# Efficient Multiple Change Point Detection and Localization For High-dimensional Quantile Regression with Heteroscedasticity

Xianru Wang[*], Bin Liu[*], Xinsheng Zhang[*], Yufeng Liu[†]

**Abstract**

Data heterogeneity is a challenging issue for modern statistical data analysis. There are different types of data heterogeneity in practice. In this paper, we consider potential structural changes and complicated tail distributions. There are various existing methods proposed to handle either structural changes or heteroscedasticity. However, it is difficult to handle them simultaneously. To overcome this limitation, we consider statistically and computationally efficient change point detection and localization in high-dimensional quantile regression models. Our proposed framework is general and flexible since the change points and the underlying regression coefficients are allowed to vary across different quantile levels. The model parameters, including the data dimension, the number of change points, and the signal jump size, can be scaled with the sample size. Under this framework, we construct a novel two-step estimation of the number and locations of the change points as well as the underlying regression coefficients. Without any moment constraints on the error term, we

[*]Department of Statistics and Data Science, School of Management at Fudan University, China; E-mail:18702192581@163.com; liubin0145@gmail.com; xszhang@fudan.edu.cn

[†]Department of Statistics and Operations Research, Department of Genetics, and Department of Biostatistics, Carolina Center for Genome Sciences, Linberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, U.S.A; E-mail: yfliu@email.unc.edu

present theoretical results, including consistency of the change point number, oracle estimation of change point locations, and estimation for the underlying regression coefficients with the optimal convergence rate. Finally, we present simulation results and an application to the S&P 100 dataset to demonstrate the advantage of the proposed method.

**Keywords:** Efficient computation, Heavy tail, High dimensions, Structure change

# 1 Introduction

Large-scaled data are ubiquitous in modern complex data analysis. Due to the complex generating mechanism of high-dimensional data, data heterogeneity often arises. One important type of heterogeneity is known as structural breaks, where the model of data generation changes, and the independent identically distributed assumption does not hold. Change point detection and localization is a powerful way to handle structural changes since the seminal work by Page (1955). It has received considerable attention in financial econometrics (Bai and Perron, 1998; Kim et al., 2005), genetics (Castro et al., 2018; Dette et al., 2022), and internet security monitoring (Peng et al., 2004). Another type of data heterogeneity is the complicated tail dependency among variables such as heteroscedasticity as well as serious heavy tails, where the variances of the random error terms in the models are different or even do not exist. In this case, heteroscedasticity increases the uncertainty of estimated parameters and decreases statistical efficiency. To capture the heteroscedasticity, quantile regression (Koenker and Bassett, 1978) is a powerful technique for studying different tails of the conditional distribution of variables of interest.

Although the two issues mentioned above are very important, to our limited knowledge, there are few techniques for handling structural changes and heteroscedasticity simultaneously. In many applications, studying structural changes in conditional quantiles is important. In this paper, we consider a non-stationary variant of high-dimensional quantile

regression models. To be specific, let $Y \in \mathbb{R}$ be a real-valued response variable and $\boldsymbol{X}$ be a covariate vector in $\mathbb{R}^p$. We are interested in the following model:

$$Q_\eta(Y|\boldsymbol{X}) = \boldsymbol{X}^\top \boldsymbol{\beta}_\eta, \quad \eta \in \Pi \subset (0,1), \tag{1.1}$$

where $\boldsymbol{\beta}_\eta = (\beta_{\eta,1}, ..., \beta_{\eta,p})^\top \in \mathbb{R}^p$ is the unknown vector of coefficients and $Q_\eta(Y|\boldsymbol{X})$ is the $\eta$-th quantile of $Y$ conditional on $\boldsymbol{X}$, i.e. $\mathbb{P}(Y \leq Q_\eta(Y|\boldsymbol{X})|\boldsymbol{X}) = \eta$. Model (1.1) can be equivalently formulated as $Y = \boldsymbol{X}^\top \boldsymbol{\beta}_\eta + \epsilon$, where $\epsilon$ is the error term, s.t. $\mathbb{P}(\epsilon \leq 0|\boldsymbol{X}) = \eta$ for some user-specified quantile level $\eta \in \Pi$, and $\Pi \subset (0,1)$ is a compact set of quantile indices. Note that, in our considered Model (1.1), the regression coefficients $\boldsymbol{\beta}_\eta$ can vary with different quantile indices. To simplify the notations, we abbreviate $\boldsymbol{\beta}_\eta$ to $\boldsymbol{\beta}$. Suppose we have a piecewise stationary time series $\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$ for $i = 1, \ldots, n$, where $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip})^\top$, and $\boldsymbol{\beta}^{(i)} = (\beta_1^{(i)}, \ldots, \beta_p^{(i)})^\top$ is the unknown regression coefficient vector for the $i$-th observation. Then we consider estimating multiple change points with piecewise constant coefficients for model (1.1) as follows:

$$Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(i)} + \epsilon_i. \tag{1.2}$$

Let $\widetilde{k} \geq 0$ be the true number of unknown change points along with the location vector $\widetilde{\boldsymbol{\tau}} = (\widetilde{\tau}_0, \widetilde{\tau}_1 ..., \widetilde{\tau}_{\widetilde{k}}, \widetilde{\tau}_{\widetilde{k}+1})^\top$ with $0 = \widetilde{\tau}_0 < \widetilde{\tau}_1 < \widetilde{\tau}_2 < \ldots < \widetilde{\tau}_{\widetilde{k}} < \widetilde{\tau}_{\widetilde{k}+1} = 1$. Then, the unknown $\widetilde{k}$ change points divide the $n$ time-ordered observations into $\widetilde{k}+1$ intervals and the underlying regression coefficients $\boldsymbol{\beta}^{(i)}$ have the following form: $\boldsymbol{\beta}^{(i)} = \boldsymbol{\beta}^0(j)$, if $\widetilde{\tau}_{j-1} < i/n \leqslant \widetilde{\tau}_j, j = 1, \ldots, \widetilde{k}+1$, where $\boldsymbol{\beta}^0(j) = (\beta_1^0(j), ..., \beta_p^0(j))^\top \in \mathbb{R}^p$ denotes the underlying true regression coefficients in the $j$-th interval. Note that we consider a high-dimensional framework with the diverging change point number in the sense that $p$ and $\widetilde{k}$ are allowed to scale with the sample size $n$. Based on this, our main goal is to develop a complete framework of change point detection for high-dimensional quantile regression models in (1.2), including estimating the number of change points ($\widetilde{k}$), the locations of change points ($\widetilde{\boldsymbol{\tau}}$) and the regression coefficients $\boldsymbol{\beta}^0(j)$ in each segmentation with good statistical convergence rates, as $p, n \to \infty$ simultaneously.

Change point detection for high-dimensional regression models is an important yet

challenging problem. The main difficulty comes from the complexity of both computation and theoretical analysis arising from the growing dimensions. Most existing studies focus on structural changes in mean regression models. Some important works for both low-dimensional and high-dimensional setups can be found in Zhang et al. (2015); Leonardi and Bühlmann (2016); Lee et al. (2016); Wang et al. (2021), among others. However, because of the curse of dimensionality, the complicated dependency among variables such as heteroscedasticity and heavy-tailed error distributions increases the uncertainty of estimated parameters and decreases the effectiveness of signal detection. In these cases, the above-mentioned mean regression based methods fail to maintain robustness to non-normality or even heteroscedasticity. As a result, they may suffer from serious overestimation of the change point number and perform poorly for change point localization (see Section 4 of the numerical performance). As a valuable alternative to the mean regression, conditional quantiles of the variable of interest can be estimated by quantile regression in a more comprehensive way. Quantile regression has distinct advantages over mean regression. It allows the error term to be heavy-tailed and does not require the existence of finite moments of the noise distribution. It can also handle heterogeneity due to either heteroscedastic variance or non-location-scale covariate effects. This motivates us to focus on the structural changes for Model (1.2), which can form a possibly infinite collection of quantile regressions including median regression as a special case.

The high-dimensional quantile regression model in (1.1) was previously considered under the homogenous sparsity framework in Belloni and Chernozhukov (2011); Fan et al. (2014); Chen et al. (2019); Belloni et al. (2019), among others. Existing results include model selection and rates of convergence based on the $\ell_1$-penalty. The assumption of homogeneity may not hold in practice. Different from the homogenous models, detection and localization of structural changes in quantile regression are very useful and can be widely applied in finance, genetics, and medical imaging. For example, in financial markets, a vast

amount of data with heavy-tailed error distributions or heteroscedasticity are available. The structures of variables may have sudden changes due to the financial crisis or policy. In the literature, a few papers considered structural changes in quantile regression models under low-dimensional setups. Qu (2008); Oka and Qu (2011) proposed procedures for testing and estimating multiple changes, respectively. Lee et al. (2011) proposed a sup-likelihood-ratio-based method for testing the existence of threshold effects in the regression model including quantile regression. Zhang et al. (2014) developed a CUSUM-based procedure for testing change points due to a covariate threshold in regression quantiles. Ciuperca and Maciak (2020) proposed an approach for change point detection and model estimation based on the quantile-fused penalty. Under high-dimensional setups, much less work has been done. Assuming that at most one change point occurred, Lee et al. (2018) considered a three-step procedure for detecting and localizing a change point in high-dimensional quantile regression models. In practice, we typically do not have prior knowledge about the number of change points, which limits the applicability of the model with at most one change point. It is possible to have no change point or multiple change points. Furthermore, the estimation of change points may be computationally challenging. Thus, it is desirable to have a computationally efficient method. Combining all considerations, we would like to develop a new approach satisfying the following four properties: (1) adaptive to the change point number including $\widetilde{k} = 0$ and $\widetilde{k} \gg 0$; (2) robust to the heavy-tailed error distributions or heteroscedasticity; (3) statistically effective under high-dimensional setups; (4) computationally efficient for large-scale data.

Our main contributions can be summarized as follows:

First, we consider a flexible and general framework in high-dimensional quantile regression in Model (1.2). The change points and the underlying coefficients are allowed to vary across different quantiles $\eta$. As a high-dimensional framework with $p \gg n$, we allow the change point number to be dense or even scale with the sample size $n$. For this flexible

model, we provide consistent two-step estimators $(\widehat{\boldsymbol{\tau}}_\eta, \widehat{\boldsymbol{\beta}}_\eta)$ of the number of change points and locations as well as the underlying regression coefficients. To simplify the notations, we abbreviate $(\widehat{\boldsymbol{\tau}}_\eta, \widehat{\boldsymbol{\beta}}_\eta)$ to $(\widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{\beta}})$ in the later contexts. As for the two-step estimators, in the first step, we obtain the initial estimators of change points and the underlying regression coefficients by solving a dynamic programming problem (Auger and Lawrence, 1989; Jackson et al., 2005; Killick et al., 2012; Rigaill, 2015). In the second step, we refine the initial estimators by a new refitting procedure to obtain improved estimators of both change points and the underlying regression coefficients with good convergence guarantee.

Second, to implement this two-step procedure, we develop two types of two-step algorithms, refined dynamic programming (RDP) and refined binary segmentation (RBS), in Algorithms 1 and 2, respectively. The refined process is based on a simplified dynamic programming. Since the refined process needs not to solve any Lasso problem (Tibshirani, 1996), it is computationally negligible as compared to the initial step. Hence, it greatly improves the estimation accuracy without increasing the computational burden. As a result, the computational complexities of RDP and RBS are $O(n^2\text{Lasso}(n,p))$ and $O(n\log(n)\text{Lasso}(n,p))$, respectively, where $\text{Lasso}(n,p)$ is the cost to compute the Lasso estimator for quantile regression models.

Third, we present theoretical results in terms of change point detection and localization as well as estimation of the underlying regression coefficients. For assumptions, we do not need any moment constraints on the error term, which is more flexible compared to the typical sub-Gaussian or normal distribution assumptions. Hence, we allow heavy-tailed error distributions as well as heteroscedasticity among errors. For the theoretical results, we prove consistency of the estimators for the change point number and locations as well as the underlying regression coefficients under weak signal conditions. Furthermore, we establish that the refined estimators of change point locations computed by RDP and RBS enjoy the oracle property and the refined estimator of the underlying regression coefficients

achieves the optimal convergence rate.

Lastly, we develop a cross-validation approach to choose the proper values of the multiple tuning parameters in a fully data-driven fashion in the Supplementary Materials. We implement extensive numerical results under various model settings and demonstrate robustness to nonnormality and heteroscedasticity of our proposed approaches. We also present interesting results of change point detection for the S&P 100 dataset.

The rest of this paper is organized as follows. In Section 2, we introduce our methodology for detecting change points. In Section 3, the corresponding theoretical results of the change points computed by RDP and RBS are established. We investigate the performance of our proposed methods by extensive numerical results as well as real data applications in Sections 4 and 5. We summarize the paper in Section 6. Proofs and additional numerical results are given in the Supplementary Materials.

# 2 Methodology

We describe our methodology for Model (1.2). In Section 2.1.1, we propose a two-step change point estimator including the number and locations of change points. Meanwhile, the regression coefficients in each segment are estimated based on the Lasso. In Sections 2.1.2 and 2.1.3, based on the dynamic programming and binary segmentation techniques, two algorithms are respectively proposed to detect multiple change points.

We first introduce some notations. For a vector $\mathbf{a} = (a_1, \ldots, a_p)^\top \in \mathbb{R}^p$, we denote $\|\boldsymbol{a}\|_1 = \sum_{i=1}^p |a_i|$, $\|\boldsymbol{a}\|_2 = (\sum_{i=1}^p a_i^2)^{1/2}$, and $\|\boldsymbol{a}\|_\infty = \max_{1\le i\le p} |a_i|$. For two real numbered sequences $a_n$ and $b_n$, we set $a_n = O(b_n)$ if there exits a constant $C$ such that $|a_n| \le C|b_n|$ for a sufficiently large $n$; $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$. For a sequence of random variables (r.v.s) $\{\xi_1, \xi_2, \ldots\}$, we set $\xi_n \xrightarrow{\mathbb{P}} \xi$ if $\xi_n$ converges to $\xi$ in probability as $n \to \infty$. We also denote $\xi_n = o_p(1)$ if $\xi_n \xrightarrow{\mathbb{P}} 0$. For a set $A$, we use $\#A$ to denote its cardinality. We use $C_1, C_2, \ldots$ to denote positive constants that may vary from line to line.

## 2.1 New Estimation and Algorithms

Consider a non-stationary variant of the high-dimensional quantile regression in Model (1.2) where the data are observed as a time series $\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$. For any given partion $\boldsymbol{\tau} = (\tau_0, \tau_1, ..., \tau_k, \tau_{k+1})^\top$ with $0 = \tau_0 < \tau_1 < ... < \tau_k < \tau_{k+1} = 1$, we denote the $j$-th interval by $I_j(\boldsymbol{\tau}) = (\tau_{j-1}, \tau_j)$, the length of the $j$-th interval by $r_j(\boldsymbol{\tau}) = \tau_j - \tau_{j-1}$, the shortest interval length by $r(\boldsymbol{\tau}) = \min_{1\le j\le k+1} r_j(\boldsymbol{\tau})$, the change point number by $l(\boldsymbol{\tau})$. Given $\boldsymbol{\tau}$, let $\boldsymbol{\beta}(\boldsymbol{\tau}) := (\boldsymbol{\beta}^\top(\boldsymbol{\tau}, 1), \ldots, \boldsymbol{\beta}^\top(\boldsymbol{\tau}, l(\boldsymbol{\tau})+1))^\top$ be the potential regression coefficients in each segment of $\boldsymbol{\tau}$.

### 2.1.1 A Two-step Refined Estimator

Next we introduce our estimation methods for Model (1.2). For quantile regression models, we take the check loss approach of Koenker and Bassett (1978) as follows:

$$\rho\big(Y_i, \boldsymbol{X}_i^\top\boldsymbol{\beta}\big) \equiv \big(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}\big)\big(\eta - 1\big\{Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta} \le 0\big\}\big), \tag{2.1}$$

where $\eta$ is given, s.t. $\eta \in \Pi$, and $\Pi \subset (0,1)$ is a compact set of quantile indices. Consider the heterogeneous data with multiple change points $\widetilde{\boldsymbol{\tau}}$. For any given candidate change point partition $\boldsymbol{\tau} = (\tau_0, \tau_1, ..., \tau_k, \tau_{k+1})^\top$, we introduce our estimator in each segment as follows:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}, j) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \left\{ P_n\rho\left(I_j(\boldsymbol{\tau}), \boldsymbol{\beta}\right) + \lambda_n\sqrt{\tau_j - \tau_{j-1}}\big\|\boldsymbol{\beta}\big\|_1 \right\}, \text{ for } j = 1, \ldots, l(\boldsymbol{\tau}) + 1, \tag{2.2}$$

where $P_n\rho\left(I_j(\boldsymbol{\tau}), \boldsymbol{\beta}\right) := \frac{1}{n}\sum_{i=n\tau_{j-1}+1}^{n\tau_j} \rho(Y_i, \boldsymbol{X}_i^\top\boldsymbol{\beta})$ and $\lambda_n$ is the non-negative regularization parameter for the sub-interval based Lasso estimation, which is discussed in Section 3.1. The main challenge of multiple change point models is that the change point number $\widetilde{k}$, locations $\widetilde{\boldsymbol{\tau}}$, and the underlying regression coefficients $\boldsymbol{\beta}^0$ are unknown. Meanwhile, the estimator of one parameter affects the estimation efficiency of the other parameter. More specifically, the estimation efficiency for the regression coefficients may be affected by the change point estimation, and vice versa. To solve this issue, we develop a two-step refined estimation method for change points and underlying regression coefficients. In the first

step, we obtain the initial estimators of change points and underlying regression coefficients simultaneously by solving a dynamic programming problem (Auger and Lawrence, 1989; Jackson et al., 2005; Killick et al., 2012; Rigaill, 2015; Maidstone et al., 2017). This is not a convex optimization problem due to the presence of change points, which makes the convergence rates of initial estimators not optimal. In the second step, based on the initial estimator of regression coefficients, we re-estimate change points to obtain the improved change point estimator. Then we re-estimate regression coefficients based on the improved change point estimator. Next, we present the detailed procedure for estimating change points and regression coefficients.

**Step 1 Obtain initial estimator $(\widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}))$**: For any candidate partition $\boldsymbol{\tau}$ with corresponding coefficients $\boldsymbol{\beta}(\boldsymbol{\tau})$, we first define the overall check loss as:

$$J_n(\boldsymbol{\tau}, \boldsymbol{\beta}(\boldsymbol{\tau})) := \sum_{j=1}^{l(\boldsymbol{\tau})+1} P_n \rho\big(I_j(\boldsymbol{\tau}), \boldsymbol{\beta}(\boldsymbol{\tau}, j)\big). \tag{2.3}$$

We find the "best" partition $\widehat{\boldsymbol{\tau}}(\widehat{k}) := (0, \widehat{\tau}_1, ..., \widehat{\tau}_{\widehat{k}}, 1)^\top$ which minimizes the following total penalized check loss function:

$$\widehat{\boldsymbol{\tau}}(\widehat{k}) = \underset{k \in \{0,1,...,k_{\max}\}}{\arg\min} \ \underset{\boldsymbol{\tau}=(\tau_0,\tau_1,...,\tau_k,\tau_{k+1})^\top}{\arg\min} \{J_n(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\tau})) + \gamma(k+1)\}, \tag{2.4}$$

where $\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}) := (\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}, 1)^\top, \ldots, \widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}, l(\boldsymbol{\tau})+1)^\top)^\top$ can be obtained by (2.2). To further explain (2.4), we break down the above optimization problem into three detailed steps as follows:

**Step 1-1 (Search for the "best" partition)**: Given a candidate number of change points $k$, we find the "best" partition $\widehat{\boldsymbol{\tau}}(k) = (\tau_0, \widehat{\tau}_1, ..., \widehat{\tau}_k, \tau_{k+1})^\top$ as follows:

$$\widehat{\boldsymbol{\tau}}(k) = \underset{\boldsymbol{\tau}=(\tau_0,...,\tau_{k+1})^\top}{\arg\min} J_n(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\tau})) + \gamma(k+1), \tag{2.5}$$

where the Lasso-based estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}) := (\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}, 1)^\top, \ldots, \widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}, k+1)^\top)^\top$ is obtained in (2.2).

**Step 1-2 (Estimate the number of change points)**: We put $\widehat{\boldsymbol{\tau}}(k)$ into $J_n(\boldsymbol{\tau}, \boldsymbol{\beta}(\boldsymbol{\tau}))$ and obtain the minimum loss function associated with $k$ as $G(k) := J_n(\widehat{\boldsymbol{\tau}}(k), \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}(k))) + \gamma(k+1)$. Then, we find the "best" estimator for the change point number as follows:

$$\widehat{k} = \underset{k \in \{0,1,...,k_{\max}\}}{\arg\min} G(k), \tag{2.6}$$

where $k_{\max}$ is an "upper bound" of the number of change points specified by users. In this paper, we allow $k_{\max}$ grow with the sample size $n$. See Section 3.1 for more details.

**Step 1-3 (Estimate locations of change points)**: We put $\widehat{k}$ into Step 1-1 and obtain the final change point estimator $\widehat{\boldsymbol{\tau}} := \widehat{\boldsymbol{\tau}}(\widehat{k}) = (\widehat{\tau}_1, ..., \widehat{\tau}_{\widehat{k}})^\top$ by

$$\widehat{\boldsymbol{\tau}} = \underset{\boldsymbol{\tau}=(\tau_0,\ldots,\tau_{\widehat{k}+1})^\top}{\arg\min} J_n(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\tau})). \tag{2.7}$$

Moreover, using $\widehat{\boldsymbol{\tau}}$, the initial estimated underlying regression coefficients $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}) := (\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, 1)^\top, \ldots, \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, \widehat{k}+1)^\top)^\top$ are obtained by

$$\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, j) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \Big\{ \frac{1}{n} \sum_{i=n\widehat{\tau}_{j-1}+1}^{n\widehat{\tau}_j} \rho\big(Y_i, \boldsymbol{X}_i^\top \boldsymbol{\beta}\big) + \lambda_n \sqrt{\widehat{\tau}_j - \widehat{\tau}_{j-1}} \big\|\boldsymbol{\beta}\big\|_1 \Big\}, \ j = 1, \ldots, \widehat{k}+1. \tag{2.8}$$

The main purpose of Step 1 is to obtain initial estimators of the number of change points $\widetilde{k}$ and underlying regression coefficients $\boldsymbol{\beta}^0(j)$, $j = 1, ..., \widetilde{k}+1$. As proved in Section 3.1, the estimator of the number is consistent. In addition, as we mentioned before, since the change points and underlying regression coefficients are estimated simultaneously, the estimation efficiency is reduced. Then the convergence rate of the initial estimator of regression coefficients and change points may not be optimal as compared to the oracle case to be introduced later. To further improve the convergence rates, we refit our initial estimators as follows.

**Step 2 Refine the initial estimators to obtain the final estimators**.

**Step 2-1 (Improve change point estimation)**: Using the estimated number of change points $\widehat{k}$ and the underlying regression coefficients $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}})$ obtained in Step 1, we obtain the improved change point estimator $\widehat{\boldsymbol{\tau}}^r := \widehat{\boldsymbol{\tau}}^r(\widehat{k})$ by:

$$\widehat{\boldsymbol{\tau}}^r(\widehat{k}) = \underset{\boldsymbol{\tau}=(\tau_0,\ldots,\tau_{\widehat{k}+1})^\top}{\arg\min} \frac{1}{n} \sum_{j=1}^{\widehat{k}+1} \sum_{i=n\tau_{j-1}+1}^{n\tau_j} \rho(Y_i, \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, j)), \tag{2.9}$$

where $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, j); j = 1, ..., \widehat{k}+1$, are obtained previously by (2.8).

**Remark 2.1.** The refined change point estimator in Step 2-1 is motivated by the oracle estimator $\ddot{\boldsymbol{\tau}}$. Consider an oracle scenario, where the true change point number $\widetilde{k}$ and underlying regression coefficients $\boldsymbol{\beta}^0 = (\boldsymbol{\beta}^0(1), ..., \boldsymbol{\beta}^0(\widetilde{k}+1))^\top$ are known. Then, we define

the oracle estimator $\ddot{\boldsymbol{\tau}}$ as follows:

$$\ddot{\boldsymbol{\tau}} = \underset{\boldsymbol{\tau}=(\tau_0,\tau_1,\ldots,\tau_{\widetilde{k}},\tau_{\widetilde{k}+1})^\top}{\arg\min} \frac{1}{n} \sum_{j=1}^{\widetilde{k}+1} \sum_{i=n\tau_{j-1}+1}^{n\tau_j} \rho(Y_i, \boldsymbol{X}_i^\top \boldsymbol{\beta}^0(j)). \tag{2.10}$$

Given the true change point number and the underlying regression coefficients, the oracle change point estimator resulting from (2.10) can be obtained by using simplified dynamic programming. The main benefit of Step 2-1 is that we use the estimator $\widehat{k}$ of the change point number and the initial Lasso estimator $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}})$ to respectively avoid screening for the change point number and the procedure for computing the Lasso.

**Step 2-2 (Update the regression coefficients)**: Based on the change point estimator $\widehat{\boldsymbol{\tau}}^r(\widehat{k})$ obtained in Step 2-1, we obtain the final refined regression coefficients by

$$\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \Big\{ \frac{1}{n} \sum_{i=n\widehat{\tau}_{j-1}^r+1}^{n\widehat{\tau}_j^r} \rho\left(Y_i, \boldsymbol{X}_i^\top \boldsymbol{\beta}\right) + \lambda_n^r \sqrt{\widehat{\tau}_j^r - \widehat{\tau}_{j-1}^r} \|\boldsymbol{\beta}\|_1 \Big\}, \ j = 1, \ldots, \widehat{k}+1, \tag{2.11}$$

where $\lambda_n^r$ is the non-negative regularization parameter for the refined step, which is discussed in Section 3.1. With the refitting procedure, we obtain our final estimators $(\widehat{\boldsymbol{\tau}}^r, \widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r))$ of change points and regression coefficients. Our next task is to develop computationally efficient algorithms in Sections 2.1.2 and 2.1.3 to obtain these estimators in (2.9) and (2.11).

### 2.1.2 The Refined Dynamic Programming Based Approach

We outline our proposed Refined Dynamic Programming (RDP) algorithm based on dynamic programming, which is a popular technique for multiple change point detection (Friedrich et al., 2008; Killick et al., 2012; Rigaill, 2015; Leonardi and Bühlmann, 2016; Maidstone et al., 2017). Note that, as a general technique, it needs to be modified to solve different specific problems. Our proposed estimation procedure includes two steps, which correspond to a two-step refined dynamic programming algorithm. In Step 1 of RDP, we modify the basic dynamic programming for solving the quantile check loss-based optimization problem, which is non-smooth and non-convex. In Step 2, we use simplified dynamic programming to refine our initial estimators obtained by Step 1 of RDP. The usefulness

of the refined process in Step 2-1 is immediate, as it does not solve any Lasso problems, resulting in negligible computational cost. More details on using dynamic programming to obtain the initial estimators are given in Section A of the Supplementary Materials .

We present RDP in Algorithm 1.

---

**Algorithm 1** : The RDP procedure for change point detection

---

**Input:** Given the data set $\{\mathbf{X}, \boldsymbol{Y}\}$ and the user-specified quantile indice $\eta$, set the value of $k_{\max}$.

**Step 1:** Use dynamic programming to obtain the initial estimators $\widehat{\boldsymbol{\tau}}(\widehat{k})$ and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}})$ as a solution to (2.4). The detailed dynamic programming to solve Steps 1-1 to 1-3 is provided in Section A of the Supplementary Materials.

**Refinement**

**Step 2:** Given the number $\widehat{k}$ of change points and the underlying regression coefficients $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}})$, obtain the improved change point estimator $\widehat{\boldsymbol{\tau}}^r(\widehat{k})$ by a simplified dynamic programming procedure (2.9). Update the estimator of the regression coefficients via Lasso $\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r)$ by (2.11).

**Output:** The algorithm provides the two-step refined change point estimator $\widehat{\boldsymbol{\tau}}^r(\widehat{k}) = (0, \widehat{\tau}_1^r, ..., \widehat{\tau}_{\widehat{k}}^r, 1)^\top$, and the updated estimator of regression coefficients $\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r)$.

---

The change point estimators obtained as a solution to RDP have good statistical accuracy. However, RDP has a computational complexity of $O(n^2\text{Lasso}(n,p))$ (Step 1) , where $\text{Lasso}(n,p)$ is the cost to compute the Lasso estimator for quantile regression models. It is desirable to propose a different approach as introduced in Section 2.1.3 to improve the computational efficiency.

### 2.1.3 The Refined Binary Segmentation Based Approach

To improve our RDP, we develop a much more efficient Refined Binary Segmentation (RBS) approach based on the binary segmentation technique (Vostrikova, 1981). Binary segmentation is a popular technique that can be modified for solving various problems about multiple change point detection. For example, Leonardi and Bühlmann (2016) proposed

a one-step algorithm based on binary segmentation for change point detection for high-dimensional regression. There are also some extensions of binary segmentation techniques. Fryzlewicz (2014) developed the wild binary segmentation algorithm. Kovács et al. (2020) proposed "optimistic search" strategies, and Kovács et al. (2023) recently proposed Seeded Binary Segmentation algorithm. Londschien et al. (2022) developed "optimistic search" based binary segmentation approaches for computationally efficient change point detection for graphical models in the presence of missing values. Different from the aforementioned literature, our two-step RBS is used to solve a non-smooth optimization problem and also includes a refitting process. In particular, in Step 1 of RBS, we consider a binary segmentation approach to obtain the initial estimator of change points. This step greatly improves the computational efficiency for change point detection compared to Step 1 of RDP. As for Step 2, we use a simplified dynamic programming step to refine the initial estimators obtained by Step 1 of RBS as input. This improves the statistical accuracy of estimators for both change points and regression coefficients. Note that the refinement procedure of RBS is the same as that of RDP, resulting in negligible computationally cost. Next we introduce RBS to implement our two-step estimation method.

We first consider Step 1 of RBS, and the key idea is local screening. Given the initial search interval $(u, v)$, we use the penalized loss function to determine whether a new change point $s$ can be added. If $s$ is identified, then the interval $(u, v)$ is split into two subintervals $(u, s)$ and $(s, v)$ and we conduct the above procedure on $(u, s)$ and $(s, v)$ separately. This algorithm is stopped until no new subintervals can be added. To be specific, for any given $u, v \in V_n := \{i/n : i = 1, \ldots, n\}$, we define

$$Z(u, v) = \begin{cases} P_n \rho\Big((u, v], \breve{\boldsymbol{\beta}}_{(u,v]}\Big) + \gamma, & \text{if } (v - u)n \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.12}$$

and

$$h(u, v) = \underset{s \in \{u\} \cup [u+\delta, v-\delta]}{\arg\min} \{Z(u, s) + Z(s, v)\}, \tag{2.13}$$

where $\breve{\boldsymbol{\beta}}_{(u,v)} = \widehat{\boldsymbol{\beta}}_{(u,v)}$ and $\delta = \min_{1\le j\le \widetilde{k}}(\widetilde{\tau}_{j+1} - \widetilde{\tau}_j)$ is the minimal segmentation length. Based on (2.12) and (2.13), we present the procedure for obtaining the initial estimators of change points and the underlying regression coefficients in Steps 1-1 and 1-2 of Algorithm 2. For Step 2 of RBS, we use a similar refined procedure as RDP with the initial estimators obtained by RBS as input.

We present the refined binary segmentation algorithm in Algorithm 2.

---

**Algorithm 2** : The RBS procedure for multiple change point detection.

---

**Input:** Given the dataset $\{\mathbf{X}, \boldsymbol{Y}\}$ and the user-specified quantile indice $\eta$, initialize the set of change point pairs $T = \{0, 1\}$.

**Step 1-1:** For each pair $\{u, v\}$ in $T$, compute $s = h(u, v)$ as defined in (2.13). If $s > u$, add new pair of nodes $\{u, s\}$ and $\{s, v\}$ to $T$ and update $T$ as $T = T \cup \{u, s\} \cup \{s, v\}$.

**Step 1-2:** Repeat Step 1 until no more new pair of nodes can be added. Denote the terminal set of change point pairs by $T_{\text{final}} = \cup_{i=1}^{q}\{u_i, v_i\}$. Obtain the initial estimator of change points $\check{\boldsymbol{\tau}} = (\check{\tau}_0, ..., \check{\tau}_{\check{k}+1})^\top$, where $\check{k} = \#T_{\text{final}}$ and the underlying regression coefficients $\check{\boldsymbol{\beta}}$ by putting $\check{\boldsymbol{\tau}}$ into (2.8).

**Refining step**

**Step 2-1:** Using $\check{k}$, obtain the improved change point estimator $\check{\boldsymbol{\tau}}^r(\check{k})$ by a dynamic program (2.9).

**Step 2-2:** Obtain the updated estimator of regression coefficients $\check{\boldsymbol{\beta}}^r(\check{\boldsymbol{\tau}}^r)$ by Lasso (2.11).

**Output:** The algorithm provides the change point estimator $\check{\boldsymbol{\tau}}^r = (0, \check{\tau}_1^r, ..., \check{\tau}_{\check{k}}^r, 1)^\top$, and the estimator of regression coefficients $\check{\boldsymbol{\beta}}^r(\check{\boldsymbol{\tau}}^r)$.

---

We obtain the initial estimator of change points and regression coefficients by Steps 1-1 and 1-2 of RBS. Note that different from dynamic programming, Steps 1-1 and 1-2 of RBS are to approximate but not compute the global estimator defined in (2.4). RBS greatly improves the computational efficiency due to searching much fewer candidate intervals for finding a new change point compared to RDP. As a result, RBS is computationally efficient with a computational complexity of $O(n\log(n)\text{Lasso}(n, p))$ (also studied in Leonardi and Bühlmann (2016)), much lower than the computational complexity of $O(n^2\text{Lasso}(n, p))$

for RDP. As the tradeoff between computational efficiency and statistical effectiveness, we prove that the initial estimators of change points and underlying regression coefficients computed by Algorithm 2 (RBS) have lower estimation accuracy than those obtained by Algorithm 1 (RDP). In addition, under relatively strong assumptions, the refined estimators of change points and underlying regression coefficients computed by Algorithm 2 enjoy the same estimation accuracy as those obtained by Algorithm 1.

# 3 Theoretical Properties

With our proposed procedure and algorithms in place, we next examine some theoretical properties. In Sections 3.1 and 3.2, the main results including consistency of estimators of change points $\widetilde{\boldsymbol{\tau}}$ and the underlying regression coefficients $\boldsymbol{\beta}^0$ computed by our two algorithms RDP and RBS are introduced. In Supplementary Section D, we introduce some basic Assumptions A-F that are needed to derive these properties. To save space, we provide brief descriptions of our assumptions. Assumption A includes some mild moment conditions for the design matrix. Assumption B is a standard condition and frequently used for theoretical analysis for the Lasso (Bühlmann and van de Geer, 2011). Assumption C is the well-known compatibility condition, which is an important condition to derive the desired estimation error bound for the Lasso estimator. Assumptions D/D$^*$ specify the minimal SNR conditions to detect change points for RDP/RBS. Assumption E is for the non-smooth quantile loss function and the error term. The error term is usually required to satisfy the Sub-Gaussian or Gaussian distribution in high-dimensional regression models (Bühlmann and van de Geer, 2011). However, we just require that the density function of errors is bounded, which allows us to handle heavy-tailed error distributions and data with heteroscedasticity. Assumptions F/F$^*$ are the technical conditions for RDP/RBS.

To better understand the following theoretical results, we define the minimal signal jump size $m^*_{n,p} = \min_{1\le j\le \widetilde{k}} \frac{\left\|\boldsymbol{\beta}^0(j)-\boldsymbol{\beta}^0(j+1)\right\|_1}{s_*}$, where $\boldsymbol{\beta}^0(j)$ is the regression coefficient vector

of the $j$-th interval and $\widetilde{k}$ is the change point number, and $s_* := \#S_*$ is the cardinality of $S_* = \cup_{j=1}^{\widetilde{k}+1} S^{(j)}$ with $S^{(j)}$ being the sparsity of $\boldsymbol{\beta}^0(j)$. Let $(m^*_{n,p})^2 s_* = \kappa^2_{n,p}$. For RDP, we require the SNR $\delta\kappa^2_{n,p} \geq C_{SNR}\frac{s_* \log^{1+\xi}(n\vee p)}{n}$, where $\xi > 0$ is needed to guarantee consistency of the estimators of change points and underlying regression coefficients, and $C_{SNR}$ is some positive constant. For RBS, we require the SNR condition $\delta\kappa^2_{n,p} \geq C_{SNR}\frac{s_* \log^{1+\xi}(n\vee p)}{\delta^2 n}$. Note that RBS requires stronger SNR conditions also examized in Fryzlewicz (2014); Wang et al. (2020), since the estimation error for RBS has a larger upper bound, with the extra factor of $\frac{1}{\delta^2}$ compared to RDP. In addition, the tuning parameters $\lambda_n$ and $\lambda^r_n$ are required to satisfy

$$\lambda_n \propto (\log p)(\log n)\sqrt{\frac{\log(p)}{n}}, \quad \lambda^r_n \propto \sqrt{\frac{\log(p)}{n}}. \tag{3.1}$$

## 3.1 Results for RDP

The following theorem demonstrates the rates of convergence for the initial (first-step) estimators of change points $\widetilde{\boldsymbol{\tau}}$ and regression coefficients $\boldsymbol{\beta}^0$ obtained by Algorithm 1.

**Theorem 3.1.** (**Consistency of initial estimators**) Suppose Assumptions A-F hold and additionally assume $\lambda_n s_*/\sqrt{\delta} = o(1)$, i.e. $s_*^2(\log p)^3(\log n)^2 = o(\delta n)$, hold. Then we have

(1) $\mathbb{P}\{\widehat{k} = \widetilde{k}\} \to 1$, as $n, p \to \infty$;

(2) $\max_{1\leq j\leq \widehat{k}} |\widehat{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{\lambda_n^2 s_*}{\kappa^2_{n,p}})$;

(3) $\|\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, j) - \boldsymbol{\beta}^0(j)\|_1 = O_P(\lambda_n s_*/\sqrt{\delta})$, for $j = 1, ..., \widehat{k}$.

**Remark 3.2.** Note that, instead of the moment condition of the error term $\epsilon$, our theoretical results require the density function $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ has a constant lower bound $C_\epsilon$. See Assumption E.1 in Section B of the Supplementary Materials. Taking the effect of the error term into account, the convergence rate of the estimators of change points is $\max_{1\leq j\leq \widehat{k}} |\widehat{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{\lambda_n^2 s_*}{C_\epsilon \kappa^2_{n,p}})$. To be specific, the heavy-tailless of the error term leads to a large lower bound $C_\epsilon$, resulting in a slower convergence rate of the estimators of change points. See the detailed discussion about the relationships between the heavy-tailless and the lower bound $C_\epsilon$ in Section D of the Supplementary Materials.

The first and second results of Theorem 3.1 demonstrate the consistency of change points detected by Algorithm 1 in terms of their total number and locations. In particular, Result (1) implies that Algorithm 1 can correctly identify the number of change points with overwhelming probability. Note that we require $\lambda_n \propto (\log p)(\log n)\sqrt{\frac{\log p}{n}}$ in (3.1). Hence, Results (2) and (3) can be further reduced to, for $j = 1, ..., \widehat{k}$,

$$\max_{1\le j\le \widehat{k}} |\widehat{\tau}_j - \widetilde{\tau}_j| = O_P(\log^2 n \frac{\log^3 p s_*}{\kappa_{n,p}^2 n}),\ \|\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, j) - \boldsymbol{\beta}^0(j)\|_1 = O_P(s_* \log p \log n \sqrt{\frac{\log p}{\delta n}}). \tag{3.2}$$

As shown in (3.2), we can obtain consistent estimators of change points and the regression coefficients as long as $s_*^2(\log p)^3(\log n)^2 = o(\delta n)$ holds. Furthermore, Result (3) implies that the achieved convergence rate of the underlying regression coefficient estimator $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}})$ shown in (3.2) is nearly minimax optimal, with additional $\log p \log n$ factors and $\delta$ compared to the rate of regular Lasso estimation (Raskutti et al., 2011). Note that change points and underlying regression coefficients are estimated simultaneously, which results in a suboptimal convergence rate of the initial estimator. In what follows, our proposed refitting procedure in Step 2 improves the convergence rate of the estimator of underlying regression coefficients $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, j)$, where the $\log p \log n$ factor can be eliminated.

We next study theoretical properties of the second-step estimator $\widehat{\boldsymbol{\tau}}^r$ in (2.9) and $\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j)$ in (2.11). To show its oracle property, we compare with the oracle estimator $\ddot{\boldsymbol{\tau}}$ defined in (2.10). Now we are ready to present the theoretical results of $\widehat{\boldsymbol{\tau}}^r$ and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}^r, j)$.

**Theorem 3.3.** (**Consistency of refitted estimators**) Suppose Assumptions A-F hold and additionally $\lambda_n s_*/\sqrt{\delta} = o(1)$, i.e. $s_*^2(\log p)^3(\log n)^2 = o(\delta n)$, hold. Then we have

(1) $\mathbb{P}\{\widehat{k} = \widetilde{k}\} \to 1$, as $n, p \to \infty$;

(2) $\max_{1\le j\le \widehat{k}} |\ddot{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{1}{\kappa_{n,p}^2 n})$;

(3) $\max_{1\le j\le \widehat{k}} |\widehat{\tau}_j^r - \widetilde{\tau}_j| = O_P(\frac{1}{\kappa_{n,p}^2 n})$;

(4) $\|\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j) - \boldsymbol{\beta}^0(j)\|_1 = O_P(\lambda_n^r s_*/\sqrt{\delta})$, for $j = 1, ..., \widehat{k}$.

Theorem 3.3 establishes the improved rate of the refitted estimators of both change points and regression coefficients. In particular, Result (1) is the same as the first result

of Theorem 3.1 since the number of change points is determined in the initial step. Result (2) shows the convergence rate of the oracle estimator $\ddot{\boldsymbol{\tau}}$. For Result (3), it is worth noting that the convergence rate of $\widehat{\boldsymbol{\tau}}^r$ is the same as that of $\ddot{\boldsymbol{\tau}}$. This suggests that, by performing consistent variable selection, our refitting procedure in Step 2 can greatly improve the efficiency of the change point estimator and obtain the oracle property. In addition, Result (3) implies two nontrivial conclusions about the convergence rate:

**Case a:** For the weak SNR case with $\kappa_{n,p} = O((\frac{s_*^2 \log p}{n})^{\frac{1}{4}})$ and $\delta = O(\frac{\log(p)}{\sqrt{n}})$, we can still ensure consistency, which results in the convergence rate in an order of $O_p(\frac{1}{s_*\sqrt{\log(p)n}})$.

**Case b:** For the strong SNR case with $\kappa_{n,p}^2 = O(1)$ and $\delta = O(1)$, we can obtain the optimal convergence rate of order $\frac{1}{n}$ for the change point estimator, as in Oka and Qu (2011) under low-dimensional setups. To our knowledge, this is a novel result developed for Model (1.2).

Combining (3.1), (3.2) and Result (4), the convergence rate of the refitted estimator $\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j)$ is greatly improved. Specifically, recall $K_{max} = 1/\delta$ is the maximum change point number. With $\delta = O(1)$ or $\delta = o(1)$, the refined estimator $\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j)$ has different properties. For the first case, we have $\delta = O(1)$, which means that the maximum number of change points is fixed and does not increase with $n$. Note that we require $\lambda_n^r \propto \sqrt{\frac{\log p}{n}}$ in (3.1). Combining (3.1), we can reduce Results (4) in Theorem 3.3 to:

$$\|\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j) - \boldsymbol{\beta}^0(j)\|_1 = O_p\Big(s_*\sqrt{\tfrac{\log(p)}{n}}\Big), \text{ for } j = 1, ..., \widehat{k}. \tag{3.3}$$

It implies that, by assuming finite changes occur, our refined estimator of regression coefficients $\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j)$ computed by Algorithm 1 is minimax optimal (Raskutti et al., 2011). We next consider the second case with $\delta = o(1)$, where we allow the number of change points to grow with $n$. Combining (3.1), in this case, Results (4) in Theorem 3.3 reduces to:

$$\|\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j) - \boldsymbol{\beta}^0(j)\|_1 = O_p\Big(s_*\sqrt{\tfrac{\log(p)}{\delta n}}\Big), \text{ for } j = 1, ..., \widehat{k}. \tag{3.4}$$

Hence, by (3.4), consistency of the refined estimator can still be obtained as long as

$s_*\sqrt{\frac{\log(p)}{\delta n}} = o(1)$ holds. In other words, we allow the number of change points $\widetilde{k}$ grow with $n$, with an order of $o\big(\frac{n}{\log(p)s_*^2}\big)$. Finally, we summarize the convergence rates of the estimators of change points and regression coefficients by Algorithm 1 in Table 1.

Table 1: Convergence rates for our proposed estimators by Algorithm 1.

| signal | | Convergence rates for the following estimators | | | |
|---|---|---|---|---|---|
| $\kappa_{n,p}$ | $\delta$ | $\|\widehat{\boldsymbol{\tau}} - \widetilde{\boldsymbol{\tau}}\|_\infty$ | $\|\widehat{\boldsymbol{\tau}}^r - \widetilde{\boldsymbol{\tau}}\|_\infty$ | $\|\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\tau}}, j) - \boldsymbol{\beta}^0(j)\|_1$ | $\|\widehat{\boldsymbol{\beta}}^r(\widehat{\boldsymbol{\tau}}^r, j) - \boldsymbol{\beta}^0(j)\|_1$ |
| $O(1)$ | sparse $O(1)$ | $\log^2 n \frac{\log^3 p}{n}$ | $\frac{1}{n}$ | $s_*(\log p)(\log n)\sqrt{\frac{\log p}{n}}$ | $s_*\sqrt{\frac{\log p}{n}}$ |
| $(\frac{s_*^2 \log p}{n})^{\frac{1}{4}}$ | sparse $O(1)$ | $\log^2 n \frac{\log^{5/2}(p)}{\sqrt{n}}$ | $\frac{1}{s_*\sqrt{n\log(p)}}$ | $s_*(\log p)(\log n)\sqrt{\frac{\log p}{n}}$ | $s_*\sqrt{\frac{\log p}{n}}$ |
| $O(1)$ | dense $o(1)$ | $\log^2 n \frac{\log^3 p}{n}$ | $\frac{1}{n}$ | $s_*(\log p)(\log n)\sqrt{\frac{\log p}{\delta n}}$ | $s_*\sqrt{\frac{\log p}{\delta n}}$ |
| $(\frac{s_*^2 \log p}{n})^{\frac{1}{4}}$ | dense $o(1)$ | $s_* \log^2 n \frac{\log^{5/2}(p)}{\sqrt{n}}$ | $\frac{1}{s_*\sqrt{n\log(p)}}$ | $s_*(\log p)(\log n)\sqrt{\frac{\log p}{\delta n}}$ | $s_*\sqrt{\frac{\log p}{\delta n}}$ |

**Remark 3.4.** As shown in Table 1, we present the convergence rate for the initial and refined estimators by Algorithm 1. For change point estimation, the convergence rate of the refined estimator is $\log^2 n \log^3 p$ faster than that of the initial estimator. For the estimation of the underlying regression coefficients, the convergence rate of the refined estimator is $\log p \log n$ faster than that of the initial estimator. Note that we consider the change point estimation problem under the high-dimensional setup with $\sqrt{\frac{\log p}{n}} = o(1)$. Then, the refinement procedure makes a significant improvement in change point estimation. As for the computational complexity, the refinement procedure does not require solving any Lasso problems, resulting in negligible computational cost compared to the initial step, which is also shown in our numerical results in Section 4.2. Hence, the refinement highly improves the estimation accuracy without increasing much computational cost.

**Remark 3.5.** Note that Leonardi and Bühlmann (2016) (denoted by L&B) considered multiple change point detection for high-dimensional mean regression models. There are several important differences between Leonardi and Bühlmann (2016) and our methods. Firstly, Leonardi and Bühlmann (2016) considers the $\ell_2$-based loss function, while we construct

our method based on the quantile loss function. Hence, we can handle both light-tailed and heavy-tailed error distributions as well as data with heteroscedasticity. In the latter two cases, the method by Leonardi and Bühlmann (2016) may not be able to estimate the change points accurately, which can result in poor variable selections and prediction in each segmentation. Secondly, compared to their one-step least-square estimation method, our estimation method includes an additional refitting procedure in Step 2, which can greatly improve the efficiency of detecting and localizing change points. To be specific, when assuming the strong SNR case with $\kappa_{n,p}^2 = O(1)$ and $\delta = O(1)$, the convergence rate of our proposed refined change point estimator $\widehat{\boldsymbol{\tau}}^r$, with the order of $O_P(\frac{1}{n})$, is a significant improvement from the rate $O_P(\sqrt{\log p/n})$ in Leonardi and Bühlmann (2016). Finally, in terms of theoretical proof, we need to overcome several technical challenges arising from the non-smoothness of the quantile loss function.

## 3.2 Results for RBS

We now present the theoretical results computed by RBS. First, we introduce the convergence rate of the first-step estimators for $\widetilde{\boldsymbol{\tau}}$ and $\boldsymbol{\beta}^0$ obtained by Algorithm 2.

**Theorem 3.6.** (Consistency of initial estimators) Suppose Assumptions A-C, D$^*$, E and F$^*$ hold and additionally assume $\frac{\lambda_n s_*/\sqrt{\delta}}{\delta^2} = o_p(1)$ and $s_*^2(\log p)^3(\log n)^2 = o(\delta^5 n)$ hold. Then we have

(1) $\mathbb{P}\{l(\check{\boldsymbol{\tau}}) = \widetilde{k}\} \to 1$, as $n, p \to \infty$;

(2) $\max_{1\le j\le \widetilde{k}} |\check{\tau}_j - \widetilde{\tau}_j| = O_P(\frac{\lambda_n^2 s_*}{\kappa_{n,p}^2 \delta^2})$;

(3) $\|\check{\boldsymbol{\beta}}(\check{\boldsymbol{\tau}}, j) - \boldsymbol{\beta}^0(j)\|_1 = O_P(\frac{\lambda_n s_*/\sqrt{\delta}}{\delta^2})$, for $j = 1, ..., \widehat{k}$.

Theorems 3.6 demonstrates the consistency and convergence rate of the estimators of change point number and locations as well as the underlying regression coefficients. To guarantee the consistency of the estimators, RBS requires stronger assumptions (Assumptions D$^*$ and F$^*$) on the minimal signal jump size $\kappa_{n,p}$ and the minimal segment length $\delta$.

Furthermore, compared to RDP, the convergence rates of the initial estimators of change point locations and underlying regression coefficients are slower, with an additional $\frac{1}{\delta^2}$ term. Hence, RBS is more computationally efficient but less statistically accurate.

Next, we present the theoretical results of the second-step estimators $\check{\boldsymbol{\tau}}^r$ and $\check{\boldsymbol{\beta}}^r(\check{\boldsymbol{\tau}}^r, j)$ computed by Algorithm 2.

**Theorem 3.7.** (Consistency of refitted estimators) Suppose Assumptions A-C, D$^*$, E and F$^*$ hold and further assume $\frac{\lambda_n s_* / \sqrt{\delta}}{\delta^2} = o_p(1)$ and $s_*^2 (\log p)^3 (\log n)^2 = o(\delta^5 n)$ hold. Then we have

(1) $\mathbb{P}\{l(\check{\boldsymbol{\tau}}^r) = \widetilde{k}\} \to 1$, as $n, p \to \infty$;

(2) $\max_{1 \le j \le \widetilde{k}} |\check{\tau}_j^r - \widetilde{\tau}_j| = O_p(\frac{1}{\kappa_{n,p}^2 n})$ ;

(3) $\max_{1 \le j \le \widetilde{k}} |\ddot{\tau}_j - \widetilde{\tau}_j| = O_p(\frac{1}{\kappa_{n,p}^2 n})$;

(4) $\|\check{\boldsymbol{\beta}}^r(\check{\boldsymbol{\tau}}^r, j) - \boldsymbol{\beta}^0(j)\|_1 = O_p(\lambda_n^r s_* / \sqrt{\delta})$, for $j = 1, ..., \widehat{k}$.

Theorem 3.7 shows the same results as those in Theorem 3.3, which include consistency and convergence rates of estimators $(\check{\boldsymbol{\tau}}^r, \check{\boldsymbol{\beta}}^r)$ computed by RBS under relatively stronger assumptions. In other words, if we allow larger segment length and higher signal jump sizes, the refined estimator obtained by RBS can enjoy the same statistical accuracy as that by RDP. More importantly, to appreciate the usefulness of RBS, the computational efficiency is greatly improved by RBS, by reducing the order of computational operations from $O(n^2 \text{Lasso}(n, p))$ to $O(n \log(n) \text{Lasso}(n, p))$. This will be extensively demonstrated by our numerical results.

**Remark 3.8.** Wang et al. (2021) proposed a projection-based algorithm, Variance Projected Wild Binary Segmentation (VPWBS), for multiple change point detection for high-dimensional linear models. Under a similar SNR condition, VPWBS is shown to achieve the localization rate $O_p(\frac{s_* \log(n)}{\kappa_{n,p}^2 n})$. In this paper, we use standard binary segmentation to obtain our loss function-based estimators for high-dimensional quantile regression models. It is well known that wild binary segmentation (WBS) and seeded binary segmentation (SBS)

Table 2: Localization error bound and computational complexity of existing methods. Note that $M$ is the number of randomly generated intervals for WBS, and QM and LM refer to quantile models and linear models.

| Methods | Localization Error $\|\widehat{\boldsymbol{\tau}} - \widetilde{\boldsymbol{\tau}}\|_\infty$ | SNR Condition | Computational Complexity |
|---|---|---|---|
| RDP/QM | $\frac{1}{\kappa_{n,p}^2 n}$ | $\delta\kappa_{n,p}^2 \geq s_* \log^{(1+\xi)}(p \vee n)/n$ | $O(n^2\text{Lasso}(n,p))$ |
| RBS/QM | $\frac{1}{\kappa_{n,p}^2 n}$ | $\delta^3\kappa_{n,p}^2 \geq s_* \log^{(1+\xi)}(p \vee n)/n$ | $O(n\log(n)\text{Lasso}(n,p))$ |
| VPWBS/LM | $\frac{s_* \log(n)}{\kappa_{n,p}^2 n}$ | $\delta\kappa_{n,p}^2 \geq s_* \log^{(1+\xi)}(pn)/n$ | $O(nM\log(n)\text{GroupLasso}(n,p))$ |
| WBS | - | - | $O(nM\log(n)\text{Lasso}(n,p))$ |
| SBS | - | - | $O(n\log(n)\text{Lasso}(n,p))$ |
| OSBS | - | - | $O(n\text{Lasso}(n,p))$ |

are useful techniques for computationally and statistically efficient change point detection. Recently, Kovács et al. (2020) proposed the optimistic seeded binary segmentation (OSBS) algorithm. There are essential difficulties to obtain corresponding theoretical properties for these three algorithms under high-dimensional quantile regression models. We provide a detailed discussion about the connection between RBS with the existing algorithms in Section B of the Supplementary Materials. Lastly, we present the localization error bound, the SNR condition and the computational complexity of our proposed RDP, RBS, and the other state-of-art algorithms in Table 2.

**Remark 3.9.** There are two main novel points of the proof. First, this paper considers the multiple change point estimation problem for the quantile regression model. We need to make sure that the number of change points is consistent, which leads to considering multiple scenarios such as overestimation and underestimation referred to Lemmas F.5 and F.6. In each case, we have to handle mixed data (that is, data from two or more different populations). Note that we consider the change point problem of quantile regression and we need to control the estimation error based on the nonsmooth quantile loss function for the mixed data. Dealing with nonsmooth quantile loss based problem is more difficult than $\ell_2$ loss based ones (Leonardi and Bühlmann (2016); Wang et al. (2021); Rinaldo et al.

(2021). Moreover, to obtain the desired results for our quantile change point estimation, we need to carefully adopt some quantile loss related techniques. For example, we need to control the empirical process arising from quantile loss for the mixed data referred to Lemmas F.1 and F.2. Second, the joint estimates of change points and regression coefficients obtained in the initial step are not optimal. This prompted us to refine the initial estimator with an additional second step. Different from the refinement procedure of the single change point case (Kaul et al. (2019); Lee et al. (2018)) and multiple change points case (Rinaldo et al. (2021)), we use dynamic programming to refine the whole change point locations. In particular, all the change points are refined at the same time, resulting in an overall optimal solution. In the proof (see details in Appendix Section G.2), we mainly control two quantities: the excess risk (the difference between theoretical mean and empirical mean of loss function) and the empirical process. To apply some techniques of M-estimation, we define a new criterion about aforementioned two quantities based on our dynamic programming refinement procedure and further verify the feasibility of this criterion.

# 4 Empirical Studies

We first introduce some basic model settings for our simulation. Consider the quantile regression models based on the following data generating process: for $i = 1, ..., n$, $Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(i)} + \epsilon_i$. For the $p$-dimensional covariate $\boldsymbol{X}_i$, we generate $\boldsymbol{X}_i = (1, \boldsymbol{Z}_i^\top)^\top$, where $\boldsymbol{Z}_i$'s are i.i.d. from $N(\mathbf{0}, \boldsymbol{\Sigma})$. We consider the covariance matrix structure with the weakly dependent setting, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^* = (\sigma_{ij})$, where $\sigma_{i,j}^* = 0.5^{|i-j|}$ for $1 \le i, j \le p-1$. Note that the conditional $\eta-$quantile of $Y_i$ given $\boldsymbol{X}_i$ has the form: $Q_\eta(Y_i|\boldsymbol{X}_i) = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(i)} + Q_\eta(\epsilon_i)$, where $Q_\eta(\epsilon_i)$ satisfies $\mathbb{P}(\epsilon_i \le Q_\eta(\epsilon_i)) = \eta$. Furthermore, various quantile indices $\eta \in (0, 1)$ may be of interest in real applications, which correspond to different loss functions in (2.1). Hence, to show their performance, we choose two types of quantiles as **Case 1** with $\eta = 0.5$ and

**Case 2** with $\eta = 0.25$. To demonstrate the robustness and adaptability of our proposed method, for each case, we consider four models in terms of the error term.

**Model 1 (light tail):** the error term $\{\epsilon_i\}_{i=1}^n$ are i.i.d generated from $N(0,1)$;

**Model 2 (heavy tail):** the error term $\{\epsilon_i\}_{i=1}^n$ are i.i.d generated from $t_3$ (Student's $t$-distribution with degrees of freedom of 3);

**Model 3 (heavy tail and heteroscedasticity):** the model has the following form: $Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(i)} + \boldsymbol{X}_i^\top \boldsymbol{\xi}\epsilon_i$, where $\boldsymbol{X}_i = (1, \boldsymbol{Z}_i)^\top$, $\boldsymbol{Z}_i$ is from the truncated normal distribution $\Psi(0, \boldsymbol{\Sigma}, -10, 10)$, and $\boldsymbol{\xi} = (0, 1, 0, ...., 0)$ and the errors $\{\epsilon_i\}_{i=1}^n$ are i.i.d generated from $t_2$. We denote this distribution by $\boldsymbol{X} * t_2$;

**Model 4 (heavy tail):** the error term $\{\epsilon_i\}_{i=1}^n$ are i.i.d generated from $t_2$.

For comparison, we evaluate the following methods. **Method a:** The lasso-based method in the context of high-dimensional linear models introduced in Leonardi and Bühlmann (2016). Note that Leonardi and Bühlmann (2016) introduced two algorithms, which are denoted by L&B-DPA and L&B-BSA. **Method b:** Our proposed method based on Algorithm 2. For the two-step estimation, we denote the first-step and second-step estimators by RBS-one and RBS-refit. Note that the numerical performance of RDP is included in the supplement.

We consider the scenario where $(\boldsymbol{\beta}^{(i)})_{1\le i\le n}$ have three change points at $\widetilde{\boldsymbol{\tau}} = (0, 0.25, 0.5, 0.75, 1)^\top$. We set $n = 1000$ and $p = 200$. The three change points divide the data into four segments with regression coefficients $\boldsymbol{\beta}(1)$, $\boldsymbol{\beta}(2)$, $\boldsymbol{\beta}(3)$ and $\boldsymbol{\beta}(4)$. Specifically, for $s_1 \in \mathcal{S}^1$, we set $\beta_{s_1}(1) \overset{\text{iid}}{\sim} \mathrm{U}(0, 2)$. Then, for $s_j \in \mathcal{S}^j$ $(j = 2, 3, 4)$, we set $\beta_{s_j}(j) = \beta_{s_j}(j-1) + (j-1)\delta_{s_j}$ with $\delta_{s_j} \overset{\text{i.i.d.}}{\sim} \mathrm{U}(0, 10\sqrt{\log(p)/(\delta n)})$. Meanwhile, the support set of regression coefficients $\mathcal{S}$ is randomly selected from the set $\{1, 2, ..., 2[\log(p)]\}$ with $\#\mathcal{S}^j = [\log(p)]$, $j = 1, 2, 3, 4$.

In Figure 1, we report the Hausdorff distance (Munkres (2000)) with the range of $(0, 1)$, which measures the similarity between estimated change points and true change points. A small value of the Hausdorff distance indicates high estimation accuracy. As shown in Figure 1, L&B-BSA has the best performance and our methods perform very closely to L&B-BSA in Model 1 with normal error distributions. The RBS with $\eta = 0.5$

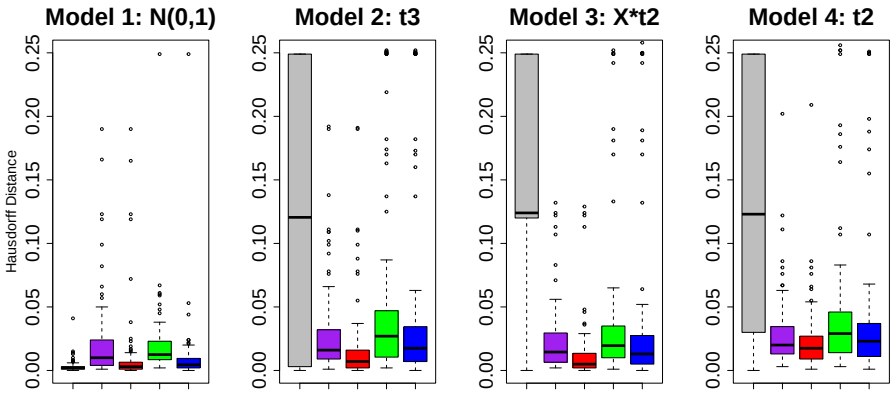


Figure 1: The Hausdorff distance between true change points and the estimators by (from left to right) L&B-BSA (gray), RBS-one with $\eta = 0.5$ (purple), RBS-refit with $\eta = 0.5$ (red), RBS-one with $\eta = 0.25$ (green) and RBS-refit with $\eta = 0.25$ (blue) in the setting of multiple change points. The results are based on 100 replications.

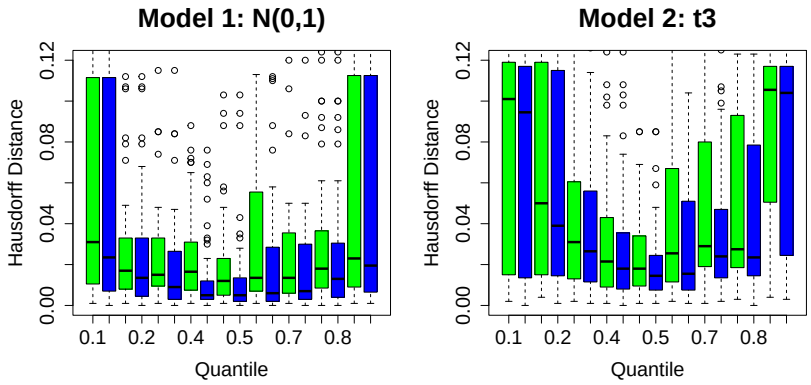


Figure 2: The accuracy of multiple change point estimation by RBS-one (green) and RBS-refit (blue) under various quantiles with two types of tails. Left panel: The Hausdorff distance of estimated change points in Model 1. Right panel: The Hausdorff distance in Model 2. The results are based on 100 replications.

performs the best and RBS with $\eta = 0.25$ performs slightly worse for three heavy-tailed error distributions. Table 3 summarizes the results of multiple change point detection. For the change point number, we record the mean and the proportion that the methods can correctly identify three change points among the 100 replications, which are named Mean and Rate (%). Meanwhile, for the locations of change points, we report mean and the root-mean-squared error (RMSE). We see that for the identification of the change point number, L&B-BSA has the best performance in normal distribution but suffers from heavy-

Table 3: Multiple change point detection for Cases 1 and 2 with $\widetilde{\boldsymbol{\tau}} = (0, 0.25, 0.5, 0.75, 1)^\top$ (corresponding to the locations 250, 500, 750). The numerical results are based on 100 replications.

| | | **Accuracy for the following methods** | | | | |
|---|---|---|---|---|---|---|
| | quantile | | $\eta = 0.5$ | | $\eta = 0.25$ | |
| | Models | L&B-BSA | RBS-one | RBS-refit | RBS-one | RBS-refit |
| **Number** | $N(0,1)$ | 3.00 \| 100 | 3.00 \| 98 | 3.00 \| 98 | 2.98 \| 98 | 2.98 \| 98 |
| | $t_3$ | 2.65 \| 43 | 3.00 \| 96 | 3.00 \| 96 | 2.94 \| 94 | 2.94 \| 94 |
| Mean \| Rate(%) | $\boldsymbol{X} * t_2$ | 3.72 \| 28 | 3.04 \| 95 | 3.04 \| 95 | 2.85 \| 90 | 2.85 \| 90 |
| | $t_2$ | 3.22 \| 32 | 3.04 \| 94 | 3.04 \| 94 | 2.91 \| 92 | 2.91 \| 92 |
| **Location 1** | $N(0,1)$ | 251 \| 4.42 | 254 \| 10.55 | 251 \| 5.28 | 252 \| 10.59 | 251 \| 5.82 |
| | $t_3$ | – | 252 \| 12.76 | 250 \| 9.15 | 253 \| 16.24 | 251 \| 12.75 |
| Mean \| RMSE | $\boldsymbol{X} * t_2$ | – | 253 \| 12.89 | 251 \| 8.97 | 252 \| 14.78 | 250 \| 11.90 |
| | $t_2$ | – | 250 \| 18.11 | 249 \| 16.74 | 253 \| 28.27 | 252 \| 26.59 |
| **Location 2** | $N(0,1)$ | 500 \| 2.51 | 499 \| 12.92 | 500 \| 6.75 | 499 \| 14.84 | 500 \| 8.91 |
| | $t_3$ | – | 501 \| 14.48 | 501 \| 8.53 | 498 \| 17.43 | 500\| 10.40 |
| Mean \| RMSE | $\boldsymbol{X} * t_2$ | – | 499 \| 11.66 | 499 \| 9.75 | 502 \| 18.00 | 501 \| 15.10 |
| | $t_2$ | – | 498 \| 20.51 | 498 \| 18.55 | 499 \| 21.87 | 499 \| 20.82 |
| **Location 3** | $N(0,1)$ | 750 \| 3.01 | 744 \| 12.92 | 747 \| 6.53 | 747 \| 13.10 | 748 \| 8.74 |
| | $t_3$ | – | 745 \| 12.14 | 745 \| 8.29 | 751 \| 16.11 | 749 \| 14.25 |
| Mean \| RMSE | $\boldsymbol{X} * t_2$ | – | 743 \| 11.43 | 745 \| 7.53 | 748 \| 10.77 | 748 \| 8.87 |
| | $t_2$ | – | 745 \| 16.37 | 745 \| 15.19 | 747 \| 24.48 | 748 \| 21.88 |

tailed and heteroscedastic error distributions. Meanwhile, RBS can correctly estimate the change point number with high accuracy for various error distributions. As for the change point locations, L&B-BSA has the best performance in Model 1 and RBS with $\eta = 0.5$ performs very similarly to L&B-BSA. While in Models 2, 3 and 4, RBS with $\eta = 0.5$ performs much better than L&B-BSA. Furthermore, Table 3 clearly shows that the refitted process significantly improves the estimation accuracy of change points compared to the initial estimation result.

To show the performance of our method with various quantile indices in multiple change point models, we set $\eta \in \{0.1, 0.2, ..., 0.8, 0.9\}$ consisting of both moderate and extreme quantile indices, which are reported in Figure 2. Based on the results, we can draw two conclusions: (1) the refined method RBS-refit can significantly improve the statistical efficiency compared to RBS-one (one-step initial estimator); (2) our method based on all these quantile indices has satisfactory performance, and the one with $\eta = 0.5$ performs the best.

In addition to the above results, we have also investigated the performance of our proposed approaches of RDP, RBS and the other existing methods (Lee et al., 2018; Leonardi and Bühlmann, 2016) under no change point and single change point models. The extensive results show the advantage of our methods including computational efficiency and statistical accuracy. Note that our proposed methods involve the selection of two parameters $\lambda_n$ and $\gamma$. To save space, we provide the detailed results and a cross-validation approach to select tuning parameters in Supplementary Section I.1.

## 4.1 Comparison with other binary segmentation based variants

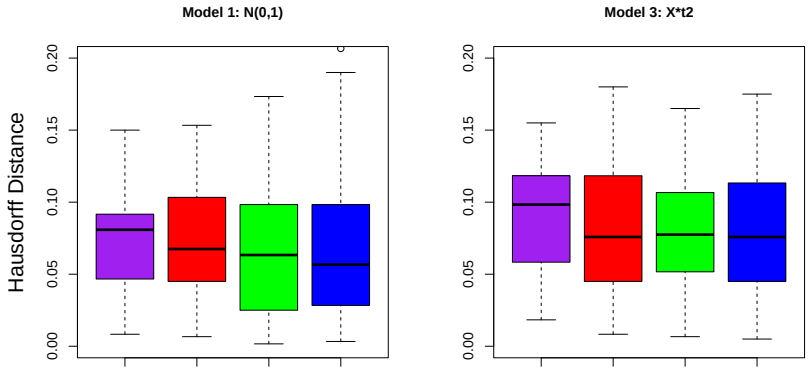


Figure 3: The Hausdorff distance between true change points and the estimators by (from left to right) OSBS (purple), RBS (red), SBS (green), WBS with $\eta = 0.5$ in multiple change point setting. The results are based on 100 replications.

To investigate the numerical performance for higher dimensional data, we consider the above mentioned three-change-point model by setting $n = 600, p = 200$. Note that the sample size in each segment is 150, which is smaller than the data dimension 200. We focus on the computational efficiency and statistical accuracy of our proposed methods and the other state-of-art algorithms including: (1) Wild Binary Segmentation algorithm (WBS) proposed by Fryzlewicz (2014); (2) Seeded Binary Segmentation algorithm (SBS) proposed by Kovács et al. (2023); (3) Optimistic Seeded Binary Segmentation algorithm (OSBS) proposed by Kovács et al. (2020). Note that these three algorithms in the original

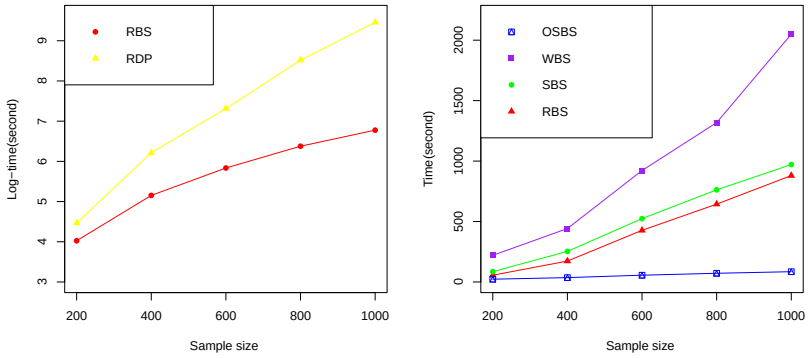


Figure 4: The computational efficiency of RBS, RDP and the existing methods with $n \in \{200, 400, 600, 800, 1000\}$ and $p = 200$. The results are based on 10 replications.

papers are designed for mean change point problems and are based on CUSUM statistics. For a better comparison under the high-dimensional quantile regression model, we modify these algorithms to detect change points based on the quantile loss function. More discussions about the detailed implementation and the theoretical difficulty of those methods are provided in the Supplementary Section B. We present the estimation results in Figure 3. Under the current model setting, all algorithms have satisfactory performance under a high-dimensional setup with $p = 200, n = 600$. Besides, WBS has the best performance and OSBS has a relatively worse performance. Note that we only present the performance of these algorithms empirically. Using these algorithms to deal with the change-point problem in quantile regression has some inherent theoretical difficulties. See more discussion about the connections and theoretical difficulties in Supplementary Section I.1.

## 4.2 Computational cost

To show the computational efficiency of our proposed approaches, we report the computational time. We implement the corresponding algorithms independently on a CPU (Linux) with 2.50GHz, 6 cores, and 4GB of RAM. We consider the three-change-point model setting in Section 4 of the main paper. As shown in Figure 4 (Left), the compu-

tational cost of RBS grows moderately (56s to 877s) with the sample size increasing from 200 to 1000, while the computational cost of RDP has a quadratic growth (87s to 12000s). Hence, RBS dramatically reduces the computational cost compared to RDP. Furthermore, we compare our proposed RBS with the aforementioned binary segmentation based algorithms: OSBS, SBS, and WBS, in terms of computational cost. As shown in Figure 4 (Right), under the current setting, OSBS has the highest computational efficiency, and WBS with a random interval $M = 20$ has the lowest. SBS and RBS have comparable computational costs. Moreover, Figure 4 shows that the computational cost for WBS grows rapidly as the number $M$ of randomly generated intervals increases, this is consistent with the observation in Kovács et al. (2023).

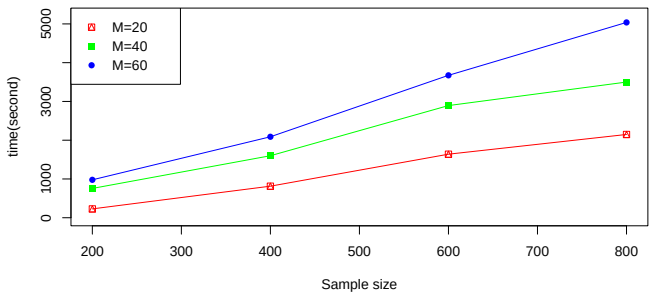


Figure 5: The computational efficiency of WBS with $n = \{200, 400, 600, 800\}$, $p = 200$ and $M \in \{20, 40, 60\}$. The results are based on 10 replications.

# 5 Application for S&P 100 data

In this section, we apply our proposed approach to the S&P 100 index data for efficient multiple change points detection. The S&P 100 dataset obtained from Yahoo! Finance (https://finance.yahoo.com/) includes the largest and most established 100 companies in the S&P 100. We collect the daily returns of 76 stocks that have remained in the S&P 100 index consistently from January 3, 2007 to December 30, 2011. This covers the financial

crisis beginning in 2008 and some other important events, resulting in $n = 1259$.

In financial marketing, predicting stock returns is of great interest (Fan et al. (2019)). Due to the frequent presence of the heavy tail and heteroscedasticity of financial stocks, quantile regression is commonly adopted to model different quantiles of stock returns. For this dataset, we consider the following quantile regression model: $Q_\eta(Y_i|\boldsymbol{X}_i) = \boldsymbol{X}_i^\top \boldsymbol{\beta}_i$, where $X_{ij}$ is the moving average (MA) of returns of the $j$-th stock on the $i$th day in the past 1 day and $Y_i$ is the return of the selected stock on the $(i+1)$th day. Hence, we have $\boldsymbol{X}_i \in \mathbb{R}^{76}, i = 1, ..., 1258$. Note that we can also let $\boldsymbol{X}_i$ be the moving average of returns of stocks in the past 3, 5, and 7 days, resulting in $\boldsymbol{X}_i \in \mathbb{R}^{76\times 2}$ (1, 3 days), $\boldsymbol{X}_i \in \mathbb{R}^{76\times 3}$ (1, 3, 5 days), and $\boldsymbol{X}_i \in \mathbb{R}^{76\times 4}$ (1, 3, 5, 7 days), respectively. To save space, we put the corresponding results in the Supplementary Materials. Our goal is to detect change points on the regression coefficient $\boldsymbol{\beta}$ during this period when we predict the target stock using the remaining ones under different quantile levels $\eta$. If changes are detected, we aim to estimate the number and locations of change points. It is worth mentioning that for simplicity, we treat the instances $(Y_i, \boldsymbol{X}_i)$ as independent replicates, even though they form a time series. This strategy has been commonly used in the literature on change point analysis (Lee et al., 2016; Fryzlewicz, 2014). We choose the tuning parameters $\lambda_n$ and $\gamma$ as suggested in Section I.1 of the Supplementary Materials. We use RBS to analyze the data since it is more computationally efficient than RDP and enjoys satisfactory accuracy. Note that we could select any stock to be the response variable $Y$. Here we present the result for "JPMorgan" to be the response variable.

As shown in Table 4, RBS with different choices of quantiles $\eta \in \{0.3, 0.5, 0.8\}$ can detect different multiple change points during this period. To better explain this result, we refer to the 'T-bills and ED' (TED) spread, which is short for the difference between the 3-month of London Inter-Bank Offer Rate (LIBOR), and the 3-month short-term U.S. government debt ("T-bills"). The TED spread, an indicator of perceived risk in the general economy,

increased significantly during the financial crisis, reflecting an increase in perceived credit risk. The plot of TED is shown in Figure 6, with the red dotted line corresponding to the estimated dates (change points) with TED Spread$^{\#}$ in Table 4. We can observe that the TED spread fluctuates tremendously near the estimated dates (change points) detected by RBS. In addition, during the financial crisis in 2008 or Global Financial Crisis (GFC), the TED spread spiked up in July 2007, remained volatile for a year, then spiked even higher in September 2008, reaching a record 4.65% on October 10, 2008. Interestingly, the estimated dates by RBS correspond to some important events. For example, the "nadir of crisis" in March 2009 (identified by the New York Times due to the Dow Jones average index reaching a trough of around 6600) has been detected by our approaches. Due to the fear of contagion of the European sovereign debt crisis, there was a sharp drop in stock prices in stock exchanges across the United States, the Middle East, Europe, and Asia. This event is detected by RBS with $\eta = 0.3$ and $\eta = 0.5$ on June 15, 2011 and June 03, 2011, respectively. Note that similar change points have also been detected in the study of Cho and Fryzlewicz (2015). In summary, the analysis of the S&P 100 dataset gives strong support to the usefulness of our proposed method.

Table 4: Change point estimation results of RBS with $\eta = 0.3, 0.5, 0.7$ for S&P 100 from 2007/01/03 to 2011/12/31. Dates with TED Spread$^{\#}$ correspond to the red dotted line in Figure 6.

| $\eta$ | $\widehat{\tau}$ | Date | |
|---|---|---|---|
| 0.3 | 410 | 2008/08/19 | Global Financial Crisis (GFC) |
| | 551 | 2009/03/19 | "nadir of the crisis" |
| | 744 | 2009/12/14 | |
| | 899 | 2010/07/28 | TED Spread$^{\#}$ |
| | 1114 | 2011/06/03 | Global stock markets fell due to fears of contagion of the European sovereign debt crisis |
| 0.5 | 215 | 2007/11/13 | TED Spread$^{\#}$ |
| | 424 | 2008/09/15 | Global Financial Crisis (GFC) |
| | 557 | 2009/03/19 | "nadir of the crisis" |
| | 689 | 2009/09/28 | |
| | 890 | 2010/07/16 | TED Spread$^{\#}$ |
| | 1120 | 2011/06/15 | Global stock markets fell due to fears of contagion of the European sovereign debt crisis |
| 0.8 | 309 | 2008/03/26 | TED Spread$^{\#}$ |
| | 439 | 2008/09/29 | TED Spread$^{\#}$ |
| | 796 | 2010/03/02 | |
| | 967 | 2010/11/02 | |
| | 1030 | 2011/02/02 | |

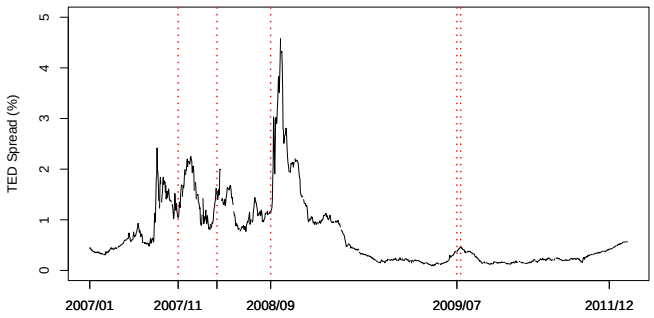

Figure 6: TED Spread during 2007 to 2011 with the estimated change points (red dotted lines) marked by # in Table 4.

# 6 Summary

This paper provides statistically and computationally efficient multiple change point detection and localization for high dimensional quantile regression models. Our quantile models are flexible and general, which contain a possibly infinite collection of quantile regressions. We construct a general two-step estimation method including estimating the change point number and locations as well as the underlying regression coefficients. For implementation, we develop two new algorithms, RDP and RBS. Different from the traditional dynamic programming and binary segmentation techniques, the additional refined step is provided in both RDP and RBS. The refining process is computationally eliminable since it does not solve any Lasso problems. Hence, it greatly improves the estimation accuracy without increasing the computational burden. Under very mild assumptions, we establish theoretical results in the high-dimensional framework with $(n, p) \to \infty$ simultaneously. We show the consistency of the change point number, the oracle estimation of change point locations, and the optimal convergence rate of the estimator for the underlying regression coefficients. The numerical results demonstrate that our proposed method is statistically and computationally efficient and robust to nonnormality and heteroscedasticity.

# References


Auger, I. E. and Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bulletin of mathematical biology* **51** 39–54.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78.

Belloni, A. and Chernozhukov, V. (2011). l1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics* **39** 82 – 130.

Belloni, A., Chernozhukov, V. and Kato, K. (2019). Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *Journal of the American Statistical Association* **114** 749–758.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data*. Springer.

Castro, B. M., Lemes, R. B., Cesar, J., Hünemeier, T. and Leonardi, F. (2018). A model selection approach for multiple sequence segmentation and dimensionality reduction. *Journal of Multivariate Analysis* **167** 319–330.

Chen, X., Liu, W. and Zhang, Y. (2019). Quantile regression under memory constraint. *The Annals of Statistics* **47** 3244–3273.

Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77** 475–507.

Ciuperca, G. and Maciak, M. (2020). Change-point detection in a linear model by adaptive fused quantile method. *Scandinavian Journal of Statistics* **47** 425–463.

Dette, H., Pan, G. and Yang, Q. (2022). Estimating a change point in a sequence of very high-dimensional covariance matrices. *Journal of the American Statistical Association* **117** 444–454.

Fan, J., Fan, Y. and Barut, E. (2014). Adaptive robust variable selection. *The Annals of Statistics* **42** 324 – 351.

Fan, J., Gong, W. and Zhu, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of Econometrics* **212** 177–202.

Friedrich, F., Kempe, A., Liebscher, V. and Winkler, G. (2008). Complexity penalized m-estimation: Fast computation. *Journal of Computational and Graphical Statistics* **17** 201–224.

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

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L. and Tsai, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters* **12** 105–108.

Kaul, A., Jandhyala, V. K. and Fotopoulos, S. B. (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *The Journal of Machine Learning Research* **20** 1–40.

Killick, R., Fearnhead, P. and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107** 1590–1598.

Kim, C.-J., Morley, J. C. and Nelson, C. R. (2005). The structural break in the equity premium. *Journal of Business & Economic Statistics* **23** 181–191.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

Kovács, S., Bühlmann, P., Li, H., and Munk, A. (2023). Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika* **110** 249–256.

Kovács, S., Li, H., Haubner, L., Munk, A. and Bühlmann, P. (2020). Optimistic search strategy: Change point detection for large-scale data via adaptive logarithmic queries. *arXiv:2010.10194* .

Lee, S., Liao, Y., Seo, M. H. and Shin, Y. (2018). Oracle estimation of a change point in high-dimensional quantile regression. *Journal of the American Statistical Association* **113** 1184–1194.

Lee, S., Seo, M. H. and Shin, Y. (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association* **106** 220–231.

Lee, S., Seo, M. H. and Shin, Y. (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78** 193–210.

Leonardi, F. and Bühlmann, P. (2016). Computationally efficient change point detection for high-dimensional regression. *Preprint arXiv:1601.03704* .

Londschien, M., Bühlmann, P. and Kovács, S. (2022). Random forests for change point detection. *Preprint arXiv:2205.04997* .

Maidstone, R., Hocking, T., Rigaill, G. and Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Statistics and Computing* **27** 519–533.

Munkres, J. (2000). *Topology*. Featured Titles for Topology, Prentice Hall, Incorporated.

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

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527.

Peng, T., Leckie, C. and Ramamohanarao, K. (2004). Proactively detecting distributed denial of service attacks using source ip address monitoring. In *International conference on research in networking*. Springer.

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

Raskutti, G., Wainwright, M. J. and Yu, B. (2011). Minimax rates of estimation for high-dimensional linear regression over $\ell_q$ -balls. *IEEE Transactions on Information Theory* **57** 6976–6994.

Rigaill, G. (2015). A pruned dynamic programming algorithm to recover the best segmentations with 1 to $k_{max}$ change-points. *Journal de la société française de statistique* **156** 180–205.

Rinaldo, A., Wang, D., Wen, Q., Willett, R. and Yu, Y. (2021). Localizing changes in high-dimensional regression models. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, vol. 130 of *Proceedings of Machine Learning Research*. PMLR.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58** 267–288.

Vostrikova, L. J. (1981). Detecting "disorder" in multidimensional random processes. *Soviet Mathematrics Doklady* **24** 55–59.

Wang, D., Yu, Y. and Rinaldo, A. (2020). Univariate mean change point detection: Penalization, CUSUM and optimality. *Electronic Journal of Statistics* **14** 1917 – 1961.

Wang, D., Zhao, Z., Lin, K. Z. and Willett, R. (2021). Statistically and computationally efficient change point localization in regression settings. *The Journal of Machine Learning Research* **22** 11255–11300.

Zhang, B., Geng, J. and Lai, L. (2015). Multiple change-points estimation in linear regression models via sparse group lasso. *IEEE Transactions on Signal Processing* **63** 2209–2224.

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