Estimation and Inference for Multi-Kink Quantile Regression

Wei Zhong[1], Chuang Wan[1] and Wenyang Zhang[2]
Xiamen University[1] and The University of York[2]

February 20, 2021

Abstract

This article is concerned with parameter estimation, kink points detection and statistical inference for a new Multi-Kink Quantile Regression (MKQR) model. It assumes different linear quantile regression forms in different regions of the domain of the threshold covariate but are still continuous at kink points. First, we investigate parameter estimation, kink point detection and statistical inference in MKQR models. We propose an iterative segmented quantile regression algorithm for estimating both the regression coefficients and the locations of kink points. The proposed algorithm is much more computationally efficient than the grid search algorithm and not sensitive to the selection of initial values. Second, asymptotic properties, such as selection consistency of the number of kink points and asymptotic normality of the estimators of both regression coefficients and kink effects, are established to justify the proposed method theoretically. Third, a score test based on partial subgradients is developed to verify whether the kink effects exist or not. Test-inversion confidence intervals for kink location parameters are also constructed. Monte Carlo simulations and two real data applications on the secondary industrial structure of China and the triceps skinfold thickness of Gambian females illustrate the excellent finite sample performances of the proposed MKQR model. Last, a new R package *MultiKink* is developed to easily implement the proposed methods.

Keywords: Change point detection, hypothesis testing, kink regression, model selection, quantile regression.

1 Introduction

The kink regression model (Hansen, 2017) or the bent line regression (Li et al., 2011) assumes that linear regression forms are separately modelled on two sides of an unknown threshold but still continuous at the threshold. It is a very useful tool to deal with the nonlinearity in data analysis. Let Y_t be a response variable of interest, X_t be a univariate threshold variable and $\mathbf{Z}_t$ be a p dimensional random vector of additional covariates, $t = 1,\ 2, \cdots,\ n$. Hansen (2017) considered the following kink regression model with an unknown threshold,

$$Y_t = \alpha_0 + \alpha_1(X_t - \delta)I(X_t \leq \delta) + \alpha_2(X_t - \delta)I(X_t > \delta) + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{Z}_t + e_t, \tag{1.1}$$

where e_t is the random error with $E(e_t|X_t,\ \mathbf{Z}_t) = 0$. The threshold variable X_t has different slopes on different segments formed by δ, but the regression function is continuous in X_t. This kind of non-linear pattern is commonly referred to as kink effect (Hansen, 2017) or bent line effect (Li et al., 2011). The parameter δ is called "kink point", "change point" or "threshold" exchangeably to represent the point where the regression function form changes. Compared with the linear models, the kink regression model relaxes the linearity assumption and thus is able to capture the necessary nonlinearity. Compared with the full nonparametric models, the kink regression model has the better interpretability by maintaining linear regression models in different regions of the domain of X_t. Thus, the kink regression model enjoys both the interpretability of linear models and the flexibility of nonparametric models.

In the literature, the kink regression models with a single unknown threshold point have been intensively studied. For example, Hansen (2017) combined the least squares estimation and a grid search algorithm to estimate the regression coefficients and the kink point in Model (1.1). The F-type statistic test was also proposed for testing $H_0 : \alpha_1 = \alpha_2$ in Model (1.1). In the application, Hansen (2017) demonstrated the famous inverted U-shaped relationship between the GDP growth rate and the ratio of debt to GDP (Reinhart and Rogoff, 2010). Li et al. (2011) proposed a bent line quantile regression model and showed that the logarithm of maximal running speed of land mammals linearly increases with the logarithm of mass up to a certain point and then decreases as the mass rises. Zhang and Li (2017) studied estimation and hypothesis testing for a continuous threshold expectile model. Fong (2019) developed fast bootstrap confidence intervals for a continuous threshold linear regression. Hidalgo et al. (2019) discussed whether there experiences a discontinuous jump or a continuous kink at the threshold point by using the quasi-likelihood-ratio test and constructed a robust confidence interval for the threshold.

However, the kink regression models with only one threshold point are not sufficient in some applications. In our second real data example, the logarithm of triceps skinfold thickness (TSF) as an important measure of body density decreases with the age in the childhood up to about 10 years old, then experiences a growth spurt at adolescence up to about 18-20 years old and finally stays almost stable for adults (see Figure 4). The kink

regression models with one threshold are clearly not appropriate for this case. It would be more sensible to consider the kink regression models with multiple threshold points. Muggeo and Adelfio (2010) studied a piecewise constant model in mean regression with multiple change points and used the penalized method to select change points. However, the mean regression is not robust to outliers and we find that the penalized method has relatively poor performance to select change points.

Moreover, to achieve the robustness to outliers and heavy-tailed errors which are often present in the data, we consider quantile regression to analyze data with heterogeneous conditional distributions. Quantile regression is able to provide a comprehensive picture of the conditional distribution of the response Y_t given the covariates, especially when its upper or lower quantiles are particularly of interest. Thus, we are motivated to study the Multi-Kink Quantile Regression (MKQR) model and establish the proper estimation and inference procedures. For a given quantile level $\tau \in (0,1)$, the MKQR model is

$$Y_t = \alpha_0 + \alpha_1 X_t + \sum_{k=1}^{K} \beta_k (X_t - \delta_k) I(X_t > \delta_k) + \boldsymbol{\gamma}^{\mathrm{T}} \mathbf{Z}_t + e_t, \tag{1.2}$$

where β_k represents the difference in slopes for X_t between two adjacent segments, $\boldsymbol{\gamma}$ is the coefficient vector of covariates $\mathbf{Z}_t$, which stays constant on the whole domain of $\mathbf{Z}_t$ and e_t are independent random errors such that $P(e_t < 0 | X_t, \mathbf{Z}_t) = \tau$. Note that $\beta_k \neq 0$ implies the existence of a kink effect at $X_t = \delta_k$ and thus $\{\delta_k, k = 1, \cdots, K\}$ represent the kink points or the locations where kink effects happen satisfying $\delta_1 < \cdots < \delta_K$. So There are $K+1$ regimes in total divided by K kink points where the number of kink points K as well as their locations are both unknown. We emphasize once again that the regression for MKQR model is continuous everywhere. Note that all the unknown parameters depend on the quantile index τ, but we omit the subscript τ for ease of notations throughout the paper. We remark that the MKQR model (1.2) is different from traditional threshold regression models (Hansen, 2000; Caner, 2002) that specify different regression functions in subsamples segmented by a threshold variable and jumps at the threshold points are allowed. However, the threshold variable X_t in the MKQR model is the predictor of interest in the regression and the regression curves are everywhere continuous on the domain of X_t. We also remark that the MKQR models can be also regarded as a special semiparametric partially linear quantile regression model where the nonparametric function of X_t is specifically modelled by a continuous piecewise linear curve.

In this paper, we focus on parameter estimation, kink points detection and statistical inference for the new MKQR model (1.2) where the number of kink points and their locations are both unknown using quantile regression. We contribute the literature in the following several aspects. First, the MKQR model (1.2) extends the existing kink regression with an unknown threshold to wider applications with unknown multiple kink points. We propose a Bootstrap Restarting Iterative Segmented Quantile (BRISQ) regression al-

gorithm for estimating both the regression coefficients and kink effects. This algorithm is much more computationally efficient than the grid search algorithm and not sensitive to the initial values due to the bootstrap restarting idea of Wood (2001). Furthermore, we suggest a backward elimination algorithm to identify the number of kinks by transforming change points detection into a model selection problem based on a strengthened quantile BIC. Second, we theoretically demonstrate that the selection consistency of the number of kink points and the asymptotical normality of the estimators for both regression coefficients and kink effects. Third, the MKQR model inherits the merits of quantile regression and thus is able to robustly model data with heterogeneous conditional distributions especially when upper or lower quantiles of the response are particularly of interest. Forth, from the statistical inference perspective, we develop a score test based on partial subgradient of quantile objective function under the null hypothesis to verify whether the kink effects exist or not. A test-inversion confidence interval based on a smoothed rank score test for a kink location parameter is also proposed and can be extended to multiple kink parameters by sample splitting. Fifth, two real data applications on the secondary industrial structure of China and the triceps skinfold thickness for Gambian females are studied to identify the kink points, which would be of interest for economists and biologists, respectively. Last, a new R package *MultiKink* is developed to implement all the estimation and inference procedures, and is free to use.

The rest of the paper is structured as follows. In Section 2, we describe the estimation procedures for the MKQR model and investigate the asymptotic properties. Section 3 presents a testing procedure for the existence of kink effects and construct the test-inversion confidence intervals for the kink locations. The finite sample performances of the proposed methods are evaluated via simulation experiments in Section 4. Section 5 presents two real data applications. Section 6 concludes the paper. The technical proofs are presented in the Appendix.

2 Estimation and Algorithm

2.1 Parameter Estimation When K is Known

From model (1.2), the τth conditional quantile of Y_t given X_t and $\mathbf{Z}_t$ is denoted by

$$Q_Y(\tau|X_t,\mathbf{Z}_t)=\alpha_0+\alpha_1X_t+\sum_{k=1}^{K}\beta_k(X_t-\delta_k)I(X_t>\delta_k)+\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{Z}_t. \tag{2.1}$$

For notation convenience, we let $\boldsymbol{\beta}=(\beta_1,\cdots,\beta_K)^{\mathrm{T}}$, $\boldsymbol{\eta}=(\alpha_0,\alpha_1,\boldsymbol{\beta}^{\mathrm{T}},\boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\delta}=(\delta_1,\cdots,\delta_K)^{\mathrm{T}}$. Notice that both the dimensions of $\boldsymbol{\eta}$ and $\boldsymbol{\delta}$ hinge on K. Let $\mathbf{W}_t=(X_t,\mathbf{Z}_t^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\theta}=(\boldsymbol{\eta}^{\mathrm{T}},\boldsymbol{\delta}^{\mathrm{T}})^{\mathrm{T}}$. We rewrite the τth conditional quantile of Y_t given $\mathbf{W}_t$ as $Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)$ to emphasize the dependence on the parameter $\boldsymbol{\theta}$. In the ideal case where the true number of kink

points denoted by K_0 is known, we can estimate $\boldsymbol{\theta}$ by the following objective function,

$$S_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^{n}\rho_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}, \tag{2.2}$$

where $\rho_\tau(u) = u\{\tau - I(u<0)\}$. A common estimator for $\boldsymbol{\theta}$ is thereby

$$\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\eta}}^{\mathrm{T}}, \widehat{\boldsymbol{\delta}}^{\mathrm{T}})^{\mathrm{T}} = \underset{\boldsymbol{\eta}\in\mathcal{B},\boldsymbol{\delta}\in\Gamma}{\arg\min} S_n(\boldsymbol{\theta}), \tag{2.3}$$

where $\mathcal{B} \subset \mathbb{R}^{2+K+p}$ and $\Gamma \subset \Omega^K$ are compact sets, in which Ω denotes the support of the threshold variable X_t.

If all kink locations parameters $\boldsymbol{\delta}$ are known, a simple quantile regression can be directly used for (2.1). For the single kink regression with an unknown threshold, the greedy grid search algorithm (Li et al., 2011; Hansen, 2017) can be used to exhaustively seek the kink point. However, without any prior information, we have to assume that both the number of kink points K and the kink locations vector $\boldsymbol{\delta}$ are unknown in Model (2.1). It makes the grid search approach inappropriate especially when K is large, because its computational cost grows at an exponential rate of K. Since the minimization problem in (2.2) is non-convex with respect to $\boldsymbol{\delta}$, the convex optimization algorithms can not directly applied. Bai and Perron (2003) proposed a novel dynamic programming algorithm to detect change points for the linear model with multiple structural changes. Oka and Qu (2011) used the dynamic programming algorithm to estimate multiple structural changes in regression quantiles. However, the dynamic programming algorithms can not be directly applied to the kink models due to the continuity constraints. To this end, we develop a new iterative segmented quantile regression algorithm to detect the number of kink points and estimate regression coefficients simultaneously.

2.1.1 Bootstrap Restarting Iterative Segmented Quantile Algorithm

Although K_0 is given, $(X-\delta_k)I(X>\delta_k)$ for $k = 1,\cdots,K$ are not observable and still non-differentiable at δ_k's. Given an initial location vector $\boldsymbol{\delta}^{(0)} = (\delta_1^{(0)},\cdots,\delta_K^{(0)})^{\mathrm{T}}$, we employ the first-order Taylor expansion to approximate $(X_t-\delta_k)I(X_t>\delta_k)$ around $\delta_k^{(0)}$,

$$(X_t-\delta_k)I(X>\delta_k) \approx (X_t-\delta_k^{(0)})I(X_t>\delta_k^{(0)}) - (\delta_k-\delta_k^{(0)})I(X_t>\delta_k^{(0)}).$$

Then, Model (2.1) can be approximated by

$$Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t) \approx \alpha_0 + \alpha_1 X_t + \sum_{k=1}^{K}\beta_k\widetilde{U}_{kt} + \sum_{k=1}^{K}\phi_k\widetilde{V}_{kt} + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{Z}_t, \tag{2.4}$$

where $\phi_k = \beta_k(\delta_k - \delta_k^{(0)})$, $\widetilde{U}_{kt} = (X_t - \delta_k^{(0)})I(X_t > \delta_k^{(0)})$ and $\widetilde{V}_{kt} = -I(X_t > \delta_k^{(0)})$ are two new covariates with coefficients β_k and ϕ_k, respectively. Denote $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_K)^{\mathrm{T}}$ and $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_K)^{\mathrm{T}}$. This local linear approximation technique has been also used in the change point detection for the piecewise constant model (Muggeo and Adelfio, 2010) and nonconcave penalized regression (Zou and Li, 2008).

By fitting the standard linear quantile model (2.4), a new estimator for δ_k can be updated by $\widehat{\delta}_k^{(1)} = \delta_k^{(0)} + \widehat{\phi}_k/\widehat{\beta}_k$, for $k = 1, \ldots, K$. The estimator could be iteratively updated. However, if initial values are not appropriately chosen, some elements of $\widehat{\boldsymbol{\delta}}^{(i)}$ at the ith iteration are possible to jump out of the support of the threshold covariate X_t or be very close to another kink point to make them hard to distinguish. We define the inadmissible set $\mathcal{T}$ as

$$\mathcal{T} = \left\{ \widehat{\delta}_k : (\widehat{\delta}_k \notin \Omega, \text{ the support of } X_t) \bigcup (|\widehat{\delta}_k - \widehat{\delta}_{k'}| < \Delta, \text{ where } \Delta \text{ is small and } k \neq k') \right\}.$$

In practice, we need to discard all $\widehat{\delta}_k$ in the inadmissible set $\mathcal{T}$. In our simulation, we choose Δ to be one thirtieth of the range of the threshold variable X_t. The underlying reason is that the local linear approximation technique is sensitive to the initial values $\boldsymbol{\delta}^{(0)}$, which makes the algorithm easy to get stuck in local optima. To deal with this drawback, we iteratively update the initial values using the bootstrap samples to make the new algorithm insensitive to the original initial values. It shares the similar spirit of the bootstrap restarting idea of Wood (2001). Thus, we call it Bootstrap Restarting Iterative Segmented Quantile (BRISQ) regression algorithm.

The main idea of the BRISQ algorithm is illustrated as follows. We first initialize parameters $\boldsymbol{\delta}^{(0)}$ evenly dispersed on the domain of X_t given K and obtain the estimator $\widehat{\boldsymbol{\theta}}_{(0)} = (\widehat{\boldsymbol{\eta}}_{(0)}^{\mathrm{T}}, \widehat{\boldsymbol{\delta}}_{(0)}^{\mathrm{T}})^{\mathrm{T}}$ by iteratively fitting the working model (2.4). Then, we generate a bootstrap sample $\mathcal{X}_n^*$ in the classic way that we randomly select original observations with replacement and estimate Model (2.4) again using $\widehat{\boldsymbol{\delta}}_{(0)}$ as the initial kink locations to obtain the bootstrap estimator $\widetilde{\boldsymbol{\theta}}_{(1)}^* = (\widetilde{\boldsymbol{\eta}}_{(1)}^{*\mathrm{T}}, \widetilde{\boldsymbol{\delta}}_{(1)}^{*\mathrm{T}})^{\mathrm{T}}$. And then, we estimate Model (2.4) again based on the original sample using the bootstrap estimator $\widetilde{\boldsymbol{\delta}}_{(1)}^*$ as initial values and obtain the new estimator $\widetilde{\boldsymbol{\theta}}_{(1)} = (\widetilde{\boldsymbol{\eta}}_{(1)}^{\mathrm{T}}, \widetilde{\boldsymbol{\delta}}_{(1)}^{\mathrm{T}})^{\mathrm{T}}$. Next, we compare $S_n(\widetilde{\boldsymbol{\theta}}_{(1)})$ with $S_n(\widehat{\boldsymbol{\theta}}_{(0)})$. If $S_n(\widetilde{\boldsymbol{\theta}}_{(1)}) < S_n(\widehat{\boldsymbol{\theta}}_{(0)})$, we update $\widehat{\boldsymbol{\theta}}_{(1)} = \widetilde{\boldsymbol{\theta}}_{(1)}$; otherwise, $\widehat{\boldsymbol{\theta}}_{(1)} = \widehat{\boldsymbol{\theta}}_{(0)}$. Last, we repeat the previous procedure until convergence. The flowchart of this BRISQ algorithm is displayed in Figure 1. In practice, this algorithm can efficiently jump out of local minima and substantially improve the stability and accuracy of estimation, therefore less sensitive to the original initial values. The detailed procedures are summarized in Algorithm 1 and easily implemented using the newly developed R package *MultiKink*. The convergence of the BRISQ algorithm is studied in Section S.1 of the Supplement.

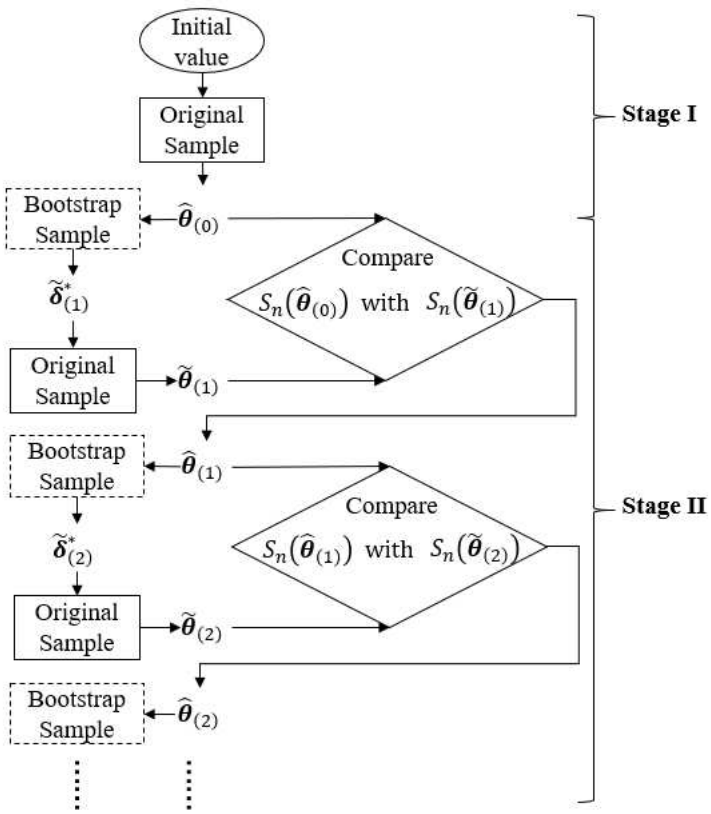

Figure 1: The flowchart of the BRISQ algorithm. Stage I contains Steps 1-3 and Stage II contains Step 4 in the Algorithm 1.

2.1.2 Limiting Distribution

We then derive the asymptotic properties for $\widehat{\boldsymbol{\theta}}$. Denote the true parameters as $\boldsymbol{\theta}_0 = (\boldsymbol{\eta}_0^{\mathrm{T}}, \boldsymbol{\delta}_0^{\mathrm{T}})^{\mathrm{T}} = \arg\min S(\boldsymbol{\theta})$, where $S(\boldsymbol{\theta}) = E\left[\rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}|\mathbf{W}_t)\}\right]$. Define

$$\mathbf{G}_n = E\left\{\left(\frac{\partial S_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial S_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}}\right\}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = n^{-1}\sum_{t=1}^{n}\tau(1-\tau)E\{h(\mathbf{W}_t;\boldsymbol{\theta}_0)h^{\mathrm{T}}(\mathbf{W}_t;\boldsymbol{\theta}_0)\},$$

$$\mathbf{D}_n = E\left\{\frac{\partial^2 S_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right\}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}E[\psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta})]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

where $\psi_\tau(u) = \tau - I(u \le 0)$. To establish the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$, we need to introduce some notations. Denote $h(\mathbf{W}_t;\boldsymbol{\theta}) = (1, X_t, (X_t-\delta_1)_+, \cdots, (X_t-\delta_K)_+, \mathbf{Z}_t^{\mathrm{T}}, -\beta_1 I(X_t > \delta_1), \cdots, -\beta_K I(X_t > \delta_K))^{\mathrm{T}}$ and the τth conditional quantiles of e_t given $\mathbf{W}_t$ as $F_t^{-1}(\tau|\mathbf{W}_t) = \inf\{u : F(u|\mathbf{W}_t) \ge 0\}$. We then make the following assumptions.

(A1) $F_t \equiv F(\cdot|\mathbf{W}_t)$ has a continuous density $f_t(\cdot|\mathbf{W}_t)$ that satisfies $0 < \inf_t f_t(\cdot) < \sup_t f_t(\cdot) < \infty$ at the point $F^{-1}(\tau|\mathbf{W}_t)$ for any sequence of values of $\mathbf{W}_t$.

Algorithm 1: Bootstrap Restarting Iterative Segmented Quantile (BRISQ) Algorithm.

Step 1. Initialize parameters $\boldsymbol{\delta}^{(0)}$ evenly dispersed on the domain of X_t;
Step 2. Fit Model (2.4) using the standard linear quantile regression based on the initial kink locations $\boldsymbol{\delta}^{(0)}$ and obtain estimators $\widehat{\boldsymbol{\beta}}^{(1)}$ and $\widehat{\boldsymbol{\phi}}^{(1)}$ for $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, respectively. Update the kink locations estimators $\widehat{\boldsymbol{\delta}}^{(1)}$ by $\widehat{\delta}_k^{(1)} = \delta_k^{(0)} + \widehat{\phi}_k^{(1)}/\widehat{\beta}_k^{(1)}$, for $k = 1, \ldots, K$.
Step 3. Repeat Step 2 iteratively until convergence or any $\widehat{\delta}_k^{(i)}$ at the ith step falls into $\mathcal{T}$. The resulting estimator is denoted by $\widehat{\boldsymbol{\theta}}_{(0)} = (\widehat{\boldsymbol{\eta}}_{(0)}^{\mathrm{T}}, \widehat{\boldsymbol{\delta}}_{(0)}^{\mathrm{T}})^{\mathrm{T}}$;
Step 4. **for** *b=1:B* **do**
  Step 4.1. Generate a bootstrap sample $\mathcal{X}_n^*$.
  Step 4.2. Find $\widetilde{\boldsymbol{\theta}}_{(b)}^* = (\widetilde{\boldsymbol{\eta}}_{(b)}^{*\mathrm{T}}, \widetilde{\boldsymbol{\delta}}_{(b)}^{*\mathrm{T}})^{\mathrm{T}}$ for the bootstrap sample $\mathcal{X}_n^*$ using $\widehat{\boldsymbol{\delta}}_{(b-1)}$ as the initial kink locations using the same procedures as Steps 2-3;
  Step 4.3. Find $\widetilde{\boldsymbol{\theta}}_{(b)} = (\widetilde{\boldsymbol{\eta}}_{(b)}^{\mathrm{T}}, \widetilde{\boldsymbol{\delta}}_{(b)}^{\mathrm{T}})^{\mathrm{T}}$ for the original sample $\mathcal{X}_n$ using $\widetilde{\boldsymbol{\delta}}_{(b)}^*$ as the initial kink locations using the same procedures as Steps 2-3;
  Step 4.4. Compare $S_n(\widetilde{\boldsymbol{\theta}}_{(b)})$ with $S_n(\widehat{\boldsymbol{\theta}}_{(b-1)})$. If $S_n(\widetilde{\boldsymbol{\theta}}_{(b)}) < S_n(\widehat{\boldsymbol{\theta}}_{(b-1)})$, $\widehat{\boldsymbol{\theta}}_{(b)} = \widetilde{\boldsymbol{\theta}}_{(b)}$; otherwise, $\widehat{\boldsymbol{\theta}}_{(b)} = \widehat{\boldsymbol{\theta}}_{(b-1)}$.
end
Step 5. Obtain the final estimators $\widehat{\boldsymbol{\theta}} = \frac{1}{B_\epsilon}\sum_{b_\epsilon} \widehat{\boldsymbol{\theta}}_{(b_\epsilon)}$, where $B_\epsilon = \#\left\{b_\epsilon : \left|\frac{S_n(\widehat{\boldsymbol{\theta}}_{(b_\epsilon)}) - S_n(\widehat{\boldsymbol{\theta}}_{\min})}{S_n(\widehat{\boldsymbol{\theta}}_{\min})}\right| \le \epsilon\right\}$, $\widehat{\boldsymbol{\theta}}_{\min} = \arg\min_{\{\widehat{\boldsymbol{\theta}}_{(b)}: 1 \le b \le B\}} S_n(\widehat{\boldsymbol{\theta}}_{(b)})$.

(A2) The objective function $S(\boldsymbol{\theta})$ has a unique global minimum at $\boldsymbol{\theta}_0$.

(A3) The threshold variable X_t has a continuous density function with a compact support $[-M, M]$, where M is a positive constant.

(A4) $\max_{1 \le t \le n} \| \mathbf{Z}_t \| = o_p(n^{1/2})$ and $E(\| \mathbf{Z} \|^3)$ is bounded.

(A5) Given K and $\boldsymbol{\beta} \neq \mathbf{0}$, there exist a nonnegative definite matrix $\mathbf{G}$ and a full rank matrix $\mathbf{D}$, such that $\lim_{n\to\infty} \mathbf{G}_n = \mathbf{G}$ and $\lim_{n\to\infty} \mathbf{D}_n = \mathbf{D}$.

Assumption (A1) is generally assumed in quantile regression. Assumption (A2) ensures the identifiability of estimation. Assumptions (A3)-(A4) impose some conditions on the threshold variable and other covariates, respectively, which can also be found in Li et al. (2011) and Zhang et al. (2017). Assumptions (A1)-(A4) are used for the proof of consistency of $\widehat{\boldsymbol{\theta}}$ and additional Assumption (A5) suffices for the asymptotical normality. The following theorem demonstrates the limiting distribution of the proposed estimator for $\boldsymbol{\theta}_0$.

Theorem 2.1. *Suppose the true number K_0 of kink points in Model (2.1) is given and Assumptions (A1)-(A5) hold, as $n \to \infty$, we have*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \mathbf{D}^{-1}\mathbf{G}\mathbf{D}^{-1}$.

According to Theorem 2.1, the regression coefficients $\boldsymbol{\eta}$ and the threshold parameters $\boldsymbol{\delta}$ are jointly asymptotically normal with $\sqrt{n}$ convergence rate. In conventional jump threshold model, the threshold parameter estimators converge to a nonstandard asymptotic distribution with n convergence rate, see Hidalgo et al. (2019) for more details.

Moreover, we estimate $\boldsymbol{\Sigma}$ by a plugging estimator $\widehat{\boldsymbol{\Sigma}}_n = \widehat{\mathbf{D}}_n^{-1}\widehat{\mathbf{G}}_n\widehat{\mathbf{D}}_n^{-1}$, where $\widehat{\mathbf{G}}_n = n^{-1}\sum_{t=1}^{n}\tau(1-\tau)h(\mathbf{W}_t;\widehat{\boldsymbol{\theta}})h^{\mathrm{T}}(\mathbf{W}_t;\widehat{\boldsymbol{\theta}})$ and $\widehat{\mathbf{D}}_n = n^{-1}\sum_{t=1}^{n}\hat{f}_t(\hat{e}_t)h(\mathbf{W}_t;\widehat{\boldsymbol{\theta}})h^{\mathrm{T}}(\mathbf{W}_t;\widehat{\boldsymbol{\theta}})$. $\widehat{\mathbf{D}}_n$ requires consistent estimate for conditional density function $f_t(\cdot)$ of error term e_t. We suggest using the method called *Hendricks-Koenker Sandwich* based on the difference quotients discussed by Hendricks and Koenker (1992). To select the bandwidth, two choices are often used. One is based on Edgeworth expansions of studentized quantiles described by Hall and Sheather (1988), the other is based on the minimum of the mean squared error of the density estimator suggested by Bofingeb (1975). In our R package *MultiKink*, we provide both versions to estimate the covariance matrix.

2.2 Parameter Estimation When K is Unknown

The aforementioned BRISQ algorithm works well when the exact number of kink points is given. However, the true number K_0 is usually unknown in practice. To estimate K_0, we first start with a large initial value $K_{\max}(\gg K_0)$ and then iteratively fit the working model (2.4) using the BRISQ algorithm in which we discard all $\widehat{\delta}_k$'s and corresponding $\widetilde{U}_k$'s and $\widetilde{V}_k$'s if $\widehat{\delta}_k \in \mathcal{T}$ at each iteration. When it stops, we obtain an estimator for the number of kink points, denoted by $K_*(< K_{\max})$.

However, we find that K_* often overestimates the true value K_0 in practice. To improve the selection and estimation accuracy, one can evaluate each MKQR model with $k = 0, 1, \cdots, K_*$ kink points according to a prescribed information criterion and find the final model with the smallest information criterion. In our algorithm, we suggest a strengthened quantile Bayesian information criterion (sBIC) to refine the kink points detection,

$$\mathrm{sBIC}(K) = \log\left(S_n(\widehat{\boldsymbol{\theta}}_K)\right) + N_K\frac{\log n}{2n}C_n, \tag{2.5}$$

where $\widehat{\boldsymbol{\theta}}_K$ denotes the estimator of parameters $\boldsymbol{\theta} = (\boldsymbol{\eta}^{\mathrm{T}}, \boldsymbol{\delta}^{\mathrm{T}})^{\mathrm{T}}$ with K kink points, N_K equals to $2 + p + 2K$ and C_n is a positive constant that allows to approach infinity as n increases. When $C_n = 1$, sBIC in (2.5) becomes the standard quantile BIC studied by Lian (2012) for consistent model selection. When $C_n > 1$, it is similar to the modified BIC of Lee et al. (2014). The selection consistency of the sBIC selector will be demonstrated in the next section. The BIC-type criteria have been widely used in model selection. For example, Wang et al. (2007) proved that the BIC tuning parameter selector is able to identify the true linear model consistently. Chen and Chen (2008) further proposed an extended BIC (EBIC) to take into account both the model complexity and the sample size for consistent model

selection. Lee et al. (2014) showed that a modified BIC is consistent in model selection for high dimensional linear quantile regression. Fryzlewicz (2014) proposed a strengthened BIC for sequential change points detection.

We can estimate the number of the kink points by $\widehat{K} = \arg\min_{K=0,1,\cdots,K_\star} \text{sBIC}(K)$. When $K_\star$ is large, it is possible to improve the computational efficiency by using the following backward elimination procedure (Chan et al., 2015). Given K_*, we re-estimate the new MKQR model with $K_* - 1$ kink points using the BRISQ algorithm, and then compare the sBIC values of two models. This procedure is repeated until the sBIC values does not decrease. Then, the final estimators for K_0 and $\boldsymbol{\theta}$ are obtained corresponding to the minimum sBIC. The detailed algorithm is summarized in Algorithm 2. Simulations will illustrate that this algorithm is able to identify the true kink points consistently.

Algorithm 2: Backward Elimination Algorithm for Estimating K.

Step 1. Given $K_{\max}$ initial kink points, repeat Step 2 of Algorithm 1 iteratively and remove any $\widehat{\delta}_k$ at each iteration if $\widehat{\delta}_k \in \mathcal{T}$ until convergence. K_* denotes the resulting estimated number of kink points.

Step 2. Estimate the working model (2.4) with $K_* - 1$ initial kink points using Algorithm 1 to obtain $\widehat{\boldsymbol{\theta}}_{K_*-1}$ and $\text{sBIC}(K_* - 1)$.

Step 3. If $\text{sBIC}(K_* - 1) < \text{sBIC}(K_*)$, then update $K_* = K_* - 1$ and go to Step 2; If $\text{sBIC}(K_* - 1) \geq \text{sBIC}(K_*)$, then stop, set $\widehat{K} = K_*$ and $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{K_*}$. If $K_* = 0$, then stop and there is no kink point.

We then show the selection consistency of $\widehat{K}$ and make the two additional assumptions.

(A6) The matrix $E\{h(\mathbf{W}_t; \boldsymbol{\theta})h^{\tau}(\mathbf{W}_t; \boldsymbol{\theta})\}$ is finite and positive definite.

(A7) $C_n \log(n)/n \to 0$ as $n \to \infty$.

Assumption (A6) is similar to Assumption (A) in Lian (2012). Assumption (A7) requires that $C_n = o(n/\log(n))$ which means C_n cannot diverge too fast to the infinity as n increases to avoid underfitting the true model.

Theorem 2.2. *Under Assumptions (A1) and A(6)-(A7), let* $\widehat{K} = \arg\min_{K=0,\cdots,K_*} sBIC(K)$, *we have* $P(\widehat{K} = K_0) \to 1$ *as* $n \to \infty$.

Theorem 2.2 shows that the quantile sBIC is able to consistently select the true number of kink points. This result plays a fundamental role since the limiting distribution of the parameter estimators $\widehat{\boldsymbol{\theta}}$ is established under the true number of kink points in Theorem 2.1.

3 Statistical Inference

3.1 Testing the Existence of Kink Effects

The kink effect estimation is meaningful if and only if the kink effect truly exists. In this section, we are interested in testing the existence of kink effects in the conditional quantiles. For $\tau \in (0,1)$, we consider the following null (H_0) and alternative (H_1) hypotheses for the quantile regression (2.1),

$$H_0 : \beta_k = 0, \text{ for all } k = 1, \cdots, K. \quad \text{v.s.} \quad H_1 : \beta_k \neq 0, \text{ for some } k = 1, \cdots, K. \tag{3.1}$$

Note that the parameters β_k's depend on τ. Under the null hypothesis, the MKQR model (2.1) degenerates to an ordinary quantile regression without any kink point. Under the alternative hypothesis, there exists at least one statistically significant kink point at the τth quantile. Thus, we suggest the following score-based test statistic based on kink quantile regression with an unknown threshold,

$$T_n(\tau) = \sup_{\delta \in \Gamma} |R_n(\delta)|, \tag{3.2}$$

where $R_n(\delta) = n^{-1/2} \sum_{t=1}^{n} \psi_\tau(Y_t - \widehat{\boldsymbol{\alpha}}^{\mathrm{T}} \mathbf{V}_t)(X_t - \delta) I(X_t \leq \delta)$, $\psi_\tau(u) = \tau - I(u \leq 0)$, δ denotes the location of an unknown threshold, $\mathbf{V}_t = (1, X_t, \mathbf{Z}_t^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{t=1}^{n} \rho_\tau(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t)$. Note that $R_n(\delta)$ is essentially the partial subgradient of the objective function with respect to β_1 evaluated at $\beta_1 = 0$ and $\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}$ up to a constant in the model (2.1) with $K = 1$ *. $T_n(\tau)$ can be viewed as a weighted CUSUM (cumulative-sum) type test statistic based on the signs of quantile residuals. Intuitively, under the null hypothesis, residuals $Y_t - \widehat{\boldsymbol{\alpha}}^{\mathrm{T}} \mathbf{V}_t$ are evenly located below or above zero which result in a relatively small value of $T_n(\tau)$. On the other hand, under the alternative hypothesis, the model is misspecified and the residuals would be consistently positive or negative which implies the large values of $T_n(\tau)$. The idea of subgradient-based tests has been studied in the literature. For example, Qu (2008) constructed the subgradient test statistic in quantile regression for testing the structural changes. Zhang et al. (2014) proposed a score test based on the subgradient to test for the jumping threshold effect in threshold models. Zhang and Li (2017) developed a related test for the continuous threshold effect in asymmetric least square regression.

Theorem 3.1 in the following derives the asymptotic behavior of $T_n(\tau)$. To derive the large-sample property for $T_n(\tau)$, we consider the local alternative model

$$Q_Y(\tau|\mathbf{W}_t) = \alpha_0 + \alpha_1 X_t + n^{-1/2} \beta (X_t - \delta) I(X_t > \delta) + \boldsymbol{\gamma}^{\mathrm{T}} \mathbf{Z}_t, \tag{3.3}$$

*In fact, $R_n(\delta)$ is the partial subgradient of the quantile objective function with respect to β_1 evaluated at $\beta = 0$ and $\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}$ up to a constant for the model $Q_Y(\tau; \boldsymbol{\theta}|\mathbf{W}_t) = \alpha_0 + \alpha_1 X_t + \beta(X_t - \delta_k) I(X_t \leq \delta) + \boldsymbol{\gamma}^{\mathrm{T}} \mathbf{Z}_t$, which is essentially same as the model (2.1) with $K = 1$ after simple reparameterizations.

where δ denotes one kink point location, $\beta \neq 0$ is a nonzero constant and $n^{-1/2}\beta \neq 0$ is the corresponding kink effect to represent the local alternative signal which diverges to zero as the rate of $n^{1/2}$. Model (3.3) implies that there exists at least one kink effect. We also introduce some notations. Define $\mathbf{H}_{1n}(\delta) = n^{-1}\sum_{t=1}^{n} E\{\mathbf{V}_t(X_t-\delta)I(X_t<\delta)f_t(e_t)\}$, $\mathbf{H}_n = n^{-1}\sum_{t=1}^{n} E\{\mathbf{V}_t\mathbf{V}_t^{\mathrm{T}} f_t(e_t)\}$ and $\mathbf{H}_{2n}(\delta,\beta) = n^{-1}\sum_{t=1}^{n} E\{\mathbf{V}_t\beta(X_t-\delta)I(X_t>\delta)f_t(e_t)\}$. We further assume that

(A8) There exists a positive definite matrix $\mathbf{H}$ such that $\lim_{n\to\infty}\mathbf{H}_n = \mathbf{H}$ and another two matrices $\mathbf{H}_1(\delta)$ and $\mathbf{H}_2(\delta,\beta)$ such that $\lim_{n\to\infty}\mathbf{H}_{1n}(\delta) = \mathbf{H}_1(\delta)$ and $\lim_{n\to\infty}\mathbf{H}_{2n}(\delta,\beta) = \mathbf{H}_2(\delta,\beta)$ uniformly hold in δ.

(A9) The density function of e_t, $f_t(\cdot)$ for $t = 1,\cdots,n$, has a bounded first-order derivative.

Theorem 3.1. *Suppose Assumptions (A1) and (A8)-(A9) hold. Under both the null hypothesis H_0 and the local alternative model (3.3), we have*

$$T_n(\tau) \Rightarrow \sup_{\delta} |R(\delta) + q(\delta,\beta)|, \tag{3.4}$$

where "$\Rightarrow$" denotes weak convergence, $R(\delta)$ is a Gaussian process with mean zeros and covariance function $W(\delta,\delta') = \tau(1-\tau)E[\{(X_t-\delta)I(X_t\leq\delta)-\mathbf{H}_1^{T}(\delta)\mathbf{H}^{-1}\mathbf{V}_t\}\{(X_t-\delta')I(X_t\leq\delta')-\mathbf{H}_1^{T}(\delta')\mathbf{H}^{-1}\mathbf{V}_t\}]$ and $q(\delta,\beta) = -\mathbf{H}_1(\delta)\mathbf{H}^{-1}\mathbf{H}_2(\delta,\beta)$.

According to Theorem 3.1, under the null hypothesis, $q(\delta,\beta) = 0$ and $R_n(\delta)$ would converge to a Gaussian process $R(\delta)$ with mean zeros. When the kink effect exists under the alternatives, $q(\delta,\beta)\neq 0$ and $T_n(\tau)$ would be significantly larger than zero. Therefore, large values of $T_n(\tau)$ provide the evidence against the null hypothesis. The score-type statistic is only built on the null hypothesis without fitting models under the alternative hypothesis, so it can also be directly used to test the existence of multiple kink points. Since the asymptotical null distribution of $T_n(\tau)$ is nonstandard, we approximate the P-values using wild bootstrap (Feng et al., 2011). The detailed procedures are relegated to Section S.2 in the Supplement.

3.2 Confidence Intervals for Kink Location Parameters

Next, we construct a test-inversion confidence interval for $\boldsymbol{\delta}$ based on a smoothed rank score test. Consider the following hypotheses for a given $\widetilde{\boldsymbol{\delta}}$ in the domain of X_t and $\tau\in(0,1)$,

$$H_0 : \boldsymbol{\delta} = \widetilde{\boldsymbol{\delta}} \quad \text{v.s.} \quad H_1 : \boldsymbol{\delta} \neq \widetilde{\boldsymbol{\delta}}. \tag{3.5}$$

Under H_0, we can obtain the estimator $\widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}})$ of the regression coefficients $\boldsymbol{\eta}$ given $\widetilde{\boldsymbol{\delta}}$ by fitting the standard linear quantile regression. Muggeo (2017) pointed that naive score statistic in threshold models may lower the test power due to the non-differentiable and non-smooth

nature. To deal with the non-smoothness of the indicator function $I(X_t > \delta_k)$, we use the smoothed Gaussian distribution function $\Phi((X_t - \delta_k)/h_k)$ to approximate $I(X_t > \delta_k)$, where h_k is the bandwidth. The smoothed objective function becomes

$$\widetilde{Q}_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta}|\mathbf{W}_t) = \alpha_0 + \alpha_1 X_t + \sum_{k=1}^{K} \beta_k (X_t - \delta_k)\Phi((X_t - \delta_k)/h_k) + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{Z}_t.$$

Then, we take the first partial derivative of $\widetilde{Q}_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta}|\mathbf{W}_t)$ with respect to $\boldsymbol{\delta}$ evaluated at $\boldsymbol{\delta} = \widetilde{\boldsymbol{\delta}}$ and $\boldsymbol{\eta} = \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}})$, denoted by $\mathbf{P}_t(\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}) = (p_{t1}, \cdots, p_{tK})^{\mathrm{T}}$, where $p_{tk} = -\widehat{\beta}_k \Phi((X_t - \widetilde{\delta}_k)/h_k) - \widehat{\beta}_k (X_t - \widetilde{\delta}_k)\phi((X_t - \widetilde{\delta}_k)/h_k)h_k^{-1}$ and $\phi(\cdot)$ is the first derivative of $\Phi(\cdot)$. Motivated by the rank score tests in Gutenbrunner and Jurečková (1992), Gutenbrunner et al. (1993) and Zhang et al. (2014), we define a smoothed rank score (SRS) test statistic as

$$\mathrm{SRS}_n(\tau) = \mathbf{S}_n^{*\mathrm{T}} \mathbf{V}_n^{-1} \mathbf{S}_n^*, \tag{3.6}$$

where $\mathbf{S}_n^* = n^{-1/2} \sum_{t=1}^n \mathbf{P}_t^*\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}\}\psi_\tau(\tilde{e}_t)$, $\tilde{e}_t = Y_t - Q_Y\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}|\mathbf{W}_t\}$ is the tth residual under H_0, and $\mathbf{V}_n = n^{-1}\sum_{t=1}^n \tau(1-\tau)\mathbf{P}_t^*\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}\}\mathbf{P}_t^*\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}\}^{\mathrm{T}}$. Here, $\mathbf{P}_t^*\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}\}$ is defined as follows. Let $\mathbf{M}_t(\widetilde{\boldsymbol{\delta}}) = (1, X_t, (X_t - \widetilde{\delta}_1)_+, \cdots, (X_t - \widetilde{\delta}_K)_+, \mathbf{Z}_t^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{M}(\widetilde{\boldsymbol{\delta}}) = (\mathbf{M}_1(\widetilde{\boldsymbol{\delta}}), \cdots, \mathbf{M}_n(\widetilde{\boldsymbol{\delta}}))^{\mathrm{T}}$. Let $\mathbf{P} = (\mathbf{P}_1\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}\}, \cdots, \mathbf{P}_n\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}\})^{\mathrm{T}}$, and define $\mathbf{P}^* = (\mathbf{I}_n - \boldsymbol{\Lambda})\mathbf{P}$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix, $\boldsymbol{\Lambda} = \mathbf{M}(\widetilde{\boldsymbol{\delta}})\{\mathbf{M}^{\mathrm{T}}(\widetilde{\boldsymbol{\delta}})\boldsymbol{\Psi}\mathbf{M}(\widetilde{\boldsymbol{\delta}})\}^{-1}\mathbf{M}^{\mathrm{T}}(\widetilde{\boldsymbol{\delta}})\boldsymbol{\Psi}$, $\boldsymbol{\Psi} = \mathrm{diag}(\hat{f}_1(\tilde{e}_1), \cdots, \hat{f}_n(\tilde{e}_n))$. $\mathbf{P}_t^*\{\tau; \widehat{\boldsymbol{\eta}}(\widetilde{\boldsymbol{\delta}}), \widetilde{\boldsymbol{\delta}}\}$ is defined as the tth row of the $n \times K$ matrix $\mathbf{P}^*$ which is considered as the residuals by projecting the partial score vector $\mathbf{P}$ on $\mathbf{M}(\widetilde{\boldsymbol{\delta}})$.

Intuitively, under H_0, the partial scores tend to be zero which implies that the test statistic is relatively small; otherwise, the large test statistic values provide the strong evidence against H_0. The following proposition demonstrates the null asymptotic distribution of $\mathrm{SRS}_n(\tau)$ and its proof is shown in Section S.3 of the Supplement.

Proposition 3.1. *Suppose that Assumptions (A1) and (A6)-(A9) hold. Under the null hypothesis H_0 in (3.5) for any $\tau \in (0,1)$, as $n \to \infty$, we have $SRS_n(\tau) \xrightarrow{d} \chi^2_K$.*

For one kink point δ, the confidence interval can be obtained by inverting the rank score test due to the fact that the test statistic $\mathrm{SRS}_n(\tau)$ is convex in δ. Specially, we first obtain the estimator $\widehat{\delta}$ for δ and then test $H_0 : \delta = \widetilde{\delta}$ for $\widetilde{\delta} = \widehat{\delta} + \varrho$, where ϱ is a small positive increment. If H_0 is not rejected, then increase $\widetilde{\delta}$ by $\widetilde{\delta} = \widetilde{\delta} + \varrho$ and test $H_0 : \delta = \widetilde{\delta}$ again. We repeat the previous testing procedure until H_0 is rejected and set the upper bound of the confident interval for δ as the minimum rejection point, denoted by $\widehat{\delta}_u$. In the similar way, we obtain the lower bound $\widehat{\delta}_l$. Thus, we can obtain a $(1-\alpha)$th confidence interval for δ, $[\widehat{\delta}_l, \widehat{\delta}_u]$. For multiple kink model (2.1), we separately construct the confidence interval for each kink location parameter by controlling other kink estimators. The details are summarized in Section S.3 of the Supplement.

4 Monte Carlo Simulations

4.1 Parameters Estimation

We generate data from the following model

$$Y_t = \alpha_0 + \alpha_1 X_t + \sum_{k=1}^{K} \beta_k (X_t - \delta_k) I(X_t > \delta_k) + \gamma Z_t + \sigma(X_t, Z_t) e_t, \quad t = 1, \ldots, n, \qquad (4.1)$$

where $X_t \sim U(-5, 5)$, $Z_t \sim N(1, 1^2)$, $e_t \sim N(0, 1)$ or t_3 distribution and $\sigma(X_t, Z_t)$ controls the heteroscedasticity. Specially, $\sigma(X_t, Z_t)$ equals to 1 for a homoscedastic model and $1 + 0.2X_t$ for a heteroscedastic model. We set $\alpha_0 = 1, \alpha_1 = 1, \gamma = 1$ and consider three different cases for kink effects: (1) $K = 1$, $\beta_1 = -3$ and $\delta_1 = 0.5$; (2) $K = 2$, $(\beta_1, \beta_2) = (-3, 4)$ and $(\delta_1, \delta_2) = (-1, 2)$; (3) $K = 3$, $(\beta_1, \beta_2, \beta_3) = (-3, 4, -4)$ and $(\delta_1, \delta_2, \delta_3) = (-3, 0, 3)$.

We first check the selection consistency of Theorem 2.2. Since Condition (A3) requires that $C_n \log(n)/n \to 0$ as $n \to \infty$, we consider $C_n = 1, \log(\log(n))$ and $\log(n)$ in the definition of sBIC. We set the sample size $n = 500$. Table 1 reports the percentages of correctly selecting $\widehat{K} = K$ based on 1000 replications under both homoscedastic and heteroscedastic models. All selection rates are very high and close to 100%. It shows that a diverging number for C_n is favorable of identifying the true model when the number of parameters is not fixed. It is in accord with the results of Fryzlewicz (2014) for sequential change points detection. These results validate the selection consistency of Theorem 2.2.

Table 1: The percentages of correctly selecting $\widehat{K} = K$ for $C_n = 1, \log(\log(n)), \log(n)$.

e_t	τ	$K=1$			$K=2$			$K=3$		
		1	log(log(n))	log(n)	1	log(log(n))	log(n)	1	log(log(n))	log(n)
Homoscedasticity										
N	0.3	91.6%	99.8%	100.0%	92.8%	99.1%	99.6%	92.3%	97.5%	98.5%
	0.5	94.2%	100.0%	100.0%	92.4%	99.6%	99.6%	95.6%	99.0%	99.6%
	0.7	91.1%	98.8%	100.0%	91.5%	99.6%	99.6%	94.8%	97.7%	98.9%
t_3	0.3	96.1%	99.8%	99.8%	95.5%	99.8%	99.6%	96.0%	98.3%	97.5%
	0.5	98.0%	100.0%	100.0%	96.8%	100.0%	99.8%	97.8%	99.4%	99.0%
	0.7	96.7%	99.8%	100.0%	95.3%	99.1%	99.8%	94.8%	99.8%	96.6%
Heteroscedasticity										
N	0.3	84.5%	97.3%	100.0%	86.8%	98.3%	99.8%	90.2%	97.7%	97.7%
	0.5	87.2%	96.2%	100.0%	87.0%	98.7%	99.8%	91.3%	98.1%	98.6%
	0.7	83.5%	97.1%	100.0%	83.8%	97.6%	99.8%	89.3%	97.1%	98.8%
t_3	0.3	89.5%	99.8%	99.8%	89.5%	98.0%	99.6%	95.4%	98.8%	99.2%
	0.5	91.6%	98.4%	100.0%	92.1%	98.1%	99.8%	96.1%	98.8%	97.9%
	0.7	91.4%	98.6%	100.0%	91.6%	98.1%	99.4%	94.0%	99.8%	99.8%

Next, we evaluate the finite sample performance of parameter estimators to check the validity of Theorem 2.1. For Case (1) with single kink effect, we compare the proposed

estimation method with the bent line quantile estimators proposed by Li et al. (2011) and the kink regression least squares estimators proposed by Hansen (2017). Both existing methods assume there is only a single kink effect. We denote two methods as SKQR and SKLS, short for Single Kink Quantile Regression and Single Kink Least Square, respectively. We conduct the simulations 500 times and report the estimation biases (Bias), the empirical standard deviations (SD) and the mean square errors (MSE) for each parameter based on 500 estimates as well as their average estimated standard errors (SE) based on the asymptotical variance in Theorem 2.1. All simulation results are summarized in Table 2. All estimates have ignorable biases, and the standard deviations (SD) are close to the estimated standard errors (SE). In the homoscedastic model with normal random errors, the SKLS estimators have smaller mean square errors (MSE) than others. However, when the model is heteroscedastic or the errors follow t_3 distribution, our method works better than the other two in terms of MSE, which demonstrate the robustness and efficiency of the proposed estimators.

Table 2: Estimation results (multiplied by a factor of 10) of three methods for Case (1).

e_t			Homoscedasticity					Heteroscedasticity				
			α_0	α_1	β_1	γ	δ	α_0	α_1	β_1	γ	δ
N	SKQR	Bias	-0.020	-0.014	-0.002	-0.058	0.089	0.054	0.001	-0.042	0.022	-0.002
		SD	1.481	0.480	0.572	0.795	0.875	1.081	0.236	0.246	1.086	0.888
		SE	1.415	0.470	0.576	0.808	0.752	1.110	0.272	0.368	1.018	0.829
		MSE	0.218	0.023	0.033	0.063	0.077	0.117	0.006	0.006	0.117	0.078
	SKLS	Bias	0.068	-0.017	-0.009	-0.028	0.019	0.096	-0.118	-0.003	-0.005	0.036
		SD	1.124	0.444	0.358	0.612	0.669	1.074	0.519	0.261	0.863	0.683
		SE	1.132	0.444	0.377	0.639	0.599	1.056	0.508	0.264	0.849	0.692
		MSE	**0.126**	**0.020**	**0.013**	**0.037**	**0.045**	0.116	0.028	0.007	**0.074**	**0.047**
	MKQR	Bias	0.051	-0.010	-0.031	-0.052	0.114	0.077	-0.043	0.007	0.011	-0.014
		SD	1.482	0.577	0.475	0.779	0.822	1.036	0.243	0.226	1.075	0.843
		SE	1.435	0.555	0.480	0.813	0.758	0.998	0.234	0.219	1.022	0.818
		MSE	0.219	0.033	0.023	0.061	0.069	**0.107**	**0.006**	**0.005**	0.115	0.071
t_3	SKQR	Bias	0.109	0.021	-0.059	0.095	-0.101	0.064	0.017	0.001	-0.026	-0.097
		SD	1.539	0.481	0.707	0.784	0.909	1.173	0.268	0.303	1.165	1.024
		SE	1.606	0.543	0.641	0.923	0.865	1.304	0.335	0.478	1.152	0.936
		MSE	0.237	**0.023**	0.050	**0.062**	0.083	0.137	0.007	0.009	0.135	0.105
	SKLS	Bias	0.025	-0.048	0.019	0.033	-0.064	-0.058	-0.004	-0.013	0.104	-0.083
		SD	1.907	0.766	0.613	1.037	1.029	1.887	0.857	0.481	1.487	1.322
		SE	1.893	0.741	0.628	1.084	1.010	1.764	0.852	0.440	1.449	1.170
		MSE	0.362	0.059	0.037	0.107	0.106	0.355	0.073	0.023	0.221	0.175
	MKQR	Bias	0.063	-0.055	0.009	0.105	-0.086	0.087	-0.001	0.023	-0.054	-0.080
		SD	1.495	0.706	0.474	0.801	0.891	1.138	0.307	0.256	1.181	0.962
		SE	1.560	0.607	0.526	0.892	0.841	1.080	0.264	0.239	1.125	0.889
		MSE	**0.223**	0.050	**0.022**	0.065	**0.080**	**0.130**	**0.007**	**0.007**	**0.135**	**0.093**

Bias: the empirical bias; SD: the empirical standard deviation; MSE: the mean square error; SE: the average estimated standard error. The minimum MSE among three estimators is highlighted in bold.

For the multi-kink models, both SKLS and SKQR methods are not able to detect the multiple kink points. Thus, we only present the simulation results of the proposed MKQR estimators for Case (2) with $K = 2$ in Table 3. All the biases are sufficiently close to zero

and the SEs are compatible with the SDs for both homoscedastic and heteroscedastic errors. To save the space, we omit the similar simulation results for Case (3) with $K = 3$. These results demonstrate the validity of Theorem 2.1 for the multiple kink effects.

4.2 Power Analysis

We now assess the power performances for testing the existence of kink effects in Section 3.1. We generate the data from Case (1) except $\beta = n^{-1/2}c$, where $n = 1000$, $c = 0, 2, 4, 6, 8, 10$ and $c = 0$ corresponds to the null hypothesis. We compare our proposed test with two existing tests, the lack-of-fit (L.O.F) test proposed by He and Zhu (2003) and the F-type test proposed by Hansen (2017). The lack-of-fit test is a general test for checking model specification, which was also used in Li et al. (2011). For our score-based test, we compute the P-values using wild bootstrap in Algorithm **??** with 300 replicates. Figure 2 displays the power curves of three tests over different signal strength values of c. Under the null hypothesis when $c = 0$, all methods have satisfactory type I errors close to the nominal significance level 5% for homoscedastic errors. However, the L.O.F test can not control the type I errors when there exists heteroscedasticity. As c increases, i.e. the kink effect gets enhanced, the empirical powers to identify the kink effect for all methods gradually increase to one for each scenario. Our proposed test have the higher empirical powers than the other two tests, especially when the errors follow the t_3 distribution or exist the heteroscedasticity.

4.3 Confidence Intervals

Next, we evaluate the finite sample performances of the smoothed rank score (SRS) test-inversion confidence intervals (CIs) for the kink location parameter $\boldsymbol{\delta}$. For a comparison purpose, we additionally consider the Wald-type CIs and the bootstrap CIs. The Wald-type $(1-\alpha)$th CIs are constructed based on the asymptotical normality in Theorem 2.1, i.e., $\widehat{\delta}_k \pm z_{\alpha/2}\mathrm{SE}(\widehat{\delta}_k)$, for $k = 1, \cdots, \widehat{K}$, where $z_{\alpha/2}$ is the $\alpha/2$ upper tailed critical value of the standard normal distribution and $\mathrm{SE}(\widehat{\delta}_k)$ is the estimated standard error of $\widehat{\delta}_k$. The Bootstrap CIs are defined as $[\widehat{\delta}^*_{k,\alpha/2}, \widehat{\delta}^*_{k,1-\alpha/2}]$, the $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of bootstrap estimators $\{\widehat{\delta}^*_{k,b}, b = 1, 2, \cdots, B\}$ with B paired bootstrap samples.

We generate data from the MKQR model of Case (2) with $K = 2$ and $n = 500$. To save the space, we only report the simulation results for the heteroscedastic model with the errors from t_3 distribution. Table 4 reports the coverage probabilities and the mean width of 95% confidence intervals as well as the average running time per replication at different quantile levels $\tau = 0.3, 0.5, 0.8$ based on 1000 simulations.

From Table 4, the coverage probabilities of Wald-type intervals are generally lower than the 95% nominal level. Hansen (2017) and Fong et al. (2017) also found that Wald-type

Table 3: Estimation results (multiplied by a factor of 10) of the proposed MKQR method for Case (2) with $K = 2$.

e_t	τ		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
					Homoscedasticity				
	0.2	Bias	0.001	0.001	0.001	0.000	0.000	0.000	0.001
		SD	0.067	0.020	0.014	0.037	0.044	0.024	0.019
		SE	0.067	0.020	0.014	0.038	0.045	0.024	0.019
		MSE	0.005	0.000	0.000	0.001	0.002	0.001	0.000
	0.5	Bias	-0.001	0.000	0.000	-0.001	0.002	0.000	-0.001
		SD	0.061	0.019	0.014	0.035	0.042	0.022	0.019
		SE	0.065	0.019	0.017	0.043	0.049	0.026	0.022
$N(0,1)$		MSE	0.004	0.000	0.000	0.001	0.002	0.000	0.000
					Heteroscedasticity				
	0.2	Bias	0.002	0.001	0.000	0.001	0.000	-0.001	0.001
		SD	0.032	0.007	0.007	0.035	0.063	0.019	0.027
		SE	0.032	0.007	0.006	0.035	0.064	0.019	0.027
		MSE	0.001	0.000	0.000	0.001	0.004	0.000	0.001
	0.5	Bias	-0.001	0.000	0.000	0.000	0.000	-0.001	-0.002
		SD	0.030	0.007	0.006	0.032	0.059	0.018	0.026
		SE	0.030	0.006	0.006	0.033	0.060	0.018	0.025
		MSE	0.001	0.000	0.000	0.001	0.003	0.000	0.001
					Homoscedasticity				
	0.2	Bias	0.007	0.001	-0.002	-0.017	0.021	0.005	-0.003
		SD	0.333	0.101	0.072	0.175	0.227	0.132	0.097
		SE	0.318	0.097	0.067	0.181	0.216	0.114	0.093
		MSE	0.111	0.010	0.005	0.031	0.052	0.017	0.009
	0.5	Bias	-0.010	-0.001	0.002	-0.007	0.013	0.004	-0.003
		SD	0.299	0.092	0.060	0.161	0.196	0.119	0.089
		SE	0.278	0.085	0.060	0.158	0.188	0.101	0.081
		MSE	0.089	0.009	0.004	0.026	0.038	0.014	0.008
t_3					Heteroscedasticity				
	0.2	Bias	0.009	0.003	-0.001	-0.016	0.035	0.001	0.002
		SD	0.147	0.032	0.033	0.174	0.320	0.098	0.142
		SE	0.152	0.033	0.030	0.168	0.308	0.089	0.130
		MSE	0.022	0.001	0.001	0.030	0.104	0.010	0.020
	0.5	Bias	-0.005	-0.001	0.000	-0.005	0.024	0.000	0.001
		SD	0.135	0.030	0.026	0.155	0.278	0.087	0.133
		SE	0.131	0.028	0.026	0.148	0.269	0.080	0.113
		MSE	0.018	0.001	0.001	0.024	0.078	0.008	0.018

Bias: the empirical bias; SD: the empirical standard deviation; MSE: the mean square error; SE: the average estimated standard error.

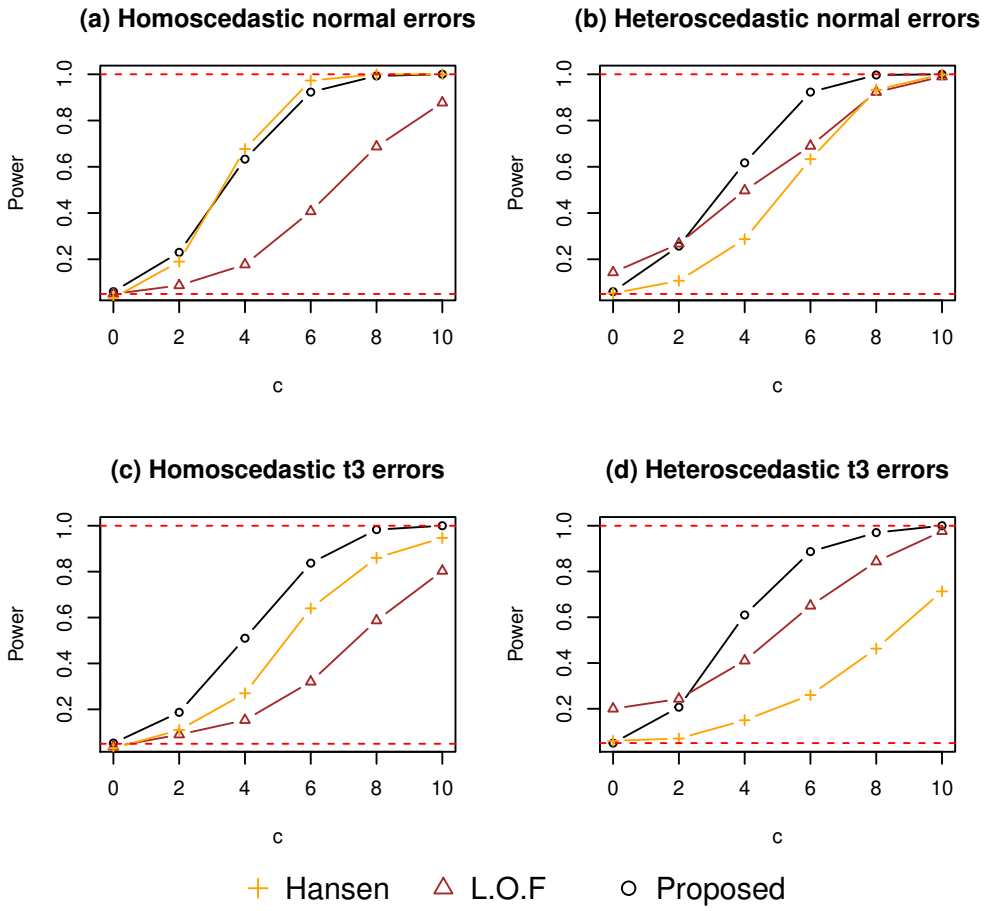

Figure 2: Power comparison of the proposed test at $\tau = 0.5$ (black circle), the lack-of-fit test at $\tau = 0.5$ (brown triangle) and the F-type test (orange plus) at the significance level 5%.

CIs have poor finite sample performance, especially for threshold parameters due to the parameter-effects curvature. The bootstrap intervals have the highest coverage rates, but they have the largest interval lengths and need much more computing time. The bootstrap method is less computationally efficient. The proposed smoothed rank score (SRS) test-inversion CIs provide a balance between the estimation accuracy and the computation efficiency. They have higher coverage probabilities than the Wald-type intervals and also need much less computing time than the Bootstrap intervals.

5 Empirical Analysis

5.1 Secondary Industrial Structure of China

The past few decades have witnessed the miracle of China's economic growth. Since China introduced the policy of reform and opening in 1978, GDP per capita has experienced a

Table 4: 95% confidence intervals for each kink point parameter in Case (2) in the heteroscedastic model with the t_3 errors.

τ	Type	Coverage probability		Mean interval length		Time(s)
		δ_1	δ_2	δ_1	δ_2	
0.3	Wald	0.933	0.923	0.353	0.518	5.07
	Boot	0.964	0.970	0.391	0.689	377.81
	Score	0.930	0.957	0.343	0.641	12.70
0.5	Wald	0.923	0.933	0.307	0.436	4.45
	Boot	0.968	0.982	0.323	0.575	375.07
	Score	0.927	0.953	0.303	0.530	11.38
0.8	Wald	0.913	0.883	0.451	0.619	4.53
	Boot	0.970	0.974	0.501	0.801	378.38
	Score	0.917	0.930	0.449	0.774	13.82

Wald: Wald-type CIs; Boot: bootstrap CIs; Score: SRS test-inversion CIs. Time is the average running time for one simulation.

considerable growth and the industrial structure has also undergone tremendous changes. Classical development economic theory tells us that in the process of development, the proportion of first industry decreases while the tertiary industry instead increases gradually for one country. Meanwhile, the proportion of secondary industry experiences a process of increasing rapidly at first and then gradually stops growing or even decreases, which implies the presence of a kink pattern. The economic development model of China, as the biggest developing country in the world, has been aroused a great of research interest, see Song et al. (2011), Brandt et al. (2013), Cao and Birchenall (2013) and etc.

In this section, we aim to investigate whether there exist kink effects between secondary industrial structure and the economic growth from the quantile regression perspective using the prefecture-level cities data in China. After removing the missing values, we collect data for 280 Chinese prefecture-level cities of year 2016 from the Organisation for Economic Co-operation and Development (OECD) database available at `https://insights.ceicdata.com/`. We consider the MKQR model

$$Q_Y(\tau|X_t, \mathbf{Z}_t) = \alpha_0 + \alpha_1 X_t + \sum_{k=1}^{K} \beta_k (X_t - \delta_k) I(X_t > \delta_k) + \boldsymbol{\gamma}^{\mathrm{T}} \mathbf{Z}_t, \quad t = 1, \cdots, 280, \qquad (5.1)$$

where Y_t represents the proportion of secondary industry of the tth city, X_t is the GDP per capita (10^4 Chinese Yuan) and $\mathbf{Z}_t$ includes the fiscal expenditure (FE) and fixed assets investment (FAI), which are generally deemed to be correlated with the industrial structure. To eliminate effect by the difference of economic scales, we divide the FE and FAI by the total GDP for each city, denoted by Z_{t1} and Z_{t2}, respectively*. We let $\tau = 0.1, 0.3, 0.5, 0.7$ and 0.9 to study the prefectural-level cities at different development levels.

*We also separately test the existence of kink effects between Y_t and Z_{t1}, Y_t and Z_{t2} at different quantiles. The resulting p-values are all greater than 0.1 across all quantiles indicating no kink effect on FE and FAI.

Table 5: Parameter estimation and test results of the MKQR model at different quantile levels for secondary industrial structure data of China.

	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
P-values	0.000	0.000	0.000	0.000	0.000
$\widehat{K}$	1	1	1	1	1
$\widehat{\alpha}_0$	$-0.012_{(0.067)}$	$0.184_{(0.057)}$	$0.090_{(0.057)}$	$0.182_{(0.057)}$	$0.273_{(0.054)}$
$\widehat{\alpha}_1$	$0.114_{(0.022)}$	$0.059_{(0.014)}$	$0.102_{(0.019)}$	$0.084_{(0.017)}$	$0.074_{(0.015)}$
$\widehat{\beta}_1$	$-0.109_{(0.023)}$	$-0.054_{(0.014)}$	$-0.098_{(0.019)}$	$-0.079_{(0.017)}$	$-0.068_{(0.015)}$
$\widehat{\delta}_1$	$3.457_{(0.307)}$	$4.490_{(0.355)}$	$3.468_{(0.227)}$	$3.571_{(0.191)}$	$3.777_{(0.301)}$
Wald	[2.856, 4.058]	[3.794, 5.184]	[3.023, 3.913]	[3.195, 3.946]	[3.187, 4.367]
Boot	[2.567, 7.377]	[2.729, 4.831]	[2.599, 4.801]	[3.027, 4.834]	[2.782, 6.552]
Score	[2.537, 4.561]	[2.786, 4.808]	[3.196, 4.967]	[2.651, 5.410]	[3.236, 4.409]
$\widehat{\gamma}_1$	$-0.985_{(0.405)}$	$-0.916_{(0.229)}$	$-0.711_{(0.159)}$	$-0.828_{(0.199)}$	$-0.687_{(0.252)}$
$\widehat{\gamma}_2$	$0.078_{(0.024)}$	$0.086_{(0.020)}$	$0.090_{(0.015)}$	$0.098_{(0.014)}$	$0.060_{(0.012)}$

The figures in parentheses denote the standard errors of estimators.

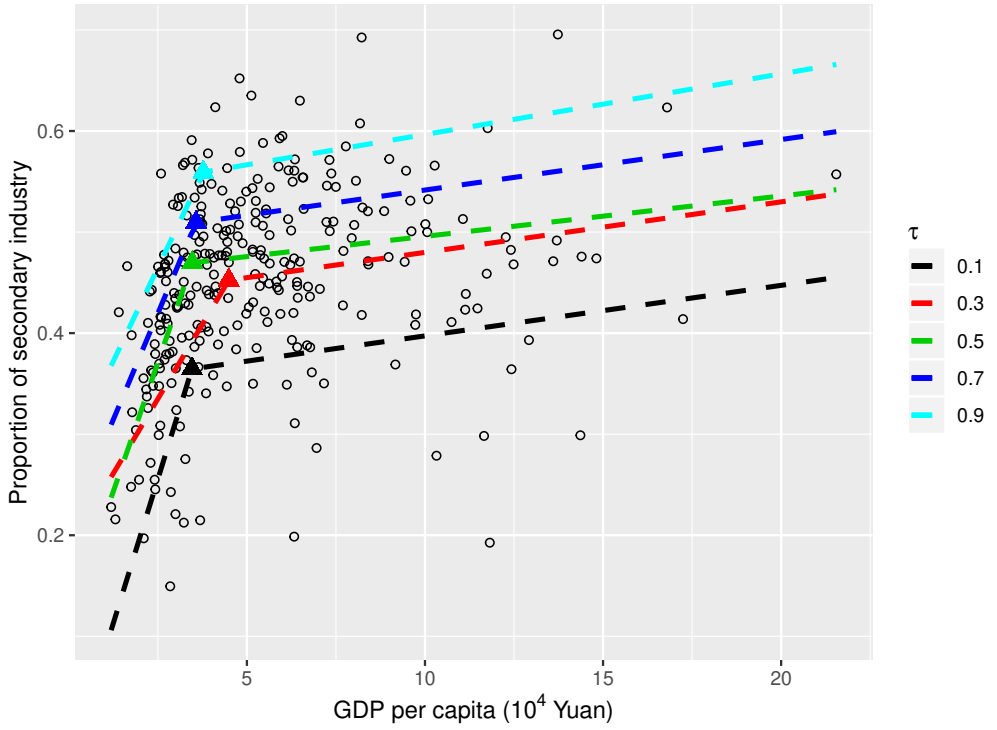

Figure 3: Scatter plot between secondary industrial structure of China and GDP per capita with the fitted MKQR curves at different quantile levels. ▲ denotes the estimated kink point.

Table 5 reports P-values for testing the existence of kink effects based on 1000 bootstrap replicates, the estimated number of kink points, the estimated parameters as well as their the standard errors, and the confidence intervals for kink locations. Figure 3 displays the scatter plot between secondary industrial proportions of 280 cities in China and their GDP per capita

with the fitted MKQR curves at different quantile levels. According to Table 5, P-values are 0 and $\widehat{K} = 1$ for all different quantiles, which indicate that there exists a significant kink point. $\widehat{\alpha}_1 > 0$ and $\widehat{\beta}_1 < 0$ for all quantiles are statistically significant which means that the second industrial proportions Y_t first quickly increase with GDP per capita and then stabilizes with a slow increasing rate of $\widehat{\beta}_1 + \widehat{\alpha}_1$ (e.g. $\widehat{\beta}_1 + \widehat{\alpha}_1 = 0.004$ for $\tau = 0.5$). This empirical finding demonstrates the classical economic theory about the process of development. It is also of interest to observe that the estimated kink points are around 35000 to 45000 Chinese Yuan (roughly 5000-6500 United States Dollar). Based on the Chenery industrialization stage theory (Chenery et al., 1986)[†], GDP per capita in this interval indicates that an economic entity is going through an important turning period. During this period, if one economy can skip the threshold value and achieve economic restructuring, it will move into high-income group. Otherwise, the middle-income trap may loom. In addition, both regressors Z_{t1} and Z_{t2} are statistically significant based on the Wald-type test. It is confirmed that the proportions of secondary industry are indeed correlated to the government fiscal expenditure and the fixed assert investment.

5.2 Triceps Skinfold Thickness of Gambian Females

Triceps skinfold thickness (TSF) as an important measure for body density experiences the dynamic changes with the increase of age. People whose TSFs are above the 85th percentile are more likely to suffer from obesity, while those whose TSFs are lower the 20th percentile are usually skinny. Exploring the relationship between TSF and age at different quantiles has been of great interest in biological and human health studies. For instance, Durnin and Womersley (1974) divided the 481 subjects aged from 16-72 into four subgroups based on the ages and used the linear regression to fit the logarithm of TSF and body densities for each subsample. The results showed that the regression coefficients of each group exhibited significant differences from the others. Cole and Green (1992) demonstrated that there existed cubic splines non-linear pattern between the logarithm of TSF and age by using the smooth fitting curves. Perperoglou et al. (2019) fitted the Gambian females dataset using the spline regression to depict the nonlinearity between TSF and the age. Although the spline regression captures the nonlinear trend, it does not provide any information concerning thresholds and is lack of interpretability in each segment. The nonparametric spline method is either not robust to the outliers and heavy-tailed data.

We consider the dataset collected by Royston and Sauerbrei (2008) from an anthropometry survey at three Gambian villages in 1989, containing 892 women between the ages of 0 and 55. To investigate the relationship between their TSFs and the age and identify the

[†] Professor Hollis B. Chenery at Harvard University believed that modern economic growth can be understood as a comprehensive transformation of the economic structure. He divided the structural transformation process of GDP per capita into three stages: Initial, Intermediate and Post-industrial stages, corresponding to the GDP per capita less than 1495 dollars, 1495-11214 dollars and greater than 11214 dollars.

potential kink points at different quantiles, the following MKQR model is considered

$$Q_Y(\tau|X_t) = \alpha_0 + \alpha_1 X_t + \sum_{k=1}^{K} \beta_k (X_t - \delta_k) I(X_t > \delta_k), \quad t = 1, \cdots, 892, \tag{5.2}$$

where Y_t is log(TSF), X_t is the age, K is the unknown number of kink points. We set $\tau = 0.1, 0.3, 0.5, 0.7$ and 0.9 to study the different conditional quantiles of log(TSF) on age.

Table 6 reports P-values for testing the existence of kink effects based on 1000 bootstrap replicates, the estimated number of kink points, the estimated parameters as well as their the standard errors, and the confidence intervals for kink locations. The resulting P-values are all close to zeros, implying that log(TSF) has significant kink effects on the age for all quantiles. We estimate the MKQR models at different quantiles by setting 10 initial kink points and identify $\widehat{K} = 2$ kink points located round 10 years and 20 years. This result is in accord with the biological intuition. Two kink points split the domain of the age into the three growth periods of human beings: childhood, adolescence and adults. Figure 4 also displays the scatter plot between log(TSF) and their age with the fitted MKQR curves at different quantile levels. One can observe that the logarithm of TSF decreases quickly with the age in the childhood up to about 8-11 years old, then experiences a growth spurt at adolescence up to about 18-21 years old and finally stays almost stable after then for adults. The variance of TSF increases with the age, which makes quantile regression necessary to handle with the heteroscedasticity. It is also interesting to notice that the kink points estimators are heterogeneous across different quantiles. For the higher quantiles such as $\tau = 0.9$, log(TSF) tend to experience the smaller kink points locations than other quantiles, for example $\widehat{\delta}_1 = 8.604$ years for $\tau = 0.9$. It means that Gambian females with obesity reach the biological limits earlier, making their TSFs get changed sooner in the growth process.

As a comparison, we also analyze the dataset by using SKLS and SKQR methods. Both methods can only detect with a single kink point at around 6-8 years, which is much lower than our first threshold estimator $\widehat{\delta}_1$. However, if the Wald-type test of Li et al. (2011) and the F-type test of Hansen (2017) are used to check the existence of kink effects using subsmaple in the second segment divided by the threshold estimator, we find that both tests reject the null hypothesis indicating that some potential kink effect is ignored. In contrast, our MKQR method is flexible and robust in practice to capture multi-kink effects.

6 Conclusion

In this article, we studied the flexible multi-kink quantile regression (MKQR) model without knowing the number of kink points. It is robust to outliers and heavy-tailed errors and more flexible for modelling data with heterogeneous conditional distributions. We proposed a BRISQ algorithm for estimating parameters. It is much more computationally efficient and

Table 6: Parameter estimation and test results of the MKQR model at different quantile levels for triceps skinkfold thickness data for Gambian females.

	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
P-values	0.000	0.000	0.000	0.000	0.007
$\widehat{K}$	2	2	2	2	2
$\widehat{\alpha}_0$	$1.895_{(0.021)}$	$2.042_{(0.029)}$	$2.183_{(0.027)}$	$2.241_{(0.023)}$	$2.426_{(0.037)}$
$\widehat{\alpha}_1$	$-0.041_{(0.002)}$	$-0.040_{(0.005)}$	$-0.046_{(0.005)}$	$-0.037_{(0.003)}$	$-0.049_{(0.008)}$
$\widehat{\beta}_1$	$0.096_{(0.013)}$	$0.106_{(0.010)}$	$0.129_{(0.010)}$	$0.136_{(0.012)}$	$0.144_{(0.015)}$
$\widehat{\beta}_2$	$-0.056_{(0.014)}$	$-0.058_{(0.010)}$	$-0.075_{(0.009)}$	$-0.090_{(0.012)}$	$-0.086_{(0.013)}$
$\widehat{\delta}_1$	$10.035_{(0.130)}$	$10.117_{(0.379)}$	$10.030_{(0.306)}$	$10.635_{(0.425)}$	$8.604_{(0.472)}$
Wald	[9.781, 10.290]	[9.373, 10.861]	[9.430, 10.630]	[9.803, 11.467]	[7.679, 9.530]
Boot	[8.206, 12.988]	[9.086, 12.235]	[9.425, 12.050]	[8.223, 12.202]	[7.737, 10.679]
Score	[8,217, 13.930]	[8.979, 12.393]	[9.418, 12.478]	[7.663, 12.758]	[7.665, 10.967]
$\widehat{\delta}_2$	$20.414_{(2.927)}$	$19.689_{(1.525)}$	$18.993_{(1.048)}$	$18.964_{(0.845)}$	$18.720_{(1.489)}$
Wald	[14.678, 26.150]	[16.700, 22.678]	[16.939, 21.047]	[17.307, 20.621]	[15.801, 21.639]
Boot	[14.530, 47.470]	[17.280, 29.735]	[17.562, 23.226]	[18.067, 24.724]	[16.588, 24.821]
Score	[17.487, 49.680]	[16.639, 42.566]	[17.945, 25.282]	[18.119, 25.728]	[15.742, 26.166]

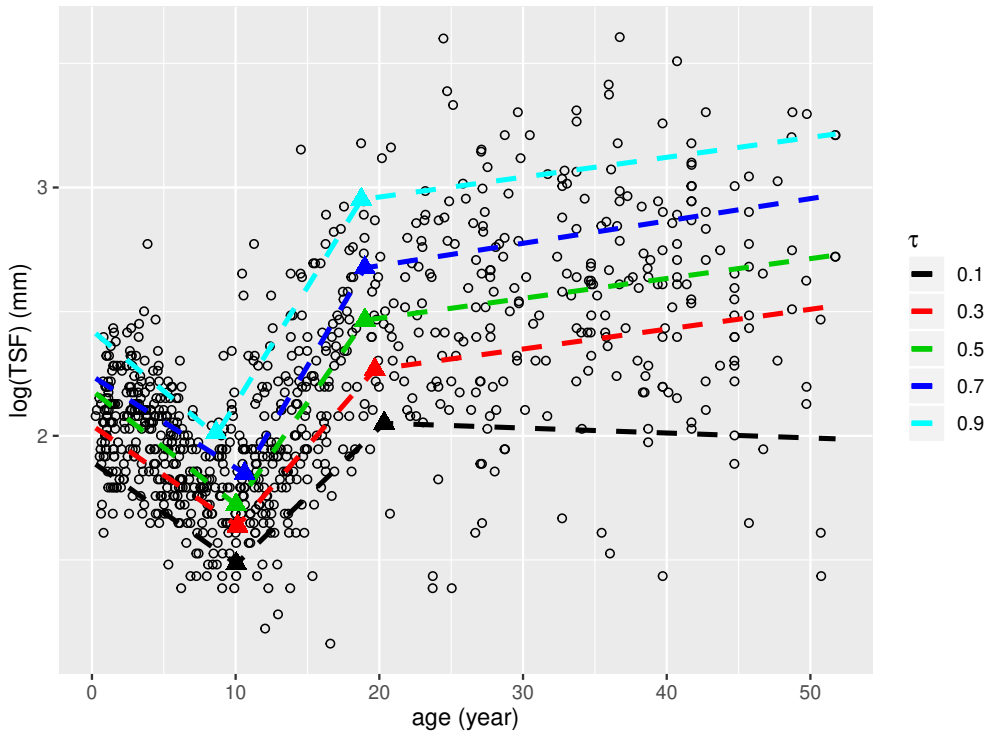

Figure 4: Scatter plot between the logarithm of TSF and ages for Gambian females with the fitted MKQR curves at different quantile levels. ▲ denotes the estimated kink point.

not sensitive to the initial values. The selection consistency and the asymptotic normality were established and the statistical inference for kink effects were also developed. A R package *MultiKink* has been developed for all the estimation and inference procedures. In

the future studies, several topics could be considered. It is interesting to test whether the regression curve is continuous or not at the change points and how to estimate the parameters if jumps happen at the change points. We may also consider some additional restrictions on the regression coefficients such as nondecreasing slopes in the MKQR model. We can also consider the composite quantile regression idea to enhance the estimation efficiency to identify the kink points under the commonality assumption of kink points across multiple quantile levels. Besides, extensions to other regressions such as generalized linear models, Cox proportional hazards models or censored models can be also considered for the future research.

Supplementary Material. The convergence of the BRISQ Algorithm, the Wild bootstrap algorithm for P-Values, the proof of Proposition 3.1 for the null asymptotic distribution of the smoothed rank score (SRS) test statistic and additional simulation results are included in a separate online supplemental file.

Acknowledgement. We thank the Editor, the Associate Editor and two referees for their encouragements and insightful comments which have substantially improved the paper. Zhong's research was supported by the National Natural Science Foundation of China (11922117), Fujian Provincial Science Fund for Distinguished Young Scholars (2019J06004), Basic Scientific Center Project of NNSFC (71988101) and the 111 Project (B13028).

7 Appendix

A.1 Proof of Theorem 2.1

To show the asymptotic normality of $\widehat{\boldsymbol{\theta}}$, we need derive its consistency at first.

Lemma A.1. *Under Assumptions (A1)-(A4),* $\widehat{\boldsymbol{\theta}}$ *is a consistent estimator of* $\boldsymbol{\theta}_0$.

Proof of Lemma A.1: We first need to show that $\sup_{\boldsymbol{\theta}\in\Theta} |S_n(\boldsymbol{\theta}) - S(\boldsymbol{\theta})| \xrightarrow{p} 0$ as $n \to \infty$. Notice that $S(\boldsymbol{\theta})$ is continuous and has the following first derivative

$$\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathrm{E}\left[\psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta})\right].$$

By the Assumptions (A3) and (A4), we can get that $\mathrm{E}\sup_{\boldsymbol{\theta}} |h(\mathbf{W}_t;\boldsymbol{\theta})| < \infty$. Together with $\psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\} \le \max(\tau, 1-\tau)$, we can show that $\mathrm{E}\sup_{\boldsymbol{\theta}\in\Theta} \psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta})$

is finite. By using the mean-value theorem, for any $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2 \in \Theta$, there exists a $\boldsymbol{\theta}^*$ such that

$$S_n(\boldsymbol{\theta}^1) - S_n(\boldsymbol{\theta}^2) = \frac{1}{n}\sum_{t=1}^{n} [\psi_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}^*|\mathbf{W}_t)\}h(\mathbf{W}_t; \boldsymbol{\theta}^*)]^{\mathrm{T}} (\boldsymbol{\theta}^1 - \boldsymbol{\theta}^2)$$

By using Assumptions (A3) and (A4) again,

$$\mathrm{E}\Big|\frac{1}{n}\sum_{t=1}^{n} \psi_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}^*|\mathbf{W}_t)\}h(\mathbf{W}_t; \boldsymbol{\theta}^*)\Big| \leq B_n < \infty,$$

where $B_n = \mathrm{E}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\max(\tau, 1-\tau)h(\mathbf{W}_t; \boldsymbol{\theta})\Big|$. Hence $B_n = O_p(1)$ and $|S_n(\boldsymbol{\theta}^1) - S_n(\boldsymbol{\theta}^2)| \leq B_n\|\boldsymbol{\theta}^1 - \boldsymbol{\theta}^2\|$ for every $\mathcal{X}_n$. By applying the Lemma 2.9 of Newey and McFadden (1994), we have $\sup_{\boldsymbol{\theta}\in\Theta}|S_n(\boldsymbol{\theta}) - S(\boldsymbol{\theta})| \xrightarrow{p} 0$ for $\boldsymbol{\theta} \in \Theta$.

Since $S_n(\boldsymbol{\theta})$ is continuous w.r.t $\boldsymbol{\theta}$, and $S(\boldsymbol{\theta})$ uniquely reaches its global minimum at $\boldsymbol{\theta}_0$ (Assumption (A4)), together with $\sup_{\boldsymbol{\theta}\in\Theta}|S_n(\boldsymbol{\theta}) - S(\boldsymbol{\theta})| \xrightarrow{p} 0$, then we can immediately induce that $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ as $n \to \infty$ by using the Theorem 2.1 of Newey and McFadden (1994).■

The following lemma is sufficient for deriving the Bahadur representation of $\widehat{\boldsymbol{\theta}}$.

Lemma A.2. *Suppose Assumptions (A1) and (A3)-(A4) hold, for any positive sequence d_n converging to zero, we have*

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq d_n} \Bigg|\Bigg| n^{-1/2}\sum_{t=1}^{n}[\psi_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t; \boldsymbol{\theta}) - \psi_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_0|\mathbf{W}_t)\}$$
$$\times h(\mathbf{W}_t; \boldsymbol{\theta}_0)] - n^{-1/2}E\left[\sum_{t=1}^{n}\psi_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t; \boldsymbol{\theta})\right]\Bigg|\Bigg| = o_p(1)$$

Proof of Lemma A.2: Define

$$\begin{aligned} u_t(\boldsymbol{\theta}, \boldsymbol{\theta}_0) &= \sum_{k=1}^{K_0+1} [\psi_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t; \boldsymbol{\theta}) - \psi_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_0|\mathbf{W}_t)\} \\ &\quad h(\mathbf{W}_t; \boldsymbol{\theta}_0)] \cdot I(\delta_{k-1,0} < X_t < \delta_{k,0}) \\ &= \sum_{k=1}^{K_0+1} u_{t,k}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \end{aligned}$$

For any $\delta_k \in (\delta_{k-1,0}, \delta_{k+1,0})$, $u_{t,k}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ can be partitioned into several parts based on the

range of X_t,

$$\begin{aligned} u_{t,k}(\boldsymbol{\theta},\boldsymbol{\theta}_0) &= u_{t,k}(\boldsymbol{\theta},\boldsymbol{\theta}_0)I\{\max(\delta_k,\delta_{k,0}) < X_t < \delta_{k+1,0}\} + u_{t,k}(\boldsymbol{\theta},\boldsymbol{\theta}_0)\cdot \\ &\quad I\{\delta_{k-1,0} < X_t \le \min(\delta_k,\delta_{k,0})\} + u_{t,k}(\boldsymbol{\theta},\boldsymbol{\theta}_0)I(\delta_k \le X_t < \delta_{k,0}) \\ &\quad + u_{t,k}(\boldsymbol{\theta},\boldsymbol{\theta}_0)I(\delta_{k,0} \le X_t < \delta_k) \\ &= u_{t,k,1}(\boldsymbol{\theta},\boldsymbol{\theta}_0) + u_{t,k,2}(\boldsymbol{\theta},\boldsymbol{\theta}_0) + u_{t,k,3}(\boldsymbol{\theta},\boldsymbol{\theta}_0) + u_{t,k,4}(\boldsymbol{\theta},\boldsymbol{\theta}_0) \end{aligned}$$

To prove Lemma A.2, it is sufficient to show $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le d_n} \|n^{-1/2}\sum_{t=1}^n [u_{t,k,j} - E(u_{t,k,j})]\| = o_p(1)$ for $k = 1,\cdots,K_0$ and $j = 1,\cdots,4$. The proofs directly follow from the result of Lemma 4.6 in He and Shao (1996). We only take $u_{t,k,1}(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ for illustration and the remaining are the same. For this, we need to check the conditions (B1), (B3) and (B5′) in He and Shao (1996).

The measurability is straightforward for (B1). For (B3), by using mean-value theorem , we have

$$E_{Y_t}[\|u_{t,k,1}(\boldsymbol{\theta},\boldsymbol{\theta}_0)\|^2|\mathbf{W}_t] \le L d_n f_t^* \|\mathbf{U}_t\|^3,$$

where L is some positive constant, $\mathbf{U}_t = (1, X_t, \mathbf{Z}_t^{\mathrm{T}})^{\mathrm{T}}$ and f_t^* is some intermediate density satisfying $f_t^* \to f_t(0)$ almost surely when $n \to \infty$. It is obvious to obtain (B3). For (B5′), let $A_n = L\sum_t f_t^* \|\mathbf{U}_t\|^3$. Under Assumptions (A5) and (A6), we have $A_n = O_p(n)$, and $\max_{1\le t\le n}\|u_{t,k,1}(\boldsymbol{\theta},\boldsymbol{\theta}_0)\| = O_p(n^{1/2})$. Thus, (B5′) is satisfied. By using Lemma 4.6 of He and Shao (1996), Lemma A.2 is therefore established. ■

Proof of Theorem 2.1: Based on Lemmas A.1 and A.2, we have

$$\begin{aligned} & n^{-1/2}\sum_{t=1}^n \Big[\psi_\tau\{Y_t - Q_Y(\tau;\widehat{\boldsymbol{\theta}}|\mathbf{W}_t)\}h(\mathbf{W}_t;\widehat{\boldsymbol{\theta}}) - \psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}_0|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta}_0)\Big] \\ & -n^{-1/2}\left[E\sum_{t=1}^n \psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta})\right]\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = o_p(1) \end{aligned} \tag{A.1}$$

Applying the Taylor expansion, we obtain

$$\left[E\sum_{t=1}^n \psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta})\right]\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = n\mathbf{D}_n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O_p(n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^2) \tag{A.2}$$

where

$$\begin{aligned}
\mathbf{D}_n &= n^{-1}\sum_{t=1}^{n}\frac{\partial E\psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= n^{-1}\sum_{t=1}^{n}\frac{\partial([\tau - F_t\{Y_t - Q_Y(\tau;\boldsymbol{\theta}|\mathbf{W}_t)\}]h(\mathbf{W}_t;\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= n^{-1}\sum_{t=1}^{n}\Big([-f_t\{Y_t - Q_Y(\tau;\boldsymbol{\theta}_0|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta}_0)h^{\mathrm{T}}(\mathbf{W}_t;\boldsymbol{\theta}_0)] \\
&\quad +[\tau - F_t\{Y_t - Q_Y(\tau;\boldsymbol{\theta}_0|\mathbf{W}_t)\}]\frac{\partial h(\mathbf{W}_t;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\Big) \\
&= n^{-1}\sum_{t=1}^{n}[-f_t\{Y_t - Q_Y(\tau;\boldsymbol{\theta}_0|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta}_0)h^{\mathrm{T}}(\mathbf{W}_t;\boldsymbol{\theta}_0)]\,.
\end{aligned}$$

Combined with the subgradient condition of quantile regression, we have

$$n^{-1/2}\sum_{t=1}^{n}\psi_\tau\{Y_t - Q_Y(\tau;\widehat{\boldsymbol{\theta}}|\mathbf{W}_t)\}h(\mathbf{W}_t;\widehat{\boldsymbol{\theta}}) = o_p(1) \tag{A.3}$$

Together with (A.1), (A.2) and (A.3), we have

$$\begin{aligned}
&-n^{-1/2}\sum_{t=1}^{n}[\psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}_0|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta}_0)] \\
=\ & n^{1/2}\mathbf{D}_n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O_p(n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^2) + o_p(1).
\end{aligned} \tag{A.4}$$

Because the left-hand side term $n^{-1/2}\sum_{t=1}^{n}[\psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}_0|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta}_0)] \xrightarrow{d} N(0,\mathbf{G})$ by using the central limit theorem and $\mathbf{D}_n$ is positive definite by Assumption (A5), we have $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(n^{-1/2})$. Therefore,

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\mathbf{D}_n^{-1}n^{-1/2}\sum_{t=1}^{n}\psi_\tau\{Y_t - Q_Y(\tau;\boldsymbol{\theta}_0|\mathbf{W}_t)\}h(\mathbf{W}_t;\boldsymbol{\theta}_0) + o_p(1).$$

By Assumption (A3), it follows that $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal with mean zero and variance matrix $\mathbf{D}^{-1}\mathbf{G}\mathbf{D}^{-1}$ using the central limit theorem. This completes the proof of Theorem 2.2. ■

A.2 Proof of Theorem 2.2:

The proof of Theorem 2.2 shares the similar strategy as Wang et al. (2007). Note that Theorem 2.2 is equivalent to

$$P\left(\min_{K\neq K_0} \text{sBIC}(K) > \text{sBIC}(K_0)\right) \to 1. \tag{A.5}$$

To prove (A.5), we identify two different cases i.e Case 1 for $K < K_0$ and Case 2 for $K > K_0$. Denote $\boldsymbol{\theta}_K^\star$ as the pseudo-true parameters for a given K, that is $\boldsymbol{\theta}_K^\star = \arg\min_{\boldsymbol{\theta}_K} ES_n(\boldsymbol{\theta}_K)$.

Case 1: when $K < K_0$, we need first prove $S_0(\boldsymbol{\theta}_K^\star) = E\rho_\tau\{Y - Q_Y(\tau; \boldsymbol{\theta}_K^\star|\mathbf{W})\} > S_0(\boldsymbol{\theta}_{K_0}^\star) = E\rho_\tau\{Y - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star|\mathbf{W})\}$. From Knight's identity, we can directly obtain that

$$\begin{aligned} & \rho_\tau\{Y - Q_Y(\tau; \boldsymbol{\theta}_K^\star|\mathbf{W})\} - \rho_\tau\{Y - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star|\mathbf{W})\} \\ = \; & \{Q_Y(\tau; \boldsymbol{\theta}_K^\star|\mathbf{W}) - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star|\mathbf{W})\}\{I(e_0 \le 0) - \tau\} \\ & + \int_0^{Q_Y(\tau;\boldsymbol{\theta}_K^\star|\mathbf{W}) - Q_Y(\tau;\boldsymbol{\theta}_{K_0}^\star|\mathbf{W})} \{I(e_0 \le s) - I(e_0 \le 0)\}ds, \end{aligned}$$

where $e_0 = Y - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star|\mathbf{W})$ and therefore

$$\begin{aligned} & E[\rho_\tau\{Y - Q_Y(\tau; \boldsymbol{\theta}_K^\star|\mathbf{W})\}] - E[\rho_\tau\{Y - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star|\mathbf{W})\}] \\ = \; & E\int_0^{Q_Y(\tau;\boldsymbol{\theta}_K^\star|\mathbf{W}) - Q_Y(\tau;\boldsymbol{\theta}_{K_0}^\star|\mathbf{W})} [F_0(s) - F_0(0)]ds. \end{aligned}$$

From Assumption (A1), density value $f_0(\cdot|\mathbf{W})$ is always bounded away from zero. We can immediately obtain that $E\int_0^{Q_Y(\tau;\boldsymbol{\theta}_K^\star|\mathbf{W}) - Q_Y(\tau;\boldsymbol{\theta}_{K_0}^\star|\mathbf{W})} [F_0(s) - F_0(0)]ds > 0$ no matter $Q_Y(\tau; \boldsymbol{\theta}_K^\star|\mathbf{W}) - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star|\mathbf{W})$ is positive or negative. Thus, $S_0(\boldsymbol{\theta}_K) > S_0(\boldsymbol{\theta}_{K_0})$ holds.

Denote $\widehat{\boldsymbol{\theta}}_K^\star = \arg\min_{\boldsymbol{\theta}_K} S_n(\boldsymbol{\theta}_K)$. By using similar argument in the proof of Lemma A.1, it is easy to show that $\widehat{\boldsymbol{\theta}}_K^\star$ is a consistent estimator for $\boldsymbol{\theta}_K^\star$. Note that in the proof of Lemma A.1, we mainly use the Theorem 2.1 of Newey and McFadden (1994), which does not requires the model to be correctly specified. Then we have

$$\text{sBIC}(K) - \text{sBIC}(K_0) = \left[\log\left(S_n(\widehat{\boldsymbol{\theta}}_K^\star)\right) - \log\left(S_n(\boldsymbol{\theta}_{K_0}^\star)\right)\right] + (K - K_0)\frac{\log n}{n}C_n$$

Since the first term can asymptotically dominate the second term and $(K-K_0)\frac{\log n}{n}C_n = o(1)$ by Assumption (A7),

$$\text{sBIC}(K) - \text{sBIC}(K_0) \to \log\left(S_0(\boldsymbol{\theta}_K^\star)\right) - \log\left(S_0(\boldsymbol{\theta}_{K_0}^\star)\right) > 0, \quad \text{as } n \to \infty.$$

Therefore, $\text{sBIC}(K) - \text{sBIC}(K_0) > 0$ for $K < K_0$ when n goes to the infinity.

Case 2: i.e. when $K > K_0$, by using the proposed BRISQ algorithm, the estimator for $\boldsymbol{\theta}_K^\star$ is denoted by $\widehat{\boldsymbol{\theta}}_K^\star = (\widehat{\boldsymbol{\eta}}_K(\widehat{\boldsymbol{\delta}}_K^\star)^{\mathrm{T}}, \widehat{\boldsymbol{\delta}}_K^{\star\mathrm{T}})^{\mathrm{T}}$. Recall that $\mathbf{M}_t(\boldsymbol{\delta}_K) = (1, X_t, (X_t - \delta_1)_+, \cdots, (X_t - \delta_K)_+, \mathbf{Z}_t^{\mathrm{T}})^{\mathrm{T}}$. Then $Q_Y(\tau; \boldsymbol{\theta}_K | \mathbf{W}_t) = \mathbf{M}_t(\boldsymbol{\delta}_K)^{\mathrm{T}} \boldsymbol{\eta}_K$ is a linear regression conditional on $\boldsymbol{\delta}_K$. By using similar arguments in pages 121-122 of Koenker (2005), we can show that $\|\widehat{\boldsymbol{\eta}}_K^\star(\boldsymbol{\delta}_K^\star) - \boldsymbol{\eta}_K^\star\| = O_p(n^{1/2})$ conditional on pseudo-true $\boldsymbol{\delta}_K^\star$. More specially, letting $\widetilde{\boldsymbol{\eta}}_K^\star = \boldsymbol{\eta}_K^\star + \mathbf{u}/\sqrt{n}$ where $\mathbf{u} \in \mathbb{R}^K$, we have

$$
\begin{aligned}
&\sum_t \rho_\tau\{Y_t - Q_Y(\tau; \widetilde{\boldsymbol{\eta}}_K^\star, \boldsymbol{\delta}_K^\star | \mathbf{W}_t)\} - \sum_t \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\eta}_K^\star, \boldsymbol{\delta}_K^\star | \mathbf{W}_t)\} \\
=&\frac{1}{\sqrt{n}} \sum_t \mathbf{M}_t(\boldsymbol{\delta}_K^\star)^{\mathrm{T}} \mathbf{u}\{I(e_t \le 0) - \tau\} + \int_0^{\mathbf{M}_t(\boldsymbol{\delta}_K^\star)^{\mathrm{T}}\mathbf{u}/\sqrt{n}} \{I(e_t \le s) - I(e_t \le 0)\} ds.
\end{aligned}
$$

For the first term of right hand, we have $\frac{1}{\sqrt{n}} \sum_t \mathbf{M}_t(\boldsymbol{\delta}_K^\star)^{\mathrm{T}} \mathbf{u}\{I(e_t \le 0) - \tau\} = O_p(1)$ by using central limit theorem. For the second term, it converges to a positive definite matrix. Thus we can obtain that $\widehat{\boldsymbol{\eta}}_K^\star(\boldsymbol{\delta}_K^\star)$ is a root-n consistent estimator and

$$
\left\| \sum_t \rho_\tau\{Y_t - Q_Y(\tau; \widehat{\boldsymbol{\eta}}_K^\star(\boldsymbol{\delta}_K^\star), \boldsymbol{\delta}_K^\star | \mathbf{W}_t)\} - \sum_t \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_K^\star | \mathbf{W}_t)\} \right\| = O_p(1).
$$

In addition, since $\widehat{\boldsymbol{\theta}}_K^\star$ is a consistent estimator for $\boldsymbol{\theta}_K^\star$ and $\sum_t \rho_\tau\{Y_t - Q_Y(\tau; \widehat{\boldsymbol{\eta}}_K^\star(\boldsymbol{\delta}_K^\star), \boldsymbol{\delta}_K^\star | \mathbf{W}_t)\}$ is continuous in $\boldsymbol{\delta}_K^\star$, we have

$$
\left\| \sum_t \rho_\tau\{Y_t - Q_Y(\tau; \widehat{\boldsymbol{\theta}}_K^\star | \mathbf{W}_t)\} - \sum_t \rho_\tau\{Y_t - Q_Y(\tau; \widehat{\boldsymbol{\eta}}_K^\star(\boldsymbol{\delta}_K^\star), \boldsymbol{\delta}_K^\star | \mathbf{W}_t)\} \right\| = O_p(1).
$$

Hence $\left\| \sum_t \rho_\tau\{Y_t - Q_Y(\tau; \widehat{\boldsymbol{\theta}}_K^\star | \mathbf{W}_t)\} - \sum_t \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_K^\star | \mathbf{W}_t)\} \right\| = O_p(1)$, and

$$
\begin{aligned}
& \mathrm{sBIC}(K) - \mathrm{sBIC}(K_0) \\
= \;& \log\left(1 + \frac{\frac{1}{n}\sum_{t=1}^n \rho_\tau\{Y_t - Q_Y(\tau; \widehat{\boldsymbol{\theta}}_K^\star, | \mathbf{W}_t)\} - \frac{1}{n}\sum_{t=1}^n \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star | \mathbf{W}_t)\}}{\frac{1}{n}\sum_{t=1}^n \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star | \mathbf{W}_t)\}} \right) \\
& + (K - K_0) \frac{\log n}{n} C_n \\
= \;& \log\left(1 + \frac{\frac{1}{n}\sum_{t=1}^n \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_K^\star | \mathbf{W}_t)\} - \frac{1}{n}\sum_{t=1}^n \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star | \mathbf{W}_t)\}}{\frac{1}{n}\sum_{t=1}^n \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star | \mathbf{W}_t)\}} \right. \\
& \left. + \frac{O(\frac{1}{n})}{\frac{1}{n}\sum_{t=1}^n \rho_\tau\{Y_t - Q_Y(\tau; \boldsymbol{\theta}_{K_0}^\star | \mathbf{W}_t)\}} \right) + (K - K_0) \frac{\log n}{n} C_n \\
= \;& O\left(\frac{1}{n}\right) + (K - K_0) \frac{\log n}{n} C_n
\end{aligned}
$$

Since $K - K_0 > 0$, then $\mathrm{sBIC}(K) - \mathrm{sBIC}(K_0) > 0$. The proof of Theorem 2.2 is completed. ■

A.3 Proof of Theorem 3.1

The following lemma is used for proving the Theorem 3.1.

Lemma A.3. *Under the Assumptions (A1), (A3)-(A4) and (A8), as $n \to \infty$, we have*

(I) $\sup_\delta |n^{-1}\sum_{t=1}^n \hat{f}_t(\hat{e}_t)\mathbf{V}_t(X_t-\delta)I(X_t<\delta) - \mathbf{H}_1(\delta)| \xrightarrow{p} 0$;

(II) $\sup_\delta |n^{-1}\sum_{t=1}^n \hat{f}_t(\hat{e}_t)\mathbf{V}_t\widehat{\beta}(X_t-\delta)I(X_t>\delta) - \mathbf{H}_2(\delta,\beta)| \xrightarrow{p} 0$;

(III) $\sup_\delta |n^{-1}\sum_{t=1}^n \mathbf{V}_t\mathbf{V}_t^T\hat{f}_t(\hat{e}_t) - \mathbf{H}| \xrightarrow{p} 0$.

Proof of Lemma A.3: We only give the proof for (I), since the proof for (II) and (III) are the same. For (I), it is sufficient to show that $\sup_\delta |\widehat{\mathbf{H}}_1(\delta) - \mathbf{H}_1(\delta)| = o_p(1)$, where $\widehat{\mathbf{H}}_1(\delta) = n^{-1}\sum_{t=1}^n \hat{f}_t(\hat{e}_t)\mathbf{V}_t(X_t-\delta)I(X_t<\delta)$. We have

$$\begin{aligned}
&\widehat{\mathbf{H}}_1(\delta) - \mathbf{H}_1(\delta) \\
=\ & n^{-1}\sum_{t=1}^n \{\hat{f}_t(\hat{e}_t) - f_t(\hat{e}_t)\}\mathbf{V}_t(X_t-\delta)I(X_t<\delta) + \\
& \left\{ n^{-1}\sum_{t=1}^n f_t(\hat{e}_t)\mathbf{V}_t(X_t-\delta)I(X_t<\delta) - \mathbf{H}_{1n}(\delta) \right\} + \{\mathbf{H}_{1n}(\delta) - \mathbf{H}_1(\delta)\} \\
=\ & (a)+(b)+(c) \qquad \text{(A.6)}
\end{aligned}$$

$\sup_\delta |(a)| = o_p(1)$ holds directly by the uniform convergence property of kernel estimator. For (b), note that

$$|(b)| \le n^{-1}\sum_{t=1}^n \mathbf{V}_t(X_t-\delta)I(X_t<\delta) \max_{1\le t\le n}\{f_t(Y_t - \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{V}_t) - f_t(Y_t - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}_t)\}.$$

By our Assumptions (A1), (A3) and (A9), and $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\| = o_p(n^{-1/2})$ in Lemma A.4, together with the mean value theorem, we have

$$\max_{1\le t\le n}\{f_t(Y_t - \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{V}_t) - f_t(Y_t - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}_t)\} \le \max_{1\le t\le n} \|\mathbf{V}_t\| \cdot |f'(Y_t - \boldsymbol{\xi}^{\mathrm{T}}\mathbf{V}_t)| \cdot \|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\| = o_p(1).$$

where $\boldsymbol{\xi}$ lies between $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}$. Hence $\sup_\delta |(b)|$ is $o_p(1)$.

Finanlly, for (c), we have $n^{-1}\sum_{t=1}^n f_t(e_t)\mathbf{V}_t(X_t-\delta)I(X_t<\delta) \xrightarrow{p} Ef_t(e_t)\mathbf{V}_t(X_t-\delta)I(X_t<\delta) = \mathbf{H}_1(\delta)$ for any given δ by using law of large numbers. Then $\sup_\delta |(c)| = o_p(1)$, whose proof follows the similar line of Lemma 1 in Hansen (1996) and thus is omitted. Since

(a), (b) and (c) are $o_p(1)$ uniformly in $\delta \in \Gamma$, then $\sup_\delta |\widehat{\mathbf{H}}_1(\delta) - \mathbf{H}_1(\delta)| = o_p(1)$. The proof is completed. ■

Lemma A.4. *Under Assumptions (A1), (A3)-(A4) and (A8), and the local alternative model (3.3),* $\widehat{\boldsymbol{\alpha}}$ *has the following Bahadur representation:*

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \mathbf{H}^{-1} \left\{ n^{-1} \sum_{t=1}^{n} \psi_\tau(e_t) \mathbf{V}_t \right\} + n^{-1/2} \mathbf{H}^{-1} \mathbf{H}_2(\delta, \beta) + o_p(1).$$

where $e_t = Y_t - \boldsymbol{\alpha}^T \mathbf{V}_t - n^{-1/2}\beta(X_t - \delta) I(X_t > \delta)$.

Proof of Lemma A.4: By using Lemma 4.1 of He and Shao (1996), we have

$$\sup_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| \le c_n} \left\| n^{-1/2} \sum_{t=1}^{n} \{\psi_\tau(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t) - \psi_\tau(e_t)\} \mathbf{V}_t - n^{-1/2} \sum_{t=1}^{n} E\{\psi_\tau(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t) \mathbf{V}_t | \mathbf{V}_t\} \right\| = O_p\left((c_n + n^{-1/2})^{1/2} \log n \right),$$

where $c_n = o(1)$ as $n \to \infty$. Since $E\{\psi_\tau(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t) | \mathbf{V}_t\} = \tau - F\{(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t) | \mathbf{V}_t\}$, then we can obtain

$$\begin{aligned} & n^{-1/2} \sum_{t=1}^{n} [\psi_\tau(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t) - \{\tau - F_t(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t)\}] \mathbf{V}_t \\ = \;& n^{-1/2} \sum_{t=1}^{n} \psi_\tau(e_t) \mathbf{V}_t + O_p\left((\| \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \| + n^{-1/2})^{1/2} \log n \right). \end{aligned} \tag{A.7}$$

Based on the subgradient condition of quantile regression, we get

$$n^{-1/2} \sum_{t=1}^{n} \psi_\tau(Y_t - \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{V}_t) \mathbf{V}_t = o_p(1).$$

Hence,

$$
\begin{aligned}
& n^{-1/2}\sum_{t=1}^{n}[\psi_\tau(Y_t-\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{V}_t)-\{\tau-F_t(Y_t-\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{V}_t)\}]\mathbf{V}_t \\
= \;& n^{-1/2}\sum_{t=1}^{n}\{F_t(Y_t-\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{V}_t)-\tau\}\mathbf{V}_t+o_p(1) \\
= \;& n^{-1/2}\sum_{t=1}^{n}f_t\{Y_t-\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}_t-n^{-1/2}\beta(X_t-\delta)I(X_t>\delta)\}\mathbf{V}_t\mathbf{V}_t^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) \\
& -n^{-1}\sum_{t=1}^{n}f_t\{Y_t-\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{V}_t-n^{1/2}\beta(X_t-\delta)I(X_t>\delta)\}\mathbf{V}_t(X_t-\delta)I(X_t>\delta) \\
& +o_p(1)+o_p\left(n^{1/2}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})\right) \\
= \;& n^{1/2}\mathbf{H}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})-\mathbf{H}_2(\delta,\beta)+o_p(1)+o_p\left(n^{1/2}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})\right).
\end{aligned}
$$

Together with (A.7), the proof of Lemma A.4 is completed. ■

Proof of Theorem 3.1: Under the null hypothesis $\beta=0$, $q(\delta,\beta)=0$ and thus Theorem 3.1 holds under H_0. It remains to show that Theorem 3.1 holds under local alternative model (3.3). By Lemmas A.3 and A.4, and after some simple algebraic manipulation, it is easy to obtain that

$$
\begin{aligned}
R_n(\delta) &= n^{-1/2}\sum_{t=1}^{n}\psi_\tau(Y_t-\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{V}_t)(X_t-\delta)I(X_t\le\delta) \\
&= n^{-1/2}\sum_{t=1}^{n}\psi_\tau\{e_t+n^{-1/2}\beta_1(X_t-\delta)I(X_t>\delta)-(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})^{\mathrm{T}}\mathbf{V}_t\}(X_t-\delta)I(X_t\le\delta) \\
&= n^{-1/2}\sum_{t=1}^{n}\psi_\tau(e_t)(X_t-\delta)I(X_t\le\delta)-\mathbf{H}_1(\delta)\mathbf{H}^{-1}n^{-1/2}\sum_{t=1}^{n}\psi_\tau(e_t)\mathbf{V}_t \\
&\quad -\mathbf{H}_1(\delta)\mathbf{H}^{-1}\mathbf{H}_2(\delta,\beta)+o_p(1) \\
&= n^{-1/2}\sum_{t=1}^{n}\psi_\tau(e_t)\{(X_t-\delta)I(X_t\le\delta)-\mathbf{H}_1(\delta)\mathbf{H}^{-1}\mathbf{V}_t\}-\mathbf{H}_1(\delta)\mathbf{H}^{-1}\mathbf{H}_2(\delta,\beta)+o_p(1) \\
&= R(\delta)+q(\delta,\beta)+o_p(1).
\end{aligned}
$$

The weak convergence of $R(\delta)$ can be obtained directly by following the proof of Stute (1997). This completes the proof of Theorem 3.1. ■

References

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Bofingeb, E. (1975). Estimation of a density function using order statistics. *Australian Journal of Statistics*, 17(1):1–7.

Brandt, L., Tombe, T., and Zhu, X. (2013). Factor market distortions across time, space and sectors in china. *Review of Economic Dynamics*, 16(1):39–58.

Caner, M. (2002). A note on least absolute deviation estimation of a threshold model. *Econometric Theory*, 18(3):800–814.

Cao, K. H. and Birchenall, J. A. (2013). Agricultural productivity, structural change, and economic growth in post-reform china. *Journal of Development Economics*, 104:165–180.

Chan, N. H., Yau, C. Y., and Zhang, R.-M. (2015). Lasso estimation of threshold autoregressive models. *Journal of Econometrics*, 189(2):285–296.

Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, 95:759–771.

Chenery, H. B., Robinson, S., Syrquin, M., and Feder, S. (1986). *Industrialization and growth*. Citeseer.

Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: the lms method and penalized likelihood. *Statistics in Medicine*, 11(10):1305–1319.

Durnin, J. V. and Womersley, J. (1974). Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *British Journal of Nutrition*, 32(1):77–97.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, 98:995–999.

Fong, Y. (2019). Fast bootstrap confidence intervals for continuous threshold linear regression. *Journal of Computational and Graphical Statistics*, 28(2):1–8.

Fong, Y., Di, C., Huang, Y., and Gilbert, P. B. (2017). Model-robust inference for continuous threshold regression models. *Biometrics*, 73(2):452–462.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281.

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*, 2:307–333.

Hall, P. and Sheather, S. J. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):381–391.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, 35(2):228–240.

He, X. and Shao, Q.-M. (1996). A general bahadur representation of m-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics*, 24(6):2608–2630.

He, X. and Zhu, L.-X. (2003). A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, 98(464):1013–1022.

Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 87(417):58–68.

Hidalgo, J., Lee, J., and Seo, M. H. (2019). Robust inference for threshold regression models. *Journal of Econometrics*, 210(2):291–309.

Koenker, R. W. (2005). *Quantile regression*. Cambridge University Press.

Lee, E. R., Noh, H., and Park, B. U. (2014). Model selection via bayesian information criterion for quantile regression models. *Journal of the American Statistical Association*, 109(505):216–229.

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(1):242–249.

Lian, H. (2012). A note on the consistency of schwarz's criterion in linear quantile regression with the scad penalty. *Statistics & Probability Letters*, 82(7):1224–1228.

Muggeo, V. M. (2017). Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach. *Australian & New Zealand Journal of Statistics*, 59(3):311–322.

Muggeo, V. M. and Adelfio, G. (2010). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27(2):161–166.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Oka, T. and Qu, Z. (2011). Estimating structural changes in regression quantiles. *Journal of Econometrics*, 162(2):248–267.

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., and Schmid, M. (2019). A review of spline function procedures in r. *BMC Medical Research Methodology*, 19(1):46.

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, 146(1):170–184.

Reinhart, C. M. and Rogoff, K. S. (2010). Growth in a time of debt. *American Economic Review*, 100(2):573–78.

Royston, P. and Sauerbrei, W. (2008). *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables*, volume 777. John Wiley & Sons.

Song, Z., Storesletten, K., and Zilibotti, F. (2011). Growing like china. *American Economic Review*, 101(1):196–233.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, 25(2):613–641.

Wang, H., Li, R., and Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94(3):553–568.

Wood, S. N. (2001). Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting. *Biometrics*, 57(1):240–244.

Zhang, F. and Li, Q. (2017). A continuous threshold expectile model. *Computational statistics & data analysis*, 116:49–66.

Zhang, L., Wang, H. J., and Zhu, Z. (2014). Testing for change points due to a covariate threshold in quantile regression. *Statistica Sinica*, 24(4):1859–1877.

Zhang, L., Wang, H. J., and Zhu, Z. (2017). Composite change point estimation for bent line quantile regression. *Annals of the Institute of Statistical Mathematics*, 69(1):145–168.

Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics*, 36(4):1509–1533.