)$; the first term is $o_p(1)$ because of Lemma B.3. ■

Lemma B.6. Under Assumptions 1–6,

$$\begin{aligned}
& \sup_{\tau \in \mathcal{T}} \left\| (nh_n^d)^{-1/2} \sum_{i=1}^n \left\{ \psi_\tau(u_i^0(\tau) - e_i(\tau)) - \psi_\tau(u_i^0(\tau)) \right\} z_{i,\tau} K_{i,\tau} \right\| \\
& = O_p(1).
\end{aligned}$$

Proof. The term inside the norm can be rewritten as

$$\begin{aligned}
& S_n(\tau, 0, 0) - S_n(\tau_j, 0, e_i(\tau)) \\
& + (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau} K_{i,\tau} \left\{ P(u_i^0(\tau) \leq 0) \right. \\
& \left. - E \left(1 \left(u_i^0(\tau) \leq e_i(\tau) \right) \middle| x_i \right) \right\} + o_p(1).
\end{aligned}$$

The second to the last line is $o_p(1)$ as shown in the previous Lemma. The remaining term is $O_p(1)$ upon a Taylor expansion using the fact that $e_i(\tau) = O(h_n^2)$. ■

References

- Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. *J. Amer. Statist. Assoc.* 97 (457), 284–292.
- 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.
- Angrist, J., Chernozhukov, V., Fernández-Val, I., 2006. Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* 74 (2), 539–563.
- Bai, J., 1996. Testing for parameter constancy in linear regressions: An empirical distribution function approach. *Econometrica* 64 (3), 597–622.
- Belloni, A., Chernozhukov, V., Fernández-Val, I., (2011) Conditional quantile processes based on series or many regressors. Working Paper, Boston University.
- Bickel, P.J., 1975. One-step huber estimates in the linear model. *J. Amer. Statist. Assoc.* 70 (350), 428–434.
- Billingsley, P., 1968. *Convergence of Probability Measures*. Wiley.
- Bondell, H.D., Reich, B.J., Wang, H., 2010. Noncrossing quantile regression curve estimation. *Biometrika* 97 (4), 825–838.
- Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *J. Econometrics* 167 (2), 413–425.
- Chaudhuri, P., 1991. Nonparametric estimates of regression quantiles and their local bahadur representation. *Ann. Statist.* 19 (2), 760–777.
- Chen, S., Khan, S., 2001. Semiparametric estimation of a partially linear censored regression model. *Econometric Theory* 17 (3), 567–590.
- Chernozhukov, V., Fernández-Val, I., 2005. Subsampling inference on quantile regression processes. *Sankhyā* 67 (2), 253–276.
- Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. *Econometrica* 78 (3), 1093–1125.
- Chernozhukov, V., Hansen, C., Jansson, M., 2009. Finite sample inference for quantile regression models. *J. Econometrics* 152 (2), 93–103.
- Delecroix, M., Simioni, M., Thomas-agnan, C., 1996. Functional estimation under shape constraints. *J. Nonparametr. Stat.* 6 (1), 69–89.
- Dette, H., Volgushev, S., 2008. Non-crossing non-parametric estimates of quantile curves. *J. R. Stat. Soc. Ser. B* 70 (3), 609–627.
- Doksum, K., 1974. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* 2 (2), 267–277.
- Escanciano, J.C., Velasco, C., 2010. Specification tests of parametric dynamic conditional quantiles. *J. Econometrics* 159 (1), 209–221.
- Fan, J., Hu, T.C., Truong, Y.K., 1994. Robust non-parametric function estimation. *Scand. J. Stat.* 21 (4), 433–446.
- Firpo, S., 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75 (1), 259–276.
- Guerre, E., Sabbah, C., 2012. Uniform bias study and bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* 28 (01), 87–129.
- Härdle, W.K., Song, S., 2010. Confidence bands in quantile regression. *Econometric Theory* 26, 1180–1200.
- He, X., 1997. Quantile curves without crossing. *Amer. Statist.* 51, 186–192.
- He, X., Shi, P., 1996. Bivariate tensor-product b-splines in a partly linear model. *J. Multivariate Anal.* 58 (2), 162–181.
- He, X., Shi, P., 1998. Monotone b-spline smoothing. *J. Amer. Statist. Assoc.* 93, 643–650.
- Heckman, J.J., Smith, J., Clements, N., 1997. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Rev. Econ. Stud.* 64 (4), 487–535.
- Honda, T., 2004. Quantile regression in varying coefficient models. *J. Statist. Plann. Inference* 121 (1), 113–125.
- Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. *J. Econometrics* 142 (2), 615–635.
- Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* 26 (2), 240–257.
- Kim, M.O., 2007. Quantile regression with varying coefficients. *Ann. Statist.* 35 (1), 92–108.
- Knight, K., 1998. Limiting distributions for L1 regression estimators under general conditions. *Ann. Statist.* 26 (2), 755–770.
- Koenker, R., 2005. *Quantile Regression*. Cambridge University Press.
- Koenker, R., (2010) Additive models for quantile regression: Model selection and confidence band-aids. Working Paper, University of Illinois at Urbana-Champaign.
- Koenker, R., Bassett Jr., G., 1978. Regression quantiles. *Econometrica* 46 (1), 33–50.
- Koenker, R., Ng, P., 2005. Inequality constrained quantile regression. *Sankhyā* 67 (2), 418–440.
- Koenker, R., Portnoy, S., 1987. L-estimation for linear models. *J. Amer. Statist. Assoc.* 82 (399), 851–857.
- Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. *Econometrica* 70 (4), 1583–1612.
- Kong, E., Linton, O., Xia, Y., 2010. Uniform bahadur representation for local polynomial estimates of m-regression and its application to the additive model. *Econometric Theory* 26 (05), 1529–1564.
- Krueger, A.B., 1999. Experimental estimates of education production functions. *The Q. J. Econ.* 114 (2), 497–532.
- Lee, S., 2003. Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory* 19 (1), 1–31.
- Lee, S., Linton, O., Whang, Y.J., 2009. Testing for stochastic monotonicity. *Econometrica* 77 (2), 585–602.
- Lehmann, E.L., 1975. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.
- Linton, O., Maasoumi, E., Whang, Y.J., 2005. Consistent testing for stochastic dominance under general sampling schemes. *Rev. Econ. Stud.* 72 (3), 735–765.
- Mammen, E., 1991. Estimating a smooth monotone regression function. *Ann. Statist.* 19 (2), 724–740.
- Mammen, E., Keilegom, I.V., Yu, K., (2013) Expansion for moments of regression quantiles with application to nonparametric testing. Working Paper, arXiv:1306.6179 [math.ST].
- Masry, E., 1996. Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal.* 17 (6), 571–599.
- Neocleous, T., Portnoy, S., 2008. On monotonicity of regression quantile functions. *Statist. Probab. Lett.* 78 (10), 1226–1229.
- Oka, T., Qu, Z., 2011. Estimating structural changes in regression quantiles. *J. Econometrics* 162, 248–267.
- Parzen, M.I., Wei, L.J., Ying, Z., 1994. A resampling method based on pivotal estimating functions. *Biometrika* 81 (2), 341–350.
- Ruppert, D., Wand, M.P., 1994. Multivariate locally weighted least squares regression. *Ann. Statist.* 22 (3), 1346–1370.
- Song, S., Ritov, Y., Härdle, W.K., 2012. Bootstrap confidence bands and partial linear quantile regression. *J. Multivariate Anal.* 107, 244–262.
- Yu, K., Jones, M.C., 1998. Local linear quantile regression. *J. Amer. Statist. Assoc.* 93 (441), 228–237.