



## Single-index Thresholding in Quantile Regression

Yingying Zhang, Huixia Judy Wang & Zhongyi Zhu

To cite this article: Yingying Zhang, Huixia Judy Wang & Zhongyi Zhu (2021): Single-index Thresholding in Quantile Regression, Journal of the American Statistical Association, DOI: [10.1080/01621459.2021.1915319](https://doi.org/10.1080/01621459.2021.1915319)

To link to this article: <https://doi.org/10.1080/01621459.2021.1915319>



View supplementary material [↗](#)



Published online: 01 Jun 2021.



Submit your article to this journal [↗](#)



Article views: 381



View related articles [↗](#)



View Crossmark data [↗](#)



# Single-index Thresholding in Quantile Regression

Yingying Zhang<sup>a</sup>, Huixia Judy Wang<sup>b</sup>, and Zhongyi Zhu<sup>c</sup> 

<sup>a</sup>Academy of Statistics and Interdisciplinary Sciences, KLATASDS-MOE, East China Normal University, Shanghai, China; <sup>b</sup>Department of Statistics, The George Washington University, Washington, DC; <sup>c</sup>Department of Statistics, Fudan University, Shanghai, China

## ABSTRACT

Threshold regression models are useful for identifying subgroups with heterogeneous parameters. The conventional threshold regression models split the sample based on a single and observed threshold variable, which enforces the threshold point to be equal for all subgroups of the population. In this article, we consider a more flexible single-index threshold model in the quantile regression setup, in which the sample is split based on a linear combination of predictors. We propose a new estimator by smoothing the indicator function in thresholding, which enables Gaussian approximation for statistical inference and allows characterizing the limiting distribution when the quantile process is interested. We further construct a mixed-bootstrap inference method with faster computation and a procedure for testing the constancy of the threshold parameters across quantiles. Finally, we demonstrate the value of the proposed methods via simulation studies, as well as through the application to an executive compensation data.

## ARTICLE HISTORY

Received December 2019  
Accepted March 2021

## KEYWORDS

Heterogeneity;  
Mixed-bootstrap; Quantile  
process; Smoothed  
estimator; Subgroup;  
Threshold regression

## 1. Introduction

Threshold regression is useful for identifying subgroups with heterogeneous effects determined by some threshold variables. Heterogeneous effects are commonly seen in many applications from economics, biomedicine and marketing. For example, research showed that countries may have different growth patterns according to large and small initial endowment or exports (Hansen 2000; Durlauf, Johnson, and Temple 2005; Christopoulos, McAdam, and Tzavalis 2021). The effect of body mass index on blood pressure was found to be different for subgroups with high and low BMI (Kaufman et al. 1997; Kerry et al. 2005; Tesfaye et al. 2007; Zhang, Wang, and Zhu 2014). In personalized medicine, the efficacy of treatment was shown to vary according to patients' various characteristics such as demographic information and pretreatment covariates (Rothwell 1995; Wang et al. 2007; VanderWeele et al. 2019). The estimation of heterogeneous effects and the identification of subgroups are important in these applications.

This article focuses on threshold quantile regression, which was first introduced by Cai and Stander (2008), Cai (2010), and Galvao, Montes-Rojas, and Olmo (2011) in the context of autoregression, and further studied by Lee, Seo and Shin (2011), Li et al. (2011), Yu (2013), and Su and Xu (2019) in other settings. Different from the mean regression, threshold quantile regression allows researchers to study the effect heterogeneity in different tails of the outcome distribution, which may be of more interest in applications such as medical cost, birth weight, income, and so on.

One limitation in the existing threshold quantile regression models is that the threshold splits the sample based on a single

and observed variable. Numerous studies have shown that subgroups may be characterized by several variables together. For example, Yu and Fan (2020) argued that the effect of experience on salary tends to vary with both education and gender. Furthermore, Shen and He (2015) and Fan, Song and Lu (2017) showed that a risk score, defined to be some function of multiple predictors, is helpful for identifying the subgroup of AIDS patients who benefit more from the treatment. Jiang et al. (2014) demonstrated that multiple SNPs (single-nucleotide polymorphism), rather than any individual SNP, tend to alter the risk for developing a particular disease. The importance of using numerous covariates to identify optimal subgroup for treatment was also discussed in He, Lin, and Tu (2018), and VanderWeele et al. (2019). These applications suggest that it would be more desirable to make use of multiple covariates to determine the subgroup and heterogeneous effects in applications.

Motivated by the above insights, we consider the following single-index threshold quantile regression model:

$$Q_{y_i}(\tau | z_i, \tilde{z}_i, x_i) = z_i^T \beta(\tau) + \tilde{z}_i^T \delta(\tau) I\{x_i^T \gamma(\tau) > 0\}, i = 1, \dots, n, \quad (1)$$

where  $y_i$  is the response from the  $i$ th subject,  $Q_{y_i}(\tau | \cdot)$  denotes the  $\tau$ th conditional quantile of the response given covariates,  $\tau \in (0, 1)$  is the quantile level of interest,  $z_i$  is the vector of design variables,  $\tilde{z}_i$  is a subset of  $z_i$ , and  $x_i$  includes the threshold variables. The parameters  $\beta(\tau)$  and  $\delta(\tau)$  are regression coefficients:  $\beta(\tau)$  measures the baseline effect of  $z_i$  on the  $\tau$ th quantile of  $y_i$ , and  $\delta(\tau)$  are the threshold effect, depicting the heterogeneous effect of  $\tilde{z}_i$  in the subgroup. The parameters  $\gamma(\tau)$  are threshold coefficients, which determine the single-index

threshold, an unknown linear combination of variables. In this article, we focus on the fixed-threshold-effect framework by assuming  $\delta(\tau) \neq 0$ .

The threshold parameters  $\gamma(\tau)$  play a central role for determining the single index and therefore the subgroup with heterogeneous effects. Inference on  $\gamma(\tau)$  could help to identify important threshold variables and determine the characteristics of the subgroup. However, the estimation and uncertainty quantification for  $\gamma(\tau)$  are complicated by the indicator function involved in Model (1). Existing work on threshold regression under the fixed-threshold-effect framework showed that direct estimators of threshold parameters have nonstandard limiting distributions, which are too complicated to be practically useful for statistical inference; see Chan (1993), Bai (1995), Yu (2013), and Yu and Fan (2020), to name a few. Some other researchers developed inference methods based on the asymptotic theory under the shrinking-effect framework, but the methods are limited to the cases with small threshold effects; see Hansen (2000), Caner (2002), Kuan, Michalopoulos and Xiao (2017), and Su and Xu (2019). The problem is even technically more challenging for Model (1) due to the nondifferentiability of the loss function involved in quantile regression (Koenker 2005) and the additional nonconstant covariates in thresholding.

To overcome the challenges, we propose a smoothed estimator by smoothing the indicator function in the thresholding. We show that the resulting estimators of both regression and threshold coefficients are asymptotically normal and they are asymptotically independent of each other. In addition, we show that the smoothed estimator of  $\gamma(\tau)$  has the convergence rate of  $\sqrt{h_n/n}$ , where  $h_n \rightarrow 0$  is the bandwidth involved in the smoothing of the indicator function. While the unsmoothed estimator of  $\gamma(\tau)$  could have a higher convergence rate, the proposed smoothing leads to three big advantages. First, the asymptotic normality of the threshold coefficient estimator facilitates the statistical inference based on the direct variance estimation. Second, the asymptotic normality theory also allows us to develop a computationally efficient mixed-bootstrap algorithm, which is convenient and less sensitive to the choice of bandwidth than the direct variance estimation. Third, under this framework, we are able to establish the asymptotic properties for the quantile process, which are useful for characterizing and comparing subgroups across quantiles. Particularly, we develop new procedures for testing the constancy of  $\gamma(\tau)$  across  $\tau$  in a region of quantiles. Such tests can help researchers to identify commonalities across quantiles, which can be subsequently used to improve the efficiency of subgroup analysis.

To our best knowledge, this article is the first that formally studies the estimation and inference for single-index threshold quantile regression models. Among the limited literature, Seo and Linton (2007) and Yu and Fan (2020) are the two that also studied single-index threshold but in the mean regression setup. Yu and Fan (2020) considered the direct estimator without smoothing, and proposed a nonparametric posterior interval method for inference on the threshold parameters. Seo and Linton (2007) considered smoothed estimation and suggested Wald-type, likelihood-type and bootstrap methods for inference. However, these inference methods either do not

apply to models with heteroscedastic errors, for which quantile regression has more appeals, or are computationally intensive and even more so for quantile regression. In this article, we develop new inference procedures to address the unique challenges rising from quantile regression, which require different techniques both computationally and theoretically.

The rest of the article is organized as follows. In the next section, we discuss the identification problem, present the proposed smoothed estimator and the profiled estimation procedure, and establish the large sample properties of the estimator. In Section 3, we propose a Wald-type method and a mixed-bootstrap method for inference. Section 4 is devoted to testing the constancy of threshold parameters across quantiles. Simulation studies are reported in Section 5, followed by an application to an executive compensation data in Section 6. Some concluding remarks are provided in Section 7. Technical proofs are relegated to the Appendix and the online supplement.

## 2. Smoothed Estimator and Asymptotic Properties

We will first address the identifiability issue of Model (1). Note that the same model continues to hold if  $\gamma(\tau)$  is divided by some positive constant. Although the normalization constraint  $\|\gamma(\tau)\|_2 = 1$  is often used in the literature to ensure the identifiability, we consider a different method for easier interpretation. The identification of  $\gamma(\tau)$  requires that there exists a variable  $x_1$  whose coefficient is nonzero, and that the probability distribution of  $x_1$  conditional on  $\mathbf{x}_2 = \mathbf{x}/x_1$  is absolutely continuous with respect to Lebesgue measure (Horowitz 1992). The notation  $\mathbf{x}/x_1$  denotes variables in  $\mathbf{x}$  excluding  $x_1$ . Let the first element in  $\mathbf{x}_2$  be the intercept. Model (1) can be rewritten as

$$Q_{\gamma_i}(\tau | \mathbf{z}_i, \tilde{\mathbf{z}}_i, \mathbf{x}_i) = \mathbf{z}_i^T \boldsymbol{\beta}(\tau) + \tilde{\mathbf{z}}_i^T \boldsymbol{\delta}(\tau) I\{x_{1i} + \mathbf{x}_{2i}^T \boldsymbol{\psi}(\tau) > 0\}, \quad (2)$$

where  $\boldsymbol{\psi}(\tau) = \boldsymbol{\gamma}_{-1}(\tau)/\gamma_1(\tau)$  with  $\gamma_1(\tau)$  and  $\boldsymbol{\gamma}_{-1}(\tau)$  corresponding to  $x_1$  and  $\mathbf{x}_2$ , respectively. If  $\gamma_1(\tau)$  is positive, the normalization does not alter  $\boldsymbol{\beta}(\tau)$  and  $\boldsymbol{\delta}(\tau)$ . If  $\gamma_1(\tau)$  is negative,  $\boldsymbol{\beta}(\tau)$  and  $\boldsymbol{\delta}(\tau)$  can be redefined to make Model (2) consistent with Model (1).

*Remark 1.* The identifiability constraint in model (2) requires the specification of one threshold variable  $x_1$  whose coefficient is nonzero. In practice, this variable can be determined by domain knowledge, for example, in the executive compensation study considered in Section 6. When the prior knowledge is not available, we can conduct a preliminary analysis to select this variable as follows. For a candidate  $x_1$ , we can test the hypothesis  $H_0 : \delta(\tau) = 0$  in the model  $Q_{\gamma}(\tau | \mathbf{w}) = \mathbf{z}^T \boldsymbol{\beta}(\tau) + \tilde{\mathbf{z}}^T \boldsymbol{\delta}(\tau) I\{x_1 + \mathbf{x}_2^T \boldsymbol{\psi}(\tau) > 0\}$ . The rejection of  $H_0$  would suggest that there exists a subgroup whose structure is influenced by  $x_1$ , so that  $x_1$  should be included in the threshold function, and not otherwise. We can test the hypothesis by extending the score-type specification testing method in Zhang, Wang, and Zhu (2014) or Yu (2013), which were designed for threshold regression with a single threshold variable. The critical values can be calculated by using similar resampling methods as in Fan, Song and Lu (2017) and Yu (2013). The score-type test is constructed under the null hypothesis and thus is computational

more convenient than the likelihood-ratio-type test in Lee, Seo and Shin (2011) and the Wald-type test in Galvao et al. (2014).

## 2.1. The Smoothed Estimator

For notational convenience, thereafter we let  $\theta(\tau) = (\beta(\tau)^T, \delta(\tau)^T)^T$  denote the regression coefficients, and let  $\eta(\tau) = (\theta(\tau)^T, \psi(\tau)^T)^T$  be the collection of all unknown parameters. Note that we can view the quantile function in Model (2) as a nonlinear function of  $\eta(\tau)$ . Therefore, a natural and direct estimator of  $\eta(\tau)$  can be obtained as follows:

$$\check{\eta}(\tau) = \underset{\eta=(\beta^T, \delta^T, \psi^T)^T}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau}\{y_i - \mathbf{z}_i^T \beta - \tilde{\mathbf{z}}_i^T \delta I(x_{1i} + \mathbf{x}_{2i}^T \psi > 0)\},$$

where  $\rho_{\tau}(u) = \{\tau - I(u < 0)\}u$  is the quantile “check” loss function (Koenker 2005).

Under the fixed-threshold-effect framework of threshold regression, it was known that the direct estimator of the threshold parameter often has a higher convergence rate than that of the regression parameter, but the former has a nonstandard limiting distribution that is too complicated to be useful for statistical inference. This challenge was observed in both least-squares estimation (Chan 1993; Yu and Fan 2020) and quantile regression (Bai 1995; Yu 2013). To overcome this challenge, some researchers established asymptotic distributions for the direct estimator under the shrinking-threshold-effect framework, which assumes that the threshold effect decays to zero as  $n$  increases; see for instance Hansen (2000), Caner (2002), and Kuan, Michalopoulos and Xiao (2017). While these asymptotic distributions can be approximated through simulation, they may be limited to the small effect case and could lead to conservative confidence intervals when the threshold effect size is fixed. In a recent article, Su and Xu (2019) thoroughly studied the conventional threshold quantile regression under the shrinking-threshold-effect framework, and proposed a likelihood-ratio-based test procedure. Unfortunately, the method cannot be applied to the single-index threshold quantile regression model, since the limiting distribution of  $\check{\psi}(\tau)$  will no longer be a two-sided Brownian motion, but rather become a two-sided Brownian field due to the additional nonconstant covariates in thresholding; see Yu and Fan (2020) for related discussions in mean regression. Consequently, it will be difficult to simulate the asymptotic distribution of the likelihood-ratio-type statistics.

For convenient statistical inference, we propose a smoothed estimator obtained by smoothing the indicator function in thresholding. Let  $G(\cdot)$  be a bounded continuous function satisfying that  $\lim_{s \rightarrow -\infty} G(s) = 0$  and  $\lim_{s \rightarrow +\infty} G(s) = 1$ . Note that  $G$  is analogous to a cumulative distribution function rather than a density function traditionally used in kernel estimation. The proposed smoothed estimator of  $\eta(\tau)$  is defined as

$$\hat{\eta}(\tau) = \underset{\eta(\tau)}{\operatorname{argmin}} S_n\{\eta(\tau); h_n\}, \quad (3)$$

with

$$S_n(\eta; h_n) = \frac{1}{n} \sum_{i=1}^n \rho_{\tau} \left\{ y_i - \mathbf{z}_i^T \beta - \tilde{\mathbf{z}}_i^T \delta G \left( \frac{x_{1i} + \mathbf{x}_{2i}^T \psi}{h_n} \right) \right\}$$

where  $\eta = (\beta^T, \delta^T, \psi^T)^T$ , and  $h_n \rightarrow 0$  is a positive sequence of bandwidth parameters.

The idea of smoothing was also used in various contexts to ease the computational or theoretical challenges. Horowitz (1992) and Seo and Linton (2007) proposed to smooth the indicator link and threshold function in binary regression and the threshold mean regression to obtain asymptotic normality. Brown and Wang (2005), Fu, Wang and Bai (2010), and Pang, Lu and Wang (2012) considered induced smoothing to avoid direct estimation of the error density involved in the asymptotic covariance matrices for rank and quantile regression. The quantile score function was smoothed in Whang (2006) and Wang and Zhu (2011) to achieve higher-order accuracy of inference, and in Wang, Stefanski and Zhu (2012) to provide appropriate correction for covariate measurement errors. Our setup involves two different sources of unsmoothness, one in the quantile loss function and the other in the thresholding. We propose to only smooth the indicator function involved in the thresholding to obtain asymptotic normality to facilitate statistical inference, so the smoothing has the similar purpose as in Horowitz (1992) and Seo and Linton (2007). The nondifferentiable quantile loss function, on the other hand, brings more challenges to both computation and theory. We address the computational challenge below, and discuss the challenges on statistical inference in Section 3.

The minimization in Equation (3) is challenging because the objective function is not differentiable everywhere and it is not convex in  $\psi(\tau)$ . However, for every given  $\psi(\tau)$ , minimizing  $S_n(\eta; h_n)$  is equivalent to solving a linear quantile regression problem. Based on this observation, we consider the following profiled estimation method. First, for a given  $\psi(\tau)$ , define

$$(\tilde{\beta}\{\psi(\tau)\}, \tilde{\delta}\{\psi(\tau)\}) = \underset{\beta(\tau), \delta(\tau)}{\operatorname{argmin}} S_n\{\eta(\tau); h_n\}.$$

Then we can estimate  $\psi(\tau)$  by

$$\hat{\psi}(\tau) = \underset{\psi(\tau)}{\operatorname{argmin}} S_n\{\tilde{\beta}(\psi(\tau)), \tilde{\delta}(\psi(\tau)), \psi(\tau); h_n\}, \quad (4)$$

The profiled estimator for the regression coefficients is thus defined as  $\hat{\theta}(\tau) = (\hat{\beta}(\tau)^T, \hat{\delta}(\tau)^T)^T$ , where  $\hat{\beta}(\tau) = \tilde{\beta}\{\hat{\psi}(\tau)\}$  and  $\hat{\delta}(\tau) = \tilde{\delta}\{\hat{\psi}(\tau)\}$ . For practical implementation, we recommend to use the Nelder–Mead algorithm (in the R function “optim”) when good initialization is available, and otherwise use the genetic algorithm for optimizing nonsmooth objective functions as used in Zhang et al. (2012) and Wang et al. (2018).

## 2.2. Asymptotics for a Single Quantile

For convenience, we define  $e_i(\tau) = y_i - \mathbf{z}_i^T \beta_0(\tau) - \tilde{\mathbf{z}}_i^T \delta_0(\tau) I(x_{1i} + \mathbf{x}_{2i}^T \psi_0(\tau) > 0)$  and let  $\mathbf{Z}_i(\tau) = [\mathbf{z}_i^T, \{\tilde{\mathbf{z}}_i^T I(x_{1i} + \mathbf{x}_{2i}^T \psi_0(\tau) > 0)\}^T]^T$  denote the design vector when threshold parameters are known. Let  $\mathbf{w}_i$  denote all the variables including  $\mathbf{z}_i, \mathbf{x}_i$ , and  $f_{e(\tau)}(\cdot|\mathbf{w})$  denote the conditional density function of  $e(\tau)$ . We begin by imposing a set of regularity conditions to obtain the asymptotic consistency.

**Assumption 1.** (a)  $\{(y_i, \mathbf{w}_i)\}_{i=1}^n$  is an independent and identically distributed (iid) sequence.



- (b) The true coefficients  $\theta_0(\tau) = (\beta_0(\tau)^T, \delta_0(\tau)^T)^T$  and  $\psi_0(\tau)$  are in the interior of compact subspaces  $\Theta$  and  $\Psi$ , respectively. Moreover,  $\delta_0(\tau) \neq \mathbf{0}$ .
- (c)  $E\|\mathbf{w}\|^\gamma < \infty$  for some  $\gamma > 1$ .
- (d) The probability distribution of  $x_1$  conditional on  $\mathbf{x}_2$  has everywhere positive density with respect to Lebesgue measure for almost every  $\mathbf{x}_2$ .
- (e) The minimum eigenvalue of  $E\{f_{e(\tau)}(0|\mathbf{w})\mathbf{Z}\mathbf{Z}^T|\mathbf{x}\}$  is bounded away from zero uniformly over  $\mathbf{x}$ .

Our article is under the fixed-threshold-effect framework indicated by Assumption 1(b), that is,  $\delta(\tau) \neq \mathbf{0}$ . Assumptions 1(c) and 1(d) are needed to establish the asymptotic equivalence between the smoothed and unsmoothed objective functions as  $h_n \rightarrow 0$ . Assumption 1(e) ensures the identification of regression coefficients. The following theorem presents the consistency of the smoothed estimator.

**Theorem 1.** Suppose that Model (2) at  $\tau \in (0, 1)$  and conditions in Assumption 1 hold. If  $h_n \rightarrow 0$ , then  $\hat{\theta}(\tau) \rightarrow \theta_0(\tau)$  and  $\hat{\psi}(\tau) \rightarrow \psi_0(\tau)$  almost surely as  $n \rightarrow \infty$ .

Given the established consistency, the next step is to derive the asymptotic distribution. We first make some linear transformation of variables. Let  $q_i(\tau) = x_{1i} + \mathbf{x}_{2i}^T \psi_0(\tau)$  and  $\mathbf{w}_2 = \mathbf{w}/x_1$ . Since there is a one-to-one relation between  $(q(\tau), \mathbf{w}_2)$  and  $\mathbf{w}$ , there must exist  $(\delta_{10}(\tau), \delta_{20}(\tau)^T)^T$  such that  $\tilde{\mathbf{z}}^T \delta_0(\tau) = q(\tau)\delta_{10}(\tau) + \mathbf{w}_2^T \delta_{20}(\tau)$ . Denote  $f_{q(\tau)|\mathbf{w}_2}(q|\mathbf{w}_2)$  as the density of  $q(\tau)$  conditional on  $\mathbf{w}_2$  and  $f^{(i)}(x) = \partial^i f(x)/\partial x^i$  where the derivative exists. We are now ready to define the following quantities involved in the asymptotic covariances of  $\hat{\psi}(\tau)$  and  $\hat{\theta}(\tau)$ :

$$\begin{aligned} V^\psi(\tau) &= a\tau(1-\tau)E\left[\mathbf{w}_2^T \delta_{20}(\tau)^2 \mathbf{x}_2 \mathbf{x}_2^T f_{q(\tau)|\mathbf{w}_2}(0|\mathbf{w}_2)\right], \\ Q^\psi(\tau) &= G'(\tau)E\left[f_{e(\tau)}(0|\mathbf{w}_2) \left\{\mathbf{w}_2^T \delta_{20}(\tau)\right\}^2 \mathbf{x}_2 \mathbf{x}_2^T f_{q(\tau)|\mathbf{w}_2}(0|\mathbf{w}_2)\right], \\ V^\theta(\tau) &= \tau(1-\tau) \times \begin{pmatrix} E[\mathbf{z}\mathbf{z}^T] & E[\mathbf{z}\mathbf{z}^T I\{q(\tau) > 0\}] \\ E[\tilde{\mathbf{z}}\mathbf{z}^T I\{q(\tau) > 0\}] & E[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T I\{q(\tau) > 0\}] \end{pmatrix}, \\ Q^\theta(\tau) &= \begin{pmatrix} E[f_{e(\tau)}(0|\mathbf{w}_2)\mathbf{z}\mathbf{z}^T] & E[f_{e(\tau)}(0|\mathbf{w}_2)\mathbf{z}\mathbf{z}^T I\{q(\tau) > 0\}] \\ E[f_{e(\tau)}(0|\mathbf{w}_2)\tilde{\mathbf{z}}\mathbf{z}^T I\{q(\tau) > 0\}] & E[f_{e(\tau)}(0|\mathbf{w}_2)\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T I\{q(\tau) > 0\}] \end{pmatrix}, \end{aligned}$$

where  $a = \int G'(s)^2 ds$ . To establish the asymptotic normality, we make the following additional assumptions. We let  $M$  denote a constant whose value may vary in different contexts.

**Assumption 2.** (a)  $\sup_{i \geq 1} \|\mathbf{w}_i\| \leq M$  (a.s.).

- (b)  $f_{q(\tau)|\mathbf{w}_2}^{(k)}(q|\mathbf{w}_2)$  is a continuous function and  $|f_{q(\tau)|\mathbf{w}_2}^{(k)}(q|\mathbf{w}_2)| < M$  uniformly over  $(q, \mathbf{w}_2)$  for each integer  $0 \leq k \leq k'$  ( $k'$  will be specified later).
- (c)  $e(\tau)$  is independent with  $x_1$  and  $|f_{e(\tau)}^{(1)}(u|\mathbf{w}_2)| < M$  uniformly over  $(u, \mathbf{w}_2)$ .
- (d) The matrices  $V^\psi(\tau)$ ,  $Q^\psi(\tau)$ ,  $V^\theta(\tau)$ ,  $Q^\theta(\tau)$  are positive definite.
- (e)  $G$  is twice differentiable and  $G(s) + G(-s) = 1$ ;  $G'$  is symmetric around zero;  $|G'(s)|$  and  $|G''(s)|$  are uniformly

bounded over  $s$ ;  $\int |G'(s)|ds < \infty$  and  $\int |G''(s)|ds < \infty$ . In addition,  $\int s\{G(s) - I(s > 0)\}ds < \infty$ .

- (f) For an integer  $k' \geq 1$  and each integer  $k(1 \leq k \leq k')$ ,  $\int s^{k-1}\{G(s) - I(s > 0)\}G'(s) = 0$  and  $\int s^{k'}\{G(s) - I(s > 0)\}G'(s) \neq 0$ .
- (g)  $nh_n^3 \rightarrow 0$  if  $k' = 1$  and  $nh_n^4 \rightarrow 0$  if  $k' > 1$ . Moreover,  $nh_n^2 \rightarrow \infty$ .

The boundedness condition in 2(a) is assumed for theoretical convenience, but it can be relaxed at the expense of lengthier proofs. Assumption 2(b) imposes restrictions on the conditional density of thresholding. In Assumption 2(c),  $e(\tau)$  is assumed to be independent with  $x_1$  so that  $f_{e(\tau)|\mathbf{w}}(u|\mathbf{w}) = f_{e(\tau)|\mathbf{w}_2}(u|\mathbf{w}_2)$  to facilitate the proof. Assumptions 2(e) and 2(f) impose conditions on the smoothing function. One candidate kernel satisfying the conditions is  $G(x) = \Phi(x) + x\phi(x)$ , where  $\Phi$  and  $\phi$  are distribution and density functions of standard normal.

**Theorem 2.** Suppose that Model (2) at  $\tau \in (0, 1)$ , and Assumptions 1 and 2 hold. Then we have:

$$\begin{aligned} \sqrt{n} \left\{ \hat{\theta}(\tau) - \theta_0(\tau) \right\} &\xrightarrow{d} N(0, Q^\theta(\tau)^{-1} V^\theta(\tau) Q^\theta(\tau)^{-1}), \\ \sqrt{\frac{n}{h_n}} \left\{ \hat{\psi}(\tau) - \psi_0(\tau) \right\} &\xrightarrow{d} N(0, Q^\psi(\tau)^{-1} V^\psi(\tau) Q^\psi(\tau)^{-1}), \end{aligned}$$

and the two are asymptotically independent.

**Remark 2.** As also observed in the threshold regression literature (Seo and Linton 2007; Yu 2013; Su and Xu 2019), the regression coefficient estimator  $\hat{\theta}(\tau)$  is not affected asymptotically by the threshold parameter estimator  $\hat{\psi}(\tau)$ , and the latter converges at a faster rate. Consequently, statistical inference for regression coefficients  $\theta_0(\tau)$  can be conducted as if  $\psi_0(\tau)$  were known, and existing inference methods such as the Wald and rank score methods for quantile regression (Koenker 2005) can be directly applied. The asymptotic independence is not surprising as it may appear. One intuitive explanation is as follows. As also argued in Yu (2012) and Yu (2015) for mean threshold regression, only local information around the neighborhood of  $\psi_0(\tau)$  is relevant to the estimation of the threshold parameter  $\psi_0(\tau)$ . On the other hand, the estimation of the regression coefficients  $\theta(\tau)$  depends on global information or information related to moments of the data. Theoretically the asymptotic independence between the estimator for regression coefficients and that for threshold parameters depends on the bandwidth parameter  $h_n$  used to smooth the threshold indicator function. The proof of Lemma 1 shows that the asymptotic covariance of the two estimators converges to zero at a rate of  $h_n^{1/2}$ . That is, a larger bandwidth means that the information in a wider neighborhood of the boundary is used and this consequently leads to larger dependence between the two estimators. Similar asymptotic independence property was also observed in the boundary literature, for instance in Chernozhukov and Hong (2004), where the boundary is defined by the jump in the conditional density of the regression error, analogous to the jump in the conditional quantile function in our framework.

**Remark 3.** The convergence rate of  $\hat{\psi}(\tau)$  is  $\sqrt{h_n/n}$ , which means that a faster convergence rate of the bandwidth will

escalate the convergence of  $\hat{\psi}(\tau)$ . By Yu and Fan (2020) and Su and Xu (2019), we would conjecture that the convergence rate of the unsmoothed estimator of threshold parameters is  $n^{-1}$  under the fixed-threshold-effect framework, and  $n^{2a-1}$  under the shrinking-threshold-effect framework with  $\|\delta_0(\tau)\| = O(n^{-a})$ . Therefore, we obtain the asymptotic normality at the expense of the convergence rate. Note that in an attempt to solve these theoretical challenges, Su and Xu (2019) has also sacrificed the convergence rate. The convergence rate of our proposed smoothed estimator satisfies  $n^{-3/4} \ll \sqrt{h_n/n} \ll n^{-2/3}$  for  $k' = 1$  and  $n^{-3/4} \ll \sqrt{h_n/n} \ll n^{-5/8}$  for  $k' > 1$ . Therefore, the convergence rate is not faster than  $n^{-3/4}$ , but is faster than  $n^{-5/8}$ .

### 2.3. Asymptotics for the Quantile Process

In practice, it is valuable to study the quantile process to obtain a complete analysis of the stochastic relationship between variables. The theoretical analysis can also help us develop testing tools to test whether the threshold parameters  $\psi(\tau)$  vary with quantiles or not. If  $\psi(\tau)$  is found to be constant across  $\tau$  (or across a region of quantiles), composite quantile estimation can be used to provide a more efficient estimator for  $\psi(\tau)$ . As a result of smoothing, when  $\psi(\tau)$  is treated as a stochastic process, the limiting process is Gaussian and can be characterized by its mean and covariance kernel.

**Theorem 3.** Suppose that Model (2), and Assumptions 1 and 2 hold for all  $\tau \in \mathcal{T}$ , where  $\mathcal{T}$  is a compact set in  $(0, 1)$ . In addition, assume that the parameter space in Assumption 1 and constant  $M$  in Assumption 2 are independent of  $\tau$ ,  $\|\eta_0(\tau) - \eta_0(\tau')\| < C|\tau - \tau'|^a$  and  $|G'(s) - G'(t)| < C|s - t|^a$  for some  $C > 0, a > 0$ . Then we have

1.  $\sqrt{n} \{\hat{\theta}(\cdot) - \theta_0(\cdot)\} \Rightarrow G^\theta(\cdot)$ , where  $G^\theta(\cdot)$  is a mean zero Gaussian process with the covariance function  $E\{G^\theta(\tau)G^\theta(\tau')'\} = Q^\theta(\tau)^{-1}V^\theta(\tau, \tau')\{Q^\theta(\tau')^{-1}\}'$ ,

$$V^\theta(\tau, \tau') = \{\min(\tau, \tau') - \tau\tau'\} \begin{pmatrix} Ezz^T & Ezz^T I(q(\tau') > 0) \\ Ezz^T I(q(\tau) > 0) & Ezz^T I(q(\tau) > 0, q(\tau') > 0) \end{pmatrix},$$

2.  $\sqrt{nh_n^{-1}} \{\hat{\psi}(\cdot) - \psi_0(\cdot)\} \Rightarrow G^\psi(\cdot)$  where  $G^\psi(\cdot)$  is a mean zero Gaussian process with the covariance function  $E\{G^\psi(\tau)G^\psi(\tau')'\} = Q^\psi(\tau)^{-1}V^\psi(\tau, \tau')\{Q^\psi(\tau')^{-1}\}'$ , where when  $\psi_0(\tau) = \psi_0(\tau')$ ,

$$V^\psi(\tau, \tau') = a \{\min(\tau, \tau') - \tau\tau'\} \times E\left[\{w_2\hat{\delta}_{20}(\tau)\} \{w_2\hat{\delta}_{20}(\tau')\} x_2 x_2^T f_{q(\tau)|w_2}(0|w_2)\right],$$

and when  $\psi_0(\tau) \neq \psi_0(\tau')$ ,  $V^\psi(\tau, \tau') = 0$ ,

and the two processes are asymptotically independent. Here, the independence between two stochastic processes is defined by that all finite-dimensional marginals of the two processes are independent of each other.

### 3. Inference on Threshold Parameters

As mentioned in Remark 1, the inference on regression parameters can be done as if  $\psi_0(\tau)$  were known. Thus, we focus on the inference for threshold parameters. The threshold parameters  $\psi(\tau)$  play a central role for determining the single index and therefore the subgroup with heterogeneous effects. Inference on  $\psi(\tau)$  could help to identify important threshold variables and determine the characteristics of the subgroup. We propose two procedures, including a Wald-type method and a mixed-bootstrap method, together with a step for finite sample bias correction to improve finite sample performance.

#### 3.1. Wald-type Method

Based on the asymptotic normality in Theorem 2, a Wald-type method can be constructed for inference on  $\psi(\tau)$  by directly estimating the asymptotic covariance matrix, denoted as

$$\Lambda(\tau) = Q^\psi(\tau)^{-1}V^\psi(\tau)Q^\psi(\tau)^{-1}.$$

From the proof for Lemmas 1 and 4(c) in the supplement, we obtain the following asymptotic representations of  $V^\psi(\tau)$  and  $Q^\psi(\tau)$ :

$$\begin{aligned} V^\psi(\tau) &= h_n E \left[ \left\{ \tau - I(e^G(\tau) < 0) \right\}^2 \right. \\ &\quad \left. \left\{ \tilde{z}^T \delta_0(\tau) G' \left( \frac{q(\tau)}{h_n} \right) \right\}^2 \frac{x_2 x_2^T}{h_n h_n} \right] + o_p(1), \\ Q^\psi(\tau) &= E \left[ f_{e(\tau)}(\Delta(\tau)) \left\{ \tilde{z}^T \delta_0(\tau) G' \left( \frac{q(\tau)}{h_n} \right) \right\}^2 \frac{x_2 x_2^T}{h_n} \right] \\ &\quad - E \left[ \left\{ \tau - F_{e(\tau)}(\Delta(\tau)) \right\} \tilde{z}^T \delta_0(\tau) G' \left( \frac{q(\tau)}{h_n} \right) \frac{x_2 x_2^T}{h_n} \right] \\ &\quad + o_p(1), \end{aligned}$$

where  $e^G(\tau) = y - z^T \beta_0(\tau) - \tilde{z}^T \delta_0(\tau) G'(\frac{q(\tau)}{h_n})$  and  $\Delta(\tau) = \tilde{z}^T \delta_0(\tau) \left\{ G'(\frac{q(\tau)}{h_n}) - I(q(\tau) > 0) \right\}$ . Thus, we can consistently estimate both  $V^\psi(\tau)$  and  $Q^\psi(\tau)$  by their sample version:

$$\begin{aligned} \hat{V}^\psi(\tau) &= \frac{h_n}{n} \sum_{i=1}^n \left\{ \tau - I(\hat{e}_i^G(\tau) < 0) \right\}^2 \\ &\quad \left\{ \tilde{z}_i^T \hat{\delta}(\tau) G' \left( \frac{\hat{q}_i(\tau)}{h_n} \right) \right\}^2 \frac{x_{2i} x_{2i}^T}{h_n h_n}, \\ \hat{Q}^\psi(\tau) &= \frac{1}{2nl_n} \sum_{i=1}^n I(|\hat{e}_i(\tau)| \leq l_n) \left\{ \tilde{z}_i^T \hat{\delta}(\tau) G' \left( \frac{\hat{q}_i(\tau)}{h_n} \right) \right\}^2 \frac{x_{2i} x_{2i}^T}{h_n} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \left\{ \tau - \hat{F}_{e(\tau)}(\hat{\Delta}_i(\tau)|w) \right\} \\ &\quad \times \tilde{z}_i^T \hat{\delta}(\tau) G' \left( \frac{\hat{q}_i(\tau)}{h_n} \right) \frac{x_{2i} x_{2i}^T}{h_n}, \end{aligned}$$

where  $l_n$  is an appropriately chosen bandwidth,  $\hat{F}_{e(\tau)}(t|w) = n^{-1} \sum_{i=1}^n I(\hat{e}_i(\tau) < t)$ ,  $\hat{e}_i^G(\tau)$ ,  $\hat{e}_i(\tau)$  and  $\hat{\Delta}_i(\tau)$  are calculated by plugging in the parameter estimation. Following Powell (1986), when  $l_n \rightarrow 0$ , and  $nl_n^2 \rightarrow \infty$ , we can verify that  $\hat{\Lambda}(\tau) = \hat{Q}^\psi(\tau)^{-1} \hat{V}^\psi(\tau) \{\hat{Q}^\psi(\tau)\}^{-1}$  is consistent with  $\Lambda(\tau)$ .

The asymptotic covariance matrix depends on a number of quantities, including the conditional density of  $e(\tau)$ , the kernel function  $G(\cdot)$ , as well as the bandwidth  $h_n$ . Our numerical study and previous research on quantile regression showed that the Wald-type method could be unstable in finite samples and it is sensitive to the choice of bandwidth used in estimating the conditional density function; see Kocherginsky et al. (2005) and Wang and Zhu (2011). Based on the theoretical properties of the smoothed estimator, we propose a new bootstrap procedure as an alternative inference method in the next subsection.

### 3.2. Mixed-bootstrap Method

Thanks to the asymptotic normality of the smoothed estimator, we can construct an alternative and consistent inference approach based on the bootstrap, which avoids direct estimation of the asymptotic covariance matrix of  $\hat{\psi}(\tau)$ . However, the standard bootstrap method is computationally demanding, and even more so in our setup, as we need to perform profile estimation and solve quantile minimization multiple times for each resampling. To reduce the computational burden, we propose a new mixed-bootstrap procedure by using the theoretical properties of the smoothed estimator. Specifically, based on the Bahadur representation of the smoothed estimator, given in Equation (A.6) in the appendix, we first perturb the score functions to estimate the regression coefficients, which requires no optimization but only a summation calculation, and then update the threshold parameters. This procedure avoids optimizing the profiled objective function and thus is computationally efficient.

The detailed mixed-bootstrap procedure is as follows:

- Step 1: Obtain the resampled counterpart of  $\hat{\theta}(\tau)$  as follows:

$$\hat{\theta}^*(\tau) = \begin{pmatrix} \hat{\beta}(\tau) \\ \hat{\delta}(\tau) \end{pmatrix} + \frac{1}{n} \sum_{i=1}^n (g_i - 1) \hat{Q}^\theta(\tau)^{-1} \begin{pmatrix} \{\tau - I(\hat{e}_i^G(\tau) < 0)\} z_i \\ \{\tau - I(\hat{e}_i^G(\tau) < 0)\} \tilde{z}_i G(\frac{\hat{q}_i(\tau)}{h_n}) \end{pmatrix},$$

where  $\{g_i\}_{i=1}^n$  are iid nonnegative random weights with mean and variance both equal to one. Multiplying the score functions by  $(g_i - 1)$  guarantees that the mean of  $\hat{\theta}^*(\tau)$  conditional on the data is  $\hat{\theta}(\tau)$ . The consistent estimation of  $Q^\theta(\tau)$  can be obtained by

$$\hat{Q}^\theta(\tau) = \frac{1}{2nl_n} \sum_{i=1}^n I(|\hat{e}_i(\tau)| < l_n) \begin{pmatrix} z_i z_i^T & z_i \tilde{z}_i^T I(\hat{q}_i(\tau) > 0) \\ \tilde{z}_i z_i^T I(\hat{q}_i(\tau) > 0) & \tilde{z}_i \tilde{z}_i^T I(\hat{q}_i(\tau) > 0) \end{pmatrix},$$

where  $l_n \rightarrow 0$  and  $\sqrt{n}l_n \rightarrow \infty$ . The estimation  $\hat{Q}^\theta(\tau)$  can be obtained by the function “summary.rq” from the R package quantreg; see page 81 of Koenker (2005) for the detailed implementation.

- Step 2: Given  $\hat{\theta}^*(\tau) = (\hat{\beta}^*(\tau)^T, \hat{\delta}^*(\tau)^T)^T$ , calculate

$$\hat{\psi}^*(\tau) = \underset{\psi(\tau)}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n g_i \rho_\tau \left\{ y_i - z_i^T \hat{\beta}^*(\tau) - \tilde{z}_i^T \hat{\delta}^*(\tau) G\left(\frac{x_{1i} + x_{2i}^T \psi(\tau)}{h_n}\right) \right\}.$$

The  $\hat{\psi}^*(\tau)$  is obtained by minimizing the weighted quantile loss function with  $\hat{\theta}^*(\tau)$  plugged in. Since the optimization is only with respect to  $\psi(\tau)$ , it is simpler than that in the profiled estimation in Equation (4). Any existing optimization algorithms including those designed for nonlinear quantile regression in Koenker and Park (1996), Hunter and Lange (2000) can be used in Step 2.

- Step 3: Repeat Steps 1 and 2 for  $B$  times and calculate  $E^{(b)}(\tau) = \sqrt{nh_n^{-1}} \{\hat{\psi}^*(\tau) - \hat{\psi}(\tau)\}$  for  $b = 1, \dots, B$ . The sample covariance of  $\{E^{(b)}(\tau)\}_{b=1}^B$  can then be used to approximate the covariance of  $\sqrt{nh_n^{-1}} \{\hat{\psi}(\tau) - \psi_0(\tau)\}$ . Based on the asymptotic normality, we can construct the  $100(1 - \alpha)\%$  confidence interval (CI) for  $\psi_j(\tau)$ , the  $j$ th component of  $\psi(\tau)$ , by

$$\hat{\psi}_j(\tau) \pm \sqrt{\frac{h_n}{n}} \hat{\sigma}_j z_{1-\alpha/2},$$

where  $z_{1-\alpha/2}$  is the  $(1 - \alpha/2)$ th quantile of standard normal distribution and  $\hat{\sigma}_j$  is the bootstrap standard deviation of  $\sqrt{nh_n^{-1}} \hat{\psi}_j(\tau)$ .

The following theorem justifies that the asymptotic covariance of the mixed-bootstrap estimator is the same as that of the smoothed estimator. We need to note that the mixed-bootstrap procedure relies on the asymptotic normality and the Bahadur representation of the smoothed estimator, and thus cannot be applied to the unsmoothed estimator.

**Theorem 4.** Suppose that Model (2) at  $\tau \in (0, 1)$ , and Assumptions 1 and 2 hold. In addition,  $\{g_i\}_{i=1}^n$  are iid weights with mean and variance both equal to one and  $\sup_{i \geq 1} |g_i| \leq M$  almost surely. Then conditional on  $\{(y_i, w_i)\}_{i=1}^n$ , we have

$$\sqrt{\frac{n}{h_n}} \{\hat{\psi}^*(\tau) - \hat{\psi}(\tau)\} \xrightarrow{p} N(0, Q^\psi(\tau)^{-1} V^\psi(\tau) Q^\psi(\tau)^{-1}).$$

### 3.3. Bias-corrected Confidence Interval

Both the Wald-type and mixed-bootstrap methods rely on the asymptotic normality of the smoothed estimator  $\hat{\psi}(\tau)$ . As discussed in Remark 2, the smoothing facilitates the convenient inference at the expense of slower convergence rate and bias. A smaller bandwidth  $h_n$  leads to smaller bias but further deviation from the normality. Although the bias is asymptotically ignorable, in finite samples this may affect the coverage of confidence intervals. To further improve the finite sample performance, we propose a bias correction method for confidence intervals of  $\psi_0(\tau)$ .

By the Bahadur representation of the smoothed estimator given in Equation (A.7) in the appendix, the bias of  $\hat{\psi}(\tau) - \psi_0(\tau)$  is

$$d(\tau) = h_n Q^\psi(\eta_0(\tau); h_n)^{-1} E \{f^\psi(\mathbf{w}, \eta_0(\tau); h_n)\},$$

where  $f^\psi(\mathbf{w}, \eta_0(\tau); h_n) = [\tau - I\{e(\tau) \leq \Delta(\tau)\}] \tilde{z}^T \delta(\tau) G'(q(\tau)h_n^{-1})\mathbf{x}_2 h_n^{-1}$ . The matrix  $Q^\psi(\eta_0(\tau); h_n)^{-1}$ , defined explicitly in Equation (A.5) in the appendix and abbreviated as  $(Q^\psi)^{-1}$ , can be estimated by the relationship

$$\Lambda(\tau) = \frac{h_n^2}{n} (Q^\psi)^{-1} \text{Cov}\{f^\psi(\mathbf{w}, \eta_0(\tau); h_n)\} (Q^\psi)^{-1},$$

where  $\Lambda(\tau) = \text{var}\{\hat{\psi}(\tau) - \psi_0(\tau)\}$  can be estimated by the mixed-bootstrap procedure, which is more stable than the Wald-type method. Replacing the mean and covariance of  $f^\psi$  by their sample counterparts, we can obtain the estimated bias, denoted as  $\hat{d}(\tau)$ .

Then the bias-corrected  $100(1 - \alpha)\%$  confidence interval of  $\psi_j(\tau)$  can be constructed as follows:

$$\left[ \hat{\psi}_j(\tau) - \hat{d}_j^+(\tau) - \sqrt{\frac{h_n}{n}} \hat{\sigma}_{jz_{\alpha/2}}, \hat{\psi}_j(\tau) - \hat{d}_j^-(\tau) + \sqrt{\frac{h_n}{n}} \hat{\sigma}_{jz_{\alpha/2}} \right],$$

where  $\hat{d}_j^+(\tau) = \max\{\hat{d}_j(\tau), 0\}$ ,  $\hat{d}_j^-(\tau) = \min\{\hat{d}_j(\tau), 0\}$  and  $\hat{\sigma}_j$  is the estimated standard deviation of  $\hat{\psi}_j(\tau)$ , computed either by the Wald-type method or the mixed-bootstrap method. Here, we replace  $\hat{d}_j(\tau)$  by  $\hat{d}_j^+(\tau)$  and  $\hat{d}_j^-(\tau)$ , since when the bias is small, treating the estimated bias as the true value can cause the confidence interval to have substantial undercoverage. The idea here, which is also employed by Qu and Yoon (2015), is to allow for, but not to enforce, a bias adjustment. By the proof of Lemmas 1 and 4 in the supplement, we can show that  $\hat{d}_j(\tau)$  converges to zero at the rate of  $h_n^{k'+2}$  ( $k' \geq 1$  is defined in Assumption 2(f)), while the term associated with the standard deviation in the confidence interval converges to zero at the rate of  $\sqrt{\frac{h_n}{n}}$ . Therefore, when  $nh_n^{2k'+3} \rightarrow 0$ , the bias term diminishes faster than the standard deviation term. So the bias correction will not affect the asymptotic coverage of the confidence interval.

#### 4. Testing for the Constancy of $\psi(\tau)$ Across $\tau$

The Model (2) allows both regression coefficients and the subgroup structure to vary with the quantile level  $\tau$ . In many applications, however, it is not unusual that the subgroup structure, determined by the threshold parameters  $\psi(\tau)$ , varies little across quantiles in a certain region. It will then be interesting to test the constancy of  $\psi(\tau)$  across  $\tau$ . If the test suggests that  $\psi(\tau)$  is constant, this not only can simplify the interpretation of the model, but also can help construct more efficient estimation of the common parameters by combining information across quantiles, for instance, then by adopting the composite quantile estimation idea as in Zhang, Wang, and Zhu (2017) and Su and Xu (2019). Based on the asymptotic properties in Theorem 3, we propose methods for testing the constancy of  $\psi(\tau)$  across  $\tau$  for two cases when the common parameter  $\psi_0$  is known or left unspecified.

##### 4.1. Constancy of $\psi(\tau)$ —Known $\psi_0$

We first consider testing the following hypotheses:

$$\begin{aligned} H_0 : \psi(\tau) &= \psi_0 \quad \text{for all } \tau \in \mathcal{T}, \quad \text{v.s.} \\ H_1 : \exists \tau' \in \mathcal{T} \text{ s.t. } \psi(\tau') &\neq \psi_0, \end{aligned}$$

where the null value  $\psi_0$  is assumed to be known.

We focus on the process  $\mathbf{v}_n(\cdot) = \sqrt{nh_n^{-1}} \{\hat{\psi}(\cdot) - \psi_0\}$ , and construct the Kolmogorov–Smirnov (KS) type statistic as

$$S_n = \sup_{\tau \in \mathcal{T}} \|\mathbf{v}_n(\tau)\|_{\hat{\Lambda}(\tau)},$$

where  $\hat{\Lambda}(\tau)$  can be any consistent estimation of  $\text{var}\{\mathbf{v}_n(\tau)\}$  and  $\|\mathbf{v}_n(\tau)\|_{\hat{\Lambda}(\tau)} = \mathbf{v}_n(\tau)^T \hat{\Lambda}(\tau) \mathbf{v}_n(\tau)$ . The following theorem states the limiting distribution of the KS statistic under the null and the alternative hypotheses.

**Theorem 5.** Under the assumptions in Theorem 3, (i) if  $H_0$  is true, we have  $S_n \xrightarrow{d} \sup_{\tau \in \mathcal{T}} \|\mathbf{B}_p(\tau)\|$ , where  $\mathbf{B}_p(\cdot)$  is a standard  $p$ -dimensional Brownian bridge with  $E\{\mathbf{B}_p(\tau)\mathbf{B}_p(\tau')'\} = (\min(\tau, \tau') - \tau\tau')I_p$  and  $p$  is the dimension of  $\psi(\tau)$ ; (ii) if  $H_1$  is true, we have  $S_n \xrightarrow{d} \infty$ .

In our implementation, we also perform a bias correction and modify  $\mathbf{v}_n(\cdot)$  as  $\mathbf{v}_n(\cdot) = \sqrt{nh_n^{-1}} \{\hat{\psi}(\cdot) - \hat{d}(\cdot) - \psi_0\}$ . As before we calculate the test statistic and critical value by adapting the mixed-bootstrap method to avoid direct estimation of  $\Lambda(\tau)$ . Specifically, in each resampling, we calculate  $\mathbf{E}^{(b)}(\tau) = \sqrt{nh_n^{-1}} \{\hat{\psi}^*(\tau) - \hat{\psi}(\tau)\}$  for all  $\tau \in \mathcal{T}$  with the same set of random weights. Then  $\hat{\Lambda}(\tau)$  is estimated by the sample variance of  $\{\mathbf{E}^{(b)}(\tau)\}_{b=1}^B$  for all  $\tau \in \mathcal{T}$ . The critical value is then calculated by the sample percentile of bootstrap statistics  $S_n^{(b)} = \sup_{\tau \in \mathcal{T}} \|\mathbf{E}^{(b)}(\tau)\|_{\hat{\Lambda}(\tau)}$ ,  $b = 1, \dots, B$ . No bias correction is needed in the mixed-bootstrap statistic since the bias of  $\mathbf{E}^{(b)}(\tau)$  conditional on the data is zero by the construction. It is practical to use a grid of equally spaced quantiles  $\{\tau_1, \dots, \tau_m\} \in \mathcal{T}$  to approximate the entire process. Our numerical results show that the test is insensitive to the choice of  $m$  as long as the mesh of the quantile grid is not too large.

##### 4.2. Constancy of $\psi(\tau)$ —Unknown $\psi_0$

In general,  $\psi_0$  is unknown. In this case, we consider the hypotheses

$$\begin{aligned} H_0 : \text{there exists } \psi_0 \text{ s.t. } \psi(\tau) &= \psi_0 \text{ for all } \tau \in \mathcal{T} \text{ v.s.} \\ H_1 : \text{there is no } \psi_0 \text{ s.t. } \psi(\tau) &= \psi_0 \text{ for all } \tau \in \mathcal{T}. \end{aligned}$$

Note that in this setup,  $\psi(\tau)$  is a function of  $\tau$ , and  $\psi_0$  represents the unknown nuisance parameter; both have to be estimated. The unknown parameters always jeopardize the distribution-free character of classical tests, which is called the “Durbin problem”. This problem was first posed in Durbin (1973) for parametric empirical process. There are two ways to overcome the problem in the literature. One solution is to use the Khmaladze martingale transformation, exemplified by Koenker and



Xiao (2002). Another solution is to use a resampling procedure, like that used by Chernozhukov and Hansen (2006). The simulation studies in Chernozhukov and Hansen (2006) suggested that resampling leads to an accurate size and better power than Khmaladzatation.

We propose a different solution. As commented in Remark 2, the unsmoothed estimator for threshold parameters would converge at a faster rate than the smoothed estimator. Motivated by this, we could replace the unknown  $\psi_0$  in  $\tilde{\nu}_n(\cdot)$  by a consistent unsmoothed estimator  $\check{\psi}$ , and define

$$\tilde{\nu}_n(\cdot) = \sqrt{nh_n^{-1}}\{\hat{\psi}(\cdot) - \hat{d}(\cdot) - \check{\psi}\}, \tilde{S}_n = \sup_{\tau \in \mathcal{T}} \|\tilde{\nu}_n(\tau)\|_{\hat{\Lambda}(\tau)}.$$

The estimator  $\check{\psi}$  can be taken as the unsmoothed estimator obtained at any quantile level  $\tau \in \mathcal{T}$ . In practice, to achieve more estimation stability, we propose to estimate  $\psi_0$  by the unsmoothed composite quantile estimator obtained under  $H_0$ , defined as

$$\begin{aligned} \check{\psi} = \operatorname{argmin} \sum_{i=1}^n \sum_{j=1}^m \rho_{\tau_j} \{y_i - \mathbf{z}_i^T \boldsymbol{\beta}(\tau_j) \\ - \tilde{\mathbf{z}}_i^T \boldsymbol{\delta}(\tau_j) I(x_{1i} + \mathbf{x}_{2i}^T \boldsymbol{\psi} > 0)\}. \end{aligned} \quad (5)$$

Since  $\check{\psi}$  converges at a faster rate than the smoothed estimator  $\hat{\psi}(\cdot)$ , we can show that  $\tilde{\nu}_n(\cdot)$  and  $\tilde{S}_n$  converge to the same asymptotic distributions as  $\nu_n(\cdot) = \sqrt{nh_n^{-1}}(\hat{\psi}(\cdot) - \psi_0)$  and  $S_n = \sup_{\tau \in \mathcal{T}} \|\nu_n(\tau)\|_{\hat{\Lambda}(\tau)}$  under  $H_0$ . The critical value of  $\tilde{S}_n$  can then be calculated by following the same procedure as in Section 4.1.

## 5. Simulation Study

We consider the following data-generating process:

$$y = \mathbf{z}^T \boldsymbol{\beta} + \tilde{\mathbf{z}}^T \boldsymbol{\delta} I\{x_1 + \mathbf{x}_2^T \boldsymbol{\psi} + d\Phi^{-1}(U) > 0\} + \Phi^{-1}(U), \quad (6)$$

where  $U \sim \text{Uniform}(0, 1)$  and  $d \geq 0$ . When  $\tilde{\mathbf{z}}^T \boldsymbol{\delta} > 0$ , the generating function for  $y$  is strictly increasing in  $U$ . Thus, the  $\tau$ th conditional quantile of  $y$  is

$$Q_y(\tau|\mathbf{w}) = \mathbf{z}^T \boldsymbol{\beta} + \tilde{\mathbf{z}}^T \boldsymbol{\delta} I\{x_1 + \mathbf{x}_2^T \boldsymbol{\psi} + d\Phi^{-1}(\tau) > 0\} + \Phi^{-1}(\tau). \quad (7)$$

We consider the following cases:

Case 1:  $\mathbf{z} = (1, z_1, z_2)^T$ ,  $\tilde{\mathbf{z}} = (z_1, z_2)^T$  with  $z_1 \sim \exp(1)$  and  $z_2 \sim \text{Uniform}(0, 1)$ .  $x_1 \sim N(0, 1)$  and  $\mathbf{x}_2 = (1, x_2)^T$ , where  $x_2 \sim N(1, 1)$ .  $\boldsymbol{\beta} = (1, 1, 1)^T$ ,  $\boldsymbol{\delta} = (1, 1)^T$ ,  $\boldsymbol{\psi} = (-1, 1)^T$  and  $d = 0$ .

Case 2: the same with Case 1 except that  $d = 0.1$ .

Under Model (6), the threshold parameters are constant across quantiles when  $d = 0$ , while they are more complicated and quantile-specific when  $d > 0$ . For Case 2, the true conditional mean function is:

$$E(y|\mathbf{w}) = \mathbf{z}^T \boldsymbol{\beta} + \tilde{\mathbf{z}}^T \boldsymbol{\delta} \{1 - \Phi(10 - 10x_1 - 10x_2)\}.$$

Therefore, for Case 2, the covariates  $(z_1, z_2)$  have threshold effects on the  $\tau$ th quantile of  $Y$  but not on the mean. For such scenarios, the proposed single-index threshold quantile regression can capture the heterogeneous effects of covariates that may

be overlooked by mean regression. Throughout our numerical studies, we let  $G(x) = \Phi(x) + x\phi(x)$  be the kernel function used in the smoothed estimator. We consider four sample sizes  $n = 200, 500, 1000$ , and  $2000$ , and repeat the simulation 1000 times for each scenario.

### 5.1. Estimation With Different Bandwidths

The proposed smoothed estimator depends on the bandwidth  $h_n$ . Based on the theoretical conditions on  $h_n$  in Assumptions 2(g), We consider the following rule of thumb  $h_n = cs \log(n)/\sqrt{n}$ , where  $c$  is a constant and  $s$  is the standard deviation of  $\mathbf{x}_2 \psi_0(\tau) + x_1$ . In practice,  $s$  can be estimated by the sample standard deviation of  $\mathbf{x}_2 \check{\psi}(\tau) + x_1$  with  $\check{\psi}(\tau)$  as the unsmoothed estimator. We assess the sensitivity of  $\hat{\psi}(\tau)$  against the bandwidth by varying  $c$ .

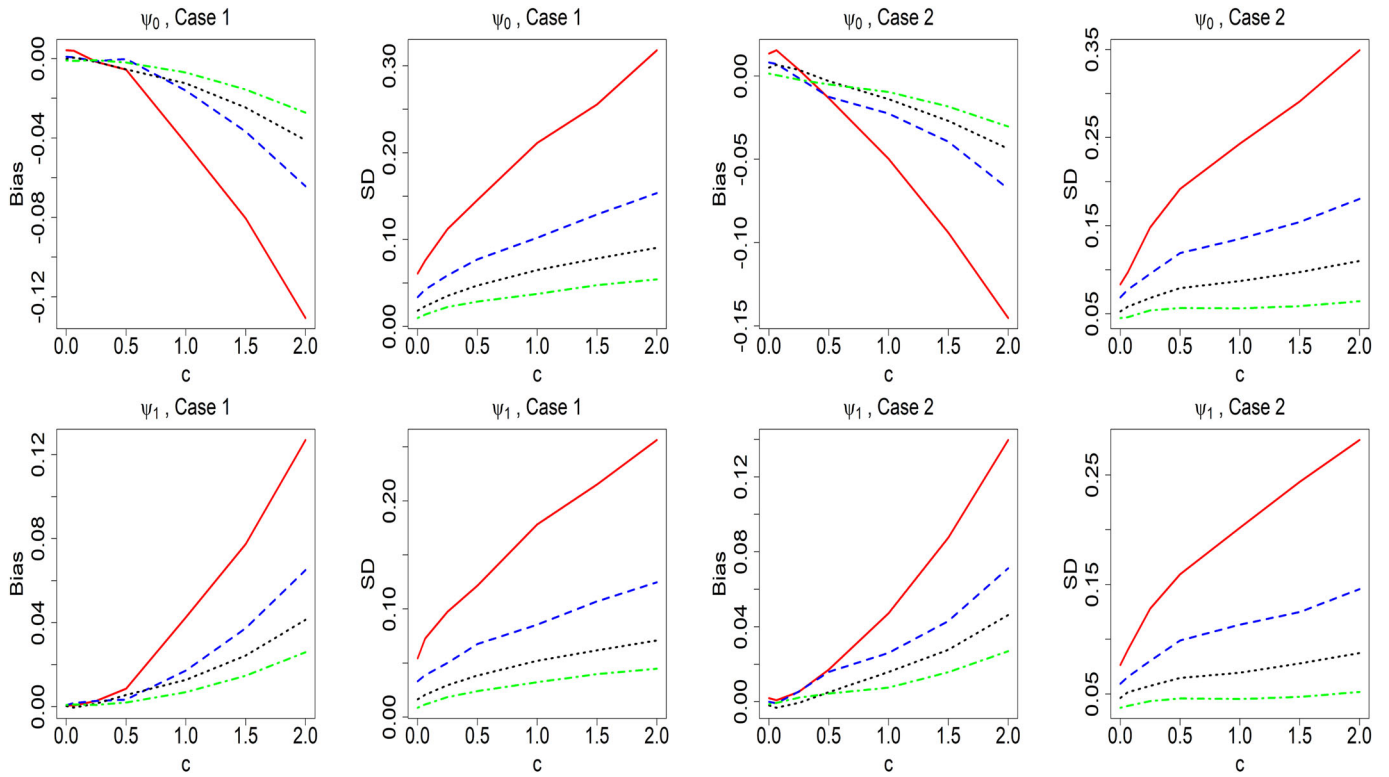
Figure 1 shows the bias and standard deviation of  $\hat{\psi}_0(\tau)$  (the intercept in the thresholding) and  $\hat{\psi}_1(\tau)$  (the threshold parameter for  $x_2$ ) at  $\tau = 0.5$ . As we expect, smaller bandwidth produces smaller bias and standard deviation, which is consistent with the discussion in Remark 3. Furthermore, Figure 2 presents the Normal Q-Q plots of  $\hat{\psi}_0(\tau)$  and  $\hat{\psi}_1(\tau)$  at  $\tau = 0.5$  for  $n = 1000$ . Results confirm that the empirical distribution of threshold parameter estimation may deviate from normal distribution when the bandwidth approaches zero. The deviation is more obvious for Case 1. As a result, if we are interested in statistical inference based on the Gaussian approximation, a larger bandwidth is preferred, such as  $c \in [0.5, 2]$ .

### 5.2. Inference for $\psi(\tau)$ at a Given Quantile Level

We first compare the Wald-type and mixed-bootstrap methods for estimating the standard deviation (SD) of  $\hat{\psi}_0(\tau)$  and  $\hat{\psi}_1(\tau)$ . As suggested in Section 5.1, we focus on the bandwidth  $h_n = cs \log(n)/\sqrt{n}$  with  $c = 0.5, 1$ , and  $2$  as such choices lead to decent Gaussian approximation. For the Wald-type method, we choose  $l_n = 1.843 \hat{\sigma} n^{-1/5}$ , where  $\hat{\sigma}$  is the sample deviation of  $\{\hat{e}_i\}_{i=1}^n$ . For the mixed-bootstrap method, the random weights  $\{g_i\}_{i=1}^n$  are generated from the Rademacher distribution, which takes values 0 and 2, each with probability 0.5, and 1000 bootstrap replicates are then generated. We conduct simulations for  $\tau = 0.5$  and  $\tau = 0.7$ .

Table 1 report the means of estimated standard deviations (SD) of  $\hat{\psi}_0(\tau)$  and  $\hat{\psi}_1(\tau)$ . The Monte Carlo SD can serve as a gold standard. We observe that the estimated SD by the mixed-bootstrap is very close to the Monte Carlo SD, while the Wald-type method shows unstable performance for Case 2.

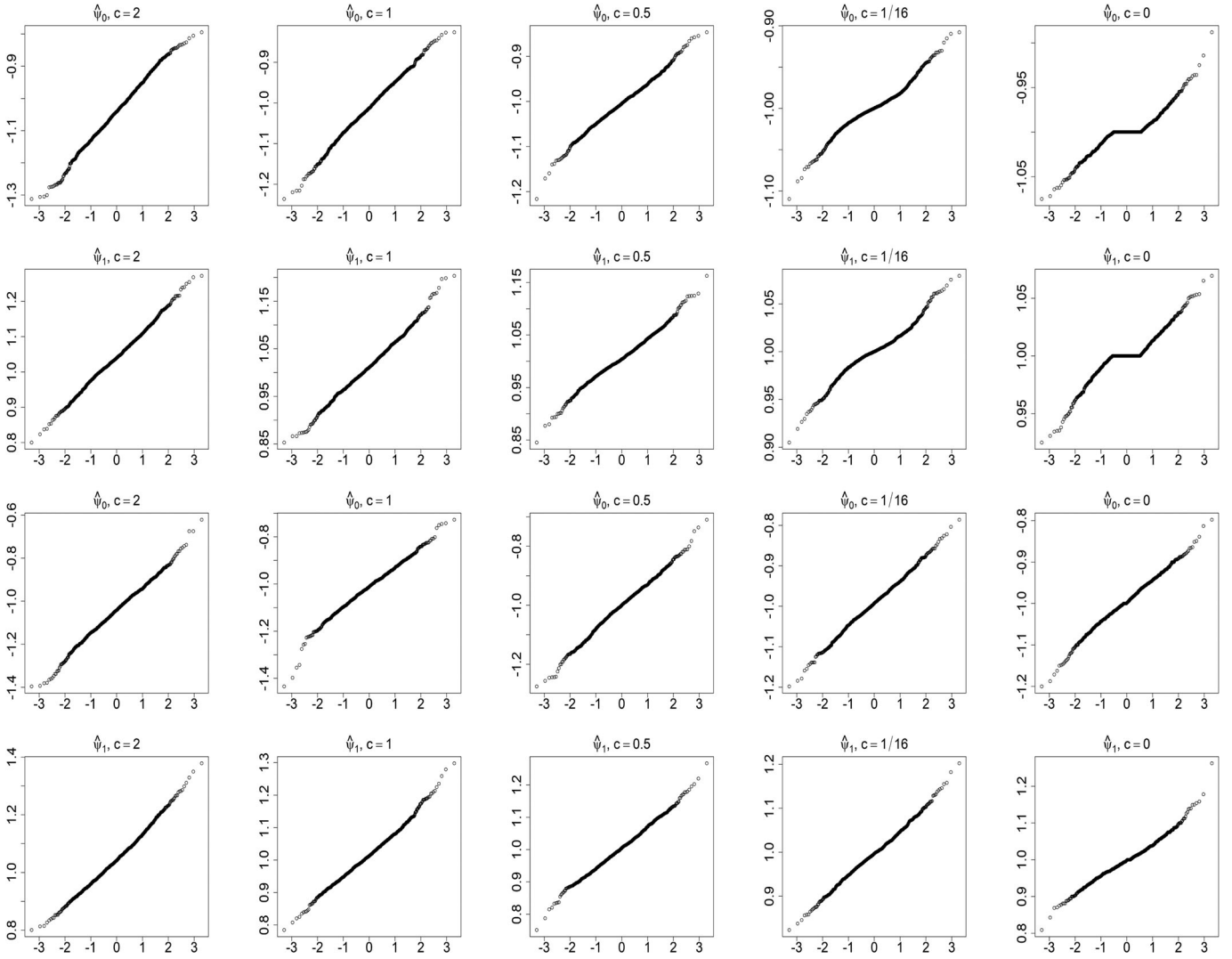
Table 2 report the coverage probabilities of the 95% confidence intervals for  $\psi_0(\tau)$  and  $\psi_1(\tau)$  obtained from different methods. The coverage probability of the Wald-type CI is generally low, especially for Case 2, partly due to the unstable SD estimation. The mixed-bootstrap method performs better, while in some cases the CIs have slight undercoverage because of the bias due to smoothing. This undercoverage is more visible for  $c = 2$ , since the larger bandwidth produces larger bias. Overall, the bias correction leads to improvement to both mixed-bootstrap and Wald-type confidence intervals. We suggest to use



**Figure 1.** Bias and standard deviation (SD) of the smoothed estimator of the threshold parameters in Cases 1-2 at  $\tau = 0.5$  with bandwidth  $h_n = cs \log(n)/\sqrt{n}$ , where  $c = 0$  represents the unsmoothed estimation. Solid:  $n = 200$ ; dashed:  $n = 500$ ; dotted:  $n = 1000$ ; dash-dot:  $n = 2000$ .

**Table 1.** Mean of estimated standard deviations for  $\hat{\psi}_0(\tau)$  and  $\hat{\psi}_1(\tau)$  in Cases 1-2 by the Wald-type (Wald) and mixed-bootstrap (MB) methods  $h_n = cs \log(n)/\sqrt{n}$ . The MC denotes the Monte Carlo standard deviation and serves as the gold standard. All quantities in the table are multiplied by 10.

$\tau = 0.5$														
Case	$c$		$n = 200$			$n = 500$			$n = 1000$			$n = 2000$		
			MC	<b>MB</b>	Wald	MC	<b>MB</b>	Wald	MC	<b>MB</b>	Wald	MC	<b>MB</b>	Wald
1	2	$\psi_0$	3.11	3.02	3.15	1.55	1.49	1.47	0.90	0.92	0.85	0.55	0.55	0.50
		$\psi_1$	2.50	2.61	2.76	1.24	1.26	1.27	0.73	0.77	0.72	0.45	0.46	0.41
	1	$\psi_0$	2.04	2.06	3.12	1.06	1.11	1.11	0.63	0.67	0.63	0.39	0.41	0.36
		$\psi_1$	1.69	1.78	2.86	0.85	0.94	0.97	0.52	0.57	0.53	0.32	0.34	0.30
	0.5	$\psi_0$	1.43	1.48	2.77	0.75	0.83	1.49	0.46	0.52	0.54	0.29	0.32	0.29
		$\psi_1$	1.26	1.29	2.32	0.63	0.72	1.26	0.39	0.45	0.48	0.24	0.27	0.25
2	2	$\psi_0$	3.42	3.41	3.45	1.77	1.69	1.55	1.04	1.07	0.91	0.67	0.68	0.53
		$\psi_1$	2.78	2.99	3.02	1.41	1.43	1.33	0.86	0.90	0.77	0.55	0.56	0.44
	1	$\psi_0$	2.34	2.36	4.22	1.37	1.35	1.20	0.85	0.90	0.70	0.56	0.59	0.40
		$\psi_1$	2.03	2.05	3.97	1.09	1.14	1.04	0.70	0.76	0.59	0.46	0.49	0.34
	0.5	$\psi_0$	1.82	1.75	7.15	1.17	1.10	1.33	0.79	0.79	0.69	0.54	0.56	0.38
		$\psi_1$	1.54	1.52	4.67	0.95	0.94	1.26	0.64	0.66	0.61	0.45	0.47	0.37
$\tau = 0.7$														
Case	$c$		$n = 200$			$n = 500$			$n = 1000$			$n = 2000$		
			MC	<b>MB</b>	Wald	MC	<b>MB</b>	Wald	MC	<b>MB</b>	Wald	MC	<b>MB</b>	Wald
1	2	$\psi_0$	3.24	3.11	3.69	1.52	1.52	1.54	0.91	0.95	0.90	0.58	0.58	0.52
		$\psi_1$	2.69	2.79	3.35	1.27	1.32	1.35	0.75	0.80	0.77	0.47	0.49	0.44
	1	$\psi_0$	2.10	2.13	2.98	1.10	1.14	1.19	0.65	0.70	0.66	0.40	0.43	0.39
		$\psi_1$	1.79	1.85	2.58	0.91	0.99	1.04	0.55	0.61	0.57	0.34	0.36	0.33
	0.5	$\psi_0$	1.54	1.54	2.71	0.82	0.86	1.11	0.48	0.54	0.55	0.30	0.33	0.31
		$\psi_1$	1.36	1.33	2.54	0.67	0.75	1.01	0.40	0.47	0.48	0.25	0.28	0.27
2	2	$\psi_0$	3.44	3.43	3.78	1.74	1.71	1.62	1.07	1.10	0.94	0.69	0.70	0.54
		$\psi_1$	2.95	3.15	3.49	1.46	1.50	1.43	0.87	0.95	0.81	0.57	0.60	0.46
	1	$\psi_0$	2.40	2.41	3.43	1.37	1.37	1.31	0.87	0.92	0.74	0.59	0.62	0.42
		$\psi_1$	2.10	2.14	3.01	1.13	1.19	1.14	0.75	0.79	0.65	0.49	0.53	0.36
	0.5	$\psi_0$	1.96	1.79	7.70	1.17	1.12	1.44	0.80	0.80	0.63	0.55	0.56	0.38
		$\psi_1$	1.69	1.58	9.45	0.96	0.97	1.42	0.68	0.69	0.58	0.46	0.48	0.34



**Figure 2.** Normal Q-Q plots for the smoothed estimator of the threshold parameters in Cases 1–2 at  $\tau = 0.5$  against the bandwidth, where  $h_n = c \log(n)/\sqrt{n}$  with  $c = 0$  representing the unsmoothed estimator. The upper two rows are for Case 1, and the bottom two rows are for Case 2.

the bias-corrected mixed-bootstrap CI, since its coverage probability results are more stable and less sensitive to the bandwidth choice across different scenarios. Please refer to the Supplementary Material for additional simulation results, including cases with more threshold variables, and the comparison with the likelihood-ratio-based inference method in Su and Xu (2019), and with the mean inference method in Seo and Linton (2007).

### 5.3. Testing the Constancy of $\psi(\tau)$ Across $\tau$

In this subsection, we focus on testing whether  $\psi(\tau)$  is constant across  $\tau$ . We generate data as in Case 1 and increase  $d$  from 0 to 0.7. When  $d = 0$ , the threshold parameters  $\psi_0(\tau) = (-1, 1)^T$  are constant across  $\tau \in (0, 1)$ , corresponding to the null model. When  $d > 0$ ,  $\psi_0(\tau) = (-1 + d\Phi^{-1}(\tau), 1)^T$  varies with  $\tau$ , and the larger value of  $d$  is associated with more deviation from the null hypothesis.

We assess the Type I error and power of the two proposed KS-type tests:  $KS_1$  and  $KS_2$ , which assume known and unknown  $\psi_0$ , respectively. We calculate test statistics based on two sets of quantile grids:  $m = 17$  and  $m = 33$  quantiles equally

spaced between  $[0.1, 0.9]$ . For both methods, the critical values are obtained by using the mixed-bootstrap with 1000 bootstrap samples. When still using  $h_n = c \log(n)/\sqrt{n}$  with  $c = 0.5, 1, 2$ , we found that  $c = 1, 0.5$  tend to be somewhat conservative. The rejection percentages at the 5% significance level with  $c = 2$  are summarized in Table 3. The two tests give similar performance: the sizes are well controlled under the null hypothesis, and the power increases gradually to one as  $d$  increases. In addition, the two tests seem to be insensitive to the number of quantile levels  $m$  used to calculate the statistics; our experience suggests that it is often sufficient to use quantile grids with the mesh less than 0.05.

## 6. An Application to the Executive Compensation

The rapid rise in executive compensation from the mid-1980s until 2005 has sparked an intense debate and investigation regarding the nature of the pay-setting process; see Bertrand and Mullainathan (2000), Garvey and Milbourn (2006), Frydman and Jenter (2010), and Edmans et al. (2017). Compensation is

**Table 2.** Coverage percentages of the 95% confidence intervals for  $\psi_0(\tau)$  and  $\psi_1(\tau)$  in Cases 1-2 by the Wald-type (Wald), Wald-type with bias correction (WBC), mixed-bootstrap (MB) and mixed-Bootstrap with bias correction (BBC), where  $h_n = cs \log(n)/\sqrt{n}$ .

$\tau = 0.5$																		
Case	c		n = 200				n = 500				n = 1000				n = 2000			
			Wald	WBC	MB	BBC	Wald	WBC	MB	BBC	Wald	WBC	MB	BBC	Wald	WBC	MB	BBC
1	2	$\psi_0$	93.6	96.4	92.4	94.9	91.0	94.7	90.5	94.2	91.0	95.5	91.6	95.6	90.8	95.6	91.0	95.3
		$\psi_1$	95.5	97.2	94.0	96.1	92.6	96.1	92.4	95.7	92.5	96.1	92.9	97.2	90.1	95.3	90.3	96.5
	1	$\psi_0$	93.5	95.1	93.4	94.6	93.6	94.5	95.1	96.3	93.5	94.8	94.4	95.6	90.9	93.5	93.9	95.7
		$\psi_1$	94.1	95.3	94.3	95.2	94.2	95.9	96.0	96.8	93.8	95.5	95.7	96.5	91.2	94.6	95.0	96.2
	0.5	$\psi_0$	91.7	92.9	93.4	93.8	93.2	94.9	96.5	96.8	92.2	94.5	96.5	97.2	90.5	92.6	95.2	96.1
		$\psi_1$	91.4	92.5	93.4	94.3	94.4	95.7	97.3	97.5	92.9	94.5	97.6	97.9	91.3	93.1	95.6	96.8
2	2	$\psi_0$	93.1	95.7	92.7	94.9	89.7	94.6	90.7	93.9	90.0	94.8	91.8	95.0	86.1	93.0	92.2	95.5
		$\psi_1$	93.7	96.2	94.5	96.6	91.8	94.9	91.6	94.9	90.3	95.6	92.9	96.8	85.7	92.6	91.4	95.5
	1	$\psi_0$	93.1	94.4	92.9	94.4	90.6	92.7	93.3	94.9	88.2	91.3	94.6	96.2	83.7	88.1	94.2	96.2
		$\psi_1$	93.2	94.2	93.3	94.7	91.6	93.7	94.7	95.7	89.0	91.4	94.9	96.1	84.1	89.6	94.9	95.7
	0.5	$\psi_0$	87.0	87.9	90.3	91.1	84.0	86.3	92.9	94.2	81.9	85.2	94.0	95.0	75.0	81.1	94.1	95.2
		$\psi_1$	88.9	89.9	91.5	92.1	86.7	89.3	93.5	94.0	85.2	87.2	93.8	94.8	75.8	82.0	95.6	96.3
$\tau = 0.7$																		
Case	c		n = 200				n = 500				n = 1000				n = 2000			
			Wald	WBC	MB	BBC	Wald	WBC	MB	BBC	Wald	WBC	MB	BBC	Wald	WBC	MB	BBC
1	2	$\psi_0$	90.7	93.6	90.6	93.2	92.4	94.6	92.3	95.1	93.7	96.7	93.8	96.3	90.1	94.4	92.3	95.6
		$\psi_1$	92.8	95.9	92.6	95.1	93.7	97.1	91.8	96.9	92.3	96.0	93.0	97.2	90.8	96.3	92.4	96.5
	1	$\psi_0$	92.2	93.9	92.2	93.4	90.8	92.9	94.6	95.9	92.5	95.1	95.4	96.8	90.9	94.9	93.3	95.8
		$\psi_1$	91.7	93.4	93.7	95.0	92.6	94.1	95.3	97.0	93.6	95.9	95.4	96.5	91.9	93.9	94.8	96.5
	0.5	$\psi_0$	90.8	92.0	92.2	93.1	90.9	92.9	95.5	96.3	91.6	93.7	95.5	96.4	90.7	93.9	94.8	96.5
		$\psi_1$	90.5	92.0	93.8	94.4	92.0	93.2	96.3	96.5	92.6	94.2	96.4	97.0	92.2	94.1	96.4	97.2
2	2	$\psi_0$	89.8	92.6	92.8	94.4	91.5	94.2	92.3	94.9	90.0	94.9	92.8	95.4	88.9	93.3	92.8	95.7
		$\psi_1$	92.1	95.5	93.6	95.1	92.4	96.4	92.8	96.4	89.6	95.3	93.5	96.1	87.8	93.8	92.4	96.5
	1	$\psi_0$	89.5	91.3	91.2	93.0	90.0	92.6	93.4	94.6	89.7	92.6	95.0	96.1	83.5	88.9	94.0	95.3
		$\psi_1$	91.2	92.7	93.3	94.7	90.9	92.7	94.8	95.8	88.6	92.2	93.6	95.2	84.8	89.0	94.8	95.8
	0.5	$\psi_0$	85.0	86.7	89.2	90.7	86.6	89.4	91.8	93.3	83.7	88.2	93.1	94.0	79.0	84.5	93.6	94.6
		$\psi_1$	86.5	88.3	91.7	92.6	88.2	90.1	93.5	95.0	83.2	86.6	94.6	95.6	79.9	84.7	94.9	95.7

**Table 3.** Rejection percentages of two proposed tests for testing the constancy of  $\psi(\tau)$  across  $\tau \in [0.1, 0.9]$  with  $h_n = 2s \log(n)/\sqrt{n}$ .  $KS_1$  and  $KS_2$  are tests assuming known and unknown  $\psi_0$ , respectively.  $m$  is the number of equally spaced quantiles in  $[0.1, 0.9]$ . The nominal level is 5%.

n	d = 0				d = 0.05				d = 0.1			
	m = 17		m = 33		m = 17		m = 33		m = 17		m = 33	
	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$
1000	4.3	4.6	3.9	3.5	8.6	7.3	7.6	5.8	15.4	13.6	14.4	13.3
2000	5.2	5.1	4.5	4.6	14.1	13.5	13.1	12.4	39.9	37.7	36.5	35.1
n	d = 0.2				d = 0.3				d = 0.4			
	m = 17		m = 33		m = 17		m = 33		m = 17		m = 33	
	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$
1000	40.7	37.5	37.7	36.0	63.2	63.0	60.6	58.4	77.3	77.4	74.2	75.4
2000	82.1	80.1	79.6	78.3	94.6	95.4	95.1	94.7	98.9	98.6	98.4	98.0
n	d = 0.5				d = 0.6				d = 0.7			
	m = 17		m = 33		m = 17		m = 33		m = 17		m = 33	
	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$	$KS_1$	$KS_2$
1000	85.7	85.5	83.6	83.3	90.7	89.4	88.0	86.9	92.0	91.6	88.8	89.4
2000	99.7	99.4	99.4	99.1	99.9	99.5	99.7	99.5	99.9	99.6	99.7	99.6

generally decided at the end of each fiscal year by the compensation committee of the board of directors, at which time they already know whether market performance is beneficial or not for the firm (Garvey and Milbourn 2006).

Some researchers such as Bertrand and Mullainathan (2001) found that executives can benefit from *luck*. *Luck* represents the

market performance, that is, changes in the firm performance that are beyond the chief executive officer's control. Further, Garvey and Milbourn (2006) found that the effect of *luck* on compensation is significantly larger when *luck* is up ( $luck > 0$ ) than when *luck* is down ( $luck < 0$ ). They refer to this phenomenon as "the asymmetry in pay for *luck*".



We will apply the proposed methods to an executive compensation data to study the heterogeneous effects of predictors on compensation. The executive's compensation data used in this article were drawn from Standard and Poor's ExecuComp, and the firm's return data were drawn from the Center for Research in Security Prices (CRSP), available from <https://wrds-www.wharton.upenn.edu/users/tou/>. Our sample period covers the years 1993-2005, when the data was available. In total there are in total 15837 observations for 2508 firms and 3709 executives. The response  $y$  is defined as the change in the logarithm of total compensation. We consider two predictors, *luck* and *skill*, which are determined as follows. First, the firm performance (measured by the firm's one-year stock return) is regressed on equal-weighted and value-weighted industry returns. Then the predicting value (centered at zero) is taken to represent *luck* (i.e., market performance), and the residual is used to represent *skill* (i.e., firm-specific performance).

As an exploratory data analysis, we plot in Figure 3 the estimation results from quantile regression of  $y$  on *luck* or *skill* for the subgroups  $luck > 0$  and  $luck < 0$  separately at three quantiles  $\tau = 0.2, 0.5$ , and  $0.8$ . At all three quantiles, the effects of *luck* and *skill* appear to differ in two subgroups, indicating the "asymmetry in pay for luck" phenomenon. In addition, the asymmetry patterns tend to vary across quantiles. As pointed out in Garvey and Milbourn (2006), the asymmetry may not necessarily occur around the zero value of *luck*. In addition, as suggested in Yu and Fan (2020), the threshold may depend on both *luck* and the firm size. These motivate us to consider the following model:

$$Q_y(\tau|L, S, F) = \beta_0(\tau) + \beta_1(\tau)L + \beta_2(\tau)S + \{\delta_1(\tau)L + \delta_2(\tau)S\} I\{L > \psi_0(\tau) + \psi_1(\tau)F\}, \quad (8)$$

where  $L$  is *luck*,  $S$  is *skill* and  $F$  is the standardized firm size, measured by the market value of the firm in millions of dollars. The model allows us to analyze the heterogeneous effect of predictors at different quantiles. In addition, the response distribution appears to be heavy-tailed (Figure 4), so median regression may be more desirable than mean regression for analyzing this data.

We fit the model in Equation (9) at 33 equally spaced quantile levels in  $[0.1, 0.9]$  by using the smoothed estimation

method. The bandwidth is chosen as  $h_n = \hat{s}(\tau) \log(n)/\sqrt{n}$ , where  $\hat{s}(\tau)$  is the interquantile range of  $x^T \check{\psi}(\tau)$ , and  $\check{\psi}(\tau)$  is the unsmoothed estimator. Since the firm size is heavily skewed, we use the interquantile range to get a robust estimate of the standard deviation. Figure 5 presents the estimates of  $\delta_1(\tau), \delta_2(\tau), \psi_0(\tau), \psi_1(\tau)$  across  $\tau \in [0.1, 0.9]$  along with the 95% pointwise confidence bands. The confidence intervals for  $\delta_1(\tau), \delta_2(\tau)$  are constructed by inverting the rank score test in Koenker (2005) with estimated threshold parameters plugged in. The confidence intervals for  $\psi_0(\tau), \psi_1(\tau)$  are constructed by the mixed-bootstrap method with bias correction.

Some interesting findings emerge from Figure 5. First, results suggest that executives in the upper quantiles ( $\tau \geq 0.65$ ) enjoy the enhanced "pay for luck," while those at the lower quantiles suffer from reduced "pay for luck." One possible explanation is that executives at the right tail of the compensation distribution may have more influence over their pay. As a result, they can insulate from bad luck and enjoy more rewards from good luck. On the other hand, executives at lower quantiles are likely to be

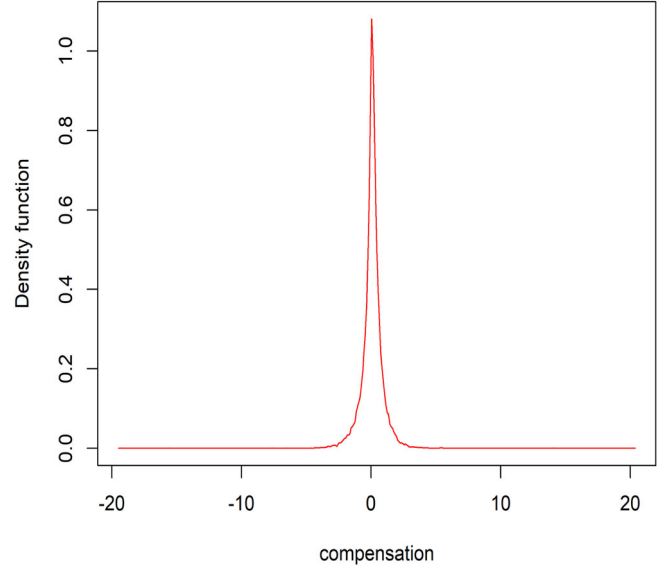


Figure 4. Kernel density estimation of the marginal response distribution in the compensation data.

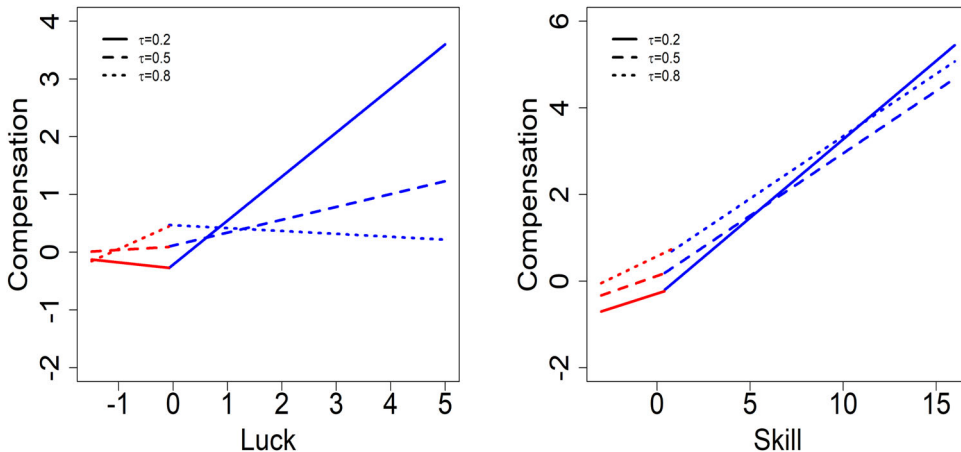
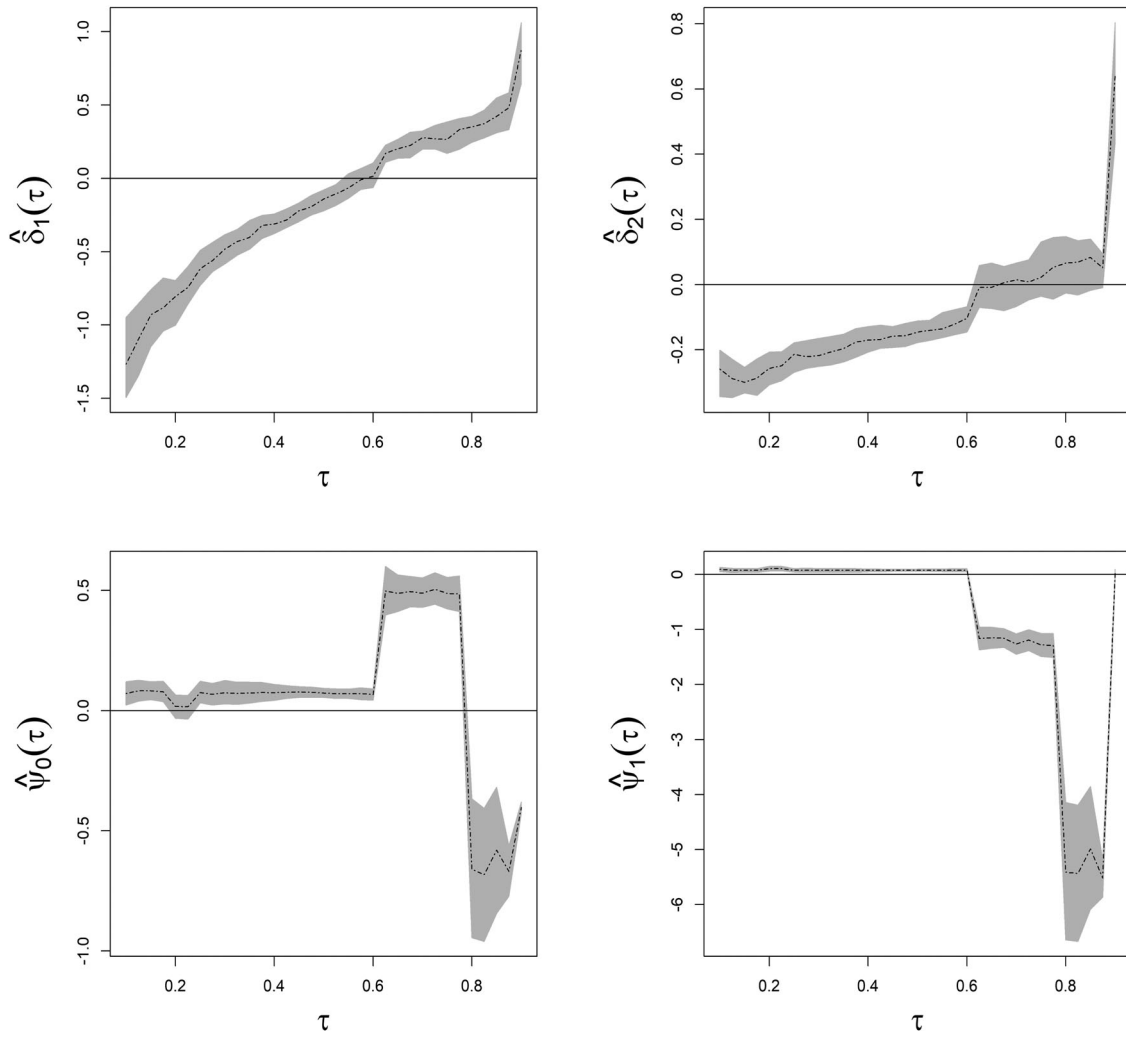


Figure 3. The estimated quantile regression lines by regressing the compensation on *luck* (left) or *skill* (right) for the two subgroups with  $luck > 0$  (blue) and  $luck < 0$  (red) separately, at the quantile levels  $\tau = 0.2, 0.5$ , and  $0.8$ .



**Figure 5.** The estimation for the regression coefficients of *luck* and *skill*,  $(\delta_1(\tau), \delta_2(\tau))$ , and the intercept and slope coefficient of firm size in the thresholding,  $(\psi_0(\tau), \psi_1(\tau))$ , across  $\tau \in [0.1, 0.9]$ . The shaded areas are the corresponding 95% pointwise confidence bands.

more risk-averse and have lower incentive contracts, exemplified by Becker (2006). Thus, these executives cannot enjoy more when the market is bull.

Second, Figure 5 suggests the threshold parameters tend to be common in three quantile regions. Thus, we apply the proposed KS-type test to test the constancy of  $\psi(\tau)$  across  $\tau \in [0.1, 0.6]$ ,  $[0.65, 0.75]$ ,  $[0.8, 0.85]$  separately. Following the suggestion from the simulation study, we calculate the KS statistic based on equally distanced grids from the quantile region with  $\text{meth}$  equals 0.025, and let  $h_n = 2\hat{s} \log(n)/\sqrt{n}$ , where  $\hat{s} = 0.32$  is the interquantile range of  $x^T \check{\psi}(\tau)$  calculated at  $\tau = 0.5$ . The resulting  $p$ -values are 0.61, 0.55, 0.41, suggesting that the thresholds do not vary significantly within each of the three quantile regions. On the other hand, testing the constancy of  $\psi(\tau)$  across  $\tau \in [0.1, 0.9]$  gives the  $p$ -value of zero, indicating that the thresholds vary significantly across this wider region.

We now use the piecewise constancy information and obtain the composite quantile estimators for  $\psi(\tau)$  within each quantile region, using the same idea as in Equation (5). The resulting composite estimations are  $(0.07, 0.08)$ ,  $(0.49, -1.28)$ , and  $(-0.51, -4.79)$  within  $[0.1, 0.6]$ ,  $[0.65, 0.75]$ , and  $[0.8, 0.85]$ , respectively. The result implies that a larger firm size would reduce the threshold of enhanced “pay for *luck*” when  $\tau >$

0.65, while it has a minimal effect on the threshold for lower quantiles, which is consistent with the observations from Figure 5. Bertrand and Mullainathan (2001) argued that larger firms typically have fewer large shareholders due to the more expensive shares. Thus, executives in larger firms can have more influence over their pay. This point was also noted in Garvey and Milbourn (2006).

## 7. Discussion

In this article, we consider single-index thresholding in quantile regression. This model completes the subgroup literature by characterizing subgroups with thresholding. Technically, the additional nonconstant variables in thresholding pose more challenges to the statistical inference of threshold parameters. We propose a smoothed estimator for convenient statistical inference both for a single quantile and for the quantile process. The proposed profile estimation algorithm works well with a modest number of threshold variables. However, when the number of candidate threshold variables is large, more efficient algorithms are needed to conduct estimation and variable selection. We develop tests to assess the constancy of the vector of threshold parameters  $\psi(\tau)$  across  $\tau$ . The proposed method

can be directly extended to test the constancy of a subset of  $\psi(\tau)$  across  $\tau$ . If the test fails to reject the null hypothesis, a more efficient composite quantile estimation can be used to estimate the common parameter by pooling information across quantiles. We leave the study of theoretical properties of the composite quantile estimator and related inference to the future work.

For the inference of threshold parameters, an interesting question is whether a bootstrap method can be constructed directly for the nonsmoothed estimator. The standard bootstrap method was proven to be inconsistent for the threshold mean regression (Yu 2014). Common solutions to remedy the limitation of standard bootstrap include the smoothed bootstrap and the subsampling. We leave it as a future topic to investigate the possibility of constructing valid bootstrap methods for the nonsmoothed estimator in the single-index threshold quantile regression framework.

While the proposed methods work for models with multiple threshold variables, the computation could be challenging when the dimension of threshold variables is high. To accommodate high-dimensional covariates, one possible solution is to adapt the path-following algorithm in the recent article Feng, Ning and Zhao (2019) that is developed for binary threshold regression. The computational and statistical guarantees of this algorithm are largely unknown and require more careful investigation.

## Appendix

The appendix includes the proof of Theorem 2. We relegate other technical proofs in the online supplement. Hereafter, every convergence is as  $n \rightarrow \infty$  and the integral  $\int$  is taken over the whole space unless specified otherwise. Denote  $\mathbb{P}_{nf} = \frac{1}{n} \sum_{i=1}^n f_i$  and  $\mathbb{G}_{nf} = \sqrt{n}(\mathbb{P}_{nf} - E f)$ . Throughout the proof of Theorem 2, we omit the argument  $(\tau)$  in notation  $\eta(\tau)$ ,  $\beta(\tau)$ ,  $\delta(\tau)$ ,  $\psi(\tau)$ ,  $q(\tau) = x_1 + x_2^T \psi_0(\tau)$  when no confusion is caused.

### A.1. Proof of Theorem 2

First define  $\Delta(\eta; h_n) = z^T(\beta - \beta_0(\tau)) + \tilde{z}^T \delta G \left\{ h_n^{-1}(x_1 + x_2^T \psi) \right\} - \tilde{z}^T \delta_0(\tau) I(q > 0)$ . Then denote  $f^\theta(w, \eta; h_n) = \left\{ f^\beta(w, \eta; h_n)^T, f^\delta(w, \eta; h_n)^T \right\}^T$  where  $f^\beta(w, \eta; h_n) = [\tau - I\{e(\tau) \leq \Delta(\eta; h_n)\}] z$  and  $f^\delta(w, \eta; h_n) = [\tau - I\{e(\tau) \leq \Delta(\eta; h_n)\}] \tilde{z} G \left\{ h_n^{-1}(x_1 + x_2^T \psi) \right\}$ . Denote  $f^\psi(w, \eta; h_n) = [\tau - I\{e(\tau) \leq \Delta(\eta; h_n)\}] \tilde{z}^T \delta G' \left\{ h_n^{-1}(x_1 + x_2^T \psi) \right\} x_2 h_n^{-1}$ . By the computational properties of the quantile regression estimator, see Theorem 3.3 in Koenker and Bassett (1978), we have that with Assumption 2(a)

$$\begin{aligned} \mathbb{P}_{nf} f^\theta(w, \hat{\eta}(\tau); h_n) &= O_p(n^{-1}), \\ \mathbb{P}_{nf} f^\psi(w, \hat{\eta}(\tau); h_n) &= O_p(n^{-1} h_n^{-1}). \end{aligned}$$

Thus, we have

$$n^{-1/2} \mathbb{G}_{nf} f^\theta(w, \eta_0(\tau); h_n) + E \left\{ f^\theta(w, \hat{\eta}; h_n) \right\} + \mathbb{I}_n^\theta = O_p(n^{-1}), \quad (\text{A.1})$$

$$n^{-1/2} \mathbb{G}_{nf} f^\psi(w, \eta_0(\tau); h_n) + E \left\{ f^\psi(w, \hat{\eta}; h_n) \right\} + \mathbb{I}_n^\psi = O_p(n^{-1} h_n^{-1}), \quad (\text{A.2})$$

where we denote  $\mathbb{I}_n^\theta = n^{-1/2} \left\{ \mathbb{G}_{nf} f^\theta(w, \hat{\eta}; h_n) - \mathbb{G}_{nf} f^\theta(w, \eta_0(\tau); h_n) \right\}$  and similarly  $\mathbb{I}_n^\psi = n^{-1/2} \left\{ \mathbb{G}_{nf} f^\psi(w, \hat{\eta}; h_n) - \mathbb{G}_{nf} f^\psi(w, \eta_0(\tau); h_n) \right\}$ . Expanding  $E \left\{ f^\theta(w, \hat{\eta}; h_n) \right\}$  in (A.1) around  $\eta_0(\tau)$  and applying Lemma 4(d), we have

$$\begin{aligned} \mathbb{P}_{nf} f^\theta(w, \eta_0(\tau); h_n) &= Q^\theta(\eta_0(\tau); h_n) \left\{ \hat{\theta} - \theta_0(\tau) \right\} \\ &\quad + \{h_n Q^{\theta\psi}(\eta_0(\tau); h_n)\} \frac{\hat{\psi} - \psi_0(\tau)}{h_n} - \mathbb{I}_n^\theta \\ &\quad + O_p\left(\frac{1}{n} \vee \|\hat{\theta} - \theta_0(\tau)\|^2 \vee \left\| \frac{\hat{\psi} - \psi_0(\tau)}{h_n} \right\|^2\right). \end{aligned} \quad (\text{A.3})$$

Expanding  $E \left\{ f^\psi(w, \hat{\eta}; h_n) \right\}$  in Equation (A.2) around  $\eta_0(\tau)$  and applying Lemma 4(d), we have

$$\begin{aligned} \mathbb{P}_{nf} f^\psi(w, \eta_0(\tau); h_n) &= Q^{\theta\psi}(\eta_0(\tau); h_n)^T \left\{ \hat{\theta} - \theta_0(\tau) \right\} \\ &\quad + \{h_n Q^\psi(\eta_0(\tau); h_n)\} \frac{\hat{\psi} - \psi_0(\tau)}{h_n} - \mathbb{I}_n^\psi \\ &\quad + O_p\left(\frac{1}{nh_n} \vee \|\hat{\theta} - \theta_0(\tau)\|^2 \vee \left\| \frac{\hat{\psi} - \psi_0(\tau)}{h_n} \right\|^2\right). \end{aligned} \quad (\text{A.4})$$

where

$$\begin{aligned} Q^\theta(\eta_0(\tau); h_n) &= -\frac{\partial}{\partial(\theta')} E \left\{ f^\theta(w, \eta; h_n) \right\}_{\eta=\eta_0(\tau)}, \\ Q^{\theta\psi}(\eta_0(\tau); h_n) &= -\frac{\partial}{\partial(\psi')} E \left\{ f^\theta(w, \eta; h_n) \right\}_{\eta=\eta_0(\tau)}, \\ Q^\psi(\eta_0(\tau); h_n) &= -\frac{\partial}{\partial(\psi')} E \left\{ f^\psi(w, \eta; h_n) \right\}_{\eta=\eta_0(\tau)}. \end{aligned} \quad (\text{A.5})$$

With  $\mathbb{P}_{nf} f^\theta(w, \eta_0(\tau); h_n) = O_p(n^{-1/2})$  (Lemma 3),  $Q^\theta(\eta_0(\tau); h_n) = O(1)$  (Lemma 4),  $Q^{\theta\psi}(\eta_0(\tau); h_n) = O(1)$  (Lemma 4),  $\mathbb{I}_n^\theta = o_p(n^{-1/2})$  (Lemma 6) and  $h_n^{-1} \left\{ \hat{\psi} - \psi_0(\tau) \right\} \rightarrow 0$  (Lemma 5), we have from Equation (A.4) that

$$\hat{\theta} - \theta_0(\tau) = O_p \left( \frac{1}{\sqrt{n}} \vee \|\hat{\psi}(\tau) - \psi_0(\tau)\| \vee \left\| \frac{\hat{\psi}(\tau) - \psi_0(\tau)}{h_n} \right\|^2 \right).$$

Plugging this result into Equation (A.5) together with  $\mathbb{P}_{nf} f^\psi(w, \eta_0(\tau); h_n) = O_p((nh_n)^{-1/2})$  (Lemma 2),  $Q^{\theta\psi}(\eta_0(\tau); h_n) = O(1)$  (Lemma 4),  $h_n Q^\psi(\eta_0(\tau); h_n) = O(1)$  (Lemma 4) and  $\mathbb{I}_n^\psi = o_p((nh_n)^{-1/2})$  (Lemma 6), we have

$$\left\| \frac{\hat{\psi} - \psi_0(\tau)}{h_n} \right\| = O_p\left(\frac{1}{\sqrt{nh_n}}\right).$$

Thus,  $\hat{\theta} - \theta_0(\tau) = O_p \left( n^{-1/2} \vee (h_n/n)^{-1/2} \vee (nh_n)^{-1} \right)$ . Because  $nh_n^2 \rightarrow \infty$ , we get  $\|\hat{\theta} - \theta_0(\tau)\| = O_p(n^{-1/2})$ . Then we can get the expansions

$$\begin{aligned} \sqrt{n} \left\{ \hat{\theta} - \theta_0(\tau) \right\} &= Q^\theta(\eta_0(\tau); h_n)^{-1} \sqrt{n} \mathbb{P}_{nf} f^\theta(w, \eta_0(\tau); h_n) \\ &\quad + o_p(1), \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} \sqrt{\frac{n}{h_n}} \left\{ \hat{\psi} - \psi_0(\tau) \right\} &= Q^\psi(\eta_0(\tau); h_n)^{-1} \sqrt{nh_n} \mathbb{P}_{nf} f^\psi(w, \eta_0(\tau); h_n) \\ &\quad + o_p(1). \end{aligned} \quad (\text{A.7})$$

The asymptotic normality then follows from Lemmas 2 and 3. The asymptotic independence result follows from Lemma 1, where we prove that the covariance between  $\sqrt{n} \mathbb{P}_{nf} f^\theta(w, \eta_0(\tau); h_n)$  and  $\sqrt{nh_n} \mathbb{P}_{nf} f^\psi(w, \eta_0(\tau); h_n)$  are asymptotically negligible.

## Supplementary Material

The online supplement contains additional simulation results, and the proofs for the theorems in the article.

## Acknowledgments

The authors are grateful to the editor, the AE, and two referees for their helpful suggestions that significantly improved the quality of the article. The authors would like to thank Drs Xiaodong Fan and Ping Yu for sharing the compensation data. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily the views of the NSF.

## Funding

The research of Zhang is supported by grant KLATASDS200204. The research of Wang was partly supported by the IR/D program from the US National Science Foundation (NSF) and the NSF grant DMS-1712760. The research of Zhu is partially supported by the National Natural Science Foundation of China 11671096, 11731011, 12071087.

## ORCID

Zhongyi Zhu  <http://orcid.org/0000-0003-0683-8106>

## References

- Bai, J. (1995), "Least Absolute Deviation Estimation of a Shift," *Econometric Theory*, 11, 403–436. [2,3]
- Becker, B. (2006), "Wealth and Executive Compensation," *The Journal of Finance*, 61, 379–397. [13]
- Bertrand, M., and Mullainathan, S. (2000), "Agents With and Without Principals," *American Economic Review*, 90, 203–208. [10]
- (2001), "Are Executives Paid for Luck? The Ones Without Principals Are," *Quarterly Journal of Economics*, 116, 901–932. [11,13]
- Brown, B., and Wang, Y.-G. (2005), "Standard Errors and Covariance Matrices for Smoothed Rank Estimators," *Biometrika*, 92, 149–158. [3]
- Cai, Y. (2010), "Forecasting for Quantile Self-exciting Threshold Autoregressive Time Series Models," *Biometrika*, 97, 199–208. [1]
- Cai, Y., and Stander, J. (2008), "Quantile Self-exciting Threshold Autoregressive Time Series Models," *Journal of Time Series Analysis*, 29, 186–202. [1]
- Caner, M. (2002), "A Note on Least Absolute Deviation Estimation of a Threshold Model," *Econometric Theory*, 18, 800–814. [2,3]
- Chan, K.-S. (1993), "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520–533. [2,3]
- Chernozhukov, V., and Hansen, C. (2006), "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491–525. [8]
- Chernozhukov, V., and Hong, H. (2004), "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72, 1445–1480. [4]
- Christopoulos, D., McAdam, P., and Tzavalis, E. (2021), "Dealing With Endogeneity in Threshold Models Using Copulas," *Journal of Business & Economic Statistics*, 39, 166–178. [1]
- Durbin, J. (1973), *Distribution Theory for Tests Based on the Sample Distribution Function*, vol. 9. Philadelphia: SIAM. [7]
- Durlauf, S. N., Johnson, P. A. and Temple, J. R. (2005), "Growth Econometrics," *Handbook of Economic Growth*, 1, 555–677. [1]
- Edmans, A., Gabaix, X. and Jenter, D. (2017), "Executive Compensation: A Survey of Theory and Evidence," in Hermalin, B., Weisbach M. (Eds.), *The Handbook of the Economics of Corporate Governance*, Vol. 1, 383–539. Amsterdam: Elsevier. [10]
- Fan, A., Song, R., and Lu, W. (2017), "Change-plane Analysis for Subgroup Detection and Sample Size Calculation," *Journal of the American Statistical Association*, 112, 769–778. [1,2]
- Feng, H., Ning, Y., and Zhao, J. (2019), "Nonregular and Minimax Estimation of Individualized Thresholds in High Dimension With Binary Responses," arXiv: Statistics Theory. [14]
- Frydman, C., and Jenter, D. (2010), "Ceo Compensation," *Annual Review of Financial Economics*, 2, 75–102. [10]
- Fu, L., Wang, Y.-G., and Bai, Z. (2010), "Rank Regression for Analysis of Clustered Data: A Natural Induced Smoothing Approach," *Computational Statistics & Data Analysis*, 54, 1036–1050. [3]
- Galvao, A. F., Kato, K., Montes-Rojas, G., and Olmo, J. (2014), "Testing Linearity Against Threshold Effects: Uniform Inference in Quantile Regression," *Annals of the Institute of Statistical Mathematics*, 66, 413–439. [3]
- Galvao, A. F., Montes-Rojas, G., and Olmo, J. (2011), "Threshold Quantile Autoregressive Models," *Journal of Time Series Analysis*, 32, 253–267. [1]
- Garvey, G. T., and Milbourn, T. T. (2006), "Asymmetric Benchmarking in Compensation: Executives are Rewarded for Good Luck But Not penalized for Bad," *Journal of Financial Economics*, 82, 197–225. [10,11,12,13]
- Hansen, B. E. (2000), "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603. [1,2,3]
- He, Y., Lin, H., and Tu, D. (2018), "A Single-index Threshold Cox Proportional Hazard Model for Identifying a Treatment-sensitive Subset based on Multiple Biomarkers," *Statistics in Medicine*, 37, 3267–3279. [1]
- Horowitz, J. L. (1992), "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531. [2,3]
- Hunter, D. R. and Lange, K. (2000), "Quantile Regression Via an MM Algorithm," *Journal of Computational and Graphical Statistics*, 9, 60–77. [6]
- Jiang, Z., Du, C., Jablensky, A., Liang, H., Lu, Z., Ma, Y., and Teo, K. L. (2014), "Analysis of Schizophrenia Data Using a Nonlinear Threshold index Logistic Model," *PloS One*, 9, e109454. [1]
- Kaufman, J. S., Asuzu, M. C., Mufunda, J., Forrester, T., Wilks, R., Luke, A., Long, A. E. and Cooper, R. S. (1997), "Relationship Between blood Pressure and Body Mass Index in Lean Populations," *Hypertension*, 30, 1511–1516. [1]
- Kerry, S. M., Micah, F. B., Plange-Rhule, J., Eastwood, J. B., and Cappuccio, F. P. (2005), "Blood Pressure and Body Mass Index in Lean rural and Semi-urban Subjects in West Africa," *Journal of Hypertension*, 23, 1645–1651. [1]
- Kocherginsky, M., He, X. and Mu, Y. (2005), "Practical Confidence Intervals for Regression Quantiles," *Journal of Computational and Graphical Statistics*, 14, 41–55. [6]
- Koenker, R. (2005), *Quantile Regression*, Cambridge: Cambridge University Press. [2,3,4,6,12]
- Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50. [14]
- Koenker, R., and Park, B. J. (1996), "An Interior Point Algorithm for Nonlinear Quantile Regression," *Journal of Econometrics*, 71, 265–283. [6]
- Koenker, R., and Xiao, Z. (2002), "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612. [8]
- Kuan, C.-M., Michalopoulos, C., and Xiao, Z. (2017), "Quantile Regression on Quantile Ranges—A Threshold Approach," *Journal of Time Series Analysis*, 38, 99–119. [2,3]
- Lee, S., Seo, M. H., and Shin, Y. (2011), "Testing for Threshold Effects in Regression Models," *Journal of the American Statistical Association*, 106, 220–231. [1,3]
- Li, C., Wei, Y., Chappell, R., and He, X. (2011), "Bent Line Quantile Regression With Application to An Allometric Study of Land Mammals' Speed and mass," *Biometrics*, 67, 242–249. [1]
- Pang, L., Lu, W., and Wang, H. J. (2012), "Variance Estimation in Censored Quantile Regression Via Induced Smoothing," *Computational Statistics & Data Analysis*, 56, 785–796. [3]
- Powell, J. L. (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155. [5]
- Qu, Z., and Yoon, J. (2015), "Nonparametric Estimation and Inference on Conditional Quantile Processes," *Journal of Econometrics*, 185, 1–19. [7]



- Rothwell, P. M. (1995), “Can Overall Results of Clinical Trials Be Applied to All Patients?” *The Lancet*, 345, 1616–1619. [1]
- Seo, M. H., and Linton, O. (2007), “A Smoothed Least Squares Estimator for Threshold Regression Models,” *Journal of Econometrics*, 141, 704–735. [2,3,4,10]
- Shen, J., and He, X. (2015), “Inference for Subgroup Analysis With a Structured Logistic-Normal Mixture Model,” *Journal of the American Statistical Association*, 110, 303–312. [1]
- Su, L., and Xu, P. (2019), “Common Threshold in Quantile Regressions With an Application to Pricing for Reputation,” *Econometric Reviews*, 38, 417–450. [1,2,3,4,5,7,10]
- Tesfaye, F., Nawi, N., Van Minh, H., Byass, P., Berhane, Y., Bonita, R., and Wall, S. (2007), “Association Between Body Mass Index and Blood Pressure Across Three Populations in Africa and Asia,” *Journal of Human Hypertension*, 21, 28–37. [1]
- VanderWeele, T. J., Luedtke, A. R., van der Laan, M. J., and Kessler, R. C. (2019), “Selecting Optimal Subgroups for Treatment Using Many Covariates,” *Epidemiology*, 30, 334–341. [1]
- Wang, H. J., Stefanski, L. A., and Zhu, Z. (2012), “Corrected-loss Estimation for Quantile Regression With Covariate Measurement Errors,” *Biometrika*, 99, 405–421. [3]
- Wang, H. J., and Zhu, Z. (2011), “Empirical Likelihood for Quantile Regression Models With Longitudinal Data,” *Journal of Statistical Planning and Inference*, 141, 1603–1615. [3,6]
- Wang, L., Zhou, Y., Song, R., and Sherwood, B. (2018), “Quantile-optimal Treatment Regimes,” *Journal of the American Statistical Association*, 113, 1243–1254. [3]
- Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., and Drazen, J. M. (2007), “Statistics in Medicine—Reporting of Subgroup Analyses in Clinical Trials,” *New England Journal of Medicine*, 357, 2189–2194. [1]
- Whang, Y.-J. (2006), “Smoothed Empirical Likelihood Methods for Quantile Regression Models,” *Econometric Theory*, 22, 173–205. [3]
- Yu, P. (2012), “Likelihood Estimation and Inference in Threshold Regression,” *Journal of Econometrics*, 167, 274–294. [4]
- (2013), “Integrated Quantile Threshold Regression and Distributional Threshold Effects,” Tech. rep. London: The Institute for Fiscal Studies Department of Economics, UCL [1,2,3,4]
- (2014), “The Bootstrap in Threshold Regression,” *Econometric Theory*, 676–714. [14]
- (2015), “Adaptive Estimation of the Threshold Point in Threshold Regression,” *Journal of Econometrics*, 189, 83–100. [4]
- Yu, P., and Fan, X. (2020), “Threshold Regression with a Threshold Boundary,” *Journal of Business & Economic Statistics*, 1–54. [1,2,3,5,12]
- Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012), “A Robust Method for Estimating Optimal Treatment Regimes,” *Biometrics*, 68, 1010–1018. [3]
- Zhang, L., Wang, H. J., and Zhu, Z. (2014), “Testing for Change Points Due to a Covariate Threshold in Quantile Regression,” *Statistica Sinica*, 24, 1859–1877. [1,2]
- (2017), “Composite Change Point Estimation for Bent Line Quantile Regression,” *Annals of the Institute of Statistical Mathematics*, 69, 145–168. [7]