Received: 21 March 2020 | Revised: 4 September 2020 | Accepted: 16 September 2020

DOI: 10.1002/bimj.202000078

Biometrical Journal

RESEARCH PAPER

Sequential change point detection for high-dimensional data using nonconvex penalized quantile regression

Suthakaran Ratnasingam[1] | **Wei Ning**[2,3]

[1] Department of Mathematics, California State University, San Bernardino, CA, USA

[2] Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, USA

[3] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, P. R. China

Correspondence
Wei Ning, Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA.
Email: wning@bgsu.edu

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

In this paper, a sequential change point detection method is developed to monitor structural change in smoothly clipped absolute deviation (SCAD) penalized quantile regression (SPQR) models. The asymptotic properties of the test statistic are derived from the null and alternative hypotheses. In order to improve the performance of the SPQR method, we propose a post-SCAD penalized quantile regression estimator (P-SPQR) for high-dimensional data. We examined the finite sample properties of the proposed methods via Monte Carlo studies under different scenarios. A real data application is provided to demonstrate the effectiveness of the method.

KEYWORDS

change point detection, high-dimensional, quantile regression, SCAD, sequential analysis

1 | INTRODUCTION

In recent years, quantile regression has been widely used in many areas due to its appealing properties contrast to the traditional ordinary least square (OLS) regression model. The quantile regression method was introduced by Koenker and Bassett (1978) as an alternative to the least square regression. This is considered as an extension of the least absolute deviation (LAD) regression or median regression. Unlike the least squares, quantile regression has been designed to model the changes in the conditional quantiles of the response variable concerning the changes in the covariates . The OLS model examines the importance of predictor X by modeling the conditional expectations of the response variable Y given X while quantile regression models attempt to estimate either the conditional median or other quantiles of the response variable. The quantile regression produces much more information about the conditional response distribution and provides a more robust analysis of data. Unlike the OLS, the quantile regression estimates are not sensitive to outliers in the response variable, see, for example, Davino, Furno, and Vistocco (2013) and Furno and Vistocco (2018). Therefore, the quantile regression can be used when the distribution of random errors is heavy-tailed, or when there are outliers in the samples.

Change point detection for quantile regression has been extensively studied, see, for example, Bai (1996) proposed tests to detect changes in quantile regression parameters as well as changes in variance. In addition, it can be used to detect error heterogeneity in the data. Furno (2007) studied a likelihood ratio test based on quantile regressions. Wang and He (2007) proposed a test for detecting differences in certain quantiles of the intensity distributions. Qu (2008) proposed

different test statistics for structural change occurring in a prespecified quantile or across quantiles. Lagrange multiplier test for structural breaks in quantile regressions was proposed by Furno (2012). Furno and Vistocco (2013) investigated the test by Qu (2008) for structural breaks in quantile regressions. Aue, Cheung, Lee, and Zhong (2014) proposed a new methodology to simultaneously (or separately) detect breakpoints, conduct variable selection, and estimate parameters in quantile regression models. Zhang, Wang, and Zhu (2014) developed a new procedure for testing change points due to a covariate threshold in regression quantiles. Their proposed test was based on the cumulative sum (CUSUM) of the subgradient of the quantile objective function and required fitting the model only under the null hypothesis. Yet, very few studies examine the use of quantile regression in sequential change point analysis. For instance, Zhou, Wang, and Tang (2015) developed a method for sequential detection of structural changes in linear quantile regression models. They established the asymptotic properties of the test statistics. Ciuperca (2017) proposed the test statistic for sequential change point detection in a nonlinear quantile model. Ciuperca (2018) proposed a test statistic based on the adaptive least absolute shrinkage and selection operator (LASSO) quantile method to detect a change in a linear model.

In real-world scenarios, we often deal with a large number of explanatory variables. For example, genomics, finance, and health care data have a large number of explanatory variables for each observation. In this paper, we study the sequential change point method for quantile regression in high-dimensional covariates. Our approach enables us to analyze the conditional distribution of the response variable at different quantile levels with large explanatory variables. To the best of our knowledge, no previous studies investigate the use of a nonconvex penalized quantile regression model in the sequential change point analysis.

This paper is organized as follows. In Section 2, the detection procedures based on smoothly clipped absolute deviation (SCAD) penalized quantile regression (SPQR) and post-SCAD penalized quantile regression estimator (P-SPQR) are proposed to detect changes sequentially under high-dimensional scenarios. Corresponding asymptotic results are established. Simulations with various settings are conducted in Section 3 to investigate the performance of the proposed methods. The proposed P-SPQR method is applied to a breast cancer gene expression data to illustrate the detection and estimation process in Section 4. Some discussion is provided in Section 5.

2 | METHODOLOGY

Let $\tau \in (0,1)$ be fixed and known quantile of interest. Suppose we have a random sample $\{Y_i, x_{i1}, \dots, x_{ip}\}, i = 1, \dots, m$ and a vector of independent identically distributed errors $\mathcal{E} = (\mathcal{E}_1, \dots, \mathcal{E}_m)$. Consider the model,

$$Y_i = \mathrm{x}_i^\top \beta + \mathcal{E}_i, \quad i = 1, \dots, m, \tag{1}$$

where $\mathrm{x}_i = (x_{i1}, \dots, x_{ip})^\top$, $i = 1, \dots, m$ and $\beta = (\beta_{01}, \dots, \beta_{0p})$ is the vector of unknown quantile regression parameters at the τth quantile level. Let $X = (\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_p)^\top$ be the $m \times p$ matrix of covariates, where $\mathrm{x}_1^\top, \dots, \mathrm{x}_m^\top$ are the rows of X and $X = (X_1, \dots, X_p)$ where $X_1, \dots, X_p$ are the columns of X. The $\mathcal{E}_i$ is the error term satisfying $P(\mathcal{E}_i < 0 | \mathrm{x}_i) = \tau$ for $i = 1, \dots, m$. The model (1) can be expressed in a similar manner by specifying the τth conditional quantile as

$$Q_y(\tau | X_i) = \mathrm{x}_i^\top \beta. \tag{2}$$

The quantile coefficient β_τ can be estimated by

$$\hat{\beta}_\tau = \underset{\beta \in \mathbb{R}^P}{\arg\min} \sum_{i=1}^{m} \psi_\tau \left(Y_i - \mathrm{x}_i^\top \beta \right), \tag{3}$$

where

$$\psi_\tau(u) = u(\tau - I(u < 0)) \tag{4}$$

is the quantile loss function introduced by Koenker and Bassett (1978). $I(\cdot)$ is the indicator function. When $\tau = 1/2$, it corresponds to the median regression. However, too many explanatory variables in the model may cause the problem of overfitting. To remedy this issue, one can consider using the penalized quantile regression estimator in (3) as suggested in Koenker, Ng, and Portnoy (1994) and Koenker (2004).

2.1 | SCAD penalized quantile regression

Several penalty functions have been proposed in the literature, including ℓ_2 penalty used in ridge regression by Hoerl and Kennard (1970), Lasso ℓ_1 penalty by Tibshirani (1996). Lasso has some rich properties which include shrinkage of the coefficients toward zero for sufficiently large tuning parameters. However, the Lasso tends to produce biased estimates for large coefficients. The studies conducted by Knight and Fu (2000), Fan and Li (2001), and Zou (2006) revealed that the variable selection in Lasso is consistent under certain conditions, but not in general. Thus, the Lasso does not possess the oracle property.

To overcome this issue, Fan and Li (2001) introduced a nonconvex penalty called smoothly clipped absolute deviation (SCAD) and they suggested that it can be used for robust methods, such as median regression. The SCAD corresponds to a quadratic spline function with knots at λ and $a\lambda$. As mentioned in Fan and Li (2001), the SCAD penalty function satisfies three requirements for variable selection, including asymptotic unbiasedness, sparsity, and continuity of the estimated parameters. It can estimate the zero coefficients as exactly zero with the probability approaching one. Regularized quantile regression with fixed p was studied by Zou and Yuan (2008), Wu and Liu (2009), and Kai, Li, and Zou (2011). Wu and Liu (2009) investigated the nonconvex penalty for penalized quantile regression, including the SCAD penalty for variable selection, and showed that the SPQR satisfies the oracle property. Furthermore, the oracle property of nonconvex (SCAD and minimax concave penalty (MCP)) penalized linear quantile regression is established by Wang, Wu, and Li (2012) under high-dimensional settings. The SPQR model is given as

$$Q(\beta,\tau)=\sum_{i=1}^{m}\psi_\tau\left(Y_i-\mathrm{x}_i^\top\beta\right)^2+\sum_{j=1}^{p}p_{\lambda_m}(|\beta_j|). \tag{5}$$

The SPQR solves the following minimization problem:

$$\hat{\beta}_m^\tau=\underset{\beta\in\mathbb{R}^P}{\arg\min}\left\{\sum_{i=1}^{m}\psi_\tau\left(Y_i-\mathrm{x}_i^\top\beta\right)^2+\sum_{j=1}^{p}p_{\lambda_m}(|\beta_j|)\right\}, \tag{6}$$

where $p_{\lambda_m}(\cdot)$ is the penalty function with tuning parameter $\lambda_m(\geq 0)$. The first derivative of the SCAD penalty function for some $a>2$ and $\beta>0$ is given as follows:

$$p'_{\lambda_m}(\beta)=\lambda_m\left\{I(\beta\leq\lambda_m)+\frac{(a\lambda_m-\beta)_+}{(a-1)\lambda_m}I(\beta>\lambda_m)\right\}. \tag{7}$$

There are two unknown parameters λ_m and a. In practice, the best pair (λ_m, a) can be obtained by using a two-dimensional grids search, for example, a cross-validation method. Fan and Li (2001) suggested $a=3.7$ is a good choice for various problems. In this research, a is set to 3.7 to reduce the computational burden. The tuning parameter in the penalty function controls the amount of shrinkage. The larger the value of λ_m, the greater the amount of shrinkage. Like all other penalized regression procedures, the performance of the penalized quantile regression depends on the selection of a tuning parameter. The tuning parameter selection methods are widely studied. Classical methods including Mallow's C_p (Mallows, 1973), Akaike information criterion (AIC; Akaike, 1974), Bayesian information criterion (BIC; Schwarz, 1978), cross-validation, and generalized cross-validation (Golub, Heath, & Wahba, 1979) have been used for the model selection. In this research, the tuning parameter is selected using cross-validation. Throughout this paper, we let $\beta_0=(\beta_{0,1},\beta_{0,2},\ldots,\beta_{0,p})$ be the true parameter value and it is assumed to be sparse. Let $\mathbb{S}_0=\{\beta_{0,j}\neq 0:j=1,\ldots,p\}$ be the index set of the nonzero coefficients for the true parameter, where $\beta_{0,j}$ is the jth component of the parameter vector β_0. Without loss of generality, we assume that the first q regression coefficients are nonzero and the remaining $(p-q)$ regression coefficients are 0. We denote the SCAD penalized quantile estimate by $\hat{\beta}_m^\tau$. Let $\mathbb{S}_*=\{\hat{\beta}_{m,j}^\tau\neq 0:j=1,\ldots,p\}$ be the index set of the SPQR estimator calculated using the historical sample size m, where $\hat{\beta}_{m,j}^\tau$ is the jth element of the SPQR estimator $\hat{\beta}_m^\tau$.

2.2 | Asymptotic properties

In this section, we establish the asymptotic properties of the proposed test statistic. We rewrite $\mathrm{x}_i^\top = (\mathrm{z}_i^\top, \mathrm{w}_i^\top)$, where $\mathrm{z}_i = (x_{i1}, \ldots, x_{iq})^\top$ and $\mathrm{w}_i = (x_{iq+1}, \ldots, x_{ip})^\top$. We take into account the situation where the covariates are fixed. To prove the asymptotic properties, we impose following conditions:

C1. The model errors $\mathcal{E}_i$ are independent and identically distributed (i.i.d.). Let $f(\cdot)$ and $F(\cdot)$ be the density function and the distribution function of $\mathcal{E}_i$, respectively, and $f(\cdot)$ uniformly bounded away from zero.

C2. For all $n \in N$, there exists a positive definite matrix $\mathbb{C}$ such that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top = \mathbb{C}$. Also, $\lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} \|X_i\|^4 < \infty$ a.s, where $\|\cdot\|$ is the Euclidean norm.

C3. There exists a positive constant $L < \infty$ such that $\max_{1\le i\le n, 1\le j\le p} |x_{ij}| \le L$, for all $n \in N$.

C4. There exist positive constants $K_1 < K_2$ such that $K_1 \le \lambda_{\min}(\frac{1}{n}X_{\mathbb{S}_0}^\top X_{\mathbb{S}_0}) \le \lambda_{\max}(\frac{1}{n}X_{\mathbb{S}_0}^\top X_{\mathbb{S}_0}) \le K_2$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of $\frac{1}{n}X_{\mathbb{S}_0}^\top X_{\mathbb{S}_0}$, respectively. We also assumed that $\max_{1\le i\le n} \|\mathrm{z}_i\| = O_p(q^{1/2})$.

C5. The true model dimension q satisfies, $q = O(n^{c_1})$ for some $0 \le c_1 < 1/2$.

C6. There exists a constant $b > 0$ such that

$$\frac{1}{m^s}\left\| \sum_{i=m+k_m^*+1}^{m+k_m^*+m^s} \mathrm{x}_{i,\mathbb{S}_0} \left\{ F(0) - F\left(\mathrm{x}_{i,\mathbb{S}_1}^\top \beta_{1,\mathbb{S}_1} - \mathrm{x}_{i,\mathbb{S}_0}^\top \beta_{0,\mathbb{S}_0}\right)\right\}\right\| > b,$$

where $k_m^* = O(m^s)$, with the constant s and for open-end procedure $s > 1$ and for closed-end procedure $0 \le s \le 1$.

The (C1), (C2), and (C3) are used in the literature on high-dimensional quantile regression models, see, for example, Koenker (2005), Wu and Liu (2009) and Wang et al. (2012). The condition (C4) is similar to Wang et al. (2012) and it requires the design matrix corresponding to the true underlying model if it is well behaved. The condition (C5) on the true model dimension is considered in Kim, Choi, and Oh (2008), Bühlmann, Kalisch, and Maathuis (2010), and Wang et al. (2012). In particular, (C5) allows dimension growth depending on the sample size. The condition (C6) is used in Ciuperca (2018). If the distribution function $F : \mathcal{B} \longrightarrow [0,1]$ Lipschitz on the set $\mathcal{B}$ and for any compact set on $\mathcal{B}$, there exists $b_1 > 0$ not depending on m such that $f(x) > b_1$, then the condition (C6) holds. In real-life applications, these conditions may not hold. For example, condition (C1) i.i.d. noise assumption may be violated and the condition (C4) could be violated for correlated predictors.

2.3 | Oracle property

Suppose that the conditions are satisfied. If $\lambda_m \longrightarrow 0$ and $(q/m)^{1/2}\lambda_m \longrightarrow \infty$ as $m \longrightarrow \infty$, then the SPQR estimator $\hat{\beta}_m^\tau$ satisfy the oracle property,

1. Sparsity property for $\hat{\beta}_m^\tau$ happens in the historical data. Then,

$$P(\mathbb{S}_* = \mathbb{S}_0) = 1.$$

2. Asymptotic normality: $(q/m)^{1/2}(\hat{\beta}_m^\tau - \beta_0)_{\mathbb{S}_0} \longrightarrow N(0, \tau(1-\tau)\mathbb{C}_{11}^{-1}/f(0)^2)$ in distribution as $m \longrightarrow \infty$ where $\mathbb{C}_{11}$ is top-left $q \times q$ matrix of $\mathbb{C}$.

Suppose that conditions hold, by adopting theorem 2.4 in Wang et al. (2012), we can conclude that the SCAD penalized quantile regression satisfies the oracle requirements for variable selection such as asymptotic unbiasedness, sparsity, and continuity of the penalized estimator.

2.4 | Sequential change point problem

We study the change point problem with known prechange coefficients but unknown postchange parameters. Suppose there exists a historical sample size m and at a given quantile level $\tau \in (0,1)$ such that $\beta_1^\tau = \cdots = \beta_m^\tau = \beta_0$. This is called a noncontamination assumption and used in Zhou et al. (2015) and Ciuperca (2017). Now the prechange parameters are obtained using the historical sample data. Let T_m be the monitoring horizon. After the historical sample size of m, we are interested in monitoring the process sequentially. The regression model after the historical observations m is,

$$Y_i = \mathrm{x}_i^\top \beta + \mathcal{E}_i, \quad i = m+1, m+2, \dots . \tag{8}$$

At each time point i, our goal is to test whether we have the same model as the first m observations. Under the null hypothesis, there is no change in the parameters,

$$H_0 : \beta_i^\tau = \beta_0 ; \quad \text{for } i = m+1, m+2, \dots .$$

Under the alternative hypothesis, we consider at an unknown time point k the parameters changing from β_0 to β_1. There exists $k \geq 1$ such that,

$$H_1 : \begin{cases} \beta_i^\tau = \beta_0 ; & i = m+1, m+2, \dots, m+k, \\ \beta_i^\tau = \beta_1 ; & i = m+k+1, \dots, m+T_m \quad \textit{and } \beta_0 \neq \beta_1, \end{cases}$$

where $\beta_1 = (\beta_{1,1}, \beta_{1,2}, \dots, \beta_{1,p})$ and it is unknown. Let $\mathbb{S}_1 = \{\beta_{1,j} \neq 0 : j = 1, \dots, p\}$ be the index set of the nonzero coefficients under the alternative hypothesis. Following Horváth, Huškova, Kokoszká, and Steinebach (2004) and Zhou et al. (2015), our monitoring process can be defined based on the following CUSUM type process:

$$S(m,k) = J_{m,\mathbb{S}_*}^{-1/2} \sum_{i=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}_*} \psi_\tau \left(Y_i - \mathrm{x}_{i,\mathbb{S}_*}^\top \hat{\beta}_{m,\mathbb{S}_*}^\tau \right), \quad k = 1, \dots, T_m, \tag{9}$$

where $J_{m,\mathbb{S}_*} = \tau(1-\tau) D_{m,\mathbb{S}_*}$ with $D_{m,\mathbb{S}_*} = \frac{1}{m} \sum_{i=1}^m \mathrm{x}_{i,\mathbb{S}_*} \mathrm{x}_{i,\mathbb{S}_*}^\top$ and $\psi_\tau(u) = \tau - I(u < 0)$. The proposed CUSUM-based test statistic for the monitoring process of the SCAD penalized quantile regression is given as

$$\Omega(m,k,\gamma) = \frac{\|S(m,k)\|_\infty}{g(m,k,\gamma)}, \tag{10}$$

where $g(m,k,\gamma)$ is called the normalizing function and defined as

$$g(m,k,\gamma) = m^{1/2} \left(1 + \frac{k}{m}\right) \left(\frac{k}{k+m}\right)^\gamma. \tag{11}$$

The γ is called the control parameter. The choice of γ plays an important role in the monitoring process. The monitoring process stops immediately for large control parameter $\gamma \in [0, 1/2)$. In particular, a larger value of γ is preferable if the structural change happens soon after the historical sample size m. The closed- and open-end procedures are discussed in Hušková and Kirch (2012) and Zhou et al. (2015). Under the open-end procedure, the monitoring process continues possibly to infinity if no change point is detected. The open-end procedure is however not realistic in many situations. We call a procedure closed-end when the monitoring process is stopped after a finite number of observations even if no change is detected. Stopping time of the monitoring process based on the open-end procedure is defined as

$$\Lambda(k) = \begin{cases} \inf\{k \geq 1; & S(m,k)/g(m,k,\gamma) \geq c_\alpha(\gamma)\}, \\ \infty & \textit{forall } k = 1, 2, \dots, \end{cases} \tag{12}$$

where $c_\alpha(\gamma)$ is the critical value. Suppose $T_m < \infty$ with $\lim_{m\to\infty} T_m/m = N(> 0)$ with the possibility $N = \infty$. Thus, the open-end procedure has the monitoring boundary $T_m = \infty$. The critical value $c_\alpha(\gamma)$ is satisfying, under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda(k) < \infty) = \alpha, \tag{13}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda(k) < \infty) = 1. \tag{14}$$

Suppose $T_m < \infty$, with $T_m/m \longrightarrow N$ $(N > 0)$ is defined as closed-end procedure and the stopping time is

$$\Lambda^*(k) = \begin{cases} \inf\{k \geq 1; & S(m,k)/g(m,k,\gamma) \geq c^*_\alpha(\gamma)\}, \\ T_m & forall\ k = 1, \ldots, T_m, \end{cases} \tag{15}$$

where $c^*_\alpha(\gamma)$ is the critical value satisfying, under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(k) < \infty) = \alpha, \tag{16}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(k) < \infty) = 1. \tag{17}$$

Theorem 1. *Under the assumptions (C1) to (C6) and for a given constant value of $\gamma \in [0, 1/2)$, if the null hypothesis holds, we have*

1. *For the open-end procedure,*

$$\lim_{m\to\infty} P\left(\sup_{1\leq k\leq T_m} \frac{S(m,k)}{g(m,k,\gamma)} \leq c_\alpha(\gamma)\right) = P\left(\sup_{0\leq t\leq 1} \frac{\|W(t)\|_\infty}{t^\gamma} \leq c_\alpha(\gamma)\right).$$

2. *For the closed-end procedure,*

$$\lim_{m\to\infty} P\left(\sup_{1\leq k\leq T_m} \frac{S(m,k)}{g(m,k,\gamma)} \leq c^*_\alpha(\gamma)\right) = P\left(\sup_{0\leq t\leq N/(N+1)} \frac{\|W(t)\|_\infty}{t^\gamma} \leq c^*_\alpha(\gamma)\right),$$

where $\{W(t), 0 \leq t < \infty\}$ denotes a ℓ-dimensional Wiener process, where, ℓ is the number of significant features in the model based on the historical data, $\alpha \in (0, 1)$, and the control parameter $0 \leq \gamma < 1/2$.

The asymptotic critical values for the proposed test statistic can be obtained based on the Theorem 1 through simulation. First, we generate a sequence of i.i.d $\ell-$dimensional random vector $e_i = (e_{i1}, e_{i2}, \ldots, e_{i\ell})$, where $e_{ij} \sim N(0, 1)$, $j = 1, \ldots, \ell$. Define $W^*(t) = M^{-1/2} \sum_{i=1}^{tM} e_i$, where M is a grid of 10,000. In each iteration, we calculate the test statistic $\max \|W^*(t)/t^\gamma\|_\infty$ for both closed- and open-end procedures obtained over $t \in (0, 1)$ and $t \in (0, N/(N + 1))$, respectively. The critical value for a level-α test can be estimated by the $(1 - \alpha)$th quantile of the test statistics.

Theorem 2. *Under the assumptions and for a given constant value of $\gamma \in [0, 1/2)$, if the alternative hypothesis holds, we have,*

$$\sup_{1\leq k\leq T_m} \frac{S(m,k)}{g(m,k,\gamma)} \longrightarrow \infty \quad \text{as} \quad m \longrightarrow \infty.$$

Proofs are given in the supplementary web material.

2.5 | Post-SCAD penalized quantile regression

In high-dimensional settings, the quantile regression model with SCAD penalized estimator is asymptotically unbiased. Huang and Xie (2007) showed that under appropriate conditions, the SPQR is consistent for variable selection. However, the direct use of this theorem will induce bias. To reduce the bias in the estimator, Belloni and Chernozhukov (2013) suggested the so-called post-Lasso estimator. They showed that the OLS post-Lasso estimator performs at least as well as the lasso under mild additional assumptions. As discussed earlier, the SPQR enjoys the oracle property thus, the P-SPQR estimator becomes the oracle estimator as well. To improve the monitoring method, a modified test statistic based on the P-SPQR estimator is proposed. The variable selection procedure plays an important role in a high-dimensional data set. In the first step, we select the important features by regularizing quantile regression with a SCAD penalty function. Using the significant predictors, let $\hat{\beta}^*_{\tau,m}$ be the quantile coefficient estimator based on the historical data and can be obtained by minimizing

$$\hat{\beta}^*_{\tau,m} = \arg\min_{\beta\in\mathbb{R}^{\ell'}} \sum_{i=1}^{m} \psi_\tau\left(Y_i - \mathrm{x}^\top_{i,\mathbb{S}^*}\beta\right), \tag{18}$$

where $\psi_\tau(\cdot)$ defined in (4) and ℓ' is the cardinality of the set $\mathbb{S}^*$. Following Horváth et al. (2004) and Zhou et al. (2015) , the subgradient-based CUSUM-type process,

$$S^*(m,k) = J^{-1/2}_{m,\mathbb{S}^*} \sum_{i=m+1}^{m+k} \mathrm{x}_{i,\mathbb{S}^*}\psi_\tau\left(Y_i - \mathrm{x}^\top_{i,\mathbb{S}^*}\hat{\beta}^*_{\tau,m}\right), \quad k = 1,\dots,T_m, \tag{19}$$

where $J_{m,\mathbb{S}^*} = \tau(1-\tau)D_{m,\mathbb{S}^*}$ with $D_{m,\mathbb{S}^*} = \frac{1}{m}\sum_{i=1}^{m}\mathrm{x}_{i,\mathbb{S}^*}\mathrm{x}^\top_{i,\mathbb{S}^*}$ and $\psi_\tau(u) = \tau - I(u<0)$. The modified CUSUM-based test statistic for the monitoring process in P-SPQR model is given as

$$\Omega^*(m,k,\gamma) = \frac{\|S^*(m,k)\|_\infty}{g(m,k,\gamma)}. \tag{20}$$

Stopping time for the open-end procedure,

$$\Lambda_{\text{modified}}(k) = \begin{Bmatrix} \inf\{k\geq 1; & S^*(m,k)/g(m,k,\gamma) \geq c_\alpha(\gamma),\}, \\ \infty & \textit{forall } k = 1,2,\dots, \end{Bmatrix} \tag{21}$$

and for the closed-end procedure,

$$\Lambda^*_{\text{modified}}(k) = \begin{Bmatrix} \inf\{k\geq 1; & S^*(m,k)/g(m,k,\gamma) \geq c^*_\alpha(\gamma),\}, \\ T_m & \textit{forall } k = 1,\dots,T_m, \end{Bmatrix} \tag{22}$$

where $c_\alpha(\gamma)$ and $c^*_\alpha(\gamma)$ are asymptotic the critical values for closed- and open-end procedures, respectively. For a given constant $\gamma \in [0,1/2)$, the $g(m,k,\gamma)$ is called the normalizing function defined in (11). Furthermore, under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda_{\text{modified}}(k) < \infty) = \alpha \quad \text{and} \quad \lim_{m\to\infty} P(\Lambda^*_{\text{modified}}(k) < \infty) = \alpha, \tag{23}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda_{\text{modified}}(k) < \infty) = 1 \quad \text{and} \quad \lim_{m\to\infty} P(\Lambda^*_{\text{modified}}(k) < \infty) = 1. \tag{24}$$

Under H_0, for the closed- and open-end procedure the test statistics given in (21) and (22) converges in distribution to $\sup_{0\leq t<1}\frac{\|W(t)\|_\infty}{t^\gamma}$ and $\sup_{0\leq t\leq N/(N+1)}\frac{\|W(t)\|_\infty}{t^\gamma}$, respectively. Under H_1, the test statistics $\Omega^*(m,k,\gamma)$ converges in probability to ∞ as $m \longrightarrow \infty$.

TABLE 1 Type I errors comparison for the closed- and open-end procedures of the SPQR and P-SPQR methods in Case 1 for various values of γ, the nominal significance level $\alpha = .05$ and $\tau = 0.5$

Method			Closed-end			Open-end		
	m	N/γ	0	0.25	0.45	0	0.25	0.45
SPQR	75	2	0.044	0.050	0.046	0.008	0.020	0.040
		4	0.057	0.063	0.062	0.030	0.038	0.058
		6	0.047	0.051	0.060	0.030	0.039	0.056
		9	0.052	0.052	0.054	0.037	0.043	0.051
	100	2	0.052	0.053	0.050	0.010	0.026	0.040
		4	0.044	0.045	0.047	0.021	0.032	0.042
		6	0.038	0.038	0.038	0.022	0.027	0.037
		9	0.044	0.046	0.054	0.034	0.038	0.052
	150	2	0.040	0.038	0.046	0.008	0.021	0.038
		4	0.037	0.042	0.044	0.019	0.027	0.040
		6	0.043	0.044	0.041	0.028	0.035	0.040
		9	0.037	0.039	0.042	0.025	0.035	0.042
P-SPQR	75	2	0.058	0.060	0.052	0.012	0.030	0.043
		4	0.073	0.075	0.062	0.033	0.051	0.057
		6	0.070	0.065	0.056	0.043	0.054	0.052
		9	0.062	0.064	0.058	0.049	0.050	0.056
	100	2	0.064	0.065	0.050	0.016	0.034	0.045
		4	0.072	0.070	0.059	0.032	0.046	0.056
		6	0.067	0.064	0.054	0.040	0.048	0.052
		9	0.075	0.074	0.062	0.056	0.060	0.061
	150	2	0.058	0.062	0.058	0.015	0.029	0.049
		4	0.056	0.056	0.053	0.026	0.040	0.045
		6	0.068	0.068	0.048	0.042	0.051	0.045
		9	0.058	0.062	0.056	0.044	0.049	0.054

3 | SIMULATION STUDY

In this section, we conduct Monte Carlo simulations to investigate the performance of the proposed method. To evaluate how well the proposed method performs, we consider three criteria that are commonly used to determine the goodness of a sequential change point detection procedure. They are

1. Type I error rate: Close to the nominal level;
2. Power of the test: Preferably close to 1;
3. Detection time under the alternative hypothesis: Stop as soon as possible after a change begin noticed.

First, we evaluate the Type I errors of the proposed test. Under the null hypothesis, the data are obtained from the model,

$$y_i = \mathrm{x}_i^\top \beta_0 + \mathcal{E}_i, \quad i = 1, \ldots, m + T_m. \tag{25}$$

The following cases are considered to calculate the Type I errors. In all cases, the true parameter vector $\beta_0 \in \{1, 0, -1, 0, -15, 0, 0, 0, 0, 0\}$.

- Case 1: Homoskedastic errors
  - $X_i \sim N(0, 1)$ for all $i \in \{1, \ldots, 10\}$, and $\mathcal{E}_i \sim N(0, 1)$.
- Case 2: Heavy-tailed errors

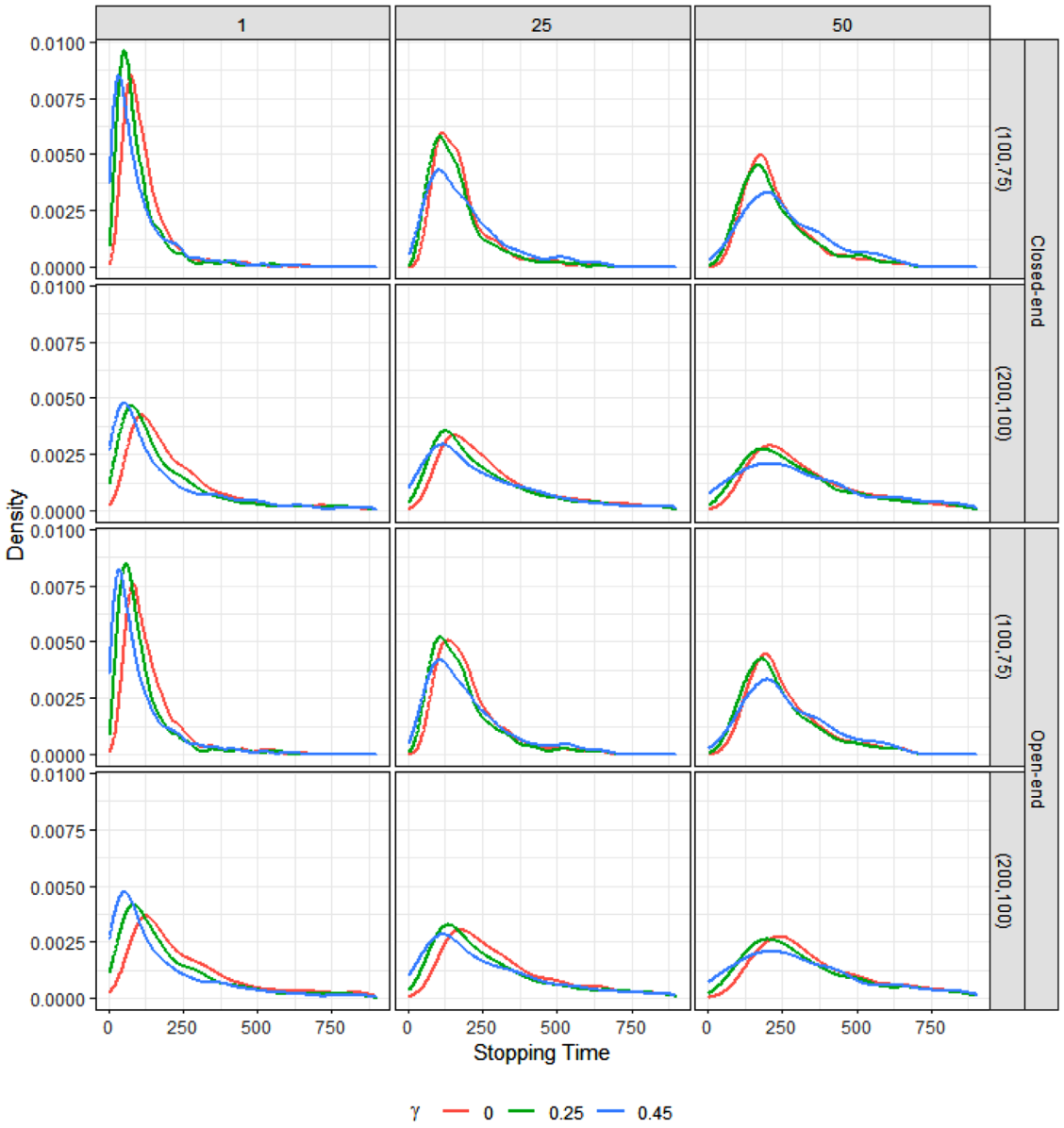

FIGURE 1 Type I errors comparison for the closed- and open-end procedures in Case 1 for the SPQR and P-SPQR methods

- $X_i \sim \text{Unif}(0, 2)$ for all $i \in \{1, \dots, 10\}$, and $\mathcal{E}_i \sim \text{Cauchy}(0, 2)$.

- Case 3: Heteroskedastic errors
  - $X_i \sim \text{Unif}(0, 2)$ for all $i \in \{1, \dots, 10\}$, $\mathcal{E}_i \sim N(0, h(z_i))$, where $h(z) = 1 + 0.2 * z$, $z = 1, \dots, m + T_m$.
- Case 4: Skewed errors
  - $X_i \sim \text{Unif}(0, 2)$ for all $i \in \{1, \dots, 10\}$, and $\mathcal{E}_i \sim SN(0, 1, 3)$,
    where the probability distribution function of a skew normal random variable X is given by

$$f_X(x) = \frac{2}{\sigma}\phi\Big(\frac{x-\mu}{\sigma}\Big)\Phi\Big(\lambda\frac{x-\mu}{\sigma}\Big), \quad x \in \mathbb{R},$$

    where ϕ and Φ are the probability distribution function and cumulative distribution function of the standard normal distribution. We denote $X \sim SN(\mu, \sigma, \lambda)$.

Next, we conduct the power analysis to illustrate the performance of the proposed test statistic. Under the null hypothesis, the true parameter vector $\beta_0 \in \{-1, 0, 1, 8, 1, 0, 0, 0, -5, 0\}$ and under the alternative hypothesis, the parameter vector

TABLE 2 Type I errors comparisons for the closed- and open-end procedures of the SPQR and P-SPQR methods in Case 2 for various values of γ, the nominal significance level $\alpha = .05$ and $\tau = \{0.5, 0.7\}$

				Closed-end			Open-end		
τ	**Method**	m	N/γ	**0**	**0.25**	**0.45**	**0**	**0.25**	**0.45**
0.5	SPQR	200	2	0.170	0.165	0.120	0.067	0.096	0.107
			4	0.229	0.213	0.163	0.160	0.174	0.154
			6	0.203	0.203	0.152	0.156	0.174	0.145
			9	0.243	0.226	0.165	0.210	0.203	0.163
		400	2	0.362	0.341	0.243	0.175	0.227	0.224
			4	0.396	0.377	0.289	0.285	0.320	0.279
			6	0.437	0.404	0.318	0.348	0.364	0.309
			9	0.442	0.412	0.331	0.393	0.390	0.327
	P-SPQR	200	2	0.066	0.066	0.042	0.005	0.023	0.032
			4	0.070	0.071	0.074	0.031	0.047	0.069
			6	0.059	0.057	0.042	0.033	0.042	0.042
			9	0.064	0.056	0.049	0.041	0.049	0.047
		400	2	0.061	0.061	0.047	0.009	0.027	0.040
			4	0.065	0.063	0.048	0.031	0.045	0.037
			6	0.073	0.067	0.060	0.047	0.053	0.053
			9	0.077	0.069	0.050	0.054	0.061	0.050
0.7	SPQR	200	2	0.386	0.358	0.295	0.191	0.264	0.269
			4	0.486	0.463	0.369	0.365	0.402	0.356
			6	0.502	0.482	0.403	0.421	0.442	0.394
			9	0.529	0.505	0.416	0.482	0.482	0.405
		400	2	0.721	0.705	0.587	0.493	0.590	0.560
			4	0.792	0.776	0.690	0.719	0.741	0.679
			6	0.824	0.805	0.716	0.763	0.772	0.710
			9	0.850	0.832	0.737	0.809	0.813	0.731
	P-SPQR	200	2	0.053	0.052	0.051	0.014	0.025	0.038
			4	0.058	0.058	0.062	0.026	0.039	0.054
			6	0.048	0.048	0.062	0.029	0.040	0.059
			9	0.063	0.060	0.071	0.046	0.054	0.071
		400	2	0.058	0.062	0.068	0.018	0.032	0.055
			4	0.057	0.055	0.060	0.028	0.040	0.058
			6	0.042	0.044	0.054	0.027	0.031	0.051
			9	0.060	0.059	0.058	0.039	0.047	0.054

$\beta_1 \in \{0, -1, 0, 2, 0, 0, 1, 0, 0, -1\}$. We consider the two different distributions of the explanatory variables $X_1, X_2, \ldots, X_{10}$. Under H_0, X_i for all $i \in \{1, \ldots, 10\}\backslash\{3, 4, 5\}$ have uniform distribution Unif$(0, 1)$ and $X_3 \sim N(2, 1), X_4 \sim N(4, 1)$ and $X_5 \sim N(5, 1)$. The second distribution for the ith explanatory variable is $X_i + 0.8$ for all $i \in \{1, \ldots, 10\}$. Under the null hypothesis, the model errors $\mathcal{E}_i \sim N(0, 1)$, and under the alternative hypothesis

- Case 1: $\mathcal{E}_i \sim N(0, 1)$.
- Case 2: $\mathcal{E}_i \sim$ Cauchy$(0, 2)$.
- Case 3: $\mathcal{E}_i \sim N(0, h(z_i))$, where $h(z_i) = 1 + 0.2 * z_i\,, \;\; z_i = 1, \ldots, m + T_m$.
- Case 4: $\mathcal{E}_i \sim SN(0, 1, 3)$.

TABLE 3 Type I errors comparisons for the closed- and open-end procedures of the SPQR and P-SPQR methods in Case 3 for various values of γ, the nominal significance level $\alpha = .05$ and $\tau = \{0.5, 0.7\}$

				Closed-end			Open-end		
τ	**Method**	m	N/γ	**0**	**0.25**	**0.45**	**0**	**0.25**	**0.45**
0.5	SPQR	200	2	0.221	0.209	0.165	0.150	0.199	0.193
			4	0.292	0.267	0.202	0.281	0.302	0.266
			6	0.273	0.265	0.210	0.286	0.293	0.251
			9	0.330	0.312	0.212	0.335	0.334	0.271
		400	2	0.460	0.437	0.349	0.342	0.428	0.396
			4	0.529	0.504	0.417	0.496	0.529	0.483
			6	0.560	0.535	0.428	0.564	0.567	0.496
			9	0.570	0.538	0.452	0.609	0.602	0.515
	P-SPQR	200	2	0.047	0.041	0.033	0.006	0.020	0.032
			4	0.051	0.045	0.040	0.029	0.036	0.042
			6	0.058	0.056	0.047	0.036	0.050	0.049
			9	0.060	0.054	0.040	0.046	0.057	0.050
		400	2	0.056	0.053	0.042	0.007	0.026	0.043
			4	0.054	0.052	0.049	0.030	0.038	0.045
			6	0.068	0.064	0.052	0.048	0.055	0.057
			9	0.068	0.072	0.046	0.041	0.048	0.039
0.7	SPQR	200	2	0.491	0.479	0.366	0.334	0.411	0.408
			4	0.581	0.560	0.479	0.545	0.577	0.531
			6	0.580	0.554	0.457	0.584	0.599	0.545
			9	0.629	0.602	0.514	0.648	0.643	0.570
		400	2	0.823	0.810	0.720	0.692	0.751	0.741
			4	0.873	0.857	0.814	0.837	0.851	0.821
			6	0.914	0.898	0.845	0.887	0.892	0.857
			9	0.914	0.906	0.852	0.907	0.909	0.871
	P-SPQR	200	2	0.051	0.049	0.053	0.013	0.023	0.051
			4	0.057	0.055	0.055	0.029	0.038	0.051
			6	0.047	0.043	0.062	0.037	0.046	0.053
			9	0.044	0.042	0.053	0.036	0.045	0.052
		400	2	0.042	0.048	0.053	0.012	0.022	0.055
			4	0.046	0.052	0.056	0.029	0.042	0.062
			6	0.057	0.051	0.061	0.031	0.048	0.060
			9	0.047	0.050	0.051	0.041	0.050	0.058

The second setting is used to compute the stopping time at different change point locations. For both power and stopping time analysis, the data are generated in the following way:

$$y_i = \begin{cases} x_i^\top \beta_0 + \mathcal{E}_i, & i = 1, \ldots, m + k^* - 1, \\ (x_i + \delta)^\top \beta_1 + \mathcal{E}_i, & i = m + k^*, \ldots, m + T_m \text{ and } \delta > 0. \end{cases} \tag{26}$$

First, we evaluate the effect on the detection procedure of various control parameter values, considering $\gamma \in \{0, 0.25, 0.45\}$. Type I errors of the open- and closed-end procedures in Case 1 at quantile level $\tau = 0.5$ are summarized in Table 1 and these results are graphed in Figure 1. The open-end procedure gives slightly deflated Type I errors and tends to increase as N increases. In general, under the open-end procedure, the P-SPQR method provides better Type I errors than the SPQR method. Besides, under the closed-end procedure, for all three γ values the SPQR method provides Type I errors close to the nominal level, however, the P-SPQR method gives slightly inflated Type I errors. In practice, we recommend the

TABLE 4 Type I errors comparison for the closed- and open-end procedures of the SPQR and P-SPQR methods in Case 4 for various values of γ, the nominal significance level $\alpha = .05$ and $\tau = \{0.5, 0.7\}$

				Closed-end			Open-end		
τ	Method	m	N/γ	0	0.25	0.45	0	0.25	0.45
0.5	SPQR	200	2	0.327	0.314	0.245	0.177	0.230	0.225
			4	0.416	0.402	0.317	0.332	0.355	0.307
			6	0.401	0.387	0.309	0.346	0.347	0.301
			9	0.444	0.421	0.331	0.400	0.389	0.326
		400	2	0.620	0.590	0.497	0.418	0.508	0.480
			4	0.673	0.651	0.569	0.579	0.608	0.554
			6	0.709	0.687	0.597	0.651	0.658	0.592
			9	0.727	0.705	0.615	0.695	0.686	0.610
	P-SPQR	200	2	0.039	0.045	0.035	0.005	0.016	0.029
			4	0.072	0.071	0.053	0.035	0.047	0.049
			6	0.076	0.069	0.059	0.045	0.055	0.057
			9	0.061	0.059	0.046	0.041	0.051	0.046
		400	2	0.059	0.048	0.045	0.010	0.021	0.035
			4	0.052	0.054	0.047	0.029	0.035	0.041
			6	0.059	0.060	0.049	0.037	0.048	0.047
			9	0.050	0.051	0.043	0.040	0.047	0.040
0.7	SPQR	200	2	0.594	0.571	0.497	0.401	0.470	0.470
			4	0.694	0.680	0.589	0.600	0.624	0.572
			6	0.724	0.712	0.626	0.663	0.674	0.617
			9	0.758	0.738	0.659	0.724	0.720	0.656
		400	2	0.868	0.854	0.802	0.742	0.797	0.786
			4	0.923	0.915	0.878	0.891	0.897	0.868
			6	0.941	0.931	0.898	0.920	0.921	0.895
			9	0.949	0.941	0.902	0.935	0.931	0.900
	P-SPQR	200	2	0.047	0.047	0.047	0.012	0.022	0.043
			4	0.055	0.058	0.059	0.028	0.039	0.054
			6	0.056	0.057	0.073	0.035	0.042	0.070
			9	0.054	0.061	0.055	0.041	0.049	0.051
		400	2	0.048	0.046	0.063	0.012	0.022	0.055
			4	0.053	0.056	0.060	0.028	0.040	0.053
			6	0.053	0.053	0.058	0.033	0.039	0.054
			9	0.056	0.056	0.059	0.039	0.047	0.057

closed-end procedure for small $N(< 6)$, and large $N(\geq 6)$ both closed- and open-end procedures work in the same manner. This supports previous findings in linear quantile regression models by Zhou et al. (2015).

In Cases 2–4, we consider the different types of error distributions including heavy tails, heteroskedastic, and skewed errors. We compute the Type I errors for both the closed- and open-end procedures at quantile levels $\tau \in \{0.5, 0.7\}$ with the historical sample sizes $m \in \{200, 400\}$. The results are summarized in Tables 2–4, and they are compared in Figure 2. The black dashed horizontal lines indicate the nominal level, $\alpha = .05$. It can be seen that the SPQR method gives larger Type I errors in all cases compared to the P-SPQR method. This could be due to the fact that the P-SPQR method removes penalization bias. Indeed, we notice that the Type I errors of the P-SPQR method are close to the nominal level for the closed-end procedure, and produces slightly deflated Type I errors for the open-end procedure especially for small control parameter values close to zero. In particular, for the P-SPQR method, when N increases the Type I errors for the open-end procedure continue to increase and become close to the nominal level in all cases. Thus, we recommend the P-SPQR method for cases such as heavy tails, heteroskedastic, and skewed error distributions.

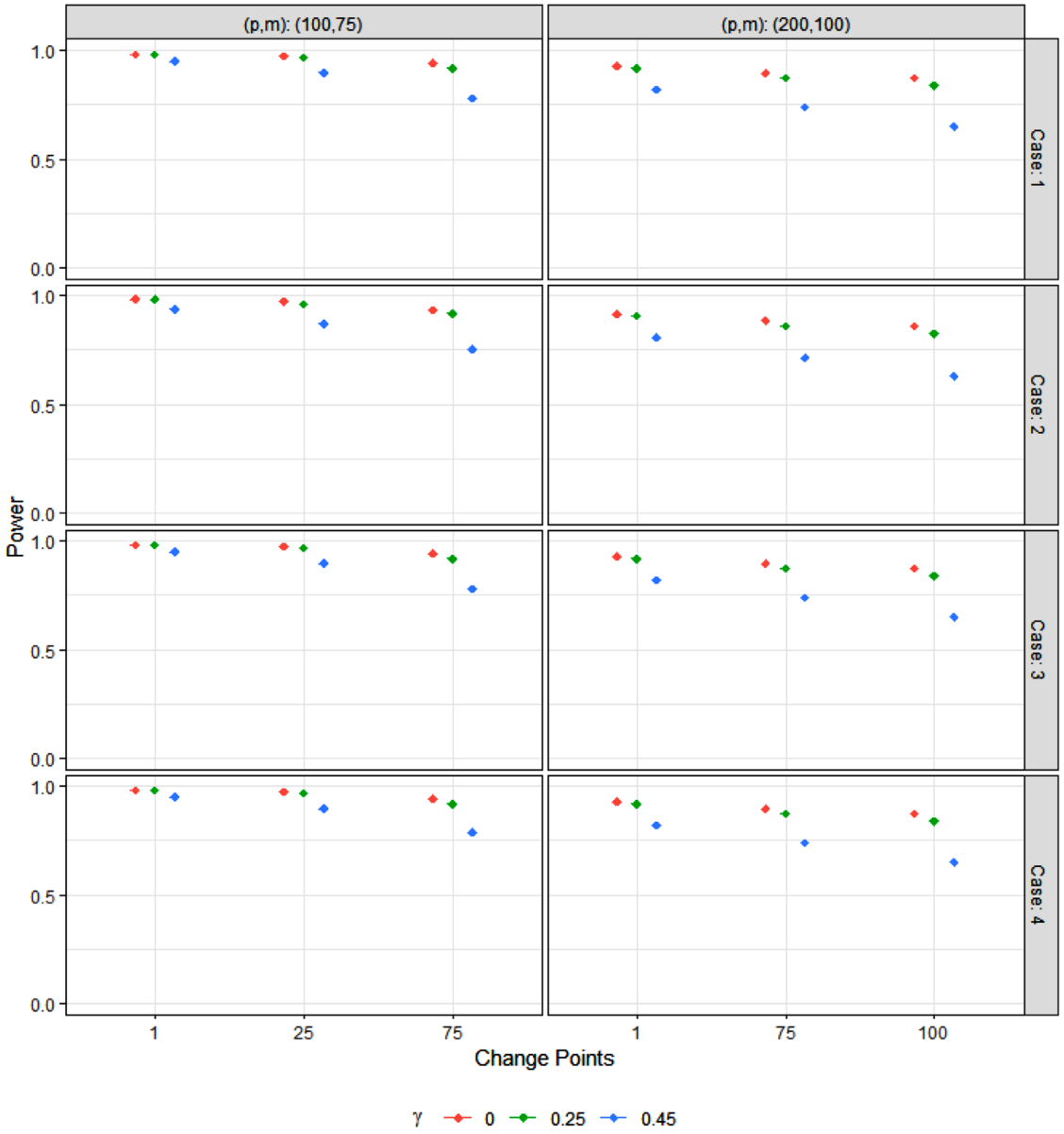

FIGURE 2 Type I errors comparison for the closed- and open-end procedures in Cases 2–4 for P-SPQR and SPQR methods

We conduct a power analysis for both SPQR and P-SPQR methods. The results are shown in Table 5. Figure 3 compares the power of both procedures. In all cases, we observe that the P-SPQR method gives better power compared to the SPQR method. Additional simulations have been carried out to evaluate power for different quantile levels such as $\tau = 0.7, 0.8$. The results are not reported here in detail. Interestingly, the P-SPQR method produces significantly higher power in all these situations. It is important to note that regardless of method, the power tends to be lower when the change location is farther away from the historical sample. Our findings confirm that the P-SPQR method performs better than the SPQR method.

We compute a five-number summary of the stopping time for the closed- and open-end procedures. The results are summarized in Table 6. In all cases, the processes are monitored from $m + 1$ until time $9m$ observations. It can be seen that the control parameter values close to 0.5 have the shortest detection delay time. In contrast, a smaller value of γ takes a longer time to detect the structural change. In cases where structural changes occur immediately after the monitoring scheme begins, larger values of γ would be preferred. Furthermore, smaller values of γ would be preferred, when the change location is farther away from the historical sample. Next, we compare the estimated density of the stopping time for various $\gamma = \{0, 0.25, 0.45\}$ for the SPQR and P-SPQR methods. They are graphed in Figure 4. We observe a slightly heavier tail for larger values of γ, for example, $\gamma = 0.45$ has a slightly heavier tail than that of the monitoring process

TABLE 5 The power comparison for the closed-end procedure with $\tau = 0.5$, $p \in \{100, 300\}$, $k^* \in \{1, 25, 100\}$, $\alpha = .05$, and $\gamma \in \{0, 0.25, 0.45\}$

		$k^* = 1$				$k^* = 25$				$k^* = 100$			
	$m \rightarrow$	100		300		100		300		100		300	
Cases	γ	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR	SPQR	P-SPQR
Case 1	0.00	0.1956	0.8036	0.4380	0.9892	0.1792	0.7768	0.4232	0.9884	0.2008	0.6844	0.3980	0.9884
	0.25	0.1852	0.7840	0.4156	0.9884	0.1792	0.7380	0.4004	0.9880	0.2396	0.6376	0.3704	0.9844
	0.45	0.1332	0.6748	0.3004	0.9812	0.1344	0.6144	0.2740	0.9764	0.1984	0.4944	0.2352	0.9720
Case 2	0.00	0.1588	0.7680	0.3776	0.9864	0.1388	0.7296	0.3676	0.9864	0.1796	0.6332	0.3364	0.9832
	0.25	0.1536	0.7420	0.3516	0.9852	0.1480	0.6964	0.3368	0.9852	0.2208	0.5840	0.3064	0.9784
	0.45	0.1144	0.6248	0.2396	0.9740	0.1264	0.5644	0.2176	0.9716	0.1900	0.4508	0.1944	0.9600
Case 3	0.00	0.1956	0.8004	0.4381	0.9844	0.1808	0.7756	0.4241	0.9844	0.1888	0.6904	0.4017	0.9844
	0.25	0.1824	0.7828	0.4141	0.9832	0.1764	0.7452	0.4037	0.9836	0.2216	0.6396	0.3757	0.9820
	0.45	0.1360	0.6728	0.2992	0.9772	0.1392	0.6172	0.2831	0.9756	0.1856	0.4904	0.2555	0.9680
Case 4	0.00	0.1956	0.8036	0.4380	0.9892	0.1792	0.7768	0.4232	0.9884	0.2008	0.6844	0.3980	0.9884
	0.25	0.1852	0.7840	0.4156	0.9884	0.1792	0.738	0.4004	0.9880	0.2396	0.6376	0.3700	0.9844
	0.45	0.1332	0.6752	0.3004	0.9812	0.1344	0.6144	0.2744	0.9764	0.1984	0.4948	0.2356	0.9720

TABLE 6 Summary statistics of the detection time for both the closed- and open-end procedures in Case I with $\tau = 0.5$, $m = 100$, $\alpha = .05$, and $\gamma \in \{0, 0.25, 0.45\}$ at various change point locations

			SPQR					P-SPQR				
Procedure	k^*	γ	Min	Q1	Med	Q3	Max	Min	Q1	Med	Q3	Max
Open-end	1	0.00	97	232	323	461	896	43	158	296	382	868
		0.25	25	121	196	318	792	23	103	181	353	896
		0.45	1	14	53	229	871	1	63	139	286	900
	25	0.00	126	270	427	536	896	59	208	306	485	895
		0.25	17	74	257	430	755	15	174	267	455	883
		0.45	1	6	16	46	661	1	116	234	393	897
	100	0.00	67	121	160	544	897	64	336	502	658	898
		0.25	17	45	71	108	832	15	297	451	624	899
		0.45	1	7	19	59	114	1	236	396	627	892
Closed-end	1	0.00	62	166	263	381	966	40	131	204	319	866
		0.25	23	107	186	294	894	14	95	165	284	886
		0.45	1	17	56	239	850	1	61	137	279	873
	25	0.00	86	249	373	517	866	52	185	273	433	897
		0.25	16	101	235	441	895	14	152	241	399	882
		0.45	1	6	17	51	648	1	114	230	393	895
	100	0.00	61	122	303	698	822	54	320	488	661	899
		0.25	16	47	77	103	897	14	280	436	621	896
		0.45	1	7	24	64	114	1	231	388	599	883

with $\gamma = 0$. Also, Figure 4 shows that the SPQR method is more likely to falsely raised an alarm even before the actual change occurred.

3.1 | Large p

In high-dimensional settings, we only consider the P-SPQR method. We conduct simulations to study the finite sample properties of the proposed P-SPQR method. A high-dimensional data set with (p, m), considering (100, 75) and (200, 100)

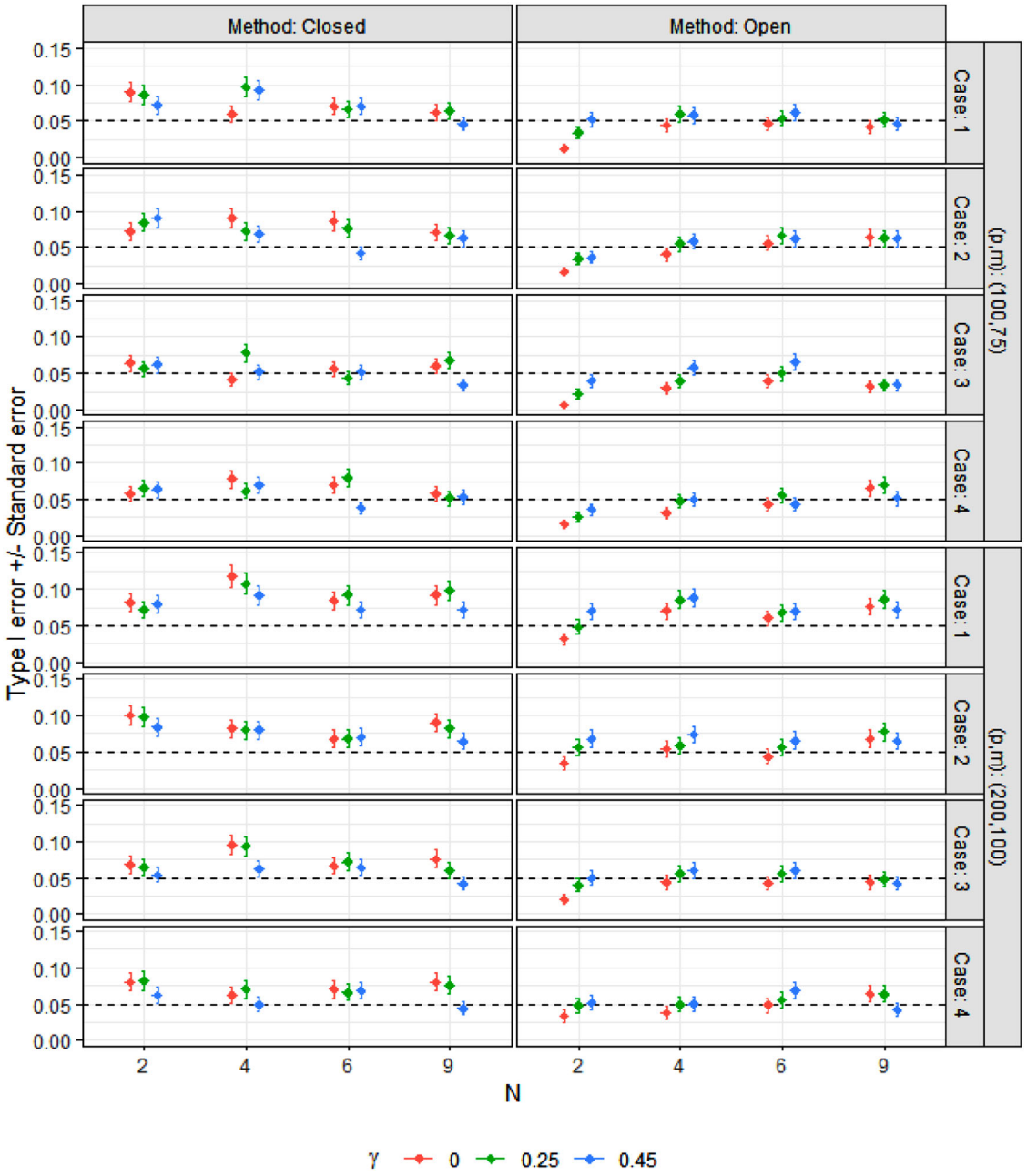

FIGURE 3 Power comparisons for P-SPQR and SPQR methods in Cases 1–4 under the closed-end procedure

are generated. We consider the following two settings. In the first setting, the nonzero components of the true parameters are $\beta_{0,j} = 0$ for $j = \{1, \dots, p\}\backslash\{1, 3, 5, 41, 52\}$ where $p \in \{100, 200\}$. $\beta_{0,1} = 1$, $\beta_{0,3} = 15$, $\beta_{0,5} = -20$, $\beta_{0,41} = -2$, and $\beta_{0,52} = -8$. The explanatory variables $X_i \sim \text{Unif}(0, 1)$ for $i = \{1, \dots, p\}\backslash\{3, 5, 74\}$, where $X_3 \sim N(2, 1)$, $X_5 \sim N(5, 1)$ and $X_{74} \sim N(8, 1)$. The model errors $\mathcal{E}_i$ are generated from the following distributions:

- Case 1: $\mathcal{E}_i \sim N(0, 1)$.
- Case 2: $\mathcal{E}_i \sim \text{Cauchy}(0, 2)$.
- Case 3: $\mathcal{E}_i \sim N(0, h(z_i))$, where $h(z_i) = 1 + 0.2 * z_i$, $z_i = 1, \dots, m + T_m$.
- Case 4: $\mathcal{E}_i \sim SN(0, 1, 3)$.

In the second setting, under H_0, the regression coefficients are $\beta_{0,1} = 1$, $\beta_{0,3} = 15$, $\beta_{0,5} = -2$, $\beta_{0,6} = -13$, $\beta_{0,41} = -2$, $\beta_{0,77} = -8$, and $\beta_{0,j} = 0$ for all $j \in \{1, \dots, p\}\backslash\{1, 3, 5, 8, 41, 77\}$ with $p \in \{100, 200\}$. Under the alternative

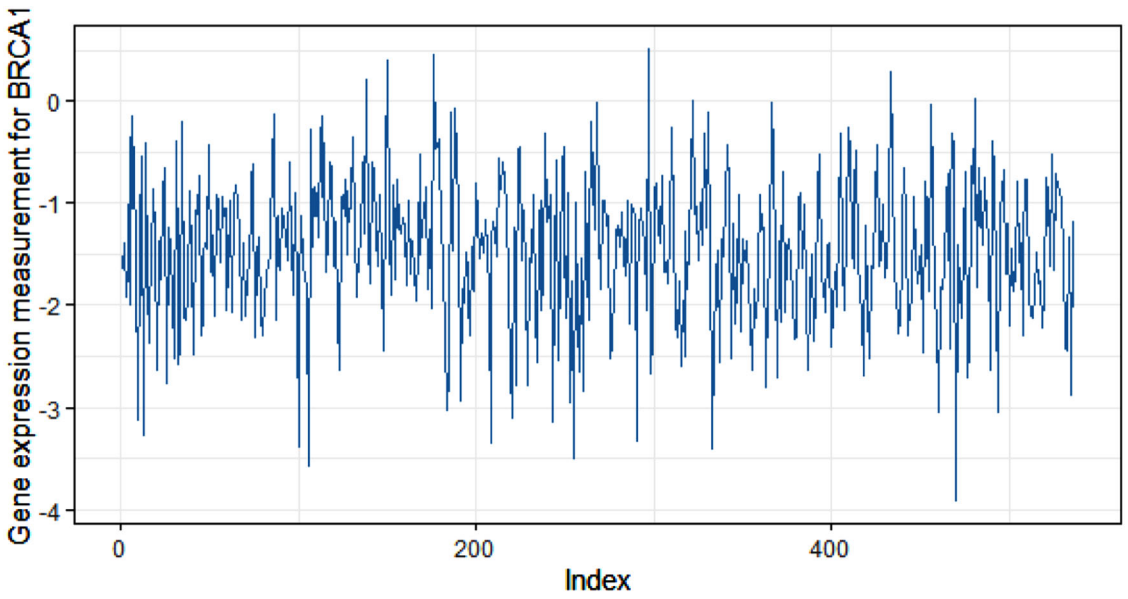

FIGURE 4 Estimated densities of the stopping time for both methods in Case 1 when$m = 100, \tau = 0.5, k^* \in \{1, 25, 100\}$, and$\gamma \in \{0,\ 0.25,\ 0.45\}$

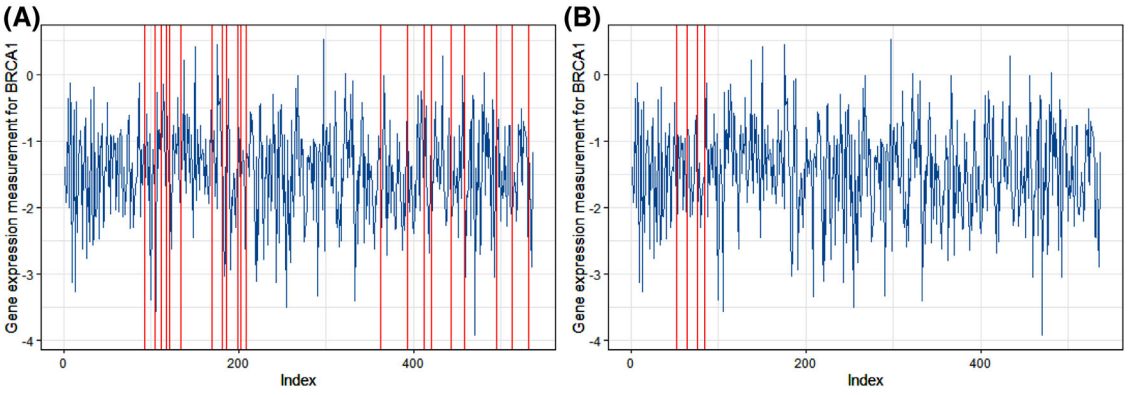

FIGURE 5 Type I errors comparison for the P-SPQR method in Cases 1–4 under the closed- and open-end procedures

hypothesis, the regression coefficients are $\beta_{1,2} = -1,\ \beta_{1,4} = 2,\ \beta_{1,7} = 1,\ \beta_{1,10} = -10.\ \beta_{1,51} = -8,\ \beta_{1,83} = -5$, and $\beta_{1,j} = 0$ for all $j \in \{1, \ldots, p\} \backslash \{2, 4, 7, 10, 51, 83\}$ with $p \in \{100, 200\}$. We consider two different distributions of the explanatory variables. Under the null hypothesis, the explanatory variables $X_i \sim \text{Unif}(0, 1)$. Under the alternative hypothesis, the ith explanatory variable is transformed to $X_i + 5$ and model errors $\mathcal{E}_i \sim N(0, 1)$ for $i \in \{1, \ldots, p\}$.

The first setting is used to calculate Type I errors. The data are generated from (25). Table 7 summarizes the Type I errors for the closed- and open-end procedures, and they are graphed in Figure 5. The various control parameter values, considering $\gamma \in \{0, 0.25, 0.45\}$ and the different size of the historical observations $m \in \{75, 100, 200\}$ are considered. The results are based on 1,000 iterations. Type I errors based on the closed-end procedure are always higher than Type I errors computed from the open-end procedure. For small values of γ, Type I errors of the open-end procedure are comparatively low, however, they are improved as N increases. When the open-end procedure is considered, smaller N provides slightly deflated Type I errors. Thus, in the cases of smaller N, we suggest that the use of larger values of γ close to 0.5. Also, in contrast to the open-end procedure, Type I errors of the closed-end procedure are stable across N in all cases.

The second setting is used to evaluate the power of the test and stopping time calculations. The data are generated from (26). First, we carry out the power analysis of the proposed P-SPQR procedure for high-dimensional data. Different change point locations have been considered under each pair of (p, m). The results are shown in Table 8. Based on Table 8, it is clear that the power tends to decrease as the change point location farther away from the historical sample. Furthermore, in all cases, the monitoring scheme with $\gamma = 0$ has higher power than that of the monitoring process with γ close to 0.5, for

TABLE 7 Type I errors for the P-SPQR method in Cases 1–4 under the closed- and open-end procedures for various values of γ, the nominal significance level $\alpha = .05$, and $\tau = 0.5$

			Closed-end			Open-end		
Cases	(p, m)	N/γ	**0**	**0.25**	**0.45**	**0**	**0.25**	**0.45**
Case 1	(100,75)	2	0.090	0.096	0.070	0.012	0.034	0.052
		4	0.086	0.092	0.062	0.044	0.060	0.058
		6	0.072	0.070	0.064	0.046	0.054	0.062
		9	0.060	0.066	0.046	0.042	0.052	0.046
	(200,100)	2	0.082	0.072	0.080	0.032	0.048	0.070
		4	0.118	0.108	0.092	0.070	0.086	0.088
		6	0.084	0.092	0.072	0.060	0.068	0.070
		9	0.092	0.098	0.072	0.076	0.086	0.072
Case 2	(100,75)	2	0.072	0.072	0.042	0.016	0.034	0.036
		4	0.084	0.068	0.070	0.040	0.054	0.058
		6	0.090	0.086	0.066	0.056	0.066	0.062
		9	0.090	0.076	0.062	0.064	0.062	0.062
	(200,100)	2	0.100	0.098	0.084	0.034	0.056	0.068
		4	0.082	0.080	0.080	0.054	0.058	0.074
		6	0.068	0.068	0.070	0.044	0.056	0.066
		9	0.090	0.082	0.064	0.068	0.078	0.064
Case 3	(100,75)	2	0.064	0.078	0.052	0.006	0.022	0.040
		4	0.056	0.052	0.060	0.030	0.040	0.058
		6	0.062	0.056	0.068	0.040	0.050	0.066
		9	0.042	0.044	0.034	0.032	0.034	0.034
	(200,100)	2	0.068	0.064	0.054	0.020	0.040	0.050
		4	0.096	0.094	0.062	0.044	0.056	0.060
		6	0.066	0.072	0.064	0.042	0.056	0.060
		9	0.076	0.060	0.042	0.044	0.048	0.042
Case 4	(100,75)	2	0.058	0.062	0.038	0.016	0.026	0.036
		4	0.066	0.070	0.058	0.032	0.048	0.050
		6	0.064	0.070	0.052	0.044	0.056	0.044
		9	0.078	0.080	0.054	0.066	0.070	0.052
	(200,100)	2	0.080	0.082	0.062	0.034	0.048	0.052
		4	0.062	0.070	0.050	0.038	0.050	0.050
		6	0.070	0.066	0.068	0.048	0.056	0.068
		9	0.080	0.076	0.044	0.064	0.064	0.042

instance, $\gamma = 0.45$. For different error distributions, the power of the test performs similarly which indicates that the proposed P-SPQR method is robust to the error distribution. Figure 6 provides the comparison for various error distributions and change locations for a given pair of (p, m) for the P-SPQR method.

Next, we compute the five-number summary of the stopping time for the P-SPQR method. The results are recorded in Table 9. Figure 7 shows the estimated density of the stopping time for the P-SPQR method in Case 1. Although in high-dimensional data, we still see that the previous findings for small p continue to hold.

4 | APPLICATION

In this section, we apply the proposed P-SPQR method to breast cancer gene expression data from The Cancer Genome Atlas (TCGA) project. The data contain expression measurements of 17,814 genes from 536 patients. All expression measurements are recorded on the log scale. The response variable y is the gene expression measurement for BRCA1 and

TABLE 8 Power comparisons for P-SPQR method in Cases 1–4 under the closed-end procedure for different pairs of (p, m), $\tau = 0.5$, $\gamma \in \{0, 0.25, 0.45\}$, and various change point locations

	(p, m)	$(100, 75)$			$(200, 100)$		
Cases	γ/k^*	**1**	**25**	**75**	**1**	**50**	**100**
Case 1	0.00	0.980	0.974	0.942	0.928	0.894	0.872
	0.25	0.980	0.966	0.918	0.918	0.872	0.836
	0.45	0.948	0.896	0.778	0.818	0.738	0.650
Case 2	0.00	0.982	0.970	0.934	0.914	0.884	0.858
	0.25	0.982	0.960	0.918	0.906	0.858	0.824
	0.45	0.938	0.870	0.754	0.808	0.714	0.628
Case 3	0.00	0.980	0.974	0.942	0.928	0.894	0.872
	0.25	0.980	0.966	0.918	0.918	0.872	0.836
	0.45	0.948	0.896	0.778	0.818	0.738	0.650
Case 4	0.00	0.980	0.974	0.942	0.928	0.894	0.872
	0.25	0.980	0.966	0.918	0.918	0.872	0.836
	0.45	0.948	0.896	0.778	0.818	0.738	0.650

TABLE 9 Summary statistics of the detection time for the P-SPQR method in Case I with $\tau = 0.5, \alpha = .05, \gamma \in \{0, 0.25, 0.45\}$, and $(p, m) = \{(100, 75), (200, 100)\}$ at various change point locations

			$(100, 75)$					$(200, 100)$				
Procedure	k^*	γ	**Min**	**Q1**	**Med**	**Q3**	**Max**	**Min**	**Q1**	**Med**	**Q3**	**Max**
Open-end	1	0.00	21	78	109	163	645	40	117	185	322	896
		0.25	9	51	79	128	669	11	77	130	244	888
		0.45	4	36	66	130	635	3	48	98	211	896
	25	0.00	43	124	169	235	665	42	154	232	363	895
		0.25	22	104	150	213	672	18	120	184	316	898
		0.45	5	102	161	252	674	3	100	169	289	893
	50	0.00	48	176	231	302	675	47	197	290	431	896
		0.25	22	157	213	292	666	18	157	248	393	899
		0.45	5	169	249	367	674	3	143	254	420	893
Closed-end	1	0.00	21	68	96	137	659	36	105	163	280	896
		0.25	9	46	69	113	597	11	72	123	220	898
		0.45	4	34	63	122	632	3	46	97	204	874
	25	0.00	41	112	151	205	670	33	138	206	327	898
		0.25	22	96	137	198	675	18	114	170	298	900
		0.45	5	100	158	247	664	3	98	166	290	890
	50	0.00	47	156	207	271	641	29	175	263	391	900
		0.25	22	147	201	273	663	18	147	233	379	899
		0.45	5	166	246	359	672	3	140	245	403	897

the explanatory variables are gene expression measurements for remaining genes. The response variable y is graphed in Figure 8. Since the dimensions are very high, we use the univariate filter method for the preprocessing step. This approach selects features according to certain criteria, for example, the correlation coefficient. Pearson correlation coefficient is used to measure the linear relationship between two variables and cannot be used when the association is nonlinear.

Székely, Rizzo, and Bakirov (2007) introduced distance correlation (DC) which is a new measure of dependence between random vectors. Let X and Y be the two random vectors. The characteristic functions of X and Y are denoted as f_X and f_Y, respectively, and let the joint characteristic function of X and Y be $f_{X,Y}$. The distance covariance (dCov) between

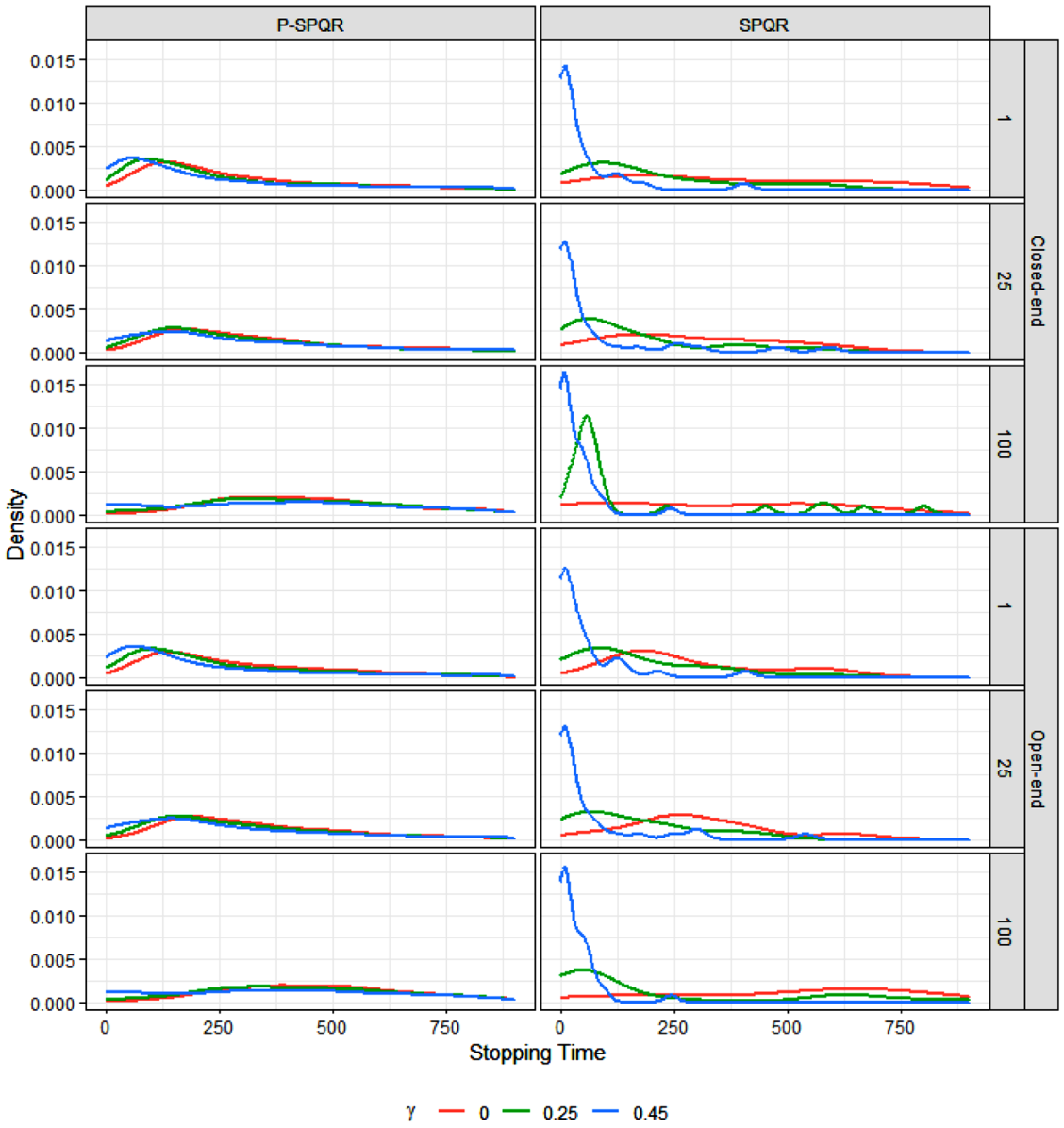

FIGURE 6 Power comparisons for the P-SPQR method in Cases 1-4 under the closed-end procedure

random vectors X and Y with finite first moments is the nonnegative number defined as

$$\mathrm{dCov}^2(X,Y) = \int_{\mathbb{R}^{p+d}} \left\| f_{X,Y}(t,s) - f_X(t) f_Y(s) \right\|^2 w(t,s)\, dt\, ds,$$

where p and d are the dimensions of X and Y, respectively, and $w(t,s)$ is a choice of weight and the weight function chosen to be

$$w(t,s) = \left(c_p c_q \|t\|_p^{1+p} \|s\|_q^{1+q} \right)^{-1},$$

with $c_d = \pi^{(1+d)/2}/\Gamma((1+d)/2)$. The DC between X and Y with finite first moments is defined as

$$\mathrm{DC} = \mathrm{dCorr}(X,Y) = \frac{\mathrm{dCov}(X,Y)}{\sqrt{\mathrm{dCov}(X,X)\,\mathrm{dCov}(Y,Y)}}.$$

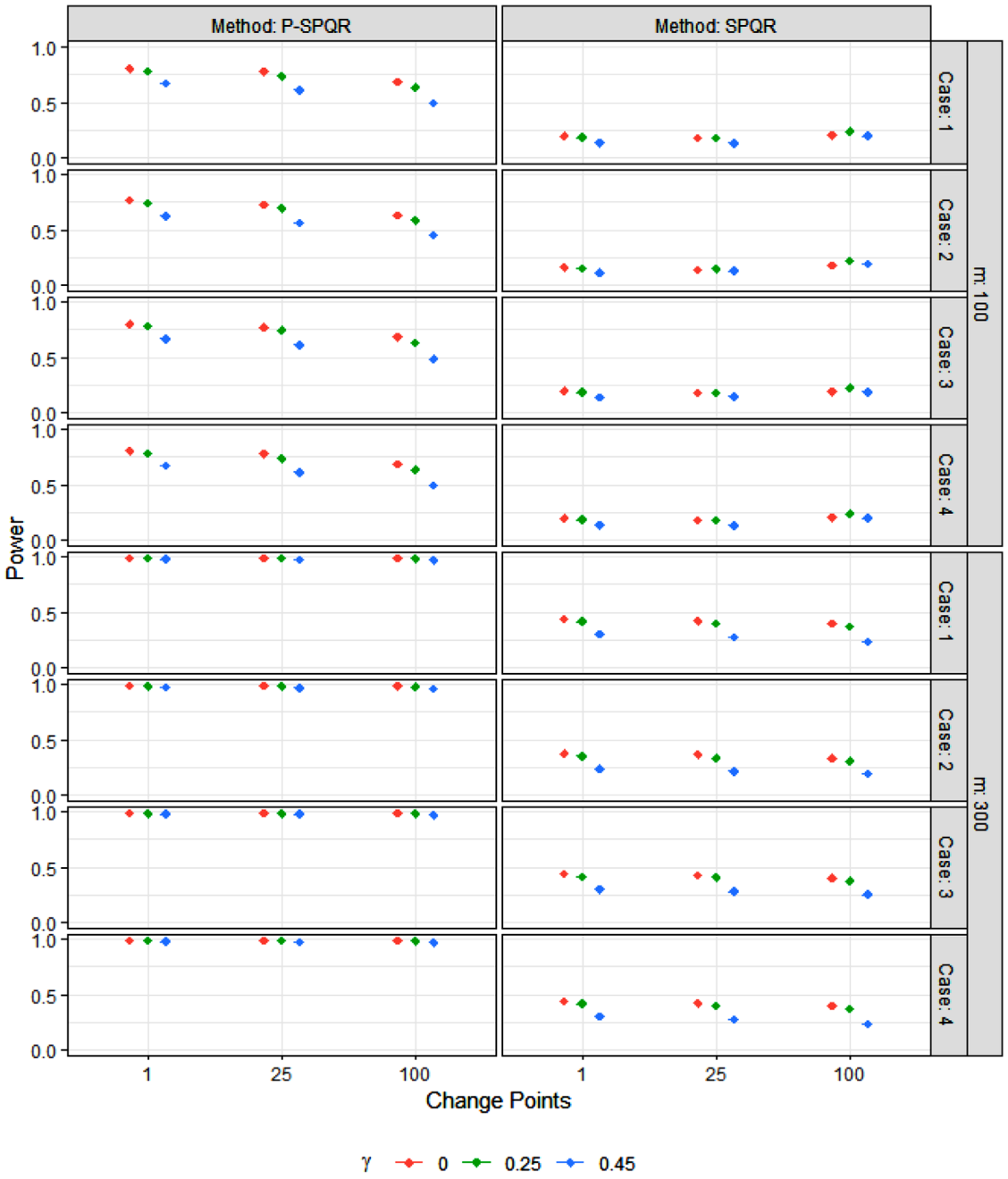

FIGURE 7 Estimated density of the stopping time for$\alpha = .05,\ \gamma \in \{0,\ 0.25,\ 0.45\}$, and various pairs of$(p, m)$for the P-SPQR method in Case 1

Theoretical properties of the DC are established by Székely et al. (2007). When compared to the Pearson correlation coefficient , the DC satisfies the following properties:

1. $0 \leq \text{dCorr}(X, Y) \leq 1$,
2. The $\text{dCorr}(X, Y) = 0$ if and only if X and Y are independent,
3. $\text{dCorr}(X, Y) = \text{dCorr}(Y, X)$.

Li, Zhong, and Zhu (2012) addressed the use of DC as a feature screening technique and its capacity to pick the very relevant features in comparison with the Pearson correlation coefficient. Furthermore, the DC is more effective than the Pearson correlation coefficient in the presence of a nonlinear association between X and Y. Thus, in the univariate filter

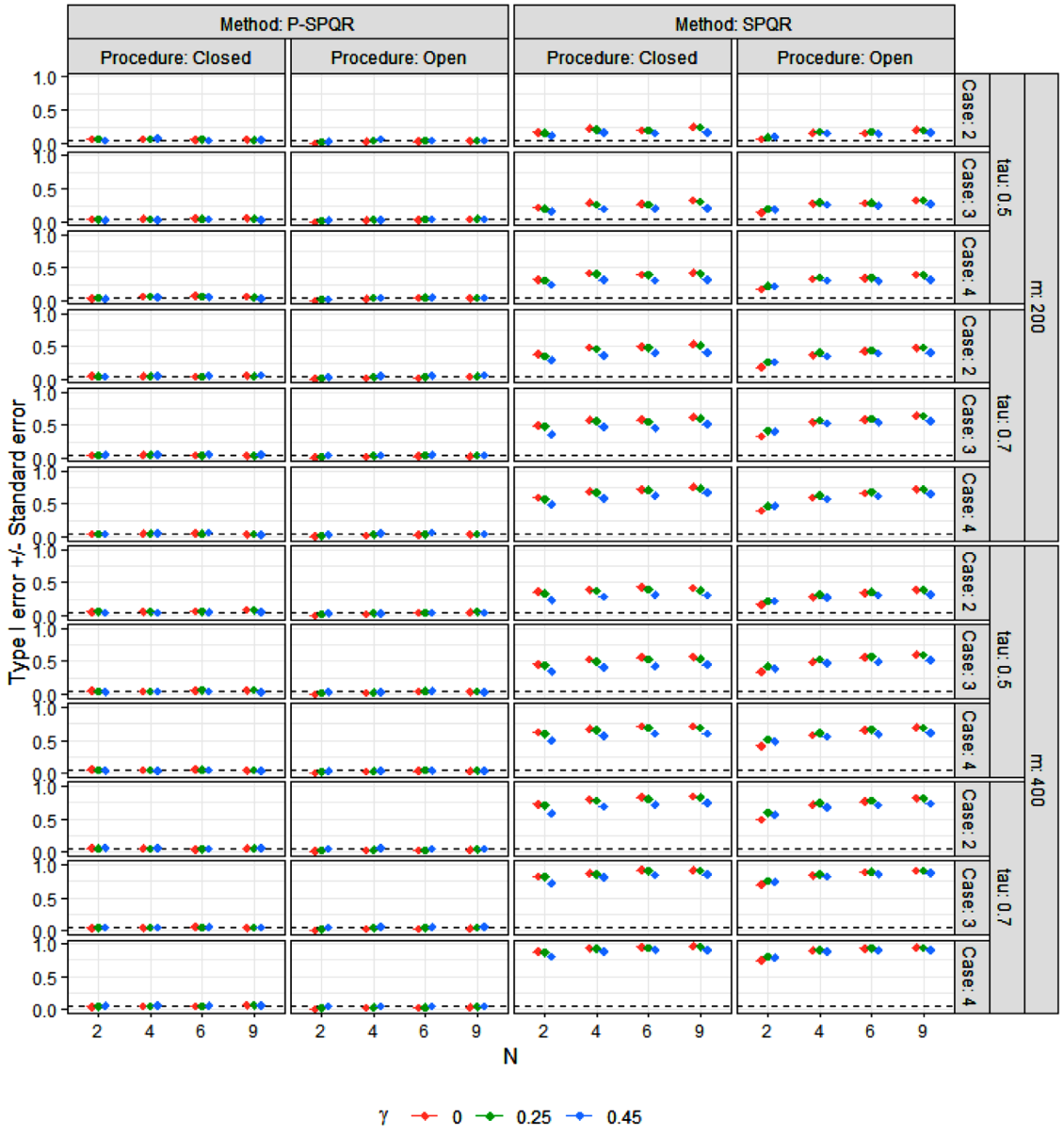

FIGURE 8 The gene expression measurement for BRCA1 of the breast cancer gene expression data

method, we use DC as a measure to identify the features that are influenced in predicting the response variable. In the first step, we compute the DC between the response variable y and each predictor variable X_i. For this analysis, the minimum DC is set to 0.50. Therefore, if the predictor variables above the threshold value of 0.50 are retained and those below 0.50 are discarded. In particular, at DC = 0.50, about 1,003 of 17,814 features are selected to build a model.

First, from the graph above we observe that there are no changes in the first 51 observations. We then apply the regular log-likelihood method to verify our guess which turns out there is no change in the first 51 observations. The residuals are assumed to be i.i.d. normal distribution with mean 0 and variance $\sigma^2 = \text{Var}(\mathcal{E}_i|X)$. Therefore, the first $m = 40$ observations are considered as the historical sample size. The significant explanatory variables are chosen using the SCAD penalized quantile regression at different quantile levels of interest, considering $\tau \in \{0.5, 0.75\}$. Our proposed P-SPQR method is used to monitor the future incoming observations sequentially with the control parameter value $\gamma = 0$. When quantile $\tau = 0.5$ our method detects 21 change points. The corresponding multiple change points are $\{92, 104, 111, 118, 121, 134, 169, 181, 186, 199, 203, 209, 363, 393, 412, 420, 443, 458, 495, 513, 532\}$. However, we detect only four change points at the quantile level $\tau = 0.75$, they are $\{52, 64, 76, 84\}$. This could be due to a lack of information available at the upper tail than at the median. All these change points are plotted in Figure 9.

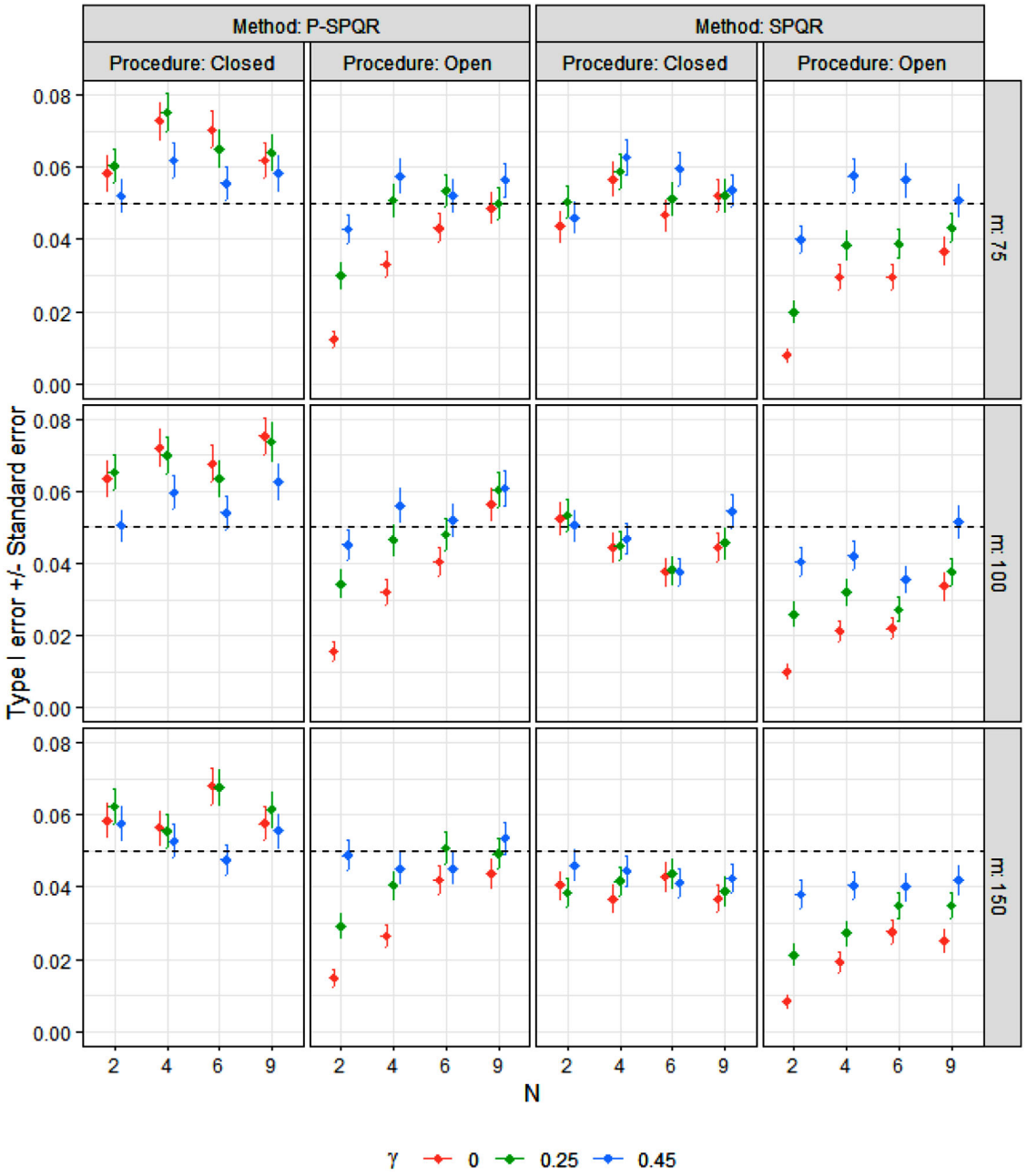

FIGURE 9 Change points at quantile level: (a)$\tau = 0.5$and (b)$\tau = 0.75$

Furthermore, we compare our proposed method with LASSO-penalized quantile regression model for high-dimensional data. The ℓ_1-penalized quantile regression is defined as

$$\widetilde{Q}(\beta, \tau) = \sum_{i=1}^{m} \psi_\tau \left(Y_i - \mathrm{x}_i^\top \beta\right)^2 + \lambda_m \sum_{j=1}^{p} |\beta_j|.$$

The ℓ_1-penalized quantile regression solves the following minimization problem:

$$\hat{\beta}^\tau_{m,\ell_1} = \underset{\beta \in \mathbb{R}^P}{\arg\min} \left\{ \sum_{i=1}^{m} \psi_\tau \left(Y_i - \mathrm{x}_i^\top \beta\right)^2 + \lambda_m \sum_{j=1}^{p} |\beta_j| \right\},$$

where $\lambda_m(\geq 0)$ is called the tuning parameter. We adopt 10-fold cross-validation to select appropriate values for the tuning parameter λ_m. Then we monitor the structural change with the ℓ_1-penalized quantile regression sequentially, however, no change point is detected. It shows that our method performs well and thus its superiority in terms of detecting structural changes is well established. Other existing methods including, Zhou et al. (2015) and Horváth et al. (2004) are not applicable since they are entirely univariate, therefore, are not suitable for high-dimensional scenarios.

5 | CONCLUSION

In this paper, we propose the SPQR procedure for the sequential change point detection for high-dimensional data. The SPQR performs variable selection and estimation simultaneously. Moreover, to improve the SPQR-based monitoring method, we develop the P-SPQR method for high-dimensional data. The asymptotic properties of the test-statistic under null and alternative hypotheses have been derived. Simulations are conducted to illustrate the performance of both methods with different historical sample sizes, various error distributions, and three different control parameters. The results show that the P-SPQR method reflects much better robustness in various error distributions including heavy-tailed, heteroskedastic, and skewed error distributions. Furthermore, in all cases, the P-SPQR method shows higher power than the SPQR method. As expected, in both methods the power tends to decrease when the change location is farther away from the historical sample size, however the power of the test increases as the historical sample size increases. The closed-end procedure is preferable for small $N(< 6)$, and both closed- and open-end procedures behave similarly for large $N(\geq 6)$. Simulation results indicate that larger control parameter values appear to detect the structural change much faster, whereas smaller control parameter values contribute more delays of detection. So for situations where we suspect or expect the true structural change may occur immediately after the historical sample, we suggest a larger control parameter and vice versa. We have only considered the P-SPQR method for high-dimensional data. Interestingly, we notice that all of the previous findings for small p continue to hold. Our proposed P-SPQR method is applied to a breast cancer gene expression data to locate multiple change points sequentially. Although we believe our work has achieved good results in sequential change point analysis for high-dimensional data, it is worth mentioning that our method can also be extended to study how correlated predictors affect the response in sequential change point analysis. This is an important topic for future research.

ACKNOWLEDGMENTS

The authors would like to thank two anonymous reviewers, an associate editor, and the editor for their constructive comments and suggestions which helped to improve this manuscript significantly.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES


This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Wei Ning  https://orcid.org/0000-0002-2056-3236

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Aue, A., Cheung, R. C. Y., Lee, T. C. M., & Zhong, M. (2014). Segmented model selection in quantile regression using the minimum description length principle. *Journal of the American Statistical Association*, *109*, 1241–1256.

Bai, J. (1996). Testing for parameter constancy in linear regressions: An empirical distribution function approach. *Econometrica*, *64*(3), 597–622.

Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, *19*(2), 521–547.

Bühlmann, P., Kalisch, M., & Maathuis, M. (2010). Variable selection in high-dimensional linear models: Partially faithful distributions and the PC-simple algorithm. *Biometrika*, *97*(2), 261–278.

Ciuperca, G. (2017). Real time change-point detection in a nonlinear quantile model. *Sequential Analysis*, *36*(1), 87–110.

Ciuperca, G. (2018). Test by adaptive LASSO quantile method for real-time detection of a change-point. *Metrika*, *81*, 689–720.
Davino, C., Furno, M., & Vistocco, D. (2013). *Quantile regression: Theory and applications*. Chichester: Wiley.
Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*, 1348–1360.
Furno, M. (2007). Parameter instability in quantile regression. *Statistical Modelling*, *7*, 345–362.
Furno, M. (2012). Tests for structural break in quantile regressions. *AStA Advances in Statistical Analysis*, *96*(4), 493–515.
Furno, M., & Vistocco, D. (2013). Qu test for structural breaks in quantile regressions. *International Journal of Statistics and Probability*, *2*(4), 42–55.
Furno, M., & Vistocco, D. (2018). *Quantile regression estimation and simulation*. Chichester: Wiley.
Golub, G. H., Heath, M., & Wahba, G. (1979). Generalised cross validation as a method for choosing a good ridge parameter. *Technometrics*, *21*(2), 215–223.
Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.
Horváth, L., Huškova, M., Kokoszká, P., & Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, *126*, 225–251.
Huang, J., & Xie, H. (2007). Asymptotic oracle properties of SCAD-penalized least squares estimators. *Asymptotics: Practices, Processes and Inverse Problems (IMS Lecture Notes-Monograph Series)*, *55*, 149–166.
Hušková, M., & Kirch, C. (2012). Bootstrapping sequential change-point tests for linear regression. *Metrika*, *75*, 673–708.
Kai, B., Li, R., & Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics*, *39*(1), 305–332.
Kim, Y., Choi, H., & Oh, H.-S. (2008). Smoothly clipped absolute deviation on high dimensions. *Journal of the American Statistical Association*, *103*, 1665–1673.
Knight, K., & Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, *28*(5), 1356–1378.
Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, *91*(1), 74–89.
Koenker, R. (2005). *Quantile regression*. Cambridge, UK: Cambridge University Press.
Koenker, R., & Bassett, G. (1978). Regrssion quantiles. *Econometrica*, *46*, 33–50.
Koenker, R., Ng, P., & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, *81*, 673–680.
Li, R., Zhong, W., & Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, *107*(499), 1129–1139.
Mallows, C. L. (1973). Some comments on c_P. *Technometrics*, *15*(4), 661–675.
Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, *146*(1), 170–184.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.
Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, *35*(6), 2769–2794.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, *58*(1), 267–288.
Wang, H. J., & He, X. (2007). Detecting differential expressions in genechip microarray studies: A quantile approach. *Journal of the American Statistical Association*, *102*, 104–112.
Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, *107*(497), 214–222.
Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, *19*, 801–817.
Zhang, L., Wang, H. J., & Zhu, Z. (2014). Testing for change points due to a covariate threshold in quantile regression. *Statistica Sinica*, *24*, 1859–1877.
Zhou, M., Wang, H. J., & Tang, Y. (2015). Sequential change point detection in linear quantile regression models. *Statistics and Probability Letters*, *100*, 98–103.
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1428.
Zou, H., & Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics*, *36*(3), 1108–1126.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ratnasingam S, Ning W. Sequential change point detection for high-dimensional data using nonconvex penalized quantile regression. *Biometrical Journal*. 2021;63:575–598.
https://doi.org/10.1002/bimj.202000078