Journal de la Société Française de Statistique
Vol. 156 No. 4 (2015)

Real time change-point detection in a model by adaptive LASSO and CUSUM

Title: Real time change-point detection in a model by adaptive LASSO and CUSUM

Gabriela Ciuperca[1]

Abstract: In this paper, the CUSUM test statistic based on adaptive LASSO residuals is proposed and studied for detecting in real time a change-point in a linear model with a large number of explanatory variables.
Under null hypothesis that the model does not change, the asymptotic distribution of the test statistic is determined. Under alternative hypothesis that at some unknown observation there is a change in model, the proposed test statistic converges in probability to ∞. These results allow to build an asymptotic critical region. Next, in order to improve the test statistic performance a modified test statistic is proposed.
Simulation results, using Monte Carlo technique, illustrate the performance of the proposed test statistic. We also compare it with the classical CUSUM test statistic.

Résumé : Dans ce papier, la statistique de test CUSUM basée sur les résidus LASSO adaptatifs est proposée et étudiée pour détecter en temps réel si un changement a lieu dans un modèle linéaire qui a un nombre grand de variables explicatives.
Sous l'hypothèse nulle que le modèle ne subit pas de changements, la distribution asymptotique de la statistique de test est déterminée. Sous l'hypothèse alternative qu'un changement se produit dans le modèle à un instant inconnu, la statistique de test proposée converge en probabilité vers ∞. Ces résultats permettent la construction d'une zone de rejet asymptotique. Ensuite, pour améliorer la performance de la statistique de test on propose une statistique de test modifiée.
Les résultats des simulations, par Monte Carlo, montrent la performance de la statistique de test proposée en la comparant aussi avec la statistique de test CUSUM classique.

Mots-clés : sequential test, adaptive LASSO, CUSUM, asymptotic behaviour
Classification AMS 2000 : 62G10 , 62G20 , 62J05 , 62F35

1. Introduction

Often in practice, for example in genetics, we want to study the influence of a large number of explanatory variables (regressors) $X_1, \cdots, X_p$ on a variable Y, with the possibility that a small number among the p variables will effectively influence Y. On the other hand, we want to know in real time if after the first m observations, the model changes. The most used technique is the cumulative sum (CUSUM) method which provides a test statistic for testing if the model changes in real time. The reference paper for this method is Horváth, L. et al., (2004), where a monitoring scheme is proposed according to the least squares (LS) residuals and a boundary function, but which has the disadvantage that the hypothesis test works only when the number of model parameters is relatively small.

[1] UMR 5208, Institut Camille Jordan, Université Lyon 1.
E-mail : Gabriela.Ciuperca@univ-lyon1.fr

 ISSN: 2102-6238

In this paper, the real time change-point detection in a linear model with a large variable number is studied. We can not conceive a hypothesis test for detecting whether change occur in the model coefficients since existing test statistics were performed for p small. As we will see in the simulations part of this paper, the CUSUM test corresponding to LS residuals gives poor results for p large. In this case, a natural idea is to first select, on the segment without change point, the significant variables. In order to select, by hypothesis tests, the significant regressors which influence the variable Y we can use a stepwise selection method, for example Forward or Backward. But all these methods have the disadvantage that they have a large variability, that the final choice is not optimal and that the correlated explanatory variables remain in the final model (see Breiman, L., , 1996). These disadvantages are accentuated when the number of regressors is large. An important advance was made by Tibshirani, R., (1996), where, for a linear model, the estimation and automatic variable selection were simultaneously treated as a single minimization problem, by the least absolute shrinkage and selection operator (LASSO) method. Then, in order to detect a change in real time in model, one would then find a test statistic based on LASSO type residuals. This justifies the interest of this paper.

Since LASSO techniques are fairly recent, there are not many papers in the literature that address the change-point problem in this framework. All existing works consider the case of the *a posteriori* changes : the data are known at the end of the experiment and then, a posteriori we consider the question if model has changed. If the model has changed, we must find the number of changes, their location and then estimate the parameters of each segment (phase). In the paper Harchaoui, Z. and Lévy-Leduc, C., (2010), the estimation of the change-points location in one-dimensional piecewise constant for a white noise is reformulated as a variable selection problem with a L_1 penalty. In Ciuperca G., (2013), a quantile change-point model with SCAD estimator and a median model with LASSO-type estimator are studied. For a linear model with change-points, estimated by LS method with LASSO-type and adaptive LASSO penalties, Ciuperca G., (2014) realize a study a posteriori proposing a model selection criterion and estimating model parameters, change-points location simultaneously. Always for a LS model with number of explanatory variables much larger than the number of observations, Lee S. et al., (2012) considers a LASSO penalty when the errors have gaussian distribution.
To the author's knowledge, the sequential hypothesis test in a model with a large number of explicative variables, has not yet been considered in literature.

The sequential change-point detection in high-volume was considered by Wang H. et al., (2002), Siris V.A. and Papagalou F., (2006), Lung-Yut-Fong A. et al., (2012) in network traffic and not in linear model. More precisely, the CUSUM method is used in Wang H. et al., (2002), Siris V.A. and Papagalou F., (2006) for detecting a change in a time series. For the same type of problem in network traffic, another approach was proposed by Lung-Yut-Fong A. et al., (2012) for detecting the change-points in doubly censored time series.

The paper is organized as follows. In Section 2, we first give the model under the null and alternative hypothesis. Assumptions are also announced at the beginning of this section. Then, we construct a CUSUM test statistic based on adaptive LASSO residuals and we study its asymptotic behaviour under hypothesis H_0 and H_1. In order to improve the test statistic performance, a modified statistic is proposed. For detecting the change location in model, we propose a stopping time from which hypothesis H_0 is rejected. In Section 3, simulation results illustrate the performance of the proposed test statistic.

 ISSN: 2102-6238

2. CUSUM test with adaptive LASSO residuals

All throughout the paper, C denotes a positive generic constant which may take different values in different formula or even in different parts of the same formula. All vectors are column and $\mathbf{v}^t$ denotes the transposed of $\mathbf{v}$. All vectors and matrices are in bold. Concerning the used norms, for a q-vector $\mathbf{v} = (v_1, \cdots, v_q)$, let us denote by $\|\mathbf{v}\|_1 = \sum_{j=1}^{q} |v_j|$ its L_1-norm. For a matrix $\mathbf{M} = (a_{ij})_{\substack{1 \leqslant i \leqslant q_1 \\ 1 \leqslant j \leqslant q_2}}$, we denote by $\|\mathbf{M}\|_1 = \max_{j=1,\cdots,q_2}(\sum_{i=1}^{q_1} |a_{ij}|)$, the subordinate norm to the vector norm $\|.\|_1$. Let $\xrightarrow[m\to\infty]{\mathscr{L}}$, $\xrightarrow[m\to\infty]{I\!P}$ represent convergence in distribution and in probability, respectively, as $m \to \infty$.

We will also use the following notations : if V_n and U_n are random variable sequences, $V_n = o_{I\!P}(U_n)$ means that $\lim_{n\to\infty} I\!P[|U_n/V_n| > e] = 0$ for any $e > 0$, $V_n = O_{I\!P}(U_n)$ means that there exists a finite $c > 0$ such that $I\!P[|U_n/V_n| > c] < e$ for any n and e. If V_n and U_n are deterministic sequences, $V_n = o(U_n)$ means that the sequence $V_n/U_n \to 0$ for $n \to \infty$, $V_n = O(U_n)$ means that the sequence V_n/U_n is bounded for sufficiently large n.

2.1. *Variable selection in a model without change-points*

We consider on the first m observations, a classical model of linear regression :

$$Y_i = \mathbf{X}_i^t \beta + \varepsilon_i, \qquad i = 1, \cdots, m \tag{1}$$

with $\beta^0 = (\beta_{,1}^0, \cdots, \beta_{,p}^0)$ the true value (unknown) of β. The parameter β belongs to $\mathbb{R}^p$, with fixed dimension p (p doesn't depend on m), but with the possibility that p is very close to m.

For model (1), Y_i denotes the corresponding observation of response variable Y. The p-column vector $\mathbf{X}_i$ is the i-th observation for the vector $(X_1, \cdots, X_p)$ of the explanatory variables (regressors).

Let us denote by $\mathbb{X}$ the $m \times p$ matrix with the m observations of the variables $X_1, \cdots, X_p$. The following notations are used for the m-column vectors $\mathbf{Y} \equiv (Y_1, \cdots, Y_m)^t$ and $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \cdots, \varepsilon_m)^t$.

In order to estimate the parameter β, we could use the ordinary least squares method

$$\hat{\beta}_m \equiv \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{m} (Y_i - \mathbf{X}_i^t \beta)^2 = (\mathbb{X}^t \mathbb{X})^{-1} \mathbb{X}^t \mathbf{Y}. \tag{2}$$

We denote the components of this vector by $\hat{\beta}_m = (\hat{\beta}_{m,1}, \cdots, \hat{\beta}_{m,p})$, with $\hat{\beta}_{m,j}$ the j-th component of $\hat{\beta}_m$.

In order to select the explanatory variables which have a significant influence on the explained variable Y, classically, we can perform hypothesis tests, or apply criteria such as BIC or AIC. As noted in Introduction, all these methods are very unstable when p is large. One could then use the LASSO method, introduced by Tibshirani, R., (1996), whereby we can at the same time, under some conditions on $\mathbb{X}$, estimate the parameters and eliminate the irrelevant variables, without using hypothesis tests. For automatically selecting the variables, Zou, H., (2006) provides an adaptive LASSO method, without these conditions on the design matrix $\mathbb{X}$:

$$\hat{\beta}_m^* \equiv \arg\min_{\beta \in \mathbb{R}^p} [\sum_{i=1}^{m} (Y_i - \mathbf{X}_i^t \beta)^2 + \lambda_m \sum_{j=1}^{p} \hat{\omega}_j |\beta_{,j}|], \tag{3}$$

 ISSN: 2102-6238

with $\hat{\omega}_j \equiv |\hat{\beta}_{m,j}|^{-g}$, $\hat{\beta}_{m,j}$ the j-th component of the least squares estimator $\hat{\beta}_m$, $\beta_{,j}$ the j-th component of β and g a known positive parameter. The positive sequence (λ_m) is the tuning parameter, a regularization parameter, such that $\lambda_m \to \infty$ as $m \to \infty$.
Let us denote by $\hat{\varepsilon}_i^* \equiv Y_i - \mathbf{X}_i^t \hat{\beta}_m^*$, for $i = 1, \cdots, m$, the corresponding residual. To exemplify the properties of the adaptive LASSO estimator $\hat{\beta}_m^*$, let us consider the following two sets of index :
$\mathscr{A} \equiv \{j \in \{1, \cdots, p\}; \beta_{,j}^0 \neq 0\}$ the index set of nonzero components of the true parameter,
$\mathscr{A}_m^* \equiv \{j \in \{1, \cdots, p\}; \hat{\beta}_{m,j}^* \neq 0\}$ the set of index of nonzero components of the adaptive LASSO estimator.
Generally, for a parameter β, we denote by $\beta_{\mathscr{A}}$ the subvector of β containing the corresponding components of $\mathscr{A}$.

Let us consider the following assumptions.
First for errors ε_i :
(A1) $\varepsilon_1, \cdots, \varepsilon_m, \varepsilon_{m+1}, \cdots$, are i.i.d. $I\!E[\varepsilon_1] = 0$, $0 < \sigma^2 = Var(\varepsilon_1) < \infty$, $I\!E[|\varepsilon_1|^\nu] < \infty$ for some $\nu > 2$.
For the explanatory variables $X_1, \cdots, X_p$, there exists a positive definite matrix $\mathbf{C}$ and a constant $\eta > 0$ such that :
if they are deterministic :
(A2) $\|m^{-1} \sum_{i=1}^m \mathbf{X}_i \mathbf{X}_i^t - \mathbf{C}\|_1 = O(m^{-\eta})$,
and if they are random :
(A2bis) $\|m^{-1} \sum_{i=1}^m \mathbf{X}_i \mathbf{X}_i^t - \mathbf{C}\|_1 = O_{I\!P}(m^{-\eta})$ a.s. Moreover, ε_i and $\mathbf{X}_i$ are independent.

Assumptions (A1), (A2) and (A2bis) are standard conditions which are used in large-dimensional linear model or in sequential change-point test, see for example Zou, H., (2006), Horváth, L. et al., (2004), Ciuperca G., (2014).

Under the condition that, the errors ε_i are i.i.d. random variables with mean zero and bounded variance σ^2 (assumption (A1)), $Card(\mathscr{A})$ doesn't depend on m, the design matrix is such that $m^{-1}\mathbb{X}^t\mathbb{X}$ converges to a positive definite matrix, and the tuning parameter (λ_m) is such that, as $m \to \infty$,

$$m^{-1/2}\lambda_m \to 0, \quad m^{(g-1)/2}\lambda_m \to \infty, \tag{4}$$

we have that adaptive LASSO estimator satisfies the *oracle properties* (see Zou, H., 2006), i.e., that :
(i) asymptotic normality : $\sqrt{m}(\hat{\beta}_m^* - \beta^0)_{\mathscr{A}} \xrightarrow[m\to\infty]{\mathscr{L}} \mathscr{N}(\mathbf{0}, \sigma^2(\mathbf{C}_{\mathscr{A}})^{-1})$.
(ii) sparsity property : $\lim_{m\to\infty} I\!P[\mathscr{A}_m^* = \mathscr{A}] = 1$.
The matrix $\mathbf{C}_{\mathscr{A}}$ contains the elements of the matrix $\mathbf{C}$ with the index in the set $\mathscr{A}$.

Recall that for $\hat{\beta}_m$, the least squares estimator, we have, $I\!P[\hat{\beta}_{m,\mathscr{A}} = \mathbf{0}] = 0$.

For σ^2 error variance, we will consider the following estimator :

$$\hat{\sigma}_m^{*2} \equiv \frac{1}{m - Card(\mathscr{A}_m^*)} \sum_{i=1}^m (Y_i - \mathbf{X}_i^t \hat{\beta}_m^*)^2. \tag{5}$$

 ISSN: 2102-6238

Since $Card(\mathscr{A})$ doesn't depend on m, taking into account the oracle properties of the adaptive LASSO estimator $\hat{\beta}^*_m$, we obtain that

$$\begin{aligned}\hat{\sigma}^{*2}_m &= \frac{1}{m-Card(\mathscr{A}^*_m)}\sum_{i=1}^{m}(\varepsilon_i^2 - 2\varepsilon_i \mathbf{X}_i^t(\hat{\beta}^*_m-\beta^0)+(\hat{\beta}^*_m-\beta^0)^t\mathbf{X}_i\mathbf{X}_i^t(\hat{\beta}^*_m-\beta^0))\\ &= \left(\frac{1}{m-Card(\mathscr{A}^*_m)}\sum_{i=1}^{m}\varepsilon_i^2\right)(1+o_{I\!P}(1)) \xrightarrow[m\to\infty]{I\!P} \sigma^2.\end{aligned}$$

2.2. *Sequential test for change detection in model*

Once the significant explanatory variables were selected, we continue to observe the variables $Y, X_1, \cdots, X_p$ after observation m, with the dependent variable Y also modeled by a linear model :

$$Y_i = \mathbf{X}_i^t\beta_i + \varepsilon_i, \qquad i = m+1, m+2, \cdots$$

At each time i, we test if we have the same model as the m first observations :
$H_0 : \beta_i = \beta^0$ for all $i = m+1, m+2, \cdots$
against the alternative hypothesis, that at some (unknown) observation k^0 there is a change, called also change-point, in model :

$$H_1 : \exists k^0 \geq 1 \text{ such that } \begin{cases} \beta_i = \beta^0, & i = m+1, \cdots, m+k^0 \\ \beta_i = \beta^0_2, & i = m+k^0+1, \cdots \end{cases}$$

with $\beta^0 \neq \beta^0_2$.
The parameters β^0, β^0_2 and the change-point k^0 are unknown.
We denote the components of β^0_2 by $(\beta^0_{2,1}, \cdots, \beta^0_{2,p})$. Recall that the components of β^0 are $\beta^0_{,j}$ for $j \in \{1, \cdots, p\}$.
We assume that model (1) is significant, i.e. at least one of the regressors affects significantly to the response variable Y :

$$\exists j \in \{1, \cdots, p\} \quad \text{such that } \beta^0_{,j} \neq 0. \tag{6}$$

In order to find a test statistic, we will consider the CUSUM test method of Horváth, L. et al., (2004), first for adaptive LASSO residuals. Then, we propose another test statistic which improves the type I error probability (size of test). The principle is to assume that there is the same model in each observation $m+k$, with $k \geq 1$, and to calculate the sum of the corresponding residuals $\hat{\varepsilon}^*_i = Y_i - \mathbf{X}_i\hat{\beta}^*_m$, for $i = m+1, \cdots, m+k$.
The cumulative sum (CUSUM) of $\hat{\varepsilon}^*_i$, for some $k \geq 1$, is $\sum_{i=m+1}^{m+k}\hat{\varepsilon}^*_i$.

For a given constant $\gamma \in [0, 1/2)$, let us consider the following normalisation function (boundary function) :

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

Let us denote by $\mathbf{X}^j$ the m-column vector with the observations $X_{j1}, \cdots, X_{jm}$ of the variable X_j for the first m observations. We use the notation $sgn(.)$ for the sign function.

 ISSN: 2102-6238

Theorem 1. *Under assumptions (A1), (A2) or (A2bis), (4) and (6), we have :*
1) If hypothesis H_0 holds, then we have for any real $c>0$:

$$\lim_{m\to\infty} I\!P\left[\frac{1}{\hat{\sigma}_m^*}\sup_{1\le k<\infty}\mid\sum_{i=m+1}^{m+k}\hat{\varepsilon}_i^*\mid/g(m,k,\gamma)\le c\right]=I\!P\left[\sup_{0\le t\le 1}\frac{|W(t)|}{t^\gamma}\le c\right],$$

with $\{W(t);0\le t<\infty\}$ a Wiener process.
2) If hypothesis H_1 holds, then

$$\frac{1}{\hat{\sigma}_m^*}\sup_{1\le k<\infty}\mid\sum_{i=m+1}^{m+k}\hat{\varepsilon}_i^*\mid/g(m,k,\gamma)\xrightarrow[m\to\infty]{I\!P}\infty.$$

Proof of Theorem 1. *1)* We have the obvious relation

$$\sum_{i=m+1}^{m+k}\hat{\varepsilon}_i^*=\sum_{i=m+1}^{m+k}\varepsilon_i-\sum_{i=m+1}^{m+k}\mathbf{X}_i^t(\hat{\beta}_m^*-\beta^0). \tag{7}$$

On the other hand, taking into account assumption (6) and KKT optimality conditions, we have for every $j\in\mathscr{A}\cap\mathscr{A}_m^*$, with probability one, that :

$$2\mathbf{X}^{j\,t}(\varepsilon-\mathbb{X}(\hat{\beta}_m^*-\beta^0))=\lambda_m\hat{\omega}_j sgn(\hat{\beta}_{m,j}^*),$$

We denote by $\mathscr{J}\equiv\mathscr{A}\cap\mathscr{A}_m^*$. Then, the last relation can be written under the form, for every $j\in\mathscr{J}$

$$2\mathbf{X}^{j\,t}\left(\varepsilon-\mathbb{X}_{\mathscr{F}}(\hat{\beta}_m^*-\beta^0)_{\mathscr{F}}-\mathbb{X}_{\mathscr{A}^c\cap\mathscr{A}_m^*}(\hat{\beta}_m^*-\beta^0)_{\mathscr{A}^c\cap\mathscr{A}_m^*}-\mathbb{X}_{\mathscr{A}\cap\mathscr{A}_m^{*c}}(\hat{\beta}_m^*-\beta^0)_{\mathscr{A}^c\cap\mathscr{A}_m^{*c}}\right)=\lambda_m\hat{\omega}_j sgn(\hat{\beta}_{m,j}^*), \tag{8}$$

with $\mathscr{A}^c$, $\mathscr{A}_m^{*c}$ the complementaries of the sets $\mathscr{A}$, $\mathscr{A}_m^*$, respectively. Relation (8) can be written in matrix form as

$$\mathbb{X}_{\mathscr{J}}^t\left(\varepsilon-\mathbb{X}_{\mathscr{J}}(\hat{\beta}_m^*-\beta^0)_{\mathscr{J}}-\mathbb{X}_{\mathscr{A}^c\cap\mathscr{A}_m^*}(\hat{\beta}_m^*-\beta^0)_{\mathscr{A}^c\cap\mathscr{A}_m^*}-\mathbb{X}_{\mathscr{A}\cap\mathscr{A}_m^{*c}}(\hat{\beta}_m^*-\beta^0)_{\mathscr{A}\cap\mathscr{A}_m^{*c}}\right)=\lambda_m\mathbf{W}_{\mathscr{F}}, \tag{9}$$

with $\mathbf{W}_{\mathscr{F}}$ a column vector of dimension $Card(\mathscr{F})$, defined by :

$$\mathbf{W}_{\mathscr{F}}\equiv\frac{1}{2}\left(\hat{\omega}_j sgn(\hat{\beta}_{m,j}^*)\right)_{j\in\mathscr{F}}.$$

and $\mathbb{X}_{\mathscr{F}}$ the submatrix $m\times Card(\mathscr{F})$ of $\mathbb{X}$, $(\hat{\beta}_m^*-\beta^0)_{\mathscr{F}}$ the $Card(\mathscr{F})$-subvector of $(\hat{\beta}_m^*-\beta^0)$ containing the corresponding components of $\mathscr{F}$. Similar notation is used for the index sets $\mathscr{A}^c\cap\mathscr{A}_m^*$, $\mathscr{A}\cap\mathscr{A}_m^{*c}$.
By sparsity property *(ii)* for $\hat{\beta}_m^*$, given in subsection 2.1, we have that $\lim_{m\to\infty}I\!P[\mathscr{A}^c\cap\mathscr{A}_m^*=\emptyset]=1$. Then, for all $e>0$, there exists $m_e\in\mathbb{N}$ such that for any $m\ge m_e$, we have $I\!P[(\hat{\beta}_m^*-\beta^0)_{\mathscr{A}^c\cap\mathscr{A}_m^*}=\mathbf{0}]>1-e$. Thus, for any $e>0$ and enough large m, we have $I\!P[\mathbb{X}_{\mathscr{A}^c\cap\mathscr{A}_m^*}(\hat{\beta}_m^*-\beta^0)_{\mathscr{A}^c\cap\mathscr{A}_m^*}=\mathbf{0}]>1-e$. We prove similarly that for any $e>0$ and enough large m, we have $I\!P[\mathbb{X}_{\mathscr{A}\cap\mathscr{A}_m^{*c}}(\hat{\beta}_m^*-\beta^0)_{\mathscr{A}\cap\mathscr{A}_m^{*c}}=\mathbf{0}]>1-e$. Then, for any $e>0$ and enough large m, we have

$$I\!P[\mathbb{X}(\hat{\beta}_m^*-\beta^0)=\mathbb{X}_{\mathscr{F}}(\hat{\beta}_m^*-\beta^0)_{\mathscr{F}}]>1-2e \tag{10}$$

 ISSN: 2102-6238

and relation (9) becomes, for any $e > 0$,

$$I\!P\left[\mathbb{X}^t_{\mathscr{J}}\left(\varepsilon - \mathbb{X}_{\mathscr{J}}(\hat{\beta}^*_m - \beta^0)_{\mathscr{J}}\right) = \lambda_m \mathbf{W}_{\mathscr{F}}\right] > 1 - 2e. \tag{11}$$

By assumption (A2), respectively (A2bis), we have that $\mathbb{X}^t_{\mathscr{J}}\mathbb{X}_{\mathscr{J}}$ is of rank $Card(\mathscr{J})$, then it is invertible. Thus, relation (11) implies that, for any $e > 0$ and enough large m,

$$I\!P\left[(\hat{\beta}^*_m - \beta^0)_{\mathscr{F}} = (\mathbb{X}^t_{\mathscr{F}}\mathbb{X}_{\mathscr{F}})^{-1}\mathbb{X}^t_{\mathscr{F}}\varepsilon - \lambda_m(\mathbb{X}^t_{\mathscr{F}}\mathbb{X}_{\mathscr{F}})^{-1}\mathbf{W}_{\mathscr{F}}\right] > 1 - 2e. \tag{12}$$

Using assumptions (A1) and (A2), respectively (A2bis), by the central limit theorem, we have for all $j \in \mathscr{F}$, that $\mathbf{X}^{j^t}\varepsilon = O_{I\!P}(m^{1/2})$. Then

$$(\mathbb{X}^t_{\mathscr{F}}\mathbb{X}_{\mathscr{F}})^{-1}\mathbb{X}^t_{\mathscr{F}}\varepsilon = O_{I\!P}(m^{-1/2}). \tag{13}$$

By assumption (A2) (or (A2bis)), the sparsity property $I\!P[\mathscr{F} = \mathscr{A}] \to 1$, as $m \to \infty$, we have that $(\mathbb{X}^t_{\mathscr{J}}\mathbb{X}_{\mathscr{J}})^{-1} = \frac{1}{m}\mathbf{C}^{-1}_{\mathscr{J}}(1 + o_{I\!P}(1))$.
Since $j \in \mathscr{A}$, we have that the random variable $\hat{\omega}_j$ is bounded with probability converging to 1 as $m \to \infty$. Moreover, since $\lambda_m = o(m^{1/2})$, we have that

$$\lambda_m \frac{\mathbf{W}_{\mathscr{F}}}{m} = o_{I\!P}(m^{-1/2}). \tag{14}$$

But, since $\mathscr{A} \neq \emptyset$ and by property oracle *(i)*, given in subsection 2.1, we have that $(\hat{\beta}^*_m - \beta^0)_{\mathscr{F}}$ converges to zero with the rate $m^{-1/2}$. Thus, taking into account of (12), (13) and (14), we have

$$(\hat{\beta}^*_m - \beta^0)_{\mathscr{F}} = (\mathbb{X}^t_{\mathscr{F}}\mathbb{X}_{\mathscr{F}})^{-1}\mathbb{X}^t_{\mathscr{F}}\varepsilon(1 + o_{I\!P}(1)). \tag{15}$$

We prove similar to (10) that, for any $e > 0$, $k \geq 1$, we have

$$I\!P[\sum_{i=m+1}^{m+k}\mathbf{X}^t_i(\hat{\beta}^*_m - \beta^0) = \sum_{i=m+1}^{m+k}\mathbf{X}^t_{i,\mathscr{F}}(\hat{\beta}^*_m - \beta^0)_{\mathscr{F}}] > 1 - 2e. \tag{16}$$

Then, taking into account (15), (16), we obtain that relation (7) becomes

$$\sum_{i=m+1}^{m+k}\hat{\varepsilon}^*_i = \sum_{i=m+1}^{m+k}\varepsilon_i - \left(\sum_{i=m+1}^{m+k}\mathbf{X}^t_{i,\mathscr{F}}\right)(\mathbb{X}^t_{\mathscr{F}}\mathbb{X}_{\mathscr{F}})^{-1}\mathbb{X}^t_{\mathscr{F}}\varepsilon(1 + o_{I\!P}(1)). \tag{17}$$

Since the CUSUM of $\hat{\varepsilon}^*_i$ can be written under form (17) and the CUSUM of LS residuals on $\mathscr{F}$ is $\sum_{i=m+1}^{m+k}\varepsilon_i - \left(\sum_{i=m+1}^{m+k}\mathbf{X}^t_{i,\mathscr{F}}\right)(\mathbb{X}^t_{\mathscr{F}}\mathbb{X}_{\mathscr{F}})^{-1}\mathbb{X}^t_{\mathscr{F}}\varepsilon$, we can then apply Theorem 2.1 of Horváth, L. et al., (2004).
2) Taking into account *1)*, we argue exactly as in Horváth, L. et al., (2004). ■

The result of Theorem 1 allows us to determine the asymptotic distribution of the test statistic under the hypothesis H_0, using adaptive LASSO residuals $\hat{\varepsilon}^*_i$ for CUSUM technique. Under hypothesis H_1, the test statistic converges in probability to ∞. Then it can be used for testing H_0

 ISSN: 2102-6238

against H_1, asymptotic reject region being $\sup_{0\leq t\leq 1}\frac{|W(t)|}{t^\gamma}\geq c_\alpha(\gamma)$, with $c_\alpha(\gamma)$, the $(1-\alpha)$ of the distribution

$$\sup_{0\leq t\leq 1}\frac{|W(t)|}{t^\gamma}, \tag{18}$$

for a size $\alpha\in(0,1)$ and fixed parameter $\gamma\in[0,1/2)$.

Corollary 1. *Theorem 1 implies that the hypothesis H_0 is rejected, for a fixed size $\alpha\in(0,1)$, in the point, called stopping time,*

$$\hat{k}_m^*\equiv\begin{cases}\inf\{k\geq 1;\quad (\hat{\sigma}_m^*)^{-1}\left|\sum_{i=m+1}^{m+k}\frac{\hat{\varepsilon}_i^*}{g(m,k,\gamma)}\right|>c_\alpha(\gamma),\\ \infty,\qquad \textit{otherwise.}\end{cases} \tag{19}$$

Remark 1. Assumption (A1) is essential to find by Theorem 1 the asymptotic distribution of CUSUM test statistic. Then, it makes no sense to estimate parameters β by another adaptive LASSO method making the variable selection, for example of type quantile, for which this assumption is not required. Thus, since we need assumption (A1), the better it is to estimate β by LS framework with an adaptive LASSO penalty.

The order $o_{I\!P}(1)$ in relation (14) can be quite large, moreover, the adaptive LASSO estimator $(\hat{\beta}_m^*)_{\mathscr{A}}$ can be biased (it is asymptotically unbiased), then, the direct application of this theorem for the hypothesis tests can lead to bad practical results (as we will see by simulations), especially under hypothesis H_0. Then, in order to improve the test statistic performance we will take into account the properties of the least-squares method and of the adaptive LASSO. The new technique, which we'll call *modified adaptive LASSO* be carried out in two steps.

BEGIN

Step I. The adaptive LASSO technique is realized on the observations $1,\cdots,m$. Following this step, since the adaptive LASSO performs consistent variable selection (sparsity property), we will consider the index set $\mathscr{A}_m^*$ for which the adaptive LASSO estimators are different from zero.

Step II. In this step we will re-estimate the nonzero parameters obtained to Step I. For this purpose, consider the set of parameters :

$$\Gamma_m^*\equiv\left\{\beta\in\mathbb{R}^p;\beta_{j\notin\mathscr{A}_m^*}=0\right\}.$$

Let us estimate $\beta\in\Gamma_m^*$ by least squares method :

$$\tilde{\beta}_m^*\equiv\arg\min_{\beta\in\Gamma_m^*}\sum_{i=1}^m(Y_i-\mathbf{X}_i^t\beta)^2$$

and there will be LS residual :

$$\tilde{\varepsilon}_i^*\equiv Y_i-\mathbf{X}_i^t\tilde{\beta}_m^*,\qquad i=1,\cdots,m.$$

 ISSN: 2102-6238

Let us consider a convergent estimator for the variance σ^2

$$(\tilde{\sigma}_m^*)^2 \equiv \frac{1}{m - Card(\mathscr{A}_m^*)} \sum_{i=1}^{m} (\tilde{\varepsilon}_i^*)^2.$$

Taking into account sparsity property *(ii)*, we obtain that this procedure provides, with probability converging to 1, unbiased estimators for $\beta_{\mathscr{A}}$ and the estimators directly equal to zero for $\beta_{\mathscr{A}^c}$. On the other hand, Theorem 1 remains valid considering $\tilde{\varepsilon}_i^* = Y_i - \mathbf{X}_i^t \tilde{\beta}_m^*$, for $i = m+1, \cdots, m+k$, as residuals and $\tilde{\sigma}_m^*$ as estimator of error standard-deviation σ. The asymptotic distribution of the test statistic is that given by (18). The stopping time in this case is

$$\tilde{k}_m^* \equiv \begin{cases} \inf\{k \geq 1; \quad (\tilde{\sigma}_m^*)^{-1} \left| \sum_{i=m+1}^{m+k} \frac{\tilde{\varepsilon}_i^*}{g(m,k,\gamma)} \right| > c_\alpha(\gamma), \\ \infty, \qquad \text{otherwise.} \end{cases} \tag{20}$$

END

Obviously, we have for $\tilde{k}_m^*$, for a fixed size $\alpha \in (0,1)$, under hypothesis H_0 :

$$\lim_{m \to \infty} I\!P[\tilde{k}_m^* < \infty] = \alpha$$

and under hypothesis H_1, the asymptotic power is one :

$$\lim_{m \to \infty} I\!P[\tilde{k}_m^* < \infty] = 1.$$

Recall that the CUSUM test statistic of Horváth, L. et al., (2004) is

$$\frac{1}{\hat{\sigma}_m} \sup_{1 \leq k < \infty} \left| \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i \right| / g(m,k,\gamma), \tag{21}$$

with the residuals $\hat{\varepsilon}_i = Y_i - \mathbf{X}_i^t \hat{\beta}_m$, the LS estimator $\hat{\beta}_m$ given by (2) and the consistent estimator of σ^2 :

$$\hat{\sigma}_m^2 = \frac{1}{m-p} \sum_{i=1}^{m} (Y_i - \mathbf{X}_i^t \hat{\beta}_m)^2.$$

This test statistic satisfies an equivalent result of Theorem 1, i.e. that, under H_0 it converges in distribution to (18) and under H_1 converges in probability to ∞.
The stopping time for this test statistic is :

$$\hat{k}_m \equiv \begin{cases} \inf\{k \geq 1; \quad (\hat{\sigma}_m)^{-1} \left| \sum_{i=m+1}^{m+k} \frac{\hat{\varepsilon}_i}{g(m,k,\gamma)} \right| > c_\alpha(\gamma), \\ \infty, \qquad \text{otherwise.} \end{cases} \tag{22}$$

3. Numerical study

In this section, Monte Carlo simulation studies are carried out to assess the performance of the proposed test statistics. Three CUSUM statistic tests are compared, corresponding to residuals :

 ISSN: 2102-6238

– of least squares :

$$(\hat{\sigma}_m)^{-1} \sup_{k\geq 1} \left| \sum_{i=m+1}^{m+k} \frac{\hat{\varepsilon}_i}{g(m,k,\gamma)} \right|,$$

– of adaptive LASSO

$$(\hat{\sigma}^*_m)^{-1} \sup_{k\geq 1} \left| \sum_{i=m+1}^{m+k} \frac{\hat{\varepsilon}^*_i}{g(m,k,\gamma)} \right|,$$

– of modified adaptive LASSO

$$(\tilde{\sigma}^*_m)^{-1} \sup_{k\geq 1} \left| \sum_{i=m+1}^{m+k} \frac{\tilde{\varepsilon}^*_i}{g(m,k,\gamma)} \right|.$$

The asymptotic distributions of the three test statistics follow same law (18).

All simulations were performed using the R language, packages *sde, lqa, MASS*.

As was noted in subsection 2.1, so that adaptive LASSO estimator (3) satisfies the oracle properties it is necessary that, tuning parameter λ_m and power g which are involved in weight $\hat{\omega}_j$ satisfy conditions : $m^{-1/2}\lambda_m \to 0$, $m^{(g-1)/2}\lambda_m \to \infty$, as $m \to \infty$. Then, in all simulations, we consider the following values for g and λ_m : $g = 1/5$ and $\lambda_m = m^{9/20}$. The model errors are Gaussian $\varepsilon_i \sim \mathcal{N}(0,1)$. For the explanatory variables $X_1, \cdots, X_p$ we will consider two distributions.
We denote by T the number of observations considered after the observation m.
For each model we generated 500 Monte Carlo random samples of size $m+T$.

3.1. Empirical size and power

Concerning number p of explanatory variables, we first consider a small number and we compare the results obtained by the three CUSUM methods. Based on these results, we will consider the two CUSUM best methods in order to detect a change in models with a large number of variables.

3.1.1. Small number of variables

In this subsection, we consider that the number of explanatory variables is 10 (p=10).
For model (1), the true values of the components of β^0 are $\beta^0_{,3} = 5$ and all others are 0 : $\mathcal{A} = \{3\}$.
Under the hypothesis H_1, the vector parameter β^0_2 has the fourth component $\beta^0_{2,4} = 15$ and the other components are 0.
For the regressors $X_1, \cdots, X_{10}$ we consider two distributions. The first law, denoted $\mathscr{L}_1$, is such that, all X_j, for $j \in \{1, \cdots, 10\} \setminus \{3\}$, have standard normal distribution $\mathcal{N}(0,1)$ and $X_3 \sim \mathcal{N}(2,1)$. A second distribution for the regressors, denoted $\mathscr{L}_2$, is obtained as follows : for all $j \in \{1, \cdots, 10\}$ the j-th regressor is $X_j^2 + j^2/m$, with X_j random variables of distribution $\mathscr{L}_1$, for all $j \neq 3$ and $X_3 \sim \mathcal{N}(2,1)$.

On the parameter γ of the normalization function $g(m,k,\gamma)$, the simulations of Horváth, L. et al., (2004) for detection of parameter changes in linear regression, showed that if the change

 ISSN: 2102-6238

occurs shortly after observation m, then γ must be chosen as close to $1/2$. But, if γ is too close to $1/2$, the CUSUM method could lead to false detection. That is why, we will vary the value of γ : considering $\gamma \in \{0, 0.25, 0.49\}$. We also vary the number of observations m, considering $m \in \{25, 100\}$, and the size $\alpha \in \{0.025, 0.05, 0.10\}$, for studying the CUSUM test statistic corresponding to residuals of least squares, adaptive LASSO and modified adaptive LASSO.

Under hypothesis H_1, we also vary the position of the change-point k^0 after m : $k^0 \in \{10, 25\}$.

TABLE 1. *Empirical test size* $\hat{\alpha}$ *and power* $\hat{\pi}$ *for* $p = 10$.

Law	k^0	(m,T)	method	$\hat{\alpha},\hat{\pi}$	$\gamma=0$			$\gamma=0.25$			$\gamma=0.49$		
					0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
$\mathscr{L}_1$	10	(25,25)	LS	$\hat{\alpha}$	0.02	0.03	0.07	0.05	0.08	0.14	0.10	0.13	0.20
				$\hat{\pi}$	0.99	1	1	1	1	1	1	1	1
			aLASSO	$\hat{\alpha}$	0.60	0.70	0.78	0.70	0.78	0.84	0.68	0.73	0.78
				$\hat{\pi}$	0.99	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0.08	0.09	0.1	0.09	0.1	0.12	0.10	0.11	0.13
				$\hat{\pi}$	0.99	1	1	1	1	1	1	1	1
	25	(100,100)	LS	$\hat{\alpha}$	0.002	0.006	0.01	0.005	0.01	0.03	0.01	0.03	0.06
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			aLASSO	$\hat{\alpha}$	0.99	1	1	0.99	1	1	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0.001	0.002	0.009	0.003	0.007	0.026	0.008	0.01	0.03
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
$\mathscr{L}_2$	10	(25,25)	LS	$\hat{\alpha}$	0.05	0.07	0.12	0.09	0.12	0.18	0.14	0.18	0.23
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			aLASSO	$\hat{\alpha}$	0.60	0.69	0.79	0.71	0.78	0.82	0.69	0.74	0.78
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0.05	0.06	0.06	0.07	0.08	0.11	0.07	0.08	0.10
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
	25	(100,100)	LS	$\hat{\alpha}$	0	0.006	0.1	0.004	0.01	0.04	0.02	0.03	0.07
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			aLASSO	$\hat{\alpha}$	0.99	0.99	1	0.99	1	1	0.99	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0.001	0.006	0.003	0.01	0.02	0.01	0.02	0.04
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1

In Table 3.1.1 we give the empirical size of tests (under H_0), denoted $\hat{\alpha}$, and the empirical power (under H_1), denoted $\hat{\pi}$, for the three methods. Recall that for the linear model of Horváth, L. et al., (2004), when p was much smaller ($p = 1$) in respect to m ($m \geq 25$), and for nonlinear model of Ciuperca, G., (2013), always by the CUSUM method for least squares residuals, there was no difference in the empirical size and power results for different values of γ. In our case, for $p = 10$ and $m = 25$, the LS framework gives worse results for $\gamma = 0.49$: empirical type I error probability $\hat{\alpha}$ is larger than the theoretical value α. For LS and modified adaptive LASSO frameworks, the empirical size decreases when the number of observations m increases. Due to the fact that the nonzero estimators by adaptive LASSO may be biased, the empirical size of the associated CUSUM test statistic is largely greater than α. The law of the explanatory variables has no influence on the results.

In all cases, by the three methods, for different values of m or of k^0, the empirical power is

 ISSN: 2102-6238

either 1 or very close to 1.

Taking into account all these elements, for the following simulations we no consider the adaptive LASSO method.

3.1.2. Large number of variables

In this subsection, two values for number of variables are considered : $p = 400$ and $p = 1000$. The index set of the true nonzero values is $\mathscr{A} = \{3, 30, 90\}$, for the both values of p. The nonzero components of β^0 have the values $\beta^0_{,3} = 5$, $\beta^0_{,30} = 2$, $\beta^0_{,90} = -1$. Under hypothesis H_1 only components 90 and 91 change : $\beta^0_{2,90} = 0$, $\beta^0_{2,91} = -1$, after $k^0 = 25$.
Always two possible distributions are considered for X_j : $\mathscr{L}_1$ or $\mathscr{L}_2$. The distribution $\mathscr{L}_1$ is so : $X_j \sim \mathscr{N}(0,1)$ for $\forall j \in \{1, \cdots, p\} \setminus \{3, 40, 75\}$, $X_3 \sim \mathscr{N}(2,1)$, $X_{40} \sim \mathscr{N}(4,1)$, $X_{75} \sim \mathscr{N}(-1,1)$. The distribution $\mathscr{L}_2$ is $X_j^2 + j^2/m$, for $\forall j \in \{1, \cdots, p\} \setminus \{3, 40, 75\}$, with X_j of law $\mathscr{L}_1$ and $X_3 \sim \mathscr{N}(2,1)$, $X_{40} \sim \mathscr{N}(4,1)$, $X_{75} \sim \mathscr{N}(-1,1)$.

The empirical sizes and powers of the CUSUM test statistics corresponding to LS and modified adaptive LASSO residuals, are given in Table 2 when $X_j \sim \mathscr{L}_1$ and in Table 3 when $X_j \sim \mathscr{L}_2$.
In all cases (see Tables 2 and 3), the empirical test size corresponding to the modified adaptive LASSO method is smaller than the corresponding theoretical value α. On the other hand, the CUSUM test with LS residuals gives many false alarms ($\hat{\alpha}_{LS} \gg \alpha$). This shows that when the number of variables is very large, this test does not work well.

We analyse more closely the results of Table 2 when $X_j \sim \mathscr{L}_1$. For testing hypotheses by modified adaptive LASSO method, the presence of the change-point is not always detected in the case $\gamma = 0$ (the empirical power is smaller than 0.95), (see also Figure 7) : if k^0 draws away from m then $\hat{\pi}$ decreases. The same trend is observed for the two CUSUM methods in the cases $\gamma \in \{0.25, 0.49\}$: when k^0 draws away from m thus $\hat{\pi}$ decreases, but, with larger values for $\hat{\pi}$. In order to study if this decreasing is not due to the close values of p and m, we considered for $k^0 = 5$, $m = 1000$, $T = 80$ a model with $p = 1000$ regressors and another with $p = 400$ regressors. For CUSUM method with LS residuals, we obtain in the case $p = 400$ and $m = 1010$, that the empirical sizes are lower than the theoretical size (excepted for $\gamma = 0.49$). For CUSUM method with modified adaptive LASSO, the empirical powers are the same in the two cases : $p = 400$ or $p = 1000$. We obtain the same conclusions for $k^0 = 250$, $m = 1010$ and $p \in \{400, 1000\}$.
We analyse now the results of Table 3, when $X_j \sim \mathscr{L}_2$. We observe that the empirical powers are always equal to 1. For modified adaptive LASSO method, the empirical sizes do not exceed the theoretical sizes. On the other hand, as for $X_j \sim \mathscr{L}_1$, the LS method gives many false alarms.

The difference in results between the two laws $\mathscr{L}_1$ and $\mathscr{L}_2$ comes from the conditioning of the matrix $\mathbf{C}_m = m^{-1} \sum_{i=1}^m \mathbf{X}_i \mathbf{X}_i^t$ of assumptions (A2) or (A2bis). Effectively, in the case $m = 1010$ and $p = 1000$, the maximum to 500 Monte Carlo replications of the largest eigenvalue of $\mathbf{C}_m$ is 23.69 when $X_j \sim \mathscr{L}_1$, while the minimum of the largest eigenvalue of $\mathbf{C}_m$ is $19.72 \cdot 10^7$ when $X_j \sim \mathscr{L}_2$. The smallest eigenvalues of $\mathbf{C}_m$, for the two laws $\mathscr{L}_1$, $\mathscr{L}_2$, are of the order of 10^{-5}. In the case $m = 410$ and $p = 400$, the maximum to 500 Monte Carlo replications of the largest

 ISSN: 2102-6238

eigenvalue of $\mathbf{C}_m$ is 24.13 when $X_j \sim \mathscr{L}_1$, while the minimum of the largest eigenvalue of $\mathbf{C}_m$ is $12.36 \cdot 10^6$ when $X_j \sim \mathscr{L}_2$. The smallest eigenvalues of $\mathbf{C}_m$, for the two laws $\mathscr{L}_1$, $\mathscr{L}_2$, are of the order of 10^{-4}.

TABLE 2. *Empirical test size* $\hat{\alpha}$ *and power* $\hat{\pi}$*, for* $p = 400$ *or 1000,* $k^0 = 5, 25, 100, 250$*,* $\gamma = 0, 0.25, 0.49$*, three sizes* $\alpha = 0.025, 0.05, 0.10$*,* $X_j \sim \mathscr{L}_1$.

k^0	(m,T)	p	method	$\hat{\alpha},\hat{\pi}$	$\gamma=0$			$\gamma=0.25$			$\gamma=0.49$		
					0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
5	(410,80)	400	LS	$\hat{\alpha}$	0.62	0.69	0.80	0.94	0.96	0.97	0.99	0.99	1
				$\hat{\pi}$	0.83	0.89	0.92	0.98	0.99	0.99	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0.002	0.002	0.002	0.004	0.01	0.03
				$\hat{\pi}$	0.77	0.86	0.92	0.97	0.98	0.99	0.99	0.99	0.99
	(1010,80)	1000	LS	$\hat{\alpha}$	0.65	0.73	0.80	0.97	0.98	0.99	1	1	1
				$\hat{\pi}$	0.77	0.84	0.89	0.99	0.99	0.99	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0	0.01	0.02
				$\hat{\pi}$	0.31	0.46	0.67	0.93	0.95	0.97	1	1	1
		400	LS	$\hat{\alpha}$	0	0	0	0	0	0.002	0.08	0.12	0.19
				$\hat{\pi}$	0.33	0.48	0.66	0.92	0.93	0.95	0.99	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.01	0.01	0.05
				$\hat{\pi}$	0.32	0.46	0.66	0.94	0.96	0.97	0.99	0.99	0.99
25	(410,100)	400	LS	$\hat{\alpha}$	0.69	0.73	0.78	0.97	0.98	0.99	1	1	1
				$\hat{\pi}$	0.84	0.88	0.93	0.98	0.99	0.99	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.01	0.02	0.04
				$\hat{\pi}$	0.74	0.82	0.90	0.96	0.98	0.98	0.96	0.97	0.98
	(1010,100)	1000	LS	$\hat{\alpha}$	0.66	0.75	0.83	0.99	1	1	1	1	1
				$\hat{\pi}$	0.81	0.88	0.92	0.99	0.99	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0	0.01	0.07
				$\hat{\pi}$	0.29	0.47	0.66	0.91	0.93	0.96	1	1	1
100	(410,175)	400	LS	$\hat{\alpha}$	0.82	0.89	0.94	0.97	0.98	0.99	0.94	0.95	0.96
				$\hat{\pi}$	0.91	0.94	0.96	0.98	0.99	0.99	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0.002	0.002	0.002	0.002	0.01	0.02	0.03	0.06
				$\hat{\pi}$	0.59	0.70	0.81	0.81	0.86	0.91	0.85	0.88	0.92
	(1010,175)	1000	LS	$\hat{\alpha}$	0.87	0.91	0.95	0.99	0.99	1	1	1	1
				$\hat{\pi}$	0.92	0.95	0.98	0.99	0.99	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0.001	0.01	0.03	0.05
				$\hat{\pi}$	0.23	0.39	0.56	0.75	0.83	0.89	0.90	0.93	0.95
250	(410,325)	400	LS	$\hat{\alpha}$	0.92	0.93	0.94	0.94	0.95	0.96	0.94	0.95	0.95
				$\hat{\pi}$	0.97	0.97	0.98	0.99	0.99	0.99	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0.004	0.004	0.004	0.006	0.02	0.02	0.04	0.05
				$\hat{\pi}$	0.33	0.45	0.56	0.48	0.58	0.68	0.48	0.54	0.63
	(1010,325)	1000	LS	$\hat{\alpha}$	0.94	0.96	0.97	0.97	0.98	0.99	0.98	0.99	0.99
				$\hat{\pi}$	0.97	0.98	0.99	0.99	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.006	0.04	0.06
				$\hat{\pi}$	0.15	0.26	0.42	0.48	0.61	0.72	0.63	0.71	0.77
		400	LS	$\hat{\alpha}$	0	0	0.005	0.01	0.02	0.07	0.15	0.18	0.26
				$\hat{\pi}$	0.26	0.37	0.50	0.54	0.61	0.68	0.65	0.70	0.81
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.01	0.03	0.05
				$\hat{\pi}$	0.19	0.28	0.48	0.54	0.60	0.73	0.64	0.71	0.77

 ISSN: 2102-6238

TABLE 3. *Empirical test size* $\hat{\alpha}$ *and power* $\hat{\pi}$*, for* $p = 400$ *or 1000,* $k^0 = 5, 25, 100, 250$*,* $\gamma = 0, 0.25, 0.49$*, three sizes* $\alpha = 0.025, 0.05, 0.10$*,* $X_j \sim \mathscr{L}_2$.

k^0	(m,T)	p	method	$\hat{\alpha},\hat{\pi}$	$\gamma=0$			$\gamma=0.25$			$\gamma=0.49$		
					0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
5	(410,80)	400	LS	$\hat{\alpha}$	0.64	0.72	0.82	0.91	0.94	0.98	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.004	0.01	0.05
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
	(1010,80)	1000	LS	$\hat{\alpha}$	0.64	0.74	0.82	0.97	0.98	0.99	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.01	0.02	0.04
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
25	(410,100)	400	LS	$\hat{\alpha}$	0.67	0.75	0.82	0.97	0.99	1	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.02	0.05	0.09
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
	(1010,100)	1000	LS	$\hat{\alpha}$	0.6	0.7	0.7	0.99	0.99	0.99	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.01	0.02	0.02
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
100	(410,175)	400	LS	$\hat{\alpha}$	0.80	0.86	0.91	0.97	0.98	0.99	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0.002	0.002	0.01	0.01	0.03	0.03
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
	(1010,175)	1000	LS	$\hat{\alpha}$	0.86	0.90	0.94	0.99	0.99	0.99	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0	0.01	0.02	0.02
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
250	(410,325)	400	LS	$\hat{\alpha}$	0.9	0.93	0.96	0.98	0.99	0.99	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0.002	0.002	0.006	0.002	0.002	0.02	0.01	0.03	0.07
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
	(1010,325)	1000	LS	$\hat{\alpha}$	0.92	0.95	0.98	0.99	0.99	0.99	1	1	1
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1
			modif aLASSO	$\hat{\alpha}$	0	0	0	0	0	0.002	0.006	0.02	0.05
				$\hat{\pi}$	1	1	1	1	1	1	1	1	1

 ISSN: 2102-6238

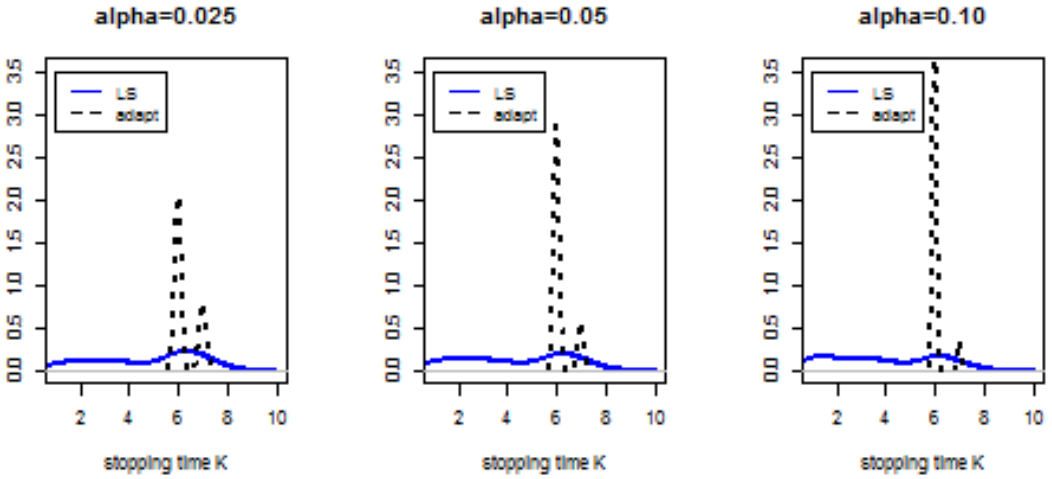

FIGURE 1. *Estimated density of the stopping time, corresponding to LS and modified adaptive LASSO residuals, for p=400,* $k^0 = 5$, $\gamma = 0.25$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.

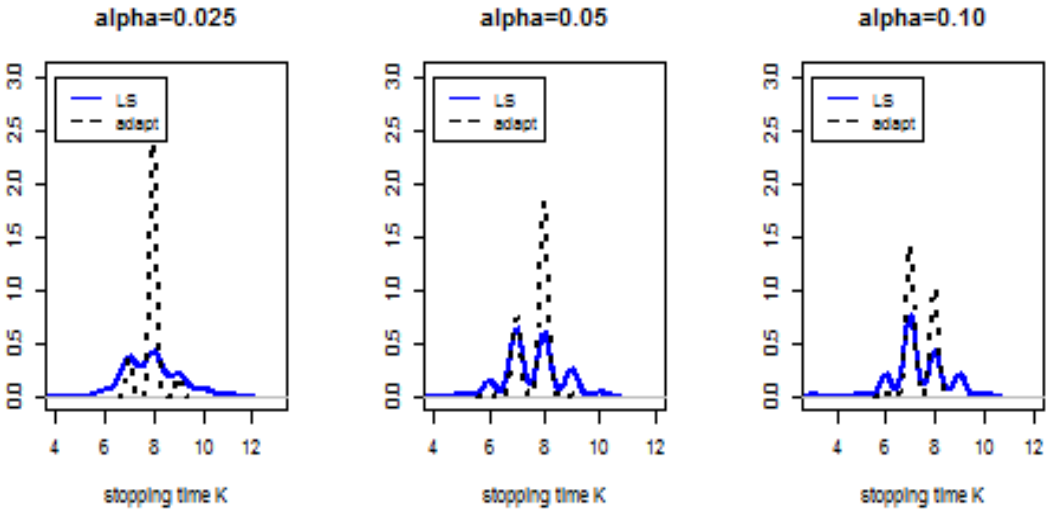

FIGURE 2. *Estimated density of the stopping time, corresponding to LS and modified adaptive LASSO residuals, for p=400,* $k^0 = 5$, $\gamma = 0$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.

3.2. Stopping time estimations

In this subsection, we estimate the stopping times $\tilde{k}_m^*$ and $\hat{k}_m$ using (20) for the modified adaptive LASSO residuals and (22) for the LS residuals (the stopping time proposed by Horváth, L. et al., , 2004). Two positions for k^0, under hypothesis H_1, are studied : $k^0 = 5$ (just after m) and $k^0 = 25$.

 ISSN: 2102-6238

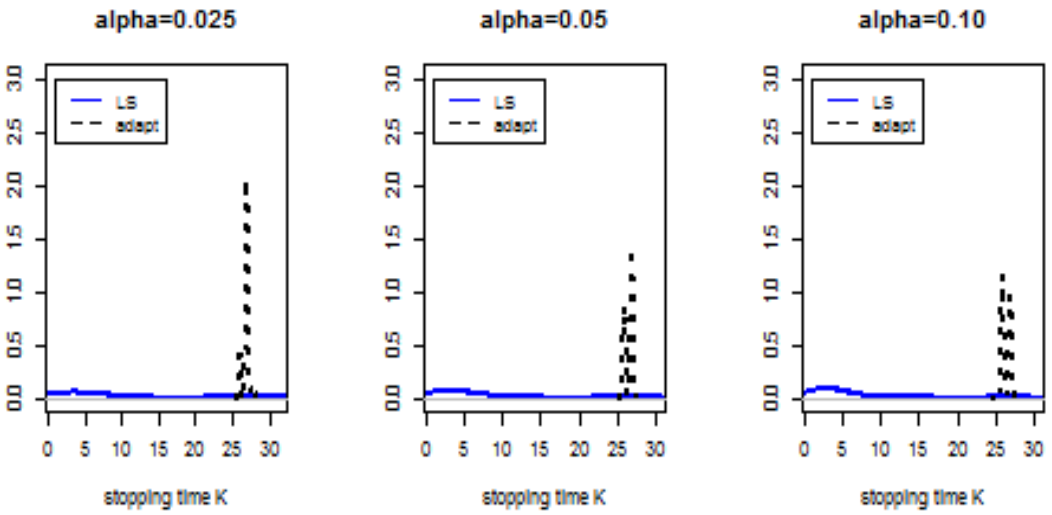

FIGURE 3. *Estimated density of the stopping time, corresponding to LS and modified adaptive LASSO residuals, for p=400,* $k^0 = 25$, $\gamma = 0.25$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.

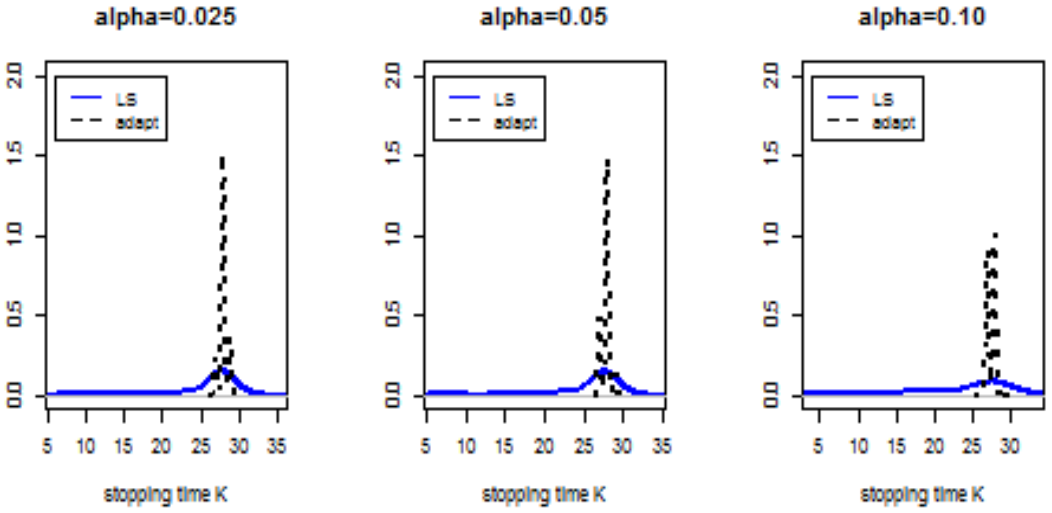

FIGURE 4. *Estimated density of the stopping time, corresponding to LS and modified adaptive LASSO residuals, for p=400,* $k^0 = 25$, $\gamma = 0$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.

In Figures 1-7 we represent the empirical density of the stopping times corresponding to the two estimation methods : solid line for the density corresponding to LS framework and dotted line for the density corresponding to modified adaptive LASSO framework.

Outside the seven empirical densities for the stopping times, represented in Figures 1-7, we

 ISSN: 2102-6238

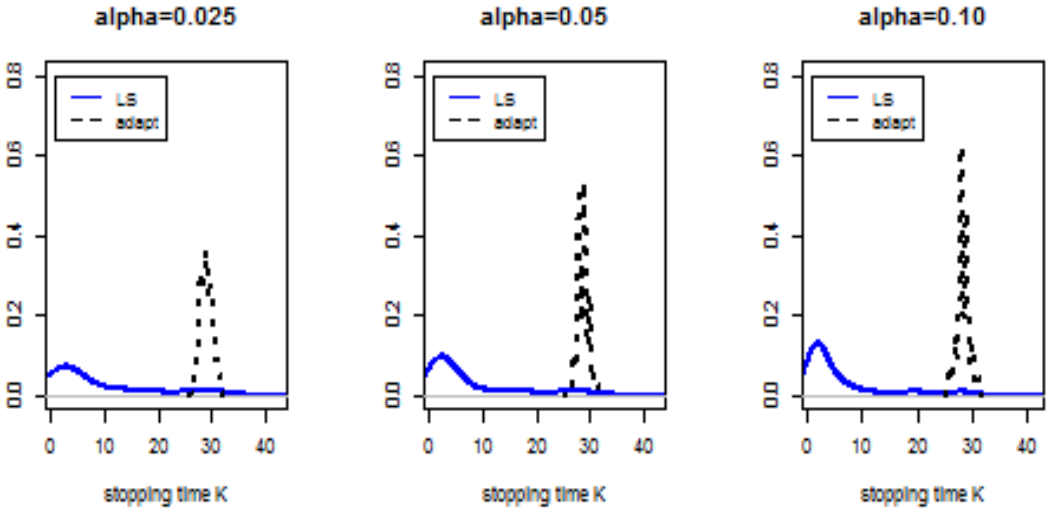

FIGURE 5. *Estimated density of the stopping time, corresponding to LS and modified adaptive LASSO residuals, for p=1000,* $k^0 = 25$, $\gamma = 0.25$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.

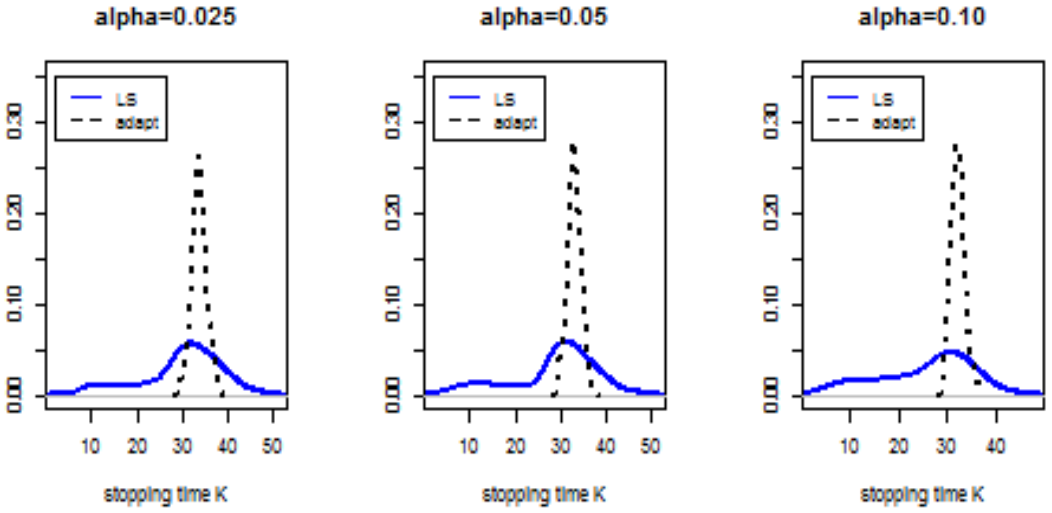

FIGURE 6. *Estimated density of the stopping time, corresponding to LS ans adaptive LASSO modified residuals, for p=1000,* $k^0 = 25$, $\gamma = 0$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.

give some elements of descriptive statistics for the stopping times estimations, for $\gamma = 0.25$ in Table 4 and for $\gamma = 0$ in Table 5. In Tables 4 and 5 we consider $p \in \{100, 400, 1000\}$, $\alpha \in \{0.025, 0.05, 0.10\}$ and $X_j \sim \mathscr{L}_2$.

The estimation of k^0 by modified adaptive LASSO is very accurate and unbiased, for $\gamma \in$

 ISSN: 2102-6238

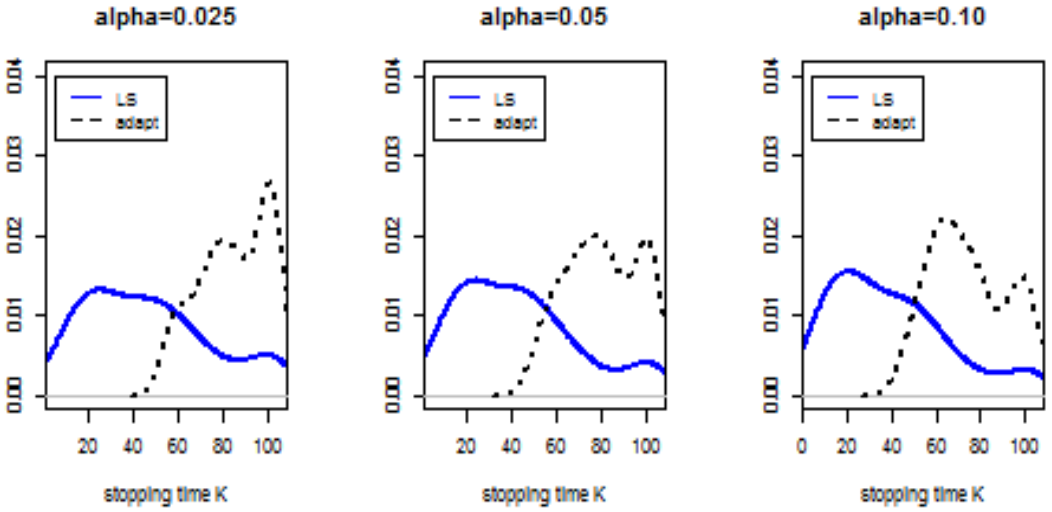

FIGURE 7. *Estimated density of the stopping time, corresponding to LS and modified adaptive LASSO residuals, for p=400,* $k^0 = 25$, $\gamma = 0$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_1$.

$\{0, 0.25\}$, if $X_j \sim \mathscr{L}_2$, whether for $k^0 = 5$ or 25 and for the number of regressors equal to 400 or 1000, . Seen the empirical densities and Table 4,we deduce that the estimation of k^0 by stopping time corresponding to the LS residuals is biased and above all detects the change before it occurs (false change). These results are consistent with those previously found in Table 3 for the empirical sizes. The empirical density shape of the stopping times corresponding to the LS residuals also indicates that the variability $\hat{k}_m$ is very high.

TABLE 4. *Some summary statistics for stopping times :* $\hat{k}_m$ *by LS and* $\tilde{k}^*_m$ *adaptive LASSO modified methods, for* $T = 100$, $k^0 = 25$, $\gamma = 0.25$, *three sizes* $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.

p	m	method	$\alpha = 0.025$		$\alpha = 0.05$		$\alpha = 0.10$	
			median($\hat{k}$)	mean($\hat{k}$)	median($\hat{k}$)	mean($\hat{k}$)	median($\hat{k}$)	mean($\hat{k}$)
1000	1010	LS	4	6	3	6	3	4
		modif aLASSO	29	29	29	28	28	28
400	410	LS	7	11	6	9	5	8
		modif aLASSO	27	26	27	26	26	26
100	110	LS	13	14	10	13	7	11
		modif aLASSO	26	26	26	26	26	26

 ISSN: 2102-6238

TABLE 5. *Some summary statistics for stopping times : $\hat{k}_m$ by LS and $\tilde{k}^*_m$ adaptive LASSO modified methods, for $T = 100$, $k^0 = 25$, $\gamma = 0$, three sizes $\alpha = 0.025, 0.05, 0.10$, $X_j \sim \mathscr{L}_2$.*

p	m	method	$\alpha = 0.025$		$\alpha = 0.05$		$\alpha = 0.10$	
			median($\hat{k}$)	mean($\hat{k}$)	median($\hat{k}$)	mean($\hat{k}$)	median($\hat{k}$)	mean($\hat{k}$)
1000	1010	LS	32	28	29	26	27	23
		modif aLASSO	33	33	33	32	32	31
400	410	LS	27	25	27	24	26	21
		modif aLASSO	28	28	28	27	27	27
100	110	LS	26	22	26	22	26	20
		modif aLASSO	26	26	26	26	26	26

TABLE 6. *Computation times of the empirical sizes $\hat{\alpha}$, for $k^0 \in \{5, 10, 250\}$, $m \in \{25, 400, 1010\}$, $T = k^0 + 75$, $p \in \{10, 400, 1000\}$ for 500 Monte Carlo replications.*

k^0	m	p	Computation time
10	25	10	4 sec
5	410	400	3 min
250	1010	1000	25 min

3.3. *Conclusions of simulations*

We presented the numerical results for three CUSUM test statistics corresponding to the model residuals, obtained, respectively, by three parametric estimation methods : LS, adaptive LASSO and modified adaptive LASSO.

Regardless of the number of variables in the model, the CUSUM statistic with adaptive LASSO residuals gives false alarms when H_0 is true and then the empirical size is greater than the theoretical size.

If the number of explanatory variables is large, by the CUSUM method with LS residuals, we get many false alarms ($\geq 60\%$) under hypothesis H_0, regardless of the distribution of the regressors, of the design matrix or of the parameter values γ. Under hypothesis H_1, for $\gamma = 0$, the stopping time given by (22) is not accurate.

For the CUSUM test with modified adaptive LASSO residuals, if matrix $\mathbf{C}_m = m^{-1} \sum_{i=1}^{m} \mathbf{X}_i \mathbf{X}_i^t$ of assumptions (A2) or (A2bis) is well-conditioned matrix, empirical type I error probability (size) is smaller than the theoretical (except for $p = 10$). The empirical test power is 1, for any required value of γ, α m, p and of k^0. The stopping time gives very similar estimates to the true value of change-point k^0.

Concerning the computation time, we give in Table 6 some examples of computation times of the empirical sizes, each model being generated for 500 Monte Carlo replications and $T = k^0 + 75$. Obviously, the computation time depends on the observation number m and on the regressor number p. These simulations were performed on a cluster computer with CPU : 800 MHz, cache size : 512 KB and RAM : 32 GB.

 ISSN: 2102-6238

Acknowledgments The author sincerely thanks the Editor, Associate Editor and two anonymous referees whose valuable comments led to an improvement of the article.

Références

Breiman, L., (1996). Heuristics of instability and stabilization in model selection, *Annals of Statistics*, **24**(6), 2350-2383.

Ciuperca G., (2013), Quantile regression in high-dimension with breaking, *Journal of Statistical Theory and Applications*, **12**(3), 288-305.

Ciuperca, G., (2013). Two tests for sequential detection of a change-point in a nonlinear model, *Journal of Statistical Planning and Inference*, **143**(10), 1621-1834.

Ciuperca G., (2014), Model selection by LASSO methods in a change-point model, *Statistical Papers*, **55**(2), 349-374.

Harchaoui, Z. and Lévy-Leduc, C.,(2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, **105**(492), 1480-1493.

Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J.,(2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, **126**, 225-251.

Lee S., Seo M.H., Shin Y., (2012), The LASSO for high-dimensional regression with a possible change-point. *arXiv :1209 :4875v2*.

Lung-Yut-Fong A., Lévy-Leduc C., Cappé O., (2012), Distributed detection/localization of change-points in high-dimensional network traffic data, *Statistics and Computing*, **22**(2), 485-496.

Siris V.A., Papagalou F., (2006), Application of anomaly detection algorithms for detecting SYN flooding attacks *Computer Communications*, **29**(9), 1433-1442. iCON 2004–12th IEEE International Conference on Network 2004.

Tibshirani, R.,(1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society*, Ser. B, **58**, 267-288.

Wang H., Zhang D., Shin K.G., (2002), Detecting SYN Flooding Attacks. *Proceedings of IEEE INFOCOM*, **3**, 1530-1539.

Zou, H.,(2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, **101**(476), 1418-1428.

 ISSN: 2102-6238