Quantile regression in high-dimension with breaking

Gabriela Ciuperca

CNRS, UMR 5208, Institut Camille Jordan, Université Lyon 1,
Bat. Braconnier, 43, blvd du 11 novembre 1918,
F - 69622 Villeurbanne Cedex, France,

E-mail: Gabriela.Ciuperca@univ-lyon1.fr

Abstract

The paper considers a linear regression model in high-dimension for which the predictive variables can change the influence on the response variable at unknown times (called change-points). Moreover, the particular case of the heavy-tailed errors is considered. In this case, least square method with LASSO or adaptive LASSO penalty can not be used since the theoretical assumptions do not occur or the estimators are not robust. Then, the quantile model with SCAD penalty or median regression with LASSO-type penalty allows, in the same time, to estimate the parameters on every segment and eliminate the irrelevant variables. We show that, for the two penalized estimation methods, the oracle properties is not affected by the change-point estimation. Convergence rates of the estimators for the change-points and for the regression parameters, by the two methods are found. Monte-Carlo simulations illustrate the performance of the methods.

Keywords: change-points; high-dimension; oracle properties; SCAD; LASSO-type estimators.

1. Introduction

A model which changes at some observations is called a change-point model. The location of these changes (called also change-points, breaks, changes) may be known or unknown. In this paper, we consider a model with multiple change-points at unknown locations. Moreover, as very often in practice, for example in genetics, the response variable is studied function of a very large number of regressors. However, only a small number of regressors is going to influence the response variable. In recent years, change-point models and high-dimension regression have received much attention in the literature, most often in the case of a model with zero mean errors and bounded variance. A L_1 or adaptive L_1 penalty in the context of least squares model can be considered. We obtain then the popular method introduced by Tibshirani and called LASSO (Least Absolute Shrinkage and Selection Operator) method[1]. On the other hand, it is well known that, the presence of outliers in model may cause a large error in a least squares estimator. This can happen especially when the error distribution is not Gaussian and distribution tail is large enough. The outliers can also create problems in the detection of the jumps. An alternative method is then the quantile estimation.

To be more precise, if the errors $(\varepsilon_i)_{1 \leqslant i \leqslant n}$ of the regression model are such that $I\!P[\varepsilon_i < 0] = \tau$, then the τth quantile regression is considered, i.e. the regression parameters are found by minimizing the function $\rho_\tau(\varepsilon) = \sum_{i=1}^n \varepsilon_i[\tau 1\!\!1_{\varepsilon_i > 0} - (1 - \tau)1\!\!1_{\varepsilon_i \leqslant 0}]$. The choice of $\tau = 1/2$ yields the median regression and the L_1-estimator, also

known as least absolute deviation (LAD) estimator.

Moreover, when the model has a very large regressor variable number, a penalty is necessary to estimate simultaneously the parameters on every segment and to eliminate the irrelevant regressors without crossing every time by a hypothesis test. The SCAD (Smoothly Clipped Absolute Deviation) and LASSO penalties have the advantage of selection and parameter estimation. It was established that these two methods have the oracle properties in a model without change-points: the zero components of the true parameters are estimated (shrunk) as 0 with probability tending to 1 (also called sparsity property) and the nonzero components have an optimal estimation rate (furthermore they are asymptotically normal). See (Ref. 2) for the SCAD method in a τth quantile regression and (Ref. 3) for the LASSO-type method in a median regression, both models without change-points. Recall also for a median regression in high dimension the paper of Wang,[4] where a L_1 penalized least absolute deviation method is considered, when the overall variable number is larger than the observation number.

In a multiple change-point model, the break estimation could affects the estimator properties. This is the main interest of this paper. The difficulty to study a change-point model results first from the dependence of the model of two parameter types: the regression parameters and the change-points.
A change-point linear model in high-dimension was also considered by Ciuperca but under stronger assumptions that the errors have mean zero and bounded variance[5]. An adaptive LASSO estimator was studied. It was proved that it has the oracle properties on each estimated segment. However, when the model contains outliers, the adaptive LASSO estimator may not be robust and moreover, the observation number should be greater than the parameter number to be estimated.

In the present work we restrict our attention to the quantile regression in high-dimension with multiple change-points when the classical conditions on the errors do not occur. The change-points and the regression parameters on each segment are first estimated by the SCAD method. After, for a median regression ($\tau = 1/2$), these parameters are estimated by the LASSO-type method. The asymptotic and oracle properties of these estimators are studied. We also carry out simulations to investigate the properties of the two proposed estimators.

The paper is organized as follows. The model and assumptions are introduced in Section 2. In Section 3, the SCAD estimator in a change-point model is proposed and its asymptotic behavior is studied. Next, LASSO-type estimator is given in Section 4. For both methods, the oracle properties and convergence rate of the estimators are obtained. Section 5 reports some simulations results which illustrate the methods interest. In Section 6 we give the proofs of Theorems. Finally, Section 7 contains some lemmas which are useful to prove the main results.

2. Model and general notations

In this section we introduce the models without and with change-points, general assumptions, notations. Some general results used required for the two estimation methods are given.
We consider the linear model without change-points

$$Y_i = \mathbf{X}_i^t \boldsymbol{\phi} + \varepsilon_i, \qquad i = 1, \cdots, n \tag{1}$$

where the response variable Y_i is an univariate random variable, $\mathbf{X}_i \in \mathbb{R}^p$ is a p-vector of regressors (covariates) and the ε_i is the error. The errors $(\varepsilon_i)_{1 \leqslant i \leqslant n}$ are independent identically distributed (i.i.d) random variables. The regression parameters are $\phi \in \Gamma \subset \mathbb{R}^p$, with Γ a compact set and $\boldsymbol{\phi}^0$ true value (unknown) of the parameter $\boldsymbol{\phi}$. Contrary to the classic suppositions for a regression model, we do not impose the condition that the mean of errors ε_i is zero or that their variance is bounded.
All throughout the paper, vector and matrices are written in bold face.

With regard to the errors ε_i and the design $\mathbf{X}_i$, we make the following assumptions:

(A1) Let f be the density of ε_i and F its distribution function. We suppose that $f(0) > 0$, $F(0) = \tau$, $|f(y) - f(0)| \leqslant c|y|^{1/2}$, for all y in a neighborhood of 0. The quantile τ is a real number in the interval $(0,1)$.

(A2) $(\mathbf{X}_i)_{1 \leqslant i \leqslant n}$ is a deterministic sequence, such that $n^{-1} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^t$ converges, as $n \to \infty$, to a non negative definite matrix;

(A3) $(\mathbf{X}_i)_{1 \leqslant i \leqslant n}$ is uniformly bounded.

These conditions are typical for a quantile regression (see e.g. (Ref. 6)). These assumptions are also classic conditions for a model estimated by LAD method: the first condition is found in (Ref. 7) and the last two in (Ref. 8).

It is of interest to note that by assumption (A1) we have $I\!P[\varepsilon_i < 0] = \tau$, but the expectation $I\!E[\varepsilon_i]$ cannot exist. A regression model (1) with the errors (ε_i) satisfying the condition $I\!P[\varepsilon_i < 0] = \tau$ is called quantile regression. In order to estimate the unknown regression parameter $\boldsymbol{\phi}$, we consider the function

$$\rho_\tau(r) = r[\tau \, 1\!\!1_{r>0} - (1-\tau) \, 1\!\!1_{r \leqslant 0}] \tag{2}$$

and the corresponding estimator

$$\widehat{\boldsymbol{\phi}}_n^{(\tau)} = \arg\min_{\boldsymbol{\phi} \in \Gamma} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_i^t \boldsymbol{\phi}). \tag{3}$$

In order to study the quantile regression and the estimator (3), let be the random processes

$$
\begin{aligned}
&G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) = \rho_\tau(\varepsilon_i - \mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi}^0)) - \rho_\tau(\varepsilon_i), \qquad \mathscr{G}_n^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) = \sum_{i=1}^{n} G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0), \\
&D_i = (1-\tau) \, 1\!\!1_{\varepsilon_i \leqslant 0} - \tau \, 1\!\!1_{\varepsilon_i > 0}, \qquad W_n = \sum_{i=1}^{n} D_i \mathbf{X}_i^t, \\
&R_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) = G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - D_i \mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi}^0).
\end{aligned}
\tag{4}
$$

Obviously $I\!E[D_i] = 0$. The relation between $\mathscr{G}_n^{(\tau)}$ and $R_i^{(\tau)}$ is

$$\mathscr{G}_n^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - I\!E[\mathscr{G}_n^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] = \sum_{i=1}^{n} [R_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - I\!E[R_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)]] + W_n(\boldsymbol{\phi} - \boldsymbol{\phi}^0). \tag{5}$$

For the parameter regression vector $\boldsymbol{\phi}$, we shall use the notation $\boldsymbol{\phi} = (\phi_{,1}, \cdots, \phi_{,p})$.

Throughout the paper, C denotes a positives generic constant not dependent on n which may take different values in different formula or even in different parts of the same formula. For a vector $\mathbf{v} = (v_1, \cdots, v_p)$ let us denote $|\mathbf{v}| = (|v_1|, \cdots, |v_p|)$ and $\frac{1}{\mathbf{v}} = (\frac{1}{v_1}, \cdots, \frac{1}{v_p})$. On the other hand, $\|\mathbf{v}\|_2$ is the Euclidean norm and $\|\mathbf{v}\|_1 = \sum_{i=1}^{p} |v_i|$ is the L_1 norm. All vectors are column and $\mathbf{v}^t$ denotes the transposed of $\mathbf{v}$.

For coherence, we try to use the same notations as in the paper of Wu and Liu[2]. By elementary calculations, we obtain, with the probability 1, $|R_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)| \leqslant |\mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi}^0)| 1\!\!1_{|\varepsilon_i| \leqslant |\mathbf{X}_i^t(\boldsymbol{\phi}-\boldsymbol{\phi}^0)|}$. This inequality, the definition of $G_i^{(\tau)}$ and the assumption (A3) allow to obtain $\mathscr{G}_n^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) \leqslant Cn\|\boldsymbol{\phi} - \boldsymbol{\phi}^0\|_2$. Following result proves that for every parameter $\boldsymbol{\phi}$ and for every quantile order τ, the process $G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)$ has positive expectation, indifferently of the design $\mathbf{X}_i$.

Proposition 1. *Under assumption (A1), we have, for all $\boldsymbol{\phi} \in \Gamma$, $I\!E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] \geqslant 0$.*

Remark 1. In (Ref. 8), the behavior in a neighborhood of $\boldsymbol{\phi}^0$ of the process $\mathscr{G}_n^{(\tau)} - I\!E[\mathscr{G}_n^{(\tau)}]$ is obtained in the particular case $\tau = 1/2$. By a similar demonstration, we can prove that the result holds in general, for any $\tau \in (0,1)$: let be a positive sequence (c_n) such that $c_n \to 0$ and $nc_n^2/\log n \to \infty$. Under the assumptions (A1)-(A3), there exists a constant $C > 0$ such that $\forall \varepsilon > 0$, $I\!P[\sup_{\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|_2 \leqslant c_n} (nc_n^2)^{-1} |\mathscr{G}_n^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - I\!E[\mathscr{G}_n^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)]| \geqslant$

$\varepsilon] \leqslant \exp(-\varepsilon^2 n c_n^2 C).$

The proof sketch of this remark is given at the end of Section 6.
As a consequence of this Remark, by the Borel-Cantelli lemma, we have for any $\varepsilon > 0$,

$$\limsup_{n \to \infty} (\sup_{\|\boldsymbol{\phi} - \boldsymbol{\phi}^0\|_2 \leqslant c_n} \frac{1}{nc_n^2} |\mathscr{G}_n^{(\tau)}(\boldsymbol{\phi}; \boldsymbol{\phi}^0) - I\!\!E[\mathscr{G}_n^{(\tau)}(\boldsymbol{\phi}; \boldsymbol{\phi}^0)]|) \leqslant \varepsilon, \qquad a.s. \tag{6}$$

It is well known that the estimator (3) has all nonzero components. For estimations and choosing the regressors simultaneously, penalized methods can be used: SCAD or LASSO-type. These estimation methods all become more interesting for a model with K change-points

$$Y_i = \mathbf{X}_i^t \boldsymbol{\phi}_1 1\!\!1_{1 \leqslant i < l_1} + \mathbf{X}_i^t \boldsymbol{\phi}_2 1\!\!1_{l_1 \leqslant i < l_2} + \cdots + \mathbf{X}_i^t \boldsymbol{\phi}_{K+1} 1\!\!1_{l_K \leqslant i \leqslant n} + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{7}$$

where $1\!\!1_{(.)}$ denotes the indicator function.
The model parameters are the regression parameters $(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{K+1})$ and the change-points $(l_1, \cdots, l_K)$. The true values (unknown) are $(\boldsymbol{\phi}_1^0, \cdots, \boldsymbol{\phi}_{K+1}^0)$, $(l_1^0, \cdots, l_K^0)$, respectively. The observations $l_{r-1} + 1, \cdots, l_r$ between two consecutive change-points will be called the rth segment (interval, phase).
Concern the distance between two consecutive change-points, we impose the assumption
(A4) $l_{r+1} - l_r \geqslant n^{3/4}$, for all $r = 0, 1, \cdots, K$.

In order to study the properties of the penalized estimators in a model with breaking, we need corresponding results obtained without change-points when $\tau = 1/2$: by (Ref. 9), (Ref. 3) and for a some $\tau \in (0, 1)$ by (Ref. 2).
In the next section we investigate theoretical properties of the smoothly clipped absolute deviation (SCAD) method in a change-point model.

3. SCAD estimator

We begin this section by recalling the SCAD estimator for the quantile regression model (1) without change-points, introduced by (Ref. 10) and developed later by (Ref. 2)

$$\widehat{\boldsymbol{\phi}}_n^{(\tau, \lambda)} \equiv \arg\min_{\boldsymbol{\phi}} (\sum_{i=1}^n [\rho_\tau(Y_i - \mathbf{X}_i^t \boldsymbol{\phi}) + \sum_{j=1}^p p_\lambda(|\phi_{,j}|)]). \tag{8}$$

The penalty $p_\lambda(\phi_{,j})$ is defined by its first derivative

$$p_\lambda'(|\phi_{,j}|) \equiv \lambda \{ 1\!\!1_{|\phi_{,j}| \leqslant \lambda} + \frac{(a\lambda - |\phi_{,j}|)_+}{(a-1)\lambda} 1\!\!1_{|\phi_{,j}| > \lambda} \}, \tag{9}$$

for all $j = 1, \cdots, p$, with $\lambda > 0$, $a > 2$ deterministic tuning parameters. For real x we use the notation $sgn(x)$ for the sign function $sgn(x) = \frac{x}{|x|}$ when $x \neq 0$ and $sgn(0) = 0$. We also denote $x_+ = \max\{0, x\}$.
In order to study the estimator $\widehat{\boldsymbol{\phi}}_n^{(\tau, \lambda)}$, introduce the function, for $i = 1, \cdots, n$,

$$G_i^{(\tau, \lambda)}(\boldsymbol{\phi}; \boldsymbol{\phi}^0) \equiv G_i^{(\tau)}(\boldsymbol{\phi}; \boldsymbol{\phi}^0) + [\mathbf{p}_\lambda(|\boldsymbol{\phi}|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}^0|)]^t \mathbf{1}_p,$$

with $\mathbf{1}_p \equiv (1, \cdots, 1)$ a $p \times 1$ vector and $\mathbf{p}_\lambda(\boldsymbol{\phi}) \equiv (p_\lambda(\phi_{,1}), \cdots p_\lambda(\phi_{,p}))$ also $p \times 1$ vector, with $\boldsymbol{\phi} = (\phi_{,1}, \cdots, \phi_{,p})$.

For this purpose, we first give the Karush-Kuhn-Tucker (KKT) conditions for the quantile model (1) without change-points.

For the estimator $\widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}$ given by the relation (8), let us consider the index set of the variables selected by the SCAD method

$$\mathscr{A}_n \equiv \left\{ j; \widehat{\phi}_{n,j}^{(\tau,\lambda)} \neq 0 \right\},$$

with $\widehat{\phi}_{n,j}^{(\tau,\lambda)}$ the jth component of $\widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}$.

Proposition 2. *For the estimator (8), the KKT conditions are*

• *for $j \in \mathscr{A}_n$:*

$$\tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} \mathbb{1}_{Y_i < \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}} = n\lambda \, sgn(\widehat{\phi}_{n,j}^{(\tau,\lambda)}) \left\{ \mathbb{1}_{|\widehat{\phi}_{n,j}^{(\tau,\lambda)}| \leqslant \lambda} + \frac{(a\lambda - |\widehat{\phi}_{n,j}^{(\tau,\lambda)}|)_+}{(a-1)\lambda} \mathbb{1}_{|\widehat{\phi}_{n,j}^{(\tau,\lambda)}| > \lambda} \right\}.$$

• *for $j \notin \mathscr{A}_n$:*

$$\left| \tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} \mathbb{1}_{Y_i < \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}} \right| \leqslant n\lambda.$$

For the model (7), in order to study the SCAD estimators of the regression parameters $(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{K+1})$, and of the change-points $(l_1, \cdots, l_K)$, let us consider the function

$$S(l_1, \cdots, l_K) \equiv \sum_{r=1}^{K+1} \inf_{(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{K+1}) \in \Gamma^{K+1}} \sum_{i=l_{r-1}+1}^{l_r} \left[\rho_\tau(Y_i - \mathbf{X}_i^t \boldsymbol{\phi}_r) + \mathbf{p}_{\lambda;(l_{r-1};l_r)}(|\boldsymbol{\phi}_r|) \mathbf{1}_p \right]. \tag{10}$$

In each interval (l_{r-1}, l_r) another penalty $\mathbf{p}_{\lambda;(l_{r-1};l_r)}$ can be considered, with $l_0 = 1$ and $l_{K+1} = n$. For simplicity of notation, we denote the penalty of (10) by $\mathbf{p}_{\lambda;(l_{r-1};l_r)}$ for some (l_{r-1}, l_r) and by $\mathbf{p}_{\lambda;(l_{r-1}^0;l_r^0)}$ for the true change-points, but it is understood that the series λ are in fact $\lambda_{(l_{r-1};l_r)}$, $\lambda_{(l_{r-1}^0;l_r^0)}$, respectively. For the interval $(1, \cdots, n)$, the tuning parameter $\lambda_{(0,n)}$ is λ_n.

We define the SCAD change-point estimator by

$$(\widehat{l}_1^{(\tau,\lambda)}, \cdots, \widehat{l}_K^{(\tau,\lambda)}) \equiv \underset{(l_1, \cdots, l_K) \in \mathbb{R}^K}{\arg\min} \, S(l_1, \cdots, l_K), \tag{11}$$

with the function S defined by (10). Between two consecutive change-points l_{r-1} and l_r, the SCAD estimator of the corresponding regression parameter $\boldsymbol{\phi}_r$ is

$$\widehat{\boldsymbol{\phi}}_{(l_{r-1};l_r)}^{(\tau,\lambda)} \equiv \underset{\boldsymbol{\phi}_r}{\arg\min} \sum_{i=l_{r-1}+1}^{l_r} \left[\rho_\tau(Y_i - \mathbf{X}_i^t \boldsymbol{\phi}_r) + \mathbf{p}_{\lambda;(l_{r-1};l_r)}(|\boldsymbol{\phi}_r|) \mathbf{1}_p \right] = \underset{\boldsymbol{\phi}_r}{\arg\min} \sum_{i=l_{r-1}+1}^{l_r} G_i^{(\tau,\lambda)}(\boldsymbol{\phi}_r; \boldsymbol{\phi}_r^0).$$

Then, the SCAD regression parameter estimator for the rth segment is obtained by considering for the change-points their corresponding estimators: $\widehat{\boldsymbol{\phi}}_{(\widehat{l}_{r-1}^{(\tau,\lambda)};\widehat{l}_r^{(\tau,\lambda)})}^{(\tau,\lambda)}$. The following two theorems state the asymptotic behaviors of the estimators (11) and of $\widehat{\boldsymbol{\phi}}_{(\widehat{l}_{r-1}^{(\tau,\lambda)};\widehat{l}_r^{(\tau,\lambda)})}^{(\tau,\lambda)}$. The first result gives the convergence rate of the change-point estimator.

Theorem 3. *Under the assumptions (A1)-(A4), with the tuning parameter $(\lambda_{(l_{r-1},l_r)})_{1 \leqslant r \leqslant K+1}$ a sequence, depending on n, converging to zero, $(l_r - l_{r-1})^{1/2}\lambda_{(l_{r-1},l_r)} \to \infty$ and for a deterministic sequence (c_n), such that*

$c_n \to 0$, $nc_n^2/\log n \to \infty$ and $\lambda_n c_n^{-2} \to 0$, as $n \to \infty$, then we have $\widehat{l}_r^{(\tau,\lambda)} - l_r^0 = O_{\mathbb{P}}(1)$, for every $r = 1, \cdots, K$.

Remark 2. We have following relations between the sequences (λ_n) and (c_n), $\lambda_n \ll c_n^2 \ll c_n$. Example of sequence (c_n) that satisfies the conditions in the Theorem 3: $c_n^2 = \lambda_n \log n$, for any sequence (λ_n) converging to zero and $n^{1/2}\lambda_n \to \infty$, as $n \to \infty$. An example of tuning parameter sequence (λ_n) is the following $\lambda_n = n^{-2/5}$.

By the Theorem 1 of Wu and Liu,[2] for tuning sequence $\lambda_{(l_{r-1},l_r)}$ converging to zero as $n \to \infty$, we have that the convergence rate of the estimators of $\boldsymbol{\phi}$ in each segment is of order $(l_r^0 - l_{r-1}^0)^{-1/2}$. Hence, taking into account Theorem 3, we deduce that $\|\widehat{\boldsymbol{\phi}}_{(\widehat{l}_{r-1}^{(\tau,\lambda)};\widehat{l}_r^{(\tau,\lambda)})}^{(\tau,\lambda)} - \boldsymbol{\phi}_r^0\|_2 = O_{\mathbb{P}}(\widehat{l}_r^{(\tau,\lambda)} - \widehat{l}_{r-1}^{(\tau,\lambda)})^{-1/2}$, for every $r = 1, \cdots, K+1$, with $\widehat{l}_0^{(\tau,\lambda)} = 1$ and $\widehat{l}_{K+1}^{(\tau,\lambda)} = n$.

We suppose that for each interval we have that the matrix $(l_r - l_{r-1})^{-1} \sum_{i=l_{r-1}+1}^{l_r} \mathbf{X}_i \mathbf{X}_i^t$ converges to $\mathbf{C}_r$, as $n \to \infty$, with $\mathbf{C}_r$ a non-negative definite matrix, which can be singular. Let us denote by $\mathbf{C}_r^0$ the limiting matrix for the true change-points l_r^0, $r = 1, \cdots, K$. We also denote by $C_{r,kj}^0$ the (k,j)th component of matrix $\mathbf{C}_r^0$.

The following result proves that on every segment, the SCAD estimator for the regression parameters has the oracle properties: nonzero parameters estimator on each estimated segment is asymptotically normal and zero parameters are shrunk directly to 0 with a probability converging to 1. Let us underline that the limiting distribution not depend on the penalty $\mathbf{p}_\lambda$, but only of the quantile order τ. For that purpose, for each two consecutive true change-points l_{r-1}^0, l_r^0 consider the set with the index of nonzero components of the true regression parameters

$$\mathscr{A}_{(l_{r-1}^0,l_r^0)} \equiv \{j; \phi_{r,j}^0 \neq 0\} =^{noted} \mathscr{A}_r^0 \tag{12}$$

and with the index of the nonzero components of the SCAD regression parameter estimator $\mathscr{A}_{n;(l_{r-1}^0,l_r^0)} \equiv \{j; \widehat{\phi}_{(l_{r-1}^0,l_r^0),j}^{(\tau,\lambda)} \neq 0\}$. Consider also the similar index set when the change-points are estimated $\mathscr{A}_{n;(\widehat{l}_{r-1}^{(\tau,\lambda)},\widehat{l}_r^{(\tau,\lambda)})} \equiv \{j; \widehat{\phi}_{(\widehat{l}_{r-1}^{(\tau,\lambda)},\widehat{l}_r^{(\tau,\lambda)}),j}^{(\tau,\lambda)} \neq 0\}$. We denoted by $\boldsymbol{\phi}_{\mathscr{A}_r^0}$ the sub-vector of $\boldsymbol{\phi}$ containing the corresponding components of $\mathscr{A}_r^0$ and by $q_r \equiv Card\{\mathscr{A}_r^0\}$ the true number of nonzero components in the rth segment.

Theorem 4. *Under the assumptions (A1)-A4), the tuning parameter sequence $(\lambda_{(l_{r-1},l_r)})_{1 \leqslant r \leqslant K+1}$ on each interval (l_{r-1},l_r) as in Theorem 3, then we have*
(i) $(\widehat{l}_r^{(\tau,\lambda)} - \widehat{l}_{r-1}^{(\tau,\lambda)})^{1/2}(\widehat{\boldsymbol{\phi}}_{(\widehat{l}_{r-1}^{(\tau,\lambda)};\widehat{l}_r^{(\tau,\lambda)})}^{(\tau,\lambda)} - \boldsymbol{\phi}_r^0)_{\mathscr{A}_r^0} = (l_r^0 - l_{r-1}^0)^{1/2}(\widehat{\boldsymbol{\phi}}_{(\widehat{l}_{r-1}^{(\tau,\lambda)};\widehat{l}_r^{(\tau,\lambda)})}^{(\tau,\lambda)} - \boldsymbol{\phi}_r^0)_{\mathscr{A}_r^0}(1 + o_{\mathbb{P}}(1)) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(\mathbf{0}, \tau(1 - \tau)/f^2(0)(\Omega_r^0)^{-1})$, where $\Omega_r^0 \equiv (C_{r,kj}^0)_{k,j\in\mathscr{A}_{(l_{r-1}^0,l_r^0)}}$ is a $q_r \times q_r$ matrix.*
(ii) $\lim_{n\to\infty} \mathbb{P}[\mathscr{A}_{n;(l_{r-1}^0,l_r^0)} = \mathscr{A}_{n;(\widehat{l}_{r-1}^{(\tau,\lambda)},\widehat{l}_r^{(\tau,\lambda)})} = \mathscr{A}_r^0] = 1.$*

4. LASSO-type estimator

An important theoretical fact is that, as Zou showed recently,[11] the oracle properties do not hold for the LASSO estimator. We have just seen that considering the SCAD method, the obtained estimators have this property. But the last method is difficult to put into practice with regard to numerical algorithms. Thus, Xu and Ying proposed,[3] for model (1), that the tuning parameter λ change from one component to the other of the parameter $\boldsymbol{\phi}$.

In this section the median model ($\tau = 1/2$) is studied.

Let us first consider, the model (1) without change-points, mentioned in Section 2. The parameter $\boldsymbol{\phi}$ is estimate by

$$\widehat{\boldsymbol{\phi}}_n^L = \arg\min_{\boldsymbol{\phi}}(\sum_{i=1}^n |Y_i - \mathbf{X}_i^t \boldsymbol{\phi}| + \boldsymbol{\lambda}_n^t |\boldsymbol{\phi}|). \tag{13}$$

Compared with the SCAD method seen in the previous section, now, the tuning parameter $\boldsymbol{\lambda}_n = (\lambda_{n,1}, \cdots, \lambda_{n,p})$ is a random p-vector with different components. The fact that $\boldsymbol{\lambda}_n$ has different components, makes possible that the estimator $\widehat{\boldsymbol{\phi}}_n^L$ have the oracle property, obviously, choosing the components of $\boldsymbol{\lambda}_n$ in a judicious way.

For the regression model (1) without change-points, and for the estimator $\widehat{\boldsymbol{\phi}}_n^L$ given by (13), consider the index set of estimator nonzero components $\mathscr{A}_n^L \equiv \{j; \widehat{\phi}_{n,j}^L \neq 0\}$ where $\widehat{\phi}_{n,j}^L$ the jth component of $\widehat{\boldsymbol{\phi}}_n^L$. Similar to the Proposition 2, we obtain that the KKT relations are in this case:

$-\sum_{i=1}^n X_{ij} sgn(Y_i - \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^L) + \lambda_{n,j} sgn(\widehat{\phi}_{n,j}^L) = 0$, for all $j \in \mathscr{A}_n^L$,

$\left| \sum_{i=1}^n X_{ij} sgn(Y_i - \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^L) \right| \leqslant \lambda_{n,j}$, for all $j \notin \mathscr{A}_n^L$,

with $\lambda_{n,j}$ the jth component of $\boldsymbol{\lambda}_n$ and $\widehat{\phi}_{n,j}^L$ of $\widehat{\boldsymbol{\phi}}_n^L$. These results will be useful to prove the oracle properties for the LASSO-type estimators of the regression parameters on each segment, in a model with change-points.

Consider now the change-point problem (7), with K (known) changes. For this estimation method, the change-point estimator is

$$(\widehat{l}_1^L, \cdots, \widehat{l}_K^L) \equiv \arg\min_{(l_1, \cdots, l_K) \in \mathbb{R}^K} \sum_{r=1}^{K+1} \inf_{(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{K+1})} \sum_{i=l_{r-1}+1}^{l_r} [|Y_i - \mathbf{X}_i^t \boldsymbol{\phi}_r| + \frac{\boldsymbol{\lambda}_{n,(l_{r-1},l_r)}^t}{l_r - l_{r-1}} |\boldsymbol{\phi}_r|].$$

The LASSO-type estimator of the regression parameters for the rth segment is $\widehat{\boldsymbol{\phi}}_{(\widehat{l}_{r-1}^L, \widehat{l}_r^L)}^L$, for each $r = 1, \cdots, K+1$, with $\widehat{l}_0^L = 1$ and $\widehat{l}_{K+1}^L = n$. Taking into account that a particular case ($\tau = 1/2$) to the quantile regression is considered, following processes are introduced

$$\begin{aligned} G_i^{(1/2)}(\boldsymbol{\phi}, \boldsymbol{\phi}^0) &\equiv |\varepsilon_i - \mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi}^0)| - |\varepsilon_i|, \qquad i = 1, \cdots, n \\ \eta_{i;(j_1,j_2)}^L(\boldsymbol{\phi}, \boldsymbol{\phi}^0) &\equiv G_i^{(1/2)}(\boldsymbol{\phi}, \boldsymbol{\phi}^0) + (j_2 - j_1)^{-1} \boldsymbol{\lambda}_{n;(j_1,j_2)}^t(|\boldsymbol{\phi}| - |\boldsymbol{\phi}^0|), \qquad i = j_1 + 1, \cdots, j_2, \end{aligned} \tag{14}$$

with $0 \leqslant j_1 < j_2 \leqslant n$ and $\boldsymbol{\phi}^0$ the true parameter. In the particular case $j_1 = 0$ and $j_2 = n$, let us denote $\boldsymbol{\lambda}_{n;(0,n)}$ by $\boldsymbol{\lambda}_n$.

Observe that, since we will study the model (7) with change-points, by the least absolute deviation method ($\tau = 1/2$) with LASSO-type penalty, the related results are obtained when there is no penalty by Bai,[8] Ciuperca are needed[9].

Following result yields that, even if the penalty is different, this estimator has the same convergence rate as the estimator obtained by the SCAD method.

Theorem 5. *If the tuning parameter $\boldsymbol{\lambda}_{n,(l_{r-1},l_r)}$ satisfies the conditions $\|\boldsymbol{\lambda}_{n,(l_{r-1},l_r)}^t\|_2 \to \infty$, $(l_r - l_{r-1})^{-1/2} \|\boldsymbol{\lambda}_{n,(l_{r-1},l_r)}^t\|_2 \xrightarrow[n\to\infty]{\mathbb{P}} M \geqslant 0$, under the assumptions (A1)-(A4), we have $\widehat{l}_r^L - l_r^0 = O_{\mathbb{P}}(1)$, for every $r = 1, \cdots, K$.*

Combining the Theorem 5 and the $\sqrt{n}$-consistency of the parameter estimator in a model without change-points (see Theorem 2 of (Ref. 3)), we have that the convergence rate of the regression parameter LASSO-type estimator on each segment is $\|\widehat{\boldsymbol{\phi}}^{L}_{(\widehat{l}^{L}_{r-1},\widehat{l}^{L}_{r})} - \boldsymbol{\phi}^{0}_r\|_2 = (l^0_r - l^0_{r-1})^{-1/2} O_{I\!\!P}(1)$, for $r = 1, \cdots, K+1$, with $l^0_0 = 1$ and $l^0_{K+1} = n$.

For this type of method, the most important is to verify that if the oracle properties are preserved in a change-point model. The sparsity property is the most interesting and it risk to be influenced by the change-point estimation. We would like to point out that, due to a penalty different, the proof of this result differs from that for the SCAD estimator.

Theorem 6. *Under the assumptions (A1)-(A4), with the tuning sequence $(\boldsymbol{\lambda}_{n,(l_{r-1},l_r)})$ as in Theorem 5 and the index set $\mathscr{A}^0_r$ defined by (12), we have:*
(i) $(\widehat{l}^{L}_r - \widehat{l}^{L}_{r-1})^{1/2}(\widehat{\boldsymbol{\phi}}^{L}_{(\widehat{l}^{L}_{r-1};\widehat{l}^{L}_r)} - \boldsymbol{\phi}^0_r)_{\mathscr{A}^0_r} = (l^0_r - l^0_{r-1})^{1/2}(\widehat{\boldsymbol{\phi}}^{L}_{(\widehat{l}^{L}_{r-1};\widehat{l}^{L}_r)} - \boldsymbol{\phi}^0_r)_{\mathscr{A}^0_r}(1 + o_{I\!\!P}(1))$ converges in distribution to the p-dimensional Gaussian vector $\mathscr{N}(\mathbf{0}, 1/(4f^2(0))(\Omega^0_r)^{-1})$, as $n \to \infty$.
(ii) $\lim_{n \to \infty} I\!\!P[\mathscr{A}^{L}_{n;(l^0_{r-1},l^0_r)} = \mathscr{A}^{L}_{n;(\widehat{l}^{L}_{r-1},\widehat{l}^{L}_r)} = \mathscr{A}^0_r] = 1$, with $\mathscr{A}^{L}_{n;(l^0_{r-1},l^0_r)} \equiv \{j; \widehat{\phi}^{L}_{(l^0_{r-1},l^0_r),j} \neq 0\}$ and $\mathscr{A}^{L}_{n;(\widehat{l}^{L}_{r-1},\widehat{l}^{L}_r)} \equiv \{j; \widehat{\phi}^{L}_{(\widehat{l}^{L}_{r-1},\widehat{l}^{L}_r),j} \neq 0\}$.

It is worthwhile to mention that, if the same model (7) is estimated by least squares, under certain conditions on design $(\mathbf{X}_i)$, with a LASSO penalty, the sparsity property (i.e. the claim (ii) of the Theorem 6), is not satisfied (see (Ref. 11)). Moreover, as the model contains change-points, this condition is more difficult to check on each interval that has random bounds. Then, an adaptive LASSO method can be considered downside to remedy this. But, it is necessary that in each segment (l_{r-1}, l_r) parameter number is smaller than observation number $l_{r-1} - l_r$. On the other hand, the adaptive LASSO for least squares method holds only under the assumptions that the errors have mean zero and bounded variance.

An example of tuning random sequence $\boldsymbol{\lambda}_{n,(l_{r-1},l_r)}$: in each segment (l_{r-1}, l_r) the LAD estimator $\widehat{\boldsymbol{\phi}}^{(1/2)}_{(l_{r-1},l_r)}$ of $\boldsymbol{\phi}_r$ is calculated by a corresponding relation to (3), for $\tau = 1/2$. Obtained estimators have all nonzero components and they have a convergence rate $v_{n,(l_{r-1},l_r)}$ to the true parameter, with $(l_r - l_{r-1})v_{n,(l_{r-1},l_r)} \to \infty$ (see Theorem 1 of (Ref. 9)). Consider then $\boldsymbol{\lambda}_{n,(l_{r-1},l_r)} = \left(\dfrac{1}{|\widehat{\phi}^{(1/2)}_{(l_{r-1},l_r),1}|}, \cdots, \dfrac{1}{|\widehat{\phi}^{(1/2)}_{(l_{r-1},l_r),p}|} \right)$.

5. Simulation study

We now give some simulation results. All simulations were performed using the R language. To calculate Least squares estimation the function *lm* was used. While, for the quantile estimations, SCAD and LASSO-type, the function *rq* of the package *quantreq* were called. To compare these estimates when the classical conditions on the error distribution do not occur, we consider also the adaptive LASSO estimation using the function *lqa* of the package *lqa* and quantile estimation with LASSO penalty.

The number of phases is assumed to be known: the models contain two change-points (three phases). We consider 10 latent variables $X_1, \cdots, X_{10}$ with $X_3 \sim \mathscr{N}(2,1)$, $X_4 \sim \mathscr{N}(4,1)$, $X_5 \sim \mathscr{N}(1,1)$ and $X_j \sim \mathscr{N}(0,1)$ for $j \in \{1,2,6,7,8,9,10\}$. The true values of the regression parameters (coefficients) on the three segments are respectively: $(1,0,4,0,-3,5,6,0,-1,0)$, $(0,3,-4,-3,0,1,2,-3,0,10)$, $(1,3,4,0,0,1,0,0,0,1)$. Three error patterns were considered: exponential, Cauchy and standard normal distributions. For the exponential er-

Table 1: Median of change-point estimations, percentage of true 0 and of false 0 by LS, QUANT, QLASSO, SCAD, LASSO-type and adaptive LASSO methods for $n = 200$, $K = 2$, $l_1^0 = 30$, $l_2^0 = 100$, $\varepsilon_i \sim \mathscr{E}xp(-1.5,1)$.

Method	LS	QUANT	QLASSO	SCAD	LASSO-type	aLASSO
median of $(\widehat{l_1}, \widehat{l_2})$	(31,100)	(31,100)	(31,100)	(30,100)	(30,100)	(30,100)
% of trues 0	0	0	46	75	97	94
% of false 0	0	0	1	3	3	8

Table 2: Median of change-point estimations, percentage of true 0 and of false 0 by LS, QUANT, QLASSO, SCAD, LASSO-type and adaptive LASSO methods for $n = 200$, $K = 2$, $l_1^0 = 30$, $l_2^0 = 100$, $\varepsilon_i \sim \mathscr{N}(0,1)$.

Method	LS	QUANT	QLASSO	SCAD	LASSO-type	aLASSO
median of $(\widehat{l_1}, \widehat{l_2})$	(31,100)	(30,100)	(30,100)	(30,100)	(30,100)	(30,100)
% of trues 0	0	0	37	65	98	94
% of false 0	0	0	0.5	5	2	8

Table 3: Median of change-point estimations, percentage of true 0 and of false 0 by LS, QUANT, QLASSO, SCAD, LASSO-type and adaptive LASSO methods for $n = 200$, $K = 2$, $l_1^0 = 30$, $l_2^0 = 100$, $\varepsilon_i \sim Cauchy$.

Method	LS	QUANT	QLASSO	SCAD	LASSO-type	aLASSO
median of $(\widehat{l_1}, \widehat{l_2})$	(31,100)	(30.5,100)	(30,100)	(30,100)	(30,100)	(30,100)
% of trues 0	0	0	36	62	95	48
% of false 0	0	0	1	3	3	12

Table 4: Median of change-point estimations, percentage of true 0 and of false 0 by LS, QUANT, QLASSO, LASSO-type and adaptive LASSO methods for $n = 60$, $K = 2$, $l_1^0 = 17$, $l_2^0 = 40$, $\varepsilon_i \sim \mathscr{E}xp(-1.5,1)$.

Method	LS	QUANT	QLASSO	LASSO-type	aLASSO
median of $(\widehat{l_1}, \widehat{l_2})$	(18,40)	(18,40)	(18,40)	(18,40)	(17,40)
% of trues 0	0	0	60	91	75
% of false 0	0	0	27	27	17

Table 5: Median of change-point estimations, percentage of true 0 and of false 0 by LS, QUANT, QLASSO, LASSO-type and adaptive LASSO methods for $n = 60$, $K = 2$, $l_1^0 = 17$, $l_2^0 = 40$, $\varepsilon_i \sim \mathscr{N}(0,1)$.

Method	LS	QUANT	QLASSO	LASSO-type	aLASSO
median of $(\widehat{l_1}, \widehat{l_2})$	(18,40)	(18,40)	(18,40)	(17,40)	(17,40)
% of trues 0	0	0	51	92.5	82
% of false 0	0	0	7	13	15

Table 6: Median of change-point estimations, percentage of true 0 and of false 0 by LS, QUANT, QLASSO, LASSO-type and adaptive LASSO methods for $n = 60$, $K = 2$, $l_1^0 = 17$, $l_2^0 = 40$, $\varepsilon_i \sim Cauchy$.

Method	LS	QUANT	QLASSO	LASSO-type	aLASSO
median of $(\widehat{l_1}, \widehat{l_2})$	(17,40)	(18,40)	(18,40)	(17,40)	(17,40)
% of trues 0	0	0	48.5	82	43
% of false 0	0	0	18	26	16

Table 7: The average of estimation error $\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^0\|_1$ in each segment under different distributions for LASSO-type and adaptive LASSO methods, $n = 200$, $K = 2$, $l_1^0 = 30$, $l_2^0 = 100$.

	LASSO-	type		adaptive	LASSO	
	$(1, l_1)$	$(l_1; l_2)$	(l_2, n)	$(1, l_1)$	$(l_1; l_2)$	(l_2, n)
$\varepsilon_i \sim \mathcal{N}(0,1)$	0.33	0.11	0.12	0.74	0.48	0.47
$\varepsilon_i \sim \mathcal{E}xp(-1.5,1)$	0.38	0.11	0.12	0.82	0.52	0.48
$\varepsilon_i \sim Cauchy$	0.51	0.17	0.13	4.6	4.84	4.83

rors, we generate a *n*-sample of distribution $\mathcal{E}xp(-1.5, 1)$, with the density $\exp(-(x+1.5))\, \mathbb{1}_{x>-1.5}$. For each model, we generated 500 Monte-Carlo random samples of size n, with $n = 60$ or $n = 200$. The percentage of zero coefficients correctly estimated to zero(true 0) and the percentage of nonzero coefficients estimated to zero(false 0) are computed (see Tables 1-6) by least squares(LS), quantile(QUANT), quantile with LASSO penalty(QLASSO), SCAD, LASSO-type and adaptive LASSO methods. The reader can find in the paper of Ciuperca more details on the adaptive LASSO method in a change-point model[5]. The adaptive LASSO estimators of the change-points and of the regression parameters are the minimizers of the following penalized sum $\sum_{r=1}^{K+1}[\sum_{i=l_{r-1}+1}^{l_r}(Y_i + \mathbf{X}_i^t\boldsymbol{\phi})^2 + \lambda_{n;(l_{r-1},l_r)}\widehat{\boldsymbol{\omega}}_{(l_{r-1},l_r)}|\boldsymbol{\phi}|]$, where the adaptive penalty *p*-vector $\widehat{\boldsymbol{\omega}}_{(l_{r-1},l_r)}$ is considered here that $|\widehat{\boldsymbol{\phi}}^{LS}_{(l_{r-1},l_r)}|^{-9/40}$. Let us specify that $\widehat{\boldsymbol{\phi}}^{LS}_{(l_{r-1},l_r)}$ is the LS estimator of $\boldsymbol{\phi}$ calculated between l_{r-1} and l_r. Recall also that the adaptive LASSO estimator of the regression parameters has the oracle properties under the assumptions for the errors ε that $\mathbb{E}[\varepsilon] = 0$ and $\mathbb{E}[\varepsilon^2] < \infty$. For the quantile method with the LASSO penalty, the sum $\sum_{i=l_{r-1}+1}^{l_r} \rho_\tau(Y_i + \mathbf{X}_i^t\boldsymbol{\phi})$ is penalized with $\lambda_{n;(l_{r-1},l_r)}|\boldsymbol{\phi}|$. The tuning parameters $\lambda_{n;(l_{r-1},l_r)}$ are $\log(l_r - l_{r-1})\mathbf{1}_p$ for the quantile estimation with LASSO penalty, $(l_r - l_{r-1})^{-2/5}$ for SCAD and $(l_r - l_{r-1})^{2/5}$ for adaptive LASSO methods. For the LASSO-type method, the tuning parameter is $(l_r - l_{r-1})^{2/5}/\widehat{\boldsymbol{\phi}}^{QLASSO}_{(l_{r-1},l_r)}$, where $\widehat{\boldsymbol{\phi}}^{QLASSO}_{(l_{r-1},l_r)}$ is the corresponding estimate by the quantile method (for the index quantile of the errors equal to τ) with LASSO penalty. Since the asymptotic distribution of the change-points estimators can not be symmetric, in each table we also give the median of the change-point estimations. Because the results by the SCAD method are poorer than by the LASSO-type method, and also because there may be convergence problem (the function *rq* not responding), in Tables 4-6 the SCAD estimator is not considered.

The outliers of the errors do not affect the precision of the change-point estimations, by all six methods, while the sparsity property of the QLASSO and adaptive LASSO are affected. More specifically, when n is large enough ($n = 200$) and the errors ε are normal, then $\mathbb{E}[\varepsilon] = 0$ and $\mathbb{E}[\varepsilon^2] < \infty$, the two methods, adaptive LASSO and LASSO-type, give the same (very satisfactory) sparsity results (as (Ref. 5), also indicates, for the adaptive LASSO estimators in a change-point model). When n or number of observations in a segment is small, the LASSO-type method is better than the adaptive LASSO method, in terms of detection of irrelevant regressors (true and false zeros). If the errors are $\mathcal{E}(-1.5, 1)$, then $\mathbb{E}[\varepsilon] \neq 0$, the results for LASSO-type are relatively better than for adaptive LASSO method (see Tables 1 and 4). This difference is accentuated when

the moments of errors don't exist, $\varepsilon \sim$ *Cauchy* (see Tables 3 and 6). Since LASSO-type and adaptive LASSO methods gave the best results, we calculate the average of estimation error $\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^0\|_1$ in each segment, for the index corresponding to the true values different to zero, over 500 simulations for different error distributions, for $n = 200$ and two change-points $l_1^0 = 30$, $l_2^0 = 100$ (see Table 7). For Gaussian and exponential distributions these two estimation methods, yield similar results. On the other hand, for Cauchy distribution, the obtained estimations by adaptive LASSO method are biased.

In conclusion, the LASSO-type method provides very satisfactory estimations in any case even for small sample size. The only less favorable result is obtained for $n = 60$ when the errors are exponential. The percentage of false zero is large enough.

6. Proofs of Theorems and Propositions

In order to simplify the proofs of theorems and propositions, we give in this section their demonstrations and in Section 6 some lemmas and their proofs which will useful.

Proof. of Proposition 1
Let us consider the notations $h_i(\boldsymbol{\phi}) = \mathbf{X}_i^t(\boldsymbol{\phi}^0 - \boldsymbol{\phi})$ and $F(x)$ for the distribution function of ε_i. By definition $I\!E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] = \int_{\mathbb{R}}[\rho_\tau(x + h_i(\boldsymbol{\phi})) - \rho_\tau(x)]dF(x)$. Using $F(0) = \tau$, a simple algebraic computation gives:

$$I\!E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] = \int_0^{-h_i(\boldsymbol{\phi})} [|h_i(\boldsymbol{\phi})| - x]dF(x), \qquad \text{if } h_i(\boldsymbol{\phi}) < 0 \tag{15}$$

and

$$I\!E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] = \int_{-h_i(\boldsymbol{\phi})}^0 [|h_i(\boldsymbol{\phi})| + x]dF(x), \qquad \text{if } h_i(\boldsymbol{\phi}) \geqslant 0. \tag{16}$$

Taking into account the relations (15) and (16) we can write $I\!E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] \geqslant \mathbb{1}_{h_i(\boldsymbol{\phi})\geqslant 0} \int_{-\frac{h_i(\boldsymbol{\phi})}{2}}^0 [|h_i(\boldsymbol{\phi})| + x]dF(x) + \mathbb{1}_{h_i(\boldsymbol{\phi})<0} \int_0^{-\frac{h_i(\boldsymbol{\phi})}{2}} [|h_i(\boldsymbol{\phi})| - x]dF(x) \geqslant \mathbb{1}_{h_i(\boldsymbol{\phi})\geqslant 0} \frac{|h_i(\boldsymbol{\phi})|}{2} \int_{-\frac{h_i(\boldsymbol{\phi})}{2}}^0 dF(x) + \mathbb{1}_{h_i(\boldsymbol{\phi})<0} \frac{|h_i(\boldsymbol{\phi})|}{2} \int_0^{-\frac{h_i(\boldsymbol{\phi})}{2}} dF(x) = \frac{|h_i(\boldsymbol{\phi})|}{2} [\mathbb{1}_{h_i(\boldsymbol{\phi})\geqslant 0}[F(0) - F(-\frac{h_i(\boldsymbol{\phi})}{2})] + \mathbb{1}_{h_i(\boldsymbol{\phi})<0}[F(-\frac{h_i(\boldsymbol{\phi})}{2}) - F(0)]] \geqslant 0$.
Hence, $I\!E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] \geqslant 0$, for all $i = 1, \cdots, n$, $\boldsymbol{\phi} \in \Gamma$. □

6.1. For SCAD estimator

Proof. of Proposition 2
    If $j \in \mathscr{A}_n$. According to the definition (8), the SCAD estimator of $\boldsymbol{\phi}$ is the solution of the following equation
$0 = \sum_{i=1}^n \frac{\partial G_i^{(\tau,\lambda)}(\widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)})}{\partial \phi_{.j}} = \sum_{i=1}^n (-\tau X_{ij} + X_{ij} \mathbb{1}_{Y_i < \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}} + \lambda \{\mathbb{1}_{|\widehat{\phi}_{n,j}^{(\tau,\lambda)}| \leqslant \lambda} sgn(\widehat{\phi}_{n,j}^{(\tau,\lambda)}) + \frac{(a\lambda - |\widehat{\phi}_{n,j}^{(\tau,\lambda)}|)_+}{(a-1)\lambda} sgn(\widehat{\phi}_{n,j}^{(\tau,\lambda)}) \mathbb{1}_{|\widehat{\phi}_{n,j}^{(\tau,\lambda)}| > \lambda}\})$.
We obtain $\tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} \mathbb{1}_{Y_i < \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}} = n\lambda sgn(\widehat{\phi}_{n,j}^{(\tau,\lambda)}) \{\mathbb{1}_{|\widehat{\phi}_{n,j}^{(\tau,\lambda)}| \leqslant \lambda} + \frac{(a\lambda - |\widehat{\phi}_{n,j}^{(\tau,\lambda)}|)_+}{(a-1)\lambda} \mathbb{1}_{|\widehat{\phi}_{n,j}^{(\tau,\lambda)}| > \lambda}\}$.

    If $j \notin \mathscr{A}_n$. In this case $0 \in \sum_{i=1}^n \frac{\partial G_{i,n}^{(\tau,\lambda)}(\widehat{\phi}_{n,j}^{(\tau,\lambda)})}{\partial \phi_{.j}} = -\tau \sum_{i=1}^n X_{ij} + \sum_{i=1}^n X_{ij} \mathbb{1}_{Y_i < \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}} + \sum_{i=1}^n p'_\lambda(|\widehat{\phi}_{n,j}^{(\tau,\lambda)}|)$. Since
$p'_\lambda(|\phi_{.j}|) \equiv \frac{\partial p_\lambda(|\phi_{.j}|)}{\partial \phi_{.j}} = \lambda sgn(\phi_{.j}) \mathbb{1}_{|\phi_{.j}| \leqslant \lambda} + \frac{(a\lambda - |\phi_{.j}|)_+ sgn(\phi_{.j})}{a-1} \mathbb{1}_{|\phi_{.j}| > \lambda}$, it follows that $0 \in \sum_{i=1}^n \frac{\partial G_{i,n}^{(\tau,\lambda)}(\widehat{\phi}_{n,j}^{(\tau,\lambda)})}{\partial \phi_{.j}} = -\tau \sum_{i=1}^n X_{ij} + \sum_{i=1}^n X_{ij} \mathbb{1}_{Y_i < \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}} + n\lambda \cdot [-1,1]$. Then, $|\tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} \mathbb{1}_{Y_i < \mathbf{X}_i^t \widehat{\boldsymbol{\phi}}_n^{(\tau,\lambda)}}| \leqslant n\lambda$. □

Proof. of Theorem 3

The proof is similar to that of Theorem 5. It is omitted. The Lemmas 8 and 11 stated in Section 6 are needed. □

Proof of Theorem 4

(i) The statement results from Theorem 3 together with Theorem 2(b) of (Ref. 2).

(ii) By the Theorem 2(a) of Wu and Liu,[2] we have: $\lim_{n\to\infty} I\!\!P[\mathscr{A}_{n;(l_{r-1}^0,l_r^0)} = \mathscr{A}_{(l_{r-1}^0,l_r^0)}] = 1$. The asymptotic normality of the estimators implies: for all $k \in \mathscr{A}_{(l_{r-1}^0,l_r^0)}$, we have $\phi_{r,k}^0 - \widehat{\phi}_{(\widehat{l}_{r-1}^{(\tau,\lambda)},\widehat{l}_r^{(\tau,\lambda)}),k}^{(\tau,\lambda)} \xrightarrow[n\to\infty]{I\!\!P} 0$. It follows that

$$\lim_{n\to\infty} I\!\!P\left[\mathscr{A}_{n;(\widehat{l}_{r-1}^{(\tau,\lambda)},\widehat{l}_r^{(\tau,\lambda)})} \supseteq \mathscr{A}_{(l_{r-1}^0,l_r^0)}\right] = 1. \tag{17}$$

By similar arguments that for the Lemma 1 of Wu and Liu,[2] we prove that

$$I\!\!P[\exists k \in \{1,\cdots,p\}, k \notin \mathscr{A}_{(l_{r-1}^0,l_r^0)}, k \in \mathscr{A}_{n;(\widehat{l}_{r-1}^{(\tau,\lambda)},\widehat{l}_r^{(\tau,\lambda)})}] \xrightarrow[n\to\infty]{} 0. \tag{18}$$

The Theorem results from the relations (17) and (18). □

6.2. For LASSO-type estimator

Proof. of Theorem 5

The proof has three steps. First, we show that the all SCAD estimators of the change-points are to a smaller distance than $n^{1/2}$ from the corresponding true value. Then, for each true change-point l_r^0, with $r \in \{1,\cdots,K\}$, we consider the function S given by (10), but calculated on the change-points $l_1,\cdots,l_K,l_1^0,\cdots,l_{r-1}^0,l_r^0 - [n^\alpha],l_r^0 + [n^\alpha],l_{r+1}^0,\cdots,l_K^0$, with $\alpha \in (1/2,1)$. For the penalized sums involving observations between l_{t-1}^0 and l_t^0, for $t \in \{1,\cdots,r-1,r+1,\cdots,K\}$, consider the change-points $k_{1,t} < \cdots < k_{J(t),t} \equiv \{l_1,\cdots,l_K\} \cap \{j; l_{r-1}^0 < j \leqslant l_r^0\}$. Then for each $t \in \{1,\cdots,r-1,r+1,\cdots,K\}$, we have $0 \geqslant \sum_{j=1}^{J(t)+1} \min_{\phi_j \in \Gamma} [\sum_{i=k_{j,t},t+1}^{k_{j,t}} |\varepsilon_i - \mathbf{X}_i^t(\phi_j - \phi_t^0)| + \lambda_{n;(k_{j-1,t};k_{j,t})}^t |\phi_j|] - \sum_{j=1}^{J(t)+1} [\sum_{i=k_{j-1,t},t+1}^{k_{j,t}} |\varepsilon_i| + \lambda_{n;(k_{j-1,t};k_{j,t})}^t |\phi_j^0|] \geqslant -2(K+1) \sup_{1\leqslant l<j\leqslant n} |\inf_\phi [\sum_{i=l+1}^j \eta_{i;(l,j)}^L(\phi,\phi^0)]|$, which is, using the Lemma 12, $-O_{I\!\!P}(\max(n^\alpha,\lambda_n))$, with $\alpha \in (1/2,1)$. The rest of proof is similar to that in (Ref. 5), Theorem 1, using also the Lemma 14 stated in Section 7 and the Remark 1. The details are omitted. □

Proof. of Theorem 6

(i) The assertion follows from the Theorem 5 and from the Theorem 3(b) of (Ref. 3).

(ii) By Xu and Ying,[3] we have: $\lim_{n\to\infty} I\!\!P[\mathscr{A}_{n;(l_{r-1}^0,l_r^0)}^L = \mathscr{A}_{(l_{r-1}^0,l_r^0)}] = 1$. The asymptotic normality of the estimators implies that, for all $k \in \mathscr{A}_{(l_{r-1}^0,l_r^0)}$ we have $\phi_{r,k}^0 - \widehat{\phi}_{n;(\widehat{l}_{r-1}^L,\widehat{l}_r^L),k}^L \xrightarrow[n\to\infty]{I\!\!P} 0$. Thus $k \in \mathscr{A}_{n;(\widehat{l}_{r-1}^L,\widehat{l}_r^L)}^L$. Hence $\lim_{n\to\infty} I\!\!P[\mathscr{A}_{n;(\widehat{l}_{r-1}^L,\widehat{l}_r^L)}^L \supseteq \mathscr{A}_{(l_{r-1}^0,l_r^0)}^L] = 1$. The proof is finished if we show the claim $I\!\!P[\exists k \in \{1,\cdots,p\}, k \notin \mathscr{A}_{(l_{r-1}^0,l_r^0)}, k \in \mathscr{A}_{n;(\widehat{l}_{r-1}^L,\widehat{l}_r^L)}^L] \to 0$, as $n \to \infty$. Since $k \in \mathscr{A}_{n;(\widehat{l}_{r-1}^L,\widehat{l}_r^L)}^L$ we have that, with the probability 1,

$$sgn(\widehat{\phi}_{(\widehat{l}_{r-1}^L,\widehat{l}_r^L),k}^L) \neq 0. \tag{19}$$

We suppose, without loss of generality, that $sgn(\widehat{\phi}_{(\widehat{l}_{r-1}^L,\widehat{l}_r^L),k}^L) = 1$. Then, using the KKT conditions, we have with the probability 1,

$$\lambda_{n,(\widehat{l}_{r-1}^L,\widehat{l}_r^L),k} sgn(\widehat{\phi}_{(\widehat{l}_{r-1}^L,\widehat{l}_r^L),k}^L) = \sum_{i=\widehat{l}_{r-1}^L+1}^{\widehat{l}_r^L} X_{ik} sgn(Y_i - \mathbf{X}_i^t \widehat{\phi}_{(\widehat{l}_{r-1}^L,\widehat{l}_r^L)}^L). \tag{20}$$

On the other hand, since $k \notin \mathscr{A}^L_{(l^0_{r-1}, l^0_r)}$, we have $\phi^0_{r,k} = 0$, then $sgn(\phi^0_{r,k}) = 0$. By the proof of the Proposition 4 in the paper of Xu and Ying,[3] for $\phi^0_{,k} = 0$, we have that for every ϕ such that $\|\phi_{,k}\| \leqslant Cn^{-1/2}$,

$$IP[sgn\left(-\sum_{i=1}^n X_{ik}sgn(Y_i - \mathbf{X}^t_i\phi) + \lambda_{n,k}sgn(\phi_{,k})\right) = sgn(\phi_{,k})]\underset{n\to\infty}{\longrightarrow} 1.$$

Then, taking into account the assertion (i), we apply the previous relation for $\widehat{\phi}^L_{(\widehat{l}^L_{r-1}; \widehat{l}^L_r), k}$ and we obtain

$$\lim_{n\to\infty} IP[sgn(-\sum_{i=1}^n X_{ik}sgn(Y_i - \mathbf{X}^t_i\widehat{\phi}^L_{(\widehat{l}^L_{r-1}; \widehat{l}^L_r)})) + \lambda_{n,(\widehat{l}^L_{r-1}; \widehat{l}^L_r),k}sgn(\widehat{\phi}^L_{(\widehat{l}^L_{r-1}; \widehat{l}^L_r),k}) = sgn(\phi_{,k})] = 1,$$

where $\widehat{\phi}^L_{(\widehat{l}^L_{r-1}; \widehat{l}^L_r),k}$ is the kth component of the random vector $\widehat{\phi}^L_{(\widehat{l}^L_{r-1}; \widehat{l}^L_r)}$. Moreover, by (20), $sgn(-\sum_{i=1}^n X_{ik}sgn(Y_i - \mathbf{X}^t_i\widehat{\phi}^L_{(\widehat{l}^L_{r-1}; \widehat{l}^L_r)}) + \lambda_{n,(\widehat{l}^L_{r-1}; \widehat{l}^L_r),k}sgn(\widehat{\phi}^L_{(\widehat{l}^L_{r-1}; \widehat{l}^L_r),k)}))$ is 0, with the probability 1. Then $\lim_{n\to\infty} IP[0 = sgn(\phi_{,k})] = 1$. Contradiction with (19). Thus the claim holds. □

The demonstration of the Remark 1 is similar to that of Bai[8]. Then we give only the main idea.

Proof. of Remark 1
Similar as for the proof of the Lemma 7, we obtain, for $\|\phi_1 - \phi_2\|_2 \leqslant c_n n^{-1/2}$, that $(nc_n^2)^{-1}|(\mathscr{G}_n^{(\tau)}(\phi_1; \phi^0) - \mathscr{G}_n^{(\tau)}(\phi_2; \phi^0) - IE[\mathscr{G}_n^{(\tau)}(\phi_1; \phi^0)] + IE[\mathscr{G}_n^{(\tau)}(\phi_2; \phi^0)])| \leqslant Cc_n^{-2}\|\phi_1 - \phi_2\|_2 \leqslant Cn^{-1/2}c_n^{-1}$, which converges to 0 for $n \to \infty$. We also have $IP[\sup_j |n^{-1}c_n^{-2}[\mathscr{G}_n^{(\tau)}(\phi_j; \phi^0) - IE[\mathscr{G}_n^{(\tau)}(\phi_j; \phi^0)]]| > \varepsilon] \leqslant \sum_j IP[|[\mathscr{G}_i^{(\tau)}(\phi_j; \phi^0) - IE[\mathscr{G}_n^{(\tau)}(\phi_j; \phi^0)]]| > nc_n^2\varepsilon]$, where $j = 1, \cdots, n^{p/2}$. By the relations (22) and (23), we have $|G_i^{(\tau)}(\phi; \phi^0) - IE[G_i^{(\tau)}(\phi; \phi^0)]| \leqslant C|\mathbf{X}^t_i(\phi - \phi^0)| < Cc_n$. The rest of proof is similar to that of the Lemma 4 of Bai[8]. □

7. Lemmas

We present in this section the lemmas with proofs, which are useful to prove the main results. Following Lemma gives the asymptotic behavior of the objective function $G_i^{(\tau)}$ without penalty. In fact, Lemma 7 will be necessary to prove the Lemmas 8 and 12, where the penalized objective functions are studied.

Lemma 7. *Under the assumptions (A1), (A3), for all $\alpha > 1/2$, we have* $\sup_{1\leqslant l<k\leqslant n}|\inf_{\phi\in\Gamma}\sum_{i=l}^k G_i^{(\tau)}(\phi; \phi^0)| = O_{IP}(n^\alpha)$.

Proof. of Lemma 7
By direct calculations $R_i^{(\tau)}(\phi_1; \phi^0) - R_i^{(\tau)}(\phi_2; \phi^0)$ can be written as

$$\{\mathbf{X}^t_i(\phi_1 - \phi_2)[(1 - \tau)1\!1_{\varepsilon_i\leqslant 0} - \tau 1\!1_{\varepsilon_i>0}])\} + \{[\varepsilon_i - \mathbf{X}^t_i(\phi_1 - \phi^0)][\tau 1\!1_{\varepsilon_i>\mathbf{X}^t_i(\phi_1-\phi^0)} - (1 - \tau)1\!1_{\varepsilon_i\leqslant\mathbf{X}^t_i(\phi_1-\phi^0)}]\}$$
$$-\{[\varepsilon_i - \mathbf{X}^t_i(\phi_2 - \phi^0)][\tau 1\!1_{\varepsilon_i>\mathbf{X}^t_i(\phi_2-\phi^0)} - (1 - \tau)1\!1_{\varepsilon_i\leqslant\mathbf{X}^t_i(\phi_2-\phi^0)}]\} \equiv S_{1,i} + S_{2,i} - S_{3,i}. \quad (21)$$

Obviously $S_{1,i} = \mathbf{X}^t_i(\phi_1 - \phi_2)D_i$. For $S_{2,i} - S_{3,i}$ we have:

If $\varepsilon_i > \mathbf{X}^t_i(\phi_1 - \phi^0)$. When $\varepsilon_i > \mathbf{X}^t_i(\phi_2 - \phi^0)$, we have $S_{2,i} - S_{3,i} = \tau\mathbf{X}^t_i(\phi_2 - \phi_1)$. In the case $\varepsilon_i \leqslant \mathbf{X}^t_i(\phi_2 - \phi^0)$, we have $S_{2,i} - S_{3,i} = \tau\mathbf{X}^t_i(\phi_2 - \phi_1) + [\varepsilon_i - \mathbf{X}^t_i(\phi_2 - \phi^0)] \leqslant \tau\mathbf{X}^t_i(\phi_2 - \phi_1)$. Then, in the both cases, $S_{2,i} - S_{3,i} \leqslant \tau\mathbf{X}^t_i(\phi_2 - \phi_1)$.

If $\varepsilon_i \leqslant \mathbf{X}^t_i(\phi_1 - \phi^0)$. When $\varepsilon_i \leqslant \mathbf{X}^t_i(\phi_2 - \phi^0)$, we have $S_{2,i} - S_{3,i} = (1 - \tau)\mathbf{X}^t_i(\phi_1 - \phi_2)$. In the case $\varepsilon_i >$

$\mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}^0)$, we have $S_{2,i} - S_{3,i} = (1-\tau)[\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2) + \varepsilon_i - \mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}^0)] \leqslant (1-\tau)\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)$. Then, in the both cases, $S_{2,i} - S_{3,i} \leqslant (1-\tau)\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)$.

In conclusion, with the probability 1,

$$S_{1,i} + S_{2,i} - S_{3,i} \leqslant \mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)D_i + \max(\tau\mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1), (1-\tau)\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)). \tag{22}$$

Similarly

$$S_{1,i} + S_{2,i} - S_{3,i} \geqslant \mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)D_i + \min(\tau\mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1), (1-\tau)\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)). \tag{23}$$

Hence, the relations (21), (22) and (23), for $\|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2\|_2 \leqslant Cn^{-1/2}$ together the assumption (A3), imply that

$$\begin{aligned}|\sum_{i=1}^n [R_i^{(\tau)}(\boldsymbol{\phi}_1; \boldsymbol{\phi}^0) - R_i^{(\tau)}(\boldsymbol{\phi}_2; \boldsymbol{\phi}^0) - E[R_i^{(\tau)}(\boldsymbol{\phi}_1; \boldsymbol{\phi}^0)] + E[R_i^{(\tau)}(\boldsymbol{\phi}_2; \boldsymbol{\phi}^0)]]| \\ \leqslant C\sum_{i=1}^n \|\mathbf{X}_i\|_2 \cdot \|\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1\|_2 \leqslant O_{I\!\!P}(n^{1/2}).\end{aligned} \tag{24}$$

By an argument similar to the one used in the Lemma 3 of Bai,[8] together the Proposition 1, we obtain

$$\sup_{1\leqslant l<k\leqslant n}|\inf_{\boldsymbol{\phi}}\sum_{i=l}^k G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)| \leqslant 2\sup_{1\leqslant k\leqslant n}\sup_{\boldsymbol{\phi}}|\sum_{i=1}^k [G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)]]|.$$

On the other hand $\sum_{i=1}^k [G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)]] = \sum_{i=1}^k [R_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - E[R_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)]] + \sum_{i=1}^k D_i\mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi}^0)$. Let us consider the random process $\xi_k = \sup_{\boldsymbol{\phi}}|\sum_{i=1}^k [G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - E[G_i^{(\tau;\boldsymbol{\phi}^0)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)]]|$. Then, since by Proposition 1 we have that $E[G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)] \geqslant 0$, follows that $\{\xi_k, \mathscr{F}_k\}_{k=1,\cdots,n}$ is a sub-martingale, where $\mathscr{F}_k = \sigma - field\{\varepsilon_1,\cdots,\varepsilon_k\}$, which implies, using Doob's inequality $I\!\!P[\sup_{1\leqslant k\leqslant n}\xi_k > n^\alpha] \leqslant n^{-\alpha m}C_m E[\xi_n^m], \quad C_m > 0$, with $m > 1$. We divide the parameter set Γ into $m^{p/2}$ cells, such that the cell diameter is $\leqslant n^{-1/2}$. Thus $|\sum_{i=1}^n [G_i^{(\tau)}(\boldsymbol{\phi}_1;\boldsymbol{\phi}^0) - E[G_i^{(\tau)}(\boldsymbol{\phi}_1;\boldsymbol{\phi}^0)] - G_i^{(\tau)}(\boldsymbol{\phi}_2;\boldsymbol{\phi}^0) + E[G_i^{(\tau)}(\boldsymbol{\phi}_2;\boldsymbol{\phi}^0)]]| \leqslant |\sum_{i=1}^n [R_i^{(\tau)}(\boldsymbol{\phi}_1;\boldsymbol{\phi}^0) - E[R_i^{(\tau)}(\boldsymbol{\phi}_1;\boldsymbol{\phi}^0)] - R_i^{(\tau)}(\boldsymbol{\phi}_2;\boldsymbol{\phi}^0) + E[R_i^{(\tau)}(\boldsymbol{\phi}_2,\boldsymbol{\phi}^0)]]| + |\sum_{i=1}^n D_i\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)|$ and using the relation (24), we obtain, with the probability 1, that the last relation is smaller than $\sum_{i=1}^n |\mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)| \leqslant Cnn^{-1/2} = Cn^{1/2}$. By an argument similar to the one used in Bai,[8] we have $E|\sum_{i=1}^n [G_i^{(\tau)}(\boldsymbol{\phi}_r;\boldsymbol{\phi}^0) - E[G_i^{(\tau)}(\boldsymbol{\phi}_r;\boldsymbol{\phi}^0)]]| \leqslant Cn^{m/2}$. The rest of proof is similar to that of the Lemma 3 of Bai[8]. $\qquad\square$

7.1. *For SCAD estimator*

Following result will be useful in the study of the convergence rate of the change-point SCAD estimator in a model with breaking.

Lemma 8. *Under the assumptions (A1), (A3), for a positive sequence $(\lambda_n)_n$ such that $\lambda_n \to 0$, we have*

$$\sup_{0\leqslant j_1<j_2\leqslant n}\left|\inf_{\boldsymbol{\phi}}\sum_{i=j_1+1}^{j_2} G_i^{(\tau,\lambda)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)\right| = O_{I\!\!P}(n^\alpha, n\lambda_n).$$

Proof. of Lemma 8

Using the triangle inequality, we deduce that

$$\sup_{0\leqslant j_1<j_2\leqslant n}|\inf_{\boldsymbol{\phi}}\sum_{i=j_1+1}^{j_2} G_i^{(\tau,\lambda)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)| \leqslant \sup_{0\leqslant j_1<j_2\leqslant n}|\inf_{\boldsymbol{\phi}}\sum_{i=j_1+1}^{j_2} G_i^{(\tau)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)| + n\sup_{\boldsymbol{\phi}}|(\mathbf{p}_\lambda(|\boldsymbol{\phi}|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}^0|))^t\mathbf{1}_p|.$$

Considering Lemma 7 and the definition of p'_λ, we have that the last quantity is smaller than $O_{I\!P}(n^\alpha) + n\lambda_n$. $\quad\square$

In the following Lemma, the behavior of $\mathscr{G}_n^{(\tau)}$ is studied in the outside of the ball center β^0 and radius c_n.

Lemma 9. *Under the assumptions (A1), (A2), with (c_n) a positive sequence such that $c_n \to 0$ and $nc_n^2/\log n \to \infty$, there exists $\varepsilon > 0$ such that we have with probability 1*

$$\liminf_{n\to\infty}\left(\inf_{\|\phi-\phi^0\|_2\geqslant c_n}\frac{1}{nc_n^2}\mathscr{G}_n^{(\tau)}(\phi;\phi^0)\right) \geqslant \varepsilon > 0.$$

Proof. of Lemma 9

Let $\mathbf{u}$ in an open subset of $\mathbb{R}^p$. By the proof of the Lemma 3 of Wu and Liu,[2] taking into account the assumptions (A1) and (A2), we have

$$I\!E[\mathscr{G}_n^{(\tau)}(\phi^0 + \frac{\mathbf{u}}{\sqrt{n}};\phi^0)] = \frac{f(0)}{2n}\mathbf{u}^t(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^t)\mathbf{u} + o(1).$$

If $c_n \to 0$ and $nc_n^2 \to \infty$, we have similarly $I\!E[\mathscr{G}_n^{(\tau)}(\phi^0 + \mathbf{u}c_n;\phi^0)] = \frac{f(0)}{2}c_n^2\mathbf{u}^t(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^t)\mathbf{u} + o_{I\!P}(1)$. The function $G_i^{(\tau)}(\phi;\phi^0)$ is convex, hence $\mathscr{G}_n^{(\tau)}(\phi;\phi^0)$ is convex in ϕ. Thus, its minimum over $\|\phi - \phi^0\|_2 \geqslant c_n$ is realized for $\|\phi - \phi^0\|_2 = c_n$. Then, for $\|\mathbf{u}\|_2 = 1$, using the assumption (A2) we obtain that $I\!E[\mathscr{G}_n^{(\tau)}(\phi^0 + \mathbf{u}c_n;\phi^0)] = \frac{f(0)}{2}nc_n^2(C + o(1))$. The rest of proof follows using the the Lemma 5 of Bai,[8] taking into account the relation (6).
$\square$

Lemma 10. *Under the assumptions (A1), (A2), for two positive sequences (c_n) and (λ_n) such that $\lambda_n \to 0$, $c_n \to 0$, $nc_n^2/\log n \to \infty$ and $\lambda_n c_n^{-2} \to 0$, we have, with probability 1,*

$$\liminf_{n\to\infty}\left(\inf_{\|\phi-\phi^0\|_2\geqslant c_n}\frac{1}{nc_n^2}\sum_{i=1}^n G_i^{(\tau,\lambda)}(\phi;\phi^0)\right) > \varepsilon.$$

Proof. of Lemma 10

Applying the mean value theorem, we write $G_i^{(\tau,\lambda)}(\phi;\phi^0) = G_i^{(\tau)}(\phi;\phi^0) + [|\phi| - |\phi^0|]^t\mathbf{p}'_\lambda(\widetilde{\phi})$, with $\widetilde{\phi} = \phi^0 - \mathbf{b}(\phi - \phi^0)$, $\mathbf{b} \in [0,1]^p$. Then, using the relation (9), we have, with probability 1,

$$\inf_{\|\phi-\phi^0\|_2\geqslant c_n}(nc_n^2)^{-1}\sum_{i=1}^n G_i^{(\tau,\lambda)}(\phi;\phi^0) \geqslant \inf_{\|\phi-\phi^0\|_2\geqslant c_n}(nc_n^2)^{-1}\mathscr{G}_n^{(\tau)}(\phi;\phi^0) - (n\lambda_n)/(nc_n^2).$$

Since $\lambda_n c_n^{-2} \to 0$, for every $\varepsilon > 0$ there exists a $n_\varepsilon \in \mathbb{N}$ such that $\lambda_n c_n^{-2} < \varepsilon/2$. An application of Lemma 9 leads to $\inf_{\|\phi-\phi^0\|_2\geqslant c_n}((nc_n^2)^{-1}\mathscr{G}_n^{(\tau)}(\phi;\phi^0)) > 3\varepsilon/2$, with probability 1, and Lemma is proved. $\quad\square$

By similar calculus as in Lemma 10 of Bai,[8] we have following result for the estimator (3) of ϕ.

Lemma 11. *For (λ_n), (c_n) as in the Lemma 10, under assumptions (A1)-(A3), for all $n_1, n_2 \in \mathbb{N}$ such that $n_1 \geqslant n^u$, with $3/4 \leqslant u \leqslant 1$, $n_2 \leqslant n^v$, $v < 1/4$, let us consider the model*

$$\begin{aligned} Y_i &= \mathbf{X}_i^t\phi_1^0 + \varepsilon_i, \quad &i = 1,\cdots,n_1 \\ Y_i &= \mathbf{X}_i^t\phi_2^0 + \varepsilon_i, \quad &i = n_1+1,\cdots,n_1+n_2 \end{aligned}$$

with the assumption $\phi_1^0 \neq \phi_2^0$. Consider $A_{n_1+n_2}^{(\tau,\lambda)}(\phi) = \sum_{i=1}^{n_1}G_i^{(\tau,\lambda)}(\phi;\phi_1^0) + \sum_{i=n_1+1}^{n_1+n_2}G_i^{(\tau,\lambda)}(\phi;\phi_2^0)$ and $\widehat{\phi}_{n_1+n_2}^{(\tau,\lambda)} \equiv \arg\min_\phi A_{n_1+n_2}^{(\tau,\lambda)}(\phi)$.

(i) $\|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{(\tau,\lambda)} - \boldsymbol{\phi}_1^0\|_2 \leqslant n_1^{-1/2} n_1^{\frac{v+\delta}{2u}} \leqslant n_1^{-(u-v-\delta)/2}$.

(ii) $\sum_{i=1}^{n_1} G_i^{(\tau,\lambda)} \left(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{(\tau,\lambda)}; \boldsymbol{\phi}_1^0 \right) = O_{I\!\!P}(1)$.

Proof. of Lemma 11

(i) $A_{n_1+n_2}^{(\tau,\lambda)}(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{\tau,\lambda}) \leqslant \sum_{i=n_1+1}^{n_1+n_2} G_i^{(\tau,\lambda)}(\boldsymbol{\phi}_1^0; \boldsymbol{\phi}_2^0) = \sum_{i=n_1+1}^{n_1+n_2} G_i^{(\tau)}(\boldsymbol{\phi}_1^0; \boldsymbol{\phi}_2^0) + n_2[\mathbf{p}_\lambda(|\boldsymbol{\phi}_1^0|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}_2^0|)]\mathbf{1}_p = o_{I\!\!P}(1) + O(n_2)$. By Lemma 10, for $G_i^{(\tau,\lambda)}$, $i = n_1 + 1, \cdots, n_2$, we arrive to a contradiction.

(ii) Let $Z_n^{(\tau)}(\boldsymbol{\phi}) \equiv \sum_{i=1}^{n_1} G_i^{(\tau)}(\boldsymbol{\phi}; \boldsymbol{\phi}_1^0), t_n^{(\tau)}(\boldsymbol{\phi}) \equiv \sum_{i=n_1+1}^{n_1+n_2}[\rho_\tau(\varepsilon_i - \mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi}_2^0)) - \rho_\tau(\varepsilon_i - \mathbf{X}_i^t(\boldsymbol{\phi}_1^0 - \boldsymbol{\phi}_2^0))], t_n^{(\tau,\lambda)}(\boldsymbol{\phi}) \equiv t_n^{(\tau)}(\boldsymbol{\phi}) + n_2[\mathbf{p}_\lambda(|\boldsymbol{\phi}|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}_1^0|)]\mathbf{1}_p, Z_n^{(\tau,\lambda)}(\boldsymbol{\phi}) \equiv Z_n^{(\tau)}(\boldsymbol{\phi}) + n_1[\mathbf{p}_\lambda(|\boldsymbol{\phi}|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}_1^0|)]\mathbf{1}_p$. Then

$$A_{n_1+n_2}^{(\tau,\lambda)}(\boldsymbol{\phi}) = Z_n^{(\tau,\lambda)}(\boldsymbol{\phi}) + t_n^{(\tau,\lambda)}(\boldsymbol{\phi}) + n_2[\mathbf{p}_\lambda(|\boldsymbol{\phi}_1^0|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}_2^0|)]\mathbf{1}_p + \sum_{i=n_1+1}^{n_1+n_2}[\rho_\tau(\varepsilon_i - \mathbf{X}_i^t(\boldsymbol{\phi}_1^0 - \boldsymbol{\phi}_2^0)) - \rho_\tau(\varepsilon_i)].$$

We have $|t_n^{(\tau,\lambda)}(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{(\tau,\lambda)})| \leqslant |t_n^{(\tau)}(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{(\tau,\lambda)})| + n_2\|\mathbf{p}_\lambda(|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{(\tau,\lambda)}|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}_1^0|)\|_1$ and similarly that for the relation (22), $\leqslant C\sum_{i=n_1+1}^{n_1+n_2} |\mathbf{X}_i|_1 \|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{(\tau,\lambda)} - \boldsymbol{\phi}_1^0\|_2 + n_2\|\mathbf{p}_\lambda(|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^{(\tau,\lambda)}|) - \mathbf{p}_\lambda(|\boldsymbol{\phi}_1^0|)\|_1$. The rest of proof is similar to that of the Lemma 3(ii) of Ciuperca,[5] taking into account the assumption (A3). $\square$

We have the equivalent of Lemma 4 of the same paper[5].

7.2. *For LASSO-type estimator*

Lemma 12. *Under the assumptions (A1), (A3), we have, for $\alpha > 1/2$, that*

$$\sup_{0 \leqslant j_1 < j_2 \leqslant n} |\inf_{\boldsymbol{\phi}} \sum_{i=j_1+1}^{j_2} \eta_{i;(j_1,j_2)}^L(\boldsymbol{\phi}, \boldsymbol{\phi}^0)| = O_{I\!\!P}(\max(n^\alpha, \sup_{0 \leqslant j_1 < j_2 \leqslant n} \|\boldsymbol{\lambda}_{n;(j_1,j_2)}\|_2)).$$

Proof. of Lemma 12

By the Lemma 3 of Bai,[8] Lemma holds for $G_i^{(1/2)}$ instead of $\eta_{i;(j_1,j_2)}^L$. For η_i^L, we have $|\eta_{i;(j_1,j_2)}^L(\boldsymbol{\phi}; \boldsymbol{\phi}^0) - G_i^{(1/2)}(\boldsymbol{\phi}; \boldsymbol{\phi}^0)| = (j_1 - j_1)^{-1/2} |\boldsymbol{\lambda}_{n;(j_1,j_2)}^t(|\boldsymbol{\phi}| - |\boldsymbol{\phi}^0|)|$. Then, by triangular inequality together the Lemma 7 and the compactness of the set Γ, we obtain

$$\sup_{0 \leqslant j_1 < j_2 \leqslant n} |\inf_{\boldsymbol{\phi}} \sum_{i=j_1+1}^{j_2} \eta_{i;(j_1,j_2)}^L(\boldsymbol{\phi}; \boldsymbol{\phi}^0)| \leqslant \sup_{0 \leqslant j_1 < j_2 \leqslant n} |\inf_{\boldsymbol{\phi}} \sum_{i=j_1+1}^{j_2} G_i^{(1/2)}(\boldsymbol{\phi}; \boldsymbol{\phi}^0)| + C \sup_{0 \leqslant j_1 < j_2 \leqslant n} \|\boldsymbol{\lambda}_{n;(j_1,j_2)}\|_1$$

$$= O_{I\!\!P}(n^\alpha) + O_{I\!\!P}(\boldsymbol{\lambda}_n). \qquad \square$$

Lemma 13. *Under the assumptions (A1), (A2), if $n^{-1/2}\|\boldsymbol{\lambda}_n\|_2 \xrightarrow[n\to\infty]{I\!\!P} \lambda_0$, with $\lambda_0 \geqslant 0$, then*

$$\liminf_{n\to\infty} \left(\inf_{\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|_2 \geqslant n^{-1/2}} n^{-1} \sum_{i=1}^{n} \eta_{i;(0,n)}^L(\boldsymbol{\phi}; \boldsymbol{\phi}^0) \right) > \varepsilon.$$

Proof. of Lemma 13

Using (A2), by the Lemma 6 of Ciuperca,[5] we have for $G_i^{(1/2)}$, with the probability 1

$$\liminf_{n\to\infty}(\inf_{\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|_2\geqslant n^{-1/2}} n^{-1}\sum_{i=1}^n G_i^{(1/2)}(\boldsymbol{\phi};\boldsymbol{\phi}^0)) > \frac{3\varepsilon}{2}. \tag{25}$$

We also have the inequality

$$\inf_{\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|_2\geqslant n^{-1/2}}\sum_{i=1}^n \eta_{i;(0,n)}^L(\boldsymbol{\phi};\boldsymbol{\phi}^0) \geqslant \inf_{\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|_2\geqslant n^{-1/2}}\sum_{i=1}^n G_i^{(1/2)}(\boldsymbol{\phi};\boldsymbol{\phi}^0) - \sup_{\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|_2\geqslant n^{-1/2}} \boldsymbol{\lambda}_n^t(|\boldsymbol{\phi}|-|\boldsymbol{\phi}^0|). \tag{26}$$

Since $\|\boldsymbol{\lambda}_n\|_2 = O_{I\!\!P}(n^{-1/2})$ and $\boldsymbol{\phi}$ belongs to a compact set, then the last term of the right-hand-side of (26) is $o_{I\!\!P}(n^{-1})$. Hence

$$\sup_{\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|_2\geqslant n^{-1/2}}(n^{-1}\boldsymbol{\lambda}_n^t(|\boldsymbol{\phi}|-|\boldsymbol{\phi}^0|)) < \frac{\varepsilon}{2}, \quad n\geqslant n_\varepsilon. \tag{27}$$

The conclusion follows, combining the relations (25), (26) and (27). $\qquad\square$

Lemma 14. *For all $n_1,n_2 \in \mathbb{N}$ such that $n_1 \geqslant n^u$, with $3/4 \leqslant u \leqslant 1$, $n_2 \leqslant n^v$, $v < 1/4$, let us consider the model*

$$\begin{aligned} Y_i &= \mathbf{X}_i^t\boldsymbol{\phi}_1^0 + \varepsilon_i, & i &= 1,\cdots,n_1 \\ Y_i &= \mathbf{X}_i^t\boldsymbol{\phi}_2^0 + \varepsilon_i, & i &= n_1+1,\cdots,n_1+n_2 \end{aligned}$$

with the assumption $\boldsymbol{\phi}_1^0 \neq \boldsymbol{\phi}_2^0$. Under the assumptions (A1)-(A3) and (λ_n) as in the Lemma 13, let us consider $A_{n_1+n_2}^L(\boldsymbol{\phi}) = \sum_{i=1}^{n_1}\eta_{i;(0,n_1)}^L(\boldsymbol{\phi};\boldsymbol{\phi}_1^0) + \sum_{i=n_1+1}^{n_1+n_2}\eta_{i;(n_1,n_1+n_2)}^L(\boldsymbol{\phi};\boldsymbol{\phi}_2^0)$ and $\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L \equiv \arg\min_{\boldsymbol{\phi}} A_{n_1+n_2}^L(\boldsymbol{\phi})$. Then
(i) $\|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L - \boldsymbol{\phi}_1^0\|_2 \leqslant n_1^{-1/2}n_1^{\frac{v+\delta}{2u}} \leqslant n^{-(u-v-\delta)/2}$.
(ii) $\sum_{i=1}^{n_1}\eta_{i;(0,n_1)}^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L,\boldsymbol{\phi}_1^0) = O_{I\!\!P}(1)$.
Proof. of Lemma 14
We denote $\widehat{\boldsymbol{\phi}}_{n_1+n_2} \equiv \arg\min_{\boldsymbol{\phi}}\sum_{i=1}^{n_1}G_i^{(1/2)}(\boldsymbol{\phi},\boldsymbol{\phi}_1^0) + \sum_{i=n_1+1}^{n_1+n_2}G_i^{(1/2)}(\boldsymbol{\phi},\boldsymbol{\phi}_2^0)$. Using the assumptions (A1) and (A3), by Lemma 10 of (Ref. 8) we have that (i) and (ii) are true for $\widehat{\boldsymbol{\phi}}_{n_1+n_2}$ and $G_i^{(1/2)}$.

(i) We suppose the contrary $\|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L - \boldsymbol{\phi}_1^0\|_2 \geqslant n_1$. On the other hand, we have by definition:

$$A_{n_1+n_2}^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L) \leqslant \sum_{i=n_1+1}^{n_1+n_2}\eta_{i;(n_1,n_1+n_2)}^L(\boldsymbol{\phi}_1^0,\boldsymbol{\phi}_2^0) = \sum_{i=n_1+1}^{n_1+n_2}G_i^{(1/2)}(\boldsymbol{\phi}_1^0,\boldsymbol{\phi}_2^0) + \boldsymbol{\lambda}_n^t(|\boldsymbol{\phi}_1^0|-|\boldsymbol{\phi}^0|). \tag{28}$$

By Lemma 10(ii) of Bai,[8] we have: $\sum_{i=n_1+1}^{n_1+n_2}G_i^{(1/2)}(\boldsymbol{\phi}_1^0,\boldsymbol{\phi}_2^0) = o_{I\!\!P}(1)$, then taking into account the relation (28), we obtain

$$A_{n_1+n_2}^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L) \leqslant o_{I\!\!P}(1) + O_{I\!\!P}(n^{1/2}). \tag{29}$$

On the other hand, using the Lemma 13, we deduce

$$\sum_{i=1}^{n_1}\eta_{i;(0,n_1)}^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L,\boldsymbol{\phi}_1^0) \geqslant O_{I\!\!P}(n_1). \tag{30}$$

There is a contradiction between the relations (29) and (30).
(ii) Introduce $v_{(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2)}(\mathbf{X}_i) \equiv \mathbf{X}_i^t(\boldsymbol{\phi}_1-\boldsymbol{\phi}_2)$. For $\tau = 1/2$, let us recall the notations given in Lemma 11:

$Z_n^{(1/2)}(\boldsymbol{\phi}) \equiv \sum_{i=1}^{n_1} G_i^{(1/2)}(\boldsymbol{\phi}, \boldsymbol{\phi}_1^0)$, $t_n^{(1/2)}(\boldsymbol{\phi}) \equiv \sum_{i=n_1+1}^{n_1+n_2}[|\varepsilon_i - \nu_{(\boldsymbol{\phi}, \boldsymbol{\phi}_2^0)}(\mathbf{X}_i)| - |\varepsilon_i - \nu_{(\boldsymbol{\phi}_1^0, \boldsymbol{\phi}_2^0)}(\mathbf{X}_i)|]$. By the Lemma 7 of Ciuperca,[9] we have that: $Z_n^{(1/2)}(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L, \boldsymbol{\phi}_1^0) = O_{I\!\!P}(1)$. Introduce now $t_n^L(\boldsymbol{\phi}) \equiv t_n^{(1/2)}(\boldsymbol{\phi}) + \boldsymbol{\lambda}_{(n_1, n_1+n_2)}^t[|\boldsymbol{\phi}| - |\boldsymbol{\phi}_1^0|]$, $Z_n^L(\boldsymbol{\phi}) \equiv Z_n^{(1/2)}(\boldsymbol{\phi}) + \boldsymbol{\lambda}_{(0, n_1)}^t[|\boldsymbol{\phi}| - |\boldsymbol{\phi}_1^0|]$. Thus $A_{n_1+n_2}^L(\boldsymbol{\phi}) = Z_n^{(1/2)}(\boldsymbol{\phi}) + \boldsymbol{\lambda}_{(0, n_1)}^t[|\boldsymbol{\phi}| - |\boldsymbol{\phi}_1^0|] + t_n^{(1/2)}(\boldsymbol{\phi}) + \sum_{i=n_1+1}^{n_1+n_2}[|\varepsilon_i - \nu_{(\boldsymbol{\phi}_1^0, \boldsymbol{\phi}_2^0)}(\mathbf{X}_i)| - |\varepsilon_i|] + \boldsymbol{\lambda}_{(n_1, n_1+n_2)}^t[|\boldsymbol{\phi}| - |\boldsymbol{\phi}_2^0|] = Z_n^L(\boldsymbol{\phi}) + t_n^L(\boldsymbol{\phi}) + \boldsymbol{\lambda}_{(n_1, n_1+n_2)}^t[|\boldsymbol{\phi}_1^0| - |\boldsymbol{\phi}_2^0|] - \sum_{i=n_1+1}^{n_1+n_2}[|\varepsilon_i - \nu_{(\boldsymbol{\phi}_1^0, \boldsymbol{\phi}_2^0)}(\mathbf{X}_i)| - |\varepsilon_i|]$. Then $\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L \equiv \arg\min_{\boldsymbol{\phi}} A_{n_1+n_2}^L(\boldsymbol{\phi}) = \arg\min_{\boldsymbol{\phi}}[Z_n^L(\boldsymbol{\phi}) + t_n^L(\boldsymbol{\phi})]$. But $|t_n^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L)| \leqslant |t_n^{(1/2)}(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L)| + |\boldsymbol{\lambda}_{(n_1, n_1+n_2)}^t[|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L| - |\boldsymbol{\phi}_1^0|]|$ and using the elementary inequality $||a| - |b|| \leqslant |a - b|$, we have $|t_n^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L)| \leqslant \sum_{i=n_1+1}^{n_1+n_2} \|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L - \boldsymbol{\phi}_1^0\|_1 \cdot \|\mathbf{X}_i^t\|_1 + |\boldsymbol{\lambda}_{(n_1, n_1+n_2)}^t[|\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L| - |\boldsymbol{\phi}_1^0|]| \leqslant o_{I\!\!P}(1)$, we have used (i) and the assumptions (A3) and $\|\boldsymbol{\lambda}_{(n_1, n_1+n_2)}\|_2 = o_{I\!\!P}(n_2^{1/2})$.

We have also $Z_n^L(\boldsymbol{\phi}_1^0) = t_n^L(\boldsymbol{\phi}_1^0)$. Thus $0 \geqslant \inf_{\boldsymbol{\phi}}(Z_n^L(\boldsymbol{\phi}) + t_n^L(\boldsymbol{\phi})) = Z_n^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L) + t_n^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L) = Z_n^L(\widehat{\boldsymbol{\phi}}_{n_1+n_2}^L) - |o_{I\!\!P}(1)| \geqslant \inf_{\boldsymbol{\phi}} Z_n^L(\boldsymbol{\phi}) - |o_{I\!\!P}(1)|$. But $\inf_{\boldsymbol{\phi}} Z_n(\boldsymbol{\phi}) = O_{I\!\!P}(1)$. The rest of proof is similar to that of the Lemma 3(ii) of Ciuperca[12]. $\square$

Acknowledgments

1. R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society*, Ser. B, **58**, 267-288 (1996).
2. Y. Wu and Y. Liu, "Variable selection in quantile regression," *Statistica Sinica*, **19**, 801-817 (2009).
3. J. Xu and Z. Ying, "Simultaneous estimation and variable selection in median regression using Lasso-type penalty," *Annals of the Institute of Statistical Mathematics*, **62**, 487-514 (2010).
4. L. Wang, "L_1 penalized LAD estimator for high dimensional linear regression," *Journal of Multivariate Analysis*, DOI 10.1016/j.jmva.2013.04.001 (2013).
5. G. Ciuperca, "Model selection by LASSO methods in a change-point model," *Statistical Papers*, DOI 10.1007/s00362-012-0482-x, (2013).
6. R. Koenker, *Quantile Regression*, Cambridge University Press (2005).
7. G.J. Babu, "Strong representations for LAD estimators in linear models," *Probability Theory and Related Fields*, **83**, 547-558 (1989).
8. J. Bai, "Estimation of multiple-regime regressions with least absolute deviation," *Journal of Statistical Planning Inference*, **74**, 103-134 (1998).
9. G. Ciuperca, "Estimating nonlinear regression with and without change-points by the LAD-method," *Annals of the Institute of Statistical Mathematics*, **63**(4), 717-743 (2011).
10. J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, **96**(456), 1348-1360 (2001).
11. H. Zou, "The adaptive Lasso and its oracle properties," *Journal of the American Statistical Association*, **101**(476), 1418-1428 (2006).
12. G. Ciuperca, "Penalized least absolute deviations estimation for nonlinear model with change-points," *Statistical Papers*, **52**(2), 371-390 (2011).