Taylor & Francis
Taylor & Francis Group

Sequential Analysis

Design Methods and Applications

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsqa20

Monitoring sequential structural changes in penalized high-dimensional linear models

Suthakaran Ratnasingam & Wei Ning

To cite this article: Suthakaran Ratnasingam & Wei Ning (2021): Monitoring sequential structural changes in penalized high-dimensional linear models, Sequential Analysis, DOI: 10.1080/07474946.2021.1940500

To link to this article: https://doi.org/10.1080/07474946.2021.1940500

Published online: 13 Sep 2021.

SEQUENTIAL ANALYSIS
https://doi.org/10.1080/07474946.2021.1940500

Taylor & Francis
Taylor & Francis Group

Monitoring sequential structural changes in penalized high-dimensional linear models

Suthakaran Ratnasingam[a] and Wei Ning[b,c]

[a]Department of Mathematics, California State University, San Bernardino, California, USA; [b]Department of Mathematics and Statistics Bowling, Green State University, Bowling Green, Ohio, USA; [c]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China

ABSTRACT

In this article, we propose a procedure to monitor the structural changes in the penalized regression model for high-dimensional data sequentially. Our approach utilizes a given historical data set to perform both variable selection and estimation simultaneously. The asymptotic properties of the test statistics are established under the null and alternative hypotheses. The finite sample behavior of the monitoring procedure is investigated with simulation studies. The proposed method is applied to a real data set to illustrate the detection procedure.

ARTICLE HISTORY
Received 28 February 2020
Revised 13 September 2020
Accepted 9 June 2021

KEYWORDS
Sequential changes; Linear regression model; High dimension; SCAD

1. INTRODUCTION

In sequential change-point analysis the observations are received sequentially and we need to decide whether to continue the process or not after every new observation. The decision has to be made solely based on previous information in real time. There is a rich literature in sequential change-point detection analysis for univariate data; see, for example, Page (1954), Shiryaev (1963), Roberts (1966), Lorden (1971), Siegmund (1985), and Tartakovsky, Nikiforov, and Basseville (2014). Horváth et al. (2004) proposed a sequential monitoring method to detect structural change based on weighted cumulative sums (CUSUMs) of residuals, in which the unknown in-control parameter has been replaced by its least squares estimate from the training observations. Furthermore, their monitoring process continues until infinity when the null hypothesis is not rejected. In practice, for most real-world applications, we cannot continue to monitor the process until infinity if no change exists. Horváth, Kokoszka, and Steinebach (2007) further investigated the monitoring process for a linear model that stops even if no change point is detected after a certain number of observations. Zhou, Wang, and Tang (2015) developed a method for sequential detection of structural changes in linear quantile regression models. Chen (2019) proposed a new sequential change-point detection based on the nearest neighbor information and applied to sequences of multivariate observations or non-Euclidean data.

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

High-dimensional data analysis is a popular research area in statistics. Generally, data for which the number of explanatory variables (p) exceeds the number of observations (n) is referred to as high-dimensional data. In a real-world scenario, we often deal with a large number of explanatory variables. For example, genomics and health care data sets have a large number of explanatory variables for each observation. For a large number of predictor variables, there is a chance that these variables may correlate with other explanatory variables. Also, too many variables could lead to overfitting in a regression model. On the other hand, one can use variable selection procedures such as forward selection, backward elimination, or the stepwise method to select the best subset of predictor variables and to attain parsimony and best fit. These methods, however, have certain limitations for collinear regressors and lead to inaccurate results for high-dimensional data.

The ordinary least squares method is not consistent in the setting of $p > n$. Penalized regression techniques have been proven to be effective for modeling high-dimensional data; see, for example, Tibshirani (1996) and Hastie, Tibshirani, and Friedman (2009). The idea behind the penalized regression method is to perform linear regression while shrinking the coefficients toward zero. The benefit of this approach is that shrinking the coefficient estimates can significantly reduce their variances. Two well-known techniques for shrinking the regression coefficients toward zero are ridge regression and least absolute shrinkage and selection operator (LASSO); see Tibshirani (1996). The major difference between them is that ridge (ℓ_2 penalty) regression shrinks all of the coefficients to a nonzero value, whereas LASSO (ℓ_1 penalty) shrinks some of the coefficients and sets other to be exactly equal to zero. Thus, the LASSO method performs both variable selection and parameter estimation simultaneously. The LASSO-type estimator was proposed by Knight and Fu (2000), where they minimized the residual sum of squares and penalty proportions to the model's parameter. Other penalized regression techniques include smoothly clipped absolute deviation (SCAD; Fan and Li 2001), Elastic Net (Zou and Hastie 2005), and adaptive LASSO (Zou 2006). The SCAD penalty function is computationally feasible and performs both variable selection and estimation simultaneously for high-dimensional data. Fan and Li (2001) established the asymptotic properties of SCAD penalized likelihood. In addition, the SCAD penalty function satisfies the oracle property; see Fan and Li (2001) and Fan and Peng (2004).

A handful of literature is available for using LASSO estimation in change-point analysis. J. Kim and Kim (2008) studied data with asymptotic behavior of the least squares estimators in segmented multiple regression with one or more change points. Harchaoui and Lévy-Leduc (2010) proposed an approach for estimation of the location of change points in one-dimensional piecewise constant signals observed with white noise. They used a penalized least squares criterion with a ℓ_1 penalty. Ciuperca (2014) studied the model selection procedure for adaptive LASSO with multiple change points and investigated its asymptotic properties. Ciuperca (2015) investigated the CUSUM test statistic based on adaptive LASSO for detecting change points sequentially in a linear model. The detection of change points in high-dimension data using low-dimensional compressive measurements in an online setting was considered in Chi and Wu (2015). Recently, Ratnasingam and Ning (2021) studied the sequential change-point detection method to monitor structural changes in penalized quantile regression models. To the

best of our knowledge, no previous study investigated the use of the SCAD penalty with a regression model for sequential change-point analysis for high-dimensional data sets. In this article, we study the sequential change-point problem under a high-dimensional scenario. We propose test statistics for sequential change-point detection procedures using a SCAD penalized regression model with a finite monitoring horizon for high-dimensional data.

This article is organized as follows. In Section 2, we describe notations, assumptions, the variable selection procedure, and oracle properties for the SCAD penalized regression model. In Section 3, we propose test statistics for open-ended and closed-ended procedures. Corresponding asymptotic results are established. In Section 4, simulations are conducted under different settings to investigate the performance of the proposed methods. A real data application is given in Section 5 to illustrate the detecting process. Some discussion and conclusions provided in Section 6.

2. SCAD PENALIZED REGRESSION

Suppose we have a random sample $\{Y_i, x_{i1}, ..., x_{ip}\}, i = 1, ..., m$. Consider the model

$$Y = X\beta + \mathcal{E}, \tag{2.1}$$

where $Y = (Y_1, ..., Y_m)$ is a vector of responses, X is an $m \times p$ matrix of predictors with ith row $x_i^\top = (x_{i1}, ..., x_{ip})$, where $i = 1, ..., m$; and jth column $X_j = (x_{1j}, ..., x_{mj})^\top$, where $j = 1, ..., p$; $\beta = (\beta_1, ..., \beta_p)^\top$ is a p-vector of unknown parameters and $\mathcal{E} = (\mathcal{E}_1, ..., \mathcal{E}_m)^\top$ represents an m-vector of independent and identically distributed (i.i.d.) random variables with mean 0 and variance σ^2.

Tibshirani (1996) introduced the LASSO method for variable selection and estimation. The LASSO method can be written as the minimization of the least squares penalized by the norm ℓ_1 of the vector β. LASSO can successfully shrink some coefficients to be exactly zero and give a sparse solution. Zhao and Yu (2006) noted that if an irrelevant predictor is highly correlated with the predictors in the true model, LASSO may not be able to distinguish it from the true predictors with any amount of data and any amount of regularization. Further, Knight and Fu (2000) proved that the LASSO estimator is only $n^{1/2}$-consistent under some regularity conditions. Thus, it cannot achieve simultaneous consistent variable selection and estimation. Therefore, the oracle property does not hold for LASSO; see for example, Fan and Li (2001) and Zou (2006). To improve the performance of LASSO, Fan and Li (2001) introduced an oracle selection procedure referred to as SCAD. The goal of SCAD is to penalize small coefficients heavily and large coefficients lightly. The penalized least squares estimator, denoted by $\hat{\beta}_m^{SCAD}$, can be defined as

$$\hat{\beta}_m^{SCAD} = \underset{\beta \in \mathbb{R}^P}{\operatorname{argmin}} \left\{ \sum_{i=1}^{m} (Y_i - x_i^\top \beta)^2 + \sum_{j=1}^{p} p_{\lambda_m}(\beta_j) \right\}, \tag{2.2}$$

where $p_{\lambda_m}(\cdot)$ is the penalty function with tuning parameter λ_m. The SCAD penalty function $p_{\lambda_m}(\cdot)$ is symmetric and continuously differentiable on $(-\infty, 0) \cup (0, \infty)$. For given $a > 2$ and $\lambda_m > 0$, the SCAD penalty $p_{\lambda_m}(\beta)$ is given by

$$p_{\lambda_m}(\beta) = \begin{cases} \lambda_m|\beta| & \text{if} \quad |\beta| \le \lambda_m, \\ -(\beta^2 - 2a\lambda_m|\beta| + \lambda_m^2)/[2(a-1)] & \text{if} \quad \lambda_m < |\beta| \le a\lambda_m, \\ (a+1)\lambda_m^2/2 & \text{if} \quad |\beta| > a\lambda_m, \end{cases}$$

where $\lambda_m (0 < \lambda_m < \infty)$ and a are two unknown parameters. The function is continuous and its first derivative is

$$p'_{\lambda_m}(\beta) = \lambda_m \left\{ I(|\beta| \le \lambda_m) + \frac{(a\lambda_m - \beta)_+}{(a-1)\lambda_m} I(|\beta| > \lambda_m) \right\}. \tag{2.3}$$

Fan and Li (2001) suggested $a = 3.7$ as a good choice for various problems. In this research, a was set to 3.7 and λ_m was selected by the cross-validation method. The tuning parameter λ_m controls the amount of shrinkage. In particular, the amount of shrinkage is proportional to the value of λ_m. Fan and Li (2001) noted that the SCAD penalty function satisfies three requirements for variable selection and coefficient estimation, including asymptotic unbiasedness, sparsity, and continuity of the estimated parameters.

For the sparse model, we consider a situation where most regression coefficients are exactly zero; that is, there are only a few predictors whose regression coefficients are nonzero. Without loss of generality, we assume that the first q regression coefficients are nonzero and the remaining $(p-q)$ regression coefficients are zero. Let $X = (X^{(1)}, X^{(2)})$, where $X^{(1)}$ is the first $m \times q$ submatrix and $X^{(2)}$ is the last $m \times (p-q)$ submatrix of X. Similarly, we denote $\beta = (\beta^{(1)}, \beta^{(2)})$. Let $C_m = \frac{1}{m} X^\top X$ and $C_m^{(u,v)} = \frac{1}{m} X^{(u)\top} X^{(v)}$, for $u, v = 1, 2$. Let $\beta_0 = (\beta_{01}, ..., \beta_{0p})$ be the true unknown parameter vector. Let $\mathcal{A} = \{j \in \{1, ..., p\} : \beta_{0j} \neq 0\}$ be the index set of the nonzero coefficients for the true parameter, where β_{0j} is the jth component of the true parameter vector β_0. We denote the SCAD penalized regression estimate by $\hat{\beta}_m^{SCAD}$. Let $\mathcal{A}^* = \{j \in \{1, ..., p\} : \hat{\beta}_{mj}^{SCAD} \neq 0\}$ be the index set of the SCAD penalized regression estimator calculated using the historical sample size m, where $\hat{\beta}_{mj}^{SCAD}$ is the jth element of the SCAD penalized regression estimator $\hat{\beta}_m^{SCAD}$. To obtain the limiting distribution, we make the following assumptions, also called the regularity conditions, which are needed to derive the asymptotics of the estimators:

A1. The model errors $\mathcal{E}_1, ..., \mathcal{E}_m, \mathcal{E}_{m+1}, ...$ are i.i.d. random variables. $E(\mathcal{E}_i) = 0, Var(\mathcal{E}_i) = \sigma^2 < \infty$ and $E(|\mathcal{E}_1|^\rho) < \infty$ for some $\rho \ge 2$.

A2. There exists a positive constant M_1 such that

$$\frac{1}{m} X_j^\top X_j \le M_1 \quad \text{for all } j = 1, ..., p \text{ and for all } m.$$

A3. There exists a positive constant M_2 such that $\alpha' C_m^{(1,1)} \alpha \ge M_2$ for all $||\alpha||_2^2 = 1$.

A4. $q = O(m^{c_1})$ for some $0 < c_1 < 1$.

A5. There exist positive constants c_2 and M_3 such that $c_1 < c_2 \le 1$ and

$$m^{(1-c_2)/2} \min_{j=1,\ldots,q} |\beta_j| \geq M_3.$$

A6. Suppose that $p \leq m$ and C_m is nonsingular, with the smallest eigenvalue and the largest eigenvalue bounded by L.

The above are the commonly used regularity conditions in a high-dimensional linear model. Assumption A1 was used in Zou (2006) and Ciuperca (2014). Zhao and Yu (2006) used A2–A5 to prove the model selection consistency of the LASSO estimator. Y. Kim, Choi, and Oh (2008) used A2–A5 to show that the SCAD estimator has the oracle property on high-dimensional problems.

Remark 2.1. Under a high-dimensional setting, where $p > m$, A6 cannot be satisfied. However, if we find a subset $\mathbb{G}$ of $\{1,\ldots,p\}$ such that $\{1,\ldots,q\} \in \mathbb{G}$ and the design matrix $X_{\mathbb{G}} = (X_j, j \in \mathbb{G})$ satisfies A6, then we can use $X_{\mathbb{G}}$ to find the oracle estimator. For more details, we refer the reader to Y. Kim, Choi, and Oh (2008).

For fixed historical sample size m, if assumptions A1–A5 are met, then the SCAD penalized estimator satisfies the oracle property, meaning

i. Consistency in variable selection, $\lim_{m\to\infty} P(\mathcal{A}^* = \mathcal{A}) = 1$.
ii. $\sqrt{m}\mathbb{A}_m(m^{-1}X_A^\top X_A/\sigma^2)^{1/2}(\hat{\beta}_m^{SCAD} - \beta_0) \to N(0, V)$, where $\mathbb{A}_m$ is an arbitrary matrix such that $\mathbb{A}_m\mathbb{A}_m^\top \to V$ and V is a $q \times q$ nonnegative symmetric matrix and contains the elements of the matrix C_m in the set $\mathcal{A}$.

3. SEQUENTIAL CHANGE-POINT PROBLEM

Let m be the size of the historical sample. We assume that there is no change in the historical sample. Similar to Chu, Stinchcombe, and White (1996), we use the historical sample to estimate the prechange coefficients of the SCAD penalized regression model. After we select the significant explanatory variables, the future incoming observations $\{Y_i, x_{i1}, \ldots, x_{ip}\}$, $i = m+1, m+2, \ldots$ are monitored sequentially following the historical sample size m. Let T_m be the monitoring horizon. The linear model after historical observations m is

$$Y_i = x_i^\top \beta_i + \mathcal{E}_i, \quad i = m+1, m+2, \ldots \tag{3.1}$$

At each time point i, our goal is to test whether we have the same model as the one using the historical sample size m. Under the null hypothesis, if there is no change in the coefficients,

$$H_0 : \beta_i = \beta_0 \quad \text{for} \quad i = m+1, m+2, \ldots \tag{3.2}$$

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

$$H_1 : \begin{cases} \beta_i = \beta_0; & i = m+1, m+1, ..., m+k, \\ \beta_i = \beta_1; & i = m+k+1, ...m+T_m \end{cases} \quad \text{and } \beta_0 \neq \beta_1. \tag{3.3}$$

According to Horváth et al. (2004),

$$\Gamma(m,k) = \frac{1}{\hat{\sigma}_m} \left| \sum_{i=m+1}^{m+k} \hat{\mathcal{E}}_i \right|, \tag{3.4}$$

where $\hat{\mathcal{E}}_i = Y_i - X_i^\top \hat{\beta}_m^{SCAD}$ for $i = m+1, m+2, ...$ and $\hat{\sigma}_m^2$ is the error variance, defined as

$$\hat{\sigma}_m^2 = \frac{1}{(m-p^*)} \sum_{i=1}^{m} (Y_i - x_i^\top \hat{\beta}_m^{SCAD})^2, \tag{3.5}$$

where p^* is the number of nonzero coefficients of the SCAD penalized estimator. For a given constant $\gamma \in [0, 1/2)$, the $g(m,k,\gamma)$ is called the normalizing function and is defined as

$$g(m,k,\gamma) = m^{1/2} \left(1 + \frac{k}{m}\right) \left(\frac{k}{k+m}\right)^{\gamma}, \tag{3.6}$$

where γ is called the control parameter. We propose the test statistic for monitoring structural change:

$$\Omega = \sup_{1 \leq k \leq T_m} \frac{\Gamma(m,k)}{g(m,k,\gamma)}. \tag{3.7}$$

3.1. Open-Ended Procedure

The monitoring process examined by Horváth et al. (2004) can continue possibly to infinity if no alarm is raised. This is referred to as an open-ended procedure. Thus, in the open-ended procedure, the monitoring horizon $T_m = \infty$. Once the monitoring process stops, subsequent rejection of the null hypothesis is defined as the stopping time of the monitoring scheme. The stopping time of the proposed test statistic is defined as

$$\Lambda(m) = \begin{cases} \inf\{k \geq 1; & \Gamma(m,k) \geq g(m,k,\gamma) c_\alpha(\gamma)\}, \\ \infty & \text{for all } k = 1, 2, 3, ..., \end{cases} \tag{3.8}$$

where $c_\alpha(\gamma)$ is the critical value that can be obtained through simulations at a given significance level $\alpha \in (0,1)$. Under the null hypothesis,

$$\lim_{m \to \infty} P(\Lambda(m) < \infty) = \alpha, \tag{3.9}$$

and under the alternative hypothesis,

$$\lim_{m \to \infty} P(\Lambda(m) < \infty) = 1. \tag{3.10}$$

3.2. Closed-Ended Procedure

In the closed-ended procedure, the monitoring process stops after a fixed number of observations even if no change is observed; see, for example, Horváth, Kokoszka, and

Steinebach (2007) and Zhou, Wang, and Tang (2015). Let $N>0$. Suppose $T_m < \infty$ with $\lim_{m\to\infty} T_m/m = N$. Under the closed-ended procedure, the stopping time of the proposed test statistic when the monitoring process stops and rejects the null hypothesis is defined as

$$\Lambda^*(m) = \begin{cases} \inf\{k \geq 1; & \Gamma(m,k) \geq g(m,k,\gamma)c_\alpha^*(\gamma)\}, \\ T_m & \text{for all } k = 1, \ldots, T_m, \end{cases} \tag{3.11}$$

where $c_\alpha^*(\gamma)$ is the $(1-\alpha)$th quantile of the asymptotic distribution obtained in Theorem 3.1. Under the null hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(m) < \infty) = \alpha, \tag{3.12}$$

and under the alternative hypothesis,

$$\lim_{m\to\infty} P(\Lambda^*(m) < \infty) = 1. \tag{3.13}$$

The monitoring process stops immediately for large γ. Thus, a large value of γ is preferred when the change in the regression coefficients happens shortly after m.

Theorem 3.1. *Under assumptions A1–A5, if the null hypothesis holds, for an open-ended procedure,*

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

and for closed-ended procedure,

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

where $\{W(t), 0 \leq t < \infty\}$ denotes the l-dimensional Wiener process, where l is the number of significant features in the model based on the historical data.

Using Theorem 3.1, under the null hypothesis, we can obtain the asymptotic distribution of the CUSUM test statistic for the open- and closed-ended procedures. The asymptotic critical value $c_\alpha(\gamma)$ is obtained from

$$P\left(\sup_{0\leq t<1} \frac{||W(t)||_\infty}{t^\gamma} \geq c_\alpha(\gamma) \right) = \alpha,$$

and the asymptotic critical value for the closed-ended procedure $c_\alpha^*(\gamma)$ can be obtained from

$$P\left(\sup_{0\leq t\leq N/(N+1)} \frac{||W(t)||_\infty}{t^\gamma} \geq c_\alpha^*(\gamma) \right) = \alpha,$$

where $\alpha \in (0,1)$ and the tuning parameter $0 \leq \gamma < 1/2$. We obtain asymptotic critical values through simulation. First, we generate a sequence of i.i.d. l-dimensional random vector $e_i = (e_{i1}, e_{i2}, \ldots, e_{il})$, where $e_{ij} \sim N(0,1)$. 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 open- and closed-ended procedures obtained over $t \in \{1/M, 2/M, \ldots, 1\}$ and $t \in \{1/M, 2/M, \ldots, N/(N+1)\}$, respectively. The critical value for a level-α test can be

Table 1. Asymptotic critical values for the open-ended procedure for various $\gamma \in \{0.00, 0.15, 0.25, 0.35, 0.45, 0.49\}$, calculated on 50,000 replications.

γ/α	0.010	0.025	0.050	0.100	0.250
0.00	3.019468	2.734264	2.486683	2.234026	1.833559
0.15	3.078669	2.795639	2.554139	2.303043	1.911031
0.25	3.135282	2.864542	2.626865	2.372679	1.990685
0.35	3.247067	2.970504	2.742768	2.498331	2.128268
0.45	3.479984	3.229682	3.008465	2.784761	2.434316
0.49	3.764611	3.508515	3.303435	3.073143	2.720441

Table 2. Asymptotic critical values for the closed-ended procedure when $N \in \{2, 4, 6, 9\}$ and $\gamma \in \{0.00, 0.15, 0.25, 0.35, 0.45, 0.49\}$, calculated on 50,000 replications.

γ	N/α	0.010	0.025	0.050	0.100	0.250
0.00	2	2.447139	2.211787	2.020908	1.815498	1.487324
	4	2.686489	2.430562	2.222431	1.990774	1.629909
	6	2.785824	2.514810	2.297350	2.057684	1.688002
	9	2.847110	2.572124	2.351546	2.108416	1.730943
0.15	2	2.649895	2.404296	2.201695	1.987356	1.647810
	4	2.829618	2.569068	2.355556	2.121323	1.759586
	6	2.902434	2.632341	2.411011	2.170361	1.801032
	9	2.948248	2.676014	2.446939	2.209239	1.833419
0.25	2	2.816933	2.561952	2.359019	2.134979	1.790156
	4	2.954011	2.681545	2.469789	2.236936	1.874830
	6	3.004398	2.735205	2.510444	2.276272	1.907793
	9	3.045116	2.767461	2.539297	2.300939	1.930600
0.35	2	3.026753	2.766572	2.563014	2.338523	1.994179
	4	3.122079	2.854851	2.635548	2.404585	2.049275
	6	3.149867	2.883356	2.661403	2.427641	2.070099
	9	3.171244	2.904029	2.679934	2.445955	2.085801
0.45	2	3.400988	3.136741	2.935825	2.713851	2.375432
	4	3.433242	3.167601	2.965047	2.739027	2.400873
	6	3.442337	3.179246	2.974512	2.748497	2.408898
	9	3.446603	3.188064	2.981583	2.754953	2.414795
0.49	2	3.732576	3.463671	3.257303	3.037973	2.688347
	4	3.744958	3.474354	3.268213	3.050684	2.701117
	6	3.747245	3.478562	3.271668	3.054774	2.706148
	9	3.749885	3.479867	3.274127	3.057681	2.708363

estimated by the $(1-\alpha)$ th quantile of the test statistics. The asymptotic critical values are given in Tables 1 and 2. The results are based on 50,000 iterations.

Theorem 3.2. *Under assumptions A1–A5, if the alternative hypothesis holds, we have*

$$\sup_{1 \le k \le T_m} \frac{\Gamma(m,k)}{g(m,k,\gamma)} \to \infty \quad \text{as} \quad m \to \infty.$$

Proofs are given in the Appendix.

4. SIMULATION STUDIES

In this section, we conduct Monte Carlo simulations to evaluate the performance of the sequential change-point detection procedure for both open- and closed-ended methods

Table 3. Type I errors of both open- and closed-ended procedures for SCAD penalized regression for various values of γ and the nominal significance level $\alpha = 0.05$.

		Closed-ended			Open-ended		
m	N/γ	0	0.25	0.45	0	0.25	0.45
75	2	0.047	0.053	0.044	0.013	0.024	0.036
	4	0.038	0.037	0.037	0.020	0.028	0.035
	10	0.040	0.042	0.036	0.031	0.033	0.035
	20	0.050	0.052	0.046	0.046	0.047	0.045
	50	0.041	0.041	0.036	0.040	0.039	0.037
100	2	0.035	0.038	0.036	0.010	0.019	0.030
	4	0.034	0.036	0.031	0.017	0.026	0.030
	10	0.039	0.038	0.038	0.030	0.034	0.036
	20	0.033	0.032	0.025	0.028	0.028	0.025
	50	0.034	0.034	0.027	0.032	0.033	0.027
200	2	0.028	0.029	0.031	0.009	0.016	0.028
	4	0.028	0.031	0.030	0.015	0.020	0.024
	10	0.041	0.039	0.037	0.030	0.034	0.036
	20	0.026	0.026	0.024	0.020	0.024	0.023
	50	0.030	0.032	0.028	0.029	0.029	0.027

for the SCAD penalized regression model. 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 is noticed.

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

$$Y_i = x_i^\top \beta_0 + \mathcal{E}_i, \quad i = 1, ..., m + T_m .$$

We consider the number of explanatory variables $p = 10$. The following two settings are considered. The first setting is used to evaluate the Type I errors of the proposed method. In the first case, the true parameter vectors $\beta_0 \in \{-2, 0, 2, 0, 10, 1, 0, 0, 8, -5\}$ and X_i for all $i \in \{1, ..., 10\}\backslash\{3, 4, 5\}$ have a standard normal distribution $N(0, 1)$ and $X_3 \sim N(2, 1)$, $X_4 \sim N(4, 1)$, and $X_5 \sim N(5, 1)$. In the second setting, under the null hypothesis, the true parameter vectors $\beta_0 \in \{0, 0, 2, 0, 0, 1, 0, 0, 1, 0\}$, and under the alternative hypothesis, we consider the parameter vector $\beta_1 \in \{0, 0, 0, 3, 0, 0, 1, 0, 0, -1\}$. We consider the two different distributions of the explanatory variables $X_1, X_2, ..., X_{10}$. Under H_0, X_i for all $i \in \{1, ..., 10\}\backslash\{3, 4, 5\}$ have a standard normal distribution $N(0, 1)$ 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 $X_i + 0.8$, where $X_i \sim N(0, 1)$ for all $i \in \{1, ..., 10\}$. Moreover, for both settings, the model errors $\mathcal{E}_i$ are i.i.d. $N(0, 1)$.

Table 3 summarizes the Type I error for both open- and closed-ended procedures. The various control parameter γ values and the different sizes of the historical observations m are considered. $\gamma \in \{0, 0.25, 0.45\}$ and $m \in \{75, 100, 200\}$. The results are based

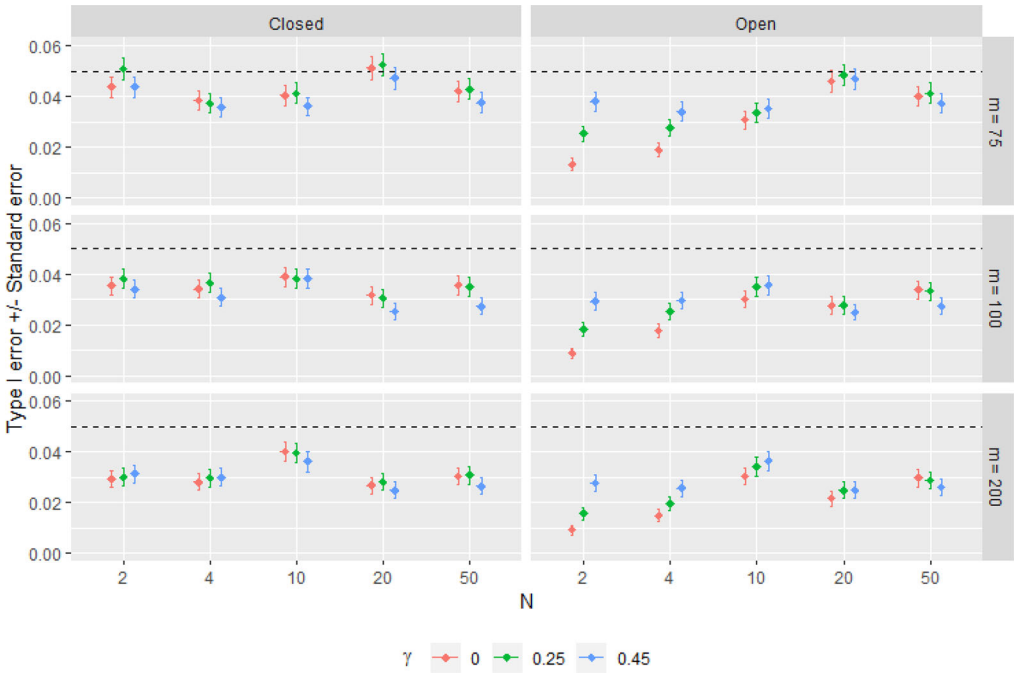

Figure 1. Type I error comparison for open- and closed-ended procedures.

on 2,500 iterations. The Type I errors based on the closed-ended procedure is always larger than the Type I errors obtained from the open-ended procedure. For small γ, the Type I errors of the open-ended procedure are below the nominal level 0.05. When the open-ended procedure is considered, smaller N provides slightly deflated Type I errors. Thus, for small N, we suggest larger values of γ close to 0.5. Type I errors are compared in Figure 1.

We conducted Monte Carlo simulation to investigate the performance of the proposed method. First, we performed power analysis for different control parameter $\gamma \in \{0, 0.25, 0.45\}$ while changing the size of a test α, considering $\alpha \in \{0.025, 0.05, 0.1\}$. Further, the simulations were carried out under various true change-point locations $k^* \in \{1, 25, 75\}$ with different historical sample sizes $m \in \{75, 100, 200\}$. We evaluated the power of the test based 1000 simulations and the results are summarized in Table 4. The results are sketched in Figure 2.

The five-number summary of the stopping time for both open- and closed-ended procedures summarized in Tables 5 and 6 and sketched in Figure 3. In all cases, the processes were monitored from $(m+1)$ until time $9\,m$. To see the effect of the historical sample size, we performed the simulations for various m, such as $m \in \{100, 300, 600\}$. Further, we changed the true change-point location, considering $k^* \in \{1, 25, 100\}$ and level $\alpha \in \{0.025, 0.05, 0.1\}$. Clearly, the selection of γ values influences the stopping time. As mentioned in Horváth et al. (2004), it is clear that smaller values of γ result in a longer time to detect the structural change, whereas larger γ values result in faster detection. We compared the estimated density of the stopping time at various change-point locations, historical sample size, control parameter, and test size (see

Table 4. Power comparison for closed-ended procedure for change points $k^* \in \{1, 25, 100\}, \alpha \in \{0.025, 0.05, 0.1\}$, and $\gamma \in \{0, 0.25, 0.45\}$.

		$k^* = 1$			$k^* = 25$			$k^* = 75$		
γ	α/m	75	100	200	75	100	200	75	100	200
0.00	0.025	0.948	0.977	0.995	0.936	0.956	0.996	0.857	0.943	0.996
	0.050	0.993	0.999	0.999	0.999	0.997	1	0.988	0.995	1
	0.100	1	1	1	1	1	1	1	1	1
0.25	0.025	0.925	0.962	0.991	0.876	0.933	0.992	0.757	0.876	0.992
	0.050	0.969	0.988	0.999	0.973	0.989	0.999	0.952	0.985	0.999
	0.100	0.999	1	1	1	1	1	0.999	1	1
0.45	0.025	0.860	0.930	0.985	0.756	0.881	0.970	0.552	0.731	0.970
	0.050	0.937	0.965	0.995	0.918	0.960	0.997	0.846	0.931	0.997
	0.100	0.991	0.997	0.999	0.992	0.999	1	0.992	0.997	1

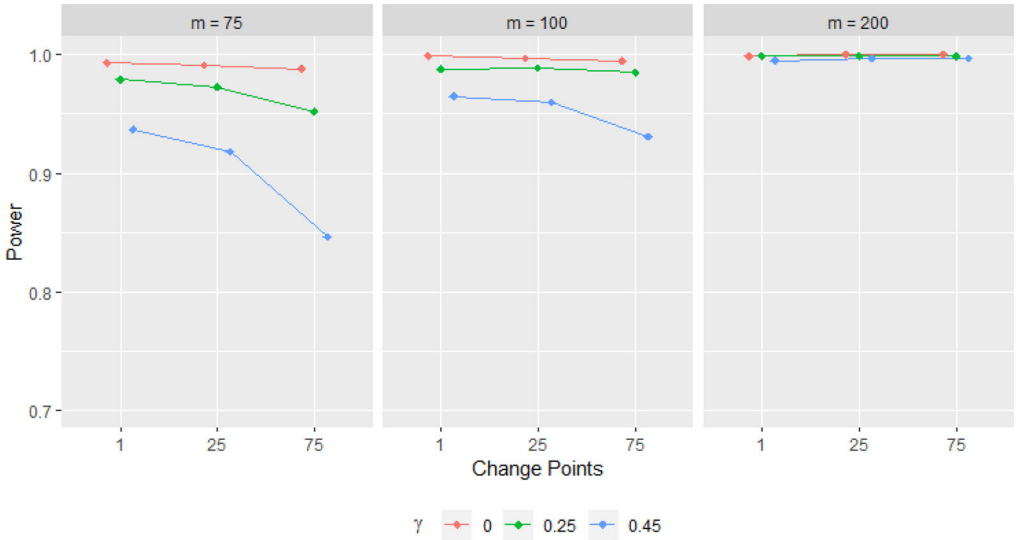

Figure 2. Power comparison for closed-ended procedure with $\alpha = 0.05$.

Table 5. Five-number summary for the detection time for the open-ended procedure with $k^* \in \{1, 25, 100\}$, $\gamma \in \{0, 0.25, 0.45\}$, and $\alpha = 0.05$.

		$m = 100$			$m = 300$			$m = 600$		
γ	Summary/k^*	1	25	100	1	25	100	1	25	100
0.00	min	4	27	72	6	31	102	10	33	109
	Q1	12	38	122	21	47	127	30	56	134
	Med	18	45	134	29	55	137	39	66	146
	Q3	26	57	141	40	67	152	48	79	161
	max	123	200	353	145	159	263	192	227	256
0.25	min	2	9	9	2	26	75	3	27	103
	Q1	5	34	119	7	36	118	9	39	121
	Med	8	40	130	12	42	128	15	45	130
	Q3	14	49	147	20	52	141	25	56	144
	max	102	197	353	95	149	244	153	187	239
0.45	min	1	1	1	2	3	3	2	13	13
	Q1	2	32	119	3	32	115	3	32	115
	Med	4	37	131	4	37	124	4	37	124
	Q3	7	46	148	8	44	137	8	44	135
	max	98	200	359	80	123	243	66	153	237

Table 6. Five-number summary for the detection time for the closed-ended procedure with $k^* \in \{1, 25, 100\}$, $\gamma \in \{0, 0.25, 0.45\}$, and $\alpha = 0.05$.

		$m = 100$			$m = 300$			$m = 600$		
γ	Summary/k^*	1	25	100	1	25	100	1	25	100
0.00	min	4	26	65	6	31	102	9	33	109
	Q1	11	37	120	20	45	125	28	54	132
	Med	16	44	132	27	53	135	37	64	144
	Q3	24	55	148	38	65	149	50	76	158
	max	122	197	335	144	157	262	192	226	255
0.25	min	2	8	8	2	26	74	3	27	102
	Q1	5	33	118	7	36	118	8	38	120
	Med	8	39	129	12	41	127	14	45	129
	Q3	14	48	145	20	51	140	24	55	142
	max	102	194	335	95	149	243	153	163	239
0.45	min	1	1	1	2	3	3	2	4	4
	Q1	2	32	119	3	32	115	3	32	115
	Med	4	37	131	4	37	124	4	36	124
	Q3	7	46	147	8	44	137	8	44	135
	max	54	200	359	80	123	243	66	151	236

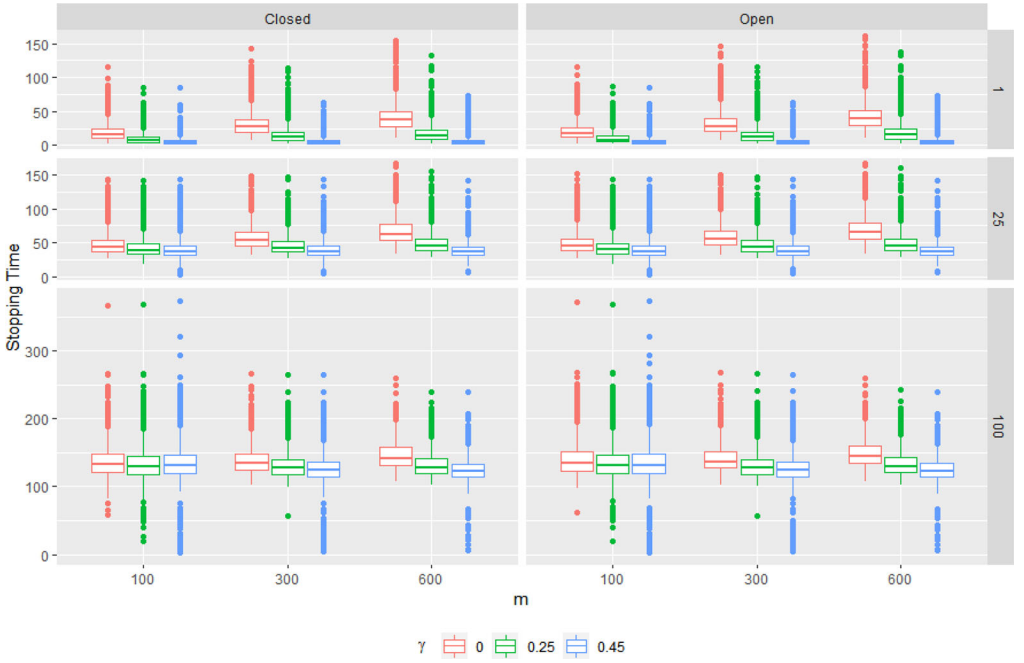

Figure 3. Stopping time comparison for closed- and open-ended procedures with $k^* \in \{1, 25, 100\}$, $\gamma \in \{0, 0.25, 0.45\}$, $\alpha = 0.05$, and various historical sample sizes.

Figure 4). When α changes from 0.025 to 0.05, the estimated densities are roughly identical. Not surprising, the historical sample m has a significant influence on the stopping time determination. When m changes from 100 to 300, we observe a high variability in the estimated densities. A small variation can be observed between the estimated densities for a fixed control value γ irrespective of the historical sample size.

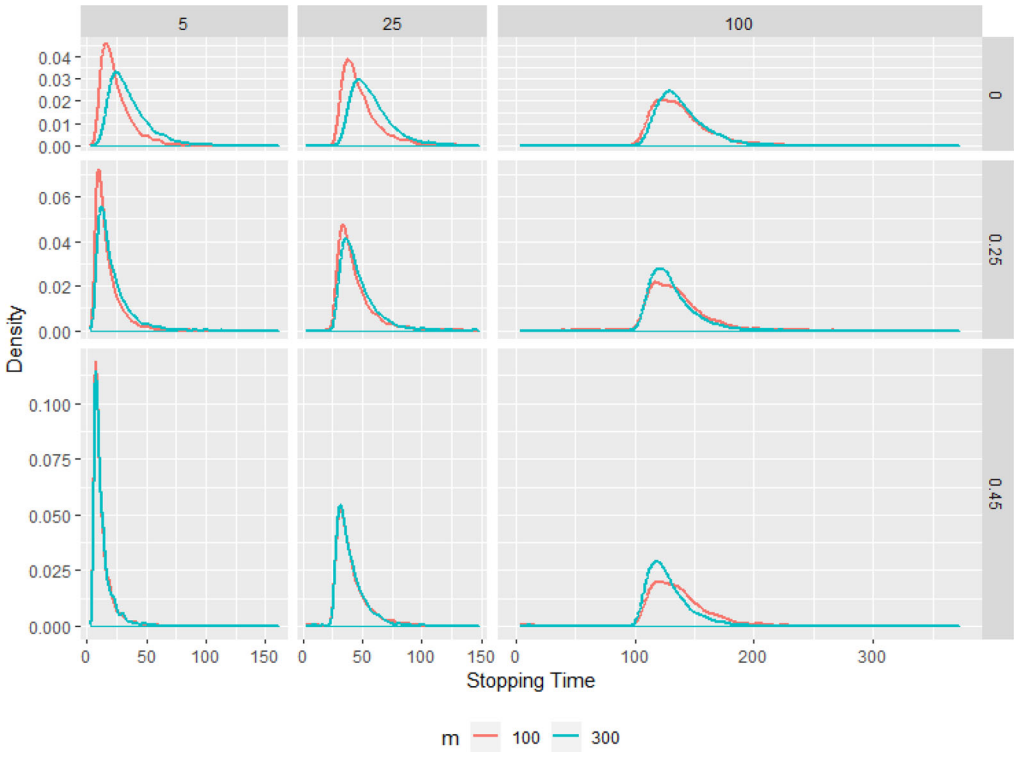

Figure 4. Estimated density of the stopping time for $k^* = \{5, 25, 100\}$ and $\gamma \in \{0,\ 0.25,\ 0.45\}$.

4.1. *Large* p

To study the performance of the monitoring process in a high-dimensional setting, we conducted another simulation study. We generated high-dimensional data sets with (p, m), considering $(100, 75)$, $(200, 100)$, and (300, 200). We considered the following two settings. In the first setting, the nonzero components of the true parameters are $\beta_{0,1} = -5$, $\beta_{0,2} = 2$, $\beta_{0,3} = 5$, $\beta_{0,4} = 1$, $\beta_{0,5} = -3$, $\beta_{0,61} = -10$, and $\beta_{0,91} = 8$. The predictor variables X_i for all $i \in \{1, \ldots, p\} \backslash \{3, 4, 5\}$ have a standard normal distribution $N(0, 1)$ and $X_3 \sim N(2, 1)$, $X_4 \sim N(4, 1)$, and $X_5 \sim N(5, 1)$.

In the second setting, under the null hypothesis, the true parameter vectors $\beta_{0,1} = -1$, $\beta_{0,2} = 1$, $\beta_{0,3} = -1$, $\beta_{0,4} = 4$, $\beta_{0,5} = -2$, $\beta_{0,58} = -3$, and $\beta_{0,86} = 2$, and under the alternative hypothesis, we consider the parameter vector $\beta_{1,1} = 3$, $\beta_{1,2} = 2$, $\beta_{1,45} = -2$ and $\beta_{1,93} = 2$ and the two different distributions of the explanatory variables $X_1, X_2, \ldots, X_p$. Under H_0, X_i for all $i \in \{1, \ldots, p\} \backslash \{3, 4, 5\}$ have a normal distribution $N(0, 1)$ 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 $X_i + 0.8$ where $X_i \sim N(0, 1)$ for all $i \in \{1, \ldots, p\}$. Moreover, for both settings, the model errors $\mathcal{E}_i$ are i.i.d. $N(0, 1)$.

Table 7 summarizes the Type I error for both open- and closed-ended procedures. The various control parameter values $\gamma \in \{0, 0.25, 0.45\}$ and the different sizes of the historical observations $m \in \{75, 100, 200\}$ are considered. The results are based on 2,500 iterations. Type I errors based on the closed-ended procedure are always larger than the Type I errors based on the open-ended procedure. In the open-ended operation, it is

Table 7. Type I errors of both open- and closed-ended procedures for SCAD penalized regression for various values of γ and the nominal significance level $\alpha = 0.05$.

		Closed-ended			Open-ended		
(p, m)	N/γ	0	0.25	0.45	0	0.25	0.45
(100, 75)	2	0.041	0.046	0.040	0.012	0.022	0.035
	4	0.043	0.045	0.038	0.025	0.034	0.034
	10	0.037	0.038	0.035	0.028	0.033	0.033
	20	0.044	0.043	0.034	0.036	0.039	0.033
	50	0.048	0.047	0.046	0.047	0.046	0.046
(200, 100)	2	0.037	0.038	0.034	0.008	0.018	0.027
	4	0.031	0.032	0.027	0.016	0.022	0.024
	10	0.036	0.033	0.027	0.026	0.028	0.026
	20	0.035	0.032	0.032	0.027	0.029	0.032
	50	0.039	0.042	0.034	0.038	0.041	0.034
(300, 200)	2	0.028	0.030	0.030	0.008	0.016	0.024
	4	0.024	0.027	0.022	0.012	0.016	0.020
	10	0.025	0.026	0.022	0.019	0.022	0.020
	20	0.022	0.019	0.026	0.017	0.018	0.025
	50	0.032	0.028	0.024	0.030	0.027	0.024

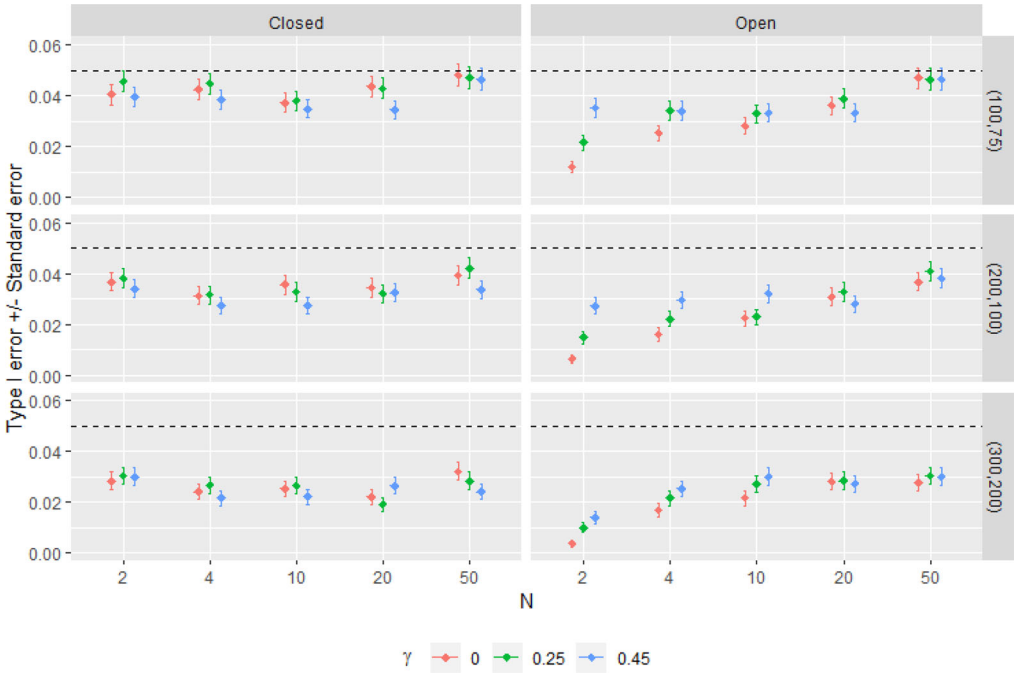

Figure 5. Type I error comparison for open- and closed-ended procedures for various (p, m).

vital to properly select the value of the control parameter. Type I errors are comparatively low in the open-ended procedure for small γ. When the open-ended procedure is considered, smaller N provides slightly deflated Type I errors and it improves for large N. Thus, for small N, we recommend larger tuning parameter values close to 0.5. For large (p, m), we deflated Type I errors for both open- and closed-ended procedures. Type I errors are sketched in Figure 5.

Table 8. Power comparison for closed-ended procedure with $k^* \in \{1, 25, 75\}, \alpha \in \{0.025, 0.05, 0.1\}, \gamma \in \{0, 0.25, 0.45\}$, and $p \in \{100, 200, 300\}$.

		$k^* = 1$			$k^* = 25$			$k^* = 75$		
γ	α/m	75	100	200	75	100	200	75	100	200
					$p = 100$					
0.00	0.025	0.962	0.980	0.999	0.935	0.970	0.999	0.875	0.938	0.996
	0.050	0.990	0.996	1	0.994	0.998	1	0.983	0.996	1
	0.100	1	1	1	1	1	1	1	1	1
0.25	0.025	0.941	0.969	0.999	0.886	0.946	0.997	0.779	0.875	0.993
	0.050	0.977	0.992	1	0.978	0.992	1	0.967	0.984	0.999
	0.100	0.998	0.998	1	0.999	1	1	1	1	1
0.45	0.025	0.874	0.936	0.999	0.768	0.879	0.995	0.599	0.727	0.976
	0.050	0.948	0.973	0.999	0.936	0.965	0.999	0.876	0.920	0.997
	0.100	0.989	0.996	1	0.995	0.998	1	0.990	0.997	1
					$p = 200$					
0.00	0.025	0.943	0.969	1	0.925	0.964	1	0.866	0.932	0.994
	0.050	0.990	0.993	1	0.993	0.998	1	0.984	0.990	1
	0.100	0.998	1	1	1	1	1	1	1	1
0.25	0.025	0.911	0.954	1	0.876	0.935	0.998	0.780	0.866	0.990
	0.050	0.969	0.986	1	0.978	0.991	1	0.963	0.981	0.999
	0.100	0.996	0.998	1	1	1	1	1	1	1
0.45	0.025	0.842	0.910	0.998	0.765	0.870	0.993	0.610	0.721	0.977
	0.050	0.925	0.958	1	0.928	0.967	0.998	0.864	0.913	0.994
	0.100	0.985	0.996	1	0.994	0.998	1	0.989	0.995	1
					$p = 300$					
0.00	0.025	0.946	0.976	1	0.921	0.966	0.999	0.887	0.944	0.999
	0.050	0.982	0.993	1	0.990	0.998	1	0.994	0.997	1
	0.100	1	1	1	1	1	1	1	1	1
0.25	0.025	0.915	0.962	0.999	0.886	0.932	0.999	0.785	0.875	0.992
	0.050	0.967	0.988	1	0.976	0.992	0.999	0.964	0.986	1
	0.100	0.998	1	1	1	1	1	1	1	1
0.45	0.025	0.859	0.928	0.996	0.793	0.869	0.996	0.624	0.746	0.968
	0.050	0.931	0.968	0.999	0.926	0.967	0.999	0.868	0.928	0.994
	0.100	0.985	0.994	1	0.994	0.998	1	0.988	0.999	1

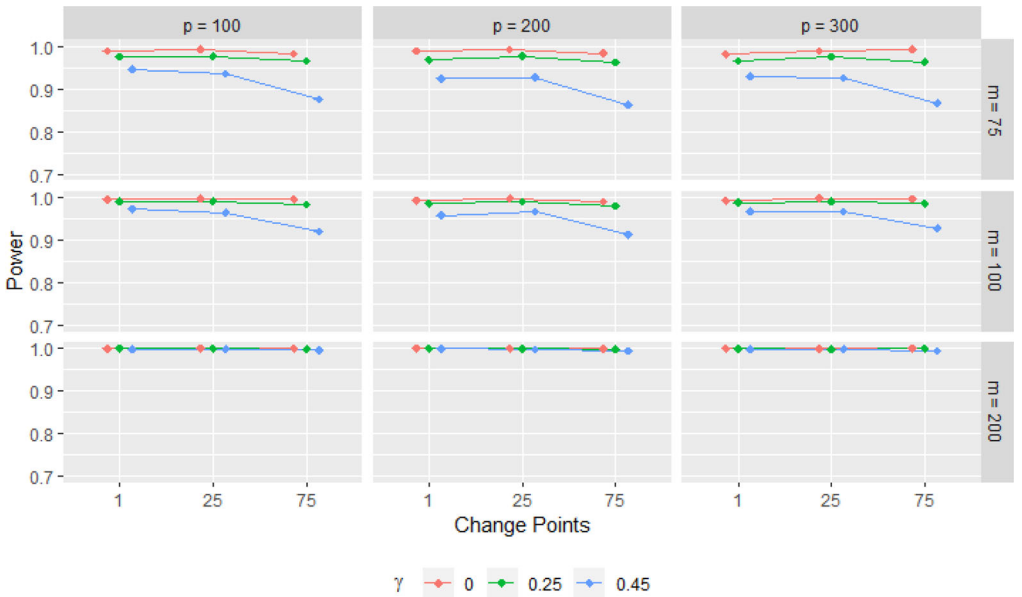

Figure 6. Power comparison for closed-ended procedure with $\alpha = 0.05$.

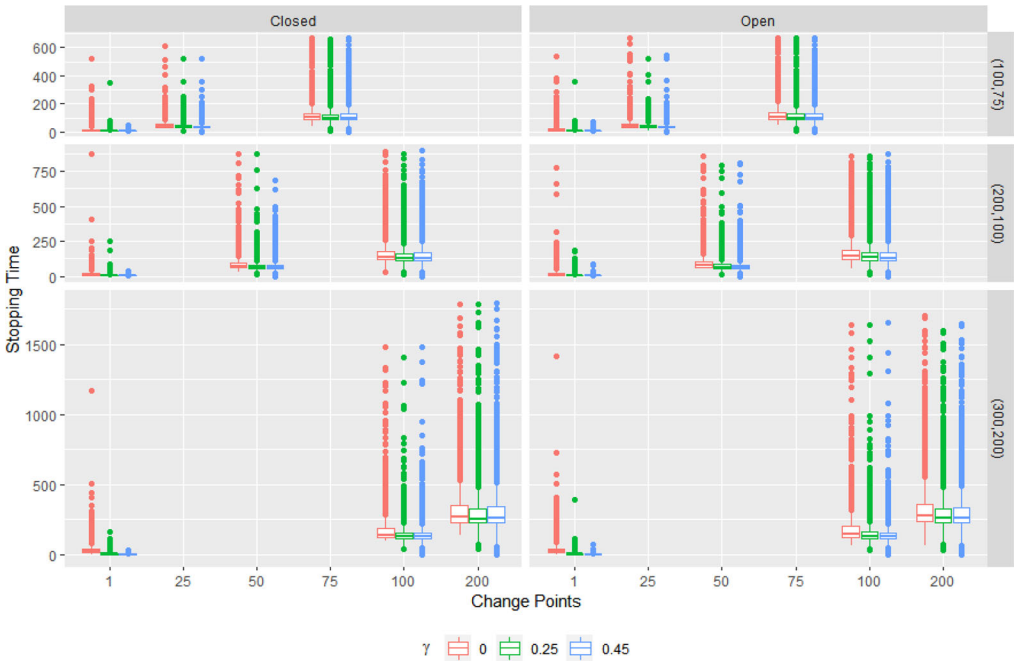

Figure 7. Stopping time comparison for closed- and open-ended procedures with $k^* \in \{1, 25, 100\}$, $\gamma \in \{0, 0.25, 0.45\}, \alpha = 0.05,$ and various historical sample sizes.

Table 9. Five-number summary for the detection time for the open-ended procedure at various change-point locations with $\gamma \in \{0, 0.25, 0.45\}$ and $\alpha = 0.05$.

	$(p, m) \rightarrow$	(100, 75)			(200, 100)			(300, 200)		
γ	Summary/k^*	1	25	75	1	50	100	1	100	200
0.00	min	2	26	44	2	33	33	3	101	151
	Q1	7	33	90	8	62	118	14	124	235
	Med	11	42	106	14	74	138	24	148	277
	Q3	20	58	139	26	99	182	45	201	369
	max	526	672	671	819	874	839	1,335	1,632	1,658
0.25	min	2	10	10	2	15	15	2	70	70
	Q1	3	30	86	3	58	114	3	115	228
	Med	4	35	99	4	65	130	6	131	260
	Q3	7	45	127	8	88	167	10	165	340
	max	354	522	664	251	761	897	166	1,433	1,653
0.45	min	1	1	1	2	1	1	2	2	2
	Q1	2	29	86	2	56	113	2	112	227
	Med	2	33	99	2	63	130	3	127	260
	Q3	4	41	130	4	78	169	4	156	343
	max	72	525	665	37	685	893	32	1,516	1,789

The power comparisons of the closed-ended procedure are given in Table 8. We monitor the process until $9\,m$ observations. Figure 6 compares the power of the closed-ended procedure. For large historical sample size m, the power is approximately equal to 1 regardless of the α level. The five-number summary of the detection time for both open- and closed-ended procedures is given in Tables 9–10 and the results are graphed in Figure 7. The results are based on 2,500 iterations. As we mentioned earlier, larger γ

Table 10. Five-number summary for the detection time for the closed-ended procedure at various change-point locations with $\gamma \in \{0, 0.25, 0.45\}$ and $\alpha = 0.05$.

	$(p, m) \rightarrow$	(100, 75)			(200, 100)			(300, 200)		
γ	Summary/k^*	1	25	75	1	50	100	1	100	200
0.00	min	2	26	42	2	30	30	3	101	135
	Q1	6	32	88	7	61	116	12	121	231
	Med	10	40	102	12	71	134	22	142	268
	Q3	18	54	132	23	94	173	41	187	350
	max	525	608	670	874	870	890	1,167	1,481	1,782
0.25	min	2	6	6	2	15	15	2	41	41
	Q1	3	29	85	3	57	113	3	114	226
	Med	4	34	97	4	64	128	5	129	255
	Q3	7	44	124	7	80	162	10	158	326
	max	353	521	671	250	874	876	166	1,385	1,784
0.45	min	1	1	1	1	1	1	2	2	2
	Q1	2	29	86	2	56	113	2	112	227
	Med	2	33	99	2	63	129	3	126	258
	Q3	4	41	128	4	78	167	4	155	341
	max	43	525	664	37	685	893	32	1,472	1,789

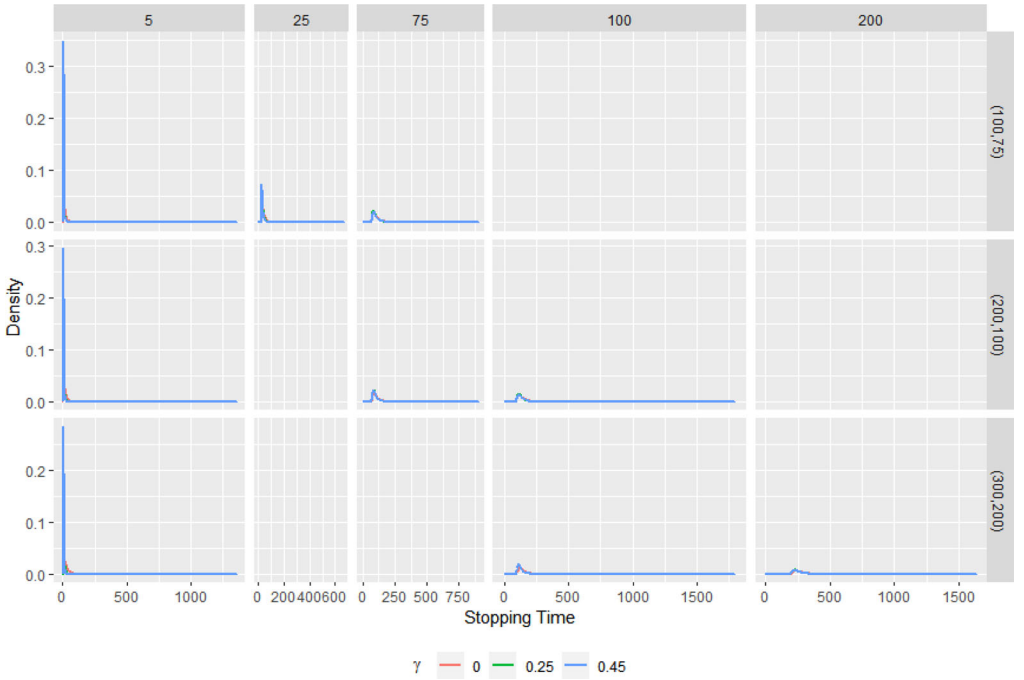

Figure 8. Estimated densities of the stopping time at various change points and various (p, m).

values result in detecting the change immediately when the change occurs shortly after the historical sample size m.

Figure 8 compares the estimated densities of the stopping time for various (p, m) and the estimated densities of the stopping time for various (p, m) and the change-point locations. We observe variation in density plots but decreases due to the large historical sample size.

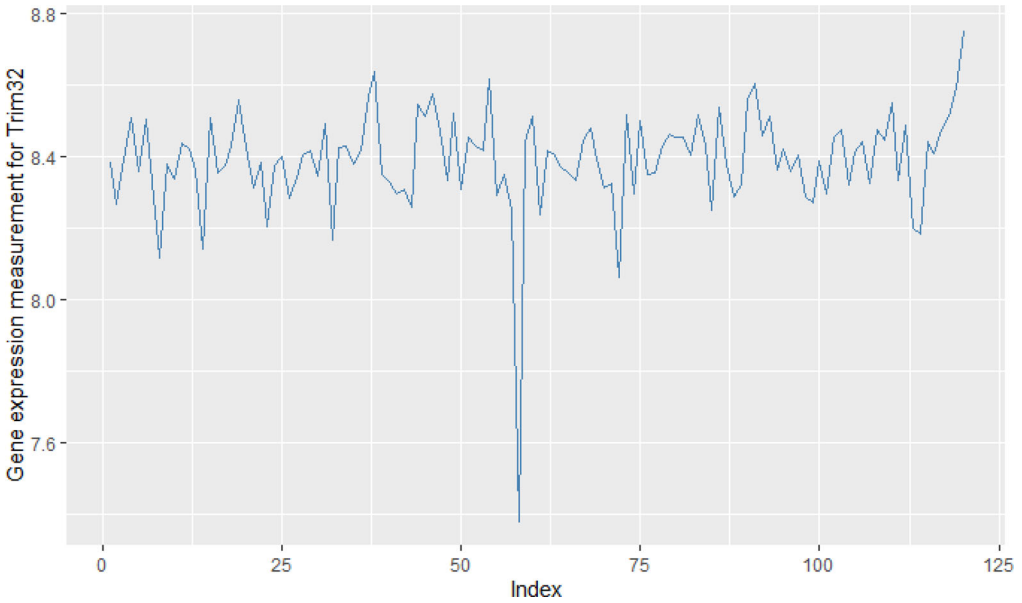

Figure 9. Outcome of the gene expression measurement for Trim32 in the mammalian eye data.

5. REAL DATA ANALYSIS

In this section, we apply the proposed method to a real data set. We consider gene expression in mammalian eye data. The data set was originally described in Scheetz et al. (2006). Laboratory rats were examined to learn about gene expression and regulation in the mammalian eye. Inbred rat strains were crossed and tissue was extracted from the eyes of 120 rats from the F2 generation. There are $n = 120$ observations and $p = 18{,}976$ explanatory variables. The outcome variable y is gene expression measurement for Trim32 and the explanatory variables are the gene expression measurements for the remaining genes. The outcome variable y is graphed in Figure 9.

According to the graph, there is no obvious jump in the data in the first 50 observations. Therefore, we consider the first 50 observations as historical data. The proposed method is applied for the control parameter value $\gamma = 0.45$ with $\alpha = 0.05$. The first change point detected after the historical sample size $m = 50$ for a given control parameter value 0.45 is 8. That is, there is a change of gene expression of 58th rat comparing to the first 57 rats. We also consider the change-point detection with the log-likelihood method by assuming the normality of the data under the fixed sample size $n = 120$ situation. The result confirms that the change occurred at 58th observation. With the binary segmentation method, there is only one change in the data. Compared to change-point detection with fixed sample size, the advantage of the sequential change-point detection method is that only a few samples are needed to make decisions. In this application, our method only requires 50 observations and our monitoring process stops after 58 samples. The traditional log-likelihood method, however, requires all observations ($n = 120$) to estimate the change location. The change-point location corresponding to $\gamma = 0.45$ is graphed in Figure 10.

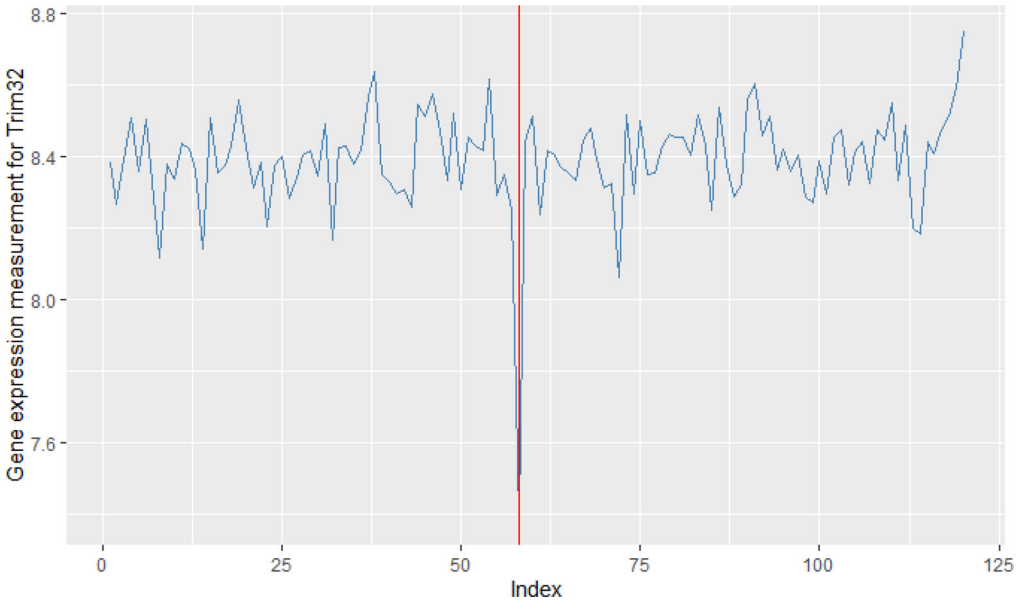

Figure 10. The first change point detected in the mammalian eye data.

6. CONCLUSION

In this article, we proposed test statistics to monitor the structural change for high-dimensional data using a SCAD penalized regression model for both open- and closed-ended procedures. The asymptotic critical values for both monitoring processes are obtained via simulation. The Type I error probability of the open-ended procedure is smaller than that of the closed-ended procedure. However, the Type I error probability improves when N increases. Thus, the closed-ended procedure is superior for small $N(< 20)$, and both closed- and open-ended procedures behave similarly for large $N(\geq 20)$. Moreover, the Type I error decreases when the historical sample size m increases. The power of the test is generally high and it is only affected by the size of a test. We computed the stopping time for both open- and closed-ended procedures. It can be seen that the larger control parameter values appear to detect the structural change much faster, whereas smaller control parameter values result in greater delays in detection. Similar to Horváth et al.'s (2004) conclusion, if a change occurs immediately after a historical sample size, we recommend a larger value of γ close to 0.5 for high-dimensional data. In comparison to an open-ended procedure, the monitoring process based on the closed-ended procedure is usually quicker. We estimated the density of the stopping time for various scenarios. We applied the proposed sequential detection procedure to analyze gene expression in the mammalian eye data to locate the change point sequentially. In the future, we would like to extend our approach using a nonparametric method. We will also compare the performance between parametric and nonparametric methods.

APPENDIX: PROOFS OF THEOREMS

We will show that Theorems 3.1–3.2 hold under the historical sample size m.

Proof of Theorem 3.1. We consider the following relation:

$$\sum_{i=m+1}^{m+k} \hat{\mathcal{E}}_i = \sum_{i=m+1}^{m+k} \mathcal{E}_i - \sum_{i=m+1}^{m+k} x_i^\top (\hat{\beta}_m^{SCAD} - \beta_0)\,. \tag{A.1}$$

Because the SCAD estimate $\hat{\beta}_{mj}^{SCAD}$ satisfies the oracle property by consistency in variable selection, under assumptions A1–A5 we have $\lim_{m\to\infty} P(\mathcal{A}^* = \mathcal{A}) = 1$. Therefore, $\lim_{m\to\infty} P(\mathcal{A}^* \cap \mathcal{A}^c = \emptyset) = 1$. Let $\mathbb{S} \equiv \mathcal{A}^* \cap \mathcal{A}$. By assumption A1 and sparsity property, we have that $(X_{\mathbb{S}}^\top X_{\mathbb{S}})^{-1} = \frac{1}{m}(C_m^{-1})_{\mathbb{S}}(1 + o_p(1))$, where $(C_m)_{\mathbb{S}}$ contains the elements of the matrix C_m with the index in the set $\mathbb{S}$.

Let $\Psi_{\mathbb{S}} = \frac{1}{2}(\ \mathrm{sign}(\hat{\beta}_{mj}^{SCAD}))_{j\in\mathbb{S}}$. In particular, $\Psi_{\mathbb{S}}$ is a column vector of dimension $Card(\mathbb{S})$. Considering the Karush-Kuhn-Tucker optimality conditions, for every $j \in \mathcal{A} \cap \mathcal{A}^*$, we have

$$\begin{aligned}
2X_j^\top (Y - X\hat{\beta}_m^{SCAD}) &= \lambda_m\ \mathrm{sign}(\hat{\beta}_{mj}^{SCAD}),\\
2X_j^\top (\mathcal{E} - X(\hat{\beta}_m^{SCAD} - \beta_0)) &= \lambda_m\ \mathrm{sign}(\hat{\beta}_{mj}^{SCAD}),\\
X_j^\top (\mathcal{E} - X(\hat{\beta}_m^{SCAD} - \beta_0)) &= \frac{1}{2}\lambda_m\ \mathrm{sign}(\hat{\beta}_{mj}^{SCAD}).
\end{aligned}$$

Let $\mathcal{A}^c$ and $\mathcal{A}^{*c}$ be the complementary sets of $\mathcal{A}^c$ and $\mathcal{A}^{*c}$, respectively. Following Ciuperca (2015), for every $j \in \mathbb{S}$, we get

$$X_j^\top (\mathcal{E} - X_{\mathbb{S}}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} - X_{\mathcal{A}^c\cap\mathcal{A}^*}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathcal{A}^c\cap\mathcal{A}^*} - X_{\mathcal{A}^c\cap\mathcal{A}^{*c}}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathcal{A}\cap\mathcal{A}^{*c}}) = \lambda_m \Psi_{\mathbb{S}}. \tag{A.2}$$

For any given $\epsilon > 0$ and large m, for the set $j \in \mathbb{S}$, we must have

$$P\left(X\left(\hat{\beta}_m^{SCAD} - \beta_0\right) = X_{\mathbb{S}}\left(\hat{\beta}_m^{SCAD} - \beta_0\right)_{\mathbb{S}}\right) > 1 - \epsilon, \tag{A.3}$$

Further, the relation (A.2) becomes

$$P\left(X_{\mathbb{S}}^\top\left(\mathcal{E} - X_{\mathbb{S}}\left(\hat{\beta}_m^{SCAD} - \beta_0\right)_{\mathbb{S}}\right) = \frac{1}{2}\lambda_m\ \mathrm{sign}\left(\hat{\beta}_{mj}^{SCAD}\right)\right) > 1 - \epsilon. \tag{A.4}$$

By assumption A2, the relation (A.4) implies that

$$P((\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = (X_{\mathbb{S}}^\top X_{\mathbb{S}})^{-1} X_{\mathbb{S}}^\top \mathcal{E} - \lambda_m (X_{\mathbb{S}}^\top X_{\mathbb{S}})^{-1} \Psi_{\mathbb{S}}) > 1 - \epsilon. \tag{A.5}$$

Let C_m be a positively definite matrix. Then, $(1/m)X_{\mathbb{S}}^\top X_{\mathbb{S}} \to C_m$. Using assumption A1 and by the central limit theorem, for every $j \in \mathbb{S}$ and for large m, we have

$$X_j^\top \mathcal{E} = O_p(m^{1/2}). \tag{A.6}$$

Following Y. Kim, Choi, and Oh (2008), we have

$$\begin{aligned}
\hat{\beta}_m^{SCAD} &= \frac{1}{m}(C_m^{-1})X^\top Y\\
&= \frac{1}{m}(C_m^{-1})X^\top(\mathcal{E} + X^\top\beta_0)\,.\\
&= \frac{1}{m}(C_m^{-1})X^\top\mathcal{E} + \beta_0
\end{aligned}$$

Thus, for the set $\mathbb{S}$,

$$(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = \frac{1}{m}(C_m^{-1})_{\mathbb{S}} X_{\mathbb{S}}^{\top} \mathcal{E}$$
$$\sqrt{m}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = \frac{1}{\sqrt{m}}(C_m^{-1})_{\mathbb{S}} X_{\mathbb{S}}^{\top} \mathcal{E}$$

Because $\mathcal{A} \neq \emptyset$ and by the oracle property, we have that

$$(q/m)^{1/2}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = O_p(1). \tag{A.7}$$

Thus, we get

$$(X_{\mathbb{S}}^{\top} X_{\mathbb{S}})^{-1} X_{\mathbb{S}}^{\top} \mathcal{E} = O_p(m^{-1/2}). \tag{A.8}$$

Thus, taking into account (A.5) and (A.6), we have

$$(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} = (X_{\mathbb{S}}^{\top} X_{\mathbb{S}})^{-1} X_{\mathbb{S}}^{\top} \mathcal{E}(1 + o_p(1)). \tag{A.9}$$

So,

$$\hat{\beta}_m^{SCAD} - \beta_0 = C_m^{-1} \frac{1}{m} \sum_{i=1}^{m} x_i \mathcal{E}_i (1 + o_p(1))$$

and we have

$$\begin{aligned} \sum_{i=m+1}^{m+k} \hat{\mathcal{E}}_i &= \sum_{i=m+1}^{m+k} (\mathcal{E}_i - x_i^{\top}(\hat{\beta}_m^{SCAD} - \beta_0)) \\ &= \sum_{i=m+1}^{m+k} \mathcal{E}_i - \left(\sum_{i=m+1}^{m+k} x_i \right)^{\top} C_m^{-1} \frac{1}{m} \sum_{j=1}^{m} x_j \mathcal{E}_j (1 + o_p(1)). \end{aligned}$$

Similarly, for all $k \geq 1$, we get

$$P\left(\sum_{i=m+1}^{m+k} x_i^{\top} \left(\hat{\beta}_m^{SCAD} - \beta_0 \right) = \sum_{i=m+1}^{m+k} x_{i,\mathbb{S}}^{\top} \left(\hat{\beta}_m^{SCAD} - \beta_0 \right)_{\mathbb{S}} \right) > 1 - \epsilon, \tag{A.10}$$

Putting together (A.5), (A.6), and (A.10) we obtain for the CUSUM of residuals

$$\begin{aligned} \sum_{i=m+1}^{m+k} \hat{\mathcal{E}}_i &= \sum_{i=m+1}^{m+k} (\mathcal{E}_i - x_{i.\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}}) \\ &= \sum_{i=m+1}^{m+k} \mathcal{E}_i - \sum_{i=m+1}^{m+k} x_{i,\mathbb{S}}^{\top}(\hat{\beta}_m^{SCAD} - \beta_0)_{\mathbb{S}} \\ &= \sum_{i=m+1}^{m+k} \mathcal{E}_i - \left(\sum_{i=m+1}^{m+k} x_{i,\mathbb{S}}^{\top} \right)(X_{\mathbb{S}}^{\top} X_{\mathbb{S}})^{-1} X_{\mathbb{S}}^{\top} \mathcal{E}(1 + o_p(1)). \end{aligned} \tag{A.11}$$

Now applying theorem 2.1 of Horváth et al. (2004) completes the proof. □

Proof of Theorem 3.2 . Let k^* be the time that the change truly occurs after a historical sample size of length m. Let $k' = m + k^*$. By the alternative hypothesis (3.3), we have

$$\begin{aligned}\sum_{i=m+1}^{m+k'}\hat{\mathcal{E}}_i &= \sum_{i=m+1}^{m+k'}\mathcal{E}_i + \sum_{i=m+1}^{m+k^*-1} x_{i,\mathbb{S}}^\top(\beta_0-\hat{\beta}_m^{SCAD})_{\mathbb{S}} + \sum_{i=m+k^*}^{m+k'} x_{i,\mathbb{S}}^\top(\beta_1-\hat{\beta}_m^{SCAD})_{\mathbb{S}},\\ &= \sum_{i=m+1}^{m+k'}\mathcal{E}_i + \sum_{i=m+1}^{m+k^*-1} x_{i,\mathbb{S}}^\top(\beta_0-\hat{\beta}_m^{SCAD})_{\mathbb{S}} + \sum_{i=m+k^*}^{m+k'} x_{i,\mathbb{S}}^\top\Big[(\beta_1-\beta_0)+(\beta_0-\hat{\beta}_m^{SCAD})\Big]_{\mathbb{S}}.\\ &= \sum_{i=m+1}^{m+k'}\mathcal{E}_i + \sum_{i=m+1}^{m+k'} x_{i,\mathbb{S}}^\top(\beta_0-\hat{\beta}_m^{SCAD})_{\mathbb{S}} + \sum_{i=m+k^*}^{m+k'} x_{i,\mathbb{S}}^\top(\beta_1-\beta_0)_{\mathbb{S}}\end{aligned}$$

But Theorem 3.1 yields that

$$\frac{\left|\sum_{i=m+1}^{m+k'}\hat{\mathcal{E}}_i + \left(\sum_{i=m+1}^{m+k'} x_{i,\mathbb{S}}\right)^\top(\beta_0-\hat{\beta}_m^{SCAD})_{\mathbb{S}}\right|}{g(m,k',\gamma)} = O_p(1).$$

Next we show that the relation (A.12) is true as $m\to\infty$.

$$\sup_{1\le k<\infty}\left|\sum_{i=m+1}^{m+k}\hat{\mathcal{E}}_i - \left(\sum_{i=m+1}^{m+k}\mathcal{E}_i - \frac{k}{m}\sum_{i=1}^{m}\mathcal{E}_i\right)\right| \Big/ m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma} = o_p(1). \tag{A.12}$$

Let c_1 be the first column of C. Combining (A.8), (A.9), and (A.10), we conclude

$$\begin{aligned}&\sup_{1\le k<\infty}\frac{1}{g(m,k,\gamma)}\left|\left\{\frac{1}{m}\left(\sum_{i=m+1}^{m+k} x_{i,\mathbb{S}}\right)^\top(C_m^{-1})_{\mathbb{S}} - \frac{k}{m}c_1^\top C_m^{-1}\right\}\sum_{j=1}^{m}x_j\mathcal{E}_j(1+o_p(1))\right|\\ &= O_p(m^{1/2})\sup_{1\le k\le\infty}\frac{1}{mg(m,k,\gamma)}\left\{\left(\frac{k}{m}\right)m^{1/2} + (k+m)^{1/2}\left(\frac{1}{m}\right) + m^{-1/2}\right\}\\ &= \sup_{1\le k\le m}\frac{(k/m)m^{-1/2}+(1+(k/m))m^{-1/2}}{(1+k/m)((k/m)/(1+k/m))^{\gamma}}\\ &= \sup_{1\le k\le m}\frac{(k/m)m^{-1/2}+(1+(k/m))m^{-1/2}}{(k/m)^{\gamma}(1+(k/m))^{1-\gamma}}\\ &\le \sup_{1\le k\le m}2^{\gamma}\left\{\left(\frac{k}{m}\right)^{1-\gamma}m^{-1/2} + \left(\frac{k}{m}\right)^{-\gamma}m^{-1/2}\right\}\\ &= \sup_{1\le k\le m}2^{\gamma}\{m^{-1/2}+m^{-1/2}\}\\ &= o(1)\quad\text{as}\quad m\to\infty.\end{aligned}$$

Similarly,

$$\sup_{m\le k\le\infty}\frac{(k/m)m^{-1/2}+(1+(k/m))m^{-1/2}}{(1+k/m)((k/m)/(1+k/m))^{\gamma}} = o(1).$$

Thus,

$$\sup_{1\le k<\infty}\frac{1}{g(m,k,\gamma)}\left|\left\{\frac{1}{m}\left(\sum_{i=m+1}^{m+k} x_{i,\mathbb{S}}\right)^\top(C_m^{-1})_{\mathbb{S}} - \frac{k}{m}c_1^\top C_m^{-1}\right\}\sum_{j=1}^{m}x_j\mathcal{E}_j(1+o_p(1))\right| = o_p(1)\quad\text{as}\quad m\to\infty.$$

Therefore, we have that

$$\left(\sum_{i=m+k^*}^{m+k'} x_{i,\mathbb{S}}\right)^\top(\beta_1-\beta_0)_{\mathbb{S}} = (k'-k^*)C_m^{(1,1)\top}(\beta_1-\beta_0)_{\mathbb{S}} + O((m+k^*)^{1/2}) + O((m+k')^{1/2}).$$

In particular, under alternative hypothesis (3.3), we have $|C_m^{(1,1)\top}(\beta_1 - \beta_0)| > 0$. Thus, we get

$$\liminf_{m\to\infty} \frac{\left|\left(\sum_{i=m+k^*}^{m+k'} x_{i,\mathbb{S}}\right)^{\top}(\beta_1 - \beta_0)_{\mathbb{S}}\right|}{g(m,k',\gamma)} > 0.$$

This completes the proof. □

ACKNOWLEDGEMENTS

The authors are grateful to the Editor-in-Chief, Professor Nitis Mukhopadhyay, for his helpful remarks on improving the layout of this article.

DISCLOSURE

The authors have no conflicts of interest to report.

FUNDING

The second author's research is partially supported by the Simons Foundation under grant 711800.

REFERENCES

Chen, H. 2019. "Sequential Change-Point Detection Based on Nearest Neighbors." *The Annals of Statistics* 47 (3):1381–407. doi:10.1214/18-AOS1718

Chi, Y., and Y. Wu. 2015. "Change-Point Estimation of High-Dimensional Streaming Data via Sketching." Paper presented at the 2015 49th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 102–106.

Chu, C.-S., M. Stinchcombe, and H. White. 1996. "Monitoring Structural Change." *Econometrica* 64 (5):1045–65. doi:10.2307/2171955

Ciuperca, G. 2014. "Model Selection by Lasso Methods in a Change-Point Model." *Statistical Papers* 55 (2):349–74. doi:10.1007/s00362-012-0482-x

Ciuperca, G. 2015. "Real Time Change-Point Detection in a Model by Adaptive Lasso and Cusum." *Journal de la Société Française de Statistique* 156 (4):113–32.

Fan, J., and R. Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties." *Journal of the American Statistical Association* 96 (456):1348–60. doi:10.1198/016214501753382273

Fan, J., and H. Peng. 2004. "Nonconcave Penalized Likelihood with a Diverging Number of Parameters." *The Annals of Statistics* 32 (3):928–61. doi:10.1214/009053604000000256

Harchaoui, Z., and C. Lévy-Leduc. 2010. "Multiple Change-Point Estimation with a Total Variation Penalty." *Journal of the American Statistical Association* 105 (492):1480–93. doi:10.1198/jasa.2010.tm09181

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*. New York, NY: Springer Series in Statistics Springer New York Inc.

Horváth, L., M. Huškova, P. Kokoszká, and J. Steinebach. 2004. "Monitoring Changes in Linear Models." *Journal of Statistical Planning and Inference* 126 (1):225–51. doi:10.1016/j.jspi.2003.07.014

Horváth, L., P. Kokoszka, and J. Steinebach. 2007. "On Sequential Detection of Parameter Changes in Linear Regression." *Statistics & Probability Letters* 77 (9):885–95. doi:10.1016/j.spl.2006.12.014

Kim, J., and H. J. Kim. 2008. "Asymptotic Results in Segmented Multiple Regression." *Journal of Multivariate Analysis* 99 (9):2016–38. doi:10.1016/j.jmva.2008.02.028

Kim, Y., H. Choi, and H. S. Oh. 2008. "Smoothly Clipped Absolute Deviation on High Dimensions." *Journal of the American Statistical Association* 103 (484):1665–73. doi:10.1198/016214508000001066

Knight, K., and W. Fu. 2000. "Asymptotics for Lasso-Type Estimators." *The Annals of Statistics* 28 (5):1356–78. doi:10.1214/aos/1015957397

Lorden, G. 1971. "Procedures for Reacting to a Change in Distribution." *The Annals of Mathematical Statistics* 42 (6):1897–908. doi:10.1214/aoms/1177693055

Page, E. S. 1954. "Continue Inspection Schemes." *Biometrika* 41 (1-2):100–35. doi:10.1093/biomet/41.1-2.100

Ratnasingam, S., and W. Ning. 2021. "Sequential Change Point Detection for High-Dimensional Data Using Nonconvex Penalized Quantile Regression." *Biometrical Journal. Biometrische Zeitschrift* 63 (3):575–98. doi:10.1002/bimj.202000078

Roberts, S. W. 1966. "A Comparison of Some Control Chart Procedures." *Technometrics* 8 (3): 411–30. doi:10.1080/00401706.1966.10490374

Scheetz, T., A.-Y. Kim, E. Swiderski, A. R. Philp, T. A. Braun, K. L. Knudtson, A. M. Dorrance, G. F. DiBon, J. Huang, T. L. Casavant, et al. 2006. "Regulation of Gene Expression in the Mammalian Eye and Its Relevance to Eye Disease." *Proceedings of the National Academy of Sciences* 103:14429–14434.

Shiryaev, A. N. 1963. "On Optimum Methods in Quickest Detection Problems." *Theory of Probability & Its Applications* 8 (1):22–46. doi:10.1137/1108002

Siegmund, D. 1985. *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer Science & Business Media.

Tartakovsky, A., I. Nikiforov, and M. Basseville. 2014. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. New York: Chapman and Hall/CRC.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1):267–88. doi:10.1111/j.2517-6161.1996.tb02080.x

Zhao, P., and B. Yu. 2006. "On Model Selection Consistency of Lasso." *Journal of Machine Learning Research* 7:2541–63.

Zhou, M., H. Wang, and Y. Tang. 2015. "Sequential Change Point Detection in Linear Quantile Regression Models." *Statistics & Probability Letters* 100:98–103. doi:10.1016/j.spl.2015.01.031

Zou, H. 2006. "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101 (476):1418–28. doi:10.1198/016214506000000735

Zou, H., and T. Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2):301–20. doi:10.1111/j.1467-9868.2005.00503.x