Ann Inst Stat Math (2011) 63:717–743
DOI 10.1007/s10463-009-0256-y

Estimating nonlinear regression with and without change-points by the LAD method

Gabriela Ciuperca

Received: 4 February 2008 / Revised: 27 January 2009 / Published online: 28 July 2009

Abstract The paper considers the least absolute deviations estimator in a nonlinear parametric regression. The interest of the LAD method is its robustness with respect to other traditional methods when the errors of model contain outliers. First, in the absence of change-points, the convergence rate of estimated parameters is found. For a model with change-points, in the case when the number of jumps is known, the convergence rate and the asymptotic distribution of estimators are obtained. Particularly, it is shown that the change-points estimator converges weakly to the minimizer of given random process. Next, when the number of jumps is unknown, its consistent estimator is proposed, via the modified Schwarz criterion.

Keywords Asymptotic properties · Change-point · LAD estimator · Parametric model

1 Introduction

We consider the model:

$$Y_i = g_\theta(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

for the step-function with K ($K \geq 0$) change-points:

$$g_\theta(x_i) = h_{\beta_1}(x_i)\mathbb{1}_{i \leq l_1} + h_{\beta_2}(x_i)\mathbb{1}_{l_1 < i \leq l_2} + \cdots + h_{\beta_{K+1}}(x_i)\mathbb{1}_{l_K < i}.$$

G. Ciuperca (✉)
Institut Camille Jordan, Université de Lyon, CNRS, UMR 5208, Université Lyon 1,
Bat. Braconnier, 43, blvd du 11 novembre 1918, 69622 Villeurbanne Cedex, France
e-mail: Gabriela.Ciuperca@univ-lyon1.fr

Let us denote the regression parameters $\theta_1 = (\beta_0, \beta_1, \ldots, \beta_K)$ and the change-points $\theta_2 = (l_1, \ldots, l_K)$ with $l_1 < l_2 < \cdots < l_K$. We set $\theta = (\theta_1, \theta_2)$. If $K = 0$, the model is without change-point.

The change-point problem arose from statistical quality control, continued by many other important applications in seismic signal processing, analysis of electrocardiograms, finance, epidemiology, etc. The statistics literature contains a vast amount of work on issues related to structural change, most of it specifically designed for the case of a single break. For recent reviews, we refer readers to Carlstein et al. (1994), Csorgo and Horvath (1997) and the references therein. Recent developments include Andrews et al. (1996) who consider optimal tests in the linear model with known variance. Garcia and Perron (1996) study the Wald test for two changes in a dynamic time series. Horváth et al. (1997) consider an estimator for the time of change in a linear model when the regression coefficients and the variance may change. Liu et al. (1997) consider multiple structural changes in a linear model estimated by least-squares and propose an information criterion for the selection of the number of changes. Lombard (1987) and Mia and Zhao (1988) propose some procedures based on a rank statistics to test for one or more change-points. Horváth et al. (2004) detect a structural change in a linear model based on weighted CUSUM of residuals. Hušková et al. (2007) study the problem of detecting change in the parameters of an autoregressive time series by using various test statistics. For testing the stationarity of a Gaussian process Epps (1988) proposes Chi-squared statistics.

To estimate the number of jumps in the mean of an independent normal sequence Yao (1988) uses the Schwarz criterion. Using this criterion, Serbinowska (1996) and Kühn (2001) proved the consistency of estimators for the number of changes in the case of the independent observations. So, for independent binomial random variables, the weak consistency of an estimator of the unknown number of changes in the parameters is proved by Serbinowska (1996). Kühn (2001) studies an estimator for the number of the change-points in the drift of a stochastic process by establishing a similar criterion that of Schwarz. Lavielle (1999) studies the a penalized contrast function for detection of changes in a sequence of dependent variables. Bai and Perron (1998) study the problem of testing multiple change-points in a linear model.

It is well known that one outlier may cause a large error in a least squares (LS) estimator or in a least absolute deviations (LAD) estimator. This occurs in the case of fatter tail distributions of the error term (see Hitomi and Kagihara 2001). On the other hand, as Bai (1998) and Kim and Choi (1995) indicate it, for heavy tailed distributions the LAD estimator is more efficient than LS estimators.

The LAD estimator of the parameter θ, for the model (1), is by definition:

$$\hat{\theta}_n = \arg\min_{\theta} \sum_{i=1}^{n} |Y_i - g_\theta(x_i)|.$$

The study of the properties of the LAD estimator is more difficult than for the LS estimator because of the nondifferentiability of the criterion function.

The LAD regression is also known by several other names, including L_1-norm regression, minimum absolute deviation regression, least absolute value regression and minimum sum of absolute errors regression.

Concerning the LAD estimator in a model without breaking points we can cite the following papers: Wu (1988) gives the conditions under which the estimator in linear regression is strongly consistent, Pollard (1991) proves that the LAD estimators in a linear regression is asymptotically normal and Babu (1989) proves that the convergence rate is $n^{-1/2}$, Oberhofer (1982) shows conditions for the consistency of the LAD estimator in nolinear regression. Richardson and Bhattacharyya (1987) extended Oberhofer's result to a more general parameter space. The LAD estimator in a dynamic nonlinear model with neither independent nor identically distributed errors is considered by Weiss (1991). The estimator is shown to be consistent and asymptotically normal. See also Dielman (2005) for a review of research on LAD regression.

For change-points estimation, to the author's knowledge, the LAD method has been only analyzed for linear model. At first, Bai (1995) studies the method for a single change-point and after, Bai (1998) extends the results for multiple-regime regression. In the case of a single change-point, each of the two regimes has one fixed and known boundary. For multiple breaks, each middle regime has boundaries completely unknown. In the linear case, many proofs are based on the convexity of the regression function $h_\beta(x) = x\beta$ with respect to the parameter vector β. That is, the extreme value of a convex function is attained on the boundary. Also, in the case of multiple change-points, the problem is much more intricate when the number of breaks is unknown.

This paper considers the estimation in a parametric nonlinear regression by the LAD method. We sometimes use the terms jump or break of instead of change-point. The study of this method was motivated by wishes to find the properties of this estimator, particularly interesting in a change-point model with outliers, the latter being able to create problems in the detection of the jumps. First, we give same properties of the LAD estimator in a model without change-points. Mainly, we improve the convergence rate of estimators in comparison to that of the linear regression (Babu 1989). Later, in a model with breaking, if the number of change-points is fixed, we study the properties of the LAD estimators. The convergence rate of the change-point estimator are derived. The asymptotic distribution of these estimators are also obtained. Next, we prove the weakly convergence of the change-points estimator to the location of the maximum of certain random processes. If the number of breaks is unknown the problem of its estimate arises. We propose, via Schwarz criterion, a consistently estimator of the number number of breaks. One illustrates by simulations that this method allows to well detect the jumps.

The paper is organized as follows. In Sect. 2 we derive results in the absence of change-points. Section 3 contains results about the rate of convergence and the asymptotic distribution of the estimators, for a model with known number of change-points. Section 4 gives a method to determine the number of breaks. In Sect. 5 we give the proofs of theorems. Finally, Sect. 6 contains some lemmas which are useful to prove the main results.

2 Results in the absence of change-points

In this section we consider a regression model (without change-points):

$$Y_i = h_\beta(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2}$$

under the hypothesis that ε_i are independent identically distributed (i.i.d.) random variables and $\beta \in \Gamma \subset I\!R^p$, with Γ a compact set. The analytical expression of the function $h_\beta(x)$ is known. In this model, Y_i is the univariate dependent random variable while the variable x_i is deterministic. For simplicity, we suppose that the x_i's are nonrandom, although the results will, typically hold for random x_i's independent of the ε_i's and if x_i independent of x_j for $i \neq j$. The set $\bar{I\!R} = I\!R \cup \{-\infty, \infty\}$ is compact under the metric $d(x, y) = |\arctan x - \arctan y|$. Thus, we can consider that the variable x_i takes values in a compact Υ of $I\!R$. The regression parameters β are unknown.

With regard to the random variable ε we make the assumptions:

- Let F be the distribution function of ε, f its density, which satisfy the conditions: $f(0) > 0$, $F(0) = 1/2$
- for one $c > 0$ we have, for all y in a neighborhood of 0: $|f(y) - f(0)| \leq c|y|^{1/2}$
- there exists $c^{(0)} > 0$ such that: $|F(y) - F(x)| \leq c^{(0)}|y - x|$, $\forall x, y \in I\!R$.

Then, $I\!E[\text{sign}(\varepsilon)] = 0$.

The purpose is to estimate the unknown regression parameters when n observations $(Y_i, x_i)_{1 \leq i \leq n}$ are available by the least absolute deviations principle. By definition, the LAD estimator is:

$$\hat{\beta}_n = \arg\min_{\beta \in \Gamma} \sum_{i=1}^{n} |Y_i - h_\beta(x_i)|. \tag{3}$$

The inference for the regression parameters based on L_1-estimation is greatly more complicated than the estimation based on smoother objective function.

For a vector, let us denote $\|.\|$ the Euclidean norm and for a matrix $A = (a_{ij})$, $\|A\| = \sum_{ij} |a_{ij}|$.

Let β^0 denote the true value (unknown) of the parameter β. We suppose that β^0 is an inner point of the set Γ. The function $h : \Upsilon \times \Gamma \to I\!R$ satisfy the conditions:

(H1) for all $x \in \Upsilon$, the function $h_\beta(x)$ is twice differentiable in β

(H2) for all $x \in \Upsilon$, $h_\beta(x)$, $\|\partial h_\beta(x)/\partial\beta\|$, $\|\partial^2 h(x, \beta)/\partial\beta^2\|$ are bounded in a neighborhood of β^0.

Moreover, we suppose that the design points (x_i) satisfy the conditions:

(H3) for n large enough, there exists $c^{(1)} > 0$ such that:

$$\frac{1}{n} \sum_{i=1}^{n} \sup_{\beta \in \Gamma} \left\| \frac{\partial h_\beta(x_i)}{\partial \beta} \right\| \leq c^{(1)} < \infty. \tag{4}$$

(H4) The following limit exists:

$$M := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left\| \frac{\partial h_{\beta^0}(x_i)}{\partial\beta} \right\|^2 .$$

Remark 1 Obviously (H1), (H2) are necessary for the Taylor expansion. The function $h_\beta(x)$ and its derivatives must be bounded only in a neighborhood of the true value β^0. The assumptions (H3) and (H4) are necessary to find the convergence rate of the LAD estimator.

Let us consider the function: $\nu_{(\beta_1,\beta_2)}(x_i) = h_{\beta_1}(x_i) - h_{\beta_2}(x_i)$. In order to study the function to be minimized: $\sum_{i=1}^n |Y_i - h_\beta(x_i)| = \sum_{i=1}^n |\varepsilon_i + h_{\beta^0}(x_i) - h_\beta(x_i)|$, let us consider the random process: $\eta_i(\beta) = |\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)| - |\varepsilon_i|$ and the centered process: $\xi_i(\beta) = \eta_i(\beta) - I\!E[\eta_i(\beta)]$.

In the following, we denote by C a generic positive finite constant not depending on n which may take different values in different formulae or even in different parts of the same formula.

The estimator $\hat{\beta}_n$ is consistent if following condition is satisfied (Oberhofer 1982): for $\Gamma_0 \subset \Gamma$ a closed set not containing β^0, then there exist numbers $\epsilon > 0$ and $n_0 \in I\!N$ such that for all $n \geq n_0$:

$$\inf_{\beta\in\Gamma_0} \frac{1}{n}\sum_{i=1}^{n} |\nu_{(\beta,\beta^0)}(x_i)| \min\left\{ F\left(\frac{|\nu_{(\beta,\beta^0)}(x_i)|}{2}\right) - \frac{1}{2}; \frac{1}{2} - F\left(-\frac{|\nu_{(\beta,\beta^0)}(x_i)|}{2}\right)\right\} \geq \epsilon. \tag{5}$$

Since $F(0) = 1/2$, the above condition implies that the functions $h_\beta(x)$ satisfy following statement: if $\|\beta - \beta^0\| > c$ for a positive c, then there exists a number $a > 0$ such that: $|h_\beta(x) - h_{\beta^0}(x)| > a$ almost everywhere on Υ.

It is well known from the literature that the LAD estimator of β is asymptotically normal, with asymptotic covariance matrix depending on the errors through the heights of their density functions at their median (Weiss 1991). On the other hand, to the author's knowledge, there are no known results concerning the rate of convergence of LAD estimators in a nonlinear regression. The convergence rate of $\hat{\beta}_n$ to β^0 is established in the following theorem.

Theorem 1 *Under the assumptions* (H1)–(H4) *and the condition* (5)*, for all monotone positive sequence* (v_n) *such that:*

$$v_n \to 0, \;\; n v_n^2 \to \infty \;\; \textit{for} \;\; n \to \infty, \tag{6}$$

we have: $\|\hat{\beta}_n - \beta^0\| = O_{I\!P}(v_n)$.

The result of the Theorem 1 is stronger than that of Babu (1989) for the linear regression which have obtained the rate of convergence $v_n = (\log n/n)^{1/2}$. The following theorem shows that the main result obtained by Babu (1989) for the linear case can be generalized for a nonlinear model.

Theorem 2 *Under the assumptions* (H1)–(H4), *for any monotone positive sequence* (v_n) *satisfying* (6), *we have:*

$$2(\hat{\beta}_n - \beta^0) \int_{-\infty}^{\infty} \frac{\partial h_{\beta^0}(x)}{\partial \beta} f(h_{\beta^0}(x))\mathrm{d}x$$
$$= \frac{1}{n}\sum_{i=1}^{n}[\mathrm{sign}(\varepsilon_i) - \mathrm{sign}(\varepsilon_i - \nu_{(\hat{\beta}_n,\beta^0)}(x_i))] + O(v_n).$$

3 Multiple change-points with K fixed

This section considers issues related to multiple structural changes, occurring at unknown times, in the nonlinear regression. We give the convergence rate and the limiting distribution of the LAD estimator under the condition that the number of change-points is known.

A concrete example of application of this type of model can be found in Pauler and Finkelstein (2002) on the recurrence in prostate cancer.

For the model (1) we consider a step-function $g_\theta(x)$ with $K (K \geq 1)$ fixed change-points.

We do not impose the restriction that the function g is continuous at the turning points.

For all $r = 1, \ldots, K, K+1$ we suppose $\beta_k \in \Gamma \subset I\!R^p$ with Γ compact and $\theta_2 \in I\!R^K$, therefore $\theta \in \Omega = \Gamma^{K+1} \times \Upsilon^K$. Let $\theta^0 = (\theta_1^0, \theta_2^0)$ be the unknown true value for θ.

To begin with we shall state the needed assumptions:

- **(A1)** We impose the condition that the change-points are sufficiently far apart: for some $u \geq 3/4$ there exists $c > 0$ such that $l^0_{r+1} - l^0_r \geq cn^u$, $\forall r = 1, \ldots, K$ with $l^0_0 = 1$ and $l^0_{K+1} = n$.
- **(A2)** Moreover, we suppose that we have two different models to the left and to the right of the every break point l^0_r: $\beta^0_r \neq \beta^0_{r+1}$, $\forall r = 1, \ldots, K$

  For the function $h_\beta(x)$ we assume (H1), (H3) of the Sect. 2 and:
- **(H2′)** for all $x \in \Upsilon$, $h_\beta(x)$, $\|\partial h_\beta(x)/\partial\beta\|$, $\|\partial^2 h_\beta(x)/\partial\beta^2\|$ are bounded in a neighborhood of β^0_r, for every $r = 1, \ldots, K$.
- **(H4′)** For every $r = 1, \ldots, K$, the following limits exist:

$$M_{r,1} := \lim_{m\to\infty} \frac{1}{m}\sum_{i=l^0_r - m}^{l^0_r} \left\| \frac{\partial h_{\beta^0_{r-1}}(x_i)}{\partial \beta}\right\|^2, \quad M_{r,2} := \lim_{m\to\infty} \frac{1}{m}\sum_{i=l^0_r + 1}^{l^0_r + m} \left\| \frac{\partial h_{\beta^0_r}(x_i)}{\partial \beta}\right\|^2,$$

  and the supplementary assumption:
- **(H5)** for n large enough, there exists $c^{(2)} > 0$ such that:

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{\beta\in\Gamma} \left\| \frac{\partial h_\beta(X_i)}{\partial\beta}\right\|^2 \leq c^{(2)} < \infty. \tag{7}$$

Remark 2 The assumption (H5) is necessary to control the variance of $n^{-1}\sum_{i=1}^n \eta_i(\beta)$. If β is not close to β^0 then the variance of the sum of absolute deviations of residuals has the order n. This assumption is used in the Lemma 6, which will allow to improve the convergence rate of the change-point estimator and to find its asymptotic distribution.

The relation (5) becomes: for every $r = 1, \ldots, K$, if $\Gamma_{0,r} \subset \Gamma$ is a closed set not containing β_r^0 then there exists numbers ϵ_r and $m_{0,r} \in I\!N$ such that for all $m \geq m_{0,r}$:

$$\inf_{\beta\in\Gamma_{0,r}} \frac{1}{m} \sum_{i=l_r^0+1}^{l_r^0+m} |\nu_{(\beta,\beta_r^0)}(x_i)| \min\left\{F\left(\frac{|\nu_{(\beta,\beta_r^0)}(x_i)|}{2}\right)-\frac{1}{2}; \frac{1}{2}-F\left(-\frac{|\nu_{(\beta,\beta_r^0)}(x_i)|}{2}\right)\right\} \geq \epsilon_r,$$

and

$$\inf_{\beta\in\Gamma_{0,r}} \frac{1}{m} \sum_{i=l_{r+1}^0-m}^{l_{r+1}^0} |\nu_{(\beta,\beta_r^0)}(x_i)| \min\left\{F\left(\frac{|\nu_{(\beta,\beta_r^0)}(x_i)|}{2}\right)-\frac{1}{2}; \frac{1}{2}-F\left(-\frac{|\nu_{(\beta,\beta_r^0)}(x_i)|}{2}\right)\right\} \geq \epsilon_r.$$

The purpose is to estimate the unknown regression parameters together with the change-points when n observations $(Y_i, x_i)_{1\leq i\leq n}$ are available. The construction of the estimator has two stages: first we search the regression parameters estimators and after we localize the change-points. For K change-points $l_1, \ldots, l_K$, let us denote $\hat{\theta}_1(\theta_2) = \hat{\theta}_1(l_1, \ldots, l_K) = \arg\min_{\theta_1} \sum_{r=1}^{K+1} \sum_{i=l_{r-1}+1}^{l_r} |Y_i - g_\theta(x_i)|$ the LAD estimator of θ_1^0 for a given parameter θ_2, with $l_0 = 0$ and $l_{K+1} = n$. Let us denote the sum of least absolute deviations of residuals for each fixed changes $l_1, \ldots, l_K$:

$$S(l_1, \ldots, l_K) := \sum_{r=1}^{K+1} \inf_{\theta_1} \sum_{i=l_{r-1}+1}^{l_r} |Y_i - g_\theta(x_i)|. \tag{8}$$

For the true parameters, this sum is: $S_0 := S(l_1^0, \ldots, l_K^0; \theta_1^0) = \sum_{i=1}^n |\varepsilon_i|$. We define the LAD estimator of the change-points by:

$$\hat{\theta}_{2n} = (\hat{l}_1, \ldots, \hat{l}_K) := \arg \min_{l_1<\cdots<l_K} S(l_1, \ldots, l_K). \tag{9}$$

The regression parameters estimators are obtained using the associated LAD estimator at estimated break points: $\hat{\theta}_{1n} := (\hat{\beta}_{1,n}, \ldots, \hat{\beta}_{K+1,n}) = \hat{\theta}_1(\hat{\theta}_{2n})$.

Between the break l_{r-1} and l_r the estimator of β is: $\hat{\beta}_{(l_{r-1},l_r)} := \arg\min_{\beta\in\Gamma} \sum_{i=l_{r-1}+1}^{l_r} |Y_i - h_\beta(x_i)|$. Let us denote: $\hat{Y}_{i,(l_{r-1},l_r)} := h_{\hat{\beta}_{(l_{r-1},l_r)}}(x_i)$.

We start with the study of the convergence rate of the change-point LAD estimator.

Theorem 3 *For all $\rho > 1/2$, under the assumptions* (A1), (A2), (H1), (H2′), (H3), (H4′) *for any sequence (v_n) as in the Theorem 1, we have for $n \to \infty$:*

$$I\!P[|\hat{l}_r - l_r^0| > \inf\{n^\rho, n^2 v_n^4\}] \to 0, \quad r = 1, \ldots, K.$$

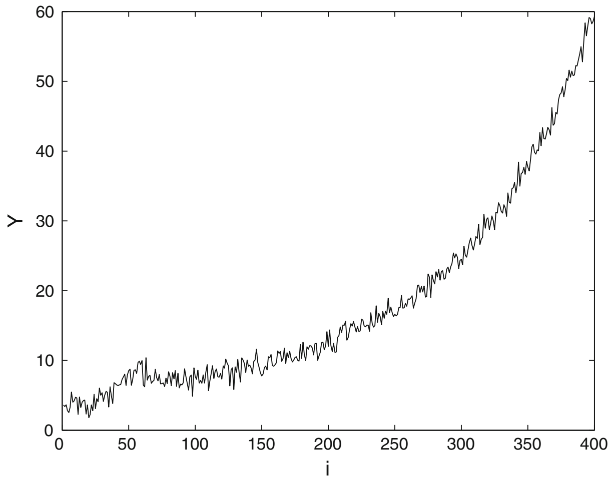

Fig. 1 Model with normal errors and with one change-point

Remark 3 The arguments used in the proof of Theorem 3 are totally different from those used in Bai (1998). The objective function study $\sum_{i=1}^n |Y_i - g_\theta(x_i)|$ when the parameter β is to the exterior of a neighborhood of β^0, in the case $h_\beta(x) = x\beta$ is based on the fact that $h_\beta(.)$ is convex and thus the extreme value of a convex function is attained on the boundary: $\inf_{|x|\geq c} h(x) = \inf_{|x|=c} h(x)$.

If we consider in more the assumption (H5), the convergence rate given by the Theorem 3 can be improved.

Theorem 4 *Under the conditions* (A1), (A2), (H1), (H2′), (H3), (H4′) *and* (H5) *for each* $r = 1, \ldots, K$ *we have:* $\hat{l}_r - l_r^0 = O_{I\!P}(1)$.

In order to work in a bounded interval, we can consider $\tau_r^0 = l_r^0/n$ and $\hat{\tau}_r = \hat{l}_r/n$ its estimator. Then, Theorem 4 implies that $\hat{\tau}_r$ converges in probability to τ_r^0 with the convergence rate n^{-1} (see Bai 1995, Theorem 1).

Example 1 Let us consider the following nonlinear model:

$$Y_i = (a_1^0 + e^{b_1^0 x_i}) 1\!\!1_{0<i\leq l_1^0} + (a_2^0 + e^{b_2^0 x_i}) 1\!\!1_{l_1^0<i\leq n} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{10}$$

where $x_i = 2i/n$ and $\{\varepsilon_i\}$ is a random sample from the standard normal distribution (see Fig. 1) or from Laplace distribution (see Fig. 2). The true values of the parameters are: $a_1^0 = 2$, $b_1^0 = 7$, $a_2^0 = 5$, $b_2^0 = 2$ and for the change-point $l_1^0 = 60$. The sample size used is $n = 400$. We realize $m = 100$ Monte Carlo replications.

From the Table 1, we see that the consistency of the estimates is shown by their means. The location of the change-point is detected well, including for the model with Laplace errors where outliers exist. Remark that for the last case, we cannot easily detect with the bare eye,the location of the change-point.

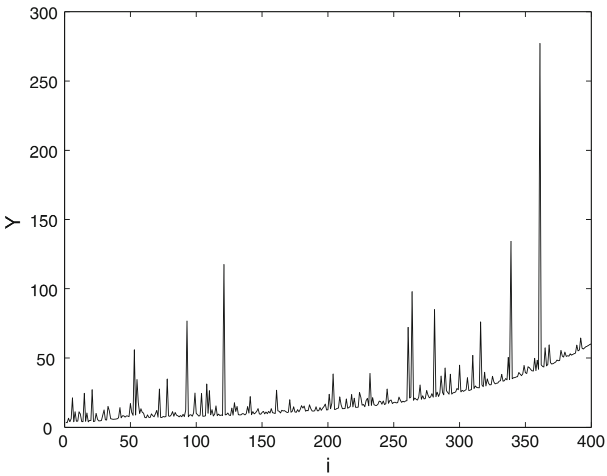

Fig. 2 Model with Laplace errors and with one change-point

Table 1 Mean of parameter estimates

Parameter	$\hat{a}_{1,n}(a_1^0 = 2)$	$\hat{b}_{1,n}(b_1^0 = 7)$	$\hat{a}_{2,n}(a_2^0 = 5)$	$\hat{b}_{2,n}(b_2^0 = 2)$	$\hat{l}_1(l_1^0 = 60)$
Normal errors	2.04	6.9	5.02	1.9998	60.93
Laplace errors	2.94	6.885	5.97	1.998	61

Let us give now the limiting distribution of the change-points estimators. Note that, due to the nonlinearity of the function $h_\beta(.)$, the proof is completely different from the similar result of Bai (1998).

Theorem 5 *Under the assumptions of Theorem 4, for each $r = 1, 2, \ldots, K$, the difference $\hat{l}_r - l_r^0$ converges in distribution for $n \to \infty$ to L_r, the location of the minima for the random processes $\{\ldots, Z_{-1}^{(r)}, Z_0^{(r)}, Z_1^{(r)}, \ldots\}$ with $Z_0^{(r)} = 0$. For $j = 1, 2, \ldots$*

$$Z_j^{(r)} = \sum_{i=l_r^0+1}^{l_r^0+j} \left\{ \left| \varepsilon_i^{(r)} - \nu_{(\beta_{r-1}^0, \beta_r^0)}(x_i) \right| - \left| \varepsilon_i^{(r)} \right| \right\},$$

and for $j = -1, -2, \ldots$:

$$Z_j^{(r)} = \sum_{i=l_r^0+j}^{l_r^0} \left\{ \left| \varepsilon_i^{(r)} + \nu_{(\beta_{r-1}^0, \beta_r^0)}(x_i) \right| - \left| \varepsilon_i^{(r)} \right| \right\},$$

where $\varepsilon_i^{(r)}$ is an independent copy of ε_i.

Remark 4 The distribution of $\hat{l}_r - l_r^0$ is symmetric about zero if ε_i has a symmetric distribution about zero. The result of the theorem indicates that the asymptotic distribution of $\hat{l}_r - l_r^0$ depends on the magnitude of shift $\nu_{(\beta^0_{r-1},\beta^0_r)}(x_i)$ and on the distribution of ε_i but not on l_r^0. This result will be confirmed by the simulations.

Remark 5 When the estimation method is the least-squares, if $h_\beta(x)$ is a step function: $h_\beta(x) = \beta$, Yao and Au (1989) proved that the change-point estimator converges in distribution to the location of the minimum of a random walk. For a linear regression: $h_\beta(x) = \beta x$, Bai and Perron (1998) proved that the asymptotic distribution of the change-point estimator is the location of the maximum of a Wiener process with drift.

Note that $\hat{\beta}_{r,n}$ depends only on change-points $\hat{l}_{r-1}$ and $\hat{l}_r$. Furthermore, by the Theorem 4, $\hat{l}_r = l_r^0 + O_{I\!P}(1)$, for each $r = 1, \ldots, K, K+1$. Moreover, again from Theorem 4, for each break point, the estimator $\hat{l}_r$ is determined by a small number of observations near l_r^0 whereas $\hat{\beta}_{r,n}$ is determinated by all set of observations between $\hat{l}_{r-1}$ and $\hat{l}_r$. Therefore, the estimated change-points are asymptotically independent of each other and of the estimated regression parameters. Taking into account the asymptotic normality of the LAD estimator for a nonlinear regression (Weiss 1991, see Theorem 3), we obtain the following result.

Theorem 6 *Under the assumptions of Theorem 4 and under the condition that the density f of the error ε is Lipschitz, we have for each $r = 1, \ldots, K+1$*

$$2f(0)(\hat{l}_r - \hat{l}_{r-1})(\hat{\beta}_{r,n} - \beta_r^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, V_r), \tag{11}$$

where $\hat{l}_{K+1} = l^0_{K+1} = n$ and

$$V_r = \frac{1}{l_r^0 - l_{r-1}^0} \sum_{i=l^0_{r-1}+1}^{l_r^0} \left(\frac{\partial h_{\beta_r^0}(x_i)}{\partial \beta}\right)\left(\frac{\partial h_{\beta_r^0}(x_i)}{\partial \beta}\right)^t.$$

Remark 6 Consequence of this theorem, using Slutsky's theorem, we obtain the asymptotic normality of the estimator $\hat{\beta}_{r,n}$.

Since $\hat{l}_r - \hat{l}_{r-1}$ converges in probability to $l_r^0 - l_{r-1}^0$, for $n \to \infty$, a consequence of the later theorem is that we can construct in the usual way the confidence interval for the regression parameters. Hypothesis tests can be realized. On the other hand, for the change-points, the properties of $\arg\min Z_j^{(r)}$ seem more difficult to study. For the moment, we cannot give a analytical solution to find the confidence interval for the change-points. We can only find a numerical approximation by using a Monte Carlo technique.

4 Estimating the number of change-points

We adapt the Schwarz criterion proposed by Yao (1988) to estimate the number K. Let K_0 be the true unknown value of K. We need to suppose that the number of change-points K_0 is not greater than a known upper bound K_U. For every

Table 2 Mean of criterion values $B(K) = n \log \hat{s}_K + K n^{5/8}$, for Normal and Laplace errors

	$K = 0$	$K = 1$	$K = 2$	$K = 3$
Normal errors	359.5	−20.2	1.43	15.3
Laplace errors	491.5	350.8	376.3	403.4

fixed change-points number K, consider $(\hat{l}_{1,K}, \ldots, \hat{l}_{K,K}) = \arg\min S(l_1, \ldots, l_K)$ and $\hat{s}_K = S(\hat{l}_{1,K}, \ldots, \hat{l}_{K,K})/n$, where the function S is defined by (8).

Theorem 7 *Under the conditions* $I\!E[|\varepsilon|] < \infty$, (A1), (A2), (H1)–(H4), *let* $\hat{K}_n$ *be the value of* K *that minimizes* $B(K) := n \log \hat{s}_K + K C_n$ *subject to* $K \le K_U$, *where* (C_n) *is any sequence satisfying* $C_n \to \infty$, $C_n n^{-3/4} \to 0$ *and* $C_n n^{-1/2} \to \infty$ *for* $n \to \infty$. *Then:*

$$I\!P[\hat{K}_n = K_0] \to 1, \quad \textit{for } n \to \infty.$$

Example 2 For the model (10), we consider the sequence $C_n = n^{5/8}$. We realize $m =$ Monte-Carlo replications. The results with the means of simulations are presented in the Table 2, for Normal, respectively, Laplace errors. For both distributions, for each replication, the criterion chooses one break point. Moreover $B(1) < B(2) < B(3) < B(0)$ in every case.

Remark 7 In case of a fixed sample size, for large n, the choice of C_n does not influence the estimation of the change-points number: $\arg\min_K B(K)$. Then, we can choose any sequence C_n converging to $+\infty$ of the form: $C_n = n^\delta$, with $\delta \in (1/2, 3/4)$.

Remark 8 In the paper of Horváth et al. (2004) a structural change is detected in a linear model using a class of procedures. The procedures are based on weighted CUSUMs of residuals, in which the unknown in-control parameter has been replaced by its least-squares estimate. Let us remind that for the distributions with outliers, the LS estimators can have large errors. On the other hand, to detect the changes in an autoregressive time series, Hušková et al. (2007) propose a test based on the partial sums of weighted residuals. The parameters are always estimated by the LS method.

5 Proofs of theorems

Proof of Theorem 1 By the Lemma 4 we have, with the probability 1, that there exists a $B > 0$ such that:

$$\limsup_{n \to \infty} \sup_{\|\beta - \beta^0\| = B v_n} \sum_{i=1}^n [|\varepsilon_i| - |\varepsilon_i - \nu_{(\beta, \beta^0)}(x_i)|] \frac{\|\beta - \beta^0\|^{-2}}{n} \le -M \frac{f(0)}{2}.$$

Then, with probability 1, for all n large enough:

$$\sum_{i=1}^n |\varepsilon_i| < \inf_{\|\beta - \beta^0\| = B v_n} \sum_{i=1}^n |\varepsilon_i - \nu_{(\beta, \beta^0)}(x_i)|. \tag{12}$$

For a fixed sequence (v_n), the Lemma 4 is valid for any positive sequence w_n satisfying (6) and such as: $w_n \geq v_n$. Since the constant B does not depend of n and since (v_n) is monotonic, we have:

$$\sum_{i=1}^{n} |\varepsilon_i| < \inf_{\|\beta-\beta^0\| \geq B v_n} \sum_{i=1}^{n} |\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)|. \tag{13}$$

Thus, the probability that $\hat{\beta}_n = \arg\min_{\beta \in \Gamma} \sum_{i=1}^{n} |\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)|$ belongs in the set $\{\beta; \|\beta - \beta^0\| \geq B v_n\}$ is zero, for n large enough. □

Proof of Theorem 2 From the Lemma 3, with $a_i = n^{-1}$, $u_n = n v_n^2$, we have:

$$\frac{2}{n} \sum_{i=1}^{n} [F(h_\beta(x_i)) - F(h_{\beta^0}(x_i))] = \frac{1}{n} \sum_{i=1}^{n} \text{sign}(\varepsilon_i) - \frac{1}{n} \sum_{i=1}^{n} \text{sign}(\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)) + O(v_n),$$

uniformly in β. On the other hand, by the dominated convergence theorem, under the hypothesis that $\partial h_\beta(x)/\partial \beta$ is bounded in a neighborhood of β^0, we get uniformly in β:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} [F(h_\beta(x_i)) - F(h_{\beta^0}(x_i))] = (\beta - \beta^0) \int_{-\infty}^{\infty} \frac{\partial h_{\beta^0}(x)}{\partial \beta} f(h_{\beta^0}(x)) dx.$$

□

Proof of Theorem 3 First, we prove that: $I\!P[|\hat{l}_r - l_r^0| > n^\rho] \to 0$.

To show this, for each change-point l_r^0, $1 \leq r \leq K$, consider the set:

$A_r(n, \rho) = \left\{(l_1, \ldots, l_K) | l_s - l_r^0| > n^\rho, \forall s = 1, \ldots, K\right\}$. We still have to prove that: $I\!P[(\hat{l}_1, \ldots, \hat{l}_K) \in A_r(n, \rho)] \to 0$ for $n \to \infty$. Consider the set $I_r = \{1, \ldots, r-1, r+2, \ldots, K+2\}$. For any fixed $(l_1, \ldots, l_K) \in A_r(n, \rho)$, we have:

$$S(l_1, \ldots, l_K) \geq S(l_1, \ldots, l_K, l_1^0, \ldots, l_{r-1}^0, l_r^0 - [n^\rho], l_r^0 + [n^\rho], l_{r+1}^0, \ldots, l_K^0). \tag{14}$$

Moreover, with the probability 1: $S(\hat{l}_1, \ldots, \hat{l}_K) \leq S(l_1^0, \ldots, l_K^0) \leq S_0 = \sum_{i=1}^{n} |\varepsilon_i|$. On the other hand:

$$S(l_1, \ldots, l_K, l_1^0, \ldots, l_{r-1}^0, l_r^0 - [n^\rho], l_r^0 + [n^\rho], l_{r+1}^0, \ldots, l_K^0) := \sum_{s=1}^{K+2} T_s, \tag{15}$$

where: T_s, $s \in I_r$, are the sums of absolute deviations involving observations for i between l_{s-1}^0 and l_s^0; T_r is the sum of absolute deviations involving observations for i

between l_{r-1}^0 and $l_r^0 - n^\rho$; T_{r+1} is the sum of absolute deviations involving observations for i between $l_r^0 + n^\rho$ and l_{r+1}^0; T_{K+2} is calculated for i between $l_r^0 - n^\rho$ and $l_r^0 + n^\rho$. Since $(l_1, \dots, l_K) \in A_r(n, \rho)$, all breaks $l_1, \dots, l_K$ are in $T_1, \dots, T_{r-1}, \dots, T_{K+1}$ but not in the sum T_{K+2}.

For each $s \in I_r$, let: $k_{1,s} < \dots < k_{J(s),s} := \{l_1, \dots, l_K\} \cap \left\{ j / l_{s-1}^0 < j \le l_s^0 \right\}$. The sum T_s can be written:

$$T_s = \sum_{j=1}^{J(s)+1} \min_{\theta_1} \sum_{i=k_{j-1,s}+1}^{k_{j,s}} |\varepsilon_i - g_{(\theta_1,\theta_2)}(x_i) + g_{(\theta_1^0,\theta_2^0)}(x_i)|,$$

with $0 \le J(s) \le K$. Using the Lemma 9 we obtain, for all $\delta > 1/2$:

$$T_s - \sum_{j=1}^{J(s)+1} \sum_{i=k_{j-1,s}+1}^{k_{j,s}} |\varepsilon_i| \ge -2(K+1) \sup_{1 \le l < k \le n} \left| \inf_\beta \sum_{i=l}^{k} (|\varepsilon_i - \nu_{(\beta,\beta_s^0)}(x_i)| - |\varepsilon_i|) \right|$$
$$= O_{I\!P}(n^\delta). \tag{16}$$

The sum: $T_r + T_{r+1} - \sum_{i=l_r^0-[n^\rho]+1}^{l_r^0+[n^\rho]} |\varepsilon_i|$ is equal to

$$\inf_\beta \sum_{i=l_r^0-[n^\rho]+1}^{l_r^0} (|\varepsilon_i - \nu_{(\beta,\beta_r^0)}(x_i)| - |\varepsilon_i|) + \inf_\beta \sum_{l_r^0+1}^{i=l_r^0+[n^\rho]} (|\varepsilon_i - \nu_{(\beta,\beta_{r+1}^0)}(x_i)| - |\varepsilon_i|). \tag{17}$$

Since $\beta_r^0 \ne \beta_{r+1}^0$ there exists a $c > 0$ such that: $\max(\|\beta - \beta_r^0\|, \|\beta - \beta_{r+1}^0\|) > c$. We suppose, without loss of generality, that $\|\beta - \beta_r^0\| \ge \|\beta - \beta_{r+1}^0\|$. Then, the infimum of $\sum_{i=l_r^0-[n^\rho]+1}^{l_r^0} (|\varepsilon_i - \nu_{(\beta,\beta_r^0)}(x_i)| - |\varepsilon_i|)$ is considered on a compact $\Gamma_{0,r} \subset \Gamma$, with $\beta_r^0 \notin \Gamma_{0,r}$. Applying the relation (4) of Oberhofer (1982) it follows that:

$$\inf_{\beta \in \Gamma_{0,r}} \sum_{i=l_r^0-[n^\rho]+1}^{l_r^0} (|\varepsilon_i - \nu_{(\beta,\beta_r^0)}(x_i)| - |\varepsilon_i|) \ge O_{I\!P}(n^\rho). \tag{18}$$

Combining the relations (15)–(18) we have:

$$S(l_1, \dots, l_K, l_1^0, \dots, l_{r-1}^0, l_r^0 - [n^\rho], l_r^0 + [n^\rho], l_{r+1}^0, \dots, l_K^0)$$
$$\ge -C O_{I\!P}(n^\delta) + O_{I\!P}(n^\rho) + \sum_{i=1}^{n} |\varepsilon_i|. \tag{19}$$

Let us consider $\delta \in (1/2, \rho)$. Then taking into account the relation (14) we have: $S(l_1, \dots l_n) \ge \sum_{i=1}^n |\varepsilon_i| + Z_n$ where Z_n is a random variable such that $Z_n \xrightarrow[n \to \infty]{I\!P} \infty$.

Hence:

$$I\!P\left[\min_{(l_1,\ldots,l_K)\in A_r(n,\rho)} S(l_1,\ldots,l_K) > S(l_1^0,\ldots,l_K^0)\right] \underset{n\to\infty}{\longrightarrow} 1,$$

what implies: $I\!P[(\hat{l}_1,\ldots,\hat{l}_K) \in A_r(n,\rho)] \underset{n\to\infty}{\longrightarrow} 0$. Now, to prove $I\!P[|\hat{l}_r - l_r^0| > n^2 v_n^4]$ $\to 0$ we follow the same steps as in Bai (1998) (Proposition 3) and combined it with the same arguments as in the proof of $I\!P[|\hat{l}_r - l_r^0| > n^\rho] \to 0$. Using the Lemma 5 (ii), in the relation (19) we have $O_{I\!P}(nv_n^2)$ in the place of $O_{I\!P}(n^\delta)$ and $O_{I\!P}(n^2 v_n^4)$ in the place of $O_{I\!P}(n^\rho)$. □

Proof of Theorem 4 For a sequence (v_n) as in Theorem 1 and $\lambda_n = O(n^2 v_n^4)$, let us consider the set: $B(n,\lambda_n) := \{(l_1,\ldots,l_K)/|l_k - l_k^0| < \lambda_n, \quad \forall k = 1,\ldots,K\}$. We also consider for a fixed change-point l_r^0 and for an arbitrary $\mathcal{M}_1 > 0$:

$$B_r(n,\lambda_n,\mathcal{M}) = \left\{(l_1,\ldots,l_K) \in B(n,\lambda_n)/l_r - l_r^0 < -\mathcal{M}_1\right\}.$$

As a consequence of the Theorem 3, for large n, the estimator $\hat{\theta}_{2n}$ belongs to $B(n,\lambda_n)$ with a probability close to 1. Let us consider two vectors of change-points $(l'_1,\ldots,l'_K)$ $\in B(n,\lambda_n)$ and $(l_1,\ldots,l_K) \in B_r(n,\lambda_n,\mathcal{M}_1)$ such that $l'_k = l_k$ for $k \neq r$ and $l'_r = l_r^0$. We shall find a $\mathcal{M}_1 > 0$ such that $\hat{\theta}_{2n}$ is in the ball with center l_r^0 and radius $\mathcal{M}_1$, with probability tending to 1. Let us consider the difference $S(l_1,\ldots,l_K) - S(l'_1,\ldots,l'_K)$ rewritten as:

$$\sum_{j=l_{r-1}+1}^{l_r} [|Y_j - \hat{Y}_{j,(l_{r-1},l_r)}| - |Y_j - \hat{Y}_{j,(l_{r-1},l_r^0)}|]$$
$$+ \sum_{j=l_r+1}^{l_r^0} [|Y_j - \hat{Y}_{j,(l_r,l_{r+1})}| - |Y_j - \hat{Y}_{j,(l_{r-1},l_r^0)}|]$$
$$+ \sum_{j=l_r^0+1}^{l_{r+1}} [|Y_j - \hat{Y}_{j,(l_r,l_{r+1})}| - |Y_j - \hat{Y}_{j,(l_r^0,l_{r+1})}|] := I_1 + I_2 + I_3.$$

Using the Lemma 8 (ii) we obtain that $I_1 = O_{I\!P}(1)$, $I_3 = O_{I\!P}(1)$ uniformly in $\mathcal{M}$. Concerning I_2, we have the decomposition:

$$I_2 = \sum_{j=l_r+1}^{l_r^0} [|\varepsilon_j - \nu_{(\beta_{r+1}^0,\beta_r^0)}(X_j)| - |\varepsilon_j|]$$
$$+ \sum_{j=l_r+1}^{l_r^0} [|\varepsilon_j - \nu_{(\hat{\beta}_{(l_r,l_{r+1})},\beta_r^0)}(X_j)| - |\varepsilon_j - \nu_{(\beta_{r+1}^0,\beta_r^0)}(X_j)|]$$

$$- \sum_{j=l_r+1}^{l_r^0} [|\varepsilon_j - \nu_{(\hat{\beta}_{(l_{r-1},l_r^0)},\beta_r^0)}(X_j)| - |\varepsilon_j|] := I_{21} + I_{22} + I_{23}.$$

For I_{22} we have that:

$$\begin{aligned}|I_{22}| &\leq \sum_{j=l_r+1}^{l_r^0} |\nu_{(\hat{\beta}_{(l_r,l_{r+1})},\beta_{r+1}^0)}(x_i)| \\ &\leq \|\hat{\beta}_{(l_r,l_{r+1})} - \beta_{r+1}^0\| \sum_{j=l_r+1}^{l_r^0} \sup_{\beta \in \mathcal{V}(\beta_{r+1}^0)} \left\| \frac{\partial h_{\beta_{r+1}^0+y}(x_i)}{\partial \beta} \right\|,\end{aligned}$$

with $\mathcal{V}(\beta_{r+1}^0)$ a neighborhood ob β_{r+1}^0.

By the Lemma 8(i) and the hypothesis (H2), considering $\rho \in [3/4, 1]$, $\upsilon \in (0, 1/4)$ and $\delta < \rho - \upsilon$, we obtain: $|I_{22}| \leq n^{-(\rho-\upsilon-\delta)/2} C n^{\upsilon} = o(1)$.

By similar arguments we prove $I_{23} = o_{I\!P}(1)$.

As regards I_{21}, using the Lemma 2 of Oberhofer (1982), we obtain that:

$$\begin{aligned}&\frac{1}{m} \sum_{j=1}^{m} I\!E[|\varepsilon_j - \nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j)| - |\varepsilon_j|] \\ &= \frac{2}{m} \sum_{\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j) \leq 0} \int_{(0,-\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j)]} [|\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j)| - x] dF(x) \\ &\quad + \frac{2}{m} \sum_{\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j) > 0} \int_{(-\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j),0]} [|\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j)| + x] dF(x).\end{aligned}$$

Moreover, since $(l_1, \ldots, l_K) \in B_r(n, \lambda_n, \mathcal{M}_1)$, $F(0) = 1/2$, we deduce that,

$$\geq \frac{1}{m} \sum_{j=1}^{m} |\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j)| \min \left\{ F(|\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j)|/2) - 1/2, 1/2 - F(-|\nu_{(\beta_r^0,\beta_{r+1}^0)}(X_j)|/2) \right\} > C.$$

Applying the Lemma 10 for $c_n = C$ we have that there exists a constant $C_1 > 0$ such that: $|I_{21}| \geq C_1(l_r^0 - l_r) \geq C_1 \mathcal{M}_1$. We choose $\mathcal{M}_1$ such that $|I_{21}|$ is bigger than $(|I_1|, |I_3|, |I_{22}|, |I_{23}|)$. Then the following relation holds:

$$\min_{(l_1,\ldots,l_K) \in B_r(n,\lambda_n,\mathcal{M}_1)} S(l_1, \ldots, l_K) - S(l'_1, \ldots, l'_K) \geq C\mathcal{M}_1 (1 + o_{I\!P}(1)).$$

This proves the claim of the theorem:

$$\lim_{n \to \infty} I\!P[(\hat{l}_1, \ldots, \hat{l}_K) \in B_r(n, \lambda_n, \mathcal{M}_1)] = 0.$$

□

Proof of Theorem 5 For $z \in I\!R$, we define:

$$G(z) := z[1-2F(-z)] + \text{sign}(z)2\int_{\max(0,-z)}^{\min(0,-z)} x dF(x).$$

For $j > 0$, there exists $c > 0$ such that:

$$I\!E[Z_j^{(r)}] = \sum_{i=l_r^0+1}^{l_r^0+j} G(\nu_{(\beta_{r-1}^0,\beta_r^0)}(x_i)) \geq \sum_{i=l_r^0+1}^{l_r^0+j} |\nu_{(\beta_{r-1}^0,\beta_r^0)}(x_i)|$$
$$\times \min\left\{F(|\nu_{(\beta_{r-1}^0,\beta_r^0)}(x_i)|/2) - 1/2, 1/2 - F(-|\nu_{(\beta_{r-1}^0,\beta_r^0)}(x_i)|/2)\right\} > cj.$$

Thus $I\!E[Z_j^{(r)}]$ converges to infinity for $j \to \infty$. With probability arbitrarily close to 1, for sufficiently large $\mathcal{M}_2$, the minimizer of $Z_j^{(r)}$ belongs to the interval $[-\mathcal{M}_2, \mathcal{M}_2]$: for all fixed $\tilde{\delta} > 0$, for each $r = 1, \ldots, K$ we have:

$$I\!P[|L_r| \leq \mathcal{M}_2] > 1 - \tilde{\delta}. \tag{20}$$

Let us consider the difference $D := S(l_1^0 + i_1, \ldots, l_K^0 + i_K) - S(l_1^0, \ldots, l_K^0)$ with i_r integer, $|i_r| \leq \mathcal{M}_1$, $\mathcal{M}_1$ as in the proof of the Theorem 4. From Theorem 4, $\hat{l}_r$ is the value of l_r that minimize D. Then, for all fixed $\tilde{\delta} > 0$, for each $r = 1, \ldots, K$ we have:

$$I\!P[\hat{l}_r - l_r^0| \leq \mathcal{M}_1] > 1 - \tilde{\delta}. \tag{21}$$

Choose $\mathcal{M} \geq \max\{\mathcal{M}_1, \mathcal{M}_2\}$.

We shall prove now that for $|i_r| \leq \mathcal{M}$, D converge jointly in i_r in distribution to $\sum_{r=1}^K Z_{i_r}^{(r)}$. Without loss of generality, we suppose that $i_{r-1}, i_r > 0$. Then, we have the decomposition:

$$\begin{aligned} D &= \sum_{r=1}^K \sum_{j=l_{r-1}^0+i_{r-1}+1}^{l_r^0} \left\{|Y_j - \hat{Y}_{(l_{r-1}^0+i_{r-1},l_r^0+i_r)}(X_j)| - |Y_j - \hat{Y}_{(l_{r-1}^0,l_r^0)}(X_j)|\right\} \\ &\quad + \sum_{r=1}^K \sum_{j=l_r^0+1}^{l_r^0+i_r} \left\{\left(|Y_j - \hat{Y}_{(l_{r-1}^0+i_{r-1},l_r^0+i_r)}(X_j)| - |\varepsilon_j|\right) - \left(|Y_j - \hat{Y}_{(l_r^0,l_{r+1}^0)}(X_j)| - |\varepsilon_j|\right)\right\} \\ &:= D_1 + (D_{21} - D_{22}). \end{aligned}$$

First, since $\partial h_{\beta_r^0}(x)/\partial\beta$ is bounded:

$$|D_1| \leq \sum_{r=1}^K \sum_{j=l_{r-1}^0+i_{r-1}+1}^{l_r^0} |\nu_{(\beta_r^0,\hat{\beta}_{(l_{r-1}^0+i_{r-1},l_r^0+i_r)})}(X_j) - \nu_{(\beta_r^0,\hat{\beta}_{(l_{r-1}^0,l_r^0)})}(X_j)| = o_{I\!P}(1).$$

On the other hand, we have $|D_{22}| \leq \sum_{r=1}^{K} \sum_{j=l_r^0+1}^{l_r^0+i_r} |\nu_{(\beta^0_{r+1}, \hat{\beta}_{(l_r^0, l_{r+1}^0)})}(X_j)| = o_{I\!P}(1)$. For D_{21}, since $\hat{\beta}_{(l^0_{r-1}+i_{r-1}, l^0_r+i_r)}$ converges in probability to β^0_{r-1} for $n \to \infty$, uniformly in i_{r-1} and i_r, we have:

$$D_{21} = \left[\sum_{r=1}^{K} \sum_{j=l_r^0+1}^{l_r^0+i_r} \left(|\varepsilon_j + \nu_{(\beta_r^0, \beta_{r-1}^0)}(X_j)| - |\varepsilon_j| \right) \right] (1 + o_{I\!P}(1)) = \sum_{r=1}^{K} Z^{(r)}_{i_r} (1 + o_{I\!P}(1)).$$

Therefore D converges in distribution to $\sum_{r=1}^{K} Z^{(r)}_{i_r}$, for $n \to \infty$. Since we have two independent sets of random variables around each of two change-points, the minimum of D is the sum of minima of $S(l_1^0, l_2^0, \ldots, l_r^0 + i_r, l_{r+1}^0, \ldots, l_K^0) - S(l_1^0, \ldots, l_K^0)$. The same thing for the minimum of the process $\sum_{r=1}^{K} Z^{(r)}_{i_r}$. Then, for all $\tilde{\delta} > 0$, taking in account (20) and (21) we obtain:

$$\left| I\!P[\hat{l}_r - l_r^0 = i_r] - I\!P[L_r = i_r] \right| < 3\tilde{\delta}. \tag{22}$$

Hence the proof is complete. □

Proof of Theorem 6 Since $\hat{\beta}_{r,n}$ depends only on change-points $\hat{l}_{r-1}$ and $\hat{l}_r$, we have: $I\!P[\hat{\beta}_{r,n} = \hat{\beta}_{(\hat{l}_{r-1}, \hat{l}_r)}] = 1$, where

$$(\hat{\beta}_{1,n}, \ldots, \hat{\beta}_{r,n}, \ldots, \hat{\beta}_{K+1,n}) = \arg \min_{(\beta_1, \ldots, \beta_r, \ldots, \beta_{K+1}) \in \Gamma^{K+1}} (\arg \min_{l_1 < \cdots < l_K} S(l_1, \ldots, l_K))$$

and

$$\hat{\beta}_{(\hat{l}_{r-1}, \hat{l}_r)} = \arg \min_{\beta \in \Gamma} \sum_{i=\hat{l}_{r-1}+1}^{\hat{l}_r} |Y_i - h_\beta(x_i)|.$$

By Theorem 4: $\hat{l}_{r-1} = l^0_{r-1} + O_{I\!P}(1)$ and $\hat{l}_r = l_r^0 + O_{I\!P}(1)$. Then, using Lemma 8: $\hat{\beta}_{(\hat{l}_{r-1}, \hat{l}_r)} = \hat{\beta}_{(l^0_{r-1}, l^0_r)}(1 + o_{I\!P}(1))$. On the other hand, we have (Weiss 1991, Theorem 3):

$$2f(0)(l_r^0 - l_{r-1}^0)(\hat{\beta}_{(l^0_{r-1}, l^0_r)} - \beta_r^0) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, V_r).$$

By Slutsky's theorem, the theorem follows. □

Proof of Theorem 7 First let us consider $K < K_0$. Under the hypothesis (A1), using similar arguments as for the proof of (18) because between two change-points the length is at least $n^{3/4}$, we have, with probability close to one: $S(\hat{l}_{1,K}, \ldots, \hat{l}_{K,K}) - \sum_{i=1}^{n} |\varepsilon_i| > C n^{3/4}$, with $C > 0$. By Lemma 13 of Bai (1998): $0 \leq S(\hat{l}_{1,K_0}, \ldots, \hat{l}_{K_0,K_0}) - \sum_{i=1}^{n} |\varepsilon_i| = O_{I\!P}(n^{1/4})$.

Then $\hat{s}_{K_0} = O_{I\!P}(n^{-3/4}) + n^{-1}\sum_{i=1}^n |\varepsilon_i| \xrightarrow[n\to\infty]{I\!P} I\!E[|\varepsilon_i|]$. Using the relation (62) of Bai (1998) we get: $B(K) - B(K_0) > [Cn^{3/4} - O_{I\!P}(n^{1/4})]/\hat{s}_{K_0} + (K - K_0)C_n = O_{I\!P}(n^{3/4}) - O(C_n) \to \infty$. Thus $I\!P[\hat{K}_n < K_0] \to 0$ for $n \to \infty$.

Now consider K, such that $K_0 < K \leq K_U$. From the definition of S we have: $S(l_1, \ldots, l_{K_0}) \geq S(\hat{l}_{1,K_0}, \ldots, \hat{l}_{K_0,K_0})$. We also have $S_0 \geq S(\hat{l}_{1,K_0}, \ldots, \hat{l}_{K_0,K_0}) \geq S(\hat{l}_{1,K}, \ldots, \hat{l}_{K,K})$. Exactly as in the proof of Theorem 3, we obtain that the last sum is greater than $S(\hat{l}_{1,K}, \ldots, \hat{l}_{K,K}, l_1, \ldots, l_{K_0})$.

By an argument similar to the one used for the proof of Theorem 3 (for T_s) we obtain that: $S(\hat{l}_{1,K}, \ldots, \hat{l}_{K,K}, l_1, \ldots, l_{K_0}) \geq S_0 - V_n$. The random variable V_n is strictly positive with probability converging to one, and $V_n = O_{I\!P}(n^\delta)$ for all $\delta > 1/2$.

These inequalities imply: $0 \leq \hat{s}_{K_0} - \hat{s}_K = O_{I\!P}(n^{\delta-1})$. Then: $n \log \hat{s}_{K_0} - n \log \hat{s}_K = O_{I\!P}(n^\delta)$. Since C_n satisfies the given conditions, we deduce:

$$n \log \hat{s}_K - n \log \hat{s}_{K_0} + KC_n - K_0 C_n = O_{I\!P}(n^\delta) + C_n(K - K_0) \to \infty, \quad \text{for } n \to \infty.$$

Therefore, $n \log \hat{s}_K + KC_n > n \log \hat{s}_{K_0} + K_0 C_n$, for large n. We conclude that, for $n \to \infty$, $I\!P[\hat{K}_n > K_0] \to 0$. □

Appendix: proofs of lemmas

Recall first a lemma due to Babu (1989).

Lemma 1 (Babu 1989, Lemma 1) *Let Z_i be a sequence of independent random variables with mean zero and $|Z_i| \leq b$ for some $b > 0$. Let also $V \geq \sum_{i=1}^n I\!E[Z_i^2]$. Then for all $0 < s < 1$ and $0 \leq a \leq V/(sb)$,*

$$I\!P\left[\sum_{i=1}^n Z_i > a\right] \leq 2\exp\left(-a^2 s(1-s)/V\right). \tag{23}$$

For the linear case, the proofs of analogue lemmas 6 and 10 are based on the convexity in ϕ of $|\varepsilon_i - x_i\phi| - |\varepsilon_i|$. The extreme value of a convex function is attained on the boundary.

To prove the asymptotical properties of the LAD estimator in a nonlinear model, without change-point, we need the auxiliary results.

Let $diam(\Gamma) = \sup_{\beta_1,\beta_2\in\Gamma} \|\beta_1 - \beta_2\|$ be the diameter of the set Γ. For the random processes defined by:

$$Z_i(\beta) := \text{sign}(\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)) - \text{sign}(\varepsilon_i) + 2F(h_\beta(x_i)) - 2F(h_{\beta^0}(x_i)), \tag{24}$$

with $\beta \in \Gamma$, we prove following two lemmas.

Lemma 2 *Let $(a_i)_{1\leq i\leq n}$ be a sequence such that $|a_i| \leq 1/c_n$, $c_n \to \infty$. Let $A_n = u_n^{1/2}\left(\sum_{i=1}^n a_i^2\right)^{1/2}$ and the sequences (u_n) and (c_n) satisfying the conditions:*

$u_n = O(\log n)$, $nc_n^{-2} = O(1)$. *Under the assumptions* (H1), (H2), (H3), *there exists two reals* $k > 0, d > 0$ *such that:*

$$\mathbb{P}\left[|\sum_{i=1}^n a_i Z_i(\beta)| > dA_n\right] = O(n^{-k}).$$

Proof of Lemma 2 We verify the conditions of Lemma 1. Consider: $b = 4/c_n$, $s = 1/2$, $a = A_n$. Thus, by Lipschitz condition for F and the assumption (H3):

$$\sum_{i=1}^n a_i^2 \mathbb{E}[Z_i^2(\beta)] \leq 4\sum_{i=1}^n a_i^2 \left|F(h_\beta(x_i)) - F(h_{\beta^0}(x_i))\right| \leq 4c^{(0)} \sum_{i=1}^n a_i^2 \left|\nu_{(\beta,\beta^0)}(x_i)\right|$$
$$\leq diam(\Gamma)\frac{4c^{(0)}}{c_n^2}\sum_{i=1}^n \sup_{\beta\in\Gamma}\left\|\frac{\partial h_\beta(x_i)}{\partial\beta}\right\| \leq diam(\Gamma)4c^{(0)}c^{(1)}\frac{n}{c_n^2}.$$

Let also consider $V = 4diam(\Gamma)c^{(0)}c^{(1)}nc_n^{-2}$. We have:

$$\frac{A_n b}{V} = \frac{diam(\Gamma)c_n u_n^{1/2}\left(\sum_{i=1}^n a_i^2\right)^{1/2}}{nc^{(0)}c^{(1)}} \leq \frac{1}{diam(\Gamma)c^{(0)}c^{(1)}}\left(\frac{u_n}{n}\right)^{1/2}.$$

We can take every d such that: $d < 2c^{(0)}c^{(1)}(n/u_n)^{1/2}$. □

Example For the sequences: $c_n = n$, $u_n = \log n$, $a_i = n^{-1}$ we have: $A_n = (n^{-1} logn)^{1/2}$.

Lemma 3 *Under the same assumptions as in Lemma 2, we have:*

$$\sup_{\beta\in\Gamma}\left|\sum_{i=1}^n a_i Z_i(\beta)\right| = O_{\mathbb{P}}\left((u_n\sum_{i=1}^n a_i^2)^{1/2}\right). \tag{25}$$

Proof of Lemma 3 First, we show that for every $k \geq 2$, $B > 0$, there exists $d > 0$ such that:

$$\mathbb{P}\left[\sup_{\|\beta_1-\beta_2\|<Bn^{-1/2}}\left|\sum_{i=1}^n a_i\left[Z_i(\beta_1) - Z_i(\beta_2)\right]\right| > dA_n\right] = O\left(n^{-k}\right), \tag{26}$$

with: $A_n = n^{-1/4}[u_n\sum_{i=1}^n a_i^2]^{1/2}$.

We observe that: $\left|\text{sign}(\varepsilon_i - \nu_{(\beta_1,\beta^0)}(x_i)) - \text{sign}(\varepsilon_i - \nu_{(\beta_2,\beta^0)}(x_i))\right| \leq 2\gamma_i(\beta_1,\beta_2)$, where $\gamma_i(\beta_1,\beta_2) = 1\!\!1_{\min(\nu_{(\beta^0,\beta_1)}(x_i),\nu_{(\beta^0,\beta_2)}(x_i))\leq\varepsilon_i\leq\max(\nu_{(\beta^0,\beta_1)}(x_i),\nu_{(\beta^0,\beta_2)}(x_i))}$.

Obviously $\gamma_i(\beta_1, \beta_2) = 0$ if $|\varepsilon_i - \nu_{(\beta_2,\beta^0)}(x_i)| > 2|\nu_{(\beta_1,\beta_2)}(x_i)|$. Let be the random event: $A_i(\beta) := \left\{\left|\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)\right| < 2Bn^{-1/2}\sup_{\beta\in\Gamma}\|\partial h_\beta(x_i)/\partial\beta\|\right\}$. Then:

$$\sup_{\|\beta_1-\beta_2\|<Bn^{-1/2}}\left|\sum_{i=1}^n a_i\left[Z_i(\beta_1) - Z_i(\beta_2)\right]\right|$$
$$\leq 2\sum_{i=1}^n |a_i|\left[1\!\!1_{A_i(\beta_2)} - I\!P\left[A_i(\beta_2)\right]\right] + 4\sum_{i=1}^n |a_i| I\!P\left[A_i(\beta_2)\right].$$

It is obvious that there exists $c>0$ such that: $I\!P[A_i(\beta_2)] \leq cn^{-1/2}\sup_{\beta\in\Gamma}\|\partial h_\beta(x_i)/\partial\beta\|$.

We apply Lemma 1 for $\sum_{i=1}^n |a_i|[1\!\!1_{A_i(\beta_2)} - I\!P\left[A_i(\beta_2)\right]]$ and $b = c_n^{-1}$. We obtain:

$$\sum_{i=1}^n a_i^2 I\!E[1\!\!1_{A_i(\beta_2)} - I\!P\left[A_i(\beta_2)\right]]^2 = \sum_{i=1}^n a_i^2 I\!P[A_i(\beta_2)]\{1 - I\!P[A_i(\beta_2)]\}$$
$$\leq 2\sum_{i=1}^n a_i^2 I\!P[A_i(\beta_2)]$$
$$\leq \frac{2}{c_n^2}\sum_{i=1}^n I\!P[A_i(\beta_2)] \leq C\frac{c^{(1)}}{c_n^2}n^{1/2},$$

and the claim (26) yields.

We divide the parameter set Γ into $\left(\frac{n^{1/2}}{u_n\sum_{i=1}^n a_i^2}\right)^{1/2}$ cells. We apply the Borel–Cantelli lemma and the relation (25) follows using also Lemma 2. □

Recall that $M = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \|\partial h_{\beta^0}(x_i)/\partial\beta\|^2$.

Lemma 4 *Under the assumptions (H1)–(H4), for any monotone positive sequence $v_n \to 0$ such that $nv_n^2 \to \infty$ for $n \to \infty$, we have the existence of a $B > 0$ such that, with the probability* 1*:*

$$\lim_{n\to\infty}\sup\left[\sup_{\|\beta-\beta^0\|=Bv_n}\frac{|\sum_{i=1}^n \xi_i(\beta)|}{n\|\beta-\beta^0\|^2}\right] \leq M\frac{f(0)}{2}.$$

Proof of Lemma 4 Divide $\|\beta-\beta^0\| = Bv_n$ into $N \leq (B+1)^p(n+2)^{p/2}$ cells. For β_1 and β_2 in the same cell, $\|\beta_1-\beta_2\| \leq n^{-1/2}v_n$, we have: $|\sum_{i=1}^n[\xi_i(\beta_1) - \xi_i(\beta_2)]| \leq C\sum_{i=1}^n|\nu_{(\beta_1,\beta_2)}(x_i)| \leq Cn^{1/2}v_n$. In the other hand, using (H3) and (H4), there exists two positive constants C_1, C_2 not dependent on n, such that:

$$Var\left[\sum_{i=1}^n \xi_i(\beta)\right] = 4\sum_{i=1}^n\int_{h_{\beta^0}(x_i)}^{f_\beta(x_i)}\int_{h_{\beta^0}(x_i)}^{f_\beta(x_i)}[F(\min(u,v)) - F(u)F(v)]\mathrm{d}u\mathrm{d}v$$
$$\leq C_1\sum_{i=1}^n[\nu_{(\beta,\beta^0)}(x_i)]^2 \leq C_2 n\|\beta-\beta^0\|^2 M(1+o(1)) = O(nv_n^2).$$

We consider: $B = (MC_2)^{1/2}$.

On the other hand, we have with the probability 1: $|\eta_i(\beta)| \le C|\nu_{(\beta,\beta^0)}(x_i)| \le C\|\beta-\beta^0\| = O(v_n)$ and by the Lemma 2 of Oberhofer (1982):

$$\begin{aligned} I\!E[\eta_i(\beta)] &= 2 1\!\!1_{\nu_{(\beta,\beta^0)}(x_i)<0} \int_0^{-\nu_{(\beta,\beta^0)}(x_i)} [|\nu_{(\beta,\beta^0)}(x_i)| - y]\mathrm{d}F(y) \\ &\quad + 2 1\!\!1_{\nu_{(\beta,\beta^0)}(x_i)>0} \int_{-\nu_{(\beta,\beta^0)}(x_i)}^{0} [|\nu_{(\beta,\beta^0)}(x_i)| + y]\mathrm{d}F(y) \\ &= O((\nu_{\beta,\beta^0)}(x_i)) = O(v_n). \end{aligned}$$

Then: $|\xi_i(\beta)| = O_{I\!P}(v_n)$. Thus, we can apply the Lemma 1 to $\sum_{i=1}^n \xi_i(\beta)$ for $s = 1/2$, $V = nv_n^2$, $b = v_n$ and $a = nv_n^2 f(0)M/2$. The lemma follows by applying the Borel–Cantelli lemma. □

The following lemma gives the behaviour of the function to be minimized. Remark that its demonstration is similar to that of Bai (1998) but the nonlinearity of $h_\beta(x)$ will be used in all the stages of proof. It is interesting to notice that we give a more precise order for the supremmum in (ii).

Lemma 5 *Under the assumptions* (H1)–(H4) *and the condition* (5)*, for each* $\delta \in (0, 1)$*, for any sequence* (v_n) *as in the Theorem* 1*, we have*

$$(i) \qquad \sup_{n\delta \le k \le n} \left| \inf_{\beta\in U(\beta^0, n^{-1/2}\phi)} \sum_{i=1}^k [|\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)| - |\varepsilon_i|] \right| = O_{I\!P}(1).$$

$$(ii) \qquad \sup_{1 \le k \le n} \left| \inf_{\beta\in U(\beta^0, n^{-1/2}\phi)} \sum_{i=1}^k |\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)| - |\varepsilon_i| \right| = O_{I\!P}(nv_n^2),$$

where $U(\beta^0, n^{-1/2}\phi)$ *is a neighborhood of* β^0 *and* $\phi \in I\!R^p$.

Proof of Lemma 5 (i) By definition: $\hat\beta_k = \arg\min_\beta \sum_{i=1}^k |\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)|$. By similar calculations that for Lemma 4 and Theorem 1, we can prove that: $\sup_{k>n\delta} \|\hat\beta_k - \beta^0\| = O_{I\!P}(n^{-1/2})$. Thus, to obtain (i) it suffices to show that the following random process:

$$G_{k,n}(\phi) := \sum_{i=1}^k [|\varepsilon_i - \nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)| - |\varepsilon_i|], \quad \phi \in I\!R^p,$$

is bounded for all $\|\phi\| \le C$. Setting

$$R_{i,n}(\phi) := |\varepsilon_i - \nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)| - |\varepsilon_i| - \nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i) \cdot \mathrm{sign}(\varepsilon_i),$$

we can write $G_{k,n}(\phi)$ as:

$$G_{k,n}(\phi) = \sum_{i=1}^k \mathrm{sign}(\varepsilon_i)\nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i) + \sum_{i=1}^k R_{i,n}(\phi).$$

By the central limit theorem, using (H2), we have $n^{-1/2}\sum_{i=1}^{k}\frac{\partial h_{\beta^0}(x_i)}{\partial\beta}\text{sign}(\varepsilon_i) = O_{I\!P}(1)$, uniformly in k. Since $||\partial^2 h_{\beta^0}(x_i)/\partial\beta^2||$ is bounded, we obtain: $|\sum_{i=1}^{k}\nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)\cdot\text{sign}(\varepsilon_i)| = O_{I\!P}(1)$, uniformly in k. Observe also that, we can rewrite $R_{i,n}(\phi)$ as:

$$\begin{aligned}&\text{sign}\left(\varepsilon_i - \nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)\right)[\varepsilon_i - \nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)]\\&\quad - \text{sign}(\varepsilon_i)\varepsilon_i - \nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)\text{sign}(\varepsilon_i).\end{aligned}$$

Thus: $|R_{i,n}(\phi)| \le 4 1\!\!1_{|\varepsilon_i|<|\nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)|}|\nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)|$. By the above inequality and (H2), we obtain:

$$\left|\sum_{i=1}^{k} R_{i,n}(\phi)\right| \le CMn^{-1/2}\sum_{i=1}^{n} 1\!\!1_{|\varepsilon_i|<|\nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)|}, \quad \forall k = 1,\ldots,n,$$

uniformly in ϕ. In addition, under the Lipshitz condition for F in a neighborhood of β^0: $I\!E[n^{-1/2}\sum_{i=1}^{n} 1\!\!1_{|\varepsilon_i|<|\nu_{(\beta^0+n^{-1/2}\phi,\beta^0)}(x_i)|}] = O(1)$ and this completes the proof of the first part of the Lemma.

(ii) By Theorem 1 we have that, for every $\epsilon > 0$ there exists a $B \in (0,\infty)$ such that for all large k:

$$I\!P[\|\hat{\beta}_k - \beta^0\| \le Bv_k] \ge 1-\epsilon, \tag{27}$$

while noting $M_k = Bv_k$, we will show that there exists $n_0 \in I\!N$ such that for all $n \ge n_0$, with the probability approaching 1:

$$\sup_{\|\beta-\beta^0\|\le M_k}\left|\sum_{i=1}^{k}\eta_i(\beta)\right| < 2Akv_k^2 \le 2Anv_n^2, \tag{28}$$

for all k such that $n_0 \le k \le n$. On the other hand, for $k < n_0$, we have: $|\eta_i(\beta)| < M_k \sup_{\|\beta-\beta^0\|\le M_k}\|\partial h_\beta(x_i)/\partial\beta\|$. Then, there exists a constant $C > 0$ such that: $|\sum_{i=1}^{k}\eta_i(\beta)| \le \sum_{i=1}^{n_0}|\eta_i(\beta)| < Cnv_n^2$. Hence, for large n, with the probability 1:

$$\sup_{\|\beta-\beta^0\|\le M_k}\left|\sum_{i=1}^{k}\eta_i(\beta)\right| < Cnv_n^2$$

and the relation (28) is also satisfied for all $k < n_0$. Hence the relation (28) holds for all $k \ge 1$. Let us observe that (27) implies:

$$\begin{aligned}&I\!P\left[\sup_{1\le k\le n}|\inf_{\beta\in\Gamma}\sum_{i=1}^{k}\eta_i(\beta)| > 2Anv_n^2\right]\\&\quad \le o(1) + I\!P\left[\sup_{1\le k\le n}\sup_{\|\beta-\beta^0\|\le M_k}|\sum_{i=1}^{k}\eta_i(\beta)| > 2Anv_n^2\right]\end{aligned}$$

and taking into account (28) for all k, the claim (ii) is deduced.

It remains to show that there exists $n_0 \in I\!N$ such that (28) holds for all $n \geq n_0$. We divide the region $\|\beta - \beta^0\| < M_k$ into $C_p k^{p/2}$ cells. For β_1 and β_2 belonging to the same cell we have $\|\beta_1 - \beta_2\| \leq k^{-1/2} M_k$. For a β_s in the sth cell:

$$\sup_{\|\beta-\beta^0\|\leq M_k} \left|\sum_{i=1}^{k} \eta_i(\beta)\right| \leq \sup_{\beta_s} \left|\sum_{i=1}^{k} \eta_i(\beta_s)\right| + \sup_{\|\beta_1-\beta_2\|\leq k^{-1/2}M_k} \left|\sum_{i=1}^{k} [\eta_i(\beta_1) - \eta_i(\beta_2)]\right|. \tag{29}$$

For the second term on the right-hand side of (29) we get:

$$\left|\sum_{i=1}^{k} [\eta_i(\beta_1) - \eta_i(\beta_2)]/(kv_k^2)\right|$$
$$\leq \|\beta_1 - \beta_2\| \sum_{i=1}^{k} \sup_{\beta} \left\|\frac{\partial h_\beta(x_i)}{\partial \beta}\right\| \Big/ (kv_k^2) \leq C \left(kv_k^2\right)^{-1/2}.$$

For the first term on the right-hand side of (29), we can write:
$\eta_i(\beta_s) = \varepsilon_i[\text{sign}(\varepsilon_i - \nu_{(\beta_s,\beta^0)}(x_i)) - \text{sign}(\varepsilon_i)] - \nu_{(\beta_s,\beta^0)}(x_i)\text{sign}(\varepsilon_i - \nu_{(\beta_s,\beta^0)}(x_i))$. Without loss of generality, we assume: $\nu_{(\beta_s,\beta^0)}(x_i) > 0$. Then: $\text{sign}(\varepsilon_i - \nu_{(\beta_s,\beta^0)}(x_i)) - \text{sign}(\varepsilon_i) = -2 1\!\!1_{0<\varepsilon_i<\nu_{(\beta_s,\beta^0)}(x_i)}$. We have also: $I\!E[\text{sign}(\varepsilon_i - \nu_{(\beta_s,\beta^0)}(x_i))] = 1 - 2F\left(\nu_{(\beta_s,\beta^0)}(x_i)\right) = -2\int_0^{\nu_{(\beta_s,\beta^0)}(x_i)} f(t)dt$. Then

$$I\!E[\eta_i(\beta_s)] = 2\int_0^{\nu_{(\beta_s,\beta^0)}(x_i)} [\nu_{(\beta_s,\beta^0)}(x_i) - t] f(t)\mathrm{d}t = \nu^2_{(\beta_s,\beta^0)}(x_i) f(0)\left(1 + o(1)\right).$$

Using dominated convergence Theorem, we obtain: $|\sum_{i=1}^{k} I\!E[\eta_i(\beta_s)]| \leq Ckf(0) \|\beta_s - \beta^0\|^2 \leq Cf(0)kv_k^2$. With the probability 1, we have: $|\xi_i(\beta_s)| \leq 2|\nu_{(\beta_s,\beta^0)}(x_i)| \leq Cv_k \sup_{\|\beta-\beta^0\|<M_k} \left\|\partial h_\beta(x_i)/\partial\beta\right\|$. Then: $Var[\xi_i(\beta_s)] \leq 4\nu^2_{(\beta_s,\beta^0)}(x_i)$. This implies, since $\partial h_\beta(x)/\partial\beta$ is bounded in the region $\|\beta - \beta^0\| < M_k$, that: $\sum_{i=1}^{k} I\!E[\xi_i^2(\beta_s)] \leq Ckv_k^2$. From Lemma 1 and the relation (29) we obtain in a similar way as in the linear case (Bai 1995, Lemma A.1.):

$$\lim\sup_{k\to\infty} \left[\sup_{\|\beta-\beta^0\|\leq M_k} \left|\sum_{i=1}^{k} \eta_i(\beta)\right| /(kv_k^2\right] \leq A, \tag{30}$$

with the probability 1, for some $A > 0$. The relation (30) implies that there exists $n_0 \in I\!N$ such that (28) holds for all $k \geq n_0$. □

Lemma 6 *Under the assumptions (H1), (H2), (H5) and the condition* (5)*, for a monotone positive sequence* $v_n \to 0$ *such that* $nv_n^2 \to \infty$*, there exists an* $\epsilon > 0$ *such that, with the probability* 1*:*

$$\liminf_{n\to\infty} \inf_{\|\beta-\beta^0\|\geq \upsilon_n} \frac{1}{n}\sum_{i=1}^n \eta_i(\beta) \geq \epsilon. \tag{31}$$

Proof of Lemma 6 Let us consider β such that $\|\beta-\beta^0\| \geq \upsilon_n$. For every compact sub-set Γ_0 of Γ not containing β^0, we have by the Theorem 1 of Oberhofer (1982) that there exists $\epsilon > 0$ such that: $n^{-1}\sum_{i=1}^n I\!E[\eta_i(\phi)] \geq \epsilon$. Moreover, using the assumption (H5):

$$\frac{1}{n}\sum_{i=1}^n Var[\eta_i(\beta)] \leq \|\beta-\beta^0\|^2 \frac{1}{n}\sum_{i=1}^n \sup_{\beta\in\Gamma} \left\| \frac{\partial h_\beta(x_i)}{\partial\beta} \right\|^2 \leq C.$$

By the strong law of large numbers:

$$I\!P\left[\left(\liminf_{n\to\infty} \inf_{\|\beta-\beta^0\|\geq \upsilon_n} \frac{1}{n}\sum_{i=1}^n \eta_i(\beta)\right) \geq \epsilon\right] = 1.$$

□

Following two lemmas prove that when the data are from two different models, the LAD estimator in the regression without change-point is close to the parameter of the model from where most of the data came.

Lemma 7 *Let n_1 and n_2 be two integers such that $n_1 \geq n^\rho$ with $1 \geq \rho \geq 3/4$, $n_2 \leq n^\upsilon$ and $\upsilon < 1/4$. Let us also consider the regressions*

$$Y_i = h_{\beta_1^0}(x_i) + \varepsilon_i, \quad i = 1, \ldots, n_1,$$
$$Y_i = h_{\beta_2^0}(x_i) + \varepsilon_i, \quad i = n_1+1, \ldots, n_1+n_2,$$

with $\beta_1^0 \neq \beta_2^0$. We set $N = n_1 + n_2$ and $\hat\beta_N := \arg\min_\beta \sum_{i=1}^N |Y_i - h_\beta(x_i)|$. Under the assumptions (H1)-(H5) and the condition (5), we have:

(i) For every $\delta \in (0, \rho-\upsilon)$, with the probability tending to 1, we have:

$$\|\hat\beta_N - \beta_1^0\| \leq n^{-(\rho-\upsilon-\delta)/2}.$$

(ii) $\sum_{i=1}^{n_1}[|\varepsilon_i - \nu_{(\hat\beta_N,\beta_1^0)}(x_i)| - |\varepsilon_i|] = O_{I\!P}(1)$.

Proof of Lemma 7

(i) By definition:

$$\hat\beta_N = \arg\min_{\beta\in\Gamma}\left[\sum_{i=1}^{n_1}\left[|\varepsilon_i - \nu_{(\beta,\beta_1^0)}(x_i)| - |\varepsilon_i|\right] + \sum_{i=n_1+1}^{n_1+n_2}\left[|\varepsilon_i - \nu_{(\beta,\beta_2^0)}(x_i)| - |\varepsilon_i|\right]\right]. \tag{32}$$

Using (H3), we have that

$$\sum_{i=n_1+1}^{n_1+n_2}\Big[|\varepsilon_i-\nu_{(\beta,\beta_2^0)}(x_i)|-|\varepsilon_i|\Big]\leq\|\beta-\beta_2^0\|\sum_{i=n_1+1}^{n_1+n_2}\sup_{\beta\in\Gamma}\left\|\frac{\partial h_\beta(x_i)}{\partial\beta}\right\|=O(n^\upsilon).$$

The rest of proof for the first term on the right-hand side of relation (32) is similar to that of the Lemma 10 of Bai (1998), using the Lemma 6.

(ii) The proof mimic the proof of the Lemma 10 of Bai (1998). For the clarity we give outline of this proof. This is based on the writing of the right-hand of the relation (32):

$$\sum_{i=1}^{n_1}[|\varepsilon_i-\nu_{(\beta,\beta_1^0)}(x_i)|-|\varepsilon_i|]+\sum_{i=n_1+1}^{n_1+n_2}[|\varepsilon_i-\nu_{(\beta,\beta_2^0)}(x_i)|-|\varepsilon_i|]$$
$$=z_n(\beta)+t_n(\beta)+\sum_{i=n_1+1}^{n_1+n_2}[|\varepsilon_i-\nu_{(\beta_1^0,\beta_2^0)}(x_i)|-|\varepsilon_i|],$$

where $z_n(\beta)=\sum_{i=1}^{n_1}[|\varepsilon_i-\nu_{(\beta,\beta_1^0)}(x_i)|-|\varepsilon_i|]$ and $t_n(\beta)=\sum_{i=n_1+1}^{n_1+n_2}[|\varepsilon_i-\nu_{(\beta,\beta_2^0)}(x_i)|-|\varepsilon_i-\nu_{(\beta_1^0,\beta_2^0)}(x_i)|]$.

For t_n we apply (i), $|t_n(\hat\beta_N)|\leq\sum_{i=n_1+1}^{n_1+n_2}||\varepsilon_i-\nu_{(\hat\beta_N,\beta_2^0)}(x_i)|-|\varepsilon_i-\nu_{(\beta_1^0,\beta_2^0)}(x_i)||\leq\sum_{i=n_1+1}^{n_1+n_2}|\nu_{(\beta_1^0,\hat\beta_N)}(x_i)|=o(1)$. On the other hand: $0\geq z_n(\hat\beta_N)+t_n(\hat\beta_N)\geq\inf_\beta z_n(\beta-|o_{I\!P}(1)|$. Then: $|z_n(\hat\beta_N)|\leq|\inf_\beta z_n(\beta|+o_{I\!P}(1)$ By Lemma 5(i) we have $\inf_\beta z_n(\beta)=O_{I\!P}(1)$ and the claim follows. □

The following three lemmas are necessary for the model with the change-points.

Lemma 8 *Let n_1 and n_2 be two integers as in the Lemma 7. Consider:*

$$Y_i=h_{\beta_1^0}(x_i)+\varepsilon_i,\quad i=1,\ldots,k,$$

$$Y_i=h_{\beta_2^0}(x_i)+\varepsilon_i,\quad i=k+1,\ldots,k+n_2,$$

where $k\in[an_1,n_1]$, $a\in(0,1)$ and $N=k+n_2$. Under the same assumptions as in the Lemma 7, we have

(i) Then $\forall a\in(0,1]$, $\forall\delta\in(0,\rho-\upsilon)$, with the probability tending to 1:

$$\sup_{n_1a\leq k\leq n_1}\|\hat\beta_{k+n_2}-\beta_1^0\|\leq n^{-(\rho-\upsilon-\delta)/2}.$$

(ii)

$$\sup_{n_1a\leq k\leq n_1}\left|\sum_{i=1}^{k}|\varepsilon_i-\nu_{(\hat\beta_{k+n_2},\beta_1^0)}(x_i)|-|\varepsilon_i|\right|=O_{I\!P}(1).$$

Proof of Lemma 8 The proof is similar to that of Lemma 11 of Bai (1998). The only difference is that we use Lemma 6 which is the analoguous for the nonlinear case of the Lemma 5 of Bai (1998). □

Lemma 9 *Under* (H1) *and* (H3), *for all* $\delta > 1/2$, *we have:*

$$\sup_{1\leq j_1\leq j_2\leq n}\left|\inf_{\beta\in\Gamma}\sum_{i=j_1}^{j_2}\left[|\varepsilon_i - \nu_{(\beta,\beta^0)}(x_i)| - |\varepsilon_i|\right]\right| = O_{I\!P}(n^{\delta}).$$

Proof of Lemma 9 The proof is similar to that of the Lemma 3 of Bai (1998). The single place where the linearity intervenes in the proof of the Lemma 3 of Bai (1998) is in the relation which is placed between Eqs. (15) and (16). The equivalent of $\sum_{i=1}^{n}|w_i||s-t|$ is $\sum_{i=1}^{n}|\nu_{(\beta_1,\beta_2)}(x_i)|$, for $\beta_1, \beta_2 \in \Gamma$, which is $O(n^{1/2})$ as consequence of (H3). □

Lemma 10 *Let us consider the sequence* $c_n \to 0$ *such that* $nc_n/\log n \to \infty$ *or* $c_n \equiv c > 0$. *Under the assumptions (H1)–(H3), there exists* $\tilde{C} > 0$ *such that* $\forall\epsilon > 0, \forall n > n_\epsilon$*:*

$$I\!P\left[\sup_{\|\beta-\beta^0\|\leq c_n}\left|\frac{1}{nc_n^2}\sum_{i=1}^{n}\xi_i(\beta)\right|\geq\epsilon\right]\leq\exp\left(-\epsilon^2 nc_n^2\tilde{C}\right).$$

Proof of Lemma 10 Divide the region $\|\beta-\beta^0\|\leq c_n$ into $c_p n^{p/2}$ cells. For β_1, β_2 belonging to a common cell: $\|\beta_1-\beta_2\|\leq c_n n^{-1/2}$, we have:

$$\frac{1}{nc_n^2}\left|\sum_{i=1}^{n}[\xi_i(\beta_1)-\xi_i(\beta_2)]\right|\leq\frac{1}{nc_n^2}\left|\sum_{i=1}^{n}\nu_{(\beta_1,\beta_2)}(x_i)\right|\leq 2n^{-1/2}c_n^{-1}C=o(1).$$

On the other hand: $|\xi_i(\beta)|\leq 2|\nu_{(\beta,\beta^0)}(x_i)|\leq c_n\tilde{C}$ for a constant $\tilde{C}>0$. We apply Lemma 1 for $\xi_i(\beta)$ and $s=1/2$, $b=2c_n\tilde{C}$, $a=nc_n^2\epsilon$, $V=4\tilde{C}^2nc_n^2$. □

Acknowledgments The author would like to thank the three anonymous referees for carefully reading the paper and for their comments which greatly improved the paper.

References

Andrews, D. W. K., Lee, I., Ploberger, W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics, 70*, 9–38.

Babu, G. J. (1989). Strong representations for LAD estimators in linear models. *Probability Theory and Related Fields, 83*, 547–558.

Bai, J. (1995). Least absolute deviation estimation of a shift. *Econometric Theory, 11*, 403–436.

Bai, J. (1998). Estimation of multiple-regime regressions with least absolute deviation. *Journal of Statistical Planning Inference, 74*, 103–134.

Bai, J., Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica, 66*, 47–78.

Carlstein, E., Muller, H. G., Siegmund, D. (1994). *Change-point problems*. Hayward: IMS.

Csōrgo, M., Horváth, L. (1997). *Limit theorems in change-point analysis*. New York: Wiley.

Dielman, T. E. (2005). Least absolute value regression: recent contributions. *Journal of Statistical Computation and Simulation, 75*(4), 263–286.

Epps, T. (1988). Testing that a Gaussian process is stationary. *The Annals of Statistics 16*, 1667–1683.

Garcia, R., Perron, P. (1996). An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics, 78*, 111–125.

Hitomi, K., Kagihara, M. (2001). Calculation method for nonlinear dynamic least-absolute deviations estimator. *Econometric Theory, 7*, 46–68.

Horváth, L., Hušková, M., Serbinowska, M. (1997). Estimators for the time of change in linear models. *Statistics, 29*(2), 109–130.

Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning Inference, 126*(1), 225–251.

Hušková, M., Prášková, Z., Steinebach, J. (2007). On the detection of changes in autoregressive time series. I. Asymptotics. *Journal of Statistical Planning Inference, 137*(4), 1243–1259.

Kim, H. K., Choi, S. H. (1995). Asymptotic properties of non-linear least absolute deviation estimators. *Journal of the Korean Statistical Society, 24*, 127–139.

Kühn, C. (2001). An estimator of the number of change points based on a weak invariance principle. *Statistics and Probability Letters, 51*(2), 189–196.

Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications, 83*, 79–102.

Liu, J., Wu, S., Zidek, J. V. (1997). On segmented multivariate regressions. *Statistic Sinica, 7*, 497–525.

Lombard, F. (1987). Rank tests for change point problems. *Biometrika, 74*, 615–624.

Mia, B., Zhao, L. (1988). Detection of change points using rank methods. *Communications in Statistics. Theory and Methods, 17*, 3207–3217.

Oberhofer, W. (1982). The consistency of nonlinear regression minimizing the L_1-norm. *The Annals of Statistics, 10*(1), 316–319.

Pauler, D. K., Finkelstein, D. M. (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine, 21*(24), 3897–3911

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory, 7*, 186–199.

Richardson, G. D., Bhattacharyya, B. B. (1987). Consistent L_1-estimators in non-linear regression for a noncompact parameter space. *Sankhya:The Indian Journal of Statistics, Series A, 49*, 377–387.

Serbinowska, M. (1996). Consistency of an estimator of the number of changes in binomial observations. *Statistics and Probability Letters, 29*(4), 337–344.

Weiss, A. A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory, 7*, 46–68.

Wu, Y. (1988). Strong consistency and exponential rate of the minimum L_1-norm estimates in linear regression models. *Computational Statistics and Data Analysis, 6*, 285–295.

Yao, Y. C. (1988). Estimating the number of change-points via Schwarz's criterion. *Statistics and Probability Letters, 6*, 181–189.

Yao, Y. C., Au, S. T. (1988). Least-squares estimation of a step function. *Sankhya, 51*, 370–381.