Adaptive LASSO model selection in a multiphase quantile regression

Gabriela Ciuperca **

Université de Lyon, Université Lyon 1, CNRS, UMR 5208, Institut Camille Jordan,

Bat. Braconnier, 43, blvd du 11 novembre 1918, F - 69622 Villeurbanne Cedex, France

October 22, 2014

Abstract

We propose a general adaptive LASSO method for a quantile regression model. Our method is very interesting when we know nothing about the first two moments of the model error. We first prove that the obtained estimators satisfy the oracle properties, which involves the relevant variable selection without using hypothesis test. Next, we study the proposed method when the (multiphase) model changes to unknown observations called change-points. Convergence rates of the change-points and of the regression parameters estimators in each phase are found. The sparsity of the adaptive LASSO quantile estimators of the regression parameters is not affected by the change-points estimation. If the phases number is unknown, a consistent criterion is proposed. Numerical studies by Monte Carlo simulations show the performance of the proposed method, compared to other existing methods in the literature, for models with a single phase or for multiphase models. The adaptive LASSO quantile method performs better than known variable selection methods, as the least squared method with adaptive LASSO penalty, L_1-method with LASSO-type penalty and quantile method with SCAD penalty.

Keywords adaptive LASSO quantile; change-point; oracle properties; variable selection; selection criterion.

AMS Subject Classification: 62J05; 62F12.

1 Introduction

The usual case investigated in literature, for the following regression model,

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

**Corresponding author. Email: Gabriela.Ciuperca@univ-lyon1.fr

is that the errors are assumed to be homoscedastic, i.e., the errors are assumed to be independent random variables with mean zero and bounded variance. In such cases, the regression parameters $\boldsymbol{\phi}$ are estimated by the method of the least squares (LS). We will call the regression in this case, the LS model. If the assumptions on the first two moments of the errors are not satisfied, then the LS method is not appropriate, because it can provide bad estimators (biased, with large variance). In the case of errors with zero mean of sign, i.e. $I\!\!E[sgn(\varepsilon_i)] = 0$, the Least Absolute Deviations (LAD) method could be used. But, often in practice, we can not known if $I\!\!E[sgn(\varepsilon_i)] = 0$, then, a generalization can be used by considering the quantile method. Here, we use the notation $sgn(.)$ for the sign function.

For a fixed quantile index(level) $\tau \in (0,1)$, the τth quantile b_τ of ε is:

$$\tau = I\!\!P[\varepsilon < b_\tau] = F(b_\tau), \qquad b_\tau \in \mathcal{B} \subseteq \mathbb{R}, \tag{2}$$

where F is the distribution function of the error ε and $\mathcal{B}$ is a real set . We use ε to denote a generic member of the sample $(\varepsilon_i)_{1 \leq i \leq n}$. In order to consider a general case for model (1), the τth quantile b_τ of ε is supposed unknown.

For the model (1), Y is the response variable, $\boldsymbol{\phi} = (\phi_{,1}, \phi_{,2}, \cdots, \phi_{,p}) \in \Gamma \subseteq \mathbb{R}^p$ the regression parameters and $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})$ the regressors. Number p of the regressors can be large. More exactly, p does not depend on the observation number n, and $p < n$.

The fact that the τth quantile of ε is unknown is also often the case in a change-point model (model with multiple phases), model defined in Section 3 by the relation (6).

We will first focus on the study of model (1), without assuming classical conditions imposed on errors ε_i. The model parameters $\boldsymbol{\phi}$ will be estimated initially by a method without penalty, taking the quantile-process as objective function. Afterwards, in order to select the variables, we propose to add to the quantile-process an adaptive general penalty of LASSO type. In these cases, we will then refer to the model (1) as quantile regression.

Under the metric $d(x, y) = |\arctan x - \arctan y|$, for $x, y \in \mathbb{R}$, the set $\bar{\mathbb{R}} \equiv \mathbb{R} \cup \{-\infty, \infty\}$ is compact set, then, without loss of generality, we suppose that the sets $\mathcal{B}$ and Γ are compacts. The τth conditional quantile function of Y_i given $\mathbf{X}_i = \mathbf{x}_i$ is $\mathbf{x}_i^t \boldsymbol{\phi}^{\mathbf{0}} + b_\tau^0$, with b_τ^0 the τth quantile of ε. Then, for model (1), the unknown parameters, to estimate, knowing $(Y_i, \mathbf{X}_i)_{1 \leq i \leq n}$ are the τth quantile b_τ and the regression parameters $\boldsymbol{\phi}$. Denote by b_τ^0, $\boldsymbol{\phi}^{\mathbf{0}} = (\phi^0_{,1}, \cdots, \phi^0_{,p})$ their true values, unknown, assumed to be inner points of the sets $\mathcal{B}$, Γ, respectively. The parameters $b_\tau, \boldsymbol{\phi}$ can be estimated by quantile method, by minimizing the quantile-process:

$$(\hat{b}_n, \hat{\boldsymbol{\phi}}_n) \equiv \underset{(b,\boldsymbol{\phi})}{\arg\min} \sum_{i=1}^n \rho_\tau(Y_i - b - \mathbf{X}_i^t \boldsymbol{\phi}), \tag{3}$$

with the function $\rho_\tau(.) : \mathbb{R} \to \mathbb{R}$ defined by $\rho_\tau(r) = r[\tau 1\!\!1_{r>0} - (1-\tau) 1\!\!1_{r \leq 0}]$. We call $(\hat{b}_n, \hat{\boldsymbol{\phi}}_n)$ the quantile estimators of $(b_\tau^0, \boldsymbol{\phi}^{\mathbf{0}})$. The components of $\hat{\boldsymbol{\phi}}_n$ are $(\hat{\phi}_{n,1}, \cdots, \hat{\phi}_{n,p})$. By [1] these estimators are strongly convergent and asymptotically normal:

$$(\hat{b}_n, \hat{\boldsymbol{\phi}}_n) \xrightarrow[n \to \infty]{a.s.} (b_\tau^0, \boldsymbol{\phi}^{\mathbf{0}}),$$

$$\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}^{\mathbf{0}}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \frac{\tau(1-\tau)}{f^2(b_\tau^0)}\boldsymbol{\Upsilon}^{-1}\right),$$

with the matrix $\boldsymbol{\Upsilon}$ defined later in the assumption (A1) and f the density function of ε.

For $\tau = 1/2$ and $b_\tau^0 = 0$ we obtain the median (LAD, L_1) regression, considered for instance by [2]. For a complete review on quantile regression and on quantile estimators, the reader can see the book of [1]. A great advantage of this method is that, compared to the classical estimation methods (least squares or likelihood methods, for example) that are sensitive to outliers, the quantile method provides more robust estimators. Furthermore, the imposed condition to the error distribution are relaxed. However, like all other estimators (for example least squares, maximum likelihood estimators) obtained by optimizing an objective function without penalty, the quantile estimators do not satisfy the oracle properties. Recall that the oracle properties are: the zero components of the true parameters are estimated (shrunk) as 0 with probability tending to 1 (also called sparsity property) and the nonzero parameters have an optimal estimation rate (and is asymptotically normal). Thus, the solution is to consider a penalized objective function. The parameters for which we want to have the selection consistency property are included in the penalty. This type of penalty was introduced by [3] for the least squares estimation framework, with a L_1 penalty (obtaining thereby LASSO estimators). Nevertheless the LASSO estimator does not always satisfy the oracle properties. To remove this inconvenience, an adaptive LASSO estimator was proposed by [4] for the LS regression case. Various procedures, based on LASSO framework, have been proposed and studied during the last few years, in order to simultaneously estimate the parameters and to select significant regressors. The considered models are either with fixed dimension p for $\boldsymbol{\phi}$ or with p depending on the sample size n, with $p > n$. Since the number of papers in these areas is very important in recent years, we mention only some of them. In a median regression, [5] consider the LASSO-type penalty and [6] proposes a L_1 penalty with the possibility that the regressors number is larger than the observation number. Always for a median model, with LASSO penalty, was considered by [7]. In a general quantile regression, [8] propose the SCAD penalty, but which is difficult into practice with regard to numerical algorithms and an adaptive LASSO penalty but under the assumption that the τth quantile b_τ^0 is known and is equal to 0. Alternatively, [9] estimate the quantile-adaptive model-free screening frameworks using a B-spline approximation. In the paper of [10], a composite quantile regression is considered with an adaptive LASSO penalty. The paper of [11] proposes an estimation procedure in a semi-parametric additive partial linear models.

In this paper we first propose, for quantile model (1), a general adaptive LASSO estimator. For this estimator, we will study the oracle properties and other behaviour properties for the estimators and for the adaptive quantile objective function.

Afterwards, we will study for a multiphase model (i.e., a model that changes the shape to unknown observations), if the break estimation affects the oracle properties of the regression parameter estimators. The main theoretical difficulty of this type of model is that the estimation of the change-point locations and of the parameters in each phase can not be performed simultaneously, but sequentially, in the sense that we first estimate regression parameters for fixed change-points

and then, the change-points. Finally, the estimator of the regression parameters is taken as that corresponding to the optimal change-points. This imply that the theoretical study of obtained estimators is very difficult, even assuming that number of phases is known.
If this number is unknown, an additional difficulty to the model study is added. In this case, a consistent estimation criterion for this number is proposed in this paper. Since the LASSO techniques are fairly recent, there are not many papers in the literature that address the breaking problem by this estimation method. In the paper of [12], LS model is estimated by LASSO-type and by adaptive LASSO techniques.

In [13], a quantile model with SCAD estimator and a median model with LASSO-type estimator are studied. Apart from the fact that when the quantile model was considered, it was supposed that the τth error quantile is known (more precisely, it was taken as 0), the SCAD method has also the disadvantage that it is difficult to put into practice with regard to numerical algorithms. Let us also remind the paper of [14], where a LS model with a single change-point is considered, with a LASSO penalty, when the errors have gaussian distribution.
It is important to emphasize that the numerical studies by Monte Carlo simulations confirm the superiority of the adaptive LASSO quantile estimators (in terms of bias, precision, of identification of the true zeros, especially in the case when moments of errors don't exist or when its median is different to zero), in comparison to other variable selection estimators. In a multiphase model, if the changes are due only to error distribution (regression parameters remain the same) the adaptive quantile method also gives the best results.

We give some general notations. 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 $|\mathbf{v}| = (|v_1|, \cdots, |v_p|)$. On the other hand, $\|\mathbf{v}\|_2$ is the Euclidean norm and $\|\mathbf{v}\|_1 = \sum_{j=1}^p |v_j|$ the L_1-norm. All vectors are column, $\mathbf{v}^t$ denotes the transposed of $\mathbf{v}$ and $\frac{1}{\mathbf{v}} = \left(\frac{1}{v_1}, \cdots, \frac{1}{v_p}\right)$. For a strictly positive constant c, let also denote by $\mathbf{v}^c$ the following vector $\left(\frac{1}{v_1^c}, \cdots, \frac{1}{v_p^c}\right)$. Let $\overset{\mathcal{L}}{\underset{n\rightarrow\infty}{\longrightarrow}}$, $\overset{I\!P}{\underset{n\rightarrow\infty}{\longrightarrow}}$, $\overset{a.s.}{\underset{n\rightarrow\infty}{\longrightarrow}}$ represent convergence in distribution, in probability and almost sure, respectively, as $n \rightarrow \infty$. For a real x, $[x]$ is its integer part.

The paper is organized as follows. In Section 2, we introduce and study especially the oracle properties of a general adaptive LASSO quantile estimator. In Section 3, the corresponding estimator in a change-point model is defined and its asymptotic behaviors (convergence rate and oracle properties) are studied. Section 4 proposes a consistent criterion to determine the breaking number. In Section 5, simulation results illustrate the performance of the proposed estimators. The adaptive LASSO quantile estimator is compared with other variable selection estimators. All proofs are given in Section 6.

2 Adaptive LASSO quantile for regression model

In this section we propose and study a general adaptive LASSO quantile estimator for the model (1), estimator defined by

$$(\hat{b}_n^*, \hat{\boldsymbol{\phi}}_n^*) \equiv \underset{(b,\boldsymbol{\phi})}{\arg\min} \left(\sum_{i=1}^n \rho_\tau(Y_i - b - \mathbf{X}_i^t \boldsymbol{\phi}) + \lambda_n \hat{\boldsymbol{\omega}}_n^t |\boldsymbol{\phi}| \right), \tag{4}$$

where $\hat{\boldsymbol{\omega}}_n = (\hat{\omega}_{n,1}, \cdots, \hat{\omega}_{n,p}) \equiv \frac{1}{|\hat{\boldsymbol{\phi}}_n|^g} = \left(\frac{1}{|\hat{\phi}_{n,1}|^g}, \cdots, \frac{1}{|\hat{\phi}_{n,p}|^g} \right)$, $\hat{\boldsymbol{\phi}}_n$ is the quantile estimator, defined by (3), and $g > 0$ is a constant which will be later specified. The components of $\hat{\boldsymbol{\phi}}_n^*$ are $(\hat{\phi}_{n,1}^*, \cdots, \hat{\phi}_{n,p}^*)$. The positive sequence $(\lambda_n)_n$, also called the tuning parameter, is a regularization parameter such that $\lambda_n \to \infty$ as $n \to \infty$. To the author's knowledge, for a quantile regression, three particular penalties of the adaptive LASSO type have been previously proposed in other papers. In two papers, the case $g = 2$ is considered. First, [10] consider that p, the number of regression parameters, is fixed and after, in [15], the case $p = p_n \to \infty$ is studied. In the paper of [8], an adaptive LASSO penalty with the same form as in the relation (4) is proposed, but under the assumption that the τth quantile b_τ^0 is known and it is equal to 0. In the paper of [16], which is a particular case to that proposed here, always for $b_\tau^0 = 0$, the weight vector considered is $\min(\sqrt{n}, |\tilde{\boldsymbol{\phi}}_n|^{-1})$, with $\tilde{\boldsymbol{\phi}}_n$ a consistent estimator of $\boldsymbol{\phi}^{\mathbf{0}}$. However, their estimator has the advantage that it also applies when the number of regressors is very large, $p = O(\exp(n^c))$, with the constant $c \in (0,1)$.

Let us consider the deterministic design matrix $\mathbb{X} \equiv (X_{ij})_{\substack{1 \le i \le n \\ 1 \le j \le p}}$, with $\mathbf{X}_i$ the ith line, corresponding to the observation i.

We give now two assumptions, denoted (A1), (A2), on the design and on the errors.

For the design $\mathbb{X}$, we suppose that:

(A1) $n^{-1} \max_{1 \le i \le n} \mathbf{X}_i^t \mathbf{X}_i \underset{n \to \infty}{\longrightarrow} 0$ and $n^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t \underset{n \to \infty}{\longrightarrow} \boldsymbol{\Upsilon}$, with $\boldsymbol{\Upsilon} = (v_{ij})_{1 \le i,j \le p}$ a positive definite matrix.

For the errors ε_i we suppose that they are independent, identically distributed, with a continuous positive density f in a neighborhood of b_τ^0 and:

(A2) For every $e \in int(\mathcal{B})$, $u_0 \in \mathbb{R}$, $\mathbf{u} \in \mathbb{R}^p$ we have

$$\lim_{n \to \infty} n^{-1} \sum_{i=1}^n \int_0^{u_0 + \mathbf{x}_i^t \mathbf{u}} \sqrt{n} [F(e + v/\sqrt{n}) - F(e)] dv = \frac{1}{2} f(e) (u_0, \mathbf{u}^t) \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \boldsymbol{\Upsilon} \end{bmatrix} (u_0, \mathbf{u}^t)^t.$$

Recall that $F : \mathcal{B} \to [0,1]$ is the distribution function of ε. The p regressors $X_{i1}, \cdots, X_{ip}$ are independent of the errors ε_i, for all $i = 1, \cdots, n$. In fact, the design can be considered either deterministic or random, in which case we will consider the conditional expectation. Obviously, the first variable X_1 can be equal to 1, in which case the model contains an intercept. Remark that the assumption (A1) is standard when LASSO methods are used, see for example [4], [8] and (A2) is classic for a quantile regression, see for example [1].

Note that, the estimators defined by (3) or (4), depend on the value of τ and on observation number n. For simplicity reasons, τ does not appear in the notation. The quantile estimators $(\hat{b}_n, \hat{\boldsymbol{\phi}}_n)$ found by (3), minimize the objective function $\sum_{i=1}^n [\rho_\tau(\varepsilon_i - b - \mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi^0})) - \rho_\tau(\varepsilon_i - b_\tau^0)]$ and the adaptive LASSO quantile estimators $(\hat{b}_n^*, \hat{\boldsymbol{\phi}}_n^*)$ given by (4), minimize the objective function $\sum_{i=1}^n [\rho_\tau(\varepsilon_i - b - \mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi^0})) - \rho_\tau(\varepsilon_i - b_\tau^0)] + \lambda_n \hat{\boldsymbol{\omega}}_n^t [|\boldsymbol{\phi}| - |\boldsymbol{\phi^0}|]$.

To simplify the notations, in what follows we denote b_τ^0 by b^0.

Then, let us first define two stochastic processes, necessary to study the objective functions and to analyze the behavior of the corresponding estimators obtained by quantile and by adaptive LASSO quantile methods, respectively. For $b \in \mathcal{B}$, $\boldsymbol{\phi} \in \Gamma$ we define, for each observation $i \in \{1, \cdots, n\}$, the following random process:

$$R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi^0}) \equiv \rho_\tau(\varepsilon_i - b - \mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi^0})) - \rho_\tau(\varepsilon_i - b^0) \tag{5}$$

and for two observations l and k between 0 and n, with $l < k$, we define also the corresponding process, taking an adaptive LASSO penalty:

$$R_{i;(l,k)}^{(\tau,\lambda)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi^0}) \equiv R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi^0}) + \frac{\lambda_{(l,k)}}{k-l} \hat{\boldsymbol{\omega}}_{(l,k)}^t (|\boldsymbol{\phi}| - |\boldsymbol{\phi^0}|), \qquad i = l+1, \cdots, k,$$

with $\lambda_{(l;k)}$ the tuning parameter, dependent of position of l and k. The weight vector $\hat{\boldsymbol{\omega}}_{(l;k)} = |\hat{\boldsymbol{\phi}}_{(l;k)}|^{-g}$ is obtained considering $\hat{\boldsymbol{\phi}}_{(l;k)}$ the quantile estimator of $\boldsymbol{\phi}$ calculated by (3) on the samples $l+1, l+2, \cdots, k$:

$$(\hat{b}_{(l;k)}, \hat{\boldsymbol{\phi}}_{(l;k)}) = \underset{(b,\boldsymbol{\phi})}{\arg\min} \sum_{i=l+1}^{k} \rho_\tau(Y_i - b - \mathbf{X}_i^t \boldsymbol{\phi}).$$

For $l = 0$ and $k = n$, the tuning parameter $\lambda_{(0,n)}$ becomes λ_n and $\hat{\boldsymbol{\omega}}_{(0,n)}$, $(\hat{b}_{(0,n)}, \hat{\boldsymbol{\phi}}_{(0,n)})$, $(\hat{b}_{(0,n)}^*, \hat{\boldsymbol{\phi}}_{(0,n)}^*)$ become $\hat{\boldsymbol{\omega}}_n$, $(\hat{b}_n, \hat{\boldsymbol{\phi}}_n)$, $(\hat{b}_n^*, \hat{\boldsymbol{\phi}}_n^*)$ respectively.

We note that the estimators obtained by (3) are in fact the parameters that minimize in $(b, \boldsymbol{\phi})$ the random process $\sum_{i=1}^n R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi^0})$ and the penalized estimators obtained by relation (4) are the ones that minimize $\sum_{i=1}^n R_{i;(0,n)}^{(\tau,\lambda)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi^0})$.

In the following Lemma we prove that, if in a model, either the τth quantile or the regression parameters are different to the true values, then the observation number for which $|b - b^0| + |\mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi^0})|$ exceeds a threshold, is large.

Lemma 2.1 *Under assumption (A1), if $b \neq b^0$ or $\boldsymbol{\phi} \neq \boldsymbol{\phi^0}$, then there exists $\delta > 0$ such that for the following set $N_n \equiv Card\{i \in \{1, \cdots, n\}; |b - b^0| + |\mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi^0})| > \delta\}$ we have that there exists an $\epsilon_0 > 0$ such that $N_n > n\epsilon_0$, for n large enough.*

In the following subsection we prove that, the adaptive LASSO estimator for the regression parameters satisfies the oracle properties: nonzero estimators are asymptotically normal and zero parameters are shrunk directly to 0 with a probability converging to 1 as n tends to infinity.

2.1 Oracle properties

First of all, let us formulate the Karush-Kuhn-Tucker (KKT) optimality conditions, needed to prove the oracle properties. Note that for a real x, we use the notation $sgn(x)$ for the sign function $sgn(x) \equiv x/|x|$ when $x \neq 0$ and $sgn(0) = 0$. Let the set index

$$\hat{\mathcal{A}}_n^* \equiv \{j \in \{1, \cdots, p\}; \hat{\phi}_{n,j}^* \neq 0\}$$

of nonzero components of the adaptive (LASSO) quantile estimator of the regression parameter.

Proposition 2.2 *Under assumptions (A1), (A2), for $g > 0$, if (λ_n) is a sequence such that $\lambda_n \to \infty$, $n^{-1/2}\lambda_n \to 0$ and $n^{(g-1)/2}\lambda_n \to \infty$, as $n \to \infty$, then:*
(i) for all $j \in \hat{\mathcal{A}}_n^*$ we have: $\tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} 1\!\!1_{Y_i < \mathbf{X}_i^t \hat{\boldsymbol{\phi}}_n^*} = n\hat{\omega}_{n,j} sgn(\hat{\phi}_{,j}^*)$.
(ii) for all $j \notin \hat{\mathcal{A}}_n^*$ we have $\left|\tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} 1\!\!1_{Y_i < \mathbf{X}_i^t \hat{\boldsymbol{\phi}}_n^*}\right| \leq n\hat{\omega}_{n,j}$.

In addition to the set $\hat{\mathcal{A}}_n^*$, let us consider the set

$$\mathcal{A}^0 \equiv \{j \in \{1, \cdots, p\}; \phi_{,j}^0 \neq 0\}$$

with the index of nonzero components of the true regression parameters. Throughout the paper, we denote by $\boldsymbol{\phi}_{\mathcal{A}^0}$ the sub-vector of $\boldsymbol{\phi}$ containing the corresponding components of $\mathcal{A}^0$. In the same way, we consider $\hat{\boldsymbol{\phi}}^*_{\hat{\mathcal{A}}_n^*}$, $\hat{\boldsymbol{\phi}}^*_{\mathcal{A}^0}$ subvectors of $\hat{\boldsymbol{\phi}}_n^*$ with the index in the sets $\hat{\mathcal{A}}_n^*$, $\mathcal{A}^0$, respectively.
With these notations and with Proposition 2.2, we can now state the main result of this section on the oracle properties of the adaptive LASSO quantile estimator. We comment that even if the τth quantile is unknown, the assumptions are the same as in the known case of [8] where b^0 was considered zero. In the paper of [16] the particular case $\hat{\omega}_{n,j} = \min(\sqrt{n}, |\tilde{\phi}_{n,j}|^{-1})$, with $\tilde{\phi}_{n,j}$ a consistent estimator of $\phi_{,j}^0$, is considered. In the present paper we consider $\hat{\omega}_{n,j} = |\hat{\phi}_{n,j}|^{-g}$, with $\hat{\phi}_{n,j}$ the quantile estimator.

Theorem 2.3 *Under assumptions (A1), (A2), for a power $g > 0$, if (λ_n) is a sequence as in the Proposition 2.2, then the estimator $\hat{\boldsymbol{\phi}}_n^*$ satisfies the oracle properties:*
(i) $\sqrt{n}(\hat{\boldsymbol{\phi}}^*_{\mathcal{A}^0} - \boldsymbol{\phi^0}_{\mathcal{A}^0}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \frac{\tau(1-\tau)}{f^2(b^0)} \boldsymbol{\Upsilon}_{\mathcal{A}^0}^{-1}\right)$, where $\boldsymbol{\Upsilon}_{\mathcal{A}^0}$ contains the elements of the matrix $\boldsymbol{\Upsilon}$, defined in assumption (A1), with the index in the set $\mathcal{A}^0$: $\boldsymbol{\Upsilon}_{\mathcal{A}^0} \equiv (v_{ij})_{i,j \in \mathcal{A}^0}$.
(ii) If moreover $n^{g/2-1}\lambda_n \to \infty$, then $I\!\!P[\hat{\mathcal{A}}_n^* = \mathcal{A}^0] \to 1$, as $n \to \infty$.

As a consequence of this Theorem, emphasize the fact that the convergence rate of the adaptive LASSO quantile estimators $\hat{b}^*_n$ and $\hat{\boldsymbol{\phi}}^*_n$ has the order $n^{-1/2}$.

The condition $n^{g/2-1}\lambda_n \rightarrow \infty$ on Theorem 2.3(ii) is stronger than that of Proposition 2.2, where $n^{(g-1)/2}\lambda_n \rightarrow \infty$, as $n \rightarrow \infty$, is supposed. The importance of this condition for sparsity property will be studied in Section 5, by Monte Carlo simulations. Note that, the condition $n^{g/2-1}\lambda_n \rightarrow \infty$ on Theorem 2.3(ii) implies $g > 1$.

Remark 2.4 *In the particular case $b^0_\tau = 0$, the Normal asymptotic law of Theorem 2.3(i) is the same that obtained by [8] by the SCAD estimation method. If, moreover, $\tau = 1/2$, we obtain the same law that [5] using a LASSO-type penalty.*

Remark 2.5 *Compared with the adaptive LASSO estimator for a LS model (see [4]), the fact that the conditions on ε are weakened (no assumption on existence of the first two order moments and nor $I\!E[\varepsilon] = 0$), implies that in our case, in order to have the sparsity, the sequence (λ_n) and the constant g must verify the additional condition $n^{g/2-1}\lambda_n \rightarrow \infty$ beside to claim (i).*

2.2 Behavior study for $R_i^{(\tau)}$ and $R_i^{(\tau,\lambda)}$

The obtained results in this subsection on the random processes $R_i^{(\tau)}$ and $R_i^{(\tau,\lambda)}$ which intervene in the corresponding objective functions, will be necessary to study the behavior of the multiphase model. By the following Proposition, we show that the process $R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})$ has a positive expected value for any parameter $(b, \boldsymbol{\phi}) \in \mathcal{B} \times \Gamma$.

Proposition 2.6 *For all parameters $b \in \mathcal{B}$, $\boldsymbol{\phi} \in \Gamma$, we have that $I\!E[R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})] \geq 0$.*

The penalty order, which is involved in the expression of $R^{(\tau,\lambda)}_{i;(l;k)}$, is obtained by the following result. It depends of the power g considered for the weights $\hat{\boldsymbol{\omega}}_{(l;k)}$.

Lemma 2.7 *Under assumptions (A1), (A2), for all $0 \leq l < k \leq n$, if $\lambda_{(l;k)} = o(n^{1/2})$, then, $\lambda_{(l;k)} \|\hat{\boldsymbol{\omega}}_{(l;k)}\|_2 = o_{I\!P}(n^{(1+g)/2})$.*

In order to study the multiphase model, the following result studies the objective function behavior between two observations sufficiently far apart.

Lemma 2.8 *For $1 \leq l < k \leq n$ such that $k - l \rightarrow \infty$, as $n \rightarrow \infty$, under assumptions (A1) and (A2), if $\lambda_{(l;k)} = o(n^{1/2})$, for all $\alpha > 1/2$, we have $\sup_{0 \leq l < k \leq n} \left| \inf_{b,\boldsymbol{\phi}} \sum_{i=l}^k R^{(\tau,\lambda)}_{i;(l;k)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) \right| = \min(o_{I\!P}(n^{(1+g)/2}), O_P(n^\alpha))$.*

For notational simplicity, we will denote the sum of the processes $R_i^{(\tau)}$ on the observations $1, \cdots, n$ by:

$$\mathcal{R}_n^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) \equiv \sum_{i=1}^n R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}).$$

Let us consider a positive deterministic sequence (c_n), such that either it converges to 0, such that $nc_n^2/\log n \to \infty$ as $n \to \infty$, or it is constant $c_n = c$. For a such sequence (c_n), consider following sub-region of the set $\mathcal{B} \times \Gamma$:

$$\Omega_n \equiv \{(b, \boldsymbol{\phi}) \in \mathcal{B} \times \Gamma; |b - b^0| \leq c_n, \|\boldsymbol{\phi} - \boldsymbol{\phi}^{\mathbf{0}}\|_2 \leq c_n\}$$

and $\Omega_n^c = \{(b, \boldsymbol{\phi}) \in \mathcal{B} \times \Gamma; \min(|b - b^0|, \|\boldsymbol{\phi} - \boldsymbol{\phi}^0\|_2) > c_n\}$ its complementary set.

The following Lemma shows that for all n large enough, the supremum of the difference between $\mathcal{R}_n^{(\tau)}$ and its expectation, when the parameters are in a neighborhood of the true values, converges to 0 in probability, with the rate $(nc_n^2)^{-1}$.

Lemma 2.9 *Under assumptions (A1), (A2), then there exists a strictly positive constant $C_1 > 0$ such that for all $\epsilon > 0$, there exists a $n_\epsilon \in \mathbb{N}$ such that, for $n \geq n_\epsilon$, following inequality holds*

$$I\!P\left[\sup_{(b,\boldsymbol{\phi}) \in \Omega_n} \left| \frac{1}{nc_n^2} \left[\mathcal{R}_n^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!E[\mathcal{R}_n^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})]\right]\right| > \epsilon\right] \leq \exp(-\epsilon^2 nc_n^2 C_1).$$

The following Lemma gives the behavior of the penalized objective function $\sum_{i=1}^n R_{i;(0,n)}^{(\tau,\lambda)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})$ on the complementary of the set Ω_n. We obtain that the infimum of this process is strictly positive.

Lemma 2.10 *Let us consider a positive deterministic sequence (c_n) such that either it converges to 0, with $nc_n^2/\log n \to \infty$ as $n \to \infty$, or $c_n = c$. If the tuning parameter sequence verifies $\lambda_n n^{-1} c_n^{-2} \to 0$, under assumptions (A1), (A2), then we have that there exists $\epsilon_1 > 0$ such that, with the probability 1:*

$$\liminf_{n \to \infty} \left(\inf_{(b,\boldsymbol{\phi}) \in \Omega_n^c} \frac{1}{nc_n^2} \sum_{i=1}^n R_{i;(0,n)}^{(\tau,\lambda)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})\right) > \epsilon_1.$$

An example of a sequence c_n and of tuning parameter λ_n is $\lambda_n = n^{2/5}$ and $c_n^2 = n^{-3/5}(\log n)^2$.

Remark 2.11 *Using Lemma 2.10, by similar technique to one used in the paper of [12], for Lemma 3 and 4, we obtain their equivalent. That is, if the data come from two different models, the adaptive LASSO quantile estimator is close to the parameter of the model from where most of the data came.*

With these results, we can now consider a model with several phases. First, the number of phases is considered known, and afterward the number of changes will be assumed unknown. We will prove that the regression parameters of each phase and the change-points l^0_{r-1}, l^0_r are estimated by consistent adaptive LASSO quantile estimators, with rate of convergence, $(l^0_r - l^0_{r-1})^{-1/2}$, $(l^0_r - l^0_{r-1})^{-1}$, respectively. A very interesting result is that oracle properties of the adaptive LASSO quantile estimators for the regression parameters are not affected by the change-point estimation.

3 Adaptive LASSO quantile for multiphase model

Let us now consider a model with $K+1$ phases, i.e. the model changes to the observations $l_1, \cdots, l_K$ with $1 < l_1 < l_2 < \cdots < l_K < n$. Initially, we suppose that the change number K is known. If K is unknown, which is the most frequent case in practice, we shall give in the following section a criterion to estimate the number K.

The model with $K+1$ phases has the form:

$$Y_i = \mathbf{X}_i^t \boldsymbol{\phi}_1 1\!\!1_{1 \leq i < l_1} + \mathbf{X}_i^t \boldsymbol{\phi}_2 1\!\!1_{l_1 \leq i < l_2} + \cdots + \mathbf{X}_i^t \boldsymbol{\phi}_{K+1} 1\!\!1_{l_K \leq i \leq n} + \varepsilon_i, \quad i = 1, \cdots, n. \tag{6}$$

The parameters of model (6) are $(b_1, \cdots, b_{K+1})$ the τth quantile of ε on each phase, $(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{K+1})$ the corresponding regression parameters and $(l_1, \cdots, l_K)$ the change-points (breaks). The true values of these parameters are $(b^0_1, \cdots, b^0_{K+1})$, $(\boldsymbol{\phi}^0_1, \cdots, \boldsymbol{\phi}^0_{K+1})$ and $(l^0_1, \cdots, l^0_K)$ respectively.

Let us notice that with regard to the paper of [13] where 0 was always the τth quantile, here quantiles can change from one phase to another. In fact, in the present paper, from one phase to the other, either the regression parameters change, or the τth quantiles change, or both types of parameters changes simultaneously. To the author knowledge, this double possibility of change has not been addressed anywhere in the literature. Moreover, in the paper of [13], the proposed penalty for a quantile multiphase model, with the τth quantile known, is SCAD (Smoothly Clipped Absolute Deviation), which produces difficulties of point of view numerical programming. In the particular case of a median multiphase model, [13] considers a LASSO-type estimator in order to avoid the numerical disadvantage generated by the SCAD method. On the other hand, in view of simulations presented in Section 5, for median multiphase model, the results obtained by SCAD method are poorer than by a LASSO-type method. This justify the interest to consider an adaptive LASSO method for a quantile multiphase model. Moreover, the simulations presented in Section 5 will show the performance of the proposed method, especially in the case of not homoscedastic error or when the change occurs in the τth quantile and not in the regression parameters.

Concerning the change-points $(l_1, \cdots, l_K)$, we suppose that a phase is long enough:
(A3) $l_{r+1} - l_r \geq n^a$, $a > 1/2$, for all $r = 0, \cdots, K$, with $l_0 = 1$ and $l_{K+1} = n$.

For fixed change-points $(l_1, \cdots, l_K)$, the objective function is minimized with respect to the τth quantiles and regression parameters of the $K+1$ phases. Let us denote the objective function value by

$$S^*(l_1, \cdots, l_K) \equiv \inf_{\substack{(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{K+1})\\ (b_1, \cdots, b_{K+1})}} \sum_{r=1}^{K+1} [\sum_{i=l_{r-1}+1}^{l_r} \rho_\tau(Y_i - b_r - \mathbf{X}_i^t \boldsymbol{\phi}_r) + \lambda_{(l_{r-1};l_r)} \hat{\boldsymbol{\omega}}^t_{(l_{r-1};l_r)} |\boldsymbol{\phi}_r|]. \quad (7)$$

The tuning parameters $\lambda_{(l_{r-1};l_r)}$ vary from a phase to the other one with the interval length $l_r - l_{r-1}$. The weight vector is $\hat{\boldsymbol{\omega}}^t_{(l_{r-1};l_r)} = |\hat{\boldsymbol{\phi}}_{(l_{r-1};l_r)}|^{-g}$, with $\hat{\boldsymbol{\phi}}_{(l_{r-1};l_r)}$ the quantile estimator of $\boldsymbol{\phi}_r$ calculated by quantile method, on the observations $l_{r-1}+1, \cdots, l_r$.

Then the adaptive LASSO quantile estimators for the change-points are the minimizers of the function S^*:

$$(\hat{l}_1^*, \cdots, \hat{l}_K^*) \equiv \underset{(l_1, \cdots, l_K)}{\arg\min}\, S^*(l_1, \cdots, l_K).$$

The adaptive LASSO quantile estimator for the regression parameters is $\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)}$ and for the τth quantile is $\hat{b}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)}$, for each $r = 1, \cdots, K+1$:

$$((\hat{b}^*_{(\hat{l}^*_0;\hat{l}^*_1)}, \hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_0;\hat{l}^*_1)}), \cdots, (\hat{b}^*_{(\hat{l}^*_K;\hat{l}^*_{K+1})}, \hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_K;\hat{l}^*_{K+1})})) =$$
$$\underset{\substack{(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{K+1})\\ (b_1, \cdots, b_{K+1})}}{\arg\min} \sum_{r=1}^{K+1} [\sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} \rho_\tau(Y_i - b_r - \mathbf{X}_i^t \boldsymbol{\phi}_r) + \lambda_{(\hat{l}^*_{r-1};\hat{l}^*_r)} \hat{\boldsymbol{\omega}}^t_{(\hat{l}^*_{r-1};\hat{l}^*_r)} |\boldsymbol{\phi}_r|].$$

In the next result we show that if in a phase we take in the place of the true parameters those of the nearby phase, then, we obtain, with a probability close to 1, different values for the objective function (without penalty).

Lemma 3.1 *Under assumptions (A1), (A2), for $r = 1, \cdots, K$, if l_r is such that $l_r < l_r^0$, $l_r - l_r^0 = O(1)$, we have that there exists two strictly positive constants η, C_1, such that*

$$I\!P\left[\sum_{i=l_r+1}^{l_r^0} [\rho_\tau(\varepsilon_i - b^0_{r+1} - \mathbf{X}_i^t(\boldsymbol{\phi}^0_{r+1} - \boldsymbol{\phi}^0_r)) - \rho_\tau(\varepsilon_i - b^0_r)] \geq \eta(l_r^0 - l_r)\right] \geq 1 - \exp(-C_1(l_r^0 - l_r)).$$

Let us now formulate the equivalent to assumption (A1) in the case of a model with $(K+1)$ phases, when the change-points are sufficiently far apart, i.e. $l_r - l_{r-1}$ converges to infinity as $n \to \infty$:

(A1bis) $(l_r - l_{r-1})^{-1} \max_{l_{r-1} < i \leq l_r} \mathbf{X}_i^t \mathbf{X}_i \to 0$ for $n \to \infty$. We suppose that for each phase 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 $\boldsymbol{\Upsilon}_r$, as $n \to \infty$, with $\boldsymbol{\Upsilon}_r$ a positive definite matrix. Let us denote by $\boldsymbol{\Upsilon}_r^0$ the limiting matrix for the true change-points l_r^0,

$r = 1, \cdots, K$. We also denote by $v^0_{r,kj}$ the (k,j)th component of matrix $\mathbf{\Upsilon}^0_r$.

Following result gives the convergence rate of the adaptive LASSO quantile estimator of the change-points. The proof is quite technical, that is why we will divide it into three parts. You can not directly show that the distance between the estimator and the true value of the corresponding change-point is finished. This is why we first show that this distance is less than $[n^{1/2}]$, afterward less than $[n^{1/4}]$ and at the end that it is bounded.

Theorem 3.2 *Under assumptions (A1bis), (A2), (A3), if the tuning parameter* $(\lambda_{(l_{r-1},l_r)})_{1 \leq r \leq K+1}$ *is a sequence, depending on* n*, converging to zero, such that* $(l_r - l_{r-1})^{1/2}\lambda_{(l_{r-1},l_r)} \to \infty$ *and if* (c_n) *is another 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* $\hat{l}^*_r - l^0_r = O_{I\!P}(1)$ *for each* $r = 1, \cdots, K$*.*

If instead of the parameters $l_1, \cdots, l_K$ we consider the reparametrization $\theta_r = l_r/n \in (0,1)$, with $0 < \theta_1 < \theta_2 < \cdots \theta_K < 1$, the adaptive LASSO quantile estimators are $\hat{\theta}^*_r = \hat{l}^*_r/n$, which have a convergence rate of order n^{-1}. This is the classical rate convergence for the change-point estimators, when the number of regressors is not dependent on n. See for example the paper of [17] for the LAD estimators, or the paper of [18] for penalized LAD estimators, but not with adaptive type or LASSO penalty. For the LS method, without penalty, the reader can see [19].

Concerning the estimators $\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)}$, in view of the results obtained in Section 2 and of Theorem 3.2, we have that its convergence rate is of order $(l^0_r - l^0_{r-1})^{1/2}$.

For $r = 1, \cdots, K+1$ and $j = 1, \cdots, p$, denote by:
$\phi_{r,j}$ the jth component of the true value $\boldsymbol{\phi}^0_r$,
$\hat{\phi}^*_{(l^0_{r-1};l^0_r),j}$ the jth component of the adaptive LASSO estimator $\hat{\boldsymbol{\phi}}^*_{(l^0_{r-1};l^0_r)}$,
$\hat{\phi}^*_{(l^*_{r-1};l^*_r),j}$ the jth component of the adaptive LASSO estimator $\hat{\boldsymbol{\phi}}^*_{(l^*_{r-1};l^*_r)}$.

The following Theorem shows that the oracle properties are preserved in a multiphase model. For each two consecutive true change-points l^0_{r-1}, l^0_r consider the set with the index of nonzero components of the true regression parameters

$$\mathcal{A}^0_r \equiv \{j \in \{1, \cdots, p\}; \phi^0_{r,j} \neq 0\}$$

and following set when only the regression parameters are estimated

$$\hat{\mathcal{A}}^0_{n,r} \equiv \{j \in \{1, \cdots, p\}; \hat{\phi}^*_{(l^0_{r-1},l^0_r),j} \neq 0\}.$$

Consider also the similar index set corresponding to the adaptive LASSO quantile estimators $\hat{l}^*_{r-1}$, $\hat{l}^*_r$ of the change-points:

$$\hat{\mathcal{A}}^*_{n,r} \equiv \{j \in \{1, \cdots, p\}; \hat{\phi}^*_{(\hat{l}^*_{r-1},\hat{l}^*_r),j} \neq 0\}.$$

In the following theorem, the matrix $\mathbf{\Upsilon}^0_{\mathcal{A}_r^0}$ contains the elements of matrix $\mathbf{\Upsilon}_r$, the vectors $\left(\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)} - \boldsymbol{\phi}^0_r\right)_{\mathcal{A}_r^0}$, $\left(\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)} - \boldsymbol{\phi}^0_r\right)_{\mathcal{A}_r^0}$ are subvectors of $\left(\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)} - \boldsymbol{\phi}^0_r\right)$, $\left(\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)} - \boldsymbol{\phi}^0_r\right)$, respectively, all with the index in the set $\mathcal{A}_r^0$.

Theorem 3.3 *Under assumptions (A1bis), (A2), (A3), for a power $g > 0$ the tuning parameter sequence $(\lambda_{(l_{r-1},l_r)})_{1 \leq r \leq K+1}$ on each interval (l_{r-1}, l_r) as in Theorem 3.2 and also $(l_r - l_{r-1})^{(g-1)/2}\lambda_{(l_{r-1},l_r)} \to \infty$, as $n \to \infty$, we have*
(i) $(\hat{l}^*_r - \hat{l}^*_{r-1})^{1/2}\left(\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)} - \boldsymbol{\phi}^0_r\right)_{\mathcal{A}_r^0} = (l^0_r - l^0_{r-1})^{1/2}\left(\hat{\boldsymbol{\phi}}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)} - \boldsymbol{\phi}^0_r\right)_{\mathcal{A}_r^0}(1 + o_{I\!P}(1)) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\Big(\mathbf{0}, (\tau(1-\tau))f^{-2}(b^0_r)(\mathbf{\Upsilon}^0_{\mathcal{A}}$
(ii) If the tuning parameter $\lambda_{(l_{r-1},l_r)}$ satisfies more $(l_r - l_{r-1})^{g/2-1}\lambda_{(l_{r-1},l_r)} \to \infty$, as $n \to \infty$, then, for every $r = 1, \cdots, K$, we have $\lim_{n\to\infty} I\!P\left[\hat{\mathcal{A}}^0_{n,r} = \hat{\mathcal{A}}^*_{n,r} = \mathcal{A}^0_r\right] = 1$.

4 Selection criterion of the number of phases

Let us now give a criterion to estimate the true change-points number, noted K^0, of the model (6). First give some notations and an additional assumption.

Concerning the distribution of error (ε_i), we suppose that
(A4) $0 < I\!E[\rho_\tau(\varepsilon - b^0_r)]$ and $I\!E[\rho^2_\tau(\varepsilon - b^0_r)] < \infty$ for all $r = 1, \cdots, K^0$.

If $b^0_r = 0$, then the first part of the assumption (A4) amounts to imposing that the mean of $|\varepsilon|$ is bounded and strictly positive. In fact, since $|\rho_\tau(\varepsilon - b^0_r) - \rho_\tau(\varepsilon)| \leq |b^0_r|$ with probability 1, and since b^0_r belongs to a compact set, the assumption (A4) implies that $I\!E[\rho_\tau(\varepsilon)] < \infty$.
The condition $I\!E[\rho^2_\tau(\varepsilon - b^0_r)] < \infty$ is necessary to define a consistent criterion in case when the error distribution changes at the observation l^0_r, for $r = 1, \cdots, K^0$.

For any number K of changes, we calculate the sum $S^*(l_1, \cdots, l_K)$, defined by the relation (7), and the corresponding change-point estimators thereby:

$$(\hat{l}^*_{1,K}, \cdots, \hat{l}^*_{K,K}) \equiv \underset{(l_1,\cdots,l_K)}{\arg\min}\, S^*(l_1, \cdots, l_K).$$

Let us consider the objective function divided by the observation number $\hat{s}^*_K \equiv n^{-1}S^*(\hat{l}^*_{1,K}, \cdots, \hat{l}^*_{K,K})$. In order to find the estimator of K, let us consider following criterion

$$B(K) \equiv n \log \hat{s}^*_K + G(K, p_K)B_n, \tag{8}$$

where (B_n) is a deterministic sequence converging to infinity such that $B_n n^{-a} \to 0$, $B_n n^{-1/2} \to \infty$ as $n \to \infty$. The constant a is that of the supposition (A3). The penalty $G(K, p)$ is a function such that $G(K_1,p) \leq G(K_2,p)$, for all $K_1 \leq K_2$ and optionally depending on the number p of parameters to be estimated.

By the proof of Theorem 4.1, we have that $\hat{s}^*_{K^0}$ converges in probability to $\sum_{r=1}^{K^0+1} \mathbb{E}[\rho_\tau(\varepsilon + b^0_r)]$. Then, in order that the proposed criterion is well defined, it is necessary to impose the condition that each $\mathbb{E}[\rho_\tau(\varepsilon + b^0_r)]$ is strictly positive and bounded.

We consider as an estimator for K^0, the number of change-points that minimizes the criterion $B(K)$, so

$$\hat{K}^*_n \equiv \arg\min_K \left(n \log \hat{s}^*_K + G(K,p) B_n \right). \tag{9}$$

This type of criterion for choosing the change-point number was introduced, as a Schwarz criterion, by [20] for a constant model, scalar, in each phase, with $G(K,p) = K$. It was after considered, for an without penalty median model, by [17]. Other information criterion are used to detect the change number in the papers [21], [22], [23]. In the paper of [24], the empirical likelihood test was considered to detect a single change against no-change in a linear regression. [25] consider likelihood ratio type statistic to test the null hypothesis of K changes, against the alternative hypothesis of $(K+1)$ changes under the assumptions that the errors have mean 0 and bounded variance. But this approach type has the disadvantage that it must perform successive test to find the true number K^0 of the change-points. The only case when two changes are tested agains no changes is the particular case of the epidemic charge model. See for example the paper of [26] for the epidemic change in a constant model.

We prove now that $\hat{K}^*_n$ obtained by the relation (9) is a weakly convergent estimator for K^0.

Theorem 4.1 *Under assumptions (A1bis), (A2)-(A4), if the deterministic sequence* (B_n) *converging to infinity is such that* $n^{-a}B_n \to 0$, $n^{-1/2}B_n \to \infty$, *as* $n \to \infty$, *then we have that* $\hat{K}^*_n \xrightarrow[n\to\infty]{\mathbb{P}} K^0$.

Remark 4.2 *In view of the proof of Theorem 4.1, for a fixed index* τ, *when the error quantile does not change from one phase to the other, then for assumption (A4) we require only* $0 < \mathbb{E}[\rho_\tau(\varepsilon - b^0_r)] < \infty$.

5 Simulations

All simulations were performed using the R language. The program codes are available from the author. First we compare the results of the proposed method with other existing in the literature for a model with a single phase. After, the best three methods are studied for a multiphase model.

5.1 Models with a single phase

The samples were generated from the following model with a single phase: $Y_i = \mathbf{X}_i^t \boldsymbol{\phi}^{\mathbf{0}} + \varepsilon_i$, $i = 1, \cdots, n$, with $\boldsymbol{\phi}^{\mathbf{0}} = (1,0,4,0,-3,5,6,0,-1,0)$, $\mathbf{X} = (X_1, \cdots, X_{10})$, $X_3 \sim \mathcal{N}(2,1)$, $X_4 \sim \mathcal{N}(-1,1)$,

$X_5 \sim \mathcal{N}(1,1)$ and $X_j \sim \mathcal{N}(0,1)$ for $j \in \{1,2,6,7,8,9,10\}$. For the errors ε, three distributions were first considered: standard Normal $\mathcal{N}(0,1)$, exponential $\mathcal{E}xp(-4.5,1)$ with the density function $\exp(-(x+4.5))1\!\!1_{x>-4.5}$, and Cauchy $\mathcal{C}(0,1)$. The sample size is $n = 200$.

The percentage of zero coefficients correctly estimated to zero (true 0) and the percentage of nonzero coefficients estimated to zero (false 0) are computed by four methods: least squares model with adaptive LASSO penalty, median model with LASSO-type penalty, quantile model with adaptive LASSO and SCAD penalties.

Recall (see the paper of [4]) that the adaptive LASSO estimators for the regression parameters in a LS model, are the minimizers of the following objective function

$$\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\boldsymbol{\phi})^2 + \lambda_n \hat{\varpi}_n |\boldsymbol{\phi}|,$$

with the adaptive weight p-vector $\hat{\varpi}_n$ considered here that $|\hat{\boldsymbol{\phi}}_n^{LS}|^{-\chi}$. Precise that $\hat{\boldsymbol{\phi}}_n^{LS}$ is the LS estimator of $\boldsymbol{\phi}$. The adaptive LASSO estimator for LS model has the sparsity property if $\lambda_n \to \infty$, $n^{-1/2}\lambda_n \to 0$ and $n^{(\chi-1)/2}\lambda_n \to \infty$, as $n \to \infty$. We consider then the tuning parameter $\lambda_n = n^{2/5}$ and the power $\chi = 9/40$.

The SCAD estimator for a quantile model (see [8]) is the minimizer of the following objective function

$$\sum_{i=1}^{n}[\rho_\tau(Y_i - \mathbf{X}^t\boldsymbol{\phi}) + \sum_{j=1}^{p} p_{\lambda_n}(|\phi_{,j}|)],$$

with penalty $p_{\lambda_n}(|\phi_{,j}|)$ defined by its first derivative

$$p'_{\lambda_n}(|\phi_{,j}|) = \lambda_n\{1\!\!1_{|\phi_{,j}|\leq\lambda_n} + \frac{(a_1\lambda_n - |\phi_{,j}|)_+}{(a-1)\lambda_n}1\!\!1_{|\phi_{,j}|>\lambda_n}\},$$

for all $j = 1, \cdots, p$, with $\lambda_n > 0$, $a_1 > 2$ deterministic tuning parameters. We take here, for the SCAD method, the tuning parameters $a_1 = 5$ and $\lambda_n = 1/|\hat{\boldsymbol{\phi}}_n^{QLASSO}|$. The estimator $\hat{\boldsymbol{\phi}}_n^{QLASSO}$ of $\boldsymbol{\phi}$ is the minimizer of the objective function $\sum_{i=1}^{n}[\rho_\tau(Y_i - \mathbf{X}^t\boldsymbol{\phi}) + \boldsymbol{\lambda}_n|\boldsymbol{\phi}|$, with $\boldsymbol{\lambda}_n = \log n \cdot \mathbf{1}_p$.

The LASSO-type estimator for a median model (see [5]), is the minimizer of the following objective function

$$\sum_{i=1}^{n}|Y_i - \mathbf{X}_i^t\boldsymbol{\phi}| + \boldsymbol{\lambda}_n^t|\boldsymbol{\phi}|,$$

with the tuning parameter $\boldsymbol{\lambda}_n$ a random p-vector. We take here $\boldsymbol{\lambda}_n = n^{2/5}\frac{1}{|\hat{\boldsymbol{\phi}}_n^{QLASSO}|}$.

For adaptive LASSO quantile method, given by relation (4), we consider two values for the power g: a value greater than 1 and another smaller than 1. The tuning parameter is $\lambda_n = n^{2/5}$.

The results obtained by these four methods are presented in Tables 1-3 for 1000 Monte Carlo replications. For some distributions or some quantile index, there is numerical problems (the

Table 1: Model with a single phase. Percentage true 0 and of false 0 by LS+adaptiveLASSO, QUANTILE+adaptiveLASSO, QUANTILE+SCAD, LAD+LASSOtype methods for $n = 200$, $\varepsilon_i \sim \mathcal{N}(0, 1)$.

$\tau\downarrow$	Method $\rightarrow$ parameters $\rightarrow$	LS+aLASSO $\chi = 9/40$	QUANT+aLASSO $g_1 = 12.25/10$	QUANT+aLASSO $g_2 = 9/40$	QUANT+SCAD	LAD+LASSOtype
0.15	% of trues 0	1	1	0.77	0.47	0.99
	% of false 0	0.005	0	0	0	0
0.50	% of trues 0	1	1	0.64	0.55	0.99
	% of false 0	0.01	0	0	0	0
0.95	% of trues 0	1	1	0.93	???	0.99
	% of false 0	0.01	0	0	???	0

Table 2: Model with a single phase. Percentage true 0 and of false 0 by adaptive LASSO least squares, adaptive LASSO quantile, QUANTILE+SCAD, LAD+LASSOtype methods for $n = 200$, $\varepsilon_i \sim \mathcal{E}xp(-4.5, 1)$.

$\tau\downarrow$	Method $\rightarrow$ parameters $\rightarrow$	LS+aLASSO $\chi = 9/40$	QUANT+aLASSO $g_1 = 12.25/10$	QUANT+aLASSO $g_2 = 9/40$	QUANT+SCAD	LAD+LASSOtype
0.15	% of trues 0	0.99	1	0.88	0.28	0.66
	% of false 0	0.01	0	0	0	0
0.50	% of trues 0	1	1	0.71	0.28	0.67
	% of false 0	0.01	0	0	0	0
0.95	% of trues 0	1	0.99	0.90	???	0.67
	% of false 0	0.01	0.02	0.01	???	0

function *rq* of the package *quantreg* of R language does not respond) for the SCAD method. Then, in Tables 1-3, this is symbolised by "???". The numerical problems of the SCAD method have been also identified by [5] who proposed the LASSO-type penalty for median regression.

5.1.1 Sparsity property

For the adaptive LASSO quantile method proposed in this paper, considering three error distributions, we deduce from Tables 1-3 that the sparsity is not satisfied when the power g of the adaptive weight is smaller than 1. This is in concordance with the condition imposed in statement (ii) of the Theorem 2.3: $n^{g/2-1}\lambda_n \to \infty$. In all of the above simulations, we have the following conclusions: for $g > 1$, the performance of the adaptive LASSO quantile estimations are always better than the SCAD, median (LAD) model with LASSO-type penalty and adaptive LASSO (for LS model) estimators, for the heavy-tailed errors.

In view of the obtained results, presented in Tables 1-3, the penalty SCAD and the parameter $g < 1$ for the adaptive LASSO quantile methods are abandoned.

Based on 1000 Monte-Carlo replications, in Figures 1, 2, 3, 4 we represent the graph of the percentages of true 0 and of false 0 obtained by three estimation methods:

- for LS model with adaptive LASSO penalty (dotted line),

Table 3: Model with a single phase. Percentage true 0 and of false 0 by LS+adaptiveLASSO, QUANTILE+adaptiveLASSO, QUANTILE+SCAD, LAD+LASSOtype methods for $n = 200$, $\varepsilon_i \sim C(0, 1)$.

$\tau \downarrow$	Method $\rightarrow$ parameters $\rightarrow$	LS+aLASSO $\chi = 9/40$	QUANT+aLASSO $g_1 = 12.25/10$	QUANT+aLASSO $g_2 = 9/40$	QUANT+SCAD	LAD+LASSOtype
0.15	% of trues 0	0.43	0.93	0.69	???	0.99
	% of false 0	0.08	0.03	0.025	???	0
0.50	% of trues 0	0.44	0.99	0.64	???	0.98
	% of false 0	0.08	0	0	???	0
0.95	% of trues 0	0.43	0.66	0.77	???	0.98
	% of false 0	0.08	0.24	0.28	???	0

- for quantile model with adaptive LASSO penalty (solid line),
- for median model with LASSO-type penalty (long dash line),

each for errors with four possible distributions: Normal, Exponential, Cauchy and $\mathcal{E}xp(-4.5, 1) + \mathcal{C}(0, 2)$.
For Gaussian errors, these three methods give very satisfactory results (see Figure 1). For Exponential errors $\mathcal{E}xp(-4.5, 1)$, the LASSO-type method does not well identify the true zeros (see Figure 2) since this method is build for median regression. For Cauchy errors $\mathcal{C}(0, 1)$ (then moments of errors don't exist) the adaptive LASSO method for LS model provides estimates that don't have the sparsity property (see Figure 3). The superiority of the adaptive LASSO quantile method is considerably higher when errors are a sum of Exponential and Cauchy laws (see Figure 4).

We note that, in Figures 1-4, for the LS model with adaptive LASSO penalty (dotted line), the quantile index τ is not useful, the represented values of the true or false zeros are in fact Monte Carlo replications.

5.1.2 Conclusion

In conclusion, for an uniphase model, adaptive LASSO quantile method provides very satisfactory results of sparsity for each distribution error, tacking as quantile index $\tau \in [0.4; 0.6]$.

For the sparsity property, the power g of the relation (4), should be greater than 1. Condition on g in concordance with the condition imposed in statement (ii) of Theorem 2.3.

The quantile model with SCAD penalty gives bad sparsity results, and moreover, it has numerical problems.

Concerning the sparsity, the adaptive LASSO quantile model, the adaptive LASSO method for LS model and LASSO-type method for median model, give very satisfactory results for Gaussian errors. The adaptive LASSO quantile method stands out to be the best method, in terms of variable selection, when moments of errors don't exist or when median of errors is different to zero.

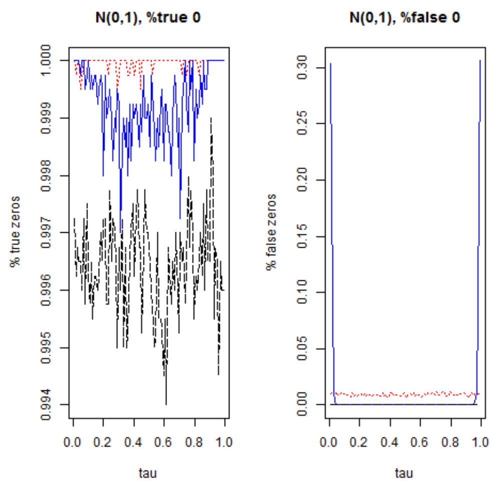

Figure 1: *Percentage of true 0 and of false 0 by LS+aLASSO, QUANTILE+aLASSO, LAD+LASSOtype methods,* $\varepsilon_i \sim \mathcal{N}(0,1)$.

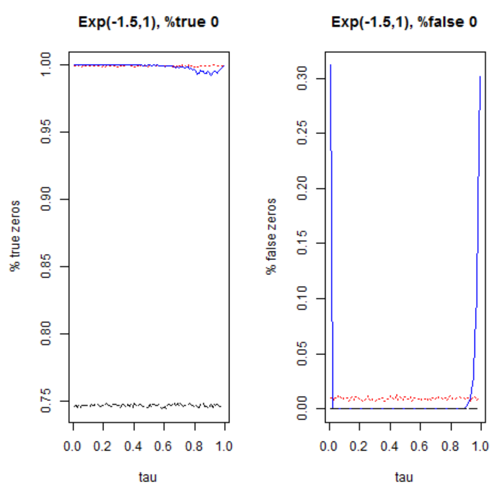

Figure 2: *Percentage of true 0 and of false 0 by LS+aLASSO, QUANTILE+aLASSO, LAD+LASSOtype methods,* $\varepsilon_i \sim \mathcal{E}xp(-4.5,1)$.

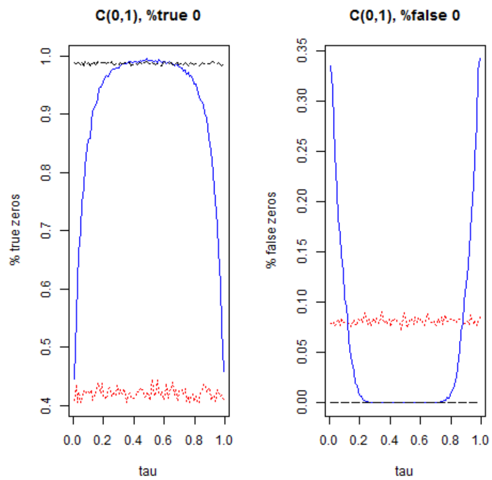

Figure 3: *Percentage of true 0 and of false 0 by LS+aLASSO, QUANTILE+aLASSO, LAD+LASSOtype methods,* $\varepsilon_i \sim \mathcal{C}(0,1)$.

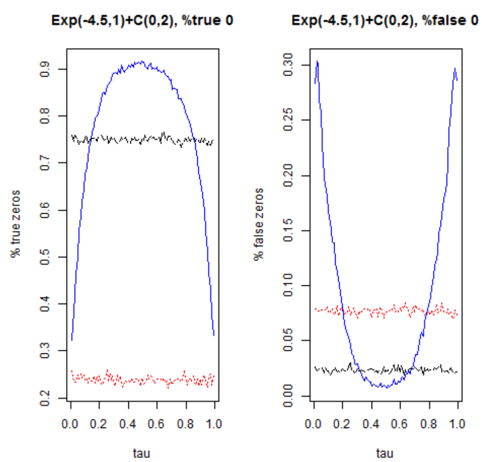

Figure 4: *Percentage of true 0 and of false 0 by LS+aLASSO, QUANTILE+aLASSO, LAD+LASSOtype methods,* $\varepsilon_i \sim \mathcal{E}xp(-4.5,1) + C(0,2)$.

5.2 Multiphase models

In view of the results for uniphase model, we will consider in this subsection only three estimation methods: adaptive LASSO quantile, adaptive LASSO for LS model and LASSO-type for median model. For each of the three methods, we have considered the same powers, tuning parameters as is the previous sub-section, with the only difference that instead of n we take $l_r - l_{r-1}$, for $r = 1, \cdots, K+1$. The quantile index τ for quantile model with adaptive LASSO penalty is considered 0.55.

5.2.1 Fixed phase number

We consider that the change number is known and it is equal to two (three phases). The true change-points are in $l_1^0 = 30$, $l_2^0 = 100$ for $n = 200$ observations. In Tables 4 and 5, we present simulation results for models with regression parameters which differ from one phase to the other: $\boldsymbol{\phi}_1^0 = (1, 0, 4, 0, -3, 5, 6, 0, -1, 0)$, $\boldsymbol{\phi}_2^0 = (0, 3, -4, -3, 0, 1, 2, -3, 0, 10)$, $\boldsymbol{\phi}_3^0 = (1, 3, 4, 0, 0, 1, 0, 0, 0, 1)$. The random vector $\mathbf{X}$ is as in the Subsection 5.1.

In Tables 6 and 7, we present simulation results when the regression parameters for the first two phases are the same ($\boldsymbol{\phi}_1^0 = \boldsymbol{\phi}_2^0$) and only the error distributions are different. In Tables 4 and 6 we present the percentage of true and of false zero, in each interval, for the three methods, when different error distributions are in each phase. In Tables 5 and 7 we give the median of the change-points estimations by the three methods. In order to simply the presentation, in these four tables, it was noted by aQ the adapted LASSO quantile method, by Lt the LASSO-type method for median model and by aLS the adaptive LASSO method for LS model. We calculate bias of the each estimates, the mean of the differences $(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{\mathbf{0}})_{\mathcal{A}^0}$ and the approximation of the estimation variances $1/M \|(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{\mathbf{0}})_{\mathcal{A}^0}\|_2^2$, with M the Monte Carlo replications number, and with $\mathcal{A}^0$ the index set of the nonzero true values. In Tables 4 - 7, the law $\mathcal{E}_1$ is $\mathcal{E}xp(-4.5, 1)$, $\mathcal{E}_2$ is $\mathcal{E}xp(1.5, 1)$, $\mathcal{E}_3$ is $\mathcal{E}xp(-6.5, 1)$. The Cauchy distribution is $\mathcal{C}(0, 1)$ and Gaussian distribution is $\mathcal{N}(0, 1)$.

Comparison of the obtained results by the three estimation methods.

For phases where error distributions are exponential or Cauchy, the adaptive LASSO quantile method gives the best results in terms of true zeros or false zeros percentage, bias and precision of the nonzero parameters.

Note that, in all situations when the regression parameters are different from one phase to the other, by the three methods, the median of the change-points estimations coincides or is very close to the true value.

Comparing the last three rows of the Tables 5 and 7, we conclude that the adaptive LASSO method for LS model gives less accurate change-points estimates when the change-point is between two phases with the same regression parameters and error distributions are two exponential law. Generally, for the LS model with adaptive LASSO penalty, satisfactory results are obtained in a phase with Gaussian errors, while for Exponential or Cauchy distributions, poor results are obtained. Recall that for a model with a single phase, the adaptive LASSO method for LS model gave satisfactory results for Gaussian and Exponential errors, and poor results for Cauchy errors (see Tables 1, 2, 3).

By LASSO-type method, the phases with Exponential errors are poorly estimated.

Comparing the Tables 4 with 6 and Tables 5 with 7, we deduce that selection percentage of the true zeros by adaptive LASSO quantile method decreases slightly for the first phase when $\boldsymbol{\phi}_1^0 = \boldsymbol{\phi}_2^0$. However, this method is better than two other methods, especially with regards to the bias and precision of estimators for the nonzero regression parameters.

Conclusion

To conclude, by the adaptive LASSO quantile and LASSO-type methods, the change-points estimation did not affect the sparsity property of the regression parameter estimates. On the other hand, the adaptive LASSO quantile estimators for the nonzero regression parameters are more accurate (in terms of bias and variance) than corresponding LASSO-type estimators.

5.2.2 Estimation of the change-point number

In Table 8 we give results, after 100 Monte Carlo replications, in order to estimate the number of phases using relation (9). There was one change-point to the observation $l_1^0 = 30$ for a total of 100 observations. The change in the model is due either to the change in the regression parameters ($\boldsymbol{\phi}_1^0 \neq \boldsymbol{\phi}_2^0$) or to the change in quantile of the error (for the same index τ).

We compare the criterion proposed in this paper for the adaptive LASSO quantile method with the criterion proposed in the paper of [12], for the adaptive LASSO method for LS model. In the paper of [13], where the LASSO-type method for a median model with change-points has been studied, there is no criterion proposed to estimate the true number of change-points. Therefore, we propose here a criterion of the same shape as for the adaptive LASSO quantile method: to the corresponding penalized objective function add a term of the form $G(K,p)B_n$ with $G(K,p) = K$ and the sequence $B_n = n^{5/8}$. The criterion values for the three methods are calculated for $K \in \{0, 1, 2, 3\}$. For the adaptive LASSO quantile method we consider $\tau = 0.55$.

From Table 8, we deduce that, the criteria associated to the three methods choose correctly the change-point number when the change is due to the regression parameters. On the other hand, when the change is due only to the quantile, i.e. $\boldsymbol{\phi}_1^0 = \boldsymbol{\phi}_2^0 = (1, 0, 4, 0, -3, 5, 6, 0, -1, 0)$, then, if the errors come from the same distribution (Exponential) but with different quantiles, the criterion for the adaptive LASSO quantile method does not identify whenever the change (only 62/100), the criterion for the LASSO-type method never identifies the change, while that for the adaptive LASSO method for LS model identifies 10/100. The results improve for adaptive LASSO quantile and LASSO-type methods when the errors of the two phases have different distributions (Exponential and Gaussian).

Table 4: Model with three phases. $\phi_1^0 \neq \phi_2^0 \neq \phi_3^0$. Percentage true 0 and of false 0 by adaptive LASSO quantile, LASSO-type for median model and adaptive LASSO for LS model.

error distribution	interval $(1, \hat{l}_1)$						interval $(\hat{l}_1, \hat{l}_2)$						interval $(\hat{l}_2, n)$					
	% true 0			% false 0			% true 0			% false 0			% true 0			% false 0		
	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS
$\varepsilon_1, \varepsilon_2, \varepsilon_3 \sim \mathcal{E}_1$	0.96	0.85	0.82	0.01	0.11	0.11	0.99	0.78	0.92	0	0.003	0.02	1	0.67	1	0	0.004	0.02
$\varepsilon_1, \varepsilon_2, \varepsilon_3 \sim \mathcal{N}$	0.97	0.97	0.88	0.01	0.04	0.12	0.98	0.98	0.96	0	0	0.03	0.99	0.98	1	0	0	0.008
$\varepsilon_1, \varepsilon_2, \sim \mathcal{E}_1, \varepsilon_3 \sim \mathcal{N}$	0.97	0.84	0.79	0.01	0.10	0.11	0.99	0.77	0.92	0	0.003	0.02	0.998	0.99	1	0	0	0.02
$\varepsilon_1, \varepsilon_2, \sim \mathcal{N}, \varepsilon_3 \sim \mathcal{E}_1$	0.97	0.97	0.85	0.02	0.04	0.11	0.99	0.99	0.96	0	0	0.03	1	0.66	1	0	0 .004	0.03
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{N}$	0.97	0.86	0.83	0.02	0.11	0.1	0.99	0.99	0.97	0	0	0.02	1	0.65	1	0	0.008	0.02
$\varepsilon_1, \varepsilon_3, \sim \mathcal{N}, \varepsilon_2 \sim \mathcal{E}_1$	0.96	0.97	0.87	0.02	0.04	0.08	0.995	0.80	0.90	0	0.005	0.02	0.997	0.99	1	0	0	0.04
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{C}$	0.97	0.845	0.72	0.01	0.12	0.12	0.81	0.88	0.22	0.03	0.04	0.06	0.998	0.64	0.96	0	0.007	0.05
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_2, \varepsilon_2 \sim \mathcal{E}_1$	0.97	0.91	0.81	0.01	0.09	0.10	0.99	0.80	0.92	0	0.005	0.002	0.998	0.69	1	0	0	0.03
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_3, \varepsilon_2 \sim \mathcal{E}_1$	0.97	0.78	0.75	0.02	0.16	0.09	1	0.76	0.92	0	0.005	0.02	1	0.55	0.99	0	0.03	0.05
$\varepsilon_1 \sim \mathcal{E}_3, \varepsilon_2, \varepsilon_3 \sim \mathcal{E}_1$	0.97	0.77	0.74	0.08	0.15	0.1	0.997	0.81	0.93	0	0.007	0.02	0.998	0.65	1	0	0.004	0.04

Table 5: Model with three phases. $\phi_1^0 \neq \phi_2^0 \neq \phi_3^0$. Summary statistics by adaptive LASSO quantile, LASSO-type for median model and adaptive LASSO for LS model.

error distribution	median($\hat{l}_1$)			median($\hat{l}_2$)			mean$(\hat{\phi}-\phi^0)_{\mathcal{A}^0}$			mean$\|(\hat{\phi}-\phi^0)_{\mathcal{A}^0}\|$			$1/M\|(\hat{\phi}-\phi^0)_{\mathcal{A}^0}\|_2^2$		
	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS
$\varepsilon_1,\varepsilon_2,\varepsilon_3 \sim \mathcal{E}_1$	30	30	30	100	100	100	-0.03	-0.25	-0.31	0.15	0.49	0.65	0.90	7.5	10.3
$\varepsilon_1,\varepsilon_2,\varepsilon_3 \sim \mathcal{N}$	30	31	31	100	100	100	-0.02	-0.03	-0.31	0.17	0.17	0.65	1	1.12	9.8
$\varepsilon_1,\varepsilon_2,\sim \mathcal{E}_1,\varepsilon_3 \sim \mathcal{N}$	30	30	30	100	100	100	-0.03	-0.17	-0.31	0.16	0.26	0.64	0.94	2.87	9.9
$\varepsilon_1,\varepsilon_3,\sim \mathcal{E}_1,\varepsilon_2 \sim \mathcal{N}$	31	30	31	100	100	100	-0.02	-0.22	-0.32	0.16	0.38	0.64	0.93	5.7	9.8
$\varepsilon_1,\varepsilon_3,\sim \mathcal{N},\varepsilon_2 \sim \mathcal{E}_1$	30	30	30	100	100	100	-0.03	-0.08	-0.29	0.16	0.28	0.64	0.96	3	9.5
$\varepsilon_1,\varepsilon_3,\sim \mathcal{E}_1,\varepsilon_2 \sim \mathcal{C}$	31	31	31	100	100	100	-0.05	-0.25	-0.37	0.26	0.49	4.8	2.7	7.5	41478
$\varepsilon_1,\varepsilon_3,\sim \mathcal{E}_2,\varepsilon_2 \sim \mathcal{E}_1$	30	30	30	100	100	100	-0.02	0.03	-0.28	0.15	0.40	0.64	0.84	5.08	10.01
$\varepsilon_1,\varepsilon_3,\sim \mathcal{E}_3,\varepsilon_2 \sim \mathcal{E}_1$	30	30	31	100	100	100	-0.03	-0.34	-0.34	0.15	0.62	0.68	0.87	13.2	11.9
$\varepsilon_1 \sim \mathcal{E}_3,\varepsilon_2,\varepsilon_3 \sim \mathcal{E}_1$	30	30	31	100	100	100	-0.02	-0.31	-0.34	0.15	0.56	0.68	0.81	10.7	11.6

6 Proofs

For the convenience of the reader, recall first a lemma due to [2].

Lemma 6.1 *([2], Lemma 1)*
Let Z_i be a sequence of independent random variables with mean zero and $|Z_i| \leq \beta$ for some $\beta > 0$. Let also $V \geq \sum_{i=1}^n I\!E[Z_i^2]$. Then for all $0 < s < 1$ and $0 \leq z \leq V/(s\beta)$, we have

$$I\!P\left[\left|\sum_{i=1}^n Z_i\right| > z\right] \leq 2\exp\left(-z^2 s(1-s)/V\right). \tag{10}$$

This section is divided into two subsections. In the first subsection we give the proofs of all Theorems and Propositions. In the second subsection, we present the lemma proofs.

6.1 Proposition and Theorem proofs

Proof of Proposition 2.2 *(i)* For all $j \in \mathcal{A}_n^*$, the estimator $\hat{\phi}^*_{n,j}$ is the solution of the equation: $0 = \sum_{i=1}^n \frac{\partial R_i^{(\tau,\lambda)}(b,\phi;b^0,\phi^0)}{\partial \phi_{,j}}$. By elementary algebra we obtain that

$$\sum_{i=1}^n \frac{\partial R_i^{(\tau,\lambda)}(b,\phi;b^0,\phi^0)}{\partial \phi_{,j}} = -\tau X_{ij} + X_{ij} 1\!\!1_{Y_i < \mathbf{X}_i^t \phi} + \lambda_n \hat{\omega}_{n,j} sgn(\phi_{,j})$$

and the assertion affirmation *(i)* follows.
(ii) In this case, the subgradient set $\partial\|0_j\|_1$ is the closed interval $[-1,1]$. Then

$$0 \in \sum_{i=1}^n \frac{\partial R_i^{(\tau,\lambda)}(b,\phi;b^0,\phi^0)}{\partial \phi_{,j}} = -\tau X_{ij} + X_{ij} 1\!\!1_{Y_i < \mathbf{X}_i^t \phi} + \lambda_n \hat{\omega}_{n,j}[-1,1].$$

Table 6: Model with three phases. $\phi_2^0 = \phi_1^0$, $\phi_2^0 \neq \phi_3^0$. Percentage true 0 and of false 0 by adaptive LASSO quantile, LASSO-type for median model and adaptive LASSO for LS model.

error distribution	interval $(1, \hat{l}_1)$						interval $(\hat{l}_1, \hat{l}_2)$						interval $(\hat{l}_2, n)$					
	% true 0			% false 0			% true 0			% false 0			% true 0			% false 0		
	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{N}$	0.87	0.86	0.84	0.07	0.11	0.12	0.98	0.99	0.98	0.001	0	0.05	1	0.67	0.996	0	0.002	0.04
$\varepsilon_1, \varepsilon_3, \sim \mathcal{N}, \varepsilon_2 \sim \mathcal{E}_1$	0.90	0.95	0.84	0.05	0.03	0.11	0.98	0.78	0.97	0.001	0.01	0.05	0.99	0.98	1	0	0	0.16
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{C}$	0.88	0.91	0.74	0.08	0.12	0.11	0.91	0.94	0.49	0.01	0.01	0.08	1	0.64	0.97	0	0	0.05
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_2, \varepsilon_2 \sim \mathcal{E}_1$	0.81	0.88	0.83	0.1	0.08	0.12	0.99	0.80	0.95	0.002	0.02	0.06	1	0.69	0.99	0	0	0.02
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_3, \varepsilon_2 \sim \mathcal{E}_1$	0.94	0.78	0.82	0.03	0.19	0.12	0.99	0.78	0.96	0	0.01	0.05	0.99	0.56	0.99	0	0.04	0.05
$\varepsilon_1 \sim \mathcal{E}_3, \varepsilon_2, \varepsilon_3 \sim \mathcal{E}_1$	0.92	0.77	0.78	0.04	0.18	0.11	1	0.8	0.95	0	0.01	0.04	1	0.64	1	0	0	0.03

Table 7: Model with three phases. $\boldsymbol{\phi}_2^0 = \boldsymbol{\phi}_1^0$, $\boldsymbol{\phi}_2^0 \neq \boldsymbol{\phi}_3^0$. Summary statistics by adaptive LASSO quantile, LASSO-type for median model and adaptive LASSO for LS model, $n = 200$, $l_1^0 = 30$, $l_2^0 = 100$.

error distribution	median($\hat{l}_1$)			median($\hat{l}_2$)			mean$(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{\mathbf{0}})_{\mathcal{A}^0}$			mean$\lvert(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{\mathbf{0}})_{\mathcal{A}^0}\rvert$			$1/M\|(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{\mathbf{0}})_{\mathcal{A}^0}\|_2^2$		
	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS	aQ	Lt	aLS
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{N}$	32	31	28	100	100	100	-0.04	-0.23	-0.33	0.15	0.43	0.63	0.79	6.1	8.8
$\varepsilon_1, \varepsilon_3, \sim \mathcal{N}, \varepsilon_2 \sim \mathcal{E}_1$	28	30	26	100	100	100	-0.04	-0.12	-0.32	0.22	0.29	0.65	2.3	3.06	9.6
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{C}$	31	31	39	100	100	99	-0.06	-0.22	-0.35	0.28	0.44	1.32	3.2	6.1	316
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_2, \varepsilon_2 \sim \mathcal{E}_1$	32	31	30	100	100	100	-0.07	-0.007	-0.29	0.25	0.42	0.64	3.2	5.2	9.2
$\varepsilon_1, \varepsilon_3, \sim \mathcal{E}_3, \varepsilon_2 \sim \mathcal{E}_1$	31	30	25	101	100	100	-0.04	-0.45	-0.38	0.19	0.71	0.72	1.66	16.9	12.45
$\varepsilon_1 \sim \mathcal{E}_3, \varepsilon_2, \varepsilon_3 \sim \mathcal{E}_1$	34	30	27	100	100	100	-0.05	-0.37	-0.38	0.22	0.63	0.74	2.9	13.4	13

Table 8: Results on the choice of the change-point number by the criteria associated to the methods: adaptive LASSO quantile, adaptive LASSO for LS model and LASSO-type for median(LAD) model, $K^0 = 1$. 100 Monte Carlo replications. Number of $\hat{K} = 0,1,2,3$ for 100 Monte Carlo replications.

$\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$	error distribution	QUANT+aLASSO number of $\hat{K} =$				LS+aLASSO number of $\hat{K} =$				LAD+LASSOtype number of $\hat{K} =$			
		0	1	2	3	0	1	2	3	0	1	2	3
$\boldsymbol{\phi}_1^0 = \boldsymbol{\phi}_2^0$	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{E}_3$	36	62	1	1	88	10	1	1	100	0	0	0
	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{N}(0,1)$	0	99	1	0	54	66	0	0	1	99	0	0
$\boldsymbol{\phi}_1^0 \neq \boldsymbol{\phi}_2^0$	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{E}_3$	0	100	0	0	0	100	0	0	0	100	0	0
	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{N}(0,1)$	0	100	0	0	0	100	0	0	0	100	0	0

This leads conclusion. ■

Proof of Theorem 2.3 *(i)* We reparameterize the model: $\mathbf{u}_n = \sqrt{n}(\hat{\boldsymbol{\phi}}_n^* - \boldsymbol{\phi}^{\mathbf{0}})$ and $u_{0,n} = \sqrt{n}(\hat{b}_n^* - b^0)$. Then $(u_{0,n}, \mathbf{u}_n)$ is the minimizer of the criterion

$$\sum_{i=1}^{n} \left[\rho_\tau \left(\varepsilon_i - b^0 - n^{-1/2}(u_{0,n} + \mathbf{X}_i^t \mathbf{u}_n)\right) - \rho_\tau(\varepsilon_i - b^0)\right] + \lambda_n \hat{\boldsymbol{\omega}}_n^t [|\boldsymbol{\phi}^{\mathbf{0}} + \mathbf{u}_n n^{-1/2}| - |\boldsymbol{\phi}^{\mathbf{0}}|].$$

The rest of the proof is similar as that of Theorem 4.1 in the paper of [10] and we omit it. With the remark that in our case (more general than in [10]) the supposition $n^{(g-1)/2}\lambda_n \to \infty$ is required to prove, in the case $\phi_{,j}^0 = 0$ and $u_{n,j} \neq 0$, that

$$\frac{\lambda_n}{\sqrt{n}|\hat{\boldsymbol{\phi}}_n|^g} \sqrt{n} \left(\left|\phi_{,j}^0 + \frac{u_{n,j}}{\sqrt{n}}\right| - |\phi_{,j}^0| \right) \xrightarrow[n \to \infty]{I\!P} \infty.$$

(ii) If $j \in \mathcal{A}^0$, then $\phi_{,j}^0 \neq 0$. Since $\hat{\phi}_{n,j}^*$ is the corresponding estimator of $\phi_{,j}^0$, and is asymptotically normal, thus $j \in \hat{\mathcal{A}}_n^*$ with probability tending to 1. Then $\lim_{n\to\infty} I\!P[\hat{\mathcal{A}}_n^* \supseteq \mathcal{A}^0] = 1$.
For the reverse inclusion, we show that if $j \notin \mathcal{A}^0$ then $j \notin \hat{\mathcal{A}}_n^*$. Let us calculate $I\!P[j \in \hat{\mathcal{A}}_n^*, j \notin \mathcal{A}^0]$.

Since $j \in \mathcal{A}_n^*$ it follows from the KKT optimality conditions of the Proposition 2.2(i), that

$$\tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} 1\!\!1_{Y_i < \mathbf{X}_i^t \hat{\phi}_n^*} = \lambda_n \hat{\omega}_{n,j} sgn(\hat{\phi}_{n,j}^*),$$

which implies that

$$\lambda_n \hat{\omega}_{n,j} < 2 \sum_{i=1}^n |X_{ij}|. \tag{11}$$

For the left member of the last inequality, we have

$$\frac{\lambda_n \hat{\omega}_{n,j}}{n} = \lambda_n \frac{1}{|n^{1/2} \hat{\phi}_{n,j}|^g} \cdot \frac{n^{g/2}}{n}, \tag{12}$$

where $\hat{\phi}_{n,j}$ is the quantile estimator of $\phi_{,j}^0$. On the other hand, since $j \not\in \mathcal{A}^0$, we have that $\phi_{,j}^0 = 0$. Using the fact that $\hat{\phi}_{n,j}$ is strongly consistent and asymptotically normal, we have that for all $\epsilon > 0$, there exists $\eta_\epsilon > 0$ such that

$$I\!\!P[n^{-1/2} |\hat{\phi}_{n,j}|^{-1} > \eta_\epsilon] > 1 - \epsilon. \tag{13}$$

Since $n^{g/2-1} \lambda_n \to \infty$, with the relation (13), we obtain that (12) converges to infinity with probability converging to 1 as $n \to \infty$. On the other hand, by the Cauchy-Schwarz inequality $n^{-1} \sum_{i=1}^n |X_{ij}| \leq \left(n^{-1} \sum_{i=1}^n X_{ij}^2 \right)^{1/2}$, which is bounded with probability converging 1, by assumption (A1). Taking into account (11), we obtain that (12) is bounded. Contradiction. This completes the proof. ■

Proof of Proposition 2.6 Let us denote $h_i = b^0 - b + \mathbf{X}_i^t (\boldsymbol{\phi}^\mathbf{0} - \boldsymbol{\phi})$. Then

$$I\!\!E[R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^\mathbf{0})] = \int_{\mathbb{R}} [\rho_\tau(u + h_i) - \rho_\tau(u)] dF(u + b^0).$$

If $h_i \leq 0$. By elementary calculations we obtain $\int_{-\infty}^0 [\rho_\tau(x - h_i) - \rho_\tau(x)] dF(x + b^0) = -\tau(1-\tau) h_i$ and also $\int_0^\infty [\rho_\tau(x - h_i) - \rho_\tau(x)] dF(x + b^0) = \tau(1-\tau) h_i - \int_0^{-h_i} [h_i(1-\tau) + \tau h_i + x] dF(x + b^0)$. Then, in this case $I\!\!E[R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^\mathbf{0})] = \int_0^{-h_i} [-x + |h_i|] dF(x + b^0)$.

If $h_i > 0$. By similar calculations as above we obtain that $I\!\!E[R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^\mathbf{0})] = \int_{-h_i}^0 [x + |h_i|] dF(x + b^0)$.

Thus, we can write all this in a more condensed rule:

$$I\!\!E[R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^\mathbf{0})] \geq 1\!\!1_{h_i > 0} \frac{|h_i|}{2} \int_{-\frac{h_i}{2}}^0 dF(x + b^0) + 1\!\!1_{h_i \leq 0} \frac{|h_i|}{2} \int_0^{-\frac{h_i}{2}} dF(x + b^0)$$

$$= \frac{|h_i|}{2} \left[1\!\!1_{h_i>0}[F(b^0) - F(b^0 - h_i/2)] + 1\!\!1_{h_i \leq 0}[F(b^0 - h_i/2) - F(b^0)] \right] \geq 0.$$

■

Proof of Theorem 3.2. <u>**Step I.**</u> We prove that, with probability approaching 1, the adaptive LASSO quantile change-point estimators are to a smaller distance than $[n^{1/2}]$. For this purpose, let us consider the constant $\varrho \in (\alpha, \min(1, (1+g)/2))$, with $\alpha > 1/2$ as in Lemma 2.8. We will prove that: $I\!\!P[|\hat{l}_r^* - l_r^0| > n^\varrho] \to 0$ as $n \to \infty$, for each $r = 1, \cdots, K$.
Consider the set of change-points, all close to the true points at a distance less $[n^\varrho]$:

$$\mathcal{L}(\varrho) \equiv \{(l_1, \cdots, l_K); \sum_{r=}^{K} |l_r - l_r^0| \leq [n^\varrho]\}.$$

Consider a subset of its complementary, for some $r \in \{1, \cdots, K\}$:

$$\mathcal{L}_r^c(\varrho) \equiv \{(l_1, \cdots, l_K); |l_t - l_r^0| > n^\varrho, \forall t = 1, \cdots, K\},$$

with all the change-points to a distance of l_r^0 greater than $[n^\varrho]$. By the definition of the objective function S^* we have, for all $(l_1, \cdots, l_K) \in \mathcal{L}_r^c(\varrho)$, with probability 1:

$$S^*(l_1, \cdots, l_K) \geq S^*(l_1, \cdots, l_K, l_1^0, \cdots, l_{r-1}^0, l_r^0 - [n^\rho], l_r^0 + [n^\rho], l_{r+1}^0, \cdots, l_K^0) \equiv \sum_{q=1}^{K+2} \mathcal{T}_q, \qquad (14)$$

where $\mathcal{T}_q$ are, for $q \in \{1, \cdots, r-1, r+1, \cdots, K+2\}$, the penalized sums involving observations between l_{q-1}^0 and l_q^0. $\mathcal{T}_r$ is the penalized sum involving observations between l_{r-1}^0 and $l_r^0 - [n^\varrho]$ and $\mathcal{T}_{r+1}$ between $l_r^0 + [n^\varrho]$ and l_{r+1}^0; $\mathcal{T}_{K+2}$ is calculated between $l_r^0 - [n^\varrho]$ and $l_r^0 + [n^\varrho]$. Note that the sum $S^*(l_1, \cdots, l_K, l_1^0, \cdots, l_{r-1}^0, l_r^0 - [n^\rho], l_r^0 + [n^\rho], l_{r+1}^0, \cdots, l_K^0)$ is the extension of the definition (7) for $2K+1$ change-points: $l_1, \cdots, l_K, l_1^0, \cdots, l_{r-1}^0$, $l_r^0 - [n^\rho], l_r^0 + [n^\rho], l_{r+1}^0, \cdots, l_K^0$.
Since $(l_1, \cdots, l_K) \in \mathcal{L}_r^c(\varrho)$, all change-points $l_1, \cdots, l_K$ are in $\mathcal{T}_1, \cdots, \mathcal{T}_{r-1}, \cdots, \mathcal{T}_{K+1}$ and none in $\mathcal{T}_{K+2}$. For each $q \in \{1, \cdots, r-1, r+1, \cdots, K+2\}$, let us consider the points between two true consecutive change-points $k_{1,q} < \cdots < k_{J(q),q} \equiv \{l_1, \cdots, l_K\} \cap \{j; l_{q-1}^0 < j \leq l_q^0\}$. The number $J(q)$ is greater (or equal) than zero and smaller (or equal) than K, with the property that for $q \neq q'$, $J(q) \neq J(q')$. The penalized sum $\mathcal{T}_q$ can be written

$$\mathcal{T}_q = \sum_{j=1}^{J(q)+1} \min_{b_j, \boldsymbol{\phi}_j} [\sum_{i=k_{j-1,q}+1}^{k_{j,q}} \rho_\tau(\varepsilon_i - b_j - \mathbf{X}_i^t(\boldsymbol{\phi}_j - \boldsymbol{\phi}_q^0)) + \lambda_{(k_{j-1,q};k_{j,q})} \hat{\boldsymbol{\omega}}_{(k_{j-1,q};k_{j,q})}^t |\boldsymbol{\phi}_j|].$$

It is obvious that

$$0 \geq \mathcal{T}_q - \sum_{j=1}^{J(q)+1} [\sum_{i=k_{j-1,q}+1}^{k_{j,q}} \rho_\tau(\varepsilon_i - b_q^0) + \lambda_{(k_{j-1,q};k_{j,q})} \hat{\boldsymbol{\omega}}_{(k_{j-1,q};k_{j,q})}^t |\boldsymbol{\phi}_t^0|]$$

$$\geq -2(K+1) \sup_{0 \leq l < k \leq n, k-l=d_n} | \inf_{b,\boldsymbol{\phi}} \sum_{i=l+1}^{k} R^{(\tau,\lambda)}_{i,(l;k)}(b,\boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})|,$$

with l and k two observations without any change between they, $b^0, \boldsymbol{\phi}^{\mathbf{0}}$ the true values of the parameters on this interval and d_n a convergent sequence to ∞ when n converges to ∞. Due to Lemma 2.8,

$$2(K+1) \sup_{0 \leq l < k \leq n, k-l=d_n} | \inf_{b,\boldsymbol{\phi}} \sum_{i=l+1}^{k} R^{(\tau,\lambda)}_{i,(l;k)}(b,\boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})| = -\min(O_{I\!P}(n^{\alpha}), o_{I\!P}(n^{(1+g)/2})). \tag{15}$$

For the observations between $l_r^0 - [n^{\varrho}]$ and $l_r^0 + [n^{\varrho}]$, we have that:

$$\mathcal{T}_r - \sum_{j=1}^{J(r)+1} [\sum_{i=k_{j_1,r}+1}^{k_{j,r}} \rho_\tau(\varepsilon_i - b_r^0) + \lambda_{(k_{j-1,r};k_{j,r})} \hat{\boldsymbol{\omega}}^t_{(k_{j-1,r};k_{j,r})} |\boldsymbol{\phi}_r^0| + \lambda_{(k_{j-1,r};k_{j,r})} \hat{\boldsymbol{\omega}}^t_{(k_{j-1,r};k_{j,r})} |\boldsymbol{\phi}_{r+1}^0|]$$

$$= \sum_{j=1}^{J(r)+1} \min_{b_j,\boldsymbol{\phi}_j} [\sum_{i=k_{j_1,r}+1}^{k_{j,r}} \rho_\tau(\varepsilon_i - b_j - \mathbf{X}_i^t(\boldsymbol{\phi}_j - \boldsymbol{\phi}_t^0)) - \rho_\tau(\varepsilon_i - b_r^0) + \lambda_{(k_{j-1,t};k_{j,t})} \hat{\boldsymbol{\omega}}^t_{(k_{j-1,t};k_{j,t})} |\boldsymbol{\phi}_j|]$$
$$- \lambda_{(k_{j-1,r};k_{j,r})} \hat{\boldsymbol{\omega}}^t_{(k_{j-1,r};k_{j,r})} [|\boldsymbol{\phi}_r^0| + |\boldsymbol{\phi}_{r+1}^0|].$$

But for $(l_1, \cdots, l_K) \in \mathcal{L}_r^c(\varrho)$, there is no other change-point between $l_r^0 - [n^{\varrho}]$ and $l_r^0 + [n^{\varrho}]$, so this last relation is in fact

$$\min_{b,\boldsymbol{\phi}} [\sum_{i=l_r^0-[n^{\varrho}]+1}^{l_r^0} R^{(\tau,\lambda)}_{i;(l_r^0-[n^{\varrho}];l_r^0)}(b,\boldsymbol{\phi}; b_r^0, \boldsymbol{\phi}_r^0) + \sum_{i=l_r^0+1}^{l_r^0+[n^{\varrho}]} R^{(\tau,\lambda)}_{i;(l_r^0;l_r^0+[n^{\varrho}])}(b,\boldsymbol{\phi}; b_{r+1}^0, \boldsymbol{\phi}_{r+1}^0)]. \tag{16}$$

Since to left and to right of l_r^0 we have different models, then $(b_r^0, \boldsymbol{\phi}_r^0) \neq (b_{r+1}^0, \boldsymbol{\phi}_{r+1}^0)$. This implies that for the $(b, \boldsymbol{\phi})$ which minimizes (16) we have either $|b - b_r^0| + \|\boldsymbol{\phi} - \boldsymbol{\phi}_r^0\|_2 > c > 0$ or $|b - b_{r+1}^0| + \|\boldsymbol{\phi} - \boldsymbol{\phi}_{r+1}^0\|_2 > c > 0$. Without loss of generality, consider the last case. Using Lemma 2.10 for $c_n = c > 0$ on the interval $(l_r^0, l_r^0 + [n^{\varrho}])$, we obtain that, there exists $\epsilon > 0$ such that

$$c^{-1}[n^{\varrho}]^{-1} \sum_{i=l_r^0+1}^{l_r^0+[n^{\varrho}]} R^{(\tau,\lambda)}_{i;(l_r^0;l_r^0+[n^{\varrho}])}(b,\boldsymbol{\phi}; b_{r+1}^0, \boldsymbol{\phi}_{r+1}^0) > \epsilon > 0.$$

Thus

$$\inf_{b,\boldsymbol{\phi}} \sum_{i=l_r^0+1}^{l_r^0+[n^{\varrho}]} R^{(\tau,\lambda)}_{i;(l_r^0;l_r^0+[n^{\varrho}])}(b,\boldsymbol{\phi}; b_{r+1}^0, \boldsymbol{\phi}_{r+1}^0) \geq O_{I\!P}(n^{\varrho}) > 0. \tag{17}$$

Figure 5:

Then, taking into account the relations (14), (15), (16), (17) and the fact that by definition of S^*, we have that $S^*(\hat{l}^*, \cdots, \hat{l}^*_K) \leq S^*_0$ with probability 1, we obtain that $S^*(l_1, \cdots, l_K) - S^*_0$ is greater than $\sum_{q=1}^{K+2} \mathcal{T}_q - S^*_0$ which is greater than $-\min(O_{I\!P}(n^\alpha), o_{I\!P}(n^{(1+g)/2})) + O_{I\!P}(n^\varrho) + \sum_{q=1}^{K+2} \sum_{j=1}^{J(q)+1} \lambda_{(k_{j-1,q}, k_{j,q})} \hat{\boldsymbol{\omega}}_{(k_{j-1,q}, k_{j,q})} - \sum_{r=1}^{K+1} \lambda_{(l^0_{r-1}, l^0_r)} \hat{\boldsymbol{\omega}}^t_{(l^0_{r-1}; l^0_r)} |\boldsymbol{\phi}^0_r| = O_{I\!P}(n^\varrho) + o_{I\!P}(n^{(g+1)/2}) = O_{I\!P}(n^\varrho) > 0$. Then, for $n \to \infty$,

$$I\!P[\min_{(l_1, \cdots, l_K) \in \mathcal{L}^c_r(\varrho)} S^*(l_1, \cdots, l_K) > S^*_0] \to 1.$$

On the other hand, since $(\hat{l}^*, \cdots, \hat{l}^*_K)$ are the change-points estimators, we have that $S^*(\hat{l}^*_1, \cdots, \hat{l}^*_K) \leq S^*_0$, with probability 1. These last two relations imply

$$I\!P[(\hat{l}^*_1, \cdots, \hat{l}^*_K) \in \mathcal{L}^c_r(\varrho)] \to 0, \text{as } n \to \infty, \text{ for every } r = 1, \cdots, K.$$

Step II. We show that for all $\nu < 1/4$, for every $r = 1, \cdots, K$ we have $I\!P[|\hat{l}^*_r - l^0_r| > n^\nu] \to 0$, as $n \to \infty$.

By Step I, we have $I\!P[(\hat{l}^*_1, \cdots, \hat{l}^*_K) \in \mathcal{L}(\varrho)] \to 1$, as $n \to \infty$. Consider then the change-points $(l_1, \cdots, l_K)$ belonging to the set $\mathcal{L}(\varrho)$ and $(l_1, \cdots, l_K) \in \mathcal{L}^c_r(\nu)$, where $\mathcal{L}^c_r(\nu)$ is similar to the set $\mathcal{L}^c_r(\varrho)$ with ν instead of ϱ. For these change-points, there is a similar relation to (14) for the objective function S^*. For $q \neq r-1, r$, by assumption (A3), using step I, we have that there are at most two points l_q and l_{q+1} between l^0_q and l^0_{q+1}. Suppose that there are two points l_q and l_{q+1} between l^0_q and l^0_{q+1} (see Figure 5). If there is a single point or no point the approach is the same. Let us define the following sums

$$\begin{aligned}
\mathcal{D}(l^0_q, l^0_{q+1}) \equiv \quad & \inf_{b,\phi}\{\textstyle\sum_{i=l^0_q+1}^{l_q} \rho_\tau(\varepsilon_i - b - \mathbf{X}^t_i(\boldsymbol{\phi} - \boldsymbol{\phi}^0_{q+1})) + \lambda_{(l^0_q; l_q)} \hat{\boldsymbol{\omega}}^t_{(l^0_q; l_q)} |\boldsymbol{\phi}|\} \\
& + \inf_{b,\phi}\{\textstyle\sum_{i=l_q+1}^{l_{q+1}} \rho_\tau(\varepsilon_i - b - \mathbf{X}^t_i(\boldsymbol{\phi} - \boldsymbol{\phi}^0_{q+1})) + \lambda_{(l_q; l_{q+1})} \hat{\boldsymbol{\omega}}^t_{(l_q; l_{q+1})} |\boldsymbol{\phi}|\} \\
& + \inf_{b,\phi}\{\textstyle\sum_{i=l_{q+1}+1}^{l^0_{q+1}} \rho_\tau(\varepsilon_i - b - \mathbf{X}^t_i(\boldsymbol{\phi} - \boldsymbol{\phi}^0_{q+1})) + \lambda_{(l_{q+1}; l^0_{q+1})} \hat{\boldsymbol{\omega}}^t_{(l_{q+1}; l^0_{q+1})} |\boldsymbol{\phi}|\}.
\end{aligned}$$

The sums $\mathcal{D}(l^0_{r-1}, l^0_r - [n^\nu])$, $\mathcal{D}(l^0_r + [n^\nu], l^0_{r+1})$, $\mathcal{D}(l^0_r - [n^\nu], l^0_r + [n^\nu])$ can be defined in the same way. Then, for the difference between the following two objective functions

$$S^*(l_1, \cdots, l_K, l^0_1, \cdots, l^0_{r-1}, l^0_r - [n^\nu], l^0_r + [n^\nu], l^0_{r+1}, \cdots, l^0_K) - S^*_0,$$

we focus on what happens between two change-points. This last difference can be written

$$
\begin{aligned}
&\textstyle\sum_{q\neq r-1,r}\left\{\mathcal{D}(l_q^0,l_{q+1}^0)-\sum_{i=l_q^0+1}^{l_{q+1}^0}\rho_\tau(\varepsilon_i-b_{q+1}^0)-\lambda_{(l_q^0;l_{q+1}^0)}\hat{\boldsymbol{\omega}}^t_{(l_q^0;l_{q+1}^0)}|\boldsymbol{\phi}_{q+1}^0|\right\}\\
&+\left\{\mathcal{D}(l_{r-1}^0,l_r^0-[n^\nu])-\textstyle\sum_{i=l_{r-1}^0+1}^{l_r^0-[n^\nu]}\rho_\tau(\varepsilon_i-b_r^0)\ -\lambda_{(l_{r-1}^0;l_r^0-[n^\nu])}\hat{\boldsymbol{\omega}}^t_{(l_{r-1}^0;l_r^0-[n^\nu])}|\boldsymbol{\phi}_r^0|\right\}\\
&+\left\{\mathcal{D}(l_r^0+[n^\nu],l_{r+1}^0)-\textstyle\sum_{i=l_r^0+[n^\nu]+1}^{l_{r+1}^0}\rho_\tau(\varepsilon_i-b_{r+1}^0)\ -\lambda_{(l_r^0+[n^\nu];l_{r+1}^0)}\hat{\boldsymbol{\omega}}^t_{(l_r^0+[n^\nu];l_{r+1}^0)}|\boldsymbol{\phi}_{r+1}^0|\right\}\\
&+\left\{\mathcal{D}(l_r^0-[n^\nu],l_r^0+[n^\nu])-\textstyle\sum_{i=l_r^0+1}^{l_r^0+[n^\nu]}\rho_\tau(\varepsilon_i-b_{r+1}^0)\ -\lambda_{(l_r^0;l_r^0+[n^\nu])}\hat{\boldsymbol{\omega}}^t_{(l_r^0;l_r^0+[n^\nu])}|\boldsymbol{\phi}_{r+1}^0|\right.\\
&\left.-\textstyle\sum_{i=l_r^0-[n^\nu]+1}^{l_r^0}\rho_\tau(\varepsilon_i-b_r^0)\ -\lambda_{(l_r^0-[n^\nu];l_r^0)}\hat{\boldsymbol{\omega}}^t_{(l_r^0-[n^\nu];l_r^0)}|\boldsymbol{\phi}_r^0|\right\}\equiv D_1+D_2+D_3+D_4.
\end{aligned}
$$

By Remark 2.11, we have that $D_1, D_2, D_3 = O_{I\!P}(1)$.
For D_4 we have by the definition of sums $\mathcal{D}$ that $\mathcal{D}(l_r^0-[n^\nu],l_r^0+[n^\nu])$ is equal to

$$
\begin{aligned}
\inf_{b,\boldsymbol{\phi}}\{&\textstyle\sum_{i=l_r^0-[n^\nu]+1}^{l_r^0}\rho_\tau(\varepsilon_i-b-\mathbf{X}_i^t(\boldsymbol{\phi}-\boldsymbol{\phi}_r^0))+\lambda_{(l_r^0-[n^\nu];l_r^0)}\hat{\boldsymbol{\omega}}^t_{(l_r^0-[n^\nu];l_r^0)}|\boldsymbol{\phi}|\\
&+\textstyle\sum_{i=l_r^0+1}^{l_r^0+[n^\nu]}\rho_\tau(\varepsilon_i-b-\mathbf{X}_i^t(\boldsymbol{\phi}_{r+1}^0))+\lambda_{(l_r^0;l_r^0+[n^\nu])}\hat{\boldsymbol{\omega}}^t_{(l_r^0;l_r^0+[n^\nu])}|\boldsymbol{\phi}|\}.
\end{aligned}
$$

Applying Lemma 2.10 for $c_n = c$ on the one of the intervals $(l_r^0-[n^\nu];l_r^0)$ or $(l_r^0;l_r^0+[n^\nu])$, (it is the one where $|b-b^0|+\|\boldsymbol{\phi}-\boldsymbol{\phi^0}\|_2>\tilde{c}>0$), we have $D_4 = O_{I\!P}(n^\nu)>0$. Then, with probability converging to 1, as $n\to\infty$, we have

$$
\inf_{(l_1,\cdots,l_K)\in\mathcal{L}_r^c(\nu)}\left[S^*(l_1,\cdots,l_K,l_1^0,\cdots,l_{r-1}^0,l_r^0-[n^\nu],l_r^0+[n^\nu],l_{r+1}^0,\cdots,l_K^0)-S_0^*\right]>O_{I\!P}(n^\nu). \quad (18)
$$

Therefore, we have proved that $I\!P[(\hat{l}_1^*,\cdots,\hat{l}_K^*)\in\mathcal{L}_r^c(\nu)]\to 0$, as $n\to\infty$.
Step III. We now prove that $\hat{l}_r^*-l_r^0=O_{I\!P}(1)$, for each $r=1,\cdots,K$.
Let the set: $\mathcal{L}(\nu)\equiv\left\{(l_1,\cdots,l_K);|l_t-l_t^0|<[n^\nu],\forall t=1,\cdots,K\right\}$ with $\nu<1/4$. As a consequence of the Step II, for n large, the estimator $(\hat{l}_1^*,\cdots,\hat{l}_K^*)$ belongs to $\mathcal{L}(\nu)$ with a probability tending to 1. We use the reduction to absurdity method, supposing that there exists a change-point estimator at an unbounded distance from the true value. Consider then, for a $\mathcal{M}_1>0$ (to be determine later) the following set:

$$
\mathcal{L}_r(\nu,\mathcal{M}_1)\equiv\left\{(l_1,\cdots l_K)\in\mathcal{L}(\nu);l_r-l_r^0<-\mathcal{M}_1\right\}.
$$

The case $l_r-l_r^0>\mathcal{M}_1$ is similar. We shall find a $\mathcal{M}_1$ such that the probability that the change-point estimator belongs to the set $\mathcal{L}_r(\nu,\mathcal{M}_1)$ converges to 0.
Consider two vectors of change-points $(m_1,\cdots,m_K)\in\mathcal{L}(\nu)$ and $(l_1,\cdots,l_K)\in\mathcal{L}_r(\nu,\mathcal{M}_1)$ such that $m_t=l_t$ for $t\neq r$ and $m_r=l_r^0$. We have for the difference of the corresponding objective

functions:

$$
\begin{aligned}
&S^*(l_1,\cdots,l_K)-S^*(m_1,\cdots,m_K)=\{\textstyle\sum_{i=l_{r-1}+1}^{l_r}[\rho_\tau(Y_i-\hat{b}^*_{(l_{r-1},l_r)}-\mathbf{X}_i^t\hat{\boldsymbol{\phi}}^*_{(l_{r-1},l_r)})\\
&-\rho_\tau(Y_i-\hat{b}^*_{(l_{r-1},l_r^0)}-\mathbf{X}_i^t\hat{\boldsymbol{\phi}}^*_{(l_{r-1},l_r^0)})]+\lambda_{(l_{r-1},l_r)}\hat{\boldsymbol{\omega}}^t_{(l_{r-1},l_r)}|\hat{\boldsymbol{\phi}}^*_{(l_{r-1},l_r)}|-\lambda_{(l_{r-1},l_r^0)}\hat{\boldsymbol{\omega}}^t_{(l_{r-1},l_r^0)}|\hat{\boldsymbol{\phi}}^*_{(l_{r-1},l_r^0)}|\}\\
&+\{\textstyle\sum_{i=l_r+1}^{l_r^0}[\rho_\tau(Y_i-\hat{b}^*_{(l_r,l_{r+1})}-\mathbf{X}_i^t\hat{\boldsymbol{\phi}}^*_{(l_r,l_{r+1})})-\rho_\tau(Y_i-\hat{b}^*_{(l_{r-1},l_r^0)}-\mathbf{X}_i^t\hat{\boldsymbol{\phi}}^*_{(l_{r-1},l_r^0)})]\}\\
&+\{\textstyle\sum_{i=l_{r+1}^0}^{l_{r+1}}[\rho_\tau(Y_i-\hat{b}^*_{(l_r,l_{r+1})}-\mathbf{X}_i^t\hat{\boldsymbol{\phi}}^*_{(l_r,l_{r+1})})-\rho_\tau(Y_i-\hat{b}^*_{(l_r^0,l_{r+1})}-\mathbf{X}_i^t\hat{\boldsymbol{\phi}}^*_{(l_r^0,l_{r+1})})]\\
&+\lambda_{(l_r,l_{r+1})}\hat{\boldsymbol{\omega}}^t_{(l_r,l_{r+1})}|\hat{\boldsymbol{\phi}}^*_{(l_r,l_{r+1})}|-\lambda_{(l_r^0,l_{r+1})}\hat{\boldsymbol{\omega}}^t_{(l_r^0,l_{r+1})}|\hat{\boldsymbol{\phi}}^*_{(l_r^0,l_{r+1})}|\}\\
&\equiv\{S_{11}+S_{12}\}+\{S_{21}\}+\{S_{31}+S_{32}\}.
\end{aligned}
$$

Taking into account Remark 2.11, is easily obtained that $S_{11}=O_{I\!P}(1)$, uniformly in $\mathcal{M}_1$. Taking into account that $\lambda_{(l_{r-1},l_r)}=O((l_r-l_{r-1})^{1/2})$ together the properties of strong convergence of the quantile estimator $\hat{\boldsymbol{\phi}}_{(l_{r-1},l_r)}$ and of adaptive quantile estimator $\hat{\boldsymbol{\phi}}^*_{(l_{r-1},l_r)}$, we obtain that $S_{12}=o_{I\!P}(l^0-l_r)$. Similarly, we have that $S_{31}=O_{I\!P}(1)$ and $S_{32}=o_{I\!P}(l^0-l_r)$.
It remains to study the most difficult part, that is, S_{21}, which can be written $S_{21}=\sum_{i=l_r+1}^{l_r^0}R_i^{(\tau)}(b^0_{r+1},\boldsymbol{\phi}^0_{r+1};b^0_r,\boldsymbol{\phi}^0_r)-\sum_{i=l_r+1}^{l_r^0}[\rho_\tau(\varepsilon_i-\hat{b}^*_{(l_r,l_{r+1})}-\mathbf{X}_i^t(\hat{\boldsymbol{\phi}}^*_{(l_r,l_{r+1})}-\boldsymbol{\phi}^0_r))-\rho_\tau(\varepsilon_i-b^0_{r+1}-\mathbf{X}_i^t(\boldsymbol{\phi}^0_{r+1}-\boldsymbol{\phi}^0_r))]-\sum_{i=l_r+1}^{l_r^0}R_i^{(\tau)}(\hat{b}^*_{(l_{r-1},l_r^0)},\hat{\boldsymbol{\phi}}^*_{(l_{r-1},l_r^0)};b^0_r,\boldsymbol{\phi}^0_r)\equiv S_{211}-S_{212}-S_{213}$.
For S_{212} and S_{213} we use the inequalities

$$\left|\frac{\rho_\tau(r_1)-\rho_\tau(r_2)}{r_1-r_2}\right|\leq\max(\tau,1-\tau)<1$$

and we obtain that $|S_{212}|$ is smaller than

$$\sum_{i=l_r+1}^{l_r^0}|b^0_{r+1}-\hat{b}^*_{(l_r,l_{r+1})}+\mathbf{X}_i^t(\boldsymbol{\phi}^0_{r+1}-\hat{\boldsymbol{\phi}}^*_{(l_r,l_{r+1})})|\leq|b^0_{r+1}-\hat{b}^*_{(l_r,l_{r+1})}|(l_r^0-l_r)+\|\boldsymbol{\phi}^0_{r+1}-\hat{\boldsymbol{\phi}}^*_{(l_r,l_{r+1})}\|_2\sum_{i=l_r+1}^{l_r^0}\|\mathbf{X}_i\|_2$$

that is, using Remark 2.11, of order $o_{I\!P}(1)$. We obtain analogously $|S_{213}|=o_{I\!P}(1)$. For S_{211}, combining the relation (43) for $c_n=c=\max(|b^0_r-b^0_{r+1}|,\|\boldsymbol{\phi}^0_r-\boldsymbol{\phi}^0_{r+1}\|_2)$ together with Lemma 3.1 yield that $S_{211}\geq(l_r^0-l_r)\eta\geq\mathcal{M}_1\eta$, with probability converging to 1 as $\mathcal{M}_1\to\infty$.
Choosing $\mathcal{M}_1>0$ such that $S_{211}\geq\max(|S_{21}|,|S_{11}|,|S_{31}|,|S_{32}|)$ we have that

$$\lim_{n\to\infty}I\!P[(\hat{l}_1^*,\cdots,\hat{l}_K^*)\in\mathcal{L}_r(\nu,\mathcal{M}_1)]=0.$$

Which proves that $\hat{l}_r^*-l_r^0=O_{I\!P}(1)$, for each $r=1,\cdots,K$. ■

Proof of Theorem 3.3 *(i)* This assertion follows immediately from Theorem 2.3(i) and Theorem 3.2.

(ii) The consistency of the variable selection in a model with one phase, property established

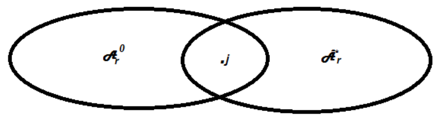

Figure 6:

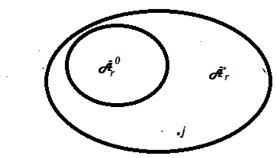

Figure 7:

by Theorem 2.3(ii), implies that

$$\lim_{n\to\infty} I\!P\left[\hat{\mathcal{A}}^0_{n,r} = \mathcal{A}^0_r\right] = 1. \tag{19}$$

It remains to prove that $\lim_{n\to\infty} I\!P\left[\hat{\mathcal{A}}^0_{n,r} = \hat{\mathcal{A}}^*_{n,r}\right] = 1$. The general case is considered in the early, that is presented in Figure 6.

If $j \in \hat{\mathcal{A}}^0_{n,r}$, thus, using (19), we have that $j \in \mathcal{A}^0_r$ with probability tending to 1, which implies that $\phi^0_{r,j} \neq 0$. Moreover, using the result proved in the previous question (i), and the fact that the adaptive LASSO quantile estimator $\hat{\phi}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}$ of the jth component of the regression parameter $\boldsymbol{\phi}^0_r$, calculated between the corresponding adaptive LASSO quantile estimators of the change-points, is asymptotically normal, we obtain that $\hat{\phi}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r);j} \overset{I\!P}{\underset{n\to\infty}{\longrightarrow}} \phi^0_{r,j} \neq 0$. Then,

$$\lim_{n\to\infty} I\!P\left[\hat{\phi}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r);j} \neq 0\right] = 1, \text{ i.e. } I\!P[j \in \hat{\mathcal{A}}^*_{n,r}] \to 1.$$

Thus $I\!P[\hat{\mathcal{A}}^0_{n,r} \subseteq \hat{\mathcal{A}}^*_{n,r}] \to 1$.

There, remains now the most difficult part: to prove that, if the index $j \notin \hat{\mathcal{A}}^0_r$, then $j \notin \hat{\mathcal{A}}^*_{n,r}$ (in fact, in view of (19), we must show that if the true component is zero, the corresponding estimators don't converge to 0 and for a fixed n, the component $\hat{\phi}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r);j} \neq 0$). We will then calculate $I\!P[j \notin \mathcal{A}^0_r, j \in \hat{\mathcal{A}}^*_{n,r}]$ (see Figure 7). Since $j \in \hat{\mathcal{A}}^*_{n,r}$, by KKT optimality condition, Proposition 2.2(i), we have

$$\tau \sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} X_{ij} - \sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} X_{ij} 1\!\!1_{Y_i < \mathbf{X}^t_i \hat{\phi}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)}} = \lambda_{(\hat{l}^*_{r-1};\hat{l}^*_r)} \hat{\omega}_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}\, sgn(\hat{\phi}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}).$$

Then

$$\lambda_{(\hat{l}^*_{r-1};\hat{l}^*_r)}\hat{\omega}_{(\hat{l}^*_{r-1};\hat{l}^*_r),j} = \left| \tau \sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} X_{ij} - \sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} X_{ij} 1\!\!1_{Y_i < \mathbf{X}_i^t \hat{\phi}^*_{(\hat{l}^*_{r-1};\hat{l}^*_r)}} \right| < 2 \sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} |X_{ij}|. \qquad (20)$$

On the other hand

$$\frac{\lambda_{(\hat{l}^*_{r-1};\hat{l}^*_r)}\hat{\omega}_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}}{\hat{l}^*_r - \hat{l}^*_{r-1}} = \frac{\lambda_{(\hat{l}^*_{r-1};\hat{l}^*_r)}}{|(\hat{l}^*_r - \hat{l}^*_{r-1})^{1/2}\hat{\phi}_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}|^g} \cdot \frac{(\hat{l}^*_r - \hat{l}^*_{r-1})^{g/2}}{\hat{l}^*_r - \hat{l}^*_{r-1}}, \qquad (21)$$

where $\hat{\phi}_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}$ is the quantile estimator of $\phi^0_{r,j}$. Since $j \in \mathcal{A}^0_r$, we have that $\phi^0_{r,j} \neq 0$. On the other hand, taking into account Theorem 3.2 and the fact that the quantile estimator $\hat{\phi}_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}$ is strongly consistent and asymptotically normal, we have that for all $\epsilon > 0$, there exists $\eta_\epsilon > 0$ such that

$$I\!\!P\Big[\Big((\hat{l}^*_r - \hat{l}^*_{r-1})^{1/2}|\hat{\phi}_{(\hat{l}^*_{r-1};\hat{l}^*_r),j}|\Big)^{-1} > \eta_\epsilon\Big] > 1 - \epsilon. \qquad (22)$$

Since $(l_r - l_{r-1})^{g/2-1}\lambda_{(l_{r-1},l_r)} \to \infty$, together with the relation (22), we obtain that (21) converges to infinity with probability converging to 1 as $n \to \infty$. On the other hand, an application of Cauchy-Schwarz's inequality yields that

$$(\hat{l}^*_r - \hat{l}^*_{r-1})^{-1} \sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} |X_{ij}| \leq \left((\hat{l}^*_r - \hat{l}^*_{r-1})^{-1} \sum_{i=\hat{l}^*_{r-1}+1}^{\hat{l}^*_r} X^2_{ij} \right)^{1/2},$$

wich is bounded with probability converging to 1, from assumption (A1). Taking into account (20), we obtain that (21) is bounded. Contradiction. Then

$$I\!\!P[j \notin \mathcal{A}^0_r, j \in \hat{\mathcal{A}}^*_{n,r}] \to 0, \text{ as } n \to \infty.$$

Which implies that $I\!\!P[\hat{\mathcal{A}}^*_{n,r} \subseteq \mathcal{A}^0_r] \to 1$ and in view of relation (19), we have that $I\!\!P[\hat{\mathcal{A}}^*_{n,r} \subseteq \hat{\mathcal{A}}^0_{n,r}] \to 1$, as $n \to \infty$. ■

Proof of Theorem 4.1 For K^0 the true number of changes, let us define the objective function calculated for the true values of the parameters. Only the weights are estimated:

$$\mathcal{S}_0 \equiv \sum_{i=1}^{n} \sum_{r=1}^{K^0+1} \rho_\tau(\varepsilon_i - b^0_r 1\!\!1_{l^0_{r-1} \leq i < l^0_r}) + \sum_{r=1}^{K^0+1} \lambda_{(l^0_{r-1};l^0_r)} \hat{\boldsymbol{\omega}}^t_{(l^0_{r-1};l^0_r)} |\boldsymbol{\phi}^0_r|, \qquad (23)$$

with $\hat{\boldsymbol{\omega}}^t_{(l^0_{r-1};l^0_r)} = |\hat{\boldsymbol{\phi}}_{(l^0_{r-1};l^0_r)}|^{-g}$, calculated on the basis of the quantile estimator, on the observations between l^0_{r-1} et l^0_r.

We will first study the behavior of $\hat{s}^*_{K^0}$ for the true number of phases. Using Lemma 2.8, for $\alpha > 1/2$, we have for the difference between the objective function S^* calculated for the adaptive LASSO change-points estimators and the sum calculated for the true values:

$$S^*(\hat{l}^*_{1,K^0}, \cdots, \hat{l}^*_{K^0,K^0}) - \mathcal{S}_0 = \min(o_{I\!P}(n^{(1+g)/2}), O_{I\!P}(n^{\alpha})). \tag{24}$$

Then

$$\hat{s}^*_{K^0} = n^{-1} S^*(\hat{l}^*_{1,K^0}, \cdots, \hat{l}^*_{K^0,K^0}) - n^{-1}\mathcal{S}_0 + n^{-1}\mathcal{S}_0 = \min(o_{I\!P}(n^{(g-1)/2}), O_{I\!P}(n^{\alpha-1/2})) + n^{-1}\mathcal{S}_0.$$

By the weak law of large numbers, using assumption (A4) and the independence of (ε_i), we have

$$n^{-1}\mathcal{S}_0 \xrightarrow[n\to\infty]{I\!P} \sum_{r=1}^{K^0+1} I\!E[\rho_\tau(\varepsilon - b^0_r)].$$

Thus,

$$\hat{s}^*_{K^0} \xrightarrow[n\to\infty]{I\!P} \sum_{r=1}^{K^0+1} I\!E[\rho_\tau(\varepsilon - b^0_r)] > 0. \tag{25}$$

Now, we show that

$$I\!P[\hat{K}^*_n < K^0] \to 0, \qquad \text{as } n \to \infty. \tag{26}$$

In order to prove (26), consider K any change-point number , with $K < K^0$. Then

$$B(K) - B(K^0) = n \log\left(1 + \frac{\hat{s}^*_K - \hat{s}^*_{K^0}}{\hat{s}^*_{K^0}}\right) + B_n[G(K,p) - G(K^0,p)].$$

Two cases are possible concerning $(\hat{s}^*_K - \hat{s}^*_{K^0})/\hat{s}^*_{K^0}$.

- If $(\hat{s}^*_K - \hat{s}^*_{K^0})/\hat{s}^*_{K^0}$ is great ($\geq C > 0$), then, since $\hat{s}^*_{K^0} > \epsilon_2 > 0$, we have that there exists $\epsilon_1 > 0$ such that $\hat{s}^*_K - \hat{s}^*_{K^0} > \epsilon_1 > 0$ for any n large enough. Then

  $$I\!P[\arg\min_K \hat{s}^*_K < K^0] \to 0, \text{ for } n \to \infty$$

  and the relation (26) follows.

- If $(\hat{s}^*_K - \hat{s}^*_{K^0})/\hat{s}^*_{K^0} = o_{I\!P}(1)$, then, using the fact that for x close to 0 we have $\log(1+x) \simeq x$, we have

  $$B(K) - B(K^0) = n\frac{\hat{s}^*_K - \hat{s}^*_{K^0}}{\hat{s}^*_{K^0}}(1 + o_{I\!P}(1)) + B_n[G(K,p) - G(K^0,p)]. \tag{27}$$

  We will study in this case the first term of the right side of the relation (27). Recall that the constant $a \in (1/2, 1)$ is that of the assumption (A3). Similarly of the relation (18), since between any two consecutive change-points l_{r-1} and l_r there are at least $[n^a]$ observations and since when $K < K^0$ there is at least a non estimated true change-point, we obtain that

$$S^*(\hat{l}^*_{1,K}, \cdots, \hat{l}^*_{K,K}) - \mathcal{S}_0 > Cn^a. \qquad (28)$$

Then, $n(\hat{s}^*_K - \hat{s}^*_{K^0})/\hat{s}^*_{K^0} = [S^*(\hat{l}^*_{1,K}, \cdots, \hat{l}^*_{K,K}) - \mathcal{S}_0 - S^*(\hat{l}^*_{1,K^0}, \cdots, \hat{l}^*_{K^0,K^0}) + \mathcal{S}_0]/\hat{s}^*_{K^0}$. Using relations (24), (25), for $a > \alpha > 1/2$, we obtain

$$n(\hat{s}^*_K - \hat{s}^*_{K^0})/\hat{s}^*_{K^0} > C[O_{I\!P}(n^a) - O_{I\!P}(n^\alpha)] = O_{I\!P}(n^a). \qquad (29)$$

Using the relations (27), (29), the fact that $B_n = o(n^a)$ and since the function G is increasing in K, we obtain that for $K < K^0$ we have $B(K) - B(K^0) > O_{I\!P}(n^a) - o_{I\!P}(n^a) \to \infty$. Thus, the relation (26) follows.

We finally consider the case $K > K^0$, cases wherein, given the definition of S^* and of the change-points estimators, we have

$$\begin{aligned}\mathcal{S}_0 \geq S^*(l^0_1, \cdots, l^0_{K^0}) &\geq S^*(\hat{l}^*_{1,K^0}, \cdots, \hat{l}^*_{K^0,K^0}) \geq S^*(\hat{l}^*_{1,K}, \cdots, \hat{l}^*_{K,K}) \\ &\geq S^*(\hat{l}^*_{1,K}, \cdots, \hat{l}^*_{K,K}, l^0_1, \cdots, l^0_{K^0})\end{aligned} \qquad (30)$$

Then, by similar calculations as for the inequality (18) of Theorem 3.2, we have that for the last term of (30):

$$S^*(\hat{l}^*_{1,K}, \cdots, \hat{l}^*_{K,K}, l^0_1, \cdots, l^0_{K^0}) - \mathcal{S}_0 > O_{I\!P}(n^\nu), \qquad \text{for } \nu < 1/4.$$

Thus $0 \geq \hat{s}^*_{K^0} - \hat{s}^*_K = O_{I\!P}(n^{\nu-1})$. Since $G(K,p)$ increases in K, together $n^\alpha \ll B_n \ll n^a$ and the relation (25), we have that

$$n \log \hat{s}^*_{K^0} - n \log \hat{s}^*_K = n(\hat{s}^*_{K^0} - \hat{s}^*_K)(1 + o_{I\!P}(1)) = O_{I\!P}(n^\nu).$$

In fact, it's the penalty that takes over in this case $K > K^0$. In this circumstance, we have $G(K,p) \geq G(K^0,p)$ and then $B_n[G(K,p) - G(K^0,p)] \geq CB_n > O(n^{1/2})$. Hence, we have for the difference between the values of the two criteria $B(K) - B(K^0) \gg -O_{I\!P}(n^\nu) + O(n^{1/2}) = O(n^{1/2}) \to \infty$ as $n \to \infty$. Then, for large enough n, $n \log \hat{s}^*_K + G(K,p) > n \log \hat{s}^*_{K^0} + G(K^0,p)$ implying

$$I\!P[\hat{K}_n > K^0] \to 0, \qquad \text{as } n \to \infty. \qquad (31)$$

The Theorem follows from relations (26) and (31). ■

6.2 Lemma proofs

Proof of Lemma 2.1 Consider the notations $\tilde{\boldsymbol{\phi}} \equiv (b - b^0, \boldsymbol{\phi} - \boldsymbol{\phi}^{\mathbf{0}})$, $w_i = (1, \mathbf{X}_i^t)$ and afterward we apply Lemma 8 of [17]. ■

Proof of Lemma 2.7

Case I: there exists a deterministic sequence (d_n) that tends to infinity, as $n \to \infty$, such that $(k-l) \geq d_n$.

By the definition of $\hat{\boldsymbol{\omega}}_{(l;k)} = |\hat{\boldsymbol{\phi}}_{(l;k)}|^{-g}$, if there exists a component j of the quantile estimator such that $\hat{\phi}_{(l;k),j} \to 0$ (or a subsequence), then, since $I\!P[\hat{\phi}_{(l;k),j} = 0] = 0$ and $\hat{\omega}_{(l;k),j} = |\hat{\phi}_{(l,k),j}|^{-g}$, using the asymptotic normality of $\hat{\boldsymbol{\phi}}_{(l;k)}$ we have that $\hat{\phi}_{(l,k),j} = O_{I\!P}(k-l)^{-1/2} = O_{I\!P}(n^{-1/2})$. Thus, $\lambda_{(l;k)}\hat{\omega}_{(l;k),j} = o(n^{1/2})O_{I\!P}(n^{g/2}) = o_{I\!P}(n^{(1+g)/2})$. If the j-th component is such that $\hat{\phi}_{(l;k),j} \geq c > 0$ then $\lambda_{(l;k)}\hat{\omega}_{(l;k),j} = o_{I\!P}(n^{1/2})$.

Case II. If $(k-l)$ does not converge with n to infinity, then, we can extract a bounded subsequence. Let us suppose that $0 \leq l < k \leq c$. Since $I\!P[\hat{\phi}_{(l;k),j} = 0] = 0$, we have in this case $\lambda_{(l;k)}\hat{\omega}_{(l;k),j} = o_{I\!P}(n^{1/2})$. ■

Proof of Lemma 2.8 Let us denote the difference $k-l$ by d_n. Taking into account Lemma 2.7, then Lemma 2.8 is proven if an equivalent result for $R_i^{(\tau)}$ is showed:

$$\sup_{0 \leq l < k \leq n} \left| \inf_{b, \boldsymbol{\phi}} \sum_{i=l+1}^{k} R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) \right| = O_{I\!P}(n^{\alpha}), \qquad \alpha > 1/2. \tag{32}$$

Let us consider the random processes

$$H_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) \equiv R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) + [\tau 1\!\!1_{\varepsilon_i > b^0} - (1-\tau)1\!\!1_{\varepsilon_i \leq b^0}]\mathbf{X}_i^t(\boldsymbol{\phi} - \boldsymbol{\phi}^{\mathbf{0}}).$$

Obviously $I\!E[\tau 1\!\!1_{\varepsilon_i > b^0} - (1-\tau)1\!\!1_{\varepsilon_i \leq b^0}] = (1-\tau)I\!P[\varepsilon_i \leq b^0] - \tau I\!P[\varepsilon_i > b^0] = 0$.

On the other hand, we have, for all $b_1, b_2 \in \mathcal{B}$ and $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2 \in \Gamma$, that $H_i^{(\tau)}(b_1, \boldsymbol{\phi}_1; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - H_i^{(\tau)}(b_2, \boldsymbol{\phi}_2; b^0, \boldsymbol{\phi}^{\mathbf{0}}) = \rho_\tau(\varepsilon_i - b_1 - \mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}^{\mathbf{0}})) - \rho_\tau(\varepsilon_i - b_2 - \mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}^{\mathbf{0}})) + [\tau 1\!\!1_{\varepsilon_i > b^0} - (1-\tau)1\!\!1_{\varepsilon_i \leq b^0}]\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)$. But, generally, for all $r_1, r_2 \in \mathbb{R}$ we have that $|\rho_\tau(r_1) - \rho_\tau(r_2)| < |r_1 - r_2|$. Thus

$$|\rho_\tau(\varepsilon_i - b_1 - \mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}^{\mathbf{0}})) - \rho_\tau(\varepsilon_i - b_2 - \mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}^{\mathbf{0}}))| < |\mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2) + b_1 - b_2|.$$

If the parameters $b_1, b_2, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2$ are such that $\|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2\|_2 \leq Cn^{-1/2}$, $|b_1 - b_2| \leq Cn^{-1/2}$, since $I\!E[\tau 1\!\!1_{\varepsilon_i > b^0} - (1-\tau)1\!\!1_{\varepsilon_i \leq b^0}] = 0$, we have, using assumption (A1), that

$$\begin{array}{l} \sum_{i=1}^n \left\{ \rho_\tau(\varepsilon_i - b_1 - \mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}^{\mathbf{0}})) - \rho_\tau(\varepsilon_i - b_2 - \mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}^{\mathbf{0}})) \right. \\ \quad \left. - I\!E[\rho_\tau(\varepsilon_i - b_1 - \mathbf{X}_i^t(\boldsymbol{\phi}_1 - \boldsymbol{\phi}^{\mathbf{0}}))] + I\!E[\rho_\tau(\varepsilon_i - b_2 - \mathbf{X}_i^t(\boldsymbol{\phi}_2 - \boldsymbol{\phi}^{\mathbf{0}}))] \right\} \\ \quad \leq C \sum_{i=1}^n |b_1 - b_2| + C \sum_{i=1}^n \|\mathbf{X}_i\|_2 \cdot \|\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1\|_2 = O_{I\!P}(n^{1/2}). \end{array} \tag{33}$$

On the other hand, by Proposition 2.6, we have that $I\!E[R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})]\geq 0$, which means, taking into account the fact that $R_i^{(\tau)}(b^0,\boldsymbol{\phi^0};b^0,\boldsymbol{\phi^0})=0$, that, for $k-l=d_n\to\infty$, we have

$$0\geq \inf_{b,\boldsymbol{\phi}}\sum_{i=l+1}^{k}R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})\geq \inf_{b,\boldsymbol{\phi}}\sum_{i=l+1}^{k}[R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})-I\!E[R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})]].$$

Thus

$$|\inf_{b,\boldsymbol{\phi}}\sum_{i=l+1}^{k}R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})|\leq \sup_{b,\boldsymbol{\phi}}|\sum_{i=l+1}^{k}[R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})-I\!E[R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})]]|.$$

Then, for $d_n\to\infty$, we have

$$\sup_{0\leq l<k\leq n,k-l=d_n}\left|\inf_{b,\boldsymbol{\phi}}\sum_{i=l+1}^{k}R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})\right|\leq 2\sup_{d_n\leq k\leq n}\zeta_k, \tag{34}$$

with $\zeta_k\equiv\sup_{b,\boldsymbol{\phi}}|\sum_{i=1}^{k}[R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})-I\!E[R_i^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})]]|$. Thus, by Proposition 2.6, $\{\zeta_k,\mathcal{F}_k\}$ is a submartingale, where $\mathcal{F}_k\equiv\sigma-field\{\varepsilon_1,\cdots,\varepsilon_k\}$. Which means, by Doob's inequality, for $\alpha>1/2$, that:

$$I\!P\left[\sup_{d_n\leq k\leq n}\zeta_k>n^\alpha\right]\leq I\!P\left[\sup_{1\leq k\leq n}\zeta_k>n^\alpha\right]\leq n^{-\alpha m}C_m I\!E[\zeta_n^m], \tag{35}$$

for some $C_m>0$, and $m>1$ determined later.
Now divide the parameter set $\mathcal{B}\times\Gamma$ in $n^{(1+p)/2}$ cells, such that the diameter of each cell is less than $n^{-1/2}$. Thus, for $(b_1,\boldsymbol{\phi}_1),(b_2,\boldsymbol{\phi}_2)$ in the same cell, we have:

$$\begin{aligned}\textstyle\sum_{i=1}^n&\Big[R_i^{(\tau)}(b_1,\boldsymbol{\phi}_1;b^0,\boldsymbol{\phi^0})-I\!E[R_i^{(\tau)}(b_1,\boldsymbol{\phi}_1;b^0,\boldsymbol{\phi^0})]-R_i^{(\tau)}(b_2,\boldsymbol{\phi}_2;b^0,\boldsymbol{\phi^0})+I\!E[R_i^{(\tau)}(b_2,\boldsymbol{\phi}_2;b^0,\boldsymbol{\phi^0})]\Big]\\ &\leq\Big|\textstyle\sum_{i=1}^n\Big[H_i^{(\tau)}(b_1,\boldsymbol{\phi}_1;b^0,\boldsymbol{\phi^0})-I\!E[H_i^{(\tau)}(b_1,\boldsymbol{\phi}_1;b^0,\boldsymbol{\phi^0})]-H_i^{(\tau)}(b_2,\boldsymbol{\phi}_2;b^0,\boldsymbol{\phi^0})\\ &\qquad+I\!E[H_i^{(\tau)}(b_2,\boldsymbol{\phi}_2;b^0,\boldsymbol{\phi^0})]\Big]\Big|+\big|\textstyle\sum_{i=1}^n D_i\mathbf{X}_i^t(\boldsymbol{\phi}_1-\boldsymbol{\phi}_2)\big|\leq Cn^{1/2},\end{aligned} \tag{36}$$

where the last inequality follows from (33). Let be now $(b_j,\boldsymbol{\phi}_j)$ in the jth cell, $j=1,\cdots,c_pn^{(1+p)/2}$. Then, as in the paper of [17], Lemma 3, we have

$$\zeta_n^m\leq C\sup_j|\sum_{i=1}^n[R_i^{(\tau)}(b_j,\boldsymbol{\phi}_j;b^0,\boldsymbol{\phi^0})-I\!E[R_i^{(\tau)}(b_j,\boldsymbol{\phi}_j;b^0,\boldsymbol{\phi^0})]]|^m+Cn^{m/2}.$$

Since $R_i^{(\tau)}(b_j,\boldsymbol{\phi}_j;b^0,\boldsymbol{\phi^0})-I\!E[R_i^{(\tau)}(b_j,\boldsymbol{\phi}_j;b^0,\boldsymbol{\phi^0})]$ is a bounded martingale difference, for each fixed $j\in\{1,\cdots,c_pn^{(1+p)/2}\}$, we have

$$I\!E[|\sum_{i=1}^n(R_i^{(\tau)}(b_j,\boldsymbol{\phi}_j;b^0,\boldsymbol{\phi^0})-I\!E[R_i^{(\tau)}(b_j,\boldsymbol{\phi}_j;b^0,\boldsymbol{\phi^0})])|^m]\leq Cn^{m/2}.$$

Then $I\!\!E[\zeta_n^m] \leq O_{I\!\!P}(n^{\frac{1+p}{2}})n^{m/2} + O_{I\!\!P}(n^{m/2}) = O_{I\!\!P}(n^{\frac{m+p+1}{2}})$.
Thus, choosing m such that $m > \frac{p+1}{2\alpha+1}$, the right-hand side of (35) converges to 0. Furthermore, taking into account (34), (35) and (36) we obtain the claim (32). ■

Proof of Lemma 2.9 Let us consider the set Ω_n written as an union of subsets $\Omega_n = \bigcup_{j=1}^{C_p n^{(p+1)/2}} \mathcal{C}_j^n$, with C_p a bounded positive constant, and

$$\mathcal{C}_j^n \equiv \{(b, \boldsymbol{\phi}) \in \Omega_n; |b - b'| + \|\boldsymbol{\phi} - \boldsymbol{\phi}'\|_2 \leq c_n n^{-1/2}, \text{ for all } (b', \boldsymbol{\phi}') \in \mathcal{C}_j^n\}.$$

For $(b_1, \boldsymbol{\phi}_1), (b_2, \boldsymbol{\phi}_2) \in \mathcal{C}_j^n$, as in the proof of Lemma 2.8, relation (36), we have that $(nc_n^2)^{-1}|\mathcal{R}_n^{(\tau)}(b_1, \boldsymbol{\phi}_1; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[\mathcal{R}_n^{(\tau)}(b_2, \boldsymbol{\phi}_2; b^0, \boldsymbol{\phi}^{\mathbf{0}})] - \mathcal{R}_n^{(\tau)}(b_2, \boldsymbol{\phi}_2; b^0, \boldsymbol{\phi}^{\mathbf{0}}) + I\!\!E[\mathcal{R}_n^{(\tau)}(b_2, \boldsymbol{\phi}_2; b^0, \boldsymbol{\phi}^{\mathbf{0}})]| \leq C(|b_1 - b_2| - \|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2\|_2)c_n^{-2} \leq Cn^{-1/2}c_n^{-1} \to 0$, as $n \to \infty$. Then, for all $(b_1, \boldsymbol{\phi}_1), (b_2, \boldsymbol{\phi}_2) \in \mathcal{C}_j^n$,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{nc_n^2} \Big| &\mathcal{R}_n^{(\tau)}(b_1, \boldsymbol{\phi}_1; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[\mathcal{R}_n^{(\tau)}(b_2, \boldsymbol{\phi}_2; b^0, \boldsymbol{\phi}^{\mathbf{0}})] - \mathcal{R}_n^{(\tau)}(b_2, \boldsymbol{\phi}_2; b^0, \boldsymbol{\phi}^{\mathbf{0}}) \\ &+ I\!\!E[\mathcal{R}_n^{(\tau)}(b_2, \boldsymbol{\phi}_2; b^0, \boldsymbol{\phi}^{\mathbf{0}})]\Big| = 0. \end{aligned} \tag{37}$$

For $(b_j, \boldsymbol{\phi}_j) \in \mathcal{C}_j^n$, for all $j = 1, \cdots, C_p n^{(1+p)/2}$, we have that the probability

$$\begin{aligned} &I\!\!P[\sup_j |(nc_n^2)^{-1}[\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}})]]| > \epsilon] \\ &\quad \leq \sum_{j=1}^{C_p n^{(1+p)/2}} I\!\!P\left[\left|\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}})]\right| > nc_n^2\epsilon\right]. \end{aligned} \tag{38}$$

But, by assumption (A1), we have that

$$R_i^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[R_i^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}})] \leq C[|b_j - b^0| + \|\boldsymbol{\phi}_j - \boldsymbol{\phi}^{\mathbf{0}}\|_2] \leq Cc_n.$$

Then $Var[R_i^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}})] \leq C^2 c_n^2$ uniformly in b_j and $\boldsymbol{\phi}_j$. We apply Lemma 6.1, for $\beta = Cc_n$, $V = C^2 nc_n^2$, $s = 1/2$, $z = nc_n^2\epsilon$ and we obtain

$$I\!\!P[|\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}})]| > nc_n^2\epsilon] \leq 2\exp(-\epsilon^2 nc_n^2 C). \tag{39}$$

The statement of Lemma 6.1 is given at the beginning of Section 6. Relation (39) and the fact that $nc_n^2/\log n \to \infty$ imply that the right-hand side of (38) is bounded by $2C_p n^{(1+p)/2}\exp(-\epsilon^2 nc_n^2 C)$ which is smaller than $\exp(-\epsilon^2 nc_n^2 C/2)$, for any large enough n.
Then, for all $(b_j, \boldsymbol{\phi}_j) \in \mathcal{C}_j^n$, for all $j = 1, \cdots, C_p n^{(1+p)/2}$, we have

$$\lim_{n\to\infty} I\!\!P\left[\sup_j \left|(nc_n^2)^{-1}[\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[\mathcal{R}_n^{(\tau)}(b_j, \boldsymbol{\phi}_j; b^0, \boldsymbol{\phi}^{\mathbf{0}})]\right]\right| > \epsilon] = 0. \tag{40}$$

The Lemma follows from (37) and (40). ■

Proof of Lemma 2.10 By the proof of Proposition 2.6, using assumption (A2), we have

$$I\!\!E[\mathcal{R}_n^{(\tau)}(b^0 + u_0/\sqrt{n}, \boldsymbol{\phi}^{\mathbf{0}} + \mathbf{u}/\sqrt{n}; b^0, \boldsymbol{\phi}^{\mathbf{0}})] = \frac{1}{2n} f(b^0)(u_0, \mathbf{u}^t) \left[\begin{array}{cc} n & \mathbf{0} \\ 0 & \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t \end{array} \right] (u_0, \mathbf{u}^t)^t (1 + o(1)),$$

with $u_0 \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^p$ in open sets. For a sequence (c_n) converging to zero but with a slower rate than $n^{-1/2}$, we have

$$I\!\!E\left[\int_0^{(u_0 + \mathbf{X}_i^t \mathbf{u}) c_n} [1\!\!1_{\varepsilon_i \leq b^0 + t} - 1\!\!1_{\varepsilon_i \leq b^0}] dt \right] = \int_0^{u_0 + \mathbf{X}_i^t \mathbf{u}} c_n [F(b^0 + c_n v) - F(b^0)] dv.$$

Thus

$$I\!\!E[\mathcal{R}_n^{(\tau)}((b^0 + u_0) c_n, (\boldsymbol{\phi}^0 + \mathbf{u}) c_n; b^0, \boldsymbol{\phi}^{\mathbf{0}})] = c_n \sum_{i=1}^n \int_0^{u_0 + \mathbf{X}_i^t \mathbf{u}} [F(b^0 + c_n v) - F(b^0)] dv$$

$$= c_n^2 \frac{f(b^0)}{2} (u_0, \mathbf{u}^t) \left[\begin{array}{cc} n & \mathbf{0} \\ 0 & \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t \end{array} \right] (u_0, \mathbf{u}^t)^t (1 + o(1)).$$

By Lemma 2.9, using the Borel-Cantelli lemma, we have that for any $\epsilon_2 > 0$

$$\limsup_{n \to \infty} \left(\sup_{(b, \boldsymbol{\phi}) \in \Omega_n} \left| \frac{1}{n c_n^2} \left[\mathcal{R}_n^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) - I\!\!E[\mathcal{R}_n^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})] \right] \right| \right) \leq \epsilon_2, \quad a.s. \tag{41}$$

But, the function $R_i^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}})$ is convex in b and $\boldsymbol{\phi}$, therefore also its sum $\mathcal{R}_n^{(\tau)}$. Since $\mathcal{R}_n^{(\tau)}(b^0, \boldsymbol{\phi}^0; b^0, \boldsymbol{\phi}^{\mathbf{0}}) = 0$, using assumption (A1) and relation (41), by a similar argument as in Remark 4 of [17], we obtain that the infimum of $(b, \boldsymbol{\phi}) \in \Omega_n^c$ is attained on the boundary: $|b - b^0| = c_n$ and $\|\boldsymbol{\phi} - \boldsymbol{\phi}^{\mathbf{0}}\|_2 = c_n$.

Then, let us consider the scalar u_0 and the p-vector $\mathbf{u}$, each of euclidean norm 1: $|u_0| = 1$, $\|\mathbf{u}\|_2 = 1$. Using assumption (A1), we have that:

$$I\!\!E\left[\mathcal{R}_n^{(\tau)}((b^0 + u_0) c_n, (\boldsymbol{\phi}^0 + \mathbf{u}) c_n; b^0, \boldsymbol{\phi}^{\mathbf{0}})] dt \right] = n c_n^2 \frac{f(b^0)}{2} (C + o(1)), \tag{42}$$

with $C > 0$.
Then, considering $\epsilon_2 = C f(b^0)/4$ in (41), taking into account the relation (42), we obtain that there exists $\epsilon_3 > 0$ (we can consider by example $\epsilon_3 = \epsilon_2$) such that, with the probability 1:

$$\liminf_{n \to \infty} \left(\inf_{(b, \boldsymbol{\phi}) \in \Omega_n^c} \frac{1}{n c_n^2} \mathcal{R}_n^{(\tau)}(b, \boldsymbol{\phi}; b^0, \boldsymbol{\phi}^{\mathbf{0}}) \right) \geq \epsilon_3, \tag{43}$$

knowing that $\inf_{(b,\phi)\in\Omega_n^c}$ is in fact the infimum on the boundary of the set Ω_n^c.
To finish the lemma, we take into account that:

$$\inf_{(b,\phi)\in\Omega_n^c} \frac{1}{nc_n^2}\sum_{i=1}^n R_i^{(\tau,\lambda)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0})$$

$$\geq \inf_{(b,\phi)\in\Omega_n^c} \frac{1}{nc_n^2}\mathcal{R}_n^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0}) - \frac{\lambda_n}{nc_n^2}\sup_{(b,\phi)\in\Omega_n^c}(|\hat{\boldsymbol{\omega}}^t_{(0;n)}|\cdot\big||\boldsymbol{\phi}|-|\boldsymbol{\phi^0}|\big|)$$

$$\geq \inf_{(b,\phi)\in\Omega_n^c} \frac{1}{nc_n^2}\mathcal{R}_n^{(\tau)}(b,\boldsymbol{\phi};b^0,\boldsymbol{\phi^0}) - \frac{\lambda_n}{nc_n^2}\|\hat{\boldsymbol{\omega}}^t_{(0;n)}\|_1 C_+, \tag{44}$$

with C_+ a positive constant. The first term of the right-hand side of (44) is greater than ϵ_3 by the relation (43). Since $\lambda_n(nc_n^2)^{-1}\to\infty$ as $n\to\infty$, Γ a compact set, thus the second term of the right-hand side of (44) converges to 0, and then it is smaller than $\epsilon_3/2$ for n large enough. Thus, the Lemma is proved by taking $\epsilon_1=\epsilon_3/2$.

Let us note that the adaptive weight $\hat{\boldsymbol{\omega}}^t_{(0;n)}$ is calculated on all observations, without imposing the constraint that $\|\boldsymbol{\phi}-\boldsymbol{\phi^0}\|_2\geq c_n$. ■

Proof of Lemma 3.1 By similar calculations to the proof of Proposition 2.6, we obtain that

$$I\!\!E[R_i^{(\tau)}(b_r^0+\mu,\boldsymbol{\phi}_r^0;b_r^0,\boldsymbol{\phi}_r^0)] = \int_{-|\mu|}^0 [x+\mu]dF(x+b_r^0)$$

is a positive function for any $\mu\in\mathcal{B}$ and a increasing function in $|\mu|$, with a single zero for $\mu=0$. By Lemma 2.1, there exists at least $(l_r^0-l_r)\epsilon_0$ observations, for some $\epsilon_0>0$, and some $\delta>0$, such that $|b_{r+1}^0-b_r^0|+|\mathbf{X}_i^t(\boldsymbol{\phi}_{r+1}^0)-\boldsymbol{\phi}_r^0)|>\delta$. Then

$$I\!\!E\left[\sum_{i=l_r+1}^{l_r^0} R_i^{(\tau)}(b_{r+1}^0,\boldsymbol{\phi}_{r+1}^0;b_r^0,\boldsymbol{\phi}_r^0)\right] \geq (l_r^0-l_r)\epsilon_0\int_{-\delta}^0[x+\delta]dF(x+b_r^0).$$

By Lemma 2.9, for $\boldsymbol{\phi^0}=\boldsymbol{\phi}_r^0$, $\boldsymbol{\phi}=\boldsymbol{\phi}_{r+1}^0$ and $c_n=\max(|b_{r+1}^0-b_r^0|,\|\boldsymbol{\phi}_{r+1}^0)-\boldsymbol{\phi}_r^0\|_2)$, we have that for all $\epsilon>0$ such that

$$I\!\!P[|\sum_{i=l_r+1}^{l_r^0} R_i^{(\tau)}(b_{r+1}^0,\boldsymbol{\phi}_{r+1}^0;b_r^0,\boldsymbol{\phi}_r^0) - I\!\!E[\sum_{i=l_r+1}^{l_r^0} R_i^{(\tau)}(b_{r+1}^0,\boldsymbol{\phi}_{r+1}^0;b_r^0,\boldsymbol{\phi}_r^0)]|>\epsilon(l_r^0-l_r)]$$

$\leq \exp(-C\epsilon(l_r^0-l_r))$. We take $\epsilon=2^{-1}\epsilon_0\int_{-\delta}^0[x+\delta]dF(x+b_r^0)$ and $\eta=2^{-1}\epsilon_0\int_{-\delta}^0[x+\delta]dF(x+b_r^0)$. Then the Lemma follows. ■

References

[1] Koenker R., Quantile Regression, Cambridge University Press; 2005.

[2] Babu G.J., Strong representations for LAD estimators in linear models. Probability Theory and Related Fields. 1989;83:547–558.

[3] Tibshirani R., Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society, Ser. B. 1996;58:267–288.

[4] Zou H., The adaptive Lasso and its oracle properties. Journal of the American Statistical Association. 2006;101:1418–1428.

[5] Xu J., Ying Z., Simultaneous estimation and variable selection in median regression using Lasso-type penalty. Annals of the Institute of Statistical Mathematics. 2010;62:487–514.

[6] Wang L., L_1 penalized LAD estimator for high dimensional linear regression. Journal of Multivariate Analysis. 2013;120:135–151.

[7] Gao X., Huang J., Asymptotic analysis of high-dimensional LAD regression with LASSO. Statistica Sinica. 2010;20:1485–1506.

[8] Wu Y., Liu Y., Variable selection in quantile regression. Statistica Sinica. 2009;19:801–817.

[9] He X., Wang L., Hong H.G., Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. The Annals of Statistics. 2013;41:342–369.

[10] Zou H., Yuan M., Composite quantile regression and the oracle model selection theory. The Annals of Statistics. 2008;36:1108–1126.

[11] Guo J., Tang M., Tian M., Zhu K., Variable selection in high-dimensional partially linear additive models for composite quantile regression. Computational Statistics and Data Analysis. 2013;65:56–67.

[12] Ciuperca G., Model selection by LASSO methods in a change-point model, Statistical Papers. 2014;55:349–374.

[13] Ciuperca G., Quantile regression in high-dimension with breaking, Journal of Statistical Theory and Applications. 2013;12:288–305.

[14] Lee S., Seo M.H., Shin Y., The LASSO for high-dimensional regression with a possible change-point. 2012;arXiv:1209:4875v2.

[15] Belloni A., Chernozhukov V., l_1-Penalized quantile regression in high-dimensional sparse models. The Annals of Statistics. 2011;39:82–130.

[16] Zheng Q., Gallagher C., Kulasekera K.B., Adaptive penalized quantile regression for high dimensional data. Journal of Statistical Planning Inference. 2013;143:1029–1038.

[17] Bai J., Estimation of multiple-regime regressions with least absolute deviation. Journal of Statistical Planning Inference. 1998;74:103–134.

[18] Ciuperca G., Penalized least absolute deviations estimation for nonlinear model with change-points. Statistical Papers. 2011;52:371–390.

[19] Bai J., Perron P., Estimating and testing linear models with multiple structural changes. Econometrica. 998;66:47–78.

[20] Yao Y.C., Estimating the number of change-points via Schwarz's criterion. Statistics and Probability Letters. 1988;6:181–189.

[21] Osorio F., Galea M., Detection of a change-point in student-t linear regression models. Statistical Papers. 2005;45:31–48.

[22] Wu Y., Simultaneous change point analysis and variable selection in a regression problem. Journal of Multivariate Analysis. 2008;99:2154–2171.

[23] Nosek K., Schwarz information criterion based tests for a change-point in regression models. Statistical Papers. 2010;51:915–929.

[24] Liu Y., Zou C., Zhang R., Empirical likelihood ratio test for a change-point in linear regression model. Communications in Statistics-Theory and Methods. 2008;37:2551–2563.

[25] Qu Z., Perron P., Estimating and testing structural changes in multivariate regressions. Econometrica. 2007;75:459–502.

[26] Ning W., Pailden J., Gupta A., Empirical likelihood ratio test for the epidemic change model. Journal of Data Science. 2012;10:107–127.