Improved Algorithms of Multi-kink Quantile Regression

Yu Yue

Abstract

The change point problem has been involved in all varieties of fields in contemporary society. The kink regression model is important for study the change point problem, which can effectively depict continuous and piecewise linear variations between variables of datasets. According to the development history of this kind of model, this paper systematically introduces several aspects of three important evolving models: single kink mean regression model, bent-line quantile regression model and multi-kink quantile regression model, including parameter estimation, tests of kink effect, algorithms and theories of the asymptotic properties. Meanwhile, mainly aiming at the characteristics of the unsmooth at the location of kink and nondifferentiability of quantile regression loss function, this paper gives some algorithm improvements on multi-kink quantile regression model by combining with some previous research methods. These improved algorithms apply the framework of the gradient descent algorithm to obtain the effective parameter estimates. A large number of numerical simulations show that the multi-kink quantile regression model and our proposed improved algorithms can commendably fulfil the task of parameter estimation. The specific experiments show that these algorithms can well control the bias and mean square error of parameter estimation for different quantile levels under homoscedasticity and heteroscedasticity and the random error term following different distributions. Finally, this paper also applies the above methods to the actual data analysis, demonstrates the practical application significance of these algorithms, and explains that these algorithms have their own advantages and defects for different datasets. Therefore, it is vital to select appropriate algorithms to analyze data according to specific problems.

Key Words: Kink Regression Model; Quantile Regression; Parameter Estimation; Smoothing

1 Introduction

As a basic model in statistics, linear regression model has been well known by the public. The idea of this model is simple and easy to implement, and it can effectively describe the small amount of data and simple relationship; as the basis of many nonlinear models, the model is easy to understand and has good interpretability, which is conducive to data analysis. But in real life, the relationship between many things is not static. It is often observed that response variables and explanatory variables show different regression relationships before and after a critical position. This critical position is called the "change point".

The concept of change point was first proposed by Page[1]. It is used to study continuous sampling inspection in product quality inspection, simulate the quality warning line of product inspection through the position of change point, and eliminate defective products in products. After that, the change point problem has attracted more and more scholars' attention in the field of statistics. At present, it has evolved into an important research direction in statistics.

According to whether the regression function of the model is continuous at the change point, the change point model is divided into bent line model and structural change point model. In this paper, we mainly study the bent line model. In the bent line model, the position of change point is often called "kink".

The history of bent line regression model can be traced back to the continuous piecewise linear model proposed by Quandt[2, 3] and the two-stage linear regression model proposed by Sprent[4]. The model is proposed when studying the continuous bent line relationship between response variables and explanatory variables. As the name suggests, this bent line relationship is linear before and after the kink, with completely different slopes, and the two linear models intersect at the kink position. Based on this model, Hinkley[5, 6], Feder[7] and Chappell[8] studied the statistical inference of this segmented bent line model. Most of their research methods are based on the least square method, aiming to use traditional methods to solve the problem of this emerging model. Later, based on this model, Tong[9] proposed the threshold effect model. This model can be used to describe nonlinear relationships. Chan and Tsay[10] proposed a continuous threshold model on this basis. This model has the same form as the piecewise linear regression model.

However, the premise of using this model is that there are obvious abrupt changes in the regression relationship of the dataset. However, in the actual operation process, it is difficult to judge whether there is obvious mutation in the dataset. Therefore, the focus of research has gradually shifted to this issue. The problem of threshold effect test begins with the test of threshold autoregressive effect by Chan and Tong[9]; Then Hansen[11] and Lee et al.[12] also made their own contributions to this issue. It is worth mentioning that Chiu et al.[13, 14] creatively proposed BentCable mean regression model from the perspective of mitigating the sharp mutation in the model. Using this model, they eased the sharp mutation at the change point of the model, thus more flexibly depicting the model changes near the change point. Das et al.[15] subsequently made relevant improvements to this method. This method is also involved in the algorithm improvement in this paper.

Based on the research of Chan and Tsay[10], Hansen[16] expanded the relevant theories. Hansen[16] systematically gives the methods of parameter estimation, variance estimation, kink effect test and statistical inference of single bent line mean regression model, which also includes the method of constructing parameter confidence interval and even the confidence interval of the whole single bent line mean regression model. It is the current integrator in the direction of bent line mean regression, which also called as "Single Kink least-square Model". However, the bent line mean regression model cannot dynamically describe the relationship between response variables and explanatory variables. Therefore, Li et al.[17] creatively proposed a bent line quantile regression model based on the quantile regression model proposed by Koenker and Bassett[18] and taking the dataset of animal quality and maximum running speed as an empirical analysis. This model can systematically study the data in the larger and smaller quantiles of the dataset with bent line effect, and well overcome the defects of single kink mean regression model.

But as mentioned before, there is likely to be more than one kink in datasets. At this time, the bent line quantile regression model is not applicable. Therefore, Zhong et al.[19] systematically proposed a multi-kink quantile regression model based on the research of Li et al[17]. The multi-kink quantile regression model is as follows:

$$Q_Y(\tau;\boldsymbol{\eta},\boldsymbol{\delta}|x_i,\boldsymbol{z_i}) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^{K}\beta_k(x_i-\delta_k)I(x_i>\delta_k) + \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{z_i}, \quad i=1,2,\cdots,n, \quad (1)$$

where $\boldsymbol{\delta}=(\delta_1,\delta_2,\cdots,\delta_K)^{\mathrm{T}}$ is the vector of the kinks. Zhong et al.[19] made use of the linearization technique of Muggeo[20] to accurately estimate the position of the break point; at the same time, they have also given their own innovative methods for the estimation of break points, parameter estimation, the test of break point effect and the construction of confidence intervals. They combined idea of self-help method in Bradley[21] and presented a Bootstrap Restarting Iterative Segmented Quantile (BRISQ) algorithm for parameter estimation of the problem. This model well achieves the balance between linear quantile regression model and nonparametric quantile regression model, which is a recommended regression analysis method.

In this article, we give some improvements to the BRISQ algorithm mainly from the perspective of smoothing. These improvements can speed up the computation efficiency of the algorithm to a certain extent, alleviate the changes between the variable relationships at the kink location, and have good parameter estimation effect. In this paper, a large number of simulation experiments are used to analyze them.

The rest of the paper is structured as follows. In Section 2, using some methods proposed by predecessors, we make a series of improvements on the parameter estimation algorithm of the multi-kink quantile regression model of Zhong et al. In Section 3, a large number of numerical simulations are carried out to evaluate the performance of the algorithms. Section 4 presents a real data application. Section 5 concludes the paper. Some additional numerical simulation results are included in the appendix. We also make a R package named "MKQRimprove" to integrate these algorithms.

2 Improved Theory

2.1 Kernel Function Smoothing

For the multi-kink quantile regression model, the linearization technique proposed by Muggeo[20] was used to estimate the kink location. This technique can effectively make up for the shortcomings of the grid search method, such as the high computational cost, the limitation on estimation of kink to discrete grid points and so on. This is a technique used to deal with the nondifferentiability of the kink location in the multi-kink quantile regression model. But for the quantile regression model with kinks, it is difficult to estimate the kinks of this model. The most difficult part of the problem is that the loss function is an unsmooth and non everywhere differentiable function, which greatly increases the difficulty of estimation. However, the previous linearization algorithm for kink location estimation also has some defects, and the kink estimation is underestimated from time to time.

For the smoothing technique of loss function, the first thing we can think of is to improve the model. Previously, the model itself was unsmooth and non everywhere differentiable due to the existence of indicative functions. At the same time, inspired by the smoothing of the indicative function of the model during the construction of SRS test statistics proposed by Zhong et.al in the statistical inference of multi-kink quantile regression, we firstly smooth the indicative function of the model. As for kernel function, Aizerman[22] first introduced this technology into the field of machine learning in the research of potential function method. Later in the field of change point, Zhou and Zhang[23] proved the limit properties of this method.

We utilize kernel function to smooth the indication function in model (**??**) and obtain the following function.

$$\tilde{Q}_K(\tau;\boldsymbol{\eta},\boldsymbol{\delta}|x_i,\boldsymbol{z_i}) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^{K} \beta_k (x_i - \delta_k)\Phi\left(\frac{x_i-\delta_k}{h_k}\right) + \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{z_i}, \quad i = 1, 2, \cdots, n \tag{2}$$

where $\Phi(\cdot)$ is a Gaussian kernel function, and h_t is a window width parameter to regulate the smoothness.

Assumption 2.1. *When the sample size $n \to \infty$, the bandwidth satisfy that $h_k \to 0 for k = 1, 2, \cdots, K$..*

Under assumption 2.1,

Theorem 1. *When $h \to 0$, the formula*

$$\Phi\left(\frac{x-t}{h}\right) = I(x > t) + o(h)$$

holds.

Under Theorem 1, the model (2) can be regarded as an approximation of the model (1). Thus, the problem of estimating the parameters in equation (1) is transformed into the problem of estimating the parameters in equation (2). At this time, the model (2) formula is smooth and everywhere differentiable relative to the kink location $\boldsymbol{\delta}$.

Let $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ be the regression coefficient vector to be estimated, $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_K)^{\mathrm{T}}$ as the location vector of kink to be evaluated, then we can regard $\boldsymbol{\theta} = (\boldsymbol{\eta}^{\mathrm{T}}, \boldsymbol{\delta}^{\mathrm{T}})^{\mathrm{T}}$ as all parameter vectors to be estimated. The loss function is

$$L_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau[y_i - \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})]. \tag{3}$$

The corresponding score function of $L_n(\boldsymbol{\theta})$ is

$$S_n(\boldsymbol{\theta}) = \frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})]q(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}), \tag{4}$$

where $q(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})$ is the derivative of $\tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})$ with respect to $\boldsymbol{\theta}$, and is given as follows:

$$q(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}) = \frac{\partial \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})}{\partial \boldsymbol{\theta}} = \left[1, x_i, \xi_{i1}, \cdots, \xi_{iK}, \boldsymbol{z_i}^{\mathrm{T}}, \upsilon_{i1}, \cdots, \upsilon_{iK}\right]^{\mathrm{T}},$$

in which, $\xi_{ik} = (x_i - \delta_k)\Phi\left(\frac{x_i - \delta_k}{h_k}\right)$, $\upsilon_{ik} = -\beta_k\Phi\left(\frac{x_i-\delta_k}{h_k}\right) - \frac{\beta_k}{h_k}(x_i - \delta_k)\phi\left(\frac{x_i-\delta_k}{h_k}\right)$. Here, $\phi(\cdot)$ is the density function of the standard normal distribution.

Gradient descent algorithm is used to solve equation (4). The formula of gradient descent $S_n(\boldsymbol{\theta})$ is as follows:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - r\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{5}$$

where $\boldsymbol{\theta}^{(t)}$ represents the estimate of $\boldsymbol{\theta}$ in the t step iteration, and r represents the learning rate of each iteration. According to formula (5) of the gradient descent method, the update iteration formula of each parameter is given as follows:

$$\begin{cases} \alpha_0^{(t+1)} = \alpha_0^{(t)} + r\frac{1}{n}\sum_{i=1}^{n}\psi_\tau\left(y_i - \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})\right) \\ \alpha_1^{(t+1)} = \alpha_1^{(t)} + r\frac{1}{n}\sum_{i=1}^{n}\psi_\tau\left(y_i - \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})\right)x_i \\ \beta_k^{(t+1)} = \beta_k^{(t)} + r\frac{1}{n}\sum_{i=1}^{n}\psi_\tau[y_i - \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})](x_i - \delta_k)\Phi\left(\frac{x_i-\delta_k}{h_k}\right) \\ \boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} + r\frac{1}{n}\sum_{i=1}^{n}\psi_\tau[y_i - \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})]\boldsymbol{z_i}^{\mathrm{T}} \\ \delta_k^{(t+1)} = \delta_k^{(t)} - r\frac{1}{n}\sum_{i=1}^{n}\psi_\tau[y_i - \tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})]\left[\beta_k\Phi\left(\frac{x_i-\delta_k}{h_k}\right) + \frac{\beta_k}{h_k}(x_i - \delta_k)\phi\left(\frac{x_i-\delta_k}{h_k}\right)\right]. \end{cases} \tag{6}$$

where $k = 1, 2, \cdots, K$. The initial value $\boldsymbol{\theta}^{(0)}$ of $\boldsymbol{\theta}$ is mainly set in the following ways: for the initial value $\boldsymbol{\delta}^{(0)}$, we select the quantile level $\frac{1}{K+1}, \frac{2}{K+1}, \cdots, \frac{K}{K+1}$ of the explanatory variable $x_1, \cdots, x_n$. For other coefficients $\boldsymbol{\eta}$, when $\boldsymbol{\delta}^{(0)}$ is selected as the initial value of $\boldsymbol{\theta}$, we fit model formula (1) to obtain as the initial value $\boldsymbol{\eta}^{(0)}$ for the given $\delta^{(0)}$.

For the learning rate r, we set the initial learning rate as r_0, the iterative learning rate at step t is r_t. After each iteration, calculate the value of the corresponding loss function. If the loss function value of iteration in step $t+1$ $L_n^{(t+1)}(\boldsymbol{\theta})$ is less than the loss function value in step t $L_n^{(t)}(\boldsymbol{\theta})$, it indicates that the function is descending in the right direction. At this time, we increase the step size (learning rate) to $r_{t+1} = ar_t$, and

$a > 1$. Conversely, if $L_n^{(t+1)}(\boldsymbol{\theta}) \geq L_n^{(t)}(\boldsymbol{\theta})$, it indicates that the function does not perform gradient descent in the correct direction. In this case, it is necessary to reduce the step size (learning rate) to $r_{t+1} = br_t$, and $b < 1$. At present, when the loss function value between the next two iterations hardly changes, that is, for an algorithm tolerance ε, the loss function meets $|L_n^{(t+1)}(\boldsymbol{\theta}) - L_n^{(t)}(\boldsymbol{\theta})| < \varepsilon$, the iteration can be considered to be terminated, and the parameter estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ in the model (2) is obtained.

For the variance estimation of $\hat{\boldsymbol{\theta}}$, the delta method is also used. Here we mainly introduce the related theorems. Firstly, the matrix related to the expectations of the first and second derivative of the loss function used in the delta method is introduced:

$$C_n = E\left\{\left(\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}}\right\}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} = \frac{1}{n}\sum_{i=1}^{n} \tau(1-\tau)q^{\mathrm{T}}(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})q(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}),$$

$$D_n = E\left\{\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}}\right\}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} = -\frac{1}{n}\sum_{i=1}^{n} \hat{f}_i[\tilde{Q}_K(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})]q^{\mathrm{T}}(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})q(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}).$$

where the conditional density function $\hat{f}_i(\cdot)$ can be estimated using the method propesed by Hendricks and Koenker[24].

$$\hat{f}_i[Q_Y(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] = \frac{2h_n}{Q_Y(\tau + h_n; \hat{\boldsymbol{\theta}}|x_i, \boldsymbol{z_i}) - Q_Y(\tau - h_n; \hat{\boldsymbol{\theta}}|x_i, \boldsymbol{z_i})}. \tag{7}$$

The specific bandwidth h_n needs to meet Assumption 2.1.

For the asymptotic theorem of variance estimation of parameter estimation, the following assumptions should be satisfied.

Assumption 2.2. *For model* (2),

(C1). The unique true value $\boldsymbol{\theta}_0$ *of* $\boldsymbol{\theta}$ *satisfies that* $\boldsymbol{\theta_0} \in \Theta$*, where* Θ *is a compact set within the definition domain of the parameter.*

(C2). For all explanatory variables x_i*,* $\sup_{\theta\in\Theta} |\beta_k(x_i - \delta_k)|, k = 1, 2, \cdots, K$ *are bounded.*

(C3). For any explanatory variable x_i*,* $\boldsymbol{z_i}$*, satisfy that* $\sup_{\theta\in\Theta} ||q(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})||^3 < \infty$.

(C4). The density function $\hat{f}_i(\cdot)$ *is bounded and continuous.*

(C5). Matrices C_n *and* D_n *are bounded.*

At this time, there is an asymptotic theorem about estimating $\hat{\boldsymbol{\theta}}$.

Theorem 2. *Set* $\boldsymbol{\theta_0}$ *as the unique true value of the parameter, and* $\hat{\boldsymbol{\theta}}$ *as the final estimation vector of the parameters. Under Assumption 2.1 and Assumption 2.2, then* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0} = O_p(n^{-1/2})$ *holds, and*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = D_n^{-1}C_nD_n^{-1}$.

The variance estimation of the parameter estimation in the algorithm can be obtained by the asymptotic distribution theorem.

According to Section 4.1, the parameter estimate algorithm flow in the model (2) is shown in the following Algorithm 1. The algorithm smoothes the original model (1), and can complete the parameter estimation of the model. At the same time, because of the combination of the gradient descent algorithm, the operation efficiency of the algorithm itself is very high. However, the improvement does change the linear part of the original model, which makes it difficult to achieve the perfect substitution of the standard normal kernel function for the indicative function when the sample size of the data is small, thus changing the essence of the original model more or less. Therefore, for the case of small sample size, if the multi-kink quantile regression model is used for analysis, it is not recommended to use this improved algorithm.

Algorithm 1: SmoothKernel Algorithm

1: Initialize the initial iteration value of the parameter to be evaluated according to the steps mentioned above, set $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\eta}^{(0)^{\mathrm{T}}}, \boldsymbol{\delta}^{(0)^{\mathrm{T}}})^{\mathrm{T}}$.
2: Calculate the loss function value $L_n(\boldsymbol{\theta}^{(0)})$ at $\boldsymbol{\theta}^{(0)}$.
3: Calculate the parameter result $\boldsymbol{\theta}^{(1)}$ through (6). Compare $L_n(\boldsymbol{\theta}^{(0)})$ and $L_n(\boldsymbol{\theta}^{(1)})$, update the learning rate according to the discussion in Section 2.1.
4: Continue the second and third steps until the algorithm converges, terminate the iteration, and output
the final parameter estimation result $\hat{\boldsymbol{\theta}}$.
5: After the final parameter estimation $\hat{\boldsymbol{\theta}}$ is obtained, the variance estimation of the parameter estimation is calculated by using the asymptotic Theorem 3.

2.2 BentCable Kink Smoothing

In the improvement in Section 2.1, we shift the primary problem of the improvement to the smoothing of the model. After smoothing the indicative function, for the loss function which is already differentiable everywhere, we can naturally obtain the parameter estimation of the model by derivation and gradient descent. The second improvement also adopts the smoothing technique, but it is different from the smoothing of the indicative function. Here, the linear function form of the original model is retained to a certain extent, but the kinks in the model (1) that make the model unable to be differentiable everywhere are focused. This leads to the BentCable Kink smoothing technology.

This technique was proposed by Chiu[13], and extended by Das et al. [15]. The idea is to take a bandwidth at the non differentiable kink, and use another smooth function to approximate the position covered by the bandwidth, so as to replace the kink. At this point, the model function is smooth and derivable on the kink and its neighborhood.

In the multi-kink quantile regression model, a quadratic function is used to approximate the position of the fold. The model expression is:

$$\tilde{Q}_B(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \boldsymbol{z_i}) = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^{K} \beta_k S_n(\delta_k, x_i) + \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{z_i}, i = 1, 2, \cdots, n, \tag{8}$$

where $S_n(\delta_k, x_i)$ is the BentCable Kink function, expressed as

$$S_n(\delta_k, x_i) = \begin{cases} 0, x_i < \delta_k - h_k \\ (x_i - \delta_k + h_k)^2/4h_k, \delta_k - h_k \le x_i \le \delta_k + h_k \\ x_i - \delta_k, x > \delta_k + h_k. \end{cases} \tag{9}$$

Thus, $(x_i - \delta_k)_+$ in the model (1) can also be approximated by (9), and h_k indicates the bandwidth required in the function. It also needs to meet the bandwidth requirements of Assumption 2.1.

It is easy to see that the formula (9) is smooth and differentiable everywhere, and the model retains the original linear form at a position far from the inflection point. The loss function of the model at this time is

$$\begin{aligned} P_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n} \rho_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] \\ &= \frac{1}{n}\sum_{i=1}^{n} \rho_\tau[y_i - \alpha_0 - \alpha_1 x_i - \sum_{k=1}^{K} \beta_k S_n(\delta_k, x_i) - \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{z_i}]. \end{aligned}$$

Take the derivative of $P_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and obtain the equation

$$p_n(\boldsymbol{\theta}) = \frac{\partial P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] T(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}) = 0, \tag{10}$$

where $T(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})$ is given as follows:

$$t(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}) = \left(1, x_i, S_n(\delta_1, x_i), \cdots, S_n(\delta_K, x_i), \boldsymbol{z_i}^{\mathrm{T}}, \beta_1 s_n(\delta_1, x_i), \cdots, \beta_K s_n(\delta_K, x_i)\right)^{\mathrm{T}},$$

where $s_n(\delta_k, x_i)$ is the first derivative of the BentCable kink function $S_n(\delta_k, x_i)$ at the kink location δ_k,

$$s_n(\delta_k, x_i) = \begin{cases} 0, x_i < \delta_k - h_k \\ -(x_i - \delta_k + h_k)/2h_k, \delta_k - h_k \le x_i \le \delta_k + h_k \\ -1, x > \delta_k + h_k. \end{cases} \tag{11}$$

Therefore, the estimation problem is transformed into a problem of solving the equation of (10).

In this part, for convenience, the gradient descent algorithm is still used to solve the equation. Combined with the formula of the gradient descent of (5), the update iteration formula for each parameter estimate of $\boldsymbol{\theta}$ are given as follows:

$$\begin{cases} \alpha_0^{(t+1)} = \alpha_0^{(t)} + r\frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] \\ \alpha_1^{(t+1)} = \alpha_1^{(t)} + r\frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] x_i \\ \beta_k^{(t+1)} = \beta_k^{(t)} + r\frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] S_n(\delta_k, x_i) \\ \boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)} + r\frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] \boldsymbol{z_i}^{\mathrm{T}} \\ \delta_k^{(t+1)} = \delta_k^{(t)} + r\frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})][\beta_k s_n(\delta_k, x_i)] \\ h_k^{(t+1)} = h_k^{(t)} + r\frac{1}{n}\sum_{i=1}^{n} \psi_\tau[y_i - \tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})] \frac{h_k^{(t)^2} - (x_i - \delta_k)^2}{4h_k^{(t)2}}, \end{cases} \tag{12}$$

where $k = 1, 2, \cdots, K$. Note that the bandwidth h_k here is also regarded as a part of the parameter vector, so the bandwidth of each step in the iteration process also needs to be iterated. This is a consideration based on the robustness of parameters estimation. In the subsequent simulation test, the initial value of h_k is $h_k^{(0)} = \hat{\sigma} \times 30^{-0.26}$. At the same time, the initial value $\boldsymbol{\theta}^{(0)}$ and the learning rate during the iteration are consistent with those in Section 2.1.

As for the variance estimation at this time, it is similar with that in Section 2.1. The Delta method and "Sandwich" formula are also used to construct the asymptotic theorem between the parameter estimation $\hat{\boldsymbol{\theta}}$ and the true value $\boldsymbol{\theta_0}$. The estimation matrices at $\boldsymbol{\theta_0}$, $C_n(\boldsymbol{\theta_0})$ and $D_{n(\boldsymbol{\theta_0})}$ are in the following form:

$$C_n(\boldsymbol{\theta_0}) = E\left\{\left(\frac{\partial P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}}\right\}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} = n^{-1}\sum_{i=1}^{n}\tau(1-\tau)T(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})T^{\mathrm{T}}(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}),$$

$$D_n(\boldsymbol{\theta_0}) = E\left\{\frac{\partial^2 P_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\}|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} = -\frac{1}{n}\sum_{i=1}^{n}\hat{f}_i[\tilde{Q}_B(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})]T(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})T^{\mathrm{T}}(x_i, \boldsymbol{z_i}; \boldsymbol{\theta}).$$

Conditional density function $\hat{f}_i(\cdot)$ is the discrete derivative form of (7). The asymptotic theorem at this time should satisfy the following Assumption 2.3.

Assumption 2.3. *For model* (8),

(C1). The unique true value of $\boldsymbol{\theta}$ *satisfies that* $\boldsymbol{\theta_0} \in \Theta$, *where* $\Theta \in \mathbb{R}^{2+2K+p}$ *is a compact set.*

(C2). $E\left[\sup_{\theta\in\Theta}||S_n(\boldsymbol{\delta}, \boldsymbol{x})||\right]$ *and* $E\left[\sup_{\theta\in\Theta}||\beta_k I(x_i > \delta_k)||\right]$ *are both limited.*

(C3). For any explanatory variable x_i *and* $\boldsymbol{z_i}$, *satisfy that* $E\left[\sup_{\theta\in\Theta}||t(x_i, \boldsymbol{z_i}; \boldsymbol{\theta})||\right] < \infty$.

(C4). The density function $\hat{f}_i(\cdot)$ *is bounded and continuous.*

(C5). Matrices $C_n(\boldsymbol{\theta_0})$ *and* $D_n(\boldsymbol{\theta_0})$ *are both finite positive definite matrices.*

Then there is an asymptotic theorem about estimating $\hat{\boldsymbol{\theta}}$.

Theorem 3. *Set* $\boldsymbol{\theta_0}$ *as the unique true value of the parameter, and* $\hat{\boldsymbol{\theta}}$ *as the final estimation vector of the parameters. Under Assumption 2.1 and Assumption 2.3,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = D_n(\boldsymbol{\theta_0})^{-1}C_n(\boldsymbol{\theta_0})D_n(\boldsymbol{\theta_0})^{-1}$.

The variance estimation of the parameter estimation in the algorithm can be obtained by the asymptotic distribution theorem. The algorithm flow is similar to Algorithm 1 that mentioned in Section 2.1. The algorithm solves the problem that the original multi-kink quantile regression model (1) can not be derived at the kinks, and estimates the parameters with the idea of function approximation. However, the drawback is that this method involves the BentCable Kink function $S_n(\delta_k, x_i)$ and its derivative $s_n(\delta_k, x_i)$ needs to iterate the bandwidth at the same time, which adds the computation burden, so the algorithm still needs to be improved.

2.3 Smoothed Loss Function Boosting

For the quantile regression model, the loss function $\rho_\tau(u) = u[\tau - I(u < 0)]$ is not differentiable at $u = 0$. Therefore, it is natural to consider the smoothing of the loss function for quantile regression.

Zheng(2011)[25] proposed a function replacing $\rho_\tau(u)$ by the following smooth function

$$S_{\tau,\kappa}(u) = \tau u + \kappa \log\left[1 + \exp\left(-\frac{u}{\kappa}\right)\right], \tag{13}$$

where κ is called the smoothing parameter.

For this smooth function, Zheng[25] gave the following two important Lemmas.

Lemma 1. *For any $\kappa > 0$, $S_{\tau,\kappa}(u)$ is a convex function about u.*

Lemma 2. *For $\forall u \in \mathbb{R}$, there is $0 < S_{\tau,\kappa}(u) - \rho_\tau(u) \le \kappa \log 2$, so $\lim\limits_{\kappa \to 0^+} S_{\tau,\kappa}(u) = \rho_\tau(u)$.*

Lemma 1 indicates function $S_{\tau,\kappa}(u)$ is a convex function, so the global optimal solution can be found. Lemma 2 indicates that $S_{\tau,\kappa}(u)$ can well approximate $\rho_\tau(u)$ when K is small. It is worth noting that the approximation relationship between the two functions does not depend on the quantile τ. Zheng recommended that κ is 0.1 in $S_{\tau,\kappa}(u)$.

According to Lemma 2, there is an approximation relationship between the coefficient estimates obtained under two different loss functions. If the model form of general quantile regression is assumed to be $Q_y(\tau; \boldsymbol{\eta}|x_i, \boldsymbol{z_i})$. For loss function $\rho_\tau(u)$, the quantile regression coefficient obtained is $\hat{\boldsymbol{\eta}} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(y_i - Q_y(\tau; \boldsymbol{\eta}|x_i, \boldsymbol{z_i}))/n$, and for the loss function $S_{\tau,\kappa}(u)$, the quantile regression coefficient obtained is $\hat{\boldsymbol{\eta}}_\kappa = \underset{\boldsymbol{\eta}}{\operatorname{argmin}} \sum_{i=1}^n S_{\tau,\kappa}(y_i - Q_y(\tau; \boldsymbol{\eta}|x_i, \boldsymbol{z_i}))/n$. Then there is the following Theorem 4.

Theorem 4. *On the basis of Lemma (1) and (2),*

$$\lim_{\kappa \to 0^+} \hat{\boldsymbol{\eta}}_\kappa = \hat{\boldsymbol{\eta}}.$$

Theorem 4 indicates that the limit relationship between the coefficient estimates of the two. So for quantile regression, using $S_{\tau,\kappa}(u)$ to replace $\rho_\tau(u)$ is feasible for coefficient estimation.

The loss function at this time is as follows:

$$Ls_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n S_{\tau,\kappa}\left(y_i - \alpha_0 - \alpha_1 x_i - \sum_{k=1}^K \beta_k (x_i - \delta_k) I(x_i > \delta_k) - \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{z_i}\right). \tag{14}$$

After the loss function is replaced, the derivative of the independent variable is

$$s_{\tau,\kappa}(u) = \frac{\partial S_{\tau,\kappa}(u)}{\partial u} = \tau - \frac{1}{1 + \exp(\frac{u}{\kappa})}. \tag{15}$$

Actually, we can still use gradient descent algorithm to get the parameter estimates. Although calculation is effcient, the parameter estimates using this algorithm in high-dimensional cases are very poor. What's more, the variance estimates of parameter estimates is far from the true value. Therefore, we use boosting method to obtain the parameter estimates as Zheng[25] suggested. This method is similar to the coordinate gradient descent method in high dimension, and is based on the boosting algorithm framework. This algorithm framework was proposed by Friedman[26] and Mason et al.[27]. We briefly introduce this method here.

The boosting algorithm is an algorithm based on weak learning machine, which can be considered as a functional gradient descent algorithm. Suppose $\mathcal{K}$ is a set of weak learning machines, $\mathcal{K} = \{k_1(x), \cdots, k_m(x)\}$, where m is the number of weak learning machines. By taking the weak learning machine as a set of bases in the space, the space formed by the weak learning machine can be expressed as

$$\mathcal{L}(\mathcal{K}) = \left\{ \sum_{t=1}^{m} c_t k_t(x) | c_t \in \mathbb{R}, t = 1, 2, \cdots, m \right\}.$$

For the multi-kink quantile regression model (1), we can see

$$\mathcal{K}(x_i, \boldsymbol{z_i}, \boldsymbol{\delta}) = [\mathbf{1}, \boldsymbol{x}, (\boldsymbol{x} - \delta_1)I(\boldsymbol{x} > \delta_1), \cdots, (\boldsymbol{x} - \delta_K)I(\boldsymbol{x} > \delta_K), \boldsymbol{z}]$$

as a set of base vectors of the space $\mathcal{L}(\mathcal{K})$ composed of weak learning machines, where $\boldsymbol{z}$ is explanatory variables. Therefore, according to the form of model (1), $\boldsymbol{\eta}$ can be regarded as the coordinates of linear representation of the $\mathcal{K}(x_i, \boldsymbol{z_i}, \boldsymbol{\delta})$ under this set of base vectors.

Therefore, the problem is transformed into the form of the following problems:

$$Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \boldsymbol{z_i}) = \underset{Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \boldsymbol{z_i}) \in \mathcal{L}(\mathcal{K})}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau [y_i - Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \boldsymbol{z_i})].$$

But given the loss function of quantile regression $\rho_\tau(u)$ for the non differentiable defect at $u = 0$ position, the gradient descent algorithm based on the framework of boosting algorithm can be used only after the original loss function is smoothed. Therefore, the formula (13) is used to replace the loss function. In this way, the problem of solving the model is transformed into the following form:

$$Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \boldsymbol{z_i}) = \underset{Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \boldsymbol{z_i}) \in \mathcal{L}(\mathcal{K})}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} S_{\tau,\kappa} [y_i - Q_Y(\tau; \boldsymbol{\eta}, \boldsymbol{\delta} | x_i, \boldsymbol{z_i})].$$

Thus, based on the boosting algorithm for quantile regression method in Zheng[25], a coefficient boosting algorithm applied to multi-kink quantile regression is given. The algorithm mainly needs to use the negative gradient based on the coefficient of the original model for calculation, but here the functional relationship between the gradient and the original explanatory variable needs to be taken as the iterative increment of the model expression. The Delta method is mainly used for variance estimation, and the specific method is consistent with that in Sections 2.1 and 2.2.

Algorithm 2 gives the flow of boosting improved algorithm based on the smoothed loss function. We call this algorithm as “BSQreg”(Boosting multi-kink Quantile regression). At the same time, Zheng[25] proved that this method has good variable selection function and can be used for the data sets that conform to the feature selection mechanism.

Algorithm 2: BSQreg Algorithm

1: Initialize an initial iteration value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\eta}^{(0)^{\mathrm{T}}}, \boldsymbol{\delta}^{(0)^{\mathrm{T}}})^{\mathrm{T}}$. At the same time set the iteration tolerance of the algorithm to ε, and the maximum number of iterations of the algorithm to T.

2: Calculate the kink location estimation $\hat{\boldsymbol{\delta}}$ with linearization technique.

3: For algorithm iterations $t = 1 : T$:

4: Calculate the gradient of the loss function in the iteration t is

$$s_i = \tau - \left[1 + \exp\left(\frac{y_i - Q_Y(\tau; \boldsymbol{\eta}^{(t)}, \boldsymbol{\delta}|x_i, \boldsymbol{z_i})}{\kappa}\right)\right]^{-1}, \quad i = 1, 2, \cdots, n$$

5: Fit the relationship function $d^{(t)}(x_i)$ between s_i and x_i. And use it as an iteration increment.

6: Update the function $Q_Y(\cdot)$ by $Q_Y(\tau; \boldsymbol{\eta}^{(t)}, \boldsymbol{\delta}|x_i, \boldsymbol{z_i}) = Q_Y(\tau; \boldsymbol{\eta}^{(t)}, \boldsymbol{\delta}|x_i, \boldsymbol{z_i}) + r^{(t)}d^{(t)}(x_i)$. The iteration of $r^{(t)}$ is consistent with the learning rate change method in Section 2.1.

7: Judge the convergence condition, that is, for the loss function (14), if $|Ls_n^{(t+1)}(\boldsymbol{\theta}) - Ls_n^{(t)}(\boldsymbol{\theta})| < \varepsilon$, the iteration is terminated.

8: End the cycle. Output model expressions.

2.4 Induced Smoothing Method

The last improved algorithm is utilizing the induced smoothing method proposed by Wang and Brown[28]. The induced Smoothing method can be used to obtain the parameter estimates and the estimate of the variance for the parameter estimates simultaneously. Many researchers have extended this method to different statistical models. For example, Wang and Fu[29] extended this method to the rank regression model with longitudinal data, Pang et al.[30] extended this method to the quantile regression with censored data, etc. In this paper, the algorithm is extended to multiple fold quantile regression model to improve the algorithm. The induced smoothing method is briefly introduced below.

For the general estimation problem, let $S(\theta)$ be the function of parameter θ, and the estimator $\hat{\theta}$ of θ can be obtained through $S(\theta) = 0$. If $S(\theta)$ is a smooth function, it can be derived naturally, and $S(\hat{\theta})$ can be expanded at the true value θ_0, i.e.

$$0 = S(\hat{\theta}) = S(\theta_0) + S'(\theta_0)(\hat{\theta} - \theta_0) + o_p(\theta_0).$$

At the same time, $S'(\theta_0)$ is usually approximate to a constant, and the variance of $S(\theta_0)$ is expressed as $\mathrm{Cov}[S(\theta_0)]$, and with the sample size $n \to \infty$, $\mathrm{Cov}(S(\theta_0))$ converges to a Positive semidefinite matrix. On this basis, the limit theorem of estimating $\hat{\theta}$ can be derived, and the asymptotic covariance matrix of estimating $\hat{\theta}$ can be constructed by using Delta method as

$$\mathrm{Cov}(\hat{\theta}) = D^{-1}CD^{-1},$$

where $C = \mathrm{Cov}[S(\theta_0)]$,$D = E[S'(\theta_0)]$.

However, the above premise is that $S(\theta)$ is a sufficiently smooth function, otherwise the estimation accuracy and estimation variance cannot be well defined.

Therefore, Wang and brown proposed the induced smoothing method based on the relevant theory of “pseudo Bayes”, which can well obtain the parameter estimation $\hat{\theta}$ and its covariance $\mathrm{Cov}[\hat{\theta}]$. The core idea is that the difference between the true value θ of the parameter and the estimated $\hat{\theta}$ usually satisfies the asymptotic normal distribution

$$\hat{\theta} - \theta \xrightarrow{d} \mathbf{N}(0, \Sigma).$$

Therefore, the focus of the problem is put on the asymptotic normal distribution. Specifically, if θ is a vector form $\boldsymbol{\theta}$, let

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + H^{1/2} Z, \tag{16}$$

where $Z \sim \mathbf{N}(\mathbf{0}, I_p)$, I_p is a p order unit matrix. For the single parameter problem, $H = s^2$, where the standard deviation s plays a role similar to the bandwidth h in kernel smoothing. Therefore, θ at this time can be regarded as the form of random disturbance for estimating $\hat{\theta}$, and the difference between the two obeys a normal distribution with zero mean. At this time, the induced smoothing function of the estimation function is

$$\tilde{S}(\boldsymbol{\theta}) = E_Z[S(\theta + H^{1/2} Z)], \tag{17}$$

Solve the estimation equation $\tilde{S}(\boldsymbol{\theta}) = \mathbf{0}$ to get the estimation $\hat{\boldsymbol{\theta}}$. At the same time, because of the smooth nature of function $\tilde{S}(\cdot)$, the variance H can also be estimated by ”sandwich formula”, i.e

$$H = \{E[\tilde{S}'(\hat{\boldsymbol{\theta}})]\}^{-1} \mathrm{Cov}[S(\theta_0)] \{E[\tilde{S}'(\hat{\boldsymbol{\theta}})]\}^{-1}. \tag{18}$$

(17) and (18) give the induced smoothing estimation method of parameter estimation $\hat{\boldsymbol{\theta}}$ and its variance H.

The iterative updating formula of the algorithm is derived below. The iterative algorithm here is mainly aimed at the coefficient $\boldsymbol{\eta}$ of the model (1). For the kinks $\boldsymbol{\delta}$ of the model, the linearization technique is mainly used for estimation. After the kink estimation $\boldsymbol{\delta}$ is obtained, the induced smoothing algorithm is used to calculate the model coefficient estimation $\hat{\boldsymbol{\eta}}$.

First, let $p(\boldsymbol{w_i}) = p(x_i, \boldsymbol{z_i}) = \left[1, x_i, (x_i - \delta_1) I(x_i > \delta_1), \cdots, (x_i - \delta_K) I(x_i > \delta_K), \boldsymbol{z_i}^{\mathrm{T}}\right]^{\mathrm{T}}$. On the basis of integrating the variables of the model into $p(\boldsymbol{w_i})$, the target loss function can be written as

$$L_n(\boldsymbol{\eta}) = n^{-1} \sum_{i=1}^{n} \rho_\tau [y_i - \boldsymbol{\eta}^{\mathrm{T}} p(\boldsymbol{w_i})],$$

In this case, the derivative of the loss function (18) can be obtained, and the derivative form of the loss function is

$$U_n(\boldsymbol{\eta}) = n^{-1} \sum_{i=1}^{n} \psi_\tau [y_i - \boldsymbol{\eta}^{\mathrm{T}} p(\boldsymbol{w_i})] p(\boldsymbol{w_i}), \tag{19}$$

Here for $U_n(\boldsymbol{\eta})$ in formula (19) can be induced to smooth. The expected part about the indicative function is calculated as follows:

$$E_V \left\{ I[y_i - (\boldsymbol{\eta} + H^{1/2} V)^{\mathrm{T}} p(\boldsymbol{w_i}) < 0] \right\} = P[y_i - \boldsymbol{\eta}^{\mathrm{T}} p(\boldsymbol{w_i}) < p(\boldsymbol{w_i})^{\mathrm{T}} H^{1/2} V].$$

Let $\tilde{V} = p(\boldsymbol{w_i})H^{1/2}V$, then $\tilde{V} \sim \mathbf{N}(\mathbf{0}, p^{\mathrm{T}}(\boldsymbol{w_i})Hp(\boldsymbol{w_i}))$. Thus

$$E_V\left\{I[y_i - (\boldsymbol{\eta} + H^{1/2}V)^{\mathrm{T}}p(\boldsymbol{w_i}) < 0]\right\} = P[\tilde{V} > y_i - \boldsymbol{\eta}^{\mathrm{T}}p(\boldsymbol{w_i})] = \Phi\left(\frac{\boldsymbol{\eta}^{\mathrm{T}}p(\boldsymbol{w_i}) - y_i}{\sqrt{p^{\mathrm{T}}(\boldsymbol{w_i})Hp(\boldsymbol{w_i})}}\right).$$

For $V \sim \mathbf{N}(\mathbf{0}, I_{(K+2+p)})$, we can obtain the induced smooth function of the estimation function

$$\begin{aligned}
\tilde{U}(\boldsymbol{\eta}, H) &= E_V[U(\boldsymbol{\eta} + H^{1/2}V)] \\
&= E_V[n^{-1}\sum_{i=1}^{n} p(\boldsymbol{w_i})\{\tau - I[y_i - (\boldsymbol{\eta} + H^{1/2}V)^{\mathrm{T}}p(\boldsymbol{w_i})]\}] \\
&= n^{-1}\sum_{i=1}^{n} p(\boldsymbol{w_i})\left[\tau - \Phi\left(\frac{\boldsymbol{\eta}^{\mathrm{T}}p(\boldsymbol{w_i}) - y_i}{\sqrt{p^{\mathrm{T}}(\boldsymbol{w_i})Hp(\boldsymbol{w_i})}}\right)\right].
\end{aligned} \tag{20}$$

At the same time, to use the gradient descent algorithm for $\boldsymbol{\eta}$, we need to derive $\tilde{U}(\boldsymbol{\eta}, H)$, and the derivative is

$$\tilde{A}(\boldsymbol{\eta}, H) = \frac{\partial \tilde{U}(\boldsymbol{\eta}, H)}{\partial \boldsymbol{\eta}} = n^{-1}\sum_{i=1}^{n}\left[-\phi\left(\frac{\boldsymbol{\eta}^{\mathrm{T}}p(\boldsymbol{w_i}) - y_i}{\sqrt{p^{\mathrm{T}}(\boldsymbol{w_i})Hp(\boldsymbol{w_i})}}\right)\frac{p(\boldsymbol{w_i})p^{\mathrm{T}}(\boldsymbol{w_i})}{\sqrt{p(\boldsymbol{w_i})^{\mathrm{T}}Hp(\boldsymbol{w_i})}}\right].$$

Thus, the iterative formula for the coefficient vector $\boldsymbol{\eta}$ is

$$\boldsymbol{\eta}^{(t+1)} = \boldsymbol{\eta}^{(t)} + [-\tilde{A}(\boldsymbol{\eta}^{(t)}, H^{(t)})]^{-1}\tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)}), \tag{21}$$

and the calculation formula for each iteration of $H^{(t)}$ can follow the calculation method of (18). The specific calculation formula is as follows:

$$H^{(t+1)} = n^{-1}\tilde{A}^{-1}(\boldsymbol{\eta}^{(t)})\mathrm{Cov}[\tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)})]\tilde{A}^{-1}(\boldsymbol{\eta}^{(t)}). \tag{22}$$

In fact, the H finally obtained by the iterative algorithm is the variance estimation of the parameter estimation. The covariance $\mathrm{Cov}[\tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)})]$ in (22) can also be calculated by using the sandwich formula in Section 4.1. At the same time, it can also be proved that when $n \to \infty$, $\mathrm{Cov}[\tilde{U}(\boldsymbol{\eta}, H)]$ and $\mathrm{Cov}[U(\boldsymbol{\eta}, H)]$ are asymptotically consistent.

The algorithm 3 gives the specific calculation flow of the induced smoothing method. The algorithm has the following obvious advantages: the final parameter estimation and variance estimation are obtained directly in the process of the algorithm, and there is no need to solve the variance estimation; The algorithm has a strong theoretical guarantee; The algorithm can converge in a few steps and has a fast convergence speed.

3 Simulation

In this part, we will exhibit some simulation experiment results of our improved algorithms. Then we give some evaluations on our different improved algorithms through these experiment resuls.

Algorithm 3: InducedSmooth Algorithm

1: Initialize the initial iteration value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\eta}^{(0)^{\mathrm{T}}}, \boldsymbol{\delta}^{(0)^{\mathrm{T}}})^{\mathrm{T}}$, of the parameters to be evaluated according to the steps mentioned in Section 4.1. Initialize matrix $H^{(0)} = n^{-1}I_n$. At the same time, set the iteration tolerance of the algorithm to ε, and the maximum number of iterations of the algorithm to T.

2: Calculate the kink location estimation $\hat{\boldsymbol{\delta}}$ with linearization technique.

3: For algorithm iterations $t = 0 : T$:

4: Calculate the matrix $H^{(t)}$ in iteration $t$and the intermediate amount $\tilde{U}(\boldsymbol{\eta}^{(t)}, H^{(t)})$ and $\tilde{A}(\boldsymbol{\eta}^{(t)}, H^{(t)})$. Calculate $\boldsymbol{\eta}^{(t+1)}$ through (21). Use the Delta method and formula (22) to calculate H^{t+1}.

5: Judge the convergence condition, that is, for the loss function

$$L_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\rho_\tau[y_i - Q_Y(\tau; \boldsymbol{\theta}|x_i, \boldsymbol{z_i})],$$

if $|L_n^{(t+1)}(\boldsymbol{\theta}) - L_n^{(t)}(\boldsymbol{\theta})| < \varepsilon$, the iteration is terminated.

6: End the cycle. Output $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\eta}}^{\mathrm{T}}, \hat{\boldsymbol{\delta}}^{\mathrm{T}})^{\mathrm{T}}$ as the final parameter estimation, and output diagonal matrix $\hat{H}$ as the variance estimation of the final parameter estimation.

3.1 Data Generation

We generate data from the following model:

$$y_i = \alpha_0 + \alpha_1 x_i + \sum_{k=1}^{K}\beta_k(x_i - \delta_k)I(x_i > \delta_k) + \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{z_i} + \sigma(x_i, z_i)e_i, \quad i = 1, 2, \cdots, n. \quad (23)$$

where $x_i \sim \mathbf{U}(-5, 5)$,$\boldsymbol{z_i} \sim \mathbf{N}(1, 1)$. In this model, we set $\alpha_0 = 1$,$\alpha_1 = 1$ and $\boldsymbol{\gamma} = 1$. We set $\sigma(x_i, z_i)$ for homoscedasticity and heteroscedasticity situations. For homoscedasticity situation, $\sigma(x_i, z_i) = 1$ and For heteroscedasticity situation, $\sigma(x_i, z_i) = 1 + 0.2x_i$. We consider three cases for kink number: $K = 1, 2, 3$. For $K = 1$, we set $\beta_1 = -3$ and $\delta_1 = 0.5$; For $K = 2$, $\beta_1 = -3, \beta_2 = 4$ and $\delta_1 = -1, \delta_2 = 2$; For $K = 3$, $\beta_1 = -3, \beta_2 = 4, \beta_3 = -4$ and $\delta_1 = -3, \delta_2 = 0, \delta_3 = 3$. For the random error term e_i, we set three different distributions: $e_i \sim \mathbf{N}(0, 1)$,$e_i \sim \mathbf{t}(3)$ and $e_i \sim \mathbf{LA}(0, 1)$. To testify the rationality of our improved algorithms in different quantile levels, we set three important quantile levels in statistics for our simulation: $\tau = 0.25, 0.5$ and 0.75.

3.2 Result Exhibition

In our experiment, we set sample size $n = 1000$. And for experiments of each situation, we all make 1000 replications to give Monte-Carlo simulation results. Due to space constraints, we only give results of several situations on $K = 2$ in Table 1 $\sim$ Table 12, and results of $K = 1$ and $K = 3$ can be seen in supplementary.

Focusing on the overall estimation results under two different scedasticities, we can see that for the homoscedasticity situation, no matter which method is used to estimate

Table 1: $K = 2$,Homoscedasticity,$e_i \sim \mathbf{N}(0,1)$,$\tau = 0.25$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.002	0.000	0.000	-0.006	0.003	0.002	-0.004
BRISQ	**SD**	0.212	0.042	0.063	0.109	0.129	0.077	0.058
	SE	0.195	0.042	0.059	0.109	0.129	0.070	0.056
	MSE	0.038	0.017	0.035	0.012	0.017	0.005	0.003
	Bias	0.029	0.006	0.000	0.131	-0.216	-0.083	0.018
SmoothKernel	**SD**	0.222	0.066	0.044	0.110	0.132	0.078	0.054
	SE	0.457	0.139	0.240	0.286	0.096	0.464	0.172
	MSE	0.210	0.019	0.057	0.099	0.056	0.222	0.030
	Bias	0.041	0.007	-0.004	0.111	-0.171	-0.077	0.022
BentCableKink	**SD**	0.232	0.067	0.043	0.113	0.133	0.080	0.052
	SE	0.476	0.144	0.251	0.298	0.097	0.514	0.191
	MSE	0.229	0.021	0.063	0.101	0.039	0.270	0.037
	Bias	-0.037	-0.007	0.000	0.029	-0.041	-0.006	-0.002
BSQreg	**SD**	0.188	0.057	0.042	0.107	0.124	0.076	0.061
	SE	0.441	0.136	0.245	0.292	0.097	0.463	0.168
	MSE	0.196	0.018	0.060	0.086	0.011	0.215	0.028
	Bias	0.007	0.006	0.013	0.036	-0.109	-0.005	-0.001
InducedSmooth	**SD**	0.211	0.060	0.041	0.100	0.116	0.076	0.053
	SE	0.226	0.064	0.047	0.101	0.118	0.079	0.055
	MSE	0.051	0.004	0.002	0.011	0.026	0.006	0.003

the multi-kink quantile regression model, the estimation effect is reasonable. The absolute value of most estimated bias is less than 0.010, and the bias of many parameter estimations is even less than 0.005, reaching about 0.001. Only the absolute value of estimated bias in several special cases of some algorithms reaches about 0.05. Even the maximum bias has not reached 0.1. Moreover, it is not difficult to find that for the simulation results of all algorithms under the homoscedasticity situation, the random error term $e_i \sim \mathbf{n}(0,1)$, the bias of simulation results will be controlled at a better level. This is predictable, because the magnitude of the noise added to the original model (23) will affect the final estimation result. The greater the noise is added, the greater the external impact on the data itself, thus covering the real data information, and there is A large degree of interference to the model identification and estimation, so the estimation effect is usually worse.For these random error noise terms with different distributions, due to the heavy tailed nature of the distribution itself, it is considered that the added noise is the largest when the random error term $e_i \sim \mathbf{t}(3)$, so the estimation effect is relatively poor. And for the condition that $e_i \sim \mathbf{LA}(0,1)$, the density function of Laplace distribution is similar to the loss function of quantile regression, so the estimation effect is also slightly better.

However, different quantiles represent different effects. The estimation effect of me-

Table 2: $K = 2$,Homoscedasticity,$e_i \sim \mathbf{t}(3)$， $\tau = 0.25$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.012	-0.001	0.003	-0.014	0.015	0.000	-0.003
BRISQ	**SD**	0.244	0.054	0.074	0.143	0.171	0.091	0.074
	SE	0.243	0.052	0.074	0.137	0.163	0.087	0.070
	MSE	0.059	0.003	0.006	0.019	0.027	0.008	0.005
	Bias	0.019	0.005	-0.003	0.186	-0.295	-0.109	0.026
SmoothKernel	**SD**	0.287	0.086	0.054	0.135	0.160	0.103	0.065
	SE	0.582	0.177	0.297	0.355	0.122	0.583	0.221
	MSE	0.339	0.031	0.088	0.161	0.102	0.352	0.050
	Bias	0.061	0.014	0.000	0.114	-0.180	-0.091	0.027
BentCableKink	**SD**	0.291	0.085	0.053	0.133	0.159	0.098	0.065
	SE	0.614	0.185	0.320	0.379	0.122	0.674	0.256
	MSE	0.381	0.034	0.103	0.156	0.047	0.463	0.066
	Bias	0.017	0.012	0.005	0.041	-0.135	-0.017	0.000
BSQreg	**SD**	0.233	0.069	0.044	0.141	0.161	0.096	0.072
	SE	0.270	0.079	0.155	0.183	0.058	0.297	0.109
	MSE	0.073	0.006	0.024	0.035	0.022	0.088	0.012
	Bias	0.017	0.012	0.005	0.041	-0.135	-0.017	0.000
InducedSmooth	**SD**	0.233	0.069	0.044	0.141	0.161	0.096	0.072
	SE	0.236	0.069	0.048	0.148	0.167	0.100	0.076
	MSE	0.056	0.005	0.002	0.024	0.046	0.010	0.006

dian regression is slightly better than 1/4 quantile regression and 3/4 quantile regression. We take InducedSmooth algorithm as an example, for Heteroscedasticity situation ,$e_i \sim \mathbf{t}(3)$ and $\tau = 0.5$ quantile level, the estimation bias of second kink coefficient β_2 is the largest, reaching 0.018, but still far less than the maximum bias of 0.076 in the case of $\tau = 0.25$, which is more than four times as large. From the data itself, the data near the median should be the most intensive, which can often better reflect the essential information of the data, so the estimation effect is better. For the quantile regression simulation of extreme position, the sample size of simulation is often increased to improve the estimation accuracy of test results, then the robustness of estimation is achieved through the large sample statistical nature of data. Due to the limitation of article length, we will not give too many unnecessary details here. But actually, for all the above algorithms, no matter how much the corresponding quantile level τ is taken, the absolute value of the estimated bias is rarely greater than 0.1, which shows that the estimation effect of the above algorithms is generally good.

The estimation results are also reasonable in the case of heteroscedasticity. At this time, the data noise and fluctuation are larger than those in the case of homoscedasticity, but the absolute value of estimation bias is still tiny. Among all the estimation

Table 3: $K = 2$,Homoscedasticity,$e_i \sim \mathbf{LA}(0, 1)$，$\tau = 0.25$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
BRISQ	**Bias**	-0.003	0.003	0.002	-0.008	0.011	0.001	0.000
	SD	0.248	0.056	0.073	0.139	0.167	0.093	0.075
	SE	0.249	0.053	0.076	0.139	0.166	0.089	0.072
	MSE	0.062	0.003	0.006	0.019	0.028	0.008	0.005
SmoothKernel	**Bias**	0.001	0.000	0.001	0.176	-0.274	-0.100	0.024
	SD	0.277	0.081	0.056	0.135	0.167	0.103	0.064
	SE	0.588	0.179	0.302	0.363	0.123	0.588	0.223
	MSE	0.346	0.032	0.091	0.163	0.091	0.356	0.050
BentCableKink	**Bias**	0.027	0.006	0.002	0.126	-0.192	-0.086	0.026
	SD	0.278	0.081	0.056	0.139	0.174	0.104	0.064
	SE	0.620	0.186	0.322	0.383	0.123	0.671	0.253
	MSE	0.385	0.035	0.104	0.162	0.052	0.458	0.065
BSQreg	**Bias**	-0.075	-0.014	0.003	0.057	-0.076	-0.010	0.000
	SD	0.238	0.069	0.052	0.129	0.159	0.095	0.073
	SE	0.562	0.173	0.310	0.369	0.123	0.587	0.215
	MSE	0.322	0.030	0.096	0.140	0.021	0.345	0.046
InducedSmooth	**Bias**	-0.009	-0.012	0.010	0.032	-0.132	-0.002	-0.004
	SD	0.241	0.074	0.050	0.122	0.115	0.088	0.064
	SE	0.254	0.076	0.054	0.124	0.117	0.076	0.058
	MSE	0.065	0.006	0.003	0.016	0.031	0.006	0.003

results, some absolute value of bias is less than 0.001, and the maximum value of the absolute value of bias does not exceed 0.1. Although the estimation bias is tiny, the estimation effect under heteroscedasticity conditions is generally not as good as that under homoscedasticity conditions, which is related to data noise itself.For different quantile levels, although the estimation effect is different, the overall parameter estimation effect is indeed very good, which shows that the above algorithms are universal.

Next, we make analysis based on different algorithms. Although all the algorithms have good performance under various conditions, there are still more or less differences in effect and efficiency. Firstly, we can see that the overall estimation effect of the BRISQ algorithms proposed by Zhong et al. is perfect. Not only the bias of parameter estimation is very small, but also the effect of variance estimation is remarkable. It can be seen that in BRISQ simulation results, the difference between the empirical standard deviations (SD) and the average estimated standard errors (SE) is miniscule. This phenomenon can explain two points: (a) BRISQ proposed by Zhong et al. can well perform the task of estimating the standard deviation of parameters in the model; (b) At this time, the parameter estimation results are more in line with the expectation that parameters have asymptotic normality. On account of little difference between truth-value and estimation,

Table 4: $K = 2$,Homoscedasticity,$e_i \sim \mathbf{N}(0, 1)$， $\tau = 0.5$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	-0.005	0.001	0.000	-0.006	0.007	0.004	-0.001
BRISQ	**SD**	0.178	0.040	0.055	0.102	0.120	0.067	0.053
	SE	0.179	0.039	0.055	0.101	0.120	0.064	0.052
	MSE	0.032	0.002	0.003	0.010	0.014	0.004	0.003
	Bias	0.024	0.004	-0.004	0.113	-0.181	-0.067	0.018
SmoothKernel	**SD**	0.203	0.061	0.041	0.099	0.117	0.071	0.051
	SE	0.365	0.110	0.193	0.232	0.078	0.375	0.138
	MSE	0.133	0.012	0.037	0.066	0.039	0.145	0.019
	Bias	0.044	0.009	-0.002	0.080	-0.128	-0.059	0.019
BentCableKink	**SD**	0.208	0.063	0.040	0.106	0.120	0.071	0.051
	SE	0.379	0.114	0.201	0.237	0.078	0.412	0.150
	MSE	0.146	0.013	0.040	0.063	0.022	0.173	0.023
	Bias	-0.030	-0.008	0.001	0.030	-0.041	-0.005	-0.002
BSQreg	**SD**	0.185	0.056	0.031	0.096	0.114	0.076	0.058
	SE	0.358	0.109	0.199	0.236	0.079	0.378	0.136
	MSE	0.129	0.012	0.040	0.057	0.008	0.143	0.019
	Bias	-0.004	-0.002	0.001	0.025	-0.040	0.003	-0.007
InducedSmooth	**SD**	0.162	0.048	0.038	0.084	0.114	0.065	0.051
	SE	0.160	0.058	0.044	0.044	0.086	0.112	0.059
	MSE	0.026	0.003	0.002	0.003	0.009	0.013	0.004

it also shows that this method better describes the effectiveness of parameter estimation, thus ensuring the precision in calculation of mean square error (MSE) of each parameter.

For other improved algorithms, it can be seen that the SmoothKernel algorithm has small bias in parameter estimation, but there is a certain difference between the empirical standard deviations (SD) and the average estimated standard errors (SE). Although the difference is small and can be basically controlled within 0.2, the average estimated standard errors (SE) is always overestimated compared with the empirical standard deviations (SD). This may also have some impact on the confidence interval construction of parameters. However, due to the smoothing of the indicative function and the use of the gradient descent algorithm, the computational efficiency of the estimation is greatly improved, for some cases we can get estimation with 2 seconds, the efficiency is much higher than that of BRISQ.

The BentCableKink algorithm also has a small parameter estimation bias. The problem still lies in the error of variance estimation. In fact, the error of variance estimation of this algorithm is predictable. BentCableKink itself is an algorithm that approximates with quadratic function at the kink locations, essentially changes the form of the model, which is bound to have a certain impact on the variance estimation of parameters. Moreover,

Table 5: $K=2$,Homoscedasticity,$e_i \sim \mathbf{t}(3)$, $\tau=0.5$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.013	0.001	0.005	-0.012	0.003	-0.001	-0.003
BRISQ	**SD**	0.193	0.043	0.059	0.112	0.129	0.075	0.060
	SE	0.198	0.043	0.060	0.110	0.130	0.070	0.056
	MSE	0.039	0.002	0.004	0.012	0.017	0.005	0.003
	Bias	0.015	0.003	-0.001	0.152	-0.246	-0.087	0.018
SmoothKernel	**SD**	0.230	0.068	0.044	0.111	0.129	0.077	0.054
	SE	0.403	0.122	0.208	0.246	0.085	0.408	0.151
	MSE	0.162	0.015	0.043	0.084	0.068	0.174	0.023
	Bias	0.045	0.010	-0.000	0.077	-0.124	-0.062	0.019
BentCableKink	**SD**	0.228	0.068	0.043	0.107	0.131	0.078	0.054
	SE	0.425	0.128	0.225	0.264	0.085	0.473	0.175
	MSE	0.183	0.016	0.051	0.075	0.023	0.228	0.031
	Bias	-0.087	-0.026	-0.007	0.100	-0.138	-0.001	-0.014
BSQreg	**SD**	0.234	0.105	0.093	0.269	0.350	0.125	0.182
	SE	0.387	0.118	0.214	0.253	0.086	0.409	0.150
	MSE	0.157	0.015	0.046	0.074	0.026	0.167	0.023
	Bias	0.014	0.006	-0.003	0.028	-0.041	-0.007	-0.001
InducedSmooth	**SD**	0.191	0.057	0.037	0.109	0.119	0.062	0.056
	SE	0.195	0.051	0.040	0.112	0.132	0.067	0.049
	MSE	0.038	0.003	0.002	0.013	0.019	0.005	0.002

the calculation cost of this algorithm is large relatively, because the algorithm involves the iteration of the bandwidth h. The advantage of this algorithm is that it can depict the sharp change of the bent line more flexibly, which makes the kink phenomenon tend to be gentle. On account of everywhere differentiability of model in this condition, it also increases the universality of the application.

The BSQreg algorithm uses a new loss function and boosting algorithm framework. The algorithm not only perform effectively in parameter estimation, but also has a good feature selection mechanism.

The InducedSmooth algorithm is the method that we highly recommend in the improvements. This method, which combines the induce smoothing framework proposed by Wang and brown, not only has a good effect on parameter estimation (the bias of parameter estimation is very small), but also has a good effect on variance estimation. This is mainly due to the fact that this method does not change the essence of the model while smoothing, so it has a good effect on all estimation in a certain sense, which indirectly explains the asymptotic normality of parameters estimation. Although the algorithm seems complex, it can converge completely in a small number of iteration steps, so the computational cost is not very huge. In addition, the algorithm can directly estimate the

Table 6: $K = 2$,Homoscedasticity,$e_i \sim \mathbf{LA}(0, 1)$， $\tau = 0.5$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	-0.006	0.001	-0.001	-0.003	0.004	0.003	-0.001
BRISQ	**SD**	0.153	0.034	0.045	0.090	0.102	0.058	0.047
	SE	0.158	0.034	0.048	0.089	0.106	0.057	0.046
	MSE	0.025	0.001	0.002	0.008	0.011	0.003	0.002
	Bias	0.005	0.001	-0.001	0.125	-0.196	-0.066	0.018
SmoothKernel	**SD**	0.173	0.051	0.035	0.088	0.102	0.062	0.045
	SE	0.324	0.098	0.171	0.203	0.070	0.332	0.123
	MSE	0.105	0.010	0.029	0.057	0.043	0.115	0.015
	Bias	0.035	0.008	0.000	0.065	-0.105	-0.049	0.016
BentCableKink	**SD**	0.170	0.050	0.034	0.092	0.109	0.063	0.046
	SE	0.339	0.102	0.182	0.215	0.070	0.377	0.139
	MSE	0.116	0.010	0.033	0.050	0.016	0.145	0.020
	Bias	-0.058	-0.015	0.002	0.060	-0.077	-0.010	0.000
BSQreg	**SD**	0.203	0.057	0.034	0.102	0.118	0.095	0.073
	SE	0.319	0.097	0.177	0.210	0.071	0.337	0.124
	MSE	0.105	0.010	0.031	0.048	0.0141	0.114	0.015
	Bias	-0.002	-0.001	-0.004	0.021	-0.032	0.004	-0.001
InducedSmooth	**SD**	0.146	0.044	0.030	0.098	0.109	0.060	0.045
	SE	0.147	0.047	0.035	0.120	0.111	0.055	0.041
	MSE	0.022	0.002	0.001	0.015	0.013	0.003	0.002

standard deviation of all parameters through the iteration of matrix $\hat{H}$ during the iteration process without additional standard error estimation At the same time, the iterative algorithm has no strict requirements for the selection of initial values. The initial value of matrix $\hat{H}$ can be selected as the unit matrix, which shows that the algorithm has good generalization property.

To sum up, the method of Zhong et al. and the InducedSmooth algorithm have obvious advantages, and they are more worthy of promotion through our simulations.

Table 7: $K = 2$,Heteroscedasticity,$e_i \sim \mathbf{N}(0, 1)$, $\tau = 0.25$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.006	0.000	0.002	-0.006	0.009	-0.001	-0.002
BRISQ	**SD**	0.095	0.018	0.020	0.106	0.018	0.056	0.084
	SE	0.090	0.016	0.019	0.101	0.185	0.054	0.078
	MSE	0.008	0.000	0.000	0.010	0.034	0.003	0.006
	Bias	0.014	0.003	-0.001	0.150	-0.028	-0.008	0.012
SmoothKernel	**SD**	0.092	0.020	0.017	0.093	0.190	0.058	0.073
	SE	0.211	0.046	0.021	0.101	0.039	0.361	0.245
	MSE	0.045	0.002	0.047	0.033	0.083	0.136	0.060
	Bias	0.035	0.008	-0.001	0.127	-0.254	-0.078	0.012
BentCableKink	**SD**	0.086	0.018	0.018	0.100	0.202	0.061	0.071
	SE	0.214	0.046	0.233	0.428	0.039	0.392	0.269
	MSE	0.047	0.002	0.054	0.199	0.066	0.160	0.072
	Bias	-0.006	0.013	0.001	0.004	-0.026	-0.004	0.002
BSQreg	**SD**	0.094	0.021	0.025	0.102	0.179	0.060	0.080
	SE	0.206	0.044	0.228	0.420	0.038	0.367	0.234
	MSE	0.043	0.002	0.052	0.176	0.002	0.134	0.055
	Bias	0.043	0.016	-0.012	0.025	-0.097	-0.002	-0.002
InducedSmooth	**SD**	0.097	0.022	0.023	0.090	0.164	0.056	0.076
	SE	0.102	0.025	0.024	0.105	0.156	0.062	0.079
	MSE	0.012	0.001	0.001	0.012	0.034	0.004	0.006

4 Empirical Analysis

4.1 Dataset Introduction

Global warming has always been a hot issue in the world. Since the industrial revolution, due to the massive combustion of coal, oil and other energy sources, the global temperature continues to rise. But what is the rising trend of global temperature? Is it static? This is an important problem in meteorology and environmental science. Here we collect 1850-2020 monthly air temperature extremum data from website http://berkeleyearth.lbl.gov/auto/Global to study this problem. We call this dataset as "AirTemp". Obviously, different trends can be reflected by different slopes. If the global temperature shows different trends in different stages, there should be the presence of several kinks. Therefore, here we use multi-kink quantile regression model to solve the problem. We use BRISQ algorithm proposed by Zhong et.al and our improved algorithms to estimate kink locations and other coefficients.

Table 8: $K=2$,Heteroscedasticity,$e_i \sim \mathbf{t}(3)$, $\tau=0.25$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.008	0.000	0.003	-0.015	0.033	0.001	0.001
BRISQ	**SD**	0.112	0.022	0.024	0.133	0.234	0.070	0.104
	SE	0.113	0.021	0.024	0.128	0.233	0.068	0.099
	MSE	0.013	0.000	0.001	0.017	0.055	0.005	0.010
	Bias	0.012	0.003	-0.002	0.194	-0.366	-0.107	0.008
SmoothKernel	**SD**	0.112	0.024	0.021	0.112	0.227	0.073	0.085
	SE	0.264	0.057	0.267	0.519	0.049	0.444	0.311
	MSE	0.047	0.011	-0.001	0.130	-0.262	-0.087	0.015
	Bias	0.061	0.014	0.000	0.114	-0.180	-0.091	0.027
BentCableKink	**SD**	0.108	0.023	0.021	0.119	0.241	0.073	0.083
	SE	0.273	0.059	0.303	0.543	0.050	0.518	0.358
	MSE	0.077	0.004	0.092	0.312	0.071	0.276	0.129
	Bias	0.045	0.018	-0.021	0.023	-0.115	-0.005	-0.003
BSQreg	**SD**	0.107	0.023	0.025	0.131	0.221	0.076	0.099
	SE	0.129	0.028	0.280	0.517	0.048	0.453	0.294
	MSE	0.069	0.003	0.079	0.271	0.013	0.205	0.087
	Bias	0.045	0.018	-0.021	0.023	-0.115	-0.005	-0.003
InducedSmooth	**SD**	0.107	0.023	0.025	0.131	0.221	0.076	0.099
	SE	0.109	0.028	0.026	0.137	0.250	0.079	0.101
	MSE	0.014	0.001	0.001	0.019	0.076	0.006	0.010

4.2 Result Analysis

Firstly, we use kink effect test proposed by Zhong et.al to test kink effect of this dataset. The p-value of the test is much less than 0.05, so there is significant kink effect in this dataset. Then we use strengthened quantile Bayesian information criterion and backward elimination algorithm proposed by Zhong et.al to estimate number of kinks $\hat{K}$, and get the result of $\hat{K}=3$, which means that the dataset can be divided into 4 segmented by the 3 kinks. In this part, in order to compare the advantages and disadvantages of the BRISQ algorithm proposed by Zhong et al. with other algorithm improvements, we use these algorithms to perform multi-kink quantile regression on this dataset, See Table 13 for the comparison of the numerical results of the parameter estimation and variance estimation of the obtained algorithms. See the following figure 1 - figure 5 for the image of multi-kink quantile regression on "AirTemp" dataset. The quantile levels selected here are $\tau=0.1,0.3,0.5,0.7,0.9$, and the color represented by the regression line of each quantile level in the figure is red for $\tau=0.1$, blue for $\tau=0.3$, green for $\tau=0.5$, yellow for $\tau=0.7$ and orange for $\tau=0.9$.

Generally speaking, the three kinks approximately appear at the location of year 1913, year 1943 and year 1976. These three kinks divide the entire dataset into four seg-

Table 9: $K = 2$,Heteroscedasticity,$e_i \sim \mathbf{LA}(0,1)$, $\tau = 0.25$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.002	0.001	0.002	-0.010	0.018	0.000	0.001
BRISQ	**SD**	0.116	0.023	0.025	0.132	0.241	0.072	0.111
	SE	0.115	0.021	0.025	0.130	0.239	0.069	0.101
	MSE	0.013	0.000	0.001	0.017	0.057	0.005	0.010
	Bias	0.007	0.003	0.000	0.191	-0.361	-0.104	0.013
SmoothKernel	**SD**	0.111	0.024	0.022	0.114	0.237	0.079	0.085
	SE	0.268	0.058	0.271	0.529	0.050	0.450	0.315
	MSE	0.072	0.003	0.074	0.316	0.133	0.214	0.099
	Bias	0.036	0.009	0.000	0.145	-0.291	-0.093	0.015
BentCableKink	**SD**	0.100	0.023	0.022	0.122	0.246	0.080	0.084
	SE	0.275	0.059	0.300	0.553	0.050	0.511	0.358
	MSE	0.077	0.004	0.090	0.327	0.087	0.270	0.128
	Bias	-0.043	0.007	0.002	0.027	-0.061	-0.006	-0.001
BSQreg	**SD**	0.118	0.027	0.029	0.125	0.231	0.073	0.105
	SE	0.262	0.056	0.286	0.534	0.048	0.460	0.298
	MSE	0.070	0.003	0.082	0.287	0.006	0.212	0.089
	Bias	0.027	0.014	-0.018	0.030	-0.130	0.002	-0.006
InducedSmooth	**SD**	0.103	0.022	0.026	0.113	0.177	0.068	0.087
	SE	0.120	0.027	0.024	0.123	0.182	0.064	0.083
	MSE	0.015	0.001	0.001	0.016	0.050	0.004	0.007

ments. From 1850 to 1913, the global monthly temperature extremes remained stable and showed a slow growth trend. From 1913 to 1943, the extreme value of global temperature showed an obvious upward trend. Since 1943, the global temperature extremes began to stabilize. Until 1976, the rising trend of global temperature extremes began to become very significant. If we associate the location of these kinks and the rising trend of global temperature with the corresponding historical period, we can roughly analyze the reasons behind them. In 1850, the first industrial revolution had just ended. Human society basically completed the transformation to an industrial society, but it did not complete the transformation of industrialization. Moreover, the industrial revolution at that time was only carried out in a few capitalist countries and did not get the promotion of globalization. Therefore, the global temperature during this period was relatively stable. Around 1913, the world had completed the second industrial revolution. The second industrial revolution has the trend and tide of globalization. During this period, non renewable energy such as coal had become a new driving force for the development of world industry and economy. Therefore, during this period, the global temperature has increased rapidly. At the same time, this period has experienced two world wars. The occurrence of the war had led to a large expenditure of energy such as oil and coal, which has also played a

Table 10: 折 $K = 2$,Heteroscedasticity,$e_i \sim \mathbf{N}(0,1)$，$\tau = 0.5$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.002	0.000	0.000	-0.007	0.014	0.002	0.001
BRISQ	**SD**	0.083	0.017	0.018	0.101	0.168	0.052	0.080
	SE	0.083	0.014	0.018	0.094	0.171	0.050	0.072
	MSE	0.007	0.000	0.000	0.009	0.029	0.003	0.005
	Bias	0.012	0.003	0.000	0.131	-0.257	0.069	0.006
SmoothKernel	**SD**	0.083	0.018	0.016	0.089	0.179	0.053	0.071
	SE	0.166	0.035	0.175	0.332	0.030	0.287	0.195
	MSE	0.028	0.001	0.031	0.127	0.067	0.087	0.038
	Bias	0.024	0.005	-0.001	0.103	-0.211	-0.062	0.006
BentCableKink	**SD**	0.080	0.017	0.016	0.092	0.177	0.054	0.067
	SE	0.169	0.036	0.188	0.340	0.030	0.316	0.213
	MSE	0.029	0.001	0.035	0.126	0.045	0.104	0.045
	Bias	-0.026	-0.006	0.000	0.022	-0.032	-0.001	0.001
BSQreg	**SD**	0.100	0.023	0.022	0.097	0.154	0.057	0.085
	SE	0.166	0.036	0.186	0.340	0.029	0.300	0.191
	MSE	0.028	0.001	0.035	0.116	0.002	0.089	0.037
	Bias	-0.000	0.000	0.002	0.028	-0.047	0.004	-0.011
InducedSmooth	**SD**	0.084	0.019	0.021	0.086	0.159	0.050	0.075
	SE	0.085	0.025	0.019	0.088	0.170	0.053	0.059
	MSE	0.007	0.001	0.000	0.030	0.010	0.007	0.004

positive feedback role in the growth of global temperature. From 1943 to 1976, the third scientific and technological revolution had just begun, some new energy sources were in the ascendant, and the development of world economy and industry still depended on the original energy system; at this time, most countries fell into the mode of post-war reconstruction, unable to develop industry and economy in a continuous way, so the global temperature growth tended to be stable during this period; From 1976 to 2020, the world experienced the third scientific and technological revolution, with rapid economic and social development, a sharp increase in the world population and a large number of cultivated land, which seriously threatened the natural ecology and forest resources, leading to increasingly serious air pollution and a sharp increase in global temperature. It can be seen from the actual results that the trend depicted by the multiple break quantile regression model is consistent with the history. It can also be found from the Table 13 that the variance estimation at $\hat{\delta}_1$ is much larger than the variance estimation at $\hat{\delta}_2$, which indicates that the dataset has great heteroscedasticity. The quantile method is obviously more appropriate for regression analysis of data with large heteroscedasticity.

After paying attention to the overall situation of this dataset, this paper compares the results of these different algorithms. It is not difficult to find that among the result

Table 11: $K = 2$,Heteroscedasticity,$e_i \sim \mathbf{t}(3)$， $\tau = 0.5$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
	Bias	0.003	0.000	0.001	-0.008	0.011	0.000	-0.003
BRISQ	**SD**	0.088	0.017	0.019	0.109	0.190	0.058	0.086
	SE	0.090	0.016	0.019	0.103	0.186	0.055	0.079
	MSE	0.008	0.000	0.000	0.011	0.035	0.003	0.006
	Bias	-0.003	-0.001	-0.002	0.174	-0.325	-0.086	0.007
SmoothKernel	**SD**	0.092	0.019	0.019	0.093	0.189	0.056	0.073
	SE	0.183	0.039	0.188	0.359	0.033	0.313	0.215
	MSE	0.033	0.002	0.035	0.159	0.107	0.106	0.046
	Bias	0.035	0.008	-0.001	0.092	-0.194	-0.060	0.008
BentCableKink	**SD**	0.090	0.019	0.018	0.101	0.197	0.060	0.076
	SE	0.186	0.040	0.215	0.380	0.033	0.364	0.248
	MSE	0.036	0.002	0.046	0.153	0.039	0.136	0.061
	Bias	-0.075	-0.024	-0.007	0.092	-0.129	-0.001	-0.019
BSQreg	**SD**	0.141	0.092	0.092	0.284	0.390	0.113	0.205
	SE	0.179	0.038	0.199	0.363	0.033	0.319	0.210
	MSE	0.038	0.002	0.040	0.140	0.018	0.102	0.044
	Bias	0.002	0.000	-0.002	0.032	-0.038	-0.001	-0.005
InducedSmooth	**SD**	0.089	0.020	0.021	0.103	0.161	0.051	0.082
	SE	0.092	0.028	0.022	0.121	0.127	0.057	0.078
	MSE	0.008	0.001	0.000	0.016	0.018	0.003	0.006

chart, BRISQ algorithm, BSQreg algorithm and InducedSmooth algorithm have better overall effects.In these algorithms, kink location is estimated by linearization technique, and the variance estimation obtained by these estimation algorithms is also small, which effectively ensures the robustness of the estimation.

For SmoothKernel algorithm and BentCableKink algorithm, although SmoothKernel algorithm uses kernel function to improve the smoothing of the overall model, BentCableKink algorithm uses quadratic function to approximate the location of the kinks estimation, and flexibly describes the sharp change of the relationship between variables at the position of the kink, the variance estimation of the parameters of these two algorithms is large, especially for the kink position estimation $\hat{\delta}_2$.For instance, at the quantile level of $\tau = 0.9$, the variance estimation reaches about 10. Obviously, it is not a robust estimation. However, the form of multi-kink quantile regression model given by the two algorithms is roughly consistent with the results presented by BRISQ algorithm, BSQreg algorithm and InducedSmooth algorithm. Another feature of these two algorithms is that they mitigate the huge change of slope at the kink. It can be seen that the slope change of the two algorithms at the first turning point (around year 1913) is not very obvious, which effectively mitigates the slope change and is a more conservative estimation algo-

Table 12: $K = 2$,Heteroscedasticity,$e_i \sim \mathbf{LA}(0, 1)$, $\tau = 0.5$

Method		α_0	α_1	γ	β_1	β_2	δ_1	δ_2
BRISQ	**Bias**	-0.004	0.001	-0.001	-0.004	0.008	0.003	-0.001
	SD	0.070	0.014	0.015	0.086	0.143	0.046	0.066
	SE	0.073	0.013	0.016	0.083	0.152	0.045	0.064
	MSE	0.005	0.000	0.000	0.007	0.023	0.002	0.004
SmoothKernel	**Bias**	-0.006	-0.002	-0.001	0.136	-0.248	-0.064	0.008
	SD	0.071	0.015	0.014	0.080	0.149	0.048	0.059
	SE	0.149	0.032	0.156	0.294	0.027	0.256	0.173
	MSE	0.022	0.001	0.024	0.105	0.062	0.070	0.030
BentCableKink	**Bias**	0.024	0.005	-0.000	0.080	-0.165	-0.048	0.007
	SD	0.069	0.014	0.014	0.087	0.159	0.050	0.060
	SE	0.151	0.032	0.173	0.307	0.027	0.291	0.196
	MSE	0.023	0.001	0.030	0.101	0.028	0.087	0.038
BSQreg	**Bias**	-0.054	-0.013	0.001	0.050	-0.068	-0.006	-0.001
	SD	0.110	0.026	0.021	0.108	0.160	0.073	0.105
	SE	0.147	0.032	0.165	0.302	0.027	0.263	0.173
	MSE	0.025	0.001	0.027	0.094	0.005	0.069	0.030
InducedSmooth	**Bias**	-0.001	-0.000	-0.002	0.026	-0.042	0.004	-0.000
	SD	0.070	0.015	0.017	0.096	0.145	0.049	0.058
	SE	0.073	0.022	0.015	0.102	0.135	0.052	0.063
	MSE	0.005	0.000	0.000	0.011	0.020	0.003	0.004

rithm. However, the depiction of the overall trend of this dataset by these algorithms is in line with the law of historical development. In fact, different algorithms have different applicable datasets. However, due to the length of the article, other datasets will not be analyzed. In real data analysis, specific problems need to be analyzed for different datasets, and appropriate algorithms should be selected for regression.

5 Conclusion

In this paper, we mainly improve the algorithm of the multi-kink quantile regression model proposed by Zhong et al[19]. We put the emphasis on the smoothing part, smoothed the model form itself, the loss function of the model and the iterative calculation of estimation, and gave several improved algorithms, which were proved to be reasonable by a large number of simulation experiments and real data applications. A R package "MKQRimprove" has been developed for all improved algorithms mentioned above. In the future studies, we can jump out of the category of "smoothing" and combine the framework of machine learning and statistical computing to find more algorithms of this

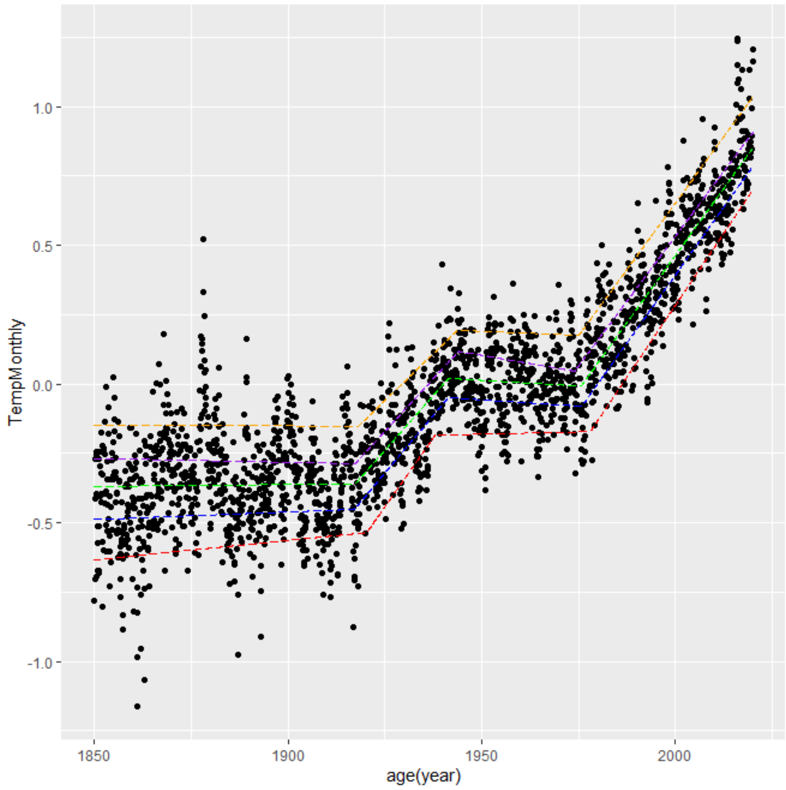

Figure 1: 1850-2020 monthly temperature extremum multi-kink quantile regression (BRISQ)

Table 13: AirTemp dataset estimation result comparison

	Method	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$
	BRISQ	$-3.222_{(1.020)}$	$0.001_{(0.001)}$	$0.018_{(0.004)}$	$-0.020_{(0.004)}$	$0.020_{(0.001)}$	$1920.304_{(3.224)}$	$1937.942_{(1.977)}$	$1978.015_{(1.573)}$
	SmoothKernel	$-6.615_{(3.899)}$	$0.003_{(0.002)}$	$0.004_{(0.006)}$	$-0.007_{(0.007)}$	$0.020_{(0.005)}$	$1915.722_{(9.664)}$	$1943.170_{(8.977)}$	$1975.998_{(5.962)}$
$\tau = 0.1$	BentCableKink	$-6.615_{(3.897)}$	$0.003_{(0.002)}$	$0.004_{(0.006)}$	$-0.007_{(0.007)}$	$0.020_{(0.005)}$	$1914.875_{(9.214)}$	$1942.476_{(8.572)}$	$1976.411_{(4.724)}$
	BSQreg	$-2.588_{(3.672)}$	$0.001_{(0.002)}$	$0.015_{(0.006)}$	$-0.017_{(0.007)}$	$0.021_{(0.003)}$	$1916.901_{(4.343)}$	$1942.370_{(6.001)}$	$1976.363_{(5.779)}$
	InducedSmooth	$-2.705_{(1.226)}$	$0.001_{(0.001)}$	$0.013_{(0.002)}$	$-0.015_{(0.002)}$	$0.021_{(0.002)}$	$1916.901_{(2.705)}$	$1942.370_{(2.439)}$	$1976.363_{(2.039)}$
	BRISQ	$-1.493_{(0.865)}$	$0.001_{(0.000)}$	$0.016_{(0.001)}$	$-0.018_{(0.002)}$	$0.020_{(0.001)}$	$1917.074_{(1.635)}$	$1941.297_{(1.347)}$	$1975.773_{(1.051)}$
	SmoothKernel	$-3.478_{(2.117)}$	$0.002_{(0.002)}$	$0.006_{(0.003)}$	$-0.006_{(0.002)}$	$0.020_{(0.001)}$	$1913.122_{(9.113)}$	$1941.519_{(8.312)}$	$1976.002_{(4.342)}$
$\tau = 0.3$	BentCableKink	$-3.478_{(2.116)}$	$0.002_{(0.001)}$	$0.005_{(0.003)}$	$-0.006_{(0.003)}$	$0.020_{(0.002)}$	$1913.263_{(9.323)}$	$1941.917_{(8.392)}$	$1975.631_{(4.588)}$
	BSQreg	$-0.616_{(1.897)}$	$-0.000_{(0.001)}$	$0.014_{(0.002)}$	$-0.014_{(0.002)}$	$0.020_{(0.002)}$	$1912.843_{(4.216)}$	$1941.892_{(3.192)}$	$1975.783_{(2.291)}$
	InducedSmooth	$-1.326_{(1.012)}$	$0.000_{(0.001)}$	$0.015_{(0.001)}$	$-0.016_{(0.002)}$	$0.020_{(0.001)}$	$1912.843_{(1.639)}$	$1941.892_{(1.218)}$	$1975.783_{(1.044)}$
	BRISQ	$-0.662_{(0.732)}$	$0.000_{(0.000)}$	$0.016_{(0.001)}$	$-0.017_{(0.002)}$	$0.020_{(0.001)}$	$1917.350_{(1.308)}$	$1941.320_{(1.866)}$	$1975.618_{(1.160)}$
	SmoothKernel	$-2.317_{(1.691)}$	$0.001_{(0.001)}$	$0.005_{(0.003)}$	$-0.005_{(0.003)}$	$0.020_{(0.002)}$	$1913.516_{(8.837)}$	$1943.189_{(7.910)}$	$1975.716_{(3.370)}$
$\tau = 0.5$	BentCableKink	$-2.272_{(1.786)}$	$0.001_{(0.001)}$	$0.005_{(0.003)}$	$-0.005_{(0.003)}$	$0.020_{(0.004)}$	$1913.311_{(8.430)}$	$1943.499_{(8.407)}$	$1975.429_{(3.378)}$
	BSQreg	$-0.086_{(1.452)}$	$-0.000_{(0.001)}$	$0.014_{(0.003)}$	$-0.014_{(0.002)}$	$0.020_{(0.003)}$	$1913.396_{(4.280)}$	$1943.085_{(3.959)}$	$1975.664_{(2.340)}$
	InducedSmooth	$0.061_{(0.818)}$	$-0.000_{(0.000)}$	$0.013_{(0.001)}$	$-0.014_{(0.002)}$	$0.020_{(0.001)}$	$1913.396_{(2.047)}$	$1943.085_{(2.197)}$	$1975.664_{(1.197)}$
	BRISQ	$0.252_{(0.782)}$	$-0.000_{(0.000)}$	$0.015_{(0.001)}$	$-0.018_{(0.001)}$	$0.021_{(0.001)}$	$1917.293_{(1.600)}$	$1943.914_{(1.117)}$	$1973.969_{(1.238)}$
	SmoothKernel	$-2.652_{(1.762)}$	$0.001_{(0.001)}$	$0.004_{(0.002)}$	$-0.004_{(0.003)}$	$0.019_{(0.003)}$	$1914.132_{(8.567)}$	$1944.576_{(8.374)}$	$1973.819_{(3.007)}$
$\tau = 0.7$	BentCableKink	$-2.694_{(2.614)}$	$0.001_{(0.001)}$	$0.004_{(0.005)}$	$-0.005_{(0.003)}$	$0.021_{(0.004)}$	$1913.791_{(8.659)}$	$1944.324_{(8.254)}$	$1974.010_{(3.168)}$
	BSQreg	$0.379_{(1.673)}$	$-0.000_{(0.001)}$	$0.013_{(0.002)}$	$-0.015_{(0.003)}$	$0.020_{(0.003)}$	$1913.641_{(3.616)}$	$1944.138_{(3.104)}$	$1974.111_{(2.807)}$
	InducedSmooth	$0.660_{(0.881)}$	$-0.001_{(0.000)}$	$0.013_{(0.001)}$	$-0.014_{(0.002)}$	$0.020_{(0.002)}$	$1913.641_{(2.293)}$	$1944.138_{(2.391)}$	$1974.111_{(1.462)}$
	BRISQ	$-0.098_{(1.210)}$	$-0.000_{(0.001)}$	$0.013_{(0.002)}$	$-0.013_{(0.002)}$	$0.020_{(0.001)}$	$1917.796_{(2.904)}$	$1943.898_{(1.284)}$	$1975.057_{(1.557)}$
	SmoothKernel	$-4.844_{(4.527)}$	$0.003_{(0.002)}$	$0.002_{(0.005)}$	$-0.003_{(0.005)}$	$0.018_{(0.005)}$	$1913.762_{(12.691)}$	$1943.957_{(4.749)}$	$1975.003_{(6.103)}$
$\tau = 0.9$	BentCableKink	$-4.932_{(4.529)}$	$0.004_{(0.002)}$	$0.002_{(0.005)}$	$-0.003_{(0.006)}$	$0.017_{(0.005)}$	$1913.189_{(11.156)}$	$1943.119_{(3.792)}$	$1976.087_{(6.009)}$
	BSQreg	$1.236_{(3.684)}$	$-0.001_{(0.002)}$	$0.013_{(0.005)}$	$-0.013_{(0.005)}$	$0.020_{(0.002)}$	$1913.350_{(9.416)}$	$1943.948_{(5.721)}$	$1975.183_{(5.352)}$
	InducedSmooth	$0.532_{(1.450)}$	$-0.000_{(0.001)}$	$0.0132_{(0.002)}$	$-0.012_{(0.002)}$	$0.019_{(0.002)}$	$1913.350_{(3.277)}$	$1943.948_{(1.032)}$	$1975.183_{(1.697)}$

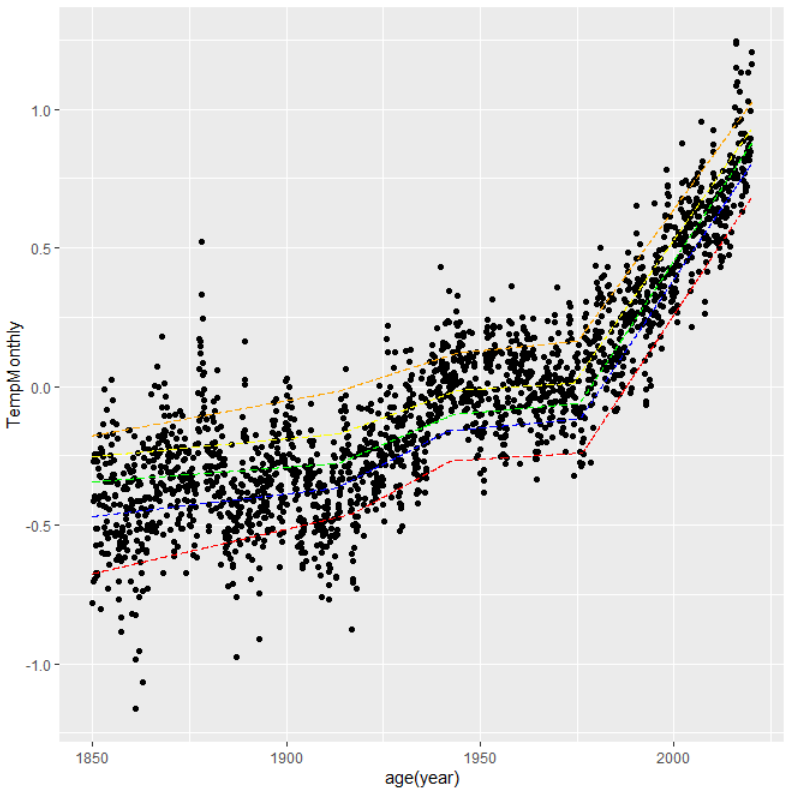

Figure 2: 1850-2020 monthly temperature extremum multi-kink quantile regression (SmoothKernel)

issue that can accurately estimate parameters and have higher efficiency.

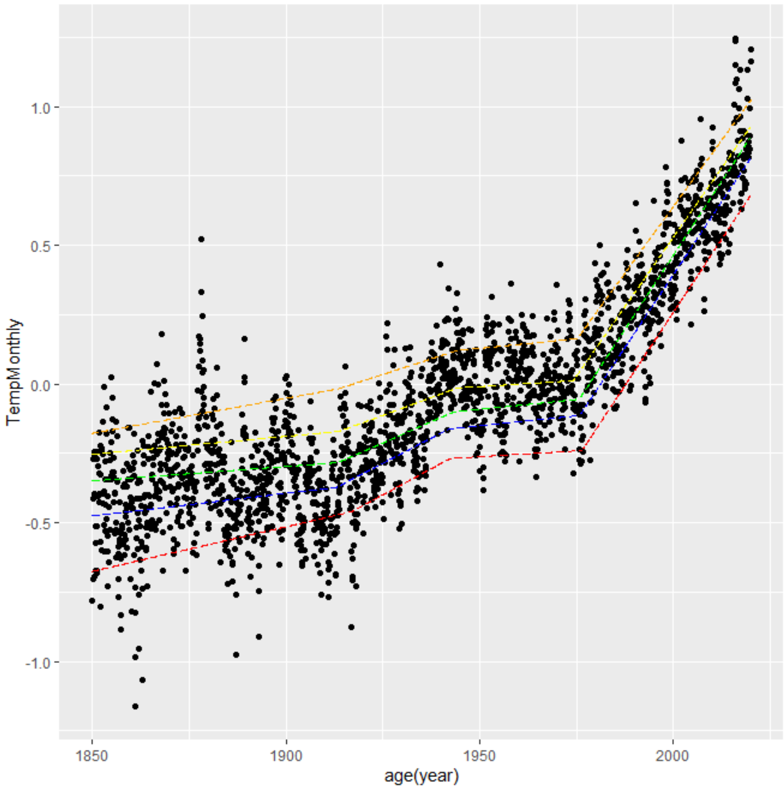

Figure 3: 1850-2020 monthly temperature extremum multi-kink quantile regression (Bent-CableKink)

References

[1] E S Page. Continuous inspection schemes. *Biometrika*, 41(1):100–115, 1954. 1

[2] Richard E Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53(284):873–880, 1958. 1

[3] Richard E Quandt. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55(290):324–330, 1960. 1

[4] P Sprent. Some hypotheses concerning two phase regression lines. *Biometrics*, 17(4):634–645, 1961. 1

[5] David V Hinkley. Inference about the intersection in two-phase regression. *Biometrika*, 56(3):495–504, 1969. 1

[6] David V Hinkley. Inference in two-phase regression. *Journal of the American Statistical Association*, 66(336):736–743, 1971. 1

[7] Paul I Feder. On asymptotic distribution theory in segmented regression problems–identified case. *The Annals of Statistics*, 3(1):49–83, 1975. 1

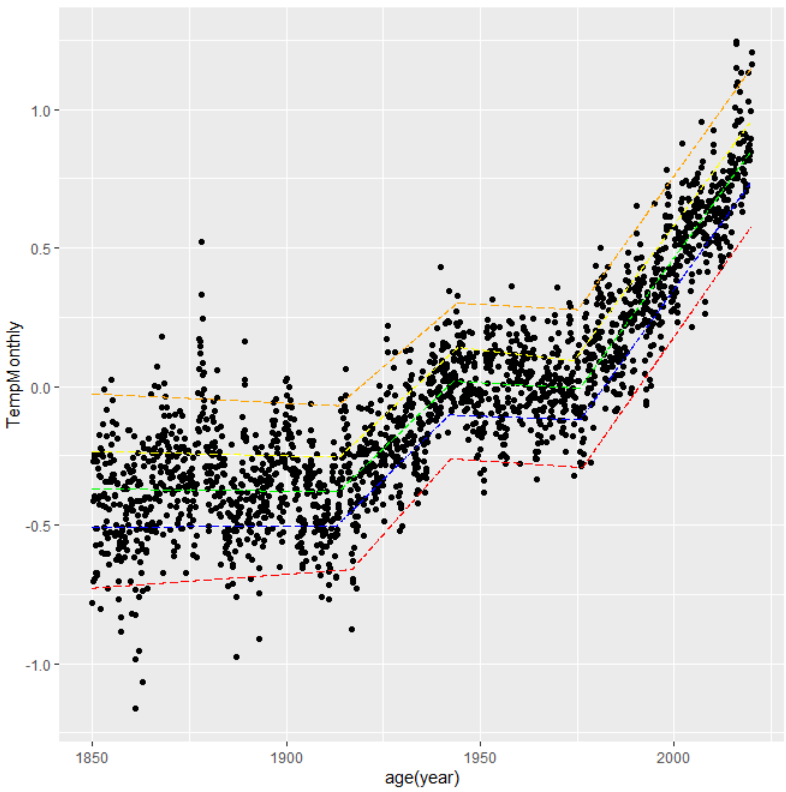

Figure 4: 1850-2020 monthly temperature extremum multi-kink quantile regression (BSQreg)

[8] Rick Chappell. Fitting bent lines to data, with applications to allometry. *Journal of Theoretical Biology*, 138(2):235–256, 1989. 1

[9] Howell Tong. *Non-linear time series: a dynamical system approach*. Oxford university press, 1990. 1

[10] Kung-Sig Chan and Ruey S Tsay. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2):413–426, 1998. 1

[11] Bruce E Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the Econometric Society*, 64(2):413–430, 1996. 1

[12] Sokbae Lee, Myung Hwan Seo, and Youngki Shin. Testing for threshold effects in regression models. *Journal of the American Statistical Association*, 106(493):220–231, 2011. 1

[13] Grace Chiu, Richard Lockhart, and Richard Routledge. Asymptotic theory for bent-cable regression—the basic case. *Journal of Statistical Planning and Inference*, 127(1-2):143–156, 2005. 1, 2.2

[14] Grace Chiu, Richard Lockhart, and Richard Routledge. Bent-cable regression theory and applications. *Journal of the American Statistical Association*, 101(474):542–553, 2006. 1

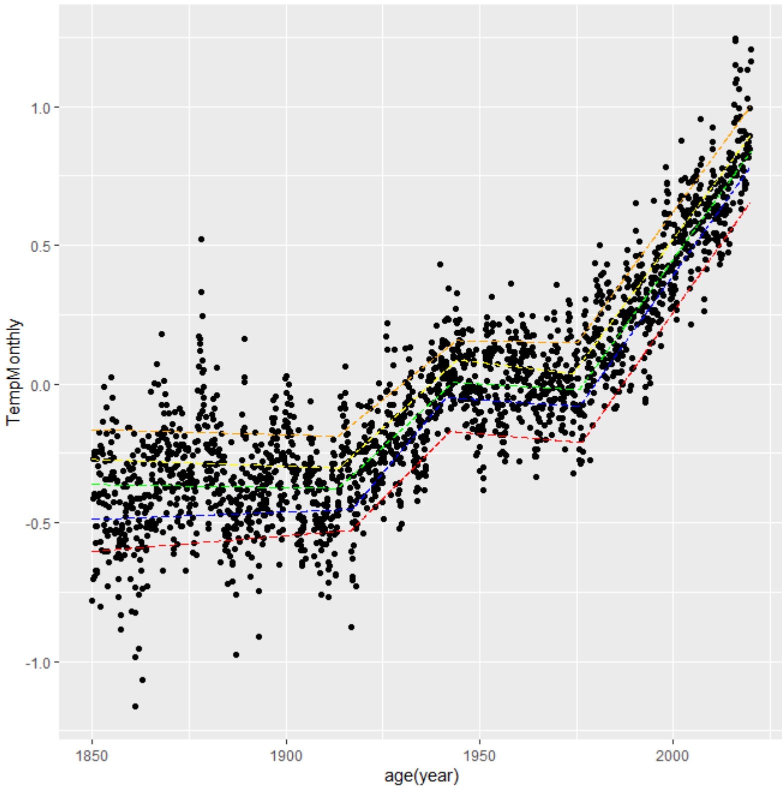

Figure 5: 1850-2020 monthly temperature extremum multi-kink quantile regression (InducedSmooth)

[15] Ritabrata Das, Moulinath Banerjee, Bin Nan, et al. Fast estimation of regression parameters in a broken-stick model for longitudinal data. *Journal of the American Statistical Association*, 111(515):1132–1143, 2016. 1, 2.2

[16] Bruce E Hansen. Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, 35(2):228–240, 2017. 1

[17] Chenxi Li, Ying Wei, Xuming He, et al. Bent line quantile regression with application to an allometric study of land mammals' speed and mass. *Biometrics*, 67(1):242–249, 2011. 1

[18] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1):33–50, 1978. 1

[19] Wei Zhong, Chuang Wan, and Wenyang Zhang. Estimation and inference for multi-kink quantile regression. *Journal of Business & Economic Statistics*, 0(0):1–17, 2021. 1, 1, 5

[20] Vito MR Muggeo. Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19):3055–3071, 2003. 1, 2.1

[21] Bradley Efron. *The jackknife, the bootstrap and other resampling plans*. SIAM, 1982. 1

[22] Mark A Aizerman, Emmanuil M Braverman, and Lev I Rozonoer. Theoretical foundation of potential functions method in pattern recognition. *Avtomat. I Telemekh*, 25(6):917–936, 1964. 2.1

[23] Xiaoying Zhou and Feipeng Zhang. Bent line quantile regression via a smoothing technique. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 13(3):216–228, 2020. 2.1

[24] Wallace Hendricks and Roger Koenker. Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 87(417):58–68, 1992. 2.1

[25] Songfeng Zheng. Gradient descent algorithms for quantile regression with smooth approximation. *International Journal of Machine Learning and Cybernetics*, 2(3):191–207, 2011. 2.3, 2.3, 2.3

[26] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001. 2.3

[27] Llew Mason, Jonathan Baxter, Peter Bartlett, et al. Boosting algorithms as gradient descent. *Advances in Neural Information Processing Systems*, 12:512–518, 1999. 2.3

[28] Bruce Maxwell Brown and You-Gan Wang. Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, 92(1):149–158, 2005. 2.4

[29] Liya Fu and You-Gan Wang. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8):2526–2538, 2012. 2.4

[30] Lei Pang, Wenbin Lu, and Huixia Judy Wang. Variance estimation in censored quantile regression via induced smoothing. *Computational Statistics & Data Analysis*, 56(4):785–796, 2012. 2.4