

# A SIMPLE NONPARAMETRIC APPROACH FOR ESTIMATION AND INFERENCE OF CONDITIONAL QUANTILE FUNCTIONS

ZHENG FANG  
*Texas A&M University*

QI LI  
*Texas A&M University*

KAREN X. YAN  
*Georgia Institute of Technology*

In this paper, we present a new nonparametric method for estimating a conditional quantile function and develop its weak convergence theory. The proposed estimator is computationally easy to implement and automatically ensures quantile monotonicity by construction. For inference, we propose to use a residual bootstrap method. Our Monte Carlo simulations show that this new estimator compares well with the check-function-based estimator in terms of estimation mean squared error. The bootstrap confidence bands yield adequate coverage probabilities. An empirical example uses a dataset of Canadian high school graduate earnings, illustrating the usefulness of the proposed method in applications.

## 1. INTRODUCTION

Quantile regression has now become a very useful tool in economics, statistics, and other social sciences. But concerns about quantile crossing and model misspecification (He, 1997; Angrist, Chernozhukov, and Fernández-Val, 2006) have led to the development of nonparametric estimation of conditional quantile functions (Chaudhuri, 1991; Koenker, Ng, and Portnoy, 1994; Yu and Jones, 1998). An important framework in this methodology is the following nonparametric location-scale model, where the outcome variable  $Y_i$  and the covariate  $X_i$  are related by

$$Y_i = m_0(X_i) + \sigma_0(X_i)\epsilon_i, \quad (1)$$

for  $i = 1, \dots, n$ , where  $\epsilon_i$  is a zero-mean error term that is independent of  $X_i$ ,  $m_0(\cdot)$  and  $\sigma_0(\cdot) > 0$  are the unknown location and scale functions, respectively.

---

We would like to thank the Editor, Peter Phillips, a Co-Editor, and three anonymous referees for helpful comments and suggestions, which have substantially improved the paper. Address correspondence to Qi Li, Department of Economics, Texas A&M University, College Station, TX, USA; e-mail [qi-li@tamu.edu](mailto:qi-li@tamu.edu).

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Model (1) generalizes the location-scale-based linear quantile regression model. In addition, compared with homoskedastic nonparametric quantile regression models (Chaudhuri, 1991; Honda, 2004), i.e.,

$$Y_i = m_0(X_i) + u_i, \quad (2)$$

where the error term  $u_i$  has zero mean and is independent of  $X_i$ , the location-scale model is more flexible in capturing heterogeneity across individuals. As a result, it has received a great deal of attention in the literature (Fan and Gijbels, 1996; Van Keilegom and Akritas, 1999; Heuchenne and Keilegom, 2007; Einmahl and Van Keilegom, 2008; Neumeyer, 2008, 2009a; Florens, Simar, and Keilegom, 2014; Birke, Neumeyer, and Volgushev, 2017). For example, Chen, Dahl, and Khan (2005) use a censored nonparametric location-scale regression model to study how the median-length unemployment insurance spell in New Jersey varies with age, whereas Florens et al. (2014) use a flexible nonparametric location-scale model to estimate a frontier function.

There is rich literature on nonparametrically estimating the error's distribution function and/or testing independence between the covariate  $X$  and the idiosyncratic error  $u$  based on the nonparametric location-scale model. For example, Akritas and Van Keilegom (2001) study the nonparametric location-scale model (1) and establish the weak convergence of the process  $\sqrt{n}(\hat{F}_\epsilon(\cdot) - F_\epsilon(\cdot))$ , where  $\hat{F}_\epsilon$  is the empirical distribution function of the nonparametrically estimated residuals  $\hat{\epsilon}_i = (Y_i - \hat{m}_n(X_i))/\hat{\sigma}_n(X_i)$ , and  $\hat{m}_n(X_i)$  and  $\hat{\sigma}_n(X_i)$  are nonparametric estimators of  $m_0(X_i)$  and  $\sigma_0(X_i)$ , respectively. Birke et al. (2017) investigate the empirical independence process of covariates and estimated errors in a nonparametric location-scale conditional quantile curve and prove a weak convergence result. Cheng (2002), Einmahl and Van Keilegom (2008), Neumeyer (2009a, b), and Kiwitt and Neumeyer (2012) consider the specification tests of the model (1), including testing the independence between the error  $\epsilon$  and the covariate  $X$ , testing monotonicity of the nonparametric conditional quantile function and establishing bootstrap versions of the weak convergence results.

In this paper, we also consider a location-scale model given in (1), but differing from the existing literature, we focus on the problem of estimating the conditional quantile function of  $Y_i$  given  $X_i = x$  based on a location-scale quantile model framework. In particular, we propose a new estimator that is theoretically and computationally simple. The simplicity of our approach is rooted in an identification strategy that is different from prior studies in the context of quantile estimation. Specifically, we assume that  $E[\epsilon_i] = 0$  and  $E[\epsilon_i^2] = 1$ , which is merely a normalization and hence is not restrictive. This assumption is not new and is, in fact, ubiquitous in *mean* regression models. When it comes to *quantile* regression, however, the literature has taken a different route. For example, He (1997) requires that  $\epsilon_i$  has a median of zero and that  $|\epsilon_i|$  has a median of one, which, of course, can also be achieved without loss of generality. Similar quantile-type normalizations have been imposed in, for instance, Chaudhuri (1991), Horowitz and Lee (2005), Chen et al. (2005), and Birke et al. (2017).

An appealing consequence of our normalization is that the conditional quantile function  $q_\tau(x)$  of  $Y_i$  given  $X_i = x$  takes a particularly simple closed-form structure:

$$q_\tau(x) = m_0(x) + \sigma_0(x)Q_\epsilon(\tau), \quad (3)$$

for all  $\tau \in (0, 1)$ , where  $Q_\epsilon(\tau)$  is the  $\tau$ th quantile of  $\epsilon_i$ . Note that  $m_0(x)$  and  $\sigma_0^2(x)$  are the conditional mean and conditional variance functions of  $Y_i$  given  $X_i = x$ , respectively, both of which can be estimated by standard nonparametric methods (Fan and Yao, 1998). Thus, if  $\hat{m}_n(x)$  and  $\hat{\sigma}_n(x)$  are estimators of  $m_0(x)$  and  $\sigma_0(x)$ , respectively (which can be obtained from mean regression), then  $q_\tau(x)$  can be estimated by

$$\hat{q}_\tau(x) = \hat{m}_n(x) + \hat{\sigma}_n(x)\hat{Q}_\epsilon(\tau), \quad (4)$$

where  $\hat{Q}_\epsilon(\tau)$  is the empirical quantile function of the regression residuals. Here, we stress that one can use any nonparametric method (kernel or series method) to estimate  $m_0(x)$  and  $\sigma_0(x)$ . As long as they are consistent, the resulting  $\hat{q}_\tau(x)$  will be consistent as well in view of equation (4).

In the literature, there are two classes of estimation strategies for the conditional quantile function. The first is tailored to the location-scale model (1). As mentioned previously, a common feature of these methods is that they adopt quantile-type normalizations. Despite the fact that these normalizations are not restrictive, the estimation steps are not as straightforward as ours. For example, the restricted regression quantile approach proposed by He (1997) consists of three different quantile regressions. Following the estimation procedure proposed by Dette and Volgushev (2008) that is designed to avoid quantile crossing, Birke et al. (2017) present an estimator constructed from a transformation of a kernel estimator of the conditional cumulative distribution function (CDF), where the transformation entails an auxiliary distribution function, a kernel function, and a bandwidth parameter (in addition to those used in kernel estimation). In contrast, our estimator automatically guarantees monotonicity in quantile index  $\tau$ . Therefore, we do not need sorting (Chernozhukov, Fernández-Val, and Galichon, 2010; Qu and Yoon, 2015), isotonization (Mammen, 1991) or transformation of any kind (Dette and Volgushev, 2008). This desired property of our approach lies in the fact that we have exploited the location-scale structure, whereas those alternative monotonicity methods have their own merits in nonlocation-scale settings.

The second class of strategies for estimating the conditional quantile function is fully nonparametric. There are two popular methods in this area: inverting an estimator of the conditional distribution function and using the check-function-based approach. The inverse-CDF-based estimator is of the form  $\tilde{q}_\tau(x) = \inf\{y : \hat{F}_n(y|x) \geq \tau\}$ , for  $\tau \in (0, 1)$ , where  $\hat{F}_n(y|x)$  is a nonparametric estimator of conditional distribution function  $F(y|x)$ . The check-function-based approach delivers  $\check{q}_\tau(x) = \arg \min_q \sum_{i=1}^n \rho_\tau(Y_i - q)K(\frac{X_i - x}{h})$ , where  $\rho_\tau(u) = u(\tau - \mathbf{1}(u \leq 0))$  is the check function,  $K(\cdot)$  is a kernel function, and  $h$  is a bandwidth parameter. It is well known that the choice of smoothing parameter is of crucial importance

in nonparametric estimation, and some data-driven methods, such as the least-squares cross-validation (LS-CV) method, can lead to optimal smoothing parameter selection. However, for the inverse-CDF-based estimator, the LS-CV method computation time is  $O(n^3)$  (Li, Lin, and Racine, 2013), where  $n$  is the sample size. By contrast, our quantile estimator requires only  $O(n^2)$  computation time. Moreover, while it is difficult to obtain a derivative estimator and derive its asymptotic theory if one uses the inverse-CDF-based method to estimate the quantile function, our method also yields the derivative function estimate when using the local linear estimation method. In addition, as in Racine and Li (2017), when the data have an unbounded support, our estimator tends to have lower finite-sample estimation mean squared errors (MSEs).

We exploit the location-scale structure in constructing our estimator and require the existence of higher-order moments in deriving the asymptotic distribution theory. By contrast, the inverse-CDF-based and check-function-based estimators do not impose these restrictions. Therefore, we admit that our model setup is restrictive relative to the aforementioned two methods. However, in practice, researchers often are prepared to make some model assumptions, and our proposed method can be useful in empirical applications, given its simplicity. Furthermore, our method is not supposed to replace the existing inverse-CDF-based estimator or the check-function-based estimator, which are well established in the literature. Instead, it is an additional tool to complement these existing methods.

Although our method avoids the quantile crossings as a result of the location-scale structure of the model, it does not prevent the model from being misspecified, i.e., the true quantile function may not have a location-scale specification. This is a somewhat more subtle issue than the indication of misspecification via quantile crossings (Chernozhukov et al. 2010; Phillips, 2015).<sup>1</sup> Following Racine and Li (2017), if the possible misspecification of a location-scale model is a concern, we suggest using a pretest procedure to detect the possible misspecification. Using a pretest estimator avoids severe misspecification with regard to the location-scale model assumption. We apply this pretest method to some data generating processes (DGPs) that violate the location-scale assumption. Simulation results show that the pretest procedure works well in guarding against severe misspecification. Note that our model structure is the same as that of Racine and Li (2017), but our estimation procedure is much simpler. In Section 6.1, we use simulations to compare the finite-sample performances of our proposed estimator with those of the check-function-based estimator and Racine and Li's (2017) estimator.

We reiterate that the simplicity of our estimation hinges on an identification strategy that is common in the context of mean regressions but (somewhat surprisingly) less appreciated in the literature of quantile regression. We derive the Bahadur expansions for our estimator, which, in turn, deliver the asymptotic distribution. Both are uniform in the quantile index. For inferential purposes, we propose a residual bootstrap procedure. These uniformity results can be useful in,

<sup>1</sup>We would like to thank Peter Phillips for pointing out the reference Phillips (2015).

for example, constructing confidence bands for the conditional quantile function and testing the distributional hypothesis, such as the homogeneity or equality of quantile treatment effects, as in Koenker and Xiao (2002) and Qu and Yoon (2015). As an illustration, we consider the problem of constructing confidence bands. Our simulation results further show that the bootstrap confidence intervals have adequate coverage probabilities under different error distributions that we consider. For an empirical illustration, we apply our method to a Canadian dataset of high school graduate earnings and examine how the conditional quantiles of logwage (defined as the logarithm of wage) vary across different quantile indexes  $\tau \in (0, 1)$ .

The remainder of the paper is organized as follows. Section 2 describes the methodology in detail. Section 3 introduces assumptions and establishes the asymptotic theory for our proposed estimator. Section 4 considers the bootstrap inference. We extend our local constant conditional quantile estimator to the local linear case in Section 5. In Section 6, we use Monte Carlo simulations to examine the finite-sample performances of the proposed quantile estimator and the bootstrap confidence interval. Section 7 contains an illustrative empirical application. Section 8 concludes. Proofs of the main theoretical results of the paper can be found in the Supplementary Material.

## 2. METHODOLOGY

In this section, we describe a two-step procedure for estimating the conditional quantile function  $q_\tau(x)$ . In the first step, we use undersmoothed smoothing parameters  $b_1$  and  $b_2$  to estimate the conditional mean function  $m_0(x)$  and the conditional standard error function  $\sigma_0(x)$ . This, in turn, allows us to construct the residuals  $\{\hat{\epsilon}_i\}_{i=1}^n$  and then use the empirical quantile function of  $\{\hat{\epsilon}_i\}_{i=1}^n$  as an estimator of  $Q_\epsilon(\tau)$ . In the second stage, we use, possibly optimally selected, smoothing parameters  $h_1$  and  $h_2$  to estimate  $m_0(x)$  and  $\sigma_0(x)$ . These estimators, together with the estimator of  $Q_\epsilon(\tau)$  in the first step, yield the final estimator of  $q_\tau(x) = m_0(x) + \sigma_0(x)Q_\epsilon(\tau)$ . We focus on the local constant estimator. The extension to the local linear case is given Section 5.

We fix  $x$  and treat  $q_\tau(x)$  as a process indexed by  $\tau \in \mathcal{T} = [\underline{\tau}, \bar{\tau}]$  with  $0 < \underline{\tau} \leq \bar{\tau} < 1$ . Let  $\mathcal{X} \subset \mathbf{R}$  denote the support of  $X$ .<sup>2</sup> We assume that  $\mathcal{X}$  is a union of bounded and closed sets. Because the local constant kernel estimator has a large estimation bias at the boundary region, we use a trimming function to eliminate observations near the boundary. For expositional simplicity, we consider a simple case that  $\mathcal{X} = [0, 1]$  to illustrate how to construct a trimming function. Let  $h_1$  and  $h_2$  be the smoothing parameters used to estimate  $m_0(x)$  and  $\sigma_0(x)$ , respectively. To avoid the boundary bias problem, similar to Birke et al. (2017), we define a trimmed set  $\mathcal{X}_n \stackrel{\text{def}}{=} [\delta_n, 1 - \delta_n]$ , where  $\delta_n = 2 \max\{h_1, h_2\}$ . We then only consider  $x \in \mathcal{X}_n$ . Note

<sup>2</sup>For expositional simplicity, we will only focus on the univariate case. Our analysis can be readily extended to the multivariate setting.

that as  $n$  gets larger, the trimmed set  $\mathcal{X}_n$  becomes closer to  $(0, 1)$ , so that we only trim out the boundary points  $\{0\}$  and  $\{1\}$  asymptotically (which has zero Lebesgue measure). In addition, we assume the kernel function  $K(\cdot)$  has support  $[-1, 1]$ . Then, in kernel estimation, when  $n$  is sufficiently large, a term like  $K((X_i - x)/h_j)$  will be zero if  $X_i$  is outside  $[\delta_n - h_j, 1 - \delta_n + h_j] \subset [\delta_n/2, 1 - \delta_n/2]$ . We will not distinguish between the trimming sets  $[\delta_n/2, 1 - \delta_n/2]$  and  $[\delta_n, 1 - \delta_n]$ , and we denote both by  $\mathcal{X}_n$ . We are now ready to present the details of the two-step estimation procedure.

**Step 1.** We estimate the quantile function of the error term based on non-parametric residuals. First, we estimate  $m_0(x)$  and  $\sigma_0(x)$  by the nonparametric local constant kernel method with the undersmoothed parameters  $b_1$  and  $b_2$  (see Assumption 3.5):

$$\hat{m}_{b_1}(x) = \frac{\sum_{j=1}^n Y_j K\left(\frac{X_j - x}{b_1}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{b_1}\right)}, \quad \hat{\sigma}_{b_2}^2(x) = \frac{\sum_{j=1}^n (Y_j - \hat{m}_{b_1}(X_j))^2 K\left(\frac{X_j - x}{b_2}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{b_2}\right)}.$$

Let  $\hat{\sigma}_{b_2}(X_i) = \sqrt{\hat{\sigma}_{b_2}^2(X_i)}$ . Using the residuals<sup>3</sup>

$$\hat{\epsilon}_i = \frac{Y_i - \hat{m}_{b_1}(X_i)}{\hat{\sigma}_{b_2}(X_i)}, \quad i = 1, \dots, n, \tag{5}$$

we estimate  $Q_\epsilon(\tau)$  by  $\hat{Q}_\epsilon(\tau) = \arg \min_s |\hat{F}_{n,\epsilon}(s) - \tau|$ , where  $\hat{F}_{n,\epsilon}(\cdot)$  is a trimmed empirical CDF of  $\epsilon$  defined by

$$\hat{F}_{n,\epsilon}(s) = \frac{\sum_{i=1}^n \mathbf{1}_{i,n} \mathbf{1}(\hat{\epsilon}_i \leq s)}{\sum_{i=1}^n \mathbf{1}_{i,n}}, \tag{6}$$

where  $\mathbf{1}_{i,n} = \mathbf{1}(X_i \in \mathcal{X}_n)$ .

**Step 2.** Estimate  $m_0(\cdot)$  and  $\sigma_0(\cdot)$  with the smoothing parameters  $h_1$  and  $h_2$  (where we allow for optimally selected smoothing parameters):

$$\hat{m}_{h_1}(x) = \frac{\sum_{j=1}^n Y_j K\left(\frac{X_j - x}{h_1}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h_1}\right)}, \quad \hat{\sigma}_{h_2}^2(x) = \frac{\sum_{j=1}^n (Y_j - \hat{m}_{h_1}(X_j))^2 K\left(\frac{X_j - x}{h_2}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h_2}\right)}.$$

Finally, the  $\tau$ th conditional quantile estimator of  $Y$  given  $X = x$  is given by

$$\hat{q}_\tau(x) = \hat{m}_{h_1}(x) + \hat{\sigma}_{h_2}(x) \hat{Q}_\epsilon(\tau), \tag{7}$$

where  $\hat{\sigma}_{h_2}(X_i) = \sqrt{\hat{\sigma}_{h_2}^2(X_i)}$ .

**Remark 2.1.** In this paper, the smoothing parameters  $b_1$  and  $b_2$  are only used for preliminary estimates of  $m_0(X_i)$  and  $\sigma_0^2(X_i)$  that are needed for estimating the

<sup>3</sup>In practice, one can also use the standardized residuals  $\tilde{\epsilon}_i = (\hat{\epsilon}_i - \bar{\tilde{\epsilon}}) / \text{std}(\hat{\epsilon}_i)$ ,  $i = 1, \dots, n$ , where  $\bar{\tilde{\epsilon}} = n^{-1} \sum_{i=1}^n \hat{\epsilon}_i$  and  $\text{std}(\hat{\epsilon}_i) = \sqrt{n^{-1} \sum_{i=1}^n (\hat{\epsilon}_i - \bar{\tilde{\epsilon}})^2}$ .

quantile  $Q_\epsilon(\tau)$ . After we obtain  $\hat{Q}_\epsilon(\tau)$ , we will estimate  $q_\tau(x)$  by estimating  $m_0(x)$  and  $\sigma_0^2(x)$  using the smoothing parameters  $h_1$  and  $h_2$ .

**Remark 2.2.** In practice, one can select  $h_1$  and  $h_2$  using the leave-one-out cross-validation method. Specifically,  $h_1 = \arg \min_h \sum_{i=1}^n (Y_i - \hat{m}_{-i,h}(X_i))^2$ , where  $\hat{m}_{-i,h}(X_i)$  is the leave-one-out kernel estimator of  $m(X_i)$  with the smoothing parameter  $h$ , and  $h_2 = \arg \min_h \sum_{i=1}^n [(Y_i - \hat{m}_{-i,h_1}(X_i))^2 - \hat{\sigma}_{-i,h}^2(X_i)]^2$ , where  $\hat{\sigma}_{-i,h}^2(X_i)$  is the leave-one-out kernel estimator of  $\sigma_0^2(X_i)$  with the smoothing parameter  $h$ . However, when we construct uniform confidence bands, we use the undersmoothed smoothing parameters  $h_j$  for  $j = 1, 2$ .

### 3. ASYMPTOTIC THEORY

We make the following regularity assumptions.

**Assumption 3.1.** (i)  $\{Y_i, X_i\}_{i=1}^n$  is independently and identically distributed (i.i.d.) as  $\{Y, X\}$ ; (ii)  $\epsilon_i$  is i.i.d. with mean zero and unit variance, and is independent of  $X_i$ , for all  $i = 1, \dots, n$ ; (iii)  $\mathcal{X}$ , the support of  $X$ , is a union of closed and bounded intervals in  $\mathbf{R}$ .

**Assumption 3.2.**  $K(\cdot)$  has the compact support over  $[-1, 1]$  with  $K(1) = K(-1) = 0$ . It is a symmetric and bounded density function, and is twice continuously differentiable on  $(-1, 1)$ .

**Assumption 3.3.** Let  $f(x)$  be the density function of  $X$  evaluated at  $x \in \mathcal{X}$ . (i) Both  $m_0(x)$  and  $\sigma_0(x)$  are twice continuously differentiable on  $x$  at the interior of  $\mathcal{X}$ . Their second derivative functions are uniformly bounded over  $x \in \mathcal{X}$ ; (ii)  $\inf_{x \in \mathcal{X}} \sigma_0(x) > 0$ ; (iii)  $\inf_{x \in \mathcal{X}} f(x) > 0$ .

**Assumption 3.4.** (i) The support  $\mathcal{E}$  of the error term  $\epsilon$  is convex; (ii) The distribution function  $F_\epsilon$  of  $\epsilon$  is twice continuously differentiable, and its second derivative function is bounded on  $\mathcal{E}$ ; (iii) The density function  $f_\epsilon$  of  $\epsilon$  is uniformly continuous on  $\mathcal{E}$  and positive on  $[F_\epsilon^{-1}(\tau) - \delta, F_\epsilon^{-1}(1 - \tau) + \delta]$ , for any  $\tau \in (0, 1/2)$  and some  $\delta > 0$ ; (iv)  $\sup_{u \in \mathcal{E}} |uf_\epsilon(u)| < \infty$ ; (v)  $E[|\epsilon|^{4+\eta}] < \infty$ , for some  $\eta > 0$ .

**Assumption 3.5.** The bandwidths  $h_j$  and  $b_j, j = 1, 2$ , satisfy that, as  $n \rightarrow \infty$ , (i)  $h_j = c_j n^{-1/5}$  for some finite constant  $c_j > 0$ ; (ii)  $b_j = c_j n^{-\beta}$  for some finite constant  $c_j > 0$  and  $\beta \in (\frac{1}{5}, \frac{1}{3})$ .

Assumption 3.1 is fairly standard. Although the *i.i.d.* assumption is maintained throughout the paper, the results can be easily extended to the weakly dependent data case. Assumption 3.2 includes standard restrictions on the kernel function. Assumption 3.3 imposes smoothness conditions on  $m_0(\cdot)$ ,  $\sigma_0(\cdot)$  and the density function  $f(\cdot)$ . We also assume  $f(\cdot)$  and the scale  $\sigma_0(\cdot)$  to be positive over the support for identification reasons. Assumption 3.4 is a sufficient condition to deliver the desired asymptotic property of the empirical quantile estimator  $\hat{Q}_\epsilon(\tau)$ . Assumption 3.5 allows for optimal smoothing on  $h_1$  and  $h_2$  but requires undersmoothing

for  $b_1$  and  $b_2$ , and the assumption implies that  $nb_j^3/(\log n)^2 \rightarrow \infty, j = 1, 2$ . We impose undersmoothing on  $b_j, j = 1, 2$ , so that the estimation error from  $\hat{Q}_\epsilon(\tau)$  is asymptotically negligible.<sup>4</sup>

Throughout this paper, we use the notation  $\dot{g}(x) = \frac{dg(x)}{dx}$  and  $\ddot{g}(x) = \frac{d^2g(x)}{dx^2}$ , where  $g$  is a function of  $x$ , such as  $m_0$  or  $\sigma_0$ , or estimators of them ( $\hat{m}, \hat{\sigma}$ , etc.).

**THEOREM 3.1.** *Under Assumptions 3.1–3.5, for any fixed  $x$  at the interior of  $\mathcal{X}$ ,*

$$\sqrt{nh_1} \{ \hat{q}_\tau(x) - q_\tau(x) - h_1^2 D_1(x) - h_2^2 D_{2,\tau}(x) \} \xrightarrow{L} \mathbb{G}(\tau) \text{ in } \ell^\infty(\mathcal{T}),$$

where

$$D_1(x) = \frac{1}{2} \mu_2 [2m_0(x)\dot{f}(x)/f(x) + m_0(x)],$$

$$D_{2,\tau}(x) = \frac{Q_\epsilon(\tau)}{4\sigma_0(x)} \mu_2 \{ 2\dot{\delta}_2(x)\dot{f}(x)/f(x) + \ddot{\delta}_2(x) \},$$

$\delta_2(x) = \sigma_0^2(x)$ , and  $\mathbb{G}(\tau)$  is a Gaussian process defined over  $\mathcal{T} = [\underline{\tau}, \bar{\tau}]$ ,  $0 < \underline{\tau} \leq \bar{\tau} < 1$ , with zero mean, and the covariance structure is given by

$$\begin{aligned} \text{Cov}(\mathbb{G}(\tau_1), \mathbb{G}(\tau_2)) = \frac{\sigma_0^2(x)}{f(x)} \left\{ \nu_0 + \frac{c\nu_0 Q_\epsilon(\tau_1) Q_\epsilon(\tau_2)}{4} [E(\epsilon^4) - 1] \right. \\ \left. + \frac{cE(\epsilon^3) \int K(u)K(cu)du}{2} [Q_\epsilon(\tau_1) + Q_\epsilon(\tau_2)] \right\}, \end{aligned}$$

with  $\nu_0 = \int K^2(u)du$ ,  $c = \lim_{n \rightarrow \infty} h_1/h_2$ .

Theorem 3.1 is the weak convergence result, and it trivially implies the pointwise convergence of  $\hat{q}_\tau(x)$  (for any fixed  $\tau \in \mathcal{T}$ ), as described in the following Corollary 3.1.

**COROLLARY 3.1.** *Under Assumptions 3.1–3.5, for any  $x$  at the interior of  $\mathcal{X}$  and any  $\tau \in \mathcal{T}$ ,*

$$\sqrt{nh_1} \{ \hat{q}_\tau(x) - q_\tau(x) - h_1^2 D_1(x) - h_2^2 D_{2,\tau}(x) \} \xrightarrow{L} N(0, V_\tau),$$

where

$$V_\tau = \frac{\sigma_0^2(x)}{f(x)} \left\{ \nu_0 + \frac{c\nu_0 Q_\epsilon^2(\tau)}{4} [E(\epsilon^4) - 1] + cQ_\epsilon(\tau)E(\epsilon^3) \int K(u)K(cu)du \right\},$$

where  $D_1(x), D_{2,\tau}(x), \nu_0$ , and  $c$  are the same as in Theorem 3.1.

<sup>4</sup>As can be seen from the proof of Theorem 3.1 in Appendix A,  $\hat{Q}_\epsilon(\tau) - Q_\epsilon(\tau) = O_p(b_1^2 + b_2^2) + o_p(n^{-2/5}) = o_p(n^{-\frac{2}{5}})$  when  $b_j = o_p(n^{-1/5})$ , for  $j = 1, 2$ , which is asymptotically negligible.



**Remark 3.1.** Note that the leading bias of the quantile estimator  $\hat{q}_\tau(x)$  includes two terms: (a)  $D_1(x)$ , the estimation bias in  $\hat{m}(x)$  and (b)  $D_{2,\tau}(x)$ , the estimation bias in  $\hat{\sigma}(x)$ . The asymptotic variance of  $\hat{q}_\tau(x)$  is related to the estimation variances of  $\hat{m}(x)$  and  $\hat{\sigma}(x)$ , as well as the covariance between the two. If one uses a higher-order kernel to estimate  $\sigma_0(x)$  as in Racine and Li (2017), then  $\hat{\sigma}(x)$  will converge faster than  $\hat{m}(x)$ . Therefore, the leading bias of our quantile estimator will simplify to  $D_1(x)$ , and the asymptotic variance will collapse to  $v_0\sigma_0^2(x)/f(x)$ . These coincide with the leading bias and variance in Racine and Li (2017). Then, by Propositions 4.4 and 4.5 of Racine and Li (2017), we know that: (1) when the error  $\epsilon$  has unbounded support, our estimator tends to have a smaller estimation MSE than the check-function-based estimator, and (2) if  $\epsilon$  has a bounded support with density function bounded away from zero, the check-function-based method is likely to have a smaller estimation MSE than our proposed method.

**Remark 3.2.** In practice, one can use the data-driven method to select the optimal smoothing parameters, and Theorem 3.1 and Corollary 3.1 still hold with the data-dependent smoothing parameters by using the stochastic equicontinuity argument as in Li and Li (2010).

**Remark 3.3.** Note that in the two-step estimation procedure described in Section 2, when estimating  $\sigma_0^2(x)$ , we use  $\hat{m}_\lambda(X_j)$ ,  $j \in \{1, \dots, n\}$  instead of  $\hat{m}_\lambda(x)$  ( $\lambda$  denotes the generic smoothing parameter). Fan and Yao (1998) recommend the same approach, as it leads to the result that the bias  $D_{2,\tau}$  only contains two terms. If one uses  $\hat{m}_\lambda(x)$ , however, the bias  $D_{2,\tau}$  will contain three terms.

**Remark 3.4.** A co-editor suggests that our identification conditions may be relaxed to  $\text{Median}(\epsilon_i) = 0$  and  $\text{Median}(|\epsilon_i|) = 1$  to allow fat-tail distribution of  $\epsilon_i$ . These normalization conditions were considered in Birke et al. (2017). One can apply results from Birke et al. (2017) to estimate the conditional quantile function  $q_\tau(x)$  under these normalization conditions. Unlike our model setup, this approach does not lead to a simple closed-form estimator. We leave the formal investigation of this approach as a future research topic.

**Remark 3.5.** Zhao and Xiao (2014) develop an innovative approach to constructing efficient estimators of conditional mean regression functions via quantile regressions. Their proposed method is based on optimally combining information over multiple quantiles and can be applied to a broad range of parametric and nonparametric settings. They use the check-function-based method to estimate the conditional quantile functions. Under the location-scale model framework, one may use our proposed method to estimate the conditional quantile function. Then, combining information over multiple quantiles, as suggested by Zhao and Xiao (2014), may lead to a more efficient estimator of the conditional mean function. A theoretical investigation of such an estimation method is beyond the scope of the present paper.

#### 4. BOOTSTRAP INFERENCE

In this section, we turn our attention to inference on the conditional quantile function estimator. Many applied researchers are interested in the distributional treatment effects of a policy intervention. The inferential theory can be used to analyze issues such as (i) constructing the confidence band for the conditional quantile estimator and (ii) testing the distributional hypotheses such as the homogeneity or equality of quantile treatment effects. In this paper, we focus on (i) and use the bootstrap method to construct confidence intervals for the proposed conditional quantile estimator. There are two motivations for using a bootstrap method for inference. First, the conventional analytic asymptotic approximation entails estimation of covariances whose forms are somewhat complicated (Calonico, Cattaneo, and Titiunik, 2014, 2018). Second, bootstrapping often yields more reliable and accurate results (Beran, 1982; Hall, 1992). We use the residual bootstrap method to construct the confidence intervals.

##### 4.1. Uniform Bootstrap Confidence Interval Algorithm

To avoid calculating complicated leading bias terms in the conditional quantile estimator, we suggest using the undersmoothed smoothing parameters  $h_j$ ,  $j = 1, 2$ , in constructing bootstrap confidence bands. We use  $h_j^*$  and  $\hat{q}_\tau^*(x)$  to denote the corresponding smoothing parameters and the quantile estimator, using the bootstrap sample, respectively. For a function  $f \in \ell^\infty(\mathcal{T})$ , where  $\mathcal{T} = [\underline{\tau}, \bar{\tau}]$ ,  $0 < \underline{\tau} \leq \bar{\tau} < 1$ , we write  $\|f\|_{\mathcal{T}} = \sup_{\tau \in \mathcal{T}} |f(\tau)|$ . In addition, define

$$T_n(x) = \|\sqrt{nh_1}\{\hat{q}_\tau(x) - q_\tau(x)\}\|_{\mathcal{T}},$$

$$T_n^*(x) = \|\sqrt{nh_1^*}\{\hat{q}_\tau^*(x) - \hat{q}_\tau(x)\}\|_{\mathcal{T}},$$

and, for  $\alpha \in (0, 1)$ , let

$$c_{n,x}(\alpha) = \inf\{c \in \mathbf{R} : P_W(T_n^*(x) \leq c) \geq \alpha\},$$

where  $P_W$  is computed with respect to the bootstrap randomness conditional on the data. Then, our level  $1 - \alpha$  confidence band is given by

$$C_n(\alpha; x) = \left\{ \left[ \hat{q}_\tau(x) \pm \frac{1}{\sqrt{nh_1}} c_{n,x}(\alpha) \right] : \tau \in \mathcal{T} \right\}.$$

We propose a three-step procedure to construct the uniform bootstrap confidence intervals for the quantile process  $q_\tau(x)$  over the set  $\mathcal{T}$ . Let  $B$  denote the number of bootstraps. For each  $b = 1, \dots, B$ :

**Step 1.** Generate the bootstrap error  $\{\epsilon_i^*\}_{i=1}^n$  by random draws with replacement from the standardized residuals  $\{\hat{\epsilon}_i\}_{i=1}^n$ , where  $\hat{\epsilon}_i$  is defined as in (5). Then, construct the bootstrap sample by

$$Y_i^* = \hat{m}_{h_1}(X_i) + \hat{\sigma}_{h_2}(X_i)\epsilon_i^*, \quad i = 1, \dots, n.$$

**Step 2.** Discretize  $\mathcal{T}$  into a grid of quantile indexes  $\{\tau_1, \dots, \tau_m\}$ , where  $m$  is the number of grids. Then, use the bootstrap sample  $\{Y_i^*, X_i\}_{i=1}^n$  to compute the bootstrap versions of the conditional quantile estimates  $\hat{m}_{h_1^*}^*(x)$ ,  $\hat{\sigma}_{h_2^*}^*(x)$  and  $\hat{q}_{\tau_j}^*(x)$ , where

$$\hat{q}_{\tau_j}^*(x) = \hat{m}_{h_1^*}^*(x) + \hat{\sigma}_{h_2^*}^*(x)\hat{Q}_\epsilon(\tau_j), \quad j = 1, \dots, m,$$

where  $\hat{m}_{h_1^*}^*(x)$  and  $\hat{\sigma}_{h_2^*}^*(x)$  denote the estimators of  $m_0(x)$  and  $\sigma_0(x)$  with the bootstrap sample;  $\hat{Q}_\epsilon(\tau_j)$  is obtained in the same way with the smoothing parameter  $b_j$ , where  $j = 1, 2$ , as in Step 1 of Section 2.

**Step 3.** Calculate the maximum distance between  $\hat{q}_{\tau_j}^*(x)$  and  $\hat{q}_{\tau_j}(x)$  over  $\tau_j$ , and let  $\hat{T}_{n,b}^*(x) = \max_{\tau_j \in \{\tau_1, \dots, \tau_m\}} \sqrt{nh_1^*} |\hat{q}_{\tau_j}^*(x) - \hat{q}_{\tau_j}(x)|$ , where  $\hat{q}_{\tau_j}(x) = \hat{m}_{h_1}(x) + \hat{\sigma}_{h_2}(x)\hat{Q}_\epsilon(\tau_j)$ .

The above three steps produce  $B$  bootstrap supremum estimators  $\hat{T}_{n,1}^*(x), \dots, \hat{T}_{n,B}^*(x)$ . Sort  $\{\hat{T}_{n,b}^*(x)\}_{b=1}^B$  in an ascending order and obtain the sequence  $\hat{T}_{n,(1)}^*(x) \leq \dots \leq \hat{T}_{n,(B)}^*(x)$ . Let  $\hat{c}_{n,x}(\alpha) = \hat{T}_{n,((1-\alpha)B)}^*(x)$  denote the empirical  $1 - \alpha$  quantile of the simulated sample  $\{\hat{T}_{n,b}^*(x)\}_{b=1}^B$ . The  $100(1 - \alpha)\%$  uniform confidence band is estimated as  $\left[ \hat{q}_\tau(x) - \frac{1}{\sqrt{nh_1}} \hat{c}_{n,x}(\alpha), \hat{q}_\tau(x) + \frac{1}{\sqrt{nh_1}} \hat{c}_{n,x}(\alpha) \right]$ ,  $\forall \tau \in \mathcal{T}$ , where  $[a]$  is the nearest integer function and  $\alpha \in (0, 1)$  is the nominal size.

### 4.2. Uniform Bootstrap Confidence Interval Theory

In this subsection, we establish the validity of the bootstrap method. We first replace Assumption 3.5 with Assumption 4.1 below.

**Assumption 4.1.** The bandwidths  $h_j, h_j^*, j = 1, 2$ , satisfy that as  $n \rightarrow \infty$ , (i)  $h_j = c_j n^{-\gamma}$  for some finite and positive constant  $c_j$ , with  $\gamma \in (\frac{1}{5}, \frac{1}{3})$ ; (ii)  $h_j^* = c_j n^{-\tilde{\gamma}}$  for some finite and positive constant  $c_j$ ,  $\tilde{\gamma} > \gamma$  and  $\tilde{\gamma} \in (\frac{1}{5}, \frac{1}{3})$ .

**Remark 4.1.** Assumption 4.1(i) implies that  $h_j^2 = o((nh_j)^{-1/2})$ ,  $nh_j^3 / (\log n)^2 \rightarrow \infty$ . That is, the bias terms are asymptotically negligible compared with the variance terms. Assumption 4.1(ii) requires that  $h_j^*$ , the smoothing parameter used in estimating the conditional quantile function with the bootstrap sample, is further undersmoothed compared with  $h_j$ , for  $j = 1, 2$ . It also implies that  $n(h_j^*)^{3+\delta} \rightarrow \infty$ , for some  $\delta > 0$ , and  $\log(n)/(nh_j) = o(1/(nh_j^*))$ , for  $j = 1, 2$ . As suggested in Hall (1992), in order to avoid estimating the complicated bias terms in the bootstrap, one can use undersmoothing in estimating the unknown functions.

**THEOREM 4.1.** *If Assumptions 3.1–3.4 and 4.1 hold, then for any  $x$  at the interior of  $\mathcal{X}$ ,*

$$\sup_{c \in \mathbf{R}} |P_W(T_n^*(x) \leq c) - P(T_n(x) \leq c)| = o_p(1).$$

Theorem 4.1 shows that the bootstrap is consistent, and hence the bootstrap confidence band has an adequate coverage ratio, which is stated in Corollary 4.1 below.

**COROLLARY 4.1.** *If Assumptions 3.1–3.4 and 4.1 hold, then for any  $x$  at the interior of  $\mathcal{X}$ ,*

$$\liminf_{n \rightarrow \infty} P(q_\tau(x) \in \mathcal{C}_n(\alpha; x), \forall \tau \in \mathcal{T}) \geq 1 - \alpha.$$

**Remark 4.2.** Given the established uniformity in  $\tau \in \mathcal{T}$ , one may choose as many grid points as possible, subject to the computational capacity. We stress, however, that the discretization here is purely of computational rather than statistical nature. This is in contrast to Qu and Yoon (2015), whose estimators as processes are obtained by extrapolating (in a particular fashion) point estimators at a finite number of grid points.

**Remark 4.3.** As a referee correctly pointed out, because the residual distribution can be estimated at the  $\sqrt{n}$ -rate (if we further impose a condition that  $b_j = O(n^{-1/4}), j = 1, 2$ ) while the nonparametric functions  $m$  and  $\sigma^2$  are estimated at a slower rate, the estimator behaves as if the residual distribution was known in advance. Therefore, the process  $\hat{q}_\tau(x)$  (indexed by  $\tau \in \mathcal{T}$ ) converges (without bootstrap) in a simple “modular” fashion: once the CDF of residuals is estimated at a (uniform)  $\sqrt{n}$ -rate and Bahadur expansions for  $\hat{m}$  and  $\hat{\sigma}^2$  are obtained, the Bahadur representation for the final estimator follows easily. Therefore, a minor contribution of the paper is to establish Bahadur expansions for  $\hat{m}$  and  $\hat{\sigma}$  under a weaker condition  $b_j = o(n^{-1/5}), j = 1, 2$ , and the main technical innovation of this paper is to establish the bootstrap consistency.

## 5. THE LOCAL LINEAR QUANTILE ESTIMATOR

Up to now, we have concentrated on the local constant conditional quantile estimator. In this section, we briefly discuss how to extend the earlier results to the case of using local linear methods to estimate the conditional quantile function. For brevity and to avoid replicating similar proofs, we choose to present a simple case of estimating a  $\tau$ th conditional quantile function,  $q_\tau(x)$ , with a fixed  $\tau \in \mathcal{T}$ . We omit the trimming indicator function  $\mathbf{1}_{i,n} = \mathbf{1}(X_i \in \mathcal{X}_n)$ . Below, we first introduce some notation.

Define  $\mathbf{a}_\tau(x) = (q_\tau(x), \dot{q}_\tau(x))^T$ ,  $\mathbf{a}_1(x) = (m_0(x), \dot{m}_0(x))^T$ , and  $\mathbf{a}_2(x) = (\sigma_0(x), \dot{\sigma}_0(x))^T$ , where the superscript “ $T$ ” denotes the transpose of a matrix. Let  $\hat{\mathbf{a}}_1(x)$ ,  $\hat{\mathbf{a}}_2(x)$ , and  $\hat{Q}_{\epsilon,LL}(\tau)$  be the local linear estimators of  $\mathbf{a}_1(x)$ ,  $\mathbf{a}_2(x)$ , and  $Q_\epsilon(\tau)$  that are defined in the Supplementary Material, respectively. Then, the local linear estimator of  $\mathbf{a}_\tau(x)$  is given by:

$$\hat{\mathbf{a}}_\tau(x) = \hat{\mathbf{a}}_1(x) + \hat{\mathbf{a}}_2(x)\hat{Q}_{\epsilon,LL}(\tau),$$

where

$$\hat{\mathbf{a}}_\tau(x) = \begin{bmatrix} \hat{q}_{\tau,LL}(x) \\ \hat{q}'_{\tau,LL}(x) \end{bmatrix}, \quad \hat{\mathbf{a}}_1(x) = \begin{bmatrix} \hat{m}_{h_1,LL}(x) \\ \hat{m}'_{h_1,LL}(x) \end{bmatrix}, \quad \hat{\mathbf{a}}_2(x) = \begin{bmatrix} \hat{\sigma}_{h_2,LL}(x) \\ \hat{\sigma}'_{h_2,LL}(x) \end{bmatrix},$$

where  $\hat{m}_{h,LL}$  and  $\hat{m}'_{h,LL}$  denote the local linear estimators of  $m_0(x)$  and  $\dot{m}_0(x)$  with the smoothing parameter  $h$ . Similarly,  $\hat{\sigma}_{h,LL}$  and  $\hat{\sigma}'_{h,LL}$  are the local linear estimators of  $\sigma_0(x)$  and  $\dot{\sigma}_0(x)$ , with the smoothing parameter  $h$ .  $\hat{Q}_\epsilon(\tau)$  is the estimator of  $Q_\epsilon(\tau)$ , derived the same way as in the local constant case, except that the estimators of  $m_0(x)$  and  $\sigma_0(x)$  are replaced with the corresponding local linear counterparts.

**PROPOSITION 5.1.** *Under Assumptions 3.1–3.5, for any  $x$  at the interior of  $\mathcal{X}$  and any  $\tau \in \mathcal{T}$ ,*

$$\sqrt{nh_1} \begin{bmatrix} 1 & 0 \\ 0 & h_1 \end{bmatrix} (\hat{\mathbf{a}}_\tau(x) - \mathbf{a}_\tau(x) - h_1^2 \mathbf{B}_{1,LL}(x) - h_2^2 \mathbf{B}_{2,LL}(x)) \xrightarrow{d} N(0, \mathbf{V}_{\tau,LL}),$$

where

$$\mathbf{B}_{1,LL}(x) = \begin{bmatrix} \frac{\mu_2 \ddot{m}_0(x)}{2} \\ 0 \end{bmatrix}, \quad \mathbf{B}_{2,LL}(x) = \begin{bmatrix} \frac{\mu_2 Q_\epsilon(\tau) \ddot{\delta}_2(x)}{4\sigma_0(x)} \\ 0 \end{bmatrix}, \quad \mathbf{V}_{\tau,LL} = \begin{bmatrix} V_{\tau,LL,1} & 0 \\ 0 & V_{\tau,LL,2} \end{bmatrix},$$

$V_{\tau,LL,1} = V_\tau$  as defined in Corollary 3.1,

$$V_{\tau,LL,2} = \frac{\sigma_0^2(x)}{4\mu_2^2 f(x)} \left\{ 4v_2 + c^3 v_2 Q_\epsilon^2(\tau) [E(\epsilon^4) - 1] + 4c^3 Q_\epsilon(\tau) E(\epsilon^3) \int u^2 K(u) K(cu) du \right\},$$

$$\delta_2(x) = \sigma_0^2(x), \quad c = \lim_{n \rightarrow \infty} h_1/h_2, \quad \mu_2 = \int u^2 K(u) du, \\ v_j = \int u^j K^2(u) du \quad \text{for } j = 0, 2.$$

The proof of Proposition 5.1 is given in Section S2 in the Supplementary Material. The asymptotic distribution of  $\hat{q}_{\tau,LL}(x)$  can be obtained as a corollary of Proposition 5.1.

**COROLLARY 5.1.** *Under Assumptions 3.1–3.5, for any  $x$  at the interior of  $\mathcal{X}$  and any  $\tau \in \mathcal{T}$ ,*

$$\sqrt{nh_1} (\hat{q}_{\tau,LL}(x) - q_\tau(x) - h_1^2 D_{1,LL}(x) - h_2^2 D_{2,\tau,LL}(x)) \xrightarrow{d} N(0, V_\tau),$$

where  $D_{1,LL}(x) = (1/2)\mu_2 \ddot{m}_0(x)$  and  $D_{2,\tau,LL}(x) = \mu_2 Q_\epsilon(\tau) \ddot{\delta}_2(x) / (4\sigma_0(x))$ .

**Remark 5.1.** Compared with the local constant result given in Corollary 3.1, the local linear estimator  $\hat{q}_{\tau,LL}(x)$  has fewer bias terms, while the asymptotic variance is the same.

With an almost identical proof method and technicality, we can show that the weak convergence result holds for the local linear estimator, i.e., for any fixed  $x$  at the interior of  $\mathcal{X}$ ,

$$\sqrt{nh_1} \{ \hat{q}_{\tau,LL}(x) - q_{\tau}(x) - h_1^2 D_{1,LL}(x) - h_2^2 D_{2,\tau,LL}(x) \} \xrightarrow{L} \mathbb{G}(\tau) \text{ in } \ell^\infty(\mathcal{T}),$$

where  $\mathbb{G}(\tau)$  is the same Gaussian process as defined in Theorem 3.1.

Using a similar calculation to the one described in Section 4.1, one can compute uniform bootstrap confidence intervals of the local linear conditional quantile estimator.

### 6. MONTE CARLO SIMULATION

We conduct numerical experiments to assess two issues: (i) the finite-sample estimation MSE performance of the proposed estimators relative to the check-function-based estimator and the conditional quantile estimator proposed by Racine and Li (2017) and (ii) the performance of the estimated uniform confidence bands.

We consider the following DGP:

$$Y_i = \sin(3\pi X_i/2) [1 + 18X_i^2 (\text{sign}(X_i) + 1)]^{-1} + \sigma(X_i)\epsilon_i, \tag{8}$$

where  $X_i \sim \text{Uniform}[-1, 1]$  and  $\sigma(X_i) = 0.2 + 0.3X_i^2$ . For  $\epsilon$ , we consider three distributions: (i)  $\epsilon \sim N(0, 1)$ , (ii)  $\epsilon \sim \exp(1)$ , and (iii)  $\epsilon \sim \chi^2(5)$  distributions. For cases (ii) and (iii),  $\epsilon_i$  is normalized to have zero mean and unit variance. The above conditional mean function  $m(X_i)$  is adopted from Calonico et al. (2018).

#### 6.1. MSE Comparison

We conduct  $R = 1,000$  Monte Carlo replications for samples of sizes  $n = 100, 200,$  and  $400$ , along with quantiles  $\tau = \{0.1, 0.15, 0.25, 0.5, 0.75, 0.85, 0.9\}$ . We report MSE, which is computed as  $n^{-1} \sum_{i=1}^n (\hat{q}_{\tau}(X_i) - q_{\tau}(X_i))^2$ , where  $\hat{q}_{\tau}(X_i)$  and  $q_{\tau}(X_i)$  are the estimated and the true quantiles, respectively.

Assumption 3.2 requires that the kernel function has a bounded support. In fact, this assumption can be relaxed by, for example, allowing for the use of a Gaussian kernel, but with a more tedious proof. Throughout the simulations in this section, we use the Gaussian kernel. In Section S3 in the Supplementary Material (Table S3.1), we also report the performance of our method using the Epanechnikov kernel to examine whether estimation results are sensitive to different kernel functions. Our simulations show that the estimation results using different kernels are quite similar. This supports our claim that, in practice, one can use a Gaussian kernel. For the check-function-based estimator, we also use a Gaussian kernel (simulation results using the Epanechnikov kernel yield similar results).

Next, we discuss how to select the bandwidth. For our proposed method, the smoothing parameters are selected by the LS-CV method, as described in Remark 2.2. For the check-function-based method, we choose the bandwidth  $h$

by minimizing the following cross-validation function:

$$CV(h) = \frac{1}{n} \sum_{j=1}^n \rho_{\tau}(Y_j - \hat{q}_{-j,\tau}(X_j; h)),$$

where  $\hat{q}_{-j,\tau}(X_j; h)$  is the leave-one-out local constant check-function-based estimator of the conditional  $\tau$ th quantile of  $Y$ , given  $X_j$  (denoted by  $q_{\tau}(X_j)$ ) with bandwidth  $h$ , i.e.,

$$\hat{q}_{-j,\tau}(X_j; h) = \underset{\alpha}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1, i \neq j}^n \rho_{\tau}(Y_i - \alpha) K\left(\frac{X_i - X_j}{h}\right).$$

Tables 1–3 report the mean MSEs for the Gaussian,  $\chi^2(5)$ , and exponential error DGPs, respectively. For the Gaussian and  $\chi^2(5)$  errors (Tables 1 and 2), it can be seen that both our proposed method and Racine and Li’s method have smaller estimation MSEs than the check-function-based method for all cases considered. For the exponential error (Table 3), we observe that, in general, the check-function-based method performs better at lower quantiles, whereas our method and Racine & Li’s method dominate the check-function-based method for middle and upper quantiles. Next, we compare the performances of our method and Racine and Li’s method. For Gaussian error, our method has a smaller MSE than Racine and Li’s method in most cases. For the other two errors, Racine and Li’s method, in general, performs better than our method at lower and middle quantiles, whereas our method works better at higher quantiles.

**TABLE 1.** Mean MSEs ( $\times 100$ ) of different methods, Gaussian kernel,  $N(0, 1)$  errors

Estimator	Sample size	Quantile index						
		$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.85$	$\tau = 0.9$
Our method	100	1.9260	1.6588	1.4560	1.2797	1.5420	1.8785	2.2285
Check func.	100	3.8622	3.0701	2.3664	1.8560	2.4223	3.1301	4.0436
Racine & Li	100	2.1349	1.8451	1.5129	1.3574	1.7217	2.0972	2.4920
Our method	200	1.0658	0.9328	0.7739	0.6958	0.8401	1.0611	1.2423
Check func.	200	2.1890	1.7806	1.3940	1.0667	1.3894	1.8447	2.3929
Racine & Li	200	1.1572	0.9799	0.8163	0.7321	0.9018	1.1021	1.2892
Our method	400	0.5613	0.4916	0.4328	0.3925	0.4772	0.5721	0.6724
Check func.	400	1.2743	1.0266	0.7986	0.6050	0.8113	1.0906	1.3759
Racine & Li	400	0.6150	0.5277	0.4459	0.3915	0.4703	0.5803	0.6777

**TABLE 2.** Mean MSEs ( $\times 100$ ) of different methods, Gaussian kernel,  $\chi^2(5)$  errors

Estimator	Sample size	Quantile index						
		$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.85$	$\tau = 0.9$
Our method	100	1.1554	1.1200	1.0900	1.1327	1.7758	2.5944	3.4620
Check func.	100	1.4978	1.3362	1.2851	1.6800	3.3525	5.5493	7.9025
Racine & Li	100	1.2246	1.1465	1.0871	1.3426	2.4982	3.7047	4.8200
Our method	200	0.6498	0.6224	0.5721	0.6333	1.0495	1.5393	2.0460
Check func.	200	0.7051	0.7388	0.7228	0.9588	1.9892	3.1701	3.9033
Racine & Li	200	0.6595	0.6319	0.6237	0.7445	1.3612	2.0286	2.7289
Our method	400	0.3456	0.3308	0.3178	0.3543	0.5933	0.8775	1.1498
Check func.	400	0.4258	0.4029	0.3992	0.5422	1.1713	1.8333	2.4980
Racine & Li	400	0.3490	0.3365	0.3328	0.4063	0.7415	1.1682	1.5860

**TABLE 3.** Mean MSEs ( $\times 100$ ) of different methods, Gaussian kernel,  $\exp(1)$  errors

Estimator	Sample size	Quantile index						
		$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.85$	$\tau = 0.9$
Our method	100	0.9692	0.9585	1.0569	1.1236	1.7851	2.9394	4.1163
Check func.	100	0.9348	0.8386	0.8501	1.5101	3.8901	6.8101	9.8956
Racine & Li	100	1.0180	0.9223	0.8961	1.1306	2.4284	4.1157	5.7246
Our method	200	0.5329	0.5327	0.5488	0.5799	1.0405	1.6938	2.4337
Check func.	200	0.4059	0.3891	0.4355	0.8538	2.2474	3.8524	5.7474
Racine & Li	200	0.5037	0.4682	0.4614	0.5679	1.2802	2.2301	3.3084
Our method	400	0.3456	0.3308	0.2967	0.3186	0.6076	0.8775	1.1498
Check func.	400	0.1916	0.2000	0.2322	0.4907	1.3047	2.2468	3.2940
Racine & Li	400	0.2570	0.2461	0.2468	0.3107	0.6752	1.2307	1.8249

## 6.2. A Pretest Estimator

Our proposed estimator utilizes the local-scale model structure, whereas the check-function-based method does not. Hence, when the location-scale model structure does not hold, our estimator can be inferior to the check-function-based estimator. In practice, researchers might be concerned about the possible model misspecification. Following Racine and Li (2017), we suggest that one conduct a pretest and proceed with the check-function-based method if the test rejects the location-scale specification, and use our proposed method otherwise. This procedure will lead to a pretest estimator. For readers' convenience, we present the pretest procedure below.



The pretest method first assesses the adequacy of the location-scale model specification using the Kolmogorov–Smirnov test proposed by Einmahl and Van Keilegom (2008). We use the resampling bootstrap to obtain the critical values of the test statistic (the null hypothesis is that  $X_i$  and  $\epsilon_i$  are independent of each other):

$$T_{KS} = \sqrt{n} \sup_{x,t} |\hat{F}_{X,\hat{\epsilon}}(x,t) - \hat{F}_X \hat{F}_{\hat{\epsilon}}(t)|, \tag{9}$$

where  $\hat{F}_{X,\hat{\epsilon}}(x,t)$ ,  $\hat{F}_X(x)$ , and  $\hat{F}_{\hat{\epsilon}}(t)$  are empirical CDFs and  $\hat{\epsilon} = \frac{y - \hat{m}(x)}{\hat{\sigma}(x)}$ . Therefore, we only need to nonparametrically estimate  $\hat{m}$ ,  $\hat{\sigma}$  once to obtain the residuals  $\{\hat{\epsilon}_i\}_{i=1}^n$ . In each bootstrap loop, we resample  $\{\hat{\epsilon}_i\}_{i=1}^n$  and compute  $T_{KS}^*$  using the bootstrap sample

$$T_{KS}^* = \sqrt{n} \sup_{x,t} |\hat{F}_{X,\hat{\epsilon}^*}(x,t) - \hat{F}_X \hat{F}_{\hat{\epsilon}^*}(t)|, \tag{10}$$

where  $\hat{\epsilon}^*$  is the bootstrap residual. The empirical distribution of  $T_{KS}^*$  is used to obtain the critical values of the null distribution of  $T_{KS}$ . As the computation of  $TS_{KS}^*$  does not involve any optimization procedures, the computation of  $T_{KS}^*$  by the resampling process is fairly fast.

We now examine the performance of the pretest estimator. We consider the following location-shape DGP that violates the location-scale model assumption:

$$Y_i = \sin(3\pi X_i/2) [1 + 18X_i^2(\text{sign}(X_i) + 1)]^{-1} + \epsilon(X_i),$$

where  $X_i \sim \text{Uniform}[-1, 1]$ . We also consider two different error DGPs: (1)  $\epsilon(X_i)$  is a Gaussian mixture of  $N(-1, 0.5)$  and  $N(3, 0.5)$  and (2)  $\epsilon(X_i)$  is a Gaussian mixture of  $N(-1, 1)$  and  $N(1, 1)$ . For both cases, the mixing probability is  $(1 + X_i)/2 \in [0, 1]$ . Figure 1 presents the two error distributions. We use the same mean function as given in (8).

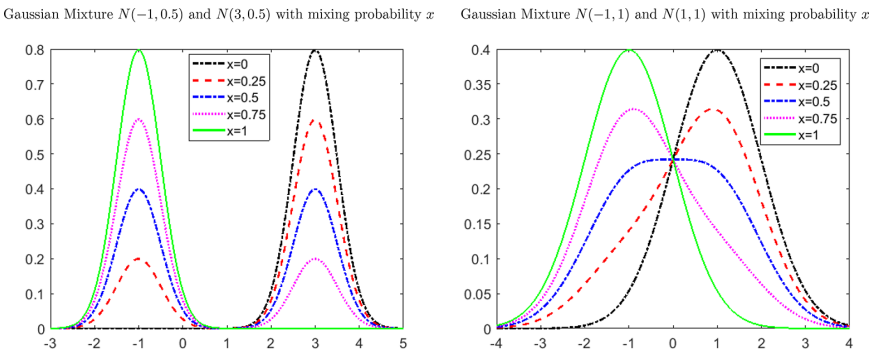


FIGURE 1. Location-shape error distributions.

**TABLE 4.** Mean MSEs of the pretest estimator, location-shape Gaussian mixture  $N(-1, 0.5)$ , and  $N(3, 0.5)$  errors

Estimator	Sample size	Quantile index						
		$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.85$	$\tau = 0.9$
Our method	100	0.8448	0.7430	0.7152	1.0358	0.6343	0.6072	0.7279
Check func.	100	0.7563	0.7166	0.5791	0.6012	0.4895	0.4962	0.5190
Pretest	100	0.7794	0.7207	0.6242	0.7478	0.5250	0.5257	0.5677
Our method	200	0.6501	0.5727	0.5734	0.9153	0.5137	0.4562	0.5667
Check func.	200	0.5040	0.4226	0.3705	0.4404	0.3332	0.3150	0.3664
Pretest	200	0.5069	0.4247	0.3731	0.4473	0.3347	0.3159	0.3681
Our method	400	0.5717	0.4910	0.5053	0.8059	0.4593	0.4206	0.5238
Check func.	400	0.3200	0.2864	0.2761	0.2629	0.2134	0.1993	0.2589
Pretest	400	0.3233	0.2872	0.2838	0.2877	0.2335	0.2165	0.2711

*Notes:* The mixing probability is  $(1 + x)/2 \in [0, 1]$ . The null is a location-scale mode, and the pretest is the check-function-based estimator if the null is rejected. Otherwise, the pretest estimator is our proposed quantile estimator. The nominal size is 0.05.

We conduct  $R = 1,000$  Monte Carlo replications for samples of size  $n = 100, 200,$  and  $400,$  along with quantiles  $\tau = \{0.1, 0.15, 0.25, 0.5, 0.75, 0.85, 0.9\}$ . For each replication, we first conduct a model specification test proposed by Einmahl and Van Keilegom (2008), with the null hypothesis being the location-scale model.<sup>5</sup> The pretest estimator is the check-function-based estimator if the null is rejected, and our proposed estimator otherwise.

Tables 4 and 5 report the mean MSEs of the pretest estimator. The MSE is calculated similar to the way it was calculated in Section 6.1. Table 4 reveals that the pretest estimator mimics the behavior of the check-function-based estimator. On the other hand, the table also reveals that the check-function-based method that does not require the location-shape assumption outperforms our estimator when the separation between the two normal distributions is large (i.e., the misspecification is severe). Table 5 suggests that our estimator holds its own when the misspecification is mild.

The simulation results show that the pretest procedure works well in guarding against severe misspecification.

### 6.3. Bootstrap Uniform Confidence Interval Coverage Ratio

In this subsection, we examine the coverage ratios of 95% and 90% uniform confidence intervals with sample sizes  $n = 100, 200,$  and  $400.$  We conduct  $R =$

<sup>5</sup>Einmahl and Van Keilegom (2008) test for the independence between the covariate(s)  $X$  and the error term  $\epsilon$  in the location-scale model (1). One constructs the Kolmogorov–Smirnov test statistic  $T_{KS} = \sqrt{nsup}_{x,t} |\hat{F}_{X,\hat{\epsilon}}(x,t) - \hat{F}_X \hat{F}_{\hat{\epsilon}}(t)|,$  where  $\hat{F}_{X,\hat{\epsilon}}(x,t), \hat{F}_X(x),$  and  $\hat{F}_{\hat{\epsilon}}(t)$  are empirical CDFs and  $\hat{\epsilon} = \frac{y - \hat{m}(x)}{\hat{\sigma}(x)}$ . We use the resampling bootstrap to obtain the critical values.

**TABLE 5.** Mean MSEs of the pretest estimator, location-shape Gaussian mixture  $N(-1, 1)$ , and  $N(1, 1)$  errors

Estimator	Sample size	Quantile index						
		$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.85$	$\tau = 0.9$
Our method	100	0.2084	0.1801	0.1569	0.1415	0.1444	0.1645	0.1907
Check func.	100	0.2731	0.2295	0.1835	0.1659	0.1674	0.1988	0.2215
Pretest	100	0.2207	0.1874	0.1603	0.1428	0.1480	0.1684	0.1947
Our method	200	0.1160	0.1010	0.0904	0.0830	0.0818	0.0882	0.0980
Check func.	200	0.1551	0.1254	0.1080	0.0981	0.1078	0.1210	0.1356
Pretest	200	0.1233	0.1056	0.0928	0.0839	0.0869	0.0966	0.1075
Our method	400	0.0747	0.0649	0.0565	0.0517	0.0525	0.0575	0.0663
Check func.	400	0.0865	0.0747	0.0627	0.0547	0.0591	0.0680	0.0770
Pretest	400	0.0763	0.0662	0.0575	0.0520	0.0531	0.0584	0.0679

*Notes:* The mixing probability is  $(1 + x)/2 \in [0, 1]$ . The null is a location-scale mode, and the pretest is the check-function-based estimator if the null is rejected. Otherwise, the pretest estimator is our proposed quantile estimator. The nominal size is 0.05.

1,000 Monte Carlo replications, and for each replication, we generate  $B = 500$  bootstrap statistics to construct uniform confidence intervals. We consider five evaluation points:  $x_0 \in \{-2/3, -1/3, 0, 1/3, 2/3\}$ , and three error distributions:  $N(0, 1)$ ,  $\chi^2(5)$ , and  $\exp(1)$ . The coverage ratio is defined as the number of times the confidence bands contain the complete quantile curve over  $\tau \in \mathcal{T} = [0.1, 0.9]$  divided by the number of bootstraps.

Our theoretical analysis assumes that  $h_j$  is undersmoothed relative to the optimal smoothing parameter that balances the estimated squared bias and variance and that  $h_j^*$  is further undersmoothed with respect to  $h_j$ , for  $j = 1, 2$ . In simulations, we use both the optimal and undersmoothed bandwidths. For undersmoothing, we choose  $h_j = n^{-\alpha_1} h_j^{opt}$ , and  $h_j^* = n^{-\alpha_2} h_j^{*opt}$ , where  $(\alpha_1, \alpha_2) = (1/20, 1/10)$ , and that  $h_j^* = o(h_j)$ ,  $h_j^{opt}$ , and  $h_j^{*opt}$  denote the optimal bandwidths obtained from leave-one-out LS-CV,  $j = 1, 2$ .

Table 6 presents the 95% coverage ratio of the uniform confidence interval using a Gaussian kernel. We observe that: first, as sample size increases, the coverage ratios get closer to nominal coverage probabilities; second, optimal smoothing can lead to undercoverage in some cases, and undersmoothing improves the coverage.<sup>6</sup> In general, our proposed bootstrap method has adequate coverage, and

<sup>6</sup>As pointed out in Calonico et al. (2014), if one uses the distributional approximation to construct the confidence interval, optimal smoothing will lead to a nonnegligible bias; conventional bias correction can also have very poor performance, because it may add to the finite-sample variability of the usual  $t$ -statistic. They proposed a robust bias-corrected  $t$ -statistic by accounting for the variability.

**TABLE 6.** The 95% coverage ratio of uniform confidence band over  $\tau \in [0.1, 0.9]$ , Gaussian kernel

Error distri.	Bandwidth	Eval. points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	Optimal	0.8370	0.8910	0.8680	0.9260	0.8930
	Undersmoothed	0.8800	0.9270	0.8830	0.9140	0.8700
$\chi^2(5)$	Optimal	0.8060	0.9310	0.8610	0.9290	0.8150
	Undersmoothed	0.8170	0.9360	0.8670	0.8980	0.8360
exp(1)	Optimal	0.7650	0.9170	0.8590	0.9260	0.8130
	Undersmoothed	0.7910	0.9310	0.8470	0.8830	0.8110
Panel B: $n = 200$						
$N(0,1)$	Optimal	0.8950	0.9290	0.9050	0.9490	0.9220
	Undersmoothed	0.9050	0.9320	0.9300	0.9330	0.9310
$\chi^2(5)$	Optimal	0.8570	0.9650	0.8900	0.9380	0.8980
	Undersmoothed	0.9100	0.9600	0.9120	0.9520	0.9280
exp(1)	Optimal	0.8610	0.9390	0.8900	0.9530	0.8740
	Undersmoothed	0.9150	0.9500	0.9120	0.9610	0.9390
Panel C: $n = 400$						
$N(0,1)$	Optimal	0.9080	0.9170	0.9070	0.9520	0.9250
	Undersmoothed	0.9410	0.9490	0.9470	0.9720	0.9600
$\chi^2(5)$	Optimal	0.9100	0.9620	0.8910	0.9700	0.9380
	Undersmoothed	0.9450	0.9730	0.9290	0.9610	0.9470
exp(1)	Optimal	0.8980	0.9560	0.9150	0.9570	0.9270
	Undersmoothed	0.9370	0.9750	0.9240	0.9440	0.9330

the results are robust to different error distributions and different kernel functions used.<sup>7</sup>

The 90% coverage ratio of the uniform confidence interval using a Gaussian kernel is shown in Table 7. We observe features similar to those in Table 6.

In Section S3 in the Supplementary Material (Tables S3.2 and S3.3), we also report the 95% and 90% coverage ratios of uniform confidence intervals over  $\tau \in \mathcal{T} = [0.2, 0.8]$ . We observe features similar to those in Tables 6 and 7. However, in general, the coverage ratios are more adequate than those over  $\tau \in \mathcal{T} = [0.1, 0.9]$ . The reason is that, as  $\tau$  approaches 0 and 1, the quantile estimation accuracy

<sup>7</sup>In Section S3 in the Supplementary Material (Table S3.4), we also report the 95% and 90% coverage ratios of the uniform confidence intervals over  $\tau \in \mathcal{T} = [0.2, 0.8]$  with an Epanechnikov kernel. The simulations show that the bootstrap confidence band coverage for using different kernels is quite similar. This suggests that, in practice, the estimated bootstrap uniform confidence interval is not sensitive to different kernel functions.

**TABLE 7.** The 90% coverage ratio of uniform confidence interval over  $\tau \in [0.1, 0.9]$ , Gaussian kernel

Error distri.	Bandwidth	Eval. points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0, 1)$	Optimal	0.7740	0.8370	0.7920	0.8790	0.8240
	Undersmoothed	0.7940	0.8720	0.8240	0.8670	0.7920
$\chi^2(5)$	Optimal	0.7350	0.8760	0.8040	0.8750	0.7460
	Undersmoothed	0.7420	0.8690	0.8060	0.8420	0.7580
exp(1)	Optimal	0.7010	0.8530	0.7920	0.8820	0.7270
	Undersmoothed	0.7050	0.8640	0.7780	0.8190	0.7230
Panel B: $n = 200$						
$N(0, 1)$	Optimal	0.8320	0.8630	0.8450	0.8900	0.8680
	Undersmoothed	0.8430	0.8810	0.8710	0.8880	0.8770
$\chi^2(5)$	Optimal	0.7830	0.9210	0.8350	0.8890	0.8280
	Undersmoothed	0.8410	0.9090	0.8520	0.9060	0.8690
exp(1)	Optimal	0.7890	0.8860	0.8310	0.9100	0.8110
	Undersmoothed	0.8530	0.8840	0.8370	0.9180	0.8690
Panel C: $n = 400$						
$N(0, 1)$	Optimal	0.8370	0.8500	0.8470	0.9000	0.8700
	Undersmoothed	0.8840	0.9130	0.9000	0.9290	0.9200
$\chi^2(5)$	Optimal	0.8440	0.9260	0.8480	0.9290	0.8760
	Undersmoothed	0.8780	0.9260	0.8630	0.9180	0.8890
exp(1)	Optimal	0.8380	0.9170	0.8540	0.9200	0.8680
	Undersmoothed	0.8690	0.9270	0.8730	0.8970	0.8780

gets less precise, which will undermine the coverage of the bootstrap confidence interval.

As a co-editor and a referee suggested, one may consider using the nonparametric resampling bootstrap method. The resampling bootstrap can be conducted in three steps, similar to the residual three-step bootstrap procedure described in Section 4.1. The first step generates the bootstrap sample  $\{Y_i^*, X_i^*\}_{i=1}^n$  by resampling with replacement from  $\{Y_i, X_i\}_{i=1}^n$ . The last two steps are the same as in the residual bootstrap procedure. In Section S3 in the Supplementary Material (Tables S3.5 and S3.6), we report the 95% and 90% coverage ratios of the resampling uniform confidence interval over  $\tau \in [0.2, 0.8]$ , under the same DGPs as in the residual bootstrap case.

One referee also suggested a score bootstrap method. We implement the score bootstrap method using the following steps.

First, we generate the bootstrap errors  $\{\epsilon_i^*\}_{i=1}^n$  by random draws with replacement from the standardized residuals  $\{\hat{\epsilon}_i\}_{i=1}^n$ , where  $\hat{\epsilon}_i$  is defined in (5). Then, we estimate  $m_{h_1}^*(x)$  and  $\sigma_{h_2}^*(x)$ , based on uniform Bahadur expansions given in Lemma A.5, i.e.,

$$\tilde{m}_{h_1}^*(x) = \hat{m}_{h_1}(x) + \frac{1}{nh_1^*} \sum_{i=1}^n \frac{\hat{\sigma}_0(X_i)\epsilon_i^*}{\hat{f}(x)} K\left(\frac{X_i - x}{h_1^*}\right), \tag{11}$$

$$\tilde{\sigma}_{h_2}^*(x) = \hat{\sigma}_{h_2}(x) + \frac{1}{nh_2^*} \sum_{i=1}^n \frac{\hat{\sigma}_0^2(X_i)(\hat{\epsilon}_i^{*2} - 1)}{2\hat{\sigma}_0(x)\hat{f}(x)} K\left(\frac{X_i - x}{h_2^*}\right). \tag{12}$$

Note that we only need to estimate  $\sigma_0(X_i)$ ,  $\sigma_0(x)$ , and  $f(x)$  once in (11) and (12). We do not have to estimate them in each bootstrap loop.

Second, we compute the score bootstrap version conditional quantile estimator  $\hat{q}_{\tau_j}^*(x)$ , which is defined by

$$\hat{q}_{\tau_j}^*(x) = \tilde{m}_{h_1}^*(x) + \tilde{\sigma}_{h_2}^*(x)\hat{Q}_\epsilon(\tau_j), \quad j = 1, \dots, m,$$

where  $\tilde{m}_{h_1}^*(x)$  and  $\tilde{\sigma}_{h_2}^*(x)$  are defined in (11) and (12), respectively, and  $\hat{Q}_\epsilon(\tau_j)$  is obtained in the same way with the smoothing parameter  $b_j, j = 1, 2$ , as in Step 1 of Section 2.

The last step is the same as Step 3 of the residual bootstrap described in Section 4.1, and we can also construct the uniform confidence interval accordingly.

In Section S3 in the Supplementary Material (Tables S3.7 and S3.8), we report the 95% and 90% coverage ratios of score bootstrapped uniform confidence intervals, under the same DGPs as in the residual bootstrap case. Simulation results show that the score bootstrap works reasonably well. In addition, in Table S3.9, we also present a computation time comparison between the residual bootstrap and the score bootstrap. The score bootstrap method only takes about 20% ~ 30% of the computation time of the residual bootstrap method for the sample sizes we considered. Therefore, the score bootstrap method is computationally more efficient. We observe that, with sample size  $n = 100$ , the score bootstrap takes about 30% of the residual bootstrap computation time, and as sample size increases to  $n = 400$ , this number drops to 20%. We expect that for larger sample sizes (say, in a big data scenario), the computational advantage of the score bootstrap over the residual-based bootstrap method will be even more substantial.

## 7. AN EMPIRICAL APPLICATION

In this section, we present an illustrative empirical example to compare the performances of our proposed method with those of the check-function-based method.

## 7.1. Canadian High School Graduate Earnings

We use a Canadian cross-section wage dataset. The data consist of a random sample taken from the 1971 Canadian Census Public Use Tapes for male individuals having common education (grade 13). The sample size is 205. The data contain two variables: “logwage” (logarithm of the wages) and “age” (years). We are interested in estimating the conditional quantiles of logwage given age,  $q_\tau(\text{age})$ , as well as its derivative function. Therefore, we choose to use the local linear method to estimate  $m_0(x)$ . However, when applying the local linear method to estimate  $\sigma_0(x)$ , sometimes we get a negative estimated value of  $\hat{\sigma}_{h_2}(x)$ . To avoid this problem, we first use the local constant method to estimate  $\sigma_0(x)$ . Then, we take the derivative of  $\hat{\sigma}_{h_2}(x)$  with respect to  $x$  to obtain an estimate of  $\hat{\sigma}_0(x)$  (we use a Gaussian kernel so that the kernel function is differentiable).

Figure 2 plots estimated Canadian logwage quantile curves using our proposed method and the check-function-based method. We see that for low quantiles ( $\tau = 0.1$  and  $0.15$ ), the wages started to decrease after age 35, whereas for upper quantiles ( $\tau = 0.85$  and  $0.9$ ), the wages started to decrease around age 55. It is likely that high-quantile people have higher unobserved characteristics (such as

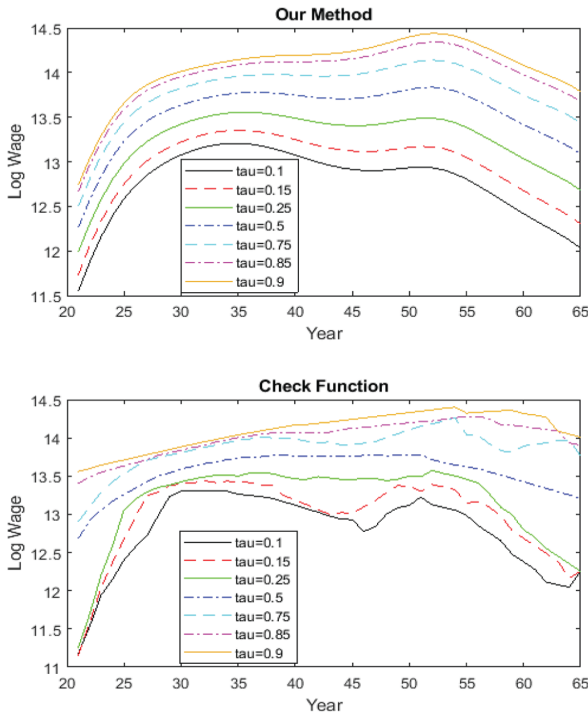


FIGURE 2. Logwage conditional quantile estimated curve.

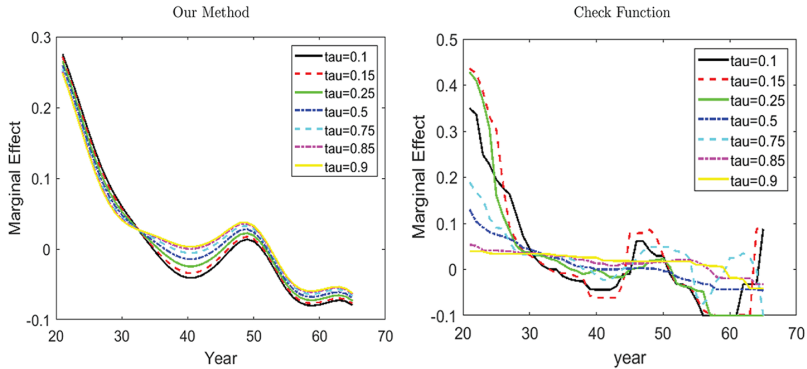


FIGURE 3. Logwage conditional quantile derivative estimated curve.

ability) that cause their wages to continue to increase until age 55, whereas at the low quantile, people may hold more labor-intensive jobs whose wages start to decrease at a much younger age compared with those of high-quantile individuals. In Section S3 in the Supplementary Material, we provide estimation results using wage (to replace logwage) as the dependent variable.

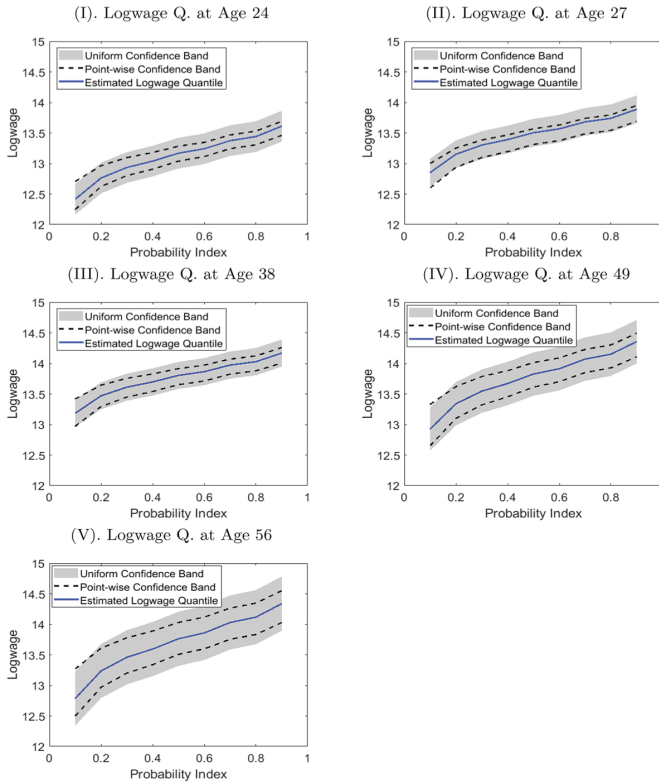
Figure 3 plots the corresponding quantile derivative curves. The check-function-based method estimated quantile derivative curve is very wiggly, whereas our proposed method gives a much smoother estimated derivative curve. From Figure 3, we can see the low quantiles' derivative functions have steeper downward slopes than those of upper quantiles. All the derivative curves reach the maximum deceleration at around age 40. After age 40, the low-quantile derivative curves still remain at the negative regions, meaning that their logwage still decreases as age goes up, although the rate of decline is less than that at age 40. In contrast, for high quantiles, the derivative curves remain positive until close to age 55. We see that smooth curves estimated by our method allow us to better interpret the estimation results, whereas it is hard to do so with the wiggly curves obtained using the check-function-based method.

Finally, Figure 4 presents the uniform confidence band (over  $\tau \in [0.1, 0.9]$ ) for the logwage conditional quantile estimated curve.

## 8. CONCLUSION

In this paper, we propose a new and easy-to-implement nonparametric method for estimating conditional quantile functions. We derive the asymptotic theory and provide a practical procedure for constructing uniform confidence bands. The proposed conditional quantile estimator compares well with the check-function counterpart, and the bootstrap confidence interval has adequate coverage probabilities. An empirical application using a Canadian cross-section wage dataset showcases the appealing feature of our proposed method in practice.





**FIGURE 4.** Uniform confidence bands of logwage conditional quantile estimated curves.  
*Notes:* Subfigures (I)–(V) plot the conditional quantiles of logwage given age, with age at 24, 27, 38, 49, and 56 (which correspond to the 10th, 25th, 50th, 75th, and 90th percentiles of age, respectively).

There are many directions in which one can extend the results of this paper to more general settings. For example, one can allow for time-series-dependent data, as in Han et al. (2016), or allow for panel nonstationary data considered in Chen and Khan (2008). Another extension is to allow for the covariate  $X_i$  to be endogenous, as in Horowitz and Lee (2007), Su and Hoshino (2016), and Kaplan and Sun (2017). We leave these as possible future research topics.

## APPENDIX

### A. Some Useful Lemmas

To avoid the slow convergence rate at the boundary region, we will estimate  $m_0(x)$  and  $\sigma_0(x)$ , for  $x \in \mathcal{X}_n$ , where  $x \in \mathcal{X}_n$  is the trimmed set defined earlier (see the beginning part of Section 2). Asymptotically,  $P(X \in \mathcal{X}_n) \rightarrow P(X \in \mathcal{X}) = 1$ . In this appendix, we present a

few lemmas that are used to prove the main results of the paper. The proofs of these lemmas are given in Section S1 in the Supplementary Material.

LEMMA A.1. *Suppose that Assumptions 3.1–3.3 and 3.5 hold. Then, we have:*

- (i)  $\sup_{x \in \mathcal{X}_n} |\hat{m}_{h_1}(x) - m_0(x)| = O_p\left(h_1^2 + \sqrt{\frac{\log(n)}{nh_1}}\right),$
- (ii)  $\sup_{x \in \mathcal{X}_n} |\hat{f}_{h_1}(x) - f(x)| = O_p\left(h_1^2 + \sqrt{\frac{\log(n)}{nh_1}}\right),$
- (iii)  $\sup_{x \in \mathcal{X}_n} |\hat{\sigma}_{h_2}(x) - \sigma_0(x)| = O_p\left(h_1^2 + h_2^2 + \sqrt{\frac{\log(n)}{nh_1}} + \sqrt{\frac{\log(n)}{nh_2}}\right).$

LEMMA A.2. *If Assumptions 3.1–3.3 and 3.5 hold, we have: uniformly in  $x \in \mathcal{X}_n$ ,*

(i)  $\hat{m}_{b_1}(x) - m_0(x) = b_1^2 B_1(x) + \frac{1}{nb_1} \sum_{i=1}^n \sigma_0(X_i) \epsilon_i K\left(\frac{X_i - x}{b_1}\right) + o_p\left(\frac{1}{\sqrt{n}}\right),$

where  $B_1(x) = \frac{1}{2} \mu_2 \left[ 2\dot{m}_0(x)\dot{f}(x)/f(x) + \ddot{m}_0(x) \right]$ , and  $\mu_2 = \int u^2 K(u)du$ , and

(ii)  $\hat{\sigma}_{b_2}(x) - \sigma_0(x) = b_2^2 B_2(x) + \frac{1}{2nb_2 f(x)\sigma_0(x)} \sum_{i=1}^n \sigma_0^2(X_i) (\epsilon_i^2 - 1) K\left(\frac{X_i - x}{b_2}\right) + o_p\left(\frac{1}{\sqrt{nb_2}}\right),$

where  $B_2(x) = \frac{\mu_2}{4\sigma_0(x)} \{ 2\dot{\delta}_2(x)\dot{f}(x)/f(x) + \ddot{\delta}_2(x) \}$ , and  $\delta_2(x) = \sigma_0^2(x)$ .

LEMMA A.3. *Recall that  $\dot{m}_0(x) = dm_0(x)/dx$ , and  $\hat{m}_h(x) = d\hat{m}_h(x)/dx$ . Under Assumptions 3.1–3.3 and 3.5, we have:*

- (i)  $\sup_{x \in \mathcal{X}_n} |\hat{m}_{h_1}(x) - \dot{m}(x)| = O_p(h_1^2 + (\log n/(nh_1^3))^{-1/2}),$
- (ii)  $\sup_{x \in \mathcal{X}_n} |\hat{\sigma}_{h_2}(x) - \dot{\sigma}_0(x)| = O_p(c_n),$

where

$c_n = h_1^2 + h_2^2 + (\log n/(nh_1^3))^{-1/2} + (\log n/(nh_2^3)).$  (A.1)

LEMMA A.4. *Under Assumptions 3.1–3.3 and 3.5, we have, for  $0 < \delta \leq 1$ :*

- (i)  $\sup_{x, x' \in \mathcal{X}_n} \frac{|\hat{m}_h(x) - \dot{m}(x) - (\hat{m}_{h_1}(x') - \dot{m}(x'))|}{|x - x'|^\delta} = O_p(h_1^2 + (\log n/(nh_1^3))^{-1/2}),$
- (ii)  $\sup_{x, x' \in \mathcal{X}_n} \frac{|\hat{\sigma}_{h_2}(x) - \dot{\sigma}_0(x) - (\hat{\sigma}_h(x') - \dot{\sigma}_0(x'))|}{|x - x'|^\delta} = O_p(c_n).$

LEMMA A.5. Suppose that Assumptions 3.1–3.4 and 4.1 hold. Then, uniformly in  $x \in \mathcal{X}_n$ , in  $P_Z$ -probability,<sup>8</sup>

$$(i) \quad \hat{m}_{h_1^*}^*(x) - \hat{m}_{h_1}(x) = \frac{1}{nh_1^*} \sum_{i=1}^n \frac{\sigma_0(X_i)\epsilon_i^*}{f(x)} K\left(\frac{X_i - x}{h_1^*}\right) + o_{P_W}\left(\frac{1}{\sqrt{nh_1^*}}\right),$$

where  $P_W$  denotes the law governing the bootstrap randomness conditional on the data, and

$$(ii) \quad \hat{\sigma}_{h_2^*}^*(x) - \hat{\sigma}_{h_2}(x) = \frac{1}{nh_2^*} \sum_{i=1}^n \frac{\sigma_0^2(X_i)(\epsilon_i^{*2} - 1)}{2\sigma_0(x)f(x)} K\left(\frac{X_i - x}{h_2^*}\right) + o_{P_W}((nh_1^*)^{-1/2} + (nh_2^*)^{-1/2}).$$

LEMMA A.6. Suppose that Assumptions 3.1–3.4 and 4.1 hold. Then, uniformly in  $x \in \mathcal{X}_n$ , in  $P_Z$ -probability,

$$(i) \quad \sup_{x \in \mathcal{X}_n} |\hat{m}_{h_1^*}^*(x) - \hat{m}_{h_1}(x)| = O_{P_W}(c_n^*),$$

$$(ii) \quad \sup_{x \in \mathcal{X}_n} |\hat{\sigma}_{h_2^*}^*(x) - \hat{\sigma}_{h_2}(x)| = O_{P_W}(c_n^*),$$

for some  $\delta > 0$ , where  $c_n^* = h_1^{*2} + h_2^{*2} + (\log n / (nh_1^{*3}))^{-1/2} + (\log n / (nh_2^{*3}))^{-1/2}$ .

LEMMA A.7. Under Assumptions 3.1–3.3 and 4.1, we have, for  $0 < \delta \leq 1$ :

$$(i) \quad \sup_{x, x' \in \mathcal{X}_n} \frac{|\hat{m}_{h_1^*}^*(x) - \hat{m}_{h_1}(x) - (\hat{m}_{h_1^*}^*(x') - \hat{m}_{h_1}(x'))|}{|x - x'|^\delta} = O_{P_W}(c_n^*),$$

$$(ii) \quad \sup_{x, x' \in \mathcal{X}_n} \frac{|\hat{\sigma}_{h_2^*}^*(x) - \hat{\sigma}_{h_2}(x) - (\hat{\sigma}_{h_2^*}^*(x') - \hat{\sigma}_{h_2}(x'))|}{|x - x'|^\delta} = O_{P_W}(c_n^*),$$

for some  $\delta > 0$ .

LEMMA A.8. Let  $P_Z$  denote the joint law of  $Z \equiv (Y, X)$  generated according to (1),  $P_X$  the marginal law of  $X$ , and  $\mathcal{G}_m$  and  $\mathcal{G}_s$  classes of real-valued functions on  $\mathcal{X}$ , respectively, uniformly bounded above and below. Define

$$\mathcal{F} \equiv \{f_{s, \eta} : f_{s, \eta}(y, x) = \mathbf{I}\{y \leq m(x) + \sigma(x)s\}, s \in \mathbf{R}, m \in \mathcal{G}_m, \sigma \in \mathcal{G}_s\}.$$

Then, under Assumptions 3.1, 3.3, and 3.4, we have, for any  $r \geq 1$ , any  $\epsilon \in (0, 1)$ , and some constant  $C > 0$ ,

$$N_{[]}(\epsilon, \mathcal{F}, L^r(P_Z)) \lesssim N_{[]}(\epsilon C^r, \mathcal{G}_m, L^r(P_X)) N_{[]}(\epsilon C^r, \mathcal{G}_s, L^r(P_X)) \frac{1}{\epsilon^2}.$$

LEMMA A.9. Suppose that Assumptions 3.1(i)(ii)–3.4 and 4.1 hold. Let  $\tilde{\mathcal{G}}_n^*(x)$  and  $\tilde{\mathcal{G}}(x)$  be defined as in the proof of Theorem 4.1, and let  $E_W(\cdot)$  denote conditional expectation given

<sup>8</sup>See, for example, Cheng and Huang (2010, p. 2891) for a precise definition of probability orders under  $P_W$  in  $P_Z$ -probability, where  $P_Z$  denotes the joint law of the data.

the data. Then, for each fixed  $x$  at the interior of  $\mathcal{X}$ ,

$$\sup_{f \in \text{BL}_1(\mathbb{R}^2)} |E_W[f(\tilde{G}_n^*(x))] - E[f(\tilde{G}(x))]| = o_P(1).$$

LEMMA A.10. Suppose that Assumptions 3.1–3.4 hold. Then,

$$(i) \quad \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^3 \xrightarrow{P} E[\epsilon_i^3], \quad (ii) \quad \frac{1}{n} \sum_{i=1}^n (\hat{\epsilon}_i^2 - 1)^2 \xrightarrow{P} E[(\epsilon_i^2 - 1)^2].$$

**B. Proofs of Main Results**

Before the proof of Theorem 3.1, we first give the following proposition.

PROPOSITION B.1. Under Assumptions 3.1–3.5, we have that, uniformly in  $\tau \in \mathcal{T}$ ,

$$\{\hat{Q}_\epsilon - Q_\epsilon\}(\tau) = O_p(n^{-1/2} + b_1^2 + b_2^2) = o_p(n^{-\frac{2}{5}}). \tag{B.1}$$

Under Assumptions 3.1–3.5 and using a similar proof as the proof of Theorem 2 in Akritas and Van Keilegom (2001), we can show that the process  $\hat{F}_\epsilon(\cdot) - F_\epsilon(\cdot)$  is of order  $o_p(n^{-2/5})$ , where  $\hat{F}_\epsilon$  is the empirical distribution function of the residuals  $\{\hat{\epsilon}_i\}_{i=1}^n$ . Then, by Theorem 3.9.4 (delta method) of van der Vaart and Wellner (1996), we will have the result (B.1).

**Proof of Theorem 3.1:** Fix  $x \in \mathcal{X}_n$  throughout. Note that

$$\hat{q}_\tau(x) - q_\tau(x) = \hat{m}_{h_1}(x) - m_0(x) + Q_\epsilon(\tau)[\hat{\sigma}_{h_2}(x) - \sigma_0(x)] + \hat{\sigma}_{h_2}(x)[\hat{Q}_\epsilon(\tau) - Q_\epsilon(\tau)].$$

By Proposition B.1, uniformly in  $\tau \in \mathcal{T}$ ,

$$\{\hat{Q}_\epsilon - Q_\epsilon\}(\tau) = o_p(n^{-\frac{2}{5}}). \tag{B.2}$$

Therefore, equation (B.2), Lemma A.2, and Assumption 3.5 allow us to further write

$$\begin{aligned} \hat{q}_\tau(x) - q_\tau(x) &= \frac{1}{nh_1 f(x)} \sum_{i=1}^n \sigma_0(X_i) \epsilon_i K\left(\frac{X_i - x}{h_1}\right) \\ &\quad + \frac{Q_\epsilon(\tau)}{2nh_2 f(x) \sigma_0(x)} \sum_{i=1}^n \sigma_0^2(X_i) (\epsilon_i^2 - 1) K\left(\frac{X_i - x}{h_2}\right) \\ &\quad + h_1^2 D_1(x) + h_2^2 D_{2,\tau}(x) + o_p(\zeta_n), \end{aligned}$$

where  $\zeta_n = h_1^2 + h_2^2 + (nh_1)^{-1/2} + (nh_2)^{-1/2}$ ,  $D_1(x) = \frac{1}{2} \mu_2[2\dot{m}_0(x)\dot{f}'(x)/f(x) + \ddot{m}_0(x)]$ ,  $D_{2,\tau}(x) = \frac{Q_\epsilon(\tau)}{4\sigma_0(x)} \mu_2\{2\dot{\delta}_2(x)\dot{f}'(x)/f(x) + \ddot{\delta}_2(x)\}$ , and  $\delta_2(x) = \sigma_0^2(x)$ .

Then, by the Lyapunov central limit theorem, we have that

$$\sqrt{nh_1}(\hat{q}_\tau(x) - q_\tau(x) - h_1^2 D_1(x) - h_2^2 D_{2,\tau}(x)) \xrightarrow{L} \mathbb{G},$$

where  $\mathbb{G}$  is a zero-mean Gaussian process, and the covariance structure by

$$\begin{aligned} \text{Cov}(\mathbb{G}(\tau_1), \mathbb{G}(\tau_2)) &= \frac{v_0 \sigma_0^2(x)}{f(x)} + \frac{c \sigma_0^2(x) E(\epsilon^3) \int K(u) K(cu) du}{2f(x)} [Q_\epsilon(\tau_1) + Q_\epsilon(\tau_2)] \\ &\quad + \frac{c v_0 Q_\epsilon(\tau_1) Q_\epsilon(\tau_2)}{4f(x)} \sigma_0^2(x) [E(\epsilon^4) - 1], \end{aligned}$$

with  $v_0 = \int K^2(u) du$ ,  $c = \lim_{n \rightarrow \infty} h_1/h_2$ .

This completes the proof of Theorem 3.1. □

**Proof of Theorem 4.1:** By the triangle inequality and Theorem 3.1 together with the undersmoothing condition in Assumption 4.1, it is equivalent to show that

$$\sup_{c \in \mathbf{R}} |P_W(T_n^*(x) \leq c) - P(T_n(x) \leq c)| = o_p(1),$$

where  $T_n(x)$  and  $T_n^*(x)$  are defined in Section 4.2.

From the proof of Theorem 3.1,

$$\mathbb{G}(x)(\tau) = \mathbb{G}_m(x) + \mathbb{G}_\sigma(x) Q_\epsilon(\tau) \equiv \phi(\bar{\mathbb{G}}(x))(\tau),$$

where  $\bar{\mathbb{G}}(x) \equiv (\mathbb{G}_m(x), \mathbb{G}_\sigma(x))$ , and  $\phi: \mathbf{R}^2 \rightarrow \ell^\infty(\mathcal{T})$  is the (unknown) map defined by, for any  $\theta = (\theta^{(1)}, \theta^{(2)}) \in \mathbf{R}^2$ ,

$$\phi(\theta)(\tau) = \theta^{(1)} + Q_\epsilon(\tau) \theta^{(2)}.$$

Next, we write

$$\bar{\mathbb{G}}_n^*(x) \equiv \begin{bmatrix} \sqrt{nh_1^*} \{ \hat{m}_{h_1^*}^*(x) - \hat{m}_{h_1}(x) \} \\ \sqrt{nh_2^*} \{ \hat{\sigma}_{h_2^*}^*(x) - \hat{\sigma}_{h_2}(x) \} \end{bmatrix},$$

and define  $\hat{\phi}_n: \mathbf{R}^2 \rightarrow \ell^\infty(\mathcal{T})$  by, for any  $\theta = (\theta^{(1)}, \theta^{(2)}) \in \mathbf{R}^2$ ,

$$\hat{\phi}_n(\theta)(\tau) = \theta^{(1)} + \hat{Q}_\epsilon(\tau) \theta^{(2)}.$$

By Lemma A.5 and Proposition B.1, the map  $\hat{\phi}_n$  trivially satisfies Assumption 3.3 in Fang and Santos (2018). This, together with Lemma A.9, allows us to conclude by Theorem 3.2 in Fang and Santos (2018) that

$$\sup_{f \in \text{BL}_1(\mathbf{R})} |E[f(\hat{\phi}_n(\bar{\mathbb{G}}_n(x))) | \{Y_i, X_i\}_{i=1}^n] - E[f(\phi(\bar{\mathbb{G}}(x)))]| = o_p(1).$$

By Proposition 10.7(i) in Kosorok (2008) and the definitions of  $T_n^*(x)$  and  $\hat{\phi}_n$ , it then follows that

$$\sup_{f \in \text{BL}_1(\mathbf{R})} |E[f(T_n^*(x)) | \{Y_i, X_i\}_{i=1}^n] - E[f(T_n(x))]| = o_p(1).$$

The theorem now follows from Lemma 10.11(i) in Kosorok (2008). □

### Supplementary Material

To view the supplementary material for this article, please visit: <https://dx.doi.org/10.1017/S0266466621000499>.

## REFERENCES

- Akritis, M.G. & I. Van Keilegom (2001) Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics* 28, 549–567.
- Angrist, J.D., V. Chernozhukov, & I. Fernández-Val (2006) Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* 74, 539–563.
- Beran, R. (1982) Estimated sampling distributions: The bootstrap and competitors. *Annals of Statistics* 10, 212–225.
- Birke, M., N. Neumeier, & S. Volgushev (2017) The independence process in conditional quantile location-scale models and an application to testing for monotonicity. *Statistica Sinica* 27, 1815–1839.
- Calonico, S., M.D. Cattaneo, & M.H. Farrell (2018) On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* 113, 767–779.
- Calonico, S., M.D. Cattaneo, & R. Titiunik (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82, 2295–2326.
- Chaudhuri, P. (1991) Nonparametric estimates of regression quantiles and their local bahadur representation. *Annals of Statistics* 19, 760–777.
- Chen, S., G.B. Dahl, & S. Khan (2005) Nonparametric identification and estimation of a censored location-scale regression model. *Journal of the American Statistical Association* 100, 212–221.
- Chen, S. & S. Khan (2008) Semiparametric estimation of nonstationary censored panel data models with time varying factor loads. *Econometric Theory* 24, 1149–1173.
- Cheng, F. (2002) Consistency of error density and distribution function estimators in nonparametric regression. *Statistics and Probability Letters* 59, 257–270.
- Cheng, G. & J. Huang (2010) Bootstrap consistency for general semiparametric  $m$ -estimation. *Annals of Statistics* 38, 2884–2915.
- Chernozhukov, V., I. Fernández-Val, & A. Galichon (2010) Quantile and probability curves without crossing. *Econometrica* 78, 1093–1125.
- Dette, H. & S. Volgushev (2008) Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70, 609–627.
- Einmahl, J.H.J. & I. Van Keilegom (2008) Specification tests in nonparametric regression. *Journal of Econometrics* 143, 88–102.
- Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. CRC Press.
- Fan, J. & Q. Yao (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85, 645–660.
- Fang, Z. & A. Santos (2018) Inference on directionally differentiable functions. *The Review of Economic Studies* 86, 377–412.
- Florens, J.-P., L. Simar, & I.V. Keilegom (2014) Frontier estimation in nonparametric location-scale models. *Journal of Econometrics* 178(Part 3), 456–470.
- Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. Springer.
- Han, H., O. Linton, T. Oka, & Y.-J. Whang (2016) The cross-quantilegram: Measuring quantile dependence and testing directional predictability between time series. *Journal of Econometrics* 193, 251–270.
- He, X. (1997) Quantile curves without crossing. *The American Statistician* 51, 186–192.
- Heuchenne, C. & I.V. Keilegom (2007) Location estimation in nonparametric regression with censored data. *Journal of Multivariate Analysis* 98, 1558–1582.
- Honda, T. (2004) Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference* 121, 113–125.
- Horowitz, J.L. & S. Lee (2005) Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* 100, 1238–1249.
- Horowitz, J.L. & S. Lee (2007) Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica* 75, 1191–1208.

- Kaplan, D.M. & Y. Sun (2017) Smoothed estimating equations for instrumental variables quantile regression search within citing articles. Technical Report 1.
- Kiwitt, S. & N. Neumeyer (2012) Estimating the conditional error distribution in non-parametric regression. *Scandinavian Journal of Statistics* 39, 259–281.
- Koenker, R., P. Ng, & S. Portnoy (1994) Quantile smoothing splines. *Biometrika* 81, 673–680.
- Koenker, R. & Z. Xiao (2002) Inference on the quantile regression process. *Econometrica* 70, 1583–1612.
- Kosorok, M. (2008) *Introduction to Empirical Processes and Semiparametric Inference*. Springer.
- Li, D. & Q. Li (2010) Nonparametric/semiparametric estimation and testing of econometric models with data dependent smoothing parameters. *Journal of Econometrics* 157, 179–190.
- Li, Q., J. Lin, & J.S. Racine (2013) Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics* 31, 57–65.
- Mammen, E. (1991) Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics* 19, 741–759.
- Neumeyer, N. (2008) A bootstrap version of the residual-based smooth empirical distribution function. *Journal of Nonparametric Statistics* 20, 153–174.
- Neumeyer, N. (2009a) Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scandinavian Journal of Statistics* 36, 204–228.
- Neumeyer, N. (2009b) Testing independence in nonparametric regression. *Journal of Multivariate Analysis* 100, 1551–1566.
- Phillips, P.C.B. (2015) Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* 13, 521–555.
- Qu, Z. & J. Yoon (2015) Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics* 185, 1–19.
- Racine, J.S. & K. Li (2017) Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach. *Journal of Econometrics* 201, 72–94.
- Su, L. & T. Hoshino (2016) Sieve instrumental variable quantile regression estimation of functional coefficient models. *Journal of Econometrics* 191, 231–254.
- van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer.
- Van Keilegom, I. & M.G. Akritas (1999) Transfer of tail information in censored regression. *Annals of Statistics* 27, 1745–1784.
- Yu, K. & M. Jones (1998) Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.
- Zhao, Z. & Z. Xiao (2014) Efficient regressions via optimally combining quantile information. *Econometric Theory* 30, 1272–1314.