

Model-based bootstrap for detection of regional quantile treatment effects

Yuan Sun & Xuming He

To cite this article: Yuan Sun & Xuming He (2021) Model-based bootstrap for detection of regional quantile treatment effects, Journal of Nonparametric Statistics, 33:2, 299-320, DOI: 10.1080/10485252.2021.1934465

To link to this article: <https://doi.org/10.1080/10485252.2021.1934465>



Published online: 07 Jun 2021.



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



Article views: 159



View related articles [↗](#)



View Crossmark data [↗](#)



Model-based bootstrap for detection of regional quantile treatment effects

Yuan Sun^{a,b} and Xuming He^b

^aSchool of Statistics, Capital University of Economics and Business, Beijing, People's Republic of China;

^bDepartment of Statistics, University of Michigan, Ann Arbor, MI, USA

ABSTRACT

Quantile treatment effects are often considered in a quantile regression framework to adjust for the effect of covariates. In this study, we focus on the problem of testing whether the treatment effect is significant at a set of quantile levels (e.g. lower quantiles). We propose a regional quantile regression rank test as a generalisation of the rank test at an individual quantile level. This test statistic allows us to detect the treatment effect for a prespecified quantile interval by integrating the regression rank scores over the region of interest. A new model-based bootstrap method is constructed to estimate the null distribution of the test statistic. A simulation study is conducted to demonstrate the validity and usefulness of the proposed test. We also demonstrate the use of the proposed method through an analysis of the 2016 US birth weight data and selected S&P 500 sector portfolio data.

ARTICLE HISTORY

Received 14 January 2020

Accepted 19 May 2021

KEYWORDS

Quantile regression; rank score; treatment effect; bootstrap

AMS SUBJECT CLASSIFICATIONS

62G09; 62G10

1. Introduction

The detection of treatment effects is an important problem in a wide variety of applications and has been studied by many researchers under different settings. In this paper, we focus on testing the hypothesis of no treatment effect against the alternative that the effect is significant for the upper or lower tail of the outcome distribution. There are at least two reasons why this particular class of alternatives is worth considering. Firstly, in some applications the evaluation of the treatment effect at one tail is of direct concern. For example, when financial institutions compare the risks among different portfolios, they need to focus on the lower tail of the return distribution so that they can be better prepared for the worst case scenarios. Secondly, there are cases where the treatment effect is minimal except at low or high quantile levels. In those cases any tests designed to detect mean or median differences may have poor power. For example, it is shown later in the paper as we analyse the 2016 US birth data that maternal hypertension is a risk factor for low birth weight, and the hypertension effect on birth weight is much more obvious at the lower tail of the birth weight distribution. In such cases, a statistical test aimed at detecting the effect in the lower tail is more useful than the conventional tests on the mean treatment effects.

Quantile regression, which was formally introduced by Koenker and Bassett (1978), is the basis of a natural solution for the above-stated problems. Similar to the least squares regression where the coefficients are estimated by minimising the squared loss of residuals, quantile regression employs an asymmetrically weighted absolute loss function, which enables us to model the conditional quantiles of the responses. Koenker and Machado (1999) discussed the likelihood ratio test, the Wald test and the rank test for inference in the quantile regression settings. A common approach is to choose a quantile level (say 0.9 quantile) and test whether the quantile regression coefficient for the treatment is significant. However, the test results may be sensitive to the choice of the individual quantile level and the test may lose power when the data are sparse around that quantile level of choice.

An improvement to individual quantile regression analysis is to consider the treatment effect over a quantile region. He, Hsu, and Hu (2010) proposed a covariate-adjusted expected shortfall test (COVES), which uses quantile regression to select the observations that lie in the upper or lower quantiles and compare the covariate-adjusted means of the selected observations. COVES has been shown to be quite powerful but the test is designed for randomised trials. Koenker (2010) suggested an alternative test using regression rank scores over a quantile region, following the quantile rank scores proposed in Gutenbrunner and Jurečková (1992) and Gutenbrunner, Jurečková, Koenker, and Portnoy (1993). The distribution of the test statistic under the null hypothesis is approximated by a chi-square distribution, but the chi-square approximation is valid only for i.i.d errors.

In this paper, we consider the regional quantile regression rank test in a more realistic case with the heterogeneous models. In this case the proposed test converges to a mixed chi-square distribution under the null hypothesis, but the mixture coefficients depend on the unknown conditional densities of the regression errors over a quantile region, whose estimates tend to be numerically unstable. An alternative way to carry out the inference is to use the bootstrap. However, commonly used bootstrap methods in regression are not directly applicable to this setting. We propose a new model-based bootstrap algorithm which aims to mimic the data generative procedure. This bootstrap algorithm enables us to generate the data under the null hypothesis and to consistently estimate the null distribution of the proposed test statistic.

Applicable beyond the proposed test, our model-based bootstrap is a general bootstrap algorithm for global quantile regression analysis and is useful for a variety of settings. For example, the proposed bootstrap can be used to build the confidence band of the quantile coefficients over certain region. It can also be used in other hypothesis testing problems because the model-based structure in our bootstrap provides the flexibility to generate data from a desired model under the null hypothesis.

The rest of the paper is arranged as follows. In Section 2, we briefly review the quantile regression model and the regression rank scores. In Section 3 we propose the regional quantile regression rank test and the model-based bootstrap algorithm along with its large sample properties. Results from simulations and real data applications are given in Sections 4 and 5, respectively. Section 6 concludes this paper with a brief discussion. We provide the proof of our main theorem in Appendix.

2. Review of quantile regression rank score

In this section we provide a brief review of quantile regression and regression rank scores. We refer to Koenker (2005) for a comprehensive introduction.

Consider a random sample of size n that follows the linear model of

$$y_i = x_i^T \beta(\tau) + e_{i,\tau}, \quad i = 1, 2, \dots, n, \quad (1)$$

where $x_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \in \mathbb{R}^p$ with $x_{i1} = 1$ and $\beta(\tau) = (\beta_1(\tau), \dots, \beta_p(\tau)) \in \mathbb{R}^p$. For identifiability, we require that at any quantile level $\tau \in (0, 1)$, the conditional τ th quantile of $e_{i,\tau}$ given x_i is 0. Letting $Q_y(\tau | x)$ be the τ th quantile of y given x , we can write (1) equivalently as $Q_{y_i}(\tau | x_i) = x_i^T \beta(\tau)$.

One can assume that Model (1) holds locally at a specific τ or globally at any $\tau \in (0, 1)$. In this paper, we work under the global quantile model since our goal is to detect the treatment effect over a region of τ . To ensure model validity, we require that $x_i^T \beta(\tau)$ be a monotone increasing function of τ given any x_i . One important observation is that under the global quantile model, the function $x_i^T \beta(\tau)$ for any $\tau \in (0, 1)$ determines the conditional distribution of y_i given x_i . Therefore at the population level, for given x_i , we can express y_i as

$$y_i = x_i^T \beta(u_i), \quad u_i \sim \text{Uniform}(0, 1). \quad (2)$$

In other words, we can view y_i as being generated from the quantile process $x_i^T \beta(u_i)$. This is an important observation for the development of our bootstrap method later in the paper.

The quantile regression estimates of $\beta(\tau)$ are obtained by

$$\hat{\beta}(\tau) = \underset{t \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(y_i - x_i^T t), \quad (3)$$

where $\rho_\tau(u) = u(\tau - \mathbb{I}(u < 0))$. This linear optimisation problem can be easily solved for all τ in $(0, 1)$ as discussed in Koenker (2005).

The optimisation problem in (3) can be transformed into a dual problem

$$\hat{a}(\tau) = \underset{a \in [0, 1]^n}{\operatorname{argmax}} \{a^T y | X^T a = (1 - \tau)X^T \mathbf{1}_n\}, \quad (4)$$

where $\hat{a}(\tau) = (\hat{a}_1(\tau), \dots, \hat{a}_n(\tau))$ is an n -dimensional vector. By the duality between (3) and (4), we have

$$\hat{a}_i(\tau) = \begin{cases} 1 & y_i > x_i^T \hat{\beta}(\tau) \\ \in (0, 1) & y_i = x_i^T \hat{\beta}(\tau) \\ 0 & y_i < x_i^T \hat{\beta}(\tau). \end{cases}$$

Thus $\hat{a}_i(\tau)$ is essentially an indicator whether the i th observation is above the fitted τ -quantile. Let $\hat{\tau}_i = \inf\{\tau : \hat{a}_i(\tau) > 0\}$, the i th observations should lie roughly at the $\hat{\tau}_i$ -quantile. Namely knowing $\hat{a}_i(\tau)$ for any $\tau \in (0, 1)$ is equivalent to knowing the relative position of the i th observation after the covariate is adjusted for. Gutenbrunner and Jurečková (1992) named $\hat{a}_i(\tau)$ as the regression rank score, because $\hat{a}_i(\tau)$ can be interpreted as a generalisation of ranks in the regression setting. Notice that $\hat{a}_i(\tau) - (1 - \tau)$ is also an approximation of the score function of quantile regression $\Psi_\tau(u) = \tau - I(u <$

0) evaluated at $x_i^T \hat{\beta}(\tau)$. The regression rank scores $\hat{a}_i(\tau)$ have been used to construct rank-based test in Koenker and Machado (1999) and Wang (2009) among others for the local quantile models.

In this paper we are interested in detecting the treatment effect over a quantile region, and we integrate the regression rank score $\hat{a}_i(\tau)$ against an non-decreasing score function $\varphi(\cdot)$. Namely, define $\hat{b} = (\hat{b}_1, \dots, \hat{b}_n)^T$ where

$$\hat{b}_i = \int_{\tau_a}^{\tau_b} \hat{a}_i(\tau) d\varphi(\tau), \quad (5)$$

on an interval $[\tau_a, \tau_b]$ that is specified by users. If an observation is above most quantiles over $[\tau_a, \tau_b]$ after the covariate adjustment, it is expected to have a relatively large $\hat{b}_i$.

The score function $\varphi(\cdot)$ provides flexibility in assigning different weights at different quantile levels. Two typical choices of $\varphi(\cdot)$ are:

- Wilcoxon score: $\varphi(t) = t$, which assigns weights evenly.
- Normal score: $\varphi(t) = \Phi^{-1}(t)$, which assigns more weights at upper and lower tails.

We use $\hat{b}_i$ to construct the regional quantile regression rank test statistic in the next section.

3. Proposed method and main results

3.1. Test statistic

In this section, we consider the following model

$$y_i = x_{i1}^T \beta_1(\tau) + x_{i2}^T \beta_2(\tau) + e_{i,\tau}, \quad i = 1, 2, \dots, n, \quad (6)$$

where x_{i1} is a p -dimensional vector, x_{i2} is a q -dimensional vector. The error $e_{i,\tau}$ are assumed to be independent but not necessarily identically distributed with the natural constraint that $Q_{e_{i,\tau}}(\tau | x_{i1}, x_{i2}) = 0$. We are interested in testing the hypothesis

$$H_0 : \beta_2(\tau) = 0 \forall \tau \in (0, 1) \quad \text{vs} \quad H_1 : \beta_2(\tau) \neq 0 \text{ for } \tau \in [\tau_a, \tau_b],$$

where $[\tau_a, \tau_b]$ is the user-specified subset of $(0, 1)$ and should be chosen to target the region of interest.

For convenience, write the design matrix of (6) as $X = [X_1, X_2] \in \mathbb{R}^{n \times (p+q)}$. Let $\hat{X}_2 = X_1(X_1^T X_1)^{-1} X_1^T X_2$, which is the projection of X_2 into the space spanned by the columns of X_1 . If we fit the quantile regression with only X_1 as the explanatory variable, $\hat{b}$ calculated under this null model represents the ranks after adjusting for X_1 . If the null hypothesis is true, $X_2 - \hat{X}_2$ is expected to be orthogonal to $\hat{b}$ asymptotically, since no variations in $\hat{b}$ can be further explained by $X_2 - \hat{X}_2$. To help understand this orthogonality, we recall that the residuals are orthogonal to the design matrix in the least squares regression. For the quantile regression $\hat{a}_i(\cdot)$ plays similar roles as the residuals and can be shown to be orthogonal to design variables used in the quantile regression. A rigorous argument follows from Lemma A.2 and Equation (A3) of Appendix.

Our test statistic is constructed based on the above observation. But instead of using the integral version of $\hat{b}$ defined in (5), we employ a grid of points in τ and replace $\hat{b}$ with a weighted sum. More precisely, consider a set of M ordered and evenly spaced grid points

$$S = (\tau_0, \tau_1, \dots, \tau_M), \quad (7)$$

where $[\tau_a, \tau_b]$ is a proper subset of $[\tau_0, \tau_M]$. With S and a differentiable score function $\varphi(\cdot)$ specified, we define

$$\tilde{b}_i = \sum_{\tau_m \in S \cap [\tau_a, \tau_b]} \hat{a}_i(\tau_m) \varphi'(\tau_m) (\tau_m - \tau_{m-1}), \quad (8)$$

where $\hat{a}_i$ is given in (4) but calculated under the null model.

The employment of these grid points in calculating $\tilde{b} = (\tilde{b}_1, \dots, \tilde{b}_n)$ is mainly to facilitate the bootstrap used later. Since $\hat{a}_i(\tau)$ is a piecewise linear function with $O(n \log n)$ break points (Portnoy 1991), $\hat{b}_i$ defined in (5) can be written as a sum of $O(n \log n)$ terms, and $\tilde{b}_i$ is an approximation of $\hat{b}_i$ with a sum of roughly M terms.

It is worth pointing out that only $\hat{a}_i$ evaluated at grid points within $[\tau_a, \tau_b]$ are used in calculating (8) to focus on our region of interest $[\tau_a, \tau_b]$. But the grid points need to be defined on $[\tau_0, \tau_M]$, which is strictly larger than $[\tau_a, \tau_b]$. To get reliable estimation of $\hat{a}(\cdot)$ at the end points using the bootstrap, $\beta(\cdot)$ should be estimated accurately over a slightly larger quantile region.

Now we define our proposed test statistic as

$$T_n = S_n^T Q_n^{-1} S_n, \quad (9)$$

where

$$S_n = n^{-1/2} (X_2 - \hat{X}_2)^T \tilde{b},$$

$$Q_n = n^{-1} (X_2 - \hat{X}_2)^T (X_2 - \hat{X}_2).$$

A larger value of T_n will be in favour of the alternative hypothesis. We shall show in Appendix that under some regularity assumptions, S_n converges to a zero mean normal distribution with variance Σ taking the form

$$\Sigma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{\tau_m \in S} c_{\tau_m} (x_{2i} - \hat{x}_{2i} - K_n^{\tau_m} x_{1i}) (x_{2i} - \hat{x}_{2i} - K_n^{\tau_m} x_{1i})^T,$$

where c_{τ_m} is a constant depending on $\varphi(\cdot)$ and $K_n^{\tau_m}$ is a matrix involves the conditional densities of y_i given x_i evaluated at τ_m -quantile. In principle, we could estimate the densities using kernel or spline methods. However, the results are often numerically unstable. Instead of estimating this covariance matrix to standardise the test statistic, we use the bootstrap as our preferred approach.

The matrix Q_n can be viewed as an approximate standardisation because it can be shown that Σ is approximately Q_n times a constant when the model is homogeneous. With the usage of Q_n , T_n behaves closer to a standard chi-square distribution asymptotically and the resulting test may have better power when the model is close to homogeneous. In theory,

many choices of Q_n would work, but the specific choice used here is consistent with the common choice for the quantile regression rank tests.

In the proposed test statistics T_n , the quantities M , $\varphi(\cdot)$ and $[\tau_a, \tau_b]$ need to be specified by the users. Therefore a discussion of how to choose them are in order.

- (1) Choice of M : The number of grid points M should be between the order of $n^{1/4}$ and of $n^{1/2}$ for our theory to work. In practice, the choice of M does not have notable influence on the result as long as M is not too extreme. For example, we find that 50 or 100 can be a suitable choice for M for a wide range of problems.
- (2) Choice of $\varphi(\cdot)$: The score function $\varphi(\cdot)$ may influence the power of the test. Koenker (2010) showed how the optimal score function can be selected under the simpler model with i.i.d errors, if the error density is known. Since the density is unknown in practise and moreover we allow heterogeneity, it is unrealistic to aim for an optimal score function. We compared the power of our test with the most commonly used Wilcoxon score and Normal score under a variety of settings by simulation and the differences are not major. We therefore recommend using the Wilcoxon score for simplicity.
- (3) Choice of $[\tau_a, \tau_b]$: The quantile region $[\tau_a, \tau_b]$ should be used to target the region of interest, such as the lower tail of birth-weight or the upper tail of the loss from an investment portfolio. In the typical quantile regression settings, we usually choose a value τ , whether a specific value of τ is better than another nearby value of τ is difficult to answer. The choice of one interval over another has the same question around it. But from the numerical results in Sections 4 and 5, we note that the power of our test is shown to be stable over a range of reasonable choices of $[\tau_a, \tau_b]$. In other words, choosing a specific value of τ in the analysis is associated with less robust analysis results than choosing an interval $[\tau_a, \tau_b]$.

3.2. Model-based bootstrap

In this subsection, we propose a model-based bootstrap method to approximate the distribution of T_n under the null hypothesis.

There are quite a few established bootstrap methods under the quantile regression setting. The paired bootstrap, the generalised bootstrap (Chatterjee and Bose 2005) and the wild bootstrap (Feng, He, and Hu 2011) are examples of those methods that have been implemented in the R package *quantreg*. However these methods cannot be directly applied here.

The paired bootstrap does not generate bootstrap samples under H_0 when the data are not from the null model. The same goes with the generalised bootstrap. One possible solution is to keep x_{i2} unchanged and sample (y_i^*, x_{i1}^*) with replacement from (y_i, x_{i1}) . The resulting bootstrap data set would be $(y_i^*, x_{i1}^*, x_{i2})$. But the correlation between x_{i1} and x_{i2} can not be preserved under such a subsampling scheme.

The wild bootstrap uses the coefficients $\hat{\beta}_1$ and residuals $\hat{e}_{i,\tau}$ obtained from τ th quantile regression fitted under H_0 . The bootstrap data set will be (y_i^*, x_{i1}, x_{i2}) where $y_i^* = x_{i1}^T \hat{\beta}(\tau) + w_i |\hat{e}_{i,\tau}|$, and w_i is generated independently from a specially designed distribution to make sure the bootstrap is consistent at the τ -quantile. The wild bootstrap is useful

for inference at a single quantile level. Since our test statistic consists of estimation from multiple quantiles, no weight distribution would work in this framework.

We propose a new bootstrap scheme that generates data globally under H_0 . The key idea is that as shown in (2), we can write our linear quantile regression model equivalently as $y_i = x_i^T \beta(u_i)$, where $u_i \sim \text{Uniform}(0, 1)$. We keep x_{i1} and x_{i2} fixed and generate bootstrap samples y_i^* from $x_{i1}^T \hat{\beta}_1(u_i)$. Namely, we view $x_{i1}^T \hat{\beta}_1(\cdot)$ as a quantile process for the bootstrap, where $\hat{\beta}_1(\cdot)$ is estimated under the null model.

Although the quantile function $x_i^T \beta(\cdot)$ is monotonously increasing at any x_i , the estimate $x^T \hat{\beta}(\cdot)$ is only guaranteed to be monotone at $x = \bar{x}$. Thus $x_{i1}^T \hat{\beta}_1(\cdot)$ may not be a valid quantile process. This is the reason why we introduce the set of grid points S defined in (7). Let $\tilde{\beta}_1(\tau)$ be the linear interpolation of $\{\hat{\beta}_1(\tau_m), m \in S\}$. Neocleous and Portnoy (2008) showed that when M increases in the order between $n^{1/4}$ and $n^{1/2}$, the probability that $x_{i1}^T \tilde{\beta}_1(\cdot)$ is monotonously increasing converges to 1. At the same time, $\tilde{\beta}_1(\cdot)$ is a good enough approximation to $\hat{\beta}_1(\cdot)$. Thus we propose to generate y_i^* from an asymptotically valid quantile process $x_{i1}^T \tilde{\beta}_1(u_i)$. The detailed algorithm of this model-based bootstrap method is given as follows:

- (1) Fit the linear quantile regression under H_0 and obtain the estimator $\hat{\beta}_1(\tau)$ for $\tau \in S \cap [\tau_0, \tau_M]$. Calculate T_n using Equation (9).
- (2) Let $\tilde{\beta}_1(\tau)$ be the linear interpolation of $\{\hat{\beta}_1(\tau_m), m \in S\}$. Namely $\tilde{\beta}_1(\tau) = \frac{\tau_{m+1}-\tau}{\tau_{m+1}-\tau_m} \hat{\beta}_1(\tau_m) + \frac{\tau-\tau_m}{\tau_{m+1}-\tau_m} \hat{\beta}_1(\tau_{m+1})$ when $\tau_m < \tau < \tau_{m+1}$, $m = 0, \dots, M-1$. Let $\tilde{\beta}_1(\tau) = \hat{\beta}_1(\tau_0)$ for $\tau < \tau_0$, and $\tilde{\beta}_1(\tau) = \hat{\beta}_1(\tau_M)$ for $\tau > \tau_M$.
- (3) For $i = 1, \dots, n$, generate $u_i \sim \text{Uniform}(0, 1)$ independently, and then construct a bootstrap sample (y_i^*, x_{i1}, x_{i2}) , where $y_i^* = x_{i1}^T \tilde{\beta}_1(u_i)$.
- (4) Calculate T_n^* from Equation (9) with the bootstrap sample.
- (5) Repeat Steps (3) and (4) for B times to get $\{T_{n1}^*, T_{n2}^*, \dots, T_{nB}^*\}$, where B is a pre-specified integer. The resulting p -value is calculated by $B^{-1} \sum_b \mathbb{I}(T_n > T_{nb}^*)$.

The model-based bootstrap can be used for other forms of test statistics. For example, the same bootstrap method can be used to approximate the distribution of $\sup_{\tau \in S \cap [\tau_a, \tau_b]} |\tilde{\beta}_2(\tau)|$ under H_0 , which may also be used as a test statistic for regional treatment effect detection. This supremum-based test is discussed in more detail in Section 4.

3.3. Asymptotic properties

Let f_i be the density of y_i given x_i . To study the asymptotic properties of the proposed test, we impose the following regularity conditions:

- (C1) $\max_i \|x_i\| \leq L$, where L is a positive constant and $\|\cdot\|$ denotes the L^2 norm.
- (C2) The densities f_i are bounded away from 0 and infinity at $x_i^T \beta(\tau)$ uniformly for i and $\tau \in [\tau_0, \tau_M]$, where $0 < \tau_0 < \tau_a$ and $\tau_b < \tau_M < 1$. Furthermore, $|f_i(c_1) - f_i(c_2)| = O(|c_1 - c_2|)$ uniformly in i as $|c_1 - c_2| \rightarrow 0$.
- (C3) The limits $Q := \lim_{n \rightarrow \infty} \frac{1}{n} \sum x_i x_i^T$ and $D_x^T := \lim_{n \rightarrow \infty} \frac{1}{n} \sum f_i(x_i^T \beta(\tau)) x_i x_i^T$ exist, and are positive definite at any $\tau \in [\tau_0, \tau_M]$.

- (C4) $\varphi(\cdot)$ is a nondecreasing differentiable function with bounded variation.
 (C5) $S = (\tau_0, \tau_1, \dots, \tau_M)$ is a set of ordered and evenly spaced grid points where $n^{1/4} \ll M \ll n^{1/2}$.

The regularity conditions are stated under fixed designs. When x_i is a random variable, all the calculations can be carried out conditioning on x_i . Replacing (C1) and (C3) with corresponding moment conditions, our results hold for random designs as well.

Condition (C1) assumes that the covariate space lies within a compact set. This assumption is necessitated by heterogeneity because if the quantile regression model is linear over an unbounded set of x at multiple τ values, the quantile functions $x^T \beta(\tau_1)$ and $x^T \beta(\tau_2)$ may cross unless they are vertical shifts. (C2) and (C3) are common sufficient conditions used to establish the uniform Bahadur representation for the quantile regression estimates. We restrict our attention to $[\tau_0, \tau_M]$, a subset of $(0, 1)$, because the asymptotic behaviour of $\hat{\beta}(\tau)$ as τ approaches 0 or 1 may be different, especially when $\beta(\tau)$ become unbounded at one or both tails. For our study, we work with a set slightly larger than our region of interest $[\tau_a, \tau_b]$, which can be chosen to be another compact subset of $(0, 1)$. On the other hand, in the cases where y_i takes values in a bounded interval, $\beta(\tau)$ stay bounded for $\tau \in [0, 1]$ by (C1). In this case, the proposed test can be applied even to all the quantile levels including $\tau_a = 0$ and $\tau_b = 1$.

Theorem 3.1: *With regularity conditions (C1)–(C4), we have under H_0 ,*

- (i) $T_n \Rightarrow \bar{\chi}^2$, a mixed chi-square distribution as a weighted sum of q chi-square variables of one degree of freedom.
 Further assume (C5) holds, then
- (ii) The bootstrap estimator $\hat{\beta}_1^*(\tau)$ is a consistent estimator of $\beta_1(\tau)$ uniformly for $\tau \in S \cap [\tau_a, \tau_b]$.
- (iii) Given the data, the conditional distribution of T_n^* converges to the same mixed chi-square distribution $\bar{\chi}^2$.

Theorem 3.1(i) shows that our test statistic converges to a mixed chi-square distribution under H_0 while Theorem 3.1(ii) and (iii) show that the conditional bootstrap distribution approximates to the same mixed chi-square distribution. Hence our model-based bootstrap is consistent for inference. The proof of these results relies on the empirical process theory and is given in Appendix.

4. Simulation

In this section, we present some empirical results of our proposed test by Monte Carlo simulations.

4.1. Settings

The number of replications in each simulation and the bootstrap replication size are both set to 1000 throughout this section. We first generated our data from the following model

that was considered in He et al. (2010),

$$y_i = 5 + x_{i1} + x_{i2} + (1 + \gamma \mathbb{I}(e_i > 0) \mathbb{I}(d_i = 0))e_i, \quad i = 1, \dots, m + n, \quad (10)$$

where the treatment indicator $d_i = 1$ for $i = 1, \dots, m$ and $d_i = 0$ for $i = m + 1, \dots, m + n$. Let $\gamma = 0$ under H_0 and $\gamma = 1.35$ under H_1 . We consider testing whether the coefficient of the treatment indicator γ is zero. By design, the treatment effect only exists in the upper tail under the alternative. We considered the following three different settings based on Model (10):

- (i) $x_{i1} \sim \text{Uniform}(5, 12)$, $x_{i2} \sim N(8, 8)$ and $e_i \sim N(0, 5)$, and they are mutually independent. This represents a randomised trial with *i.i.d* errors.
- (ii) $x_{i1} \sim \text{Uniform}(5, 12)$ when $d_i = 1$, but $x_{i1} \sim \text{Uniform}(5, 20)$ when $d_i = 0$. In addition, $x_{i2} \sim N(8, 8)$ and $e_i \sim N(0, x_{i1})$ are independently generated. This represents a non-randomised trial with heterogeneous errors.
- (iii) $x_{i1} \sim \text{Uniform}(5, 12)$ when $d_i = 1$; otherwise x_{i1} is generated from the t distribution truncated to $[0, 250]$ with 2 degrees of freedom and non-centrality parameter equal to 15. The variables $\{x_{i2}\}$ and $\{e_i\}$ are generated from the same distributions as (ii). Compared to (ii), x_{i1} is generated from a distribution with heavier tails.

Under these settings, $\{x_{i2}\}$ is generated from a normal distribution, which violates (C1) that $\{x_i\}$ should lie in a compact set. However, since the coefficient of x_{i2} is a constant of τ in these settings, we still have valid quantile functions even when the range of x_{i2} extends to the whole line, so our theory applies to the model with trivial modifications.

In addition, we also evaluated the performance of the proposed method when the effect of multiple covariates are simultaneously tested in the following model:

(iv)

$$y_i = \beta_0(u_i) + x_{i1}\beta_1(u_i) + x_{i2}\beta_2(u_i) + x_{i3}\beta_3(u_i), \quad i = 1, \dots, n,$$

where $u_i \sim \text{Uniform}(0, 1)$, $x_{i2} \sim \text{Uniform}(0, 2)$, $x_{i3} \sim \text{Uniform}(0, 2)$, and $x_{i1} \sim \text{Uniform}(1, 3)$ when $x_{i2} < 1$ but $x_{i1} \sim \text{Uniform}(0, 2)$ when $x_{i2} \geq 1$. Furthermore, let $\beta_0(\tau) = \Phi^{-1}(\tau)$, $\beta_1(\tau) = \tau^2$. Under H_0 , we use $\beta_2(\tau) = \beta_3(\tau) = 0$. Under H_1 , we use $\beta_2(\tau) = \frac{\exp(15(\tau-0.5))}{1+\exp(15(\tau-0.5))}$ and $\beta_3(\tau) = \frac{\exp(10(\tau-0.5))}{1+\exp(10(\tau-0.5))}$. As shown in Figure 1, the effect of x_{i1} and x_{i2} are larger at the upper tail under the alternative by design.

We consider the problem of testing the null hypothesis $H_0: \beta_2(\tau) = \beta_3(\tau) = 0, \forall \tau \in (0, 1)$, but the test focus on upper quantiles.

We first compared the proposed regional quantile regression rank (RQRR) test with the quantile regression rank (QRR) test that focuses on one fixed quantile proposed by Koenker and Machado (1999) to see if we can benefit from considering a quantile region. To show the necessity of the proposed bootstrap method, we also considered the proposed RQRR test statistic with the critical value approximated by the chi-square distribution based on the working assumption of *i.i.d* errors.

When $q = 1$, we further compared the performance of our test to other three methods that focus on the overall treatment effect: the COVES test proposed by He et al. (2010);

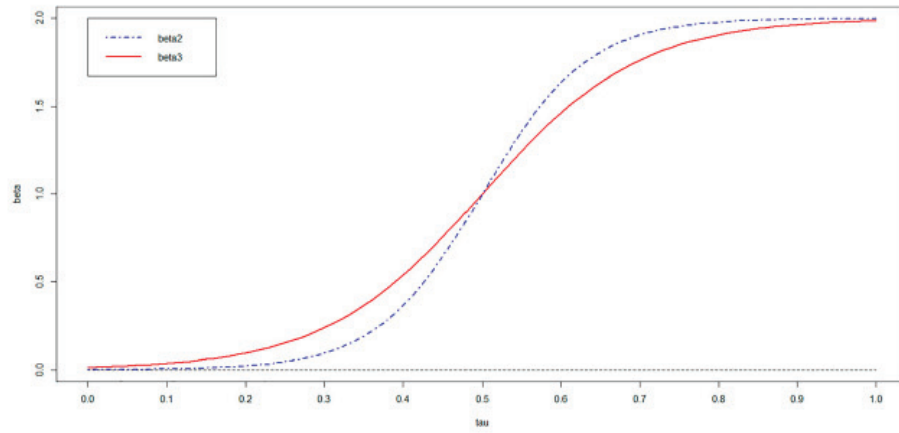


Figure 1. Curves of quantile coefficients of Model (iv) in simulation under the alternative.

the test based on simultaneous confidence band; the supremum-based test. The latter two methods are described as follows.

To build simultaneous confidence bands, we use a method similar to what is considered in Chernozhukov and Fernández-Val (2005). A level $1 - \alpha$ confidence band of $\beta_2(\cdot)$ over $[\tau_a, \tau_b]$ can be built based on the statistic

$$T_n^{\text{sup}} = \sup_{\tau \in [\tau_a, \tau_b]} |\sqrt{n} \tilde{\beta}_2(\tau)|,$$

where $\tilde{\beta}_2(\tau)$ is the linear interpolation of the coefficient estimate $\hat{\beta}_2(\tau)$. The distribution of $\tilde{\beta}_2(\tau)$ is approximated by the m out of n bootstrap, where $m = 20 + n^{1/2}$. The null hypothesis is rejected if 0 is contained nowhere in the confidence band.

To carry out the supremum-based test, we use the model-based bootstrap scheme introduced in Section 3.2 to generate the bootstrap sample (y_i^*, x_{1i}, x_{2i}) where $y_i^* = x_{1i}^T \tilde{\beta}_1(u_i)$, with $\tilde{\beta}_1(\cdot)$ estimated under the restricted model. The bootstrap test statistic $T_n^{\text{sup},*} = \sup_{\tau \in [\tau_a, \tau_b]} |\sqrt{n} \tilde{\beta}_2^*(\tau)|$ can then be obtained from the bootstrap sample. The null distribution of T_n^{sup} is approximated by the empirical distribution of $T_n^{\text{sup},*}$.

Both the test based on simultaneous confidence bands and the supremum-based test utilise the test statistics T_n^{sup} . The difference is that the bootstrap is conducted under the full model to build the confidence band while the bootstrap is conducted under the null hypothesis for the supremum-based test. Also notice that the difference between the supremum-based test and the RQRR test lies in the test statistics. Thus comparing these two tests is basically comparing the performance of a supremum-based statistic versus a rank-based statistic under our settings.

4.2. Results

We first set the quantile region to be $[0.7, 0.99]$ and $[0.85, 0.99]$ to compare the performance of the proposed RQRR test with the QRR test at one quantile level, the RQRR test with the chi-square approximation, and the COVES test. The results are summarised in Table 1.

For the randomised trial we considered in Model (i), all the tests have reasonable type I error rates. For Model (ii), both the COVES and the RQRR with chi-square approximations are not valid theoretically. According to our simulation results, COVES fails to control

Table 1. Comparison of the empirical type I error rate and the power out of 1000 simulation samples.

	Model (i)				Model (ii)			
	$m = n = 50$		$m = n = 100$		$m = n = 50$		$m = n = 100$	
	α level	power	α level	power	α level	power	α level	power
QRR(0.70)	0.050	0.327	0.047	0.586	0.041	0.273	0.052	0.524
QRR(0.80)	0.050	0.567	0.046	0.889	0.047	0.493	0.037	0.797
QRR(0.85)	0.041	0.682	0.046	0.953	0.041	0.576	0.049	0.889
QRR(0.90)	0.035	0.723	0.042	0.983	0.038	0.622	0.043	0.926
QRR(0.95)	0.036	0.431	0.025	0.974	0.024	0.488	0.034	0.891
COVES	0.066	0.909	0.052	0.998	0.070	0.917	0.088	0.999
RQRR(0.70,0.99)	0.043	0.712	0.048	0.960	0.045	0.628	0.046	0.917
RQRR _b (0.70,0.99)	0.047	0.707	0.049	0.956	0.047	0.627	0.046	0.908
RQRR(0.85,0.99)	0.041	0.816	0.044	0.995	0.041	0.702	0.046	0.970
RQRR _b (0.85,0.99)	0.045	0.834	0.046	0.994	0.048	0.725	0.051	0.972

	Model (iii)				Model (iv)			
	$m = n = 100$		$m = n = 300$		$n = 100$		$n = 200$	
	α level	power	α level	power	α level	power	α level	power
QRR(0.70)	0.064	0.162	0.048	0.338	0.045	0.356	0.048	0.657
QRR(0.80)	0.058	0.290	0.065	0.571	0.042	0.494	0.040	0.844
QRR(0.85)	0.062	0.358	0.064	0.675	0.042	0.532	0.043	0.892
QRR(0.90)	0.053	0.428	0.074	0.752	0.041	0.537	0.040	0.897
QRR(0.95)	0.046	0.438	0.057	0.789	0.034	0.384	0.037	0.784
COVES	0.522	0.943	0.915	1.000	NA	NA	NA	NA
RQRR(0.70,0.99)	0.115	0.557	0.148	0.868	0.042	0.619	0.047	0.930
RQRR _b (0.70,0.99)	0.056	0.381	0.058	0.724	0.048	0.629	0.047	0.930
RQRR(0.85,0.99)	0.102	0.678	0.128	0.930	0.049	0.625	0.040	0.944
RQRR _b (0.85,0.99)	0.061	0.551	0.063	0.853	0.052	0.657	0.040	0.947

Note: In the table, QRR(τ) stands for the quantile regression rank test conducted at the τ th quantile proposed in Koenker and Machado (1999); RQRR(τ_a, τ_b) stands for the regional quantile regression rank test with chi-square approximations at the quantile region $[\tau_a, \tau_b]$, while RQRR_b(τ_a, τ_b) stands for the proposed regional quantile regression rank test with the model-based bootstrap. For COVES, the cutoff quantile level is set to be 0.75 as in He et al. (2010).

the type I errors, and the RQRR test with chi-square approximations is acceptable. This is actually consistent with our knowledge that the RQRR with chi-square approximations is reasonably robust under heterogeneity (Kocherginsky, He, and Mu 2005). For the more extreme example where x_{i1} has a heavy right tail in Model (iii), however, it is obvious that both COVES and the RQRR with chi-square approximations are not valid anymore, while our proposed RQRR has empirical type I errors close to the nominal level. The results from Model (iv) show that our proposed test also has satisfactory performance when testing the effect of multiple continuous covariates simultaneously. Overall our proposed test works more broadly than the COVES and the RQRR with the chi-square approximation. In particular, the proposed test remains valid under heterogeneous cases where the other two tests may fail.

Table 1 also illustrates the advantages of the proposed RQRR over the QRR test at one quantile level in terms of power stability. Firstly, we observe that the empirical power of the QRR test heavily depends on the choice of τ . Its power tends to increase as τ increases to some value because the magnitude of treatment effect also increases. But the power will decrease if we further increase τ due to the inflation in variance. On the other hand, it is quite obvious from our results that the proposed RQRR test is less sensitive to the choice of the quantile region. Secondly, the proposed RQRR test with a reasonably-chosen quantile

Table 2. Comparison of the empirical type I error rate and power out of 1000 simulation samples for the tests based on simultaneous confidence bands (CF), the supremum-based test (Max), and the RQRR test over $[\tau_a, \tau_b]$.

	Model (ii)			
	$m = n = 50$		$m = n = 100$	
	α	power	α	power
RQRR _b (0.70,0.95)	0.047	0.571	0.042	0.895
RQRR _b (0.85,0.95)	0.043	0.701	0.051	0.955
RQRR _b (0.70,0.99)	0.047	0.627	0.046	0.908
RQRR _b (0.85,0.99)	0.048	0.725	0.051	0.972
CF(0.70,0.95)	0.015	0.399	0.018	0.767
CF(0.85,0.95)	0.021	0.448	0.021	0.785
CF(0.70,0.99)	0.071	0.516	0.083	0.810
CF(0.85,0.99)	0.074	0.530	0.084	0.811
Max(0.70,0.95)	0.052	0.605	0.045	0.924
Max(0.85,0.95)	0.057	0.607	0.049	0.923
Max(0.70,0.99)	0.090	0.546	0.147	0.871
Max(0.85,0.99)	0.092	0.547	0.150	0.871

interval is more powerful than the QRR test at many individual quantile levels. For example, the power for the proposed RQRR test with quantile region $[0.85, 0.99]$ is higher than the QRR test with $\tau = 0.90$ for all the settings we considered. Therefore we can benefit from utilising the extra information provided over a quantile region to achieve more stable statistical power.

We then set the quantile region to be $[0.7, 0.95]$, $[0.85, 0.95]$, $[0.7, 0.99]$ and $[0.85, 0.99]$ to compare the performance of the proposed RQRR test with the methods based on simultaneous confidence bands and the supreme-based test. For this comparison, we only present the results under Model (ii) in Table 2. The results under Model (i) and (iii) tell a similar story.

From Table 2, we notice that the methods based on simultaneous confidence bands and the supremum-based test do not control the type I error well when we set the upper quantile level to be 0.99. This is because the estimation of the coefficients are unreliable when τ is close to one for data with moderate sample sizes. The RQRR test can be roughly seen as analysing the average treatment effect over the quantile region, so it is able to handle relatively extreme tails better.

5. Real data

5.1. Birth weight data

In this section, we illustrate the power of the proposed RQRR test with the 2016 US birth weight data. Because the size of the full data is large, we are able to conduct the proposed RQRR test and the QRR test at one quantile level with sub-samples of the full data set and compare their number of rejections.

The 2016 US birth weight data set is produced by the National Center for Health Statistics and is available to the public online.¹ The data set contains the infant and maternal health characteristics along with paternal demographic information of the births occurred in the US during 2016. In particular, we restrict our focus to 32,169 white mothers whose

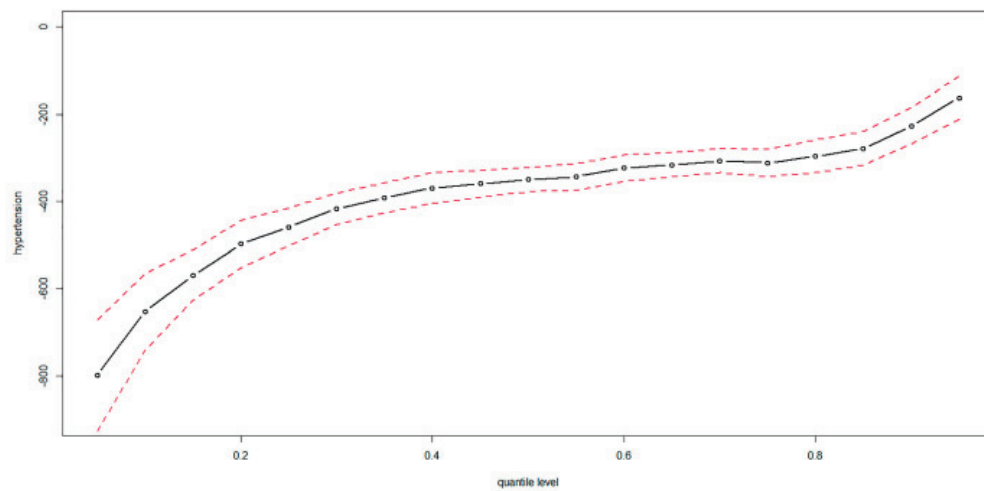


Figure 2. 95% pointwise confidence band of the hypertension effect.

ages are between 36 and 40 and we aim to study the relationship between birth weight and the maternal history of hypertension. Besides the indicator for maternal hypertension, mothers' education level, mothers' weight before delivery and indicator for smoking during pregnancy are included as confounding variables. Notice that these variables were also considered in the birthweight data collected at Baystate Medical Center during 1986 (Hosmer and Lemeshow 2000).

We first fit linear quantile regressions with the full data set at different τ . From Figure 2, we can see that the coefficient of hypertension is significantly less than zero at all the quantile levels and the hypertension effect decreases in magnitude as the quantile level increases. Namely hypertension has a negative effect on birth weight and its effect is more severe at the lower tail. Given the size of the data, we view the full data estimates as good proxies to the true parameters.

In the next step, we subtract the median hypertension effect estimated with the full data from the birth weight to check whether the quantile coefficient varies with τ . To compare the performance of the proposed RQRR test with the QRR test at on quantile level when we have limited sample sizes, we sub-sample ($n = 200$ to 800) from the full data and compare the number of rejections out of 500 sub-sampled data sets.

The results are summarised in Table 3, which is consistent to what we observe in the simulations. The power of both the proposed RQRR test and the QRR test depends on the choice of quantile interval/level. Though the treatment of hypertension becomes more significant at lower tails, the QRR test suffers from low power if we choose τ to be as small as 0.01 due to higher variances in those quantile estimates. For the RQRR test, the intervals $[0.01, 0.1]$ and $[0.01, 0.15]$ are better choices compared to the interval $[0.05, 0.25]$. Comparing the RQRR test with the QRR test, the former is less sensitive to the choice of the quantile interval/level. Even when the quantile interval is chosen to be $[0.05, 0.25]$, the power of the RQRR test is not much worse. When both the quantile interval and the quantile level are reasonably chosen, the RQRR test tends to perform better than the QRR test in general. The least squares regression is included in the comparison, and its power is clearly lower than the RQRR test, because it aims to detect the difference in the mean, which is less obvious than in the lower tail of the birthweight distribution.

Table 3. Number of rejections out of 500 sub-sampled birthweight data sets.

Test	Counts of rejection at 0.05 level		
	$n = 200$	$n = 400$	$n = 800$
RQRR _b (0.01,0.10)	103	150	223
RQRR _b (0.01,0.15)	94	141	216
RQRR _b (0.05,0.25)	72	92	171
QRR(0.01)	44	54	76
QRR(0.05)	81	120	194
QRR(0.10)	69	108	185
QRR(0.20)	40	65	104
Least squares	50	63	69

5.2. Value-at-risk for S&P 500 sectors

In this subsection, we examine some of the market factors that impact the value-at-risk (VaR) of the S&P 500 sectors with our proposed method. An $100\alpha\%$ VaR of a portfolio is given by the α -quantile of its return distribution, and is an important risk measure with small values of α .

Fama and French (1993) proposed the following three-factor model to describe the relationship between portfolio return and market factors:

$$R_t - R_{Ft} = a + b(R_{Mt} - R_{Ft}) + sSMB_t + hHML_t + e_t, \quad (11)$$

where R_t is the stock/portfolio return, R_{Ft} is the risk-free return, R_{Mt} is the return on the value-weighted market portfolio, SMB_t is the difference between the returns on diversified portfolios of small and big stocks, HML_t is the difference between the returns on diversified portfolios of high and low book to market ratio stocks, and e_t represents idiosyncratic error. Fama and French (2015) proposed a five-factor model and added two extra factors RMW_t and CMA_t to (11), where RMW_t is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMA_t is the difference between the returns on diversified portfolios of the stocks of conservatively invested and aggressive invested firms.

We wish to test if the extra two factors added in Fama and French (2015) are useful in explaining the VaR of different S&P 500 sectors. We collected the S&P 500 indices of the energy, information technology and health care sectors between January 2, 2014 to December 29, 2018 from Yahoo Finance, which has a total of 1007 trading days. The return R_t for a given portfolio is calculated by $100 \log \frac{x_t}{x_{t-1}}$, where x_t be the value of the portfolio (or index) at time t . The market factors are obtained from the data library of Kenneth R. French's website.²

To be more specific, we consider the following Fama-French five-factor model in the quantile regression framework:

$$\begin{aligned} R_t - R_{Ft} = & a(\tau) + b(\tau)(R_{Mt} - R_{Ft}) + s(\tau)SMB_t + h(\tau)HML_t \\ & + r(\tau)RMW_t + c(\tau)CMA_t + e_{t,\tau}. \end{aligned} \quad (12)$$

We assume that $e_{t,\tau}$ are independent over time given the explanatory variables. This assumption is reasonable as the autocorrelation in R_t can be fully explained by the market

Table 4. *P*-values from the of the proposed RQRR test at a quantile region and the QRR test at individual quantile levels for the significance of two additional factors in the Fama-French model.

	Energy	Information	Health
RQRR _b (0.01,0.05)	0.014	0.001	0.000
QRR(0.01)	0.102	0.014	0.295
QRR(0.05)	0.003	0.000	0.000

Note: The columns represent S&P 500 sector portfolios.

factors, analogous to the common assumptions in autoregressive models. We are interested in testing

$$H_0 : r(\tau) = c(\tau) = 0 \quad \forall \tau \in (0, 1) \quad \text{vs}$$

$$H_1 : r(\tau) \neq 0 \quad \text{or} \quad c(\tau) \neq 0 \quad \text{for some } \tau \in [0.01, 0.05].$$

We focus on the lower tail of $\tau \in [0.01, 0.05]$, and use the proposed RQRR test at the quantile interval $[0.01, 0.05]$ as well as the QRR test at $\tau = 0.01$ and 0.05 ; see the results summarised in Table 4.

The results from Table 4 show that the QRR test at $\tau = 0.01$ often lacks the power to detect the effect of the two additional factors in the five-factor model, but the RQRR test on the quantile interval $[0.01, 0.05]$ confirms the statistical significance, at the level of 0.05, of the five-factor model at the lower tail for each of the S&P 500 sectors considered here. This example further validates the power and stability of the proposed RQRR test discussed earlier.

6. Conclusion

We have proposed a regional quantile regression rank test that is designed to detect the treatment effect in the upper or lower tails of the outcome distribution. Our test is valid under heterogeneous models, and outperforms existing methods in a variety of settings. One innovation of the work is the model-based bootstrap algorithm proposed for the implementation for the test. The main idea is to sample from the conditional quantile functions estimated under the null hypothesis but over a slightly wider region than the target quantile region in the alternative hypothesis.

Because our model-based bootstrap is able to approximate the data generative procedure consistently and uniformly over a given region of τ , it can be used to approximate the null distribution for other test statistics. For example, in the simulation we combined the model-based bootstrap with an alternative supremum-type statistics. There are scenarios when such test statistics are preferred. For example, if the treatment effect is negative at lower τ but positive at upper τ , the power of the proposed rank test may suffer when it targets the average over this region, but the supremum-based test may work well. We hope that the model-based bootstrap method becomes a valuable tool in a variety of hypothesis testing problems in quantile regression.

Notes

1. The data set is available for download at https://www.cdc.gov/nchs/data_access/vitalstats_online.htm
2. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Acknowledgments

The authors thank the two reviewers, the Associate Editor and the Editor in Chief for their constructive comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research was partially supported by the National Science Foundation (NSF) under Grants DMS-1607840, DMS-1914496 and DMS-1951980.

References

- Chatterjee, S., and Bose, A. (2005), 'Generalized Bootstrap for Estimating Equations', *The Annals of Statistics*, 33, 414–436.
- Chernozhukov, V., and Fernández-Val, I. (2005), 'Subsampling Inference on Quantile Regression Processes', *The Indian Journal of Statistics*, 67, 253–276.
- Fama, E., and French, K. (1993), 'Common Risk Factors in the Returns on Stocks and Bonds', *Journal of Financial Economics*, 33, 3–56.
- Fama, E., and French, K. (2015), 'A Five-Factor Asset Pricing Model', *Journal of Financial Economics*, 116, 1–22.
- Feng, X., He, X., and Hu, J. (2011), 'Wild Bootstrap for Quantile Regression', *Biometrika*, 98, 995–999.
- Gutenbrunner, C., and Jurečková, J. (1992), 'Regression Rank Scores and Regression Quantiles', *The Annals of Statistics*, 20, 305–330.
- Gutenbrunner, C., Jurečková, J., Koenker, R., and Portnoy, S. (1993), 'Tests of Linear Hypotheses Based on Regression Rank Scores', *Journal of Nonparametric Statistics*, 2, 307–331.
- He, X., Hsu, Y.H., and Hu, M. (2010), 'Detection of Treatment Effects by Covariate-Adjusted Expected Shortfall', *The Annals of Applied Statistics*, 4, 2114–2125.
- Hosmer, D., and Lemeshow, S. (2000), *Applied Logistic Regression* (2nd ed.), New York: Wiley-Interscience Publication.
- Kocherginsky, M., He, X., and Mu, Y. (2005), 'Practical Confidence Intervals for Regression Quantiles', *Journal of Computational and Graphical Statistics*, 14, 41–55.
- Koenker, R. (2005), *Quantile Regression*, New York: Cambridge University Press.
- Koenker, R. (2010), 'Rank Tests for Heterogeneous Treatment Effects with Covariates', *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jureckova*, 7, 134–142.
- Koenker, R., and Bassett, G. (1978), 'Regression Quantiles', *Econometrica*, 46, 33–50.
- Koenker, R., and Machado, J.A.F. (1999), 'Goodness of Fit and Related Inference Processes for Quantile Regression', *Journal of the American Statistical Association*, 94, 1296–1310.
- Neocleous, T., and Portnoy, S. (2008), 'On Monotonicity of Regression Quantile Functions', *Statistics & Probability Letters*, 78, 1226–1229.
- Portnoy, S. (1991), 'Asymptotic Behavior of the Number of Regression Quantile Breakpoints', *SIAM Journal on Scientific and Statistical Computing*, 12, 867–883.
- van der Vaart, A.W., and Wellner, J.A. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer.

Wang, H. (2009), 'Inference on Quantile Regression for Heteroscedastic Mixed Models', *Statistica Sinica*, 19, 1247–1261.

Appendix. Proof of Theorem 3.1

We first study the limiting distribution of the test statistic T_n as shown in Theorem 3.1(i). Let $Z_n = \sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$. Let d_i be any l -dimensional vector that is uniformly bounded and write $D_{nd}^T = \frac{1}{n} \sum f_i(x_i^T \beta(\tau)) d_i x_i^T$. We plug in $d_i = x_i$ and $d_i = \hat{x}_{i2} - x_{i2}$ in the latter part of the proof. Notice that by conditions (C1) and (C3), the limit of D_{nd}^T exists and is positive definite when $d_i = x_i$ and $d_i = \hat{x}_{i2} - x_{i2}$.

Let $\hat{G}_n^d(t) = n^{-1/2} \sum d_i \mathbb{I}(y_i \leq x_i^T t)$ and $G_n^d(t) = n^{-1/2} \sum d_i F_i(x_i^T t)$. Lemma A.1 below shows that $G_n^d(t)$ is a good approximation of $\hat{G}_n^d(t)$ using results from the empirical process theory (van der Vaart and Wellner 1996). Furthermore define $\hat{W}_n^d = n^{-1/2} \sum d_i (\hat{a}_i(\tau) - (1 - \tau))$ and $W_n^d = n^{-1/2} \sum d_i (\tilde{a}_i(\tau) - (1 - \tau))$ where $\tilde{a}_i(\tau) = \mathbb{I}(y_i \geq x_i^T \beta(\tau))$. Recall $\hat{a}_i(\tau) \approx \mathbb{I}(y_i \geq x_i^T \hat{\beta}(\tau))$ is the main component of our test statistic and $\tilde{a}_i(\tau)$ follows i.i.d binomial distributions which is easy to analyse. Lemma A.2 establishes the relationship between $\hat{W}_n^d$ and W_n^d . Theorem 3.1(i) then follows from Lemmas A.1 and A.2.

Lemma A.1: $\sup_{\tau_0 \leq \tau \leq \tau_M} \|\hat{G}_n^d(\hat{\beta}(\tau)) - \hat{G}_n^d(\beta(\tau)) - G_n^d(\hat{\beta}(\tau)) + G_n^d(\beta(\tau))\| = o_p(1)$.

Proof: For any l -dimensional vector v , define the class of function $\mathcal{G}$ over a compact set $\mathcal{T} \in \mathbb{R}^{p+q}$ as

$$\mathcal{G} = \left\{ v^T d_i \mathbb{I}(y_i \leq x_i^T t), t \in \mathcal{T} \right\}.$$

Note that $\mathcal{G}$ is a VC subgraph class and $E(g^2)$ is bounded for any $g \in \mathcal{G}$.

Thus $v^T (\hat{G}_n^d(t) - G_n^d(t))$ is stochastically equicontinuous over $\mathcal{T}$ with semi-metric

$$\rho(t_1, t_2) = \left\{ E \left(v^T d_i \mathbb{I}(y_i \leq x_i^T t_1) - v^T d_i \mathbb{I}(y_i \leq x_i^T t_2) \right)^2 \right\}^{1/2}.$$

Since

$$\begin{aligned} \rho(t_1, t_2)^2 &\leq (v^T d_i)^2 E \left(\mathbb{I}(x_i^T t_2 \leq y_i \leq x_i^T t_1) + \mathbb{I}(x_i^T t_1 \leq y_i \leq x_i^T t_2) \right) \\ &= 2(v^T d_i)^2 O(\|t_1 - t_2\|) \\ &= O(\|t_1 - t_2\|), \end{aligned}$$

and $\hat{\beta}(\tau)$ is a consistent estimator of $\beta(\tau)$ uniformly for $\tau \in [\tau_0, \tau_M]$ (This result can be proved using similar and easier argument to that used in the proof of Theorem 3.1(ii).), we have

$$\sup_{\tau_0 \leq \tau \leq \tau_M} \left| v^T \left(\hat{G}_n^d(\hat{\beta}(\tau)) - G_n^d(\hat{\beta}(\tau)) \right) - v^T \left(\hat{G}_n^d(\beta(\tau)) + G_n^d(\beta(\tau)) \right) \right| = o_p(1),$$

by the definition of equicontinuity. The lemma is hence proved since v is arbitrary. ■

Lemma A.2: $\|\hat{W}_n^d - W_n^d + D_{nd}^T Z_n\| = O(\sqrt{n} \|\hat{\beta}(\tau) - \beta(\tau)\|^2) + o_p(1)$ uniformly over $\tau \in [\tau_0, \tau_M]$.

Proof: By simple manipulation, we can write

$$\hat{W}_n^d = W_n^d - D_{nd}^T Z_n + R_1 - R_2 - R_3,$$

where

$$\begin{aligned} R_1 &= n^{-1/2} \sum d_i \mathbb{I}(y_i = x_i^T \hat{\beta}(\tau)) \hat{a}_i(\tau), \\ R_2 &= \hat{G}_n^d(\hat{\beta}(\tau)) - \hat{G}_n^d(\beta(\tau)) - G_n^d(\hat{\beta}(\tau)) + G_n^d(\beta(\tau)), \end{aligned}$$

$$R_3 = G_n^d(\hat{\beta}(\tau)) - G_n^d(\beta(\tau)) - D_{nd}^\tau Z_n.$$

Because y_i is continuous, $\sum \mathbb{I}(y_i = x_i^T \hat{\beta}(\tau)) = p + q$ almost surely for any τ . Since $|\hat{a}_i(\tau)| \leq 1$ and d_i bounded, $R_1 = O(n^{-1/2})$ uniformly.

By Lemma A.1, R_2 is uniformly $o_p(1)$.

Now consider R_3 . By Taylor expansion,

$$\begin{aligned} & \left\| G_n^d(\beta(\tau) + n^{-1/2} \Delta) - G_n^d(\beta(\tau)) - D_{nd}^\tau \Delta \right\| \\ &= \left\| \frac{1}{\sqrt{n}} \sum d_i n^{-1/2} (x_i^T \Delta) \int_0^1 \left(f_i(x_i^T \beta(\tau) + n^{-1/2} (x_i^T \Delta) s) - f_i(x_i^T \beta(\tau)) \right) ds \right\| \\ &= \left\| \frac{1}{\sqrt{n}} \sum d_i n^{-1/2} (x_i^T \Delta) \int_0^1 O(n^{-1/2} (x_i^T \Delta) s) ds \right\| \\ &= \left\| \frac{1}{\sqrt{n}} \sum d_i O(n^{-1} (x_i^T \Delta)^2) \right\| \\ &= O(n^{-1/2} \|\Delta\|^2). \end{aligned}$$

We have the desired result letting $\Delta = \sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$. ■

Proof of Theorem 3.1(i): Set $d_i = x_i$. By the constraints in (4), $\hat{W}_n^x = n^{-1/2} \sum x_i(\hat{a}_i(\tau) - (1 - \tau)) = 0$. Thus from Lemma A.2, we have

$$D_{nx}^\tau Z_n = W_n^x + O(\sqrt{n} \|\hat{\beta}(\tau) - \beta(\tau)\|^2) + o_p(1).$$

Namely,

$$Z_n(1 + o_p(1)) = (D_{nx}^\tau)^{-1} W_n^x + o_p(1).$$

By similar argument as in Lemma A.1, $\mathcal{W} = \{x_i(\tilde{a}_i(\tau) - (1 - \tau)), \tau \in [\tau_0, \tau_M]\}$ is a VC subgraph class with bounded envelope. Then we have $Z_n = O_p(1)$ since the limit of D_{nx}^τ is positive definite by (C3). Therefore we have the uniform Bahadur representation

$$Z_n = (D_{nx}^\tau)^{-1} W_n^x + o_p(1). \quad (\text{A1})$$

By Lemma A.2 and (A1),

$$\hat{W}_n^d = W_n^d - D_{nd}^\tau (D_{nx}^\tau)^{-1} W_n^x + o_p(1). \quad (\text{A2})$$

Notice that the above derivation holds for linear quantile regression model in (1) generally. Now we consider the model under H_0 where only x_{i1} is included. Set $d_i = x_{i2} - \hat{x}_{i2}$, from (A2) we get

$$n^{-1/2} \sum_{i=1}^n (x_{2i} - \hat{x}_{2i})(\hat{a}_i(\tau) - (1 - \tau)) = n^{-1/2} \sum_{i=1}^n (x_{2i} - \hat{x}_{2i} - K_n^\tau x_{1i})(\tilde{a}_i(\tau) - (1 - \tau)) + o_p(1), \quad (\text{A3})$$

where $K_n^\tau = (X_2 - \hat{X}_2)^T \Gamma_n^\tau X_1 (X_1^T \Gamma_n^\tau X_1)^{-1}$ and $\Gamma_n^\tau = \text{diag}(f_i(x_{i1}^T \beta_1(\tau)))$.

Since (A3) holds uniformly for $\tau \in [\tau_0, \tau_M]$,

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n (x_{2i} - \hat{x}_{2i}) \tilde{b}_i \\ &= n^{-1/2} \sum_{i=1}^n \sum_{\tau_m \in S} (x_{2i} - \hat{x}_{2i} - K_n^{\tau_m} x_{1i}) (\tilde{a}_i(\tau_m) - (1 - \tau_m)) \varphi'(\tau_m) (\tau_m - \tau_{m-1}) + o_p(1). \quad (\text{A4}) \end{aligned}$$

By the Lindeberg-Feller centre limit theorem, (A4) converges to a normal distribution of mean 0 and variance

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{\tau_m \in S} \varphi^2(\tau_m) (\tau_m - \tau_{m-1})^2 (1 - \tau_m) \tau_m (x_{2i} - \hat{x}_{2i} - K_n^{\tau_m} x_{1i}) (x_{2i} - \hat{x}_{2i} - K_n^{\tau_m} x_{1i})^T.$$

Thus $T_n = S_n^T Q_n^{-1} S_n$ converges to a mixed chi-square distribution. ■

We now want to study the consistency of our model-based bootstrap. Parallel to the notations in the original sample space, we have the following notations in the bootstrap space:

$$\begin{aligned} \hat{W}_n^{d*} &= n^{-1/2} \sum d_i(\hat{a}_i^*(\tau) - (1 - \tau)) \text{ where } \hat{a}_i^*(\tau) \text{ is the regression rank score under } H_0 \text{ for the} \\ &\quad \text{bootstrap sample.} \\ W_n^{d*} &= n^{-1/2} \sum d_i(\tilde{a}_i^*(\tau) - (1 - \tau)) \text{ where } \tilde{a}_i^*(\tau) = \mathbb{I}(y_i^* \geq x_{i1}^T \tilde{\beta}_1(\tau)). \\ \hat{G}_n^{d*}(t) &= n^{-1/2} \sum d_i \mathbb{I}(y_i^* \leq x_i^T t). \\ G_n^{d*}(t) &= n^{-1/2} \sum d_i E^* \mathbb{I}(y_i^* \leq x_i^T t). \\ Z_n^* &= \sqrt{n}(\hat{\beta}_1^*(\tau) - \hat{\beta}_1(\tau)). \end{aligned}$$

We first show that $\hat{\beta}_1^*(\tau)$ is a consistent estimator of $\beta_1(\tau)$. The relationship corresponding to Lemma A.2 under the bootstrap space is given in Lemma A.3. Combining the above results, we can establish the consistency of our bootstrap algorithm in Theorem 3.1(iii).

Proof of Theorem 3.1(ii): Write $\tilde{y}_i = x_{i1}^T \bar{\beta}_1(u_i)$, where $\bar{\beta}_1(\cdot)$ is the linear interpolation of $\{\beta_1(\tau_m), m \in S\}$. By the law of large numbers,

$$\begin{aligned} &\left\| \frac{1}{n} E^* \left(\sum \rho_\tau(\tilde{y}_i - x_{i1}^T \beta_1) - \sum \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)) \right) \right. \\ &\quad \left. - \frac{1}{n} \left(\sum \rho_\tau(\tilde{y}_i - x_{i1}^T \beta_1) - \sum \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)) \right) \right\| = o_{p^*}(1). \end{aligned} \quad (A5)$$

Note that the expectation above is taken with respect to u_i . Because

$$\begin{aligned} &\left\| \frac{1}{n} \left(\sum \rho_\tau(\tilde{y}_i - x_{i1}^T \beta_1) - \sum \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)) \right) \right. \\ &\quad \left. - \frac{1}{n} \left(\sum \rho_{\tau'}(\tilde{y}_i - x_{i1}^T \beta_1') - \sum \rho_{\tau'}(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau')) \right) \right\| \\ &\leq c_1 |\tau - \tau'| + c_2 \|\beta_1 - \beta_1'\|, \end{aligned}$$

$\frac{1}{n} (\sum \rho_\tau(\tilde{y}_i - x_{i1}^T \beta_1) - \sum \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)))$ is stochastically equicontinuous. Thus the convergence in (A5) is uniform over $\tau \in [\tau_a, \tau_b]$ and β_1 in a compact set $\mathcal{B}$. We know that

$$\hat{\beta}_1^*(\tau) = \underset{\beta_1}{\operatorname{argmin}} \sum \rho_\tau(y_i^* - x_{i1}^T \beta_1) - \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)),$$

and

$$\bar{\beta}_1(\tau) = \underset{\beta_1}{\operatorname{argmin}} E^* \left(\sum \rho_\tau(\tilde{y}_i - x_{i1}^T \beta_1) - \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)) \right).$$

The minimiser $\bar{\beta}_1(\tau)$ is also unique for $\tau \in [\tau_a, \tau_b]$. Notice that

$$\begin{aligned} &\left\| \frac{1}{n} \sum \rho_\tau(\tilde{y}_i - x_{i1}^T \beta_1) - \frac{1}{n} \sum \rho_\tau(y_i^* - x_{i1}^T \beta_1) \right\| \\ &= O\left(\frac{1}{n} \sum |y_i^* - \tilde{y}_i|\right) \end{aligned}$$

$$= O(\|\tilde{\beta}_1(\tau) - \bar{\beta}_1(\tau)\|). \quad (\text{A6})$$

Since $\sup_{\tau_a \leq \tau \leq \tau_b} \|\hat{\beta}_1(\tau) - \beta_1(\tau)\| = o_p(1)$, $\sup_{\tau_a \leq \tau \leq \tau_b} \|\tilde{\beta}_1(\tau) - \bar{\beta}_1(\tau)\| = o_p(1)$. Thus by (A5) and (A6),

$$\begin{aligned} & \sup_{\tau \in [\tau_a, \tau_b], \beta_1 \in \mathcal{B}} \left\| \frac{1}{n} E^* \left(\sum \rho_\tau(\tilde{y}_i - x_{i1}^T \beta_1) - \sum \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)) \right) \right. \\ & \quad \left. - \frac{1}{n} \left(\sum \rho_\tau(y_i^* - x_{i1}^T \beta_1) - \sum \rho_\tau(\tilde{y}_i - x_{i1}^T \bar{\beta}_1(\tau)) \right) \right\| = o_p^*(1) + o_p(1). \end{aligned} \quad (\text{A7})$$

By writing $o_p^*(1) + o_p(1)$, we mean that the left hand side of Equation (A7) is $o_p^*(1)$ with probability going to 1 in the original sample space. Let $\mathcal{B}_\delta(\bar{\beta}_1(\tau))$ be a ball of radius δ centred at $\bar{\beta}_1(\tau)$ with L^∞ norm. For any $b(\tau)$ in the boundary of $\mathcal{B}_\delta(\bar{\beta}_1(\tau))$,

$$\begin{aligned} & \frac{1}{n} \sum \rho_\tau(y_i^* - x_{i1}^T b(\tau)) - \frac{1}{n} \sum \rho_\tau(y_i^* - x_{i1}^T \bar{\beta}_1(\tau)) \\ & \geq \frac{1}{n} E^* \sum \rho_\tau(y_i^* - x_{i1}^T b(\tau)) - \frac{1}{n} E^* \sum \rho_\tau(y_i^* - x_{i1}^T \bar{\beta}_1(\tau)) - o_p^*(1) - o_p(1) \\ & \geq \epsilon(\tau) - o_p^*(1) - o_p(1), \end{aligned}$$

where $\epsilon(\tau) \geq 0$ and the inequality is strict for some $\tau \in [\tau_0, \tau_M]$. Namely,

$$P^* \left(\inf_{\sup_{|\bar{\beta}_1(\tau) - b(\tau)| = \delta} \tau \in [\tau_a, \tau_b]} \left(\sum \rho_\tau(y_i^* - x_{i1}^T b(\tau)) - \sum \rho_\tau(y_i^* - x_{i1}^T \bar{\beta}_1(\tau)) \leq 0 \right) \right) \rightarrow 0$$

in P . By the convexity of ρ_τ ,

$$P^* \left(\inf_{\sup_{|\bar{\beta}_1(\tau) - b(\tau)| \geq \delta} \tau \in [\tau_a, \tau_b]} \left(\sum \rho_\tau(y_i^* - x_{i1}^T b(\tau)) - \sum \rho_\tau(y_i^* - x_{i1}^T \bar{\beta}_1(\tau)) \leq 0 \right) \right) \rightarrow 0$$

in P . Also notice that $\bar{\beta}_1(\tau) = \beta_1(\tau)$ for $\tau \in S$. Thus we have the desired result. $\blacksquare$

Lemma A.3: With probability going to 1 in the sample space, $G_n^{d*}(\hat{\beta}_1^*(\tau)) = G_n^{d*}(\hat{\beta}_1(\tau)) - D_{nd}^\tau Z_n^* + O(\sqrt{n} \|\hat{\beta}_1^*(\tau) - \hat{\beta}_1(\tau)\|^2) + o_p^*(1)$ uniformly for $\tau \in S \cap [\tau_a, \tau_b]$.

Proof: Write $G_n^{d*}(\hat{\beta}_1(\tau) + \delta) - G_n^{d*}(\hat{\beta}_1(\tau)) = A_1 + A_2$ where

$$\begin{aligned} A_1 &= n^{-1/2} \sum d_i E^* \left(\mathbb{I}(u_i \leq \tau_0) (\mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau_0) \leq x_{i1}^T (\hat{\beta}_1(\tau) + \delta)) \right. \\ & \quad \left. - \mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau_0) \leq x_{i1}^T \hat{\beta}_1(\tau))) \right) \\ & \quad + n^{-1/2} \sum d_i E^* \left(\mathbb{I}(u_i \geq \tau_M) (\mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau_M) \leq x_{i1}^T (\hat{\beta}_1(\tau) + \delta)) \right. \\ & \quad \left. - \mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau_M) \leq x_{i1}^T \hat{\beta}_1(\tau))) \right), \\ A_2 &= n^{-1/2} \sum d_i E^* \left(\mathbb{I}(\tau_0 < u_i < \tau_M) (\mathbb{I}(x_{i1}^T \tilde{\beta}_1(u_i) < x_{i1}^T (\hat{\beta}_1(\tau) + \delta)) \right. \\ & \quad \left. - \mathbb{I}(x_{i1}^T \tilde{\beta}_1(u_i) < x_{i1}^T \hat{\beta}_1(\tau))) \right). \end{aligned}$$

From Theorem 1 of Neocleous and Portnoy (2008), $x_{i1}^T \tilde{\beta}_1(\tau)$ is strictly monotone uniformly on $[\tau_0, \tau_M]$ with probability tending to 1. Therefore for $\tau \in [\tau_a, \tau_b]$, A_1 is $o_p(1)$ for any $\delta \rightarrow 0$.

Let $\Delta = \tilde{\beta}_1(u_i) - \beta_1(u_i)$, write A_2 as

$$\begin{aligned} & n^{-1/2} \sum d_i E^* \left(\mathbb{I}(\tau_0 < u_i < \tau_M) (\mathbb{I}(x_{i1}^T \beta_1(u_i) + x_{i1}^T \Delta < x_{i1}^T \hat{\beta}_1(\tau) + x_{i1}^T \delta) \right. \\ & \quad \left. - \mathbb{I}(x_{i1}^T \beta_1(u_i) + x_{i1}^T \Delta < x_{i1}^T \hat{\beta}_1(\tau)) \right) \\ & = n^{-1/2} \sum d_i E^* \left(\mathbb{I}(\tau_0 < u_i < \tau_M) \mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta < x_{i1}^T \beta_1(u_i) \right. \\ & \quad \left. < x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta - x_{i1}^T \delta) \mathbb{I}(x_{i1}^T \delta < 0) \right) \\ & \quad + n^{-1/2} \sum d_i E^* \left(\mathbb{I}(\tau_0 < u_i < \tau_M) \mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta - x_{i1}^T \delta < x_{i1}^T \beta_1(u_i) \right. \\ & \quad \left. < x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta) \mathbb{I}(x_{i1}^T \delta \geq 0) \right). \end{aligned}$$

We only need to consider the case when $x_{i1}^T \delta < 0$, since the situation when $x_{i1}^T \delta \geq 0$ is symmetric.

When $x_{i1}^T \delta < 0$,

$$\begin{aligned} & n^{-1/2} \sum d_i E^* \left(\mathbb{I}(\tau_0 < u_i < \tau_M) \mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta < x_{i1}^T \beta_1(u_i) < x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta - x_{i1}^T \delta) \right) \\ & = n^{-1/2} \sum d_i \int_{\max\{x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta, x_{i1}^T \beta_1(\tau_0)\}}^{\min\{x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta - x_{i1}^T \delta, x_{i1}^T \beta_1(\tau_M)\}} f_i(c) dc \\ & = n^{-1/2} \sum d_i \int_{\max\{x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta, x_{i1}^T \beta_1(\tau_0)\}}^{\min\{x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta - x_{i1}^T \delta, x_{i1}^T \beta_1(\tau_M)\}} f_i(x_{i1}^T \beta_1(\tau)) + O(|c - x_{i1}^T \beta_1(\tau)|) dc \\ & = n^{-1/2} \sum \left(d_i f_i(x_{i1}^T \beta_1(\tau)) (-\max\{x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta, x_{i1}^T \beta_1(\tau_0)\} \right. \\ & \quad \left. + \min\{x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta - x_{i1}^T \delta, x_{i1}^T \beta_1(\tau_M)\}) \right. \\ & \quad \left. + O(\|\Delta\| \|\delta\|) + O(\|\delta\|^2) + O(\|\hat{\beta}_1(\tau) - \beta_1(\tau)\| \|\delta\|) \right) \\ & = n^{-1/2} \sum d_i f_i(x_{i1}^T \beta_1(\tau)) (-x_{i1}^T \delta) + O(\sqrt{n} \|\Delta\| \|\delta\|) + O(\sqrt{n} \|\delta\|^2) \\ & \quad + O(\sqrt{n} \|\hat{\beta}_1(\tau) - \beta_1(\tau)\| \|\delta\|) + R_1, \end{aligned}$$

where

$$\begin{aligned} R_1 = O \left(n^{-1/2} \sum d_i f_i(x_{i1}^T \beta_1(\tau)) (\mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta < x_{i1}^T \beta_1(\tau_0)) \right. \\ \left. + \mathbb{I}(x_{i1}^T \hat{\beta}_1(\tau) - x_{i1}^T \Delta - x_{i1}^T \delta > x_{i1}^T \beta_1(\tau_M))) \right). \end{aligned}$$

For $\tau \in [\tau_a, \tau_b]$, R_1 converges to zero in probability if Δ and δ are $o(1)$.

Recall $\Delta = \tilde{\beta}_1(u_i) - \beta_1(u_i)$, which is $O_p(n^{-1/2})$ uniformly over $u_i \in [\tau_0, \tau_M]$ by Theorem 1 of Neocleous and Portnoy (2008). Letting $\delta = \hat{\beta}_1^*(\tau) - \hat{\beta}_1(\tau) = o_{p^*}(1) + o_p(1)$ for $\tau \in S \cap [\tau_a, \tau_b]$, we have

$$A_2 = D_{nd}^T Z_n^* + O_p(\sqrt{n} \|\hat{\beta}_1^*(\tau) - \hat{\beta}_1(\tau)\|^2) + o_{p^*}(1) + o_p(1).$$

■

Proof of Theorem 3.1(iii): Similar to Lemma A.1,

$$\sup_{\tau \in S \cap [\tau_a, \tau_b]} \left\| \hat{G}_n^{d*}(\hat{\beta}_1^*(\tau)) - \hat{G}_n^{d*}(\hat{\beta}_1(\tau)) - G_n^{d*}(\hat{\beta}_1^*(\tau)) + G_n^{d*}(\hat{\beta}_1(\tau)) \right\| = o_{p^*}(1) + o_p(1).$$

This is because

$$\mathcal{G}^* = \{v^T d_i \mathbb{I}(y_i^* \leq x_{i1}^T t), t \in T\}$$

is a VC subgraph class and $\hat{\beta}_1^*(\tau)$ is consistent for $\hat{\beta}_1(\tau)$ uniformly over $\tau \in S \cap [\tau_a, \tau_b]$. Thus we have

$$\hat{W}_n^{d*} = W_n^{d*} - D_{nd}^\tau Z_n^* + R_1^* - R_2^* - R_3^*,$$

where

$$\begin{aligned} R_1^* &= n^{-1/2} \sum d_i \mathbb{I}(y_i^* = x_{i1}^T \hat{\beta}_1^*(\tau)) \hat{a}_i^*(\tau), \\ R_2^* &= \hat{G}_n^{d*}(\hat{\beta}_1^*(\tau)) - \hat{G}_n^{d*}(\hat{\beta}_1(\tau)) - G_n^{d*}(\hat{\beta}_1^*(\tau)) + G_n^{d*}(\hat{\beta}_1(\tau)), \\ R_3^* &= G_n^{d*}(\hat{\beta}_1^*(\tau)) - G_n^d(\hat{\beta}_1(\tau)) - D_{nd}^\tau Z_n^*. \end{aligned}$$

Since R_1^* and R_2^* are $o_{p^*}(1) + o_p(1)$, by Lemma A.4, we have

$$\hat{W}_n^{d*} = W_n^{d*} - D_{nd}^\tau Z_n^* + O_p(\sqrt{n} \|\hat{\beta}_1^*(\tau) - \hat{\beta}_1(\tau)\|^2) + o_{p^*}(1) + o_p(1).$$

Set $d_i = x_{i1}$,

$$D_{nx}^\tau Z_n^* = W_n^{x*} + O_p(\sqrt{n} \|\hat{\beta}^*(\tau) - \hat{\beta}(\tau)\|^2) + o_{p^*}(1) + o_p(1).$$

Thus $Z_n^* = (D_{nx}^\tau)^{-1} W_n^{x*} + o_{p^*}(1) + o_p(1)$ and

$$\hat{W}_n^{d*} = W_n^{d*} - D_{nd}^\tau (D_{nx}^\tau)^{-1} W_n^{x*} + o_{p^*}(1) + o_p(1).$$

Set $d_i = \hat{x}_{i2} - x_{i2}$,

$$\begin{aligned} & n^{-1/2} \sum_i (x_{i2} - \hat{x}_{i2})(\hat{a}_i^*(\tau) - (1 - \tau)) \\ &= n^{-1/2} \sum_i (x_{i2} - \hat{x}_{i2} - K_n^\tau x_{i1})(\tilde{a}_i^*(\tau) - (1 - \tau)) + o_{p^*}(1) + o_p(1). \end{aligned}$$

Therefore

$$S_n^* = n^{-1/2} \sum_{i=1}^n \sum_{\tau_m \in S} (x_{i2} - \hat{x}_{i2} - K_n^{\tau_m} x_{i1})(\tilde{a}_i^*(\tau_m) - (1 - \tau_m)) \varphi'(\tau_m)(\tau_m - \tau_{m-1}) + o_{p^*}(1) + o_p(1). \quad (\text{A8})$$

Comparing Equation (A3) with (A8), their right hand sides are exactly the same except that we have $\tilde{a}_i^*$ instead of $\tilde{a}_i$ for the bootstrapped test statistics. Recall $\tilde{a}_i(\tau) = \mathbb{I}(y_i \geq x_{i1}^T \beta(\tau))$ and $\tilde{a}_i^*(\tau) = \mathbb{I}(y_i^* \geq x_{i1}^T \tilde{\beta}_1(\tau))$. Consider a set $\mathcal{D}$ where $x_{i1}^T \tilde{\beta}_1(\tau)$ is strictly monotone for $\tau \in [\tau_0, \tau_M]$. On $\mathcal{D}$, $\tilde{a}_i^*(\tau)$ given data independently follows the same binary distribution as $\tilde{a}_i(\tau)$. Therefore the conditional distribution of T_n^* given data will convergence to the same limiting distribution as T_n on $\mathcal{D}$. We then have the desired results since $P(\mathcal{D}) \rightarrow 1$ as $n \rightarrow \infty$ by Theorem 1 of Neocleous and Portnoy (2008). ■