



HAL
open science

A general criterion to determine the number of change-points

Gabriela Ciuperca

► **To cite this version:**

Gabriela Ciuperca. A general criterion to determine the number of change-points. *Statistics and Probability Letters*, 2011, <10.1016/j.spl.2011.03.027>. <hal-00753945>

HAL Id: hal-00753945

<https://hal.science/hal-00753945v1>

Submitted on 20 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Accepted Manuscript

A general criterion to determine the number of change-points

Gabriela Ciuperca

PII: S0167-7152(11)00110-6
DOI: 10.1016/j.spl.2011.03.027
Reference: STAPRO 5962

To appear in: *Statistics and Probability Letters*

Received date: 24 June 2010
Revised date: 14 March 2011
Accepted date: 14 March 2011



Please cite this article as: Ciuperca, G., A general criterion to determine the number of change-points. *Statistics and Probability Letters* (2011), doi:10.1016/j.spl.2011.03.027

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A general criterion to determine the number of change-points

Gabriela CIUPERCA¹

Université Lyon 1, France

Abstract

A general criterion is proposed to determine the number K of the change-points in a parametric nonlinear multi-response model. Schwarz criterion is a particular case. The change-points depend of regressors values and not of measure instant. We prove that the proposed estimator for K is consistent. Simulation results, using Monte Carlo technique, for nonlinear models which have numerous applications, support the relevance of the theory.

keywords: selection criterion, multiple change-points, M-estimator, consistency.

1 Introduction

This paper continues the article of Ciuperca (2009) where the properties of the M-estimator in a multi-phase random nonlinear model are studied when the number of change-points is fixed(known). The consistency, convergence rate, asymptotic distribution of regression parameters and change-points M-estimators were obtained. In this paper we propose a information criterion,

¹ Université de Lyon, Université Lyon 1, UMR 5208, blvd du 11 novembre 1918, F - 69622 Villeurbanne Cedex, France

that is a generalization of Schwarz criterion, for the choice of the number of breaks in a nonlinear random model, using M-estimation method in order to estimate the model parameters. The maximum likelihood (ML), least squares (LS) are particular cases of estimation method. Once the number of change-points is determined, their locations and the regression parameters on the segments defined by the change-points can be estimated using the results of Ciuperca (2009).

The interest of the proposed criterion is that the change-points τ_j depend on the regressors X (i.e. $\tau_j < X \leq \tau_{j+1}$), which was no more handled in the literature for the nonlinear parametric models. What was previously proposed are criteria for models with change τ_j in the time (i.e. $\tau_j < i \leq \tau_{j+1}$). From practical view point this is a different matter: the changes in the model are due to the values of the regressor X and not to the measure instant.

The BIC (Schwarz) criterion for a model without change-points is: $BIC = -2 \log(L_n(\hat{\theta}_n)) + \dim(\hat{\theta}_n) \log n$, where $L_n(\theta)$ is likelihood function and $\hat{\theta}_n$ is the maximizer of $L_n(\theta)$. If X_1, \dots, X_n is a sequence of independent random variables with a change in their density: $g(x, \theta_1)$ density of X_i when $i \leq k$ and $g(x, \theta_2)$ for $i > k$, the Schwarz information criterion is: $BIC(k) = -2 \log L_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) + [2\dim(\theta_1) + 1] \log n$, where $\hat{\theta}_{1k}, \hat{\theta}_{2k}$ maximize $L_n(\theta_1, \theta_2, k)$ for given k . Chen et al. (2006) consider the following criterion to detect the change in a sequence of independent random variables: $S(n) = MIC(n) - \min_{1 \leq k \leq n} MIC(k) + \dim(\theta) \log n$, with $MIC(k) = -2 \log L_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) + \left[2\dim(\theta_1) + \left(\frac{2k}{n} - 1 \right)^2 \right] \log n$, $MIC(n) = -2 \log L_n(\hat{\theta}_n, \hat{\theta}_n, n) + \dim(\theta) \log(n)$, $\hat{\theta}_n = \arg \max_{\theta} L_n(\theta, \theta, n)$. This criterion was generalized by Pan and Chen (2006) to detect multiple changes in the density of a sequence of independent random variables. In the case of a parametric model with change-points we can give the following references. For a constant model with K change-

points, a consistent estimator for K is proposed by Yao and Au (1989), using LS estimation method. Bai (1999) proposes a test based on the likelihood for a linear model. Always to detect a change in a linear model we can remind papers of Osorio and Galea (2005), Wu (2008) or still Nosek (2010). For a parametric nonlinear model, to the author's knowledge, the only available criterion is the one of Ciuperca (2011a) but for deterministic instant of measure: $Y_i = h_{\beta_1}(X_i)\mathbb{1}_{\tau_1 < i} + h_{\beta_2}(X_i)\mathbb{1}_{\tau_1 < i \leq \tau_2} + \dots + h_{\beta_{K+1}}(X_i)\mathbb{1}_{i > \tau_K}$. A modified Schwarz criterion is proposed using L_1 method for estimation. After, the method is ameliorated in Ciuperca (2011b) by penalization.

This paper is organized as follows. Model, assumptions, criterion and its consistency are presented in Section 2. Section 3 reports some simulation results in order to illustrate the theoretical results. The proofs of results are given in Section 4.

2 Main results

Consider the nonlinear random design model with K change-points:

$$Y_i = f_{(\theta_{1,K}, \theta_{2,K})}(X_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where $(X_i, \varepsilon_i)_{1 \leq i \leq n}$ is a sequence of continuous independent random variables with the same joint distribution as (X, ε) . Function $f_{(\theta_{1,K}, \theta_{2,K})}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is defined, for $x \in \mathbb{R}$, by:

$$f_{(\theta_{1,K}, \theta_{2,K})}(x) = h_{\beta_1}(x)\mathbb{1}_{x \leq \tau_1} + h_{\beta_2}(x)\mathbb{1}_{\tau_1 < x \leq \tau_2} + \dots + h_{\beta_{K+1}}(x)\mathbb{1}_{x > \tau_K},$$

with the functions h_{β_k} known up to the parameters β_k . Therefore, the instant of measure coincides with the regressor X which is random. For every

$k = 1, \dots, K + 1$, we suppose that the parameter β_k belongs to some compact $\Gamma \subseteq \mathbb{R}^p$, $p \geq 1$. Let us denote the regression parameters $\theta_{1,K} = (\beta_1, \dots, \beta_{K+1})$ and the change-points $\theta_{2,K} = (\tau_1, \dots, \tau_K)$ with $\tau_1 < \dots < \tau_K$. We consider that the vector $\theta_{2,K} \in \mathbb{R}^K$ and we set $\theta_K = (\theta_{1,K}, \theta_{2,K}) \in \Omega = \Gamma^{K+1} \times \mathbb{R}^K$. In the case $K = 0$ the model is without change-point. For $\theta_K \in \Omega$, let us denote by $\dot{f}_{\theta_K}(x) = \partial f_{\theta_K}(x) / \partial \theta_{1,K}$. If the number of change-points is known ($K = K_0$), in Ciuperca (2009)'s paper, under following assumptions for ε , X and function h , the parameter θ_{K_0} is estimated by M-estimation method when n observations of (Y, X) are available.

(A1) X has a density absolutely continuous and positive everywhere on \mathbb{R} . Moreover, $\mathbb{E}(X^2) < \infty$;

(A2) ε has a density absolutely continuous and positive everywhere on \mathbb{R} . Moreover, $\mathbb{E}(\varepsilon) = 0$, $\mathbb{E}(\varepsilon^2) < \infty$;

(A3) the random variables X_i and ε_i are independent.

For the true value K_0 of K , to simplify notations, we denote by $\theta_1^0 = (\beta_1^0, \beta_2^0, \dots, \beta_{K_0+1}^0)$ and $\theta_2^0 = (\tau_1^0, \dots, \tau_{K_0}^0)$, respectively, the true values of the regression parameters and of the change-points. Let be also $\theta^0 = (\theta_1^0, \theta_2^0)$.

Nonlinear function h_β satisfies the conditions:

(B1) for all $x \in \mathbb{R}$, $h_\beta(x)$ is three times differentiable with respect to β ;

(B2) for all $x \in \mathbb{R}$, $k \in \{1, \dots, K_0, K_0 + 1\}$, $\|\partial h_{\beta_k}(x) / \partial \beta\| \neq 0$;

(B3) the derivatives $\partial^3 h_\beta(x) / \partial \beta^3$, exist for $x \in \mathbb{R}$ and there exists the functions $F_0, F_1, F_2 \in L^2(\varphi)$ such that: $\sup_{\beta \in \Gamma} |h_\beta(x)| \leq F_0(x)$, $\sup_{\beta \in \Gamma} \|\partial^j h_\beta(x) / \partial \beta^j\| \leq F_j(x)$, $j = 1, 2$.

We suppose that model (1) make a non-zero jump at each true change-point τ_k^0 : $d_{(\beta_{k+1}, \beta_k)}(\tau_k^0) \neq 0$, $\forall \beta_k, \beta_{k+1} \in \Gamma, \beta_k \neq \beta_{k+1}, k \in \{1, \dots, K_0\}$, where we denoted $d_{(\beta_k, \beta_j)}(x) = h_{\beta_k}(x) - h_{\beta_j}(x)$, $x \in \mathbb{R}, \beta_k, \beta_j \in \Gamma$.

In practice, the number of breaks K is unknown. In order to estimate the

true value K_0 and afterwards the regression parameters, the location of the change-points, let us consider the function $\rho : \mathbb{R} \rightarrow \mathbb{R}_+$. Associated M-process for model (1) is defined as: $M_n(\theta) = \sum_{i=1}^n \rho(Y_i - f_\theta(X_i))$, with the following assumptions for function ρ :

(C1) ρ is convex on \mathbb{R} with right-continuous non-decreasing almost everywhere derivative ψ satisfying $\mathbb{E}_\varepsilon[\psi^2(\varepsilon + y)] < \infty, \forall y \in \mathbb{R}$. The function $\lambda(y) = \mathbb{E}_\varepsilon[\psi(\varepsilon + y)], y \in \mathbb{R}$, is strictly increasing on \mathbb{R} and λ is continuous at 0 with $\lambda(0) = 0$.

(C2) for K fixed, for all $c \in \mathbb{R}$, $\mathbb{E}_{(\varepsilon, X)}[\psi^2(\varepsilon + c \sup_{\theta_K, \theta_K^* \in \bar{\Omega}} (f_{\theta_K}(X) - f_{\theta_K^*}(X)))] < \infty$, where $\bar{\Omega}$ is the closure of Ω .

(C3) the function $y \rightarrow \mathbb{E}[|\psi(\varepsilon + c + y) - \psi(\varepsilon)|]$ is continuous at 0, $\forall c \in \mathbb{R}$.

(C4) the function λ is differentiable in a neighborhood of 0, with derivative λ' satisfying $\lambda'(0) \neq 0$, and $\lim_{a \rightarrow 0} a^{-1} \int_0^a |\lambda'(s) - \lambda'(0)| ds = 0$.

(C5) the random variables $\rho(\varepsilon \pm d_{(\beta_{k+1}^0, \beta_k^0)}(\tau_k^0)) - \rho(\varepsilon), \forall k = 1, \dots, K_0$, are continuous.

(C6) the function ψ is differentiable on \mathbb{R} .

For a fixed K , the M-estimator is defined by: $\hat{\theta}_{n,K} = (\hat{\theta}_{1n,K}, \hat{\theta}_{2n,K}) = \arg \min_{\theta_K \in \bar{\Omega}} M_n(\theta_K)$ a.s. For the true value K_0 of K , under assumptions (A1)-(A3), (B1)-(B3), (C1)-(C6), $\hat{\theta}_{n,K_0}$ is strongly consistent. The change-points M-estimator $\hat{\theta}_{2n,K_0}$ converges with n^{-1} rate to the vector whose components are the left end points of the minimizing interval of independent compound Poisson process (see Ciuperca (2009)). The regression parameters M-estimator $\hat{\theta}_{1n,K_0}$ has a Gaussian asymptotic distribution.

We shall give a criterion to find the number of change-points, under the condition that, for K fixed, two breaking points are a far distance:

$$\mathcal{C}_{k,n} = \text{Card}\{i; \tau_{k-1} < X_i \leq \tau_k\}, \quad \min_{1 \leq k \leq K+1} \mathcal{C}_{k,n} = A_n, \quad A_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \infty, \quad (2)$$

with $\tau_0 = -\infty$ and $\tau_{K+1} = \infty$. For every fixed change-points number K , consider the sum:

$$S(\tau_1, \dots, \tau_K) = \inf_{\theta_{1,K} \in \Gamma^{K+1}} \sum_{i=1}^n \sum_{r=1}^{K+1} \rho(Y_i - f_{(\theta_{1,K}, \theta_{2,K})}(X_i)) \mathbb{1}_{\tau_{r-1} < X_i \leq \tau_r} \quad (3)$$

and corresponding the change-point M-estimator: $(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}) = \arg \min_{\theta_{2,K} \in \mathbb{R}} S(\theta_{2,K})$.

Following lemma gives the behaviour of (3) when the model is estimated supposing that there is no change-points but in reality there is a point of break.

Lemma 2.1 *We suppose that $K_0 = 1$. Under assumptions (A1)-(A3), (B1), (B3), (C1), (C2), (C4), if $K = 0$ then we have with arbitrarily large probability for all sufficiently large n :*

$$\inf_{\beta \in \Gamma} \left[\sum_{i=1}^n \rho(\varepsilon_i + d_{(\beta_1^0, \beta)}(X_i)) \mathbb{1}_{X_i \leq \tau_1^0} + \sum_{i=1}^n \rho(\varepsilon_i + d_{(\beta_2^0, \beta)}(X_i)) \mathbb{1}_{X_i > \tau_1^0} \right] - \sum_{i=1}^n \rho(\varepsilon_i) \geq CA_n, \text{ with } C \text{ a positive constant.}$$

We consider an additional assumption, needed for the estimator consistency of the change-points number.

$$(C7) \quad 0 < \mathbb{E}[\rho^2(\varepsilon)] < \infty.$$

This condition, together Bienaymé-Tchebychev inequality, imply that random variable $\rho(\varepsilon)$ is bounded with probability close to one: $\forall \gamma > 0, \exists \gamma_1 > 0$ such that $\mathbb{P}[\rho(\varepsilon) > \gamma_1] < 1 - \gamma$.

Particular cases. 1) If $\rho(y) = y^2$ we obtain the criterion for the LS estimator. When a linear model is considered $h_\beta(x) = \beta^t x$, assumption (C7) holds if $\mathbb{E}(\varepsilon^4) < \infty$.

2) If $\rho(y) = -\log \varphi_\varepsilon(y)$, the ML method is considered. If for example error density is $\varphi_\varepsilon(y) = \exp(-U(y))$, with U a function $U : \mathbb{R} \rightarrow \mathbb{R}$, then if $\text{Var}(U(\varepsilon)) < \infty$, assumption (C7) holds. Example, other one than the normal distribution: double exponential(Laplace) density $\varphi_\varepsilon(y) = \exp(-|y|)$.

In order to find the change-points number, let us consider the criterion, that

we are going to call M-criterion:

$$B(K) = n \log \hat{s}_K + G(K, p)C_n, \quad (4)$$

with $\hat{s}_K = n^{-1}S(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K})$, (C_n) any sequence satisfying $C_n \rightarrow \infty$ and $(C_n/A_n) \rightarrow 0$ as $n \rightarrow \infty$. Assumption (C7) is necessary so that \hat{s}_{K_0} is bounded: $\forall \epsilon > 0, \exists M_\epsilon > 0$ such that $\mathbb{P}[\hat{s}_{k_0} \geq M_\epsilon] \leq \epsilon$. Function G is such that $G(K_1, p) > G(K_2, p)$ for $K_1 > K_2$. This function depend of p to take into account the model complexity. Remind that p is the dimension of space Γ . The criteria proposed by Yao and Au (1989), Nosek (2010) are particular cases, using LS method estimation and $G(K, p)$ depending only of K . Other particular cases are in Wu (2008), Osorio and Galea (2005), to test the presence of a change-point in a linear model always by LS method, respectively ML method, with G depending of p .

Let \hat{K}_n be the value of K that minimizes criterion $B(K)$:

$$\hat{K}_n = \arg \min_{0 \leq K \leq K_U} B(K),$$

with K_U the possible upper bound for K . Next result shows the consistency of estimator \hat{K}_n .

Theorem 2.1 *Under assumptions (A1)-(A3), B1)-(B3), (C1)-(C7), we have $\mathbb{P}[\hat{K}_n = K_0] \rightarrow 1$ for $n \rightarrow \infty$.*

3 Simulations

In order to illustrate the theoretical result, a simulation study is realized using R language. The programs in R for computing \hat{K}_n and the M-estimations of the parameters are available to author. The M-criteria $B(0)$, $B(1)$, $B(2)$, $B(3)$

are calculated using LS method and $C_n = \log n$. For each model, 1000 Monte Carlo samples of size n are generated for regressor X and error ε .

3.1 Nonlinear models

Let us consider growth function, or more exactly the mono-molecular model (see Seber and Wild (2003)): $h_\beta(x) = b_1 - \exp(-b_2x)$ with $\beta = (b_1; b_2)$. Then $p = 2$. We generate M replications of size $n = 500$ for regressor $X \sim \mathcal{N}(1, 1)$ and for errors $\varepsilon \sim \mathcal{N}(0, 0.5)$ or Laplace $\varepsilon \sim \mathcal{L}(0, 1)$. The model have a change-point ($K_0 = 1$) in $\tau_1^0 = 1$:

$$Y_i = h_{\beta_1^0}(X_i)\mathbb{1}_{X_i < 1} + h_{\beta_2^0}(X_i)\mathbb{1}_{X_i \geq 1} + \varepsilon_i, \quad i = 1, \dots, n \quad (5)$$

with $\beta_1^0 = (0.5, 1)$, $\beta_2^0 = (1, -0.5)$. For the criterion $B(K)$ of (4) we consider $G(K, p) = K(K + 1)p$. In Figures 1 and 2 a simulation of this model is plotted for $\varepsilon \sim \mathcal{N}(0, 0.5)$ and $\mathcal{L}(0, 1)$, respectively. In order to point out the change in the model, we represented by '•' the values of Y on the second segment and by '△' the values of Y without change-point. For the 1000 replications of the model with the size of sample $n = 500$ we obtain that $\min_{j \in \{0, 1, 2\}} B(j) = B(1)$ in both cases $\varepsilon \sim \mathcal{N}(0, 0.5)$ or $\varepsilon \sim \mathcal{L}(0, 1)$.

On the other hand, if we consider small sample sizes $n=50$, always for $G(K, p) = K(K + 1)p$, the number of change-points is very well detected in the case of Normal errors, but, due to the presence of outliers, the results are worst for Laplace errors (see Table 1).

Another very interesting nonlinear model, with numerous applications, is the compartmental model. Examples and references of important applications for these models are given in Seber and Wild (2003) (see also the references therein): it describes the movement of lead in the human body, the kinetics

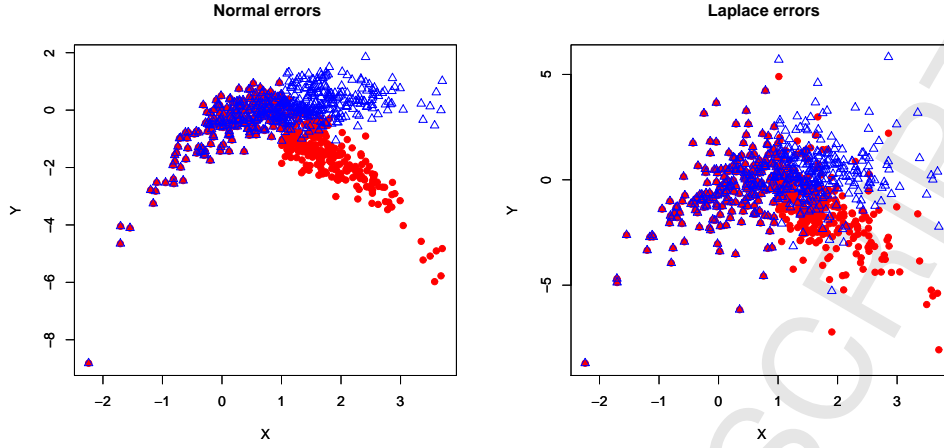


Fig. 1. A simulation for growth model with a change-point, for $\varepsilon_i \sim \mathcal{N}(0, 0.5)$. Fig. 2. A simulation for growth model with a change-point, for $\varepsilon_i \sim \mathcal{L}(0, 1)$.

Table 1

Model (5), $n=500$ or $n=50$, $M=1000$ Monte Carlo replications.

	K=0		K=1		K=2	
$n =$	500	50	500	50	500	50
$\varepsilon \sim \mathcal{N}(0, 0.5)$	0	5	1000	995	0	0
$\varepsilon \sim \mathcal{L}(0, 1)$	0	447	1000	553	0	0

of drug movement when the drug is injected at an intramuscular site, etc...

Consider two-compartment function: $h_\beta(x) = b_1 \exp(-b_1 x) + b_2 \exp(-b_2 x)$,

$\beta = (b_1, b_2) \in \Gamma \subseteq \mathbb{R}^2$ ($p = 2$). We first generate samples of size $n = 500$ for $X \sim \mathcal{N}(1, 1)$, $\varepsilon \sim \mathcal{N}(0, 0.5)$ for a model with a change-point:

$$Y_i = h_{\beta_1^0}(X_i) \mathbb{1}_{X_i < 1} + h_{\beta_2^0}(X_i) \mathbb{1}_{X_i > 1} + \varepsilon_i, \quad i = 1, \dots, n,$$

for $\beta_1^0 = (1.2, 1)$, $\beta_2^0 = (-0.5, 2)$ (see Figure 3, where ' Δ ' is for model without change-point and ' \bullet ' for model with a change in $\tau_1^0 = 1$). As the theory predicts, for all 1000 replications we obtain $\min_{j \in \{0, 1, 2\}} B(j) = B(1)$.

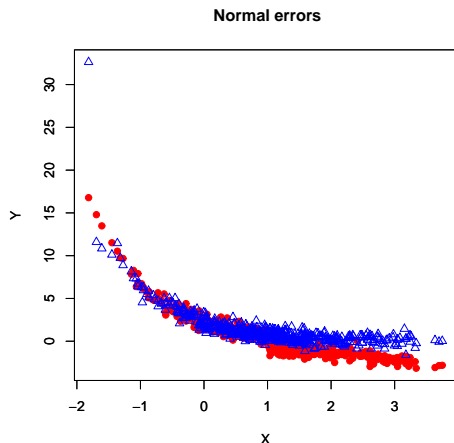


Fig. 3. *Compartmental model with one change-point, Gaussian error.*

When we consider small sample sizes, $n=50$, the M-criterion detects 88% of cases one change-points and in 12% of cases it chooses a model without change-points.

3.2 Linear model

We simulate samples of size $n = 150$ for following linear model with two change-points:

$$Y_i = (a_1^0 + b_1^0 X_i) \mathbb{1}_{X_i \leq 0} + (a_2^0 + b_2^0 X_i) \mathbb{1}_{0 < X_i < 2} + (a_3^0 + b_3^0 X_i) \mathbb{1}_{2 \leq X_i} + \varepsilon_i \quad (6)$$

with $X_i \sim \mathcal{N}(1, 1)$, $\varepsilon_i \sim \mathcal{N}(0, 0.5)$ or $\varepsilon_i \sim \mathcal{L}(0, 1)$. The true values of parameters are: $(a_1^0, b_1^0) = (-1, 1)$, $(a_2^0, b_2^0) = (1, 2.5)$, $(a_3^0, b_3^0) = (2, 3)$, $\tau_1^0 = 0$, $\tau_2^0 = 2$. We consider $G(K, p) = Kp$ or $G(K, p) = K(K+1)p$, with $p = 2$. In Table 2 we report the occurrence number when every number $K \in \{0, 1, 2, 3\}$ was chosen by the criterion, when ε is Gaussian or Laplace for model (6). We observe that if the errors are Gaussian, the change-points number is very well estimated,

Table 2

Model (6), occurrence numbers for each K , $G_1 = Kp$, $G_2 = K(K + 1)p$, $n=150$.

$G =$	K=0		K=1		K=2		K=3	
	G_1	G_2	G_1	G_2	G_1	G_2	G_1	G_2
$\varepsilon \sim \mathcal{N}(0, 0.5)$	0	0	0	0	975	1000	5	0
$\varepsilon \sim \mathcal{L}(0, 1)$	0	0	9	96	972	904	19	0

Table 3

Model (6), occurrence numbers for each K , $G_1 = Kp$, $G_2 = K(K + 1)p$, $n=50$.

$G =$	K=0		K=1		K=2		K=3	
	G_1	G_2	G_1	G_2	G_1	G_2	G_1	G_2
$\varepsilon \sim \mathcal{N}(0, 0.5)$	0	0	0	59	987	941	13	0
$\varepsilon \sim \mathcal{L}(0, 1)$	0	0	23	847	945	153	22	0

while for Laplace errors, the estimation is less good due to a bigger variability of data, especially in the case of function $G(K, p) = K(K + 1)p$.

This tendency becomes more marked when we consider small sample size: $n=50$ (see Table 3). Nevertheless, the results remain satisfactory since we detect correctly more than 9/10 change-points number for $G(K, p) = Kp$.

Consider now a linear model with one change-point:

$$Y_i = (a_1^0 + b_1^0 X_i) \mathbb{1}_{X_i \leq 1} + (a_2^0 + b_2^0 X_i) \mathbb{1}_{1 < X_i} + \varepsilon_i, \quad i = 1, \dots, n \quad (7)$$

with $X_i \sim \mathcal{N}(1, 1)$, $\varepsilon_i \sim \mathcal{N}(0, 0.5)$ or $\varepsilon_i \sim \mathcal{L}(0, 1)$ and different values for the parameters $\beta_1^0 = (a_1^0, b_1^0)$, $\beta_2^0 = (a_2^0, b_2^0)$. For $\beta_1^0 = (-2, 1)$ and $\beta_2^0 = (1, 0)$ we compare our results with the results obtained by a Schwarz information

Table 4

Model (7), occurrence numbers for each K , $G_1 = Kp$, $G_2 = K(K + 1)p$, $n=50$.

$G =$	K=0		K=1		K=2		K=3	
	G_1	G_2	G_1	G_2	G_1	G_2	G_1	G_2
$\varepsilon \sim \mathcal{N}$, $\beta_1^0 = (-2, 1)$, $\beta_2^0 = (1, -2.5)$	0	0	638	1000	358	0	4	0
$\varepsilon \sim \mathcal{N}$, $\beta_1^0 = (-2, 10)$, $\beta_2^0 = (1, -2.5)$	0	0	700	1000	295	0	5	0
$\varepsilon \sim \mathcal{N}$, $\beta_1^0 = (-2, 1)$, $\beta_2^0 = (1, 0)$	0	0	625	1000	369	0	6	0
$\varepsilon \sim \mathcal{L}$, $\beta_1^0 = (-2, 1)$, $\beta_2^0 = (1, -2.5)$	2	94	580	905	417	1	1	0
$\varepsilon \sim \mathcal{L}$, $\beta_1^0 = (-2, 10)$, $\beta_2^0 = (1, -2.5)$	0	0	625	1000	374	0	1	0
$\varepsilon \sim \mathcal{L}$, $\beta_1^0 = (-2, 1)$, $\beta_2^0 = (1, 0)$	110	630	445	370	440	0	5	0

criterion(SIC) proposed in the paper of Nosek (2010).

For this, we calculate: $\hat{\sigma}_T^2 = n^{-1} \sum_{i=1}^T (Y_i - \hat{b}_T X_i - \hat{a}_T)$ with \hat{a}_T et \hat{b}_T the coefficients of the straight line fitted to the first T points by the least squares method. Consider also: $SIC(n) = -n/2 \log(2\pi) - n/2 \log(\hat{\sigma}_n^2) - n/2 - 3/2 \log n$, $SIC(T) = -n/2 \log(2\pi) - n/2 \log(\hat{\sigma}_T^2) - n/2 - (3+n-T)/2 \log n$. If there exists $T \in \{2, \dots, n-1\}$ such that $SIC(n) < SIC(T)$ then the model have a change-point. Remark that the SIC criterion proposed by Nosek (2010) supposes that the errors $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, the regressors are deterministic and especially the change-point depends of observation i not of values of X .

In Table 4, we give the results for $n = 50$ and three values for the regression parameters. In the third case, for Normal errors, the SIC criterion gives only 185/1000 one change-points and for Laplace errors it detects 1000/1000 one change-point. More the parameters of the two segments are different, better

Table 5

Model (7), occurrence numbers for each K , $G_1 = Kp$, $G_2 = K(K + 1)p$, $n=150$.

$G =$	K=0		K=1		K=2		K=3	
	G_1	G_2	G_1	G_2	G_1	G_2	G_1	G_2
$\varepsilon \sim \mathcal{N}(0, 0.5)$	0	0	756	1000	234	0	10	0
$\varepsilon \sim \mathcal{L}(0, 1)$	0	0	686	1000	300	0	16	0

M-criterion detects the number of the change-points. On the other hand, the function $G(K, p) = K(K + 1)p$ gives generally better results (except in the last example for Laplace errors).

For $\beta_1^0 = (-1, 1)$, $\beta_2^0 = (1, -2.5)$, consider now $n = 150$. Results by M-criterion are given in Table 5. Since there are more observations, the M-criterion for $G(K, p) = Kp$ better detects the number of change-points than in the case $n = 50$. For $G(K, p) = K(K + 1)p$ the results are excellent in both cases of laws for ε . When $\varepsilon \sim \mathcal{N}(0, 0.5)$, the SIC criterion chooses 23/1000 a model with one change-point and when $\varepsilon \sim \mathcal{L}(0, 1)$ it chooses 1000/1000. The fact that there are more observations has a negative effect on the SIC criterion.

Consider now, a model without change-point:

$$Y_i = (a^0 + b^0 X_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (8)$$

for $\beta^0 = (a^0, b^0) = (-2, 1)$, $G(K, p) = Kp$ or $G(K, p) = K(K + 1)p$. We compare our results with SIC criterion. For $n = 50$, the results are in Table 6 and for $n = 150$ in Table 7. For $\varepsilon \sim$ Normal or Laplace, for all 1000 Monte-Carlo replications, the M-criterion chooses a model without change-points.

Table 6

Model (8), occurrence numbers for each K , $G_1 = Kp$, $G_2 = K(K + 1)p$, $n=50$.

$G =$	K=0		K=1		K=2		K=3	
	G_1	G_2	G_1	G_2	G_1	G_2	G_1	G_2
$\varepsilon \sim \mathcal{N}$, M-crit.	1000	1000	0	0	0	0	0	0
$\varepsilon \sim \mathcal{N}$, SIC		994		6		–		–
$\varepsilon \sim \mathcal{L}$, M-crit.	1000	1000	0	0	0	0	0	0
$\varepsilon \sim \mathcal{L}$, SIC		0		1000		–		–

Table 7

Model (8), occurrence numbers for each K , $G_1 = Kp$, $G_2 = K(K + 1)p$, $n=150$.

$G =$	K=0		K=1		K=2		K=3	
	G_1	G_2	G_1	G_2	G_1	G_2	G_1	G_2
$\varepsilon \sim \mathcal{N}$, M-crit.	900	1000	105	0	45	0	0	0
$\varepsilon \sim \mathcal{N}$, SIC		1000		0		–		–
$\varepsilon \sim \mathcal{L}$, M-crit.	750	1000	95	0	135	0	20	0
$\varepsilon \sim \mathcal{L}$, SIC		0		1000		–		–

3.3 Conclusion on the simulations

The simulation results show that, for large sample size ($n = 500$ for nonlinear models and $n = 150$ for linear models) the proposed M-criterion detects very well the change-points number, when the errors are Normal or Laplace. When we consider small sample size, $n = 50$, for nonlinear models, K_0 is well

estimated in the case of Normal errors, but, due to the presence of outliers, the results are less good for Laplace errors. For linear model, simulations were realized for $K_0 = 0, 1, 2$, function $G(K, p) = Kp$ or $G(K, p) = K(K + 1)p$, $n = 150$ or $n = 50$ and different values of β and τ . The results are not affected by the number of observations when $G(K, p) = K(K + 1)p$, but they slightly degrade when the model presents outliers. From Tables 2 and 3, remark that for small sample sizes, there are few observations such that $X_i > 2 = \tau_2^0$, then result of M-criterion is less good. This is not the case when $K_0 = 1$ and $\tau_1^0 = 0$ (see Tables 4 and 5). The best choice for G is $G(K, p) = K(K + 1)p$, for nonlinear but also for linear models.

Other very important conclusion is that the M-criterion outperforms the SIC criterion proposed by Nosek (2010). This last one prefers to choose systematically a model without change-points when the errors are Gaussian and a model with one change-point when the errors are Laplace. A possible explanation would be that the change-point depends on random design X_i while the SIC criterion proposed by Nosek (2010) is studied under assumption that it depends of the number of observation i .

4 Appendix: proofs

We give proofs for the results in Section 2. In the followings, we denote by C a generic positive finite constant not dependent on n , that may changes from line to line.

Proof of Lemma 2.1 In this case $\theta^0 = ((\beta_1^0, \beta_2^0); \tau_1^0) = (\theta_1^0; \theta_2^0)$. Let $\hat{\beta}_n$ be the value of β that minimizes the sum of S given by (3) for $K = 0$. Since $\beta_1^0 \neq \beta_2^0$, there exists a constant $c > 0$ such that: $\|\hat{\beta}_n - \beta_1^0\| > c$ or $\|\hat{\beta}_n - \beta_2^0\| > c$ with

probability 1. Without loss of generality, we take $\|\hat{\beta}_n - \beta_2^0\| > c$.

For $x \in \mathbb{R}$ let us consider the function: $l(x) = \mathbb{E}[\rho(\varepsilon + x) - \rho(\varepsilon)]$ positive for all $x \neq 0$. Then the function $e(\beta) = \mathbb{E}[l(d_{(\beta_2^0, \beta)}(X))]$ > 0 for all $\beta \neq \beta_2^0$. Under (A1), (B1), (B3), (C2), continuity on Γ of function e (see Lemma 3.1 of Ciuperca (2009)) and compactness of Γ , by an argument similar to the one used in the proof of Theorem 3.1 in Koul et al. (2003), we obtain that there exists a neighborhood $\mathcal{V}(\beta_2^0)$ of β_2^0 such that: $\inf_{\beta \in \mathcal{V}^c(\beta_2^0)} e(\beta) = \delta_0 > 0$. Hence, for all $\beta \in \mathcal{V}(\beta_2^0)$, there exists a neighborhood $\mathcal{V}(\beta)$ of β such that: $\mathbb{E} \left[\inf_{\beta^* \in \mathcal{V}(\beta)} [\rho(\varepsilon + d_{(\beta_2^0, \beta^*)}(X)) - \rho(\varepsilon)] \right] \geq 2\delta_0$. Thus, there exists a finite number m , such that $\beta_j \in \mathcal{V}^c(\beta_2^0)$, $j = 1, \dots, m$ and by the strong law of large number, we have almost surely for sufficiently large n :

$$\inf_{\beta \in \mathcal{V}(\beta_j)} \frac{1}{\mathcal{C}_{2,n}^0} \sum_{i=1}^n [\rho(\varepsilon_i + d_{(\beta_2^0, \beta)}(X_i)) - \rho(\varepsilon_i)] \mathbb{1}_{X_i > \tau_1^0} \geq \mathbb{E} \left[\inf_{\beta \in \mathcal{V}(\beta_j)} [\rho(\varepsilon + d_{(\beta_2^0, \beta)}(X)) - \rho(\varepsilon)] \right] - \delta_0$$

$\geq \delta_0$, where $\mathcal{C}_{2,n}^0 = \sum_{i=1}^n \mathbb{1}_{X_i > \tau_1^0}$. Thus, using also assumption (2):

$$\sum_{i=1}^n [\rho(\varepsilon_i + d_{(\beta_2^0, \hat{\beta}_n)}(X_i)) - \rho(\varepsilon_i)] \mathbb{1}_{X_i > \tau_1^0} \geq O_{\mathbb{P}}(\mathcal{C}_{2,n}^0) \geq A_n. \quad (9)$$

If $\|\beta_1^0 - \hat{\beta}_n\| > c$, with c a positive constant, then, by a similar proof of (9) we have: $\sum_{i=1}^n [\rho(\varepsilon_i + d_{(\beta_1^0, \hat{\beta}_n)}(X_i)) - \rho(\varepsilon_i)] \mathbb{1}_{X_i \leq \tau_1^0} \geq CA_n$. If $\|\beta_1^0 - \hat{\beta}_n\| = o_{\mathbb{P}}(1)$, by Taylor's expansion combined with (B3), (C1), (C4), we have :

$$\left| \sum_{i=1}^n [\rho(\varepsilon_i + d_{(\beta_1^0, \hat{\beta}_n)}(X_i)) - \rho(\varepsilon_i)] \mathbb{1}_{X_i \leq \tau_1^0} \right| \leq \|\hat{\beta}_n - \beta_1^0\| \sum_{i=1}^n \|h_{\beta_1^0}(X_i)\| \psi(\varepsilon_i) \mathbb{1}_{X_i \leq \tau_1^0} (1 + o_{\mathbb{P}}(1))$$

$= o_{\mathbb{P}}(A_n)$. Lemma follows. ■

Proof of Theorem 2.1 We will show that for $K \neq K_0$, $\mathbb{P}[\hat{K}_n = K] \rightarrow 0$ as $n \rightarrow \infty$. We will prove the Theorem in two steps: $\mathbb{P}[\hat{K}_n < K_0] \rightarrow 0$ and $\mathbb{P}[\hat{K}_n > K_0] \rightarrow 0$ as $n \rightarrow \infty$. Let us set $S_0 = \sum_{i=1}^n \rho(\varepsilon_i)$.

Case $K < K_0$. In this case there exists at least a true change-point τ_k^0 which

cannot be estimated. We consider the $j \in \{1, \dots, K\}$ such that $\hat{\tau}_{j,K}$ and $\hat{\tau}_{j+1,K}$ are the most close of τ_k^0 . Then $\sum_{i=1}^n \mathbb{1}_{\hat{\tau}_{j,K} < X_i \leq \tau_k^0} > A_n/2$ and $\sum_{i=1}^n \mathbb{1}_{\tau_k^0 < X_i \leq \hat{\tau}_{j+1,K}} > A_n/2$ with probability 1. By a similar argument used in the proof of Lemma 2.1, we obtain:

$$\inf_{\beta \in \Gamma} \left[\sum_{i=1}^n \rho(\varepsilon_i + d_{(\beta_k^0, \beta)}(X_i)) \mathbb{1}_{\hat{\tau}_{j,K} < X_i \leq \tau_k^0} + \sum_{i=1}^n \rho(\varepsilon_i + d_{(\beta_{k+1}^0, \beta)}(X_i)) \mathbb{1}_{\tau_k^0 < X_i \leq \hat{\tau}_{j+1,K}} \right] - \sum_{i=1}^n \rho(\varepsilon_i) \mathbb{1}_{\hat{\tau}_{j,K} < X_i \leq \hat{\tau}_{j+1,K}} \geq CA_n,$$

almost surely for sufficiently large n , and in general:

$$S(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}) - S_0 \geq CA_n. \quad (10)$$

Let us now study: $S(\hat{\tau}_{1,K_0}, \dots, \hat{\tau}_{K_0,K_0}) - S_0 = \min_{(\theta_{1,K_0}, \theta_{2,K_0})} [M_n(\theta_{1,K_0}, \theta_{2,K_0}) - M_n(\theta_1^0, \theta_2^0)]$. Since the convergence rate of the M-estimators is $n^{-1/2}$ for regression parameters estimator and n^{-1} for change-points estimator, using Lemma 3.3 and relation (11) of Ciuperca (2009), we get:

$$S(\hat{\tau}_{1,K_0}, \dots, \hat{\tau}_{K_0,K_0}) - S_0 = [\min_{w_1 \in \Gamma^{K_0+1}} (-n^{-1/2} w_1^t \sum_{i=1}^n \dot{f}_{\theta^0}(X_i) \psi(\varepsilon_i) + \frac{\lambda'(0)}{2} w_1 V_0 w_1^t) + \min_{t \in \mathbb{R}^{K_0}} \sum_{i=1}^n \sum_{k=1}^{K_0} z_{n,k}(X_i, \varepsilon_i) \mathbb{1}_{\tau_k^0 < X_i \leq \tau_k^0 + t_k/n}] (1 + o_{\mathbb{P}}(1)), \quad (11)$$

with $V_0 = \mathbb{E}[\dot{f}_{\theta^0}(X) \dot{f}_{\theta^0}^t(X)]$, $t = (t_1, \dots, t_{K_0})$ and $z_{n,k}(X_i, \varepsilon_i) = \rho(\varepsilon_i + d_{(\beta_{k+1}^0, \beta_k^0)}(X_i)) - \rho(\varepsilon_i)$. The quadratic form in w_1 has its minimum: $n^{-1} \left(\sum_{i=1}^n \dot{f}_{\theta^0}(X_i) \dot{f}_{\theta^0}^t(X_i) \psi^2(\varepsilon_i) \right) [\lambda'(0)]^{-1} V_0^{-1}$ which is $O_{\mathbb{P}}(1)$ by (A3), (C4), (B3), (C1). Then

$$\min_{w_1 \in \Gamma^{K_0+1}} \left(-n^{-1/2} w_1^t \sum_{i=1}^n \dot{f}_{\theta^0}(X_i) \psi(\varepsilon_i) + \frac{\lambda'(0)}{2} w_1 V_0 w_1^t \right) = O_{\mathbb{P}}(1). \quad (12)$$

Using, the Cauchy-Schwarz inequality, assumptions (C1) and (B3), we have:

$$\mathbb{E}[z_{n,k}(X, \varepsilon)] \leq \mathbb{E}[|d_{(\beta_{k+1}^0, \beta_k^0)}(X)| \cdot |\sup_y \psi(\varepsilon + y)|]$$

$\leq \mathbb{E}^{1/2}[d_{(\beta_{k+1}^0, \beta_k^0)}^2(X)] \mathbb{E}^{1/2}[\sup_y \psi^2(\varepsilon + y)] \leq C$ and since ε is independent of X :
 $\mathbb{E}[z_{n,k}^2(X, \varepsilon)] \leq \mathbb{E}[d_{(\beta_{k+1}^0, \beta_k^0)}^2(X)] \mathbb{E}[\sup_y \psi^2(\varepsilon + y)] \leq C$. Then by Bienaymé-Tchebychev inequality we obtain that $z_{n,k}(X_i, \varepsilon_i)$ is bounded with probability close to one for $X_i \in [\tau_k^0, \tau_k^0 + t_k/n]$. Since $n^{-1} \sum_{k=1}^n \mathbb{1}_{|X_i - \tau_k^0| \leq B/n} = O_{\mathbb{P}}(n^{-1})$:

$$\min_{t \in \mathbb{R}^{K_0}} \sum_{i=1}^n \sum_{k=1}^{K_0} z_{n,k}(X_i, \varepsilon_i) \mathbb{1}_{\tau_k^0 < X_i \leq \tau_k^0 + t_k/n} = O_{\mathbb{P}}(1). \quad (13)$$

Taking into account (11), (12), (13), we obtain:

$$S(\hat{\tau}_{1, K_0}, \dots, \hat{\tau}_{K_0, K_0}) - S_0 = O_{\mathbb{P}}(1). \quad (14)$$

Moreover, this last relation together the strong law of large number for $n^{-1} \sum_{i=1}^n \rho(\varepsilon_i)$, using assumption (C7) imply:

$$\hat{s}_{K_0} = n^{-1} [S(\hat{\tau}_{1, K_0}, \dots, \hat{\tau}_{K_0, K_0}) - S_0] + n^{-1} \sum_{i=1}^n \rho(\varepsilon_i) \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E}[\rho(\varepsilon)] \neq 0. \quad (15)$$

Thus, when $K < K_0$:

$$\begin{aligned} B(K) - B(K_0) &= n \log \left(1 + \frac{\hat{s}_K - \hat{s}_{K_0}}{\hat{s}_{K_0}} \right) + [G(K, p) - G(K_0, p)] C_n \\ &= \left(n \frac{\hat{s}_K - \hat{s}_{K_0}}{\hat{s}_{K_0}} + [G(K, p) - G(K_0, p)] C_n \right) (1 + o_{\mathbb{P}}(1)). \end{aligned}$$

On the other hand, using (10), (14) and (15), we have with arbitrarily large probability, for all sufficiently large n :

$$\begin{aligned} n(\hat{s}_K - \hat{s}_{K_0}) / \hat{s}_{K_0} &= ([S(\hat{\tau}_{1, K}, \dots, \hat{\tau}_{K, K}) - S_0] - [S(\hat{\tau}_{1, K_0}, \dots, \hat{\tau}_{K_0, K_0}) - S_0]) / \hat{s}_{K_0} \\ &\geq (CA_n - O_{\mathbb{P}}(1)) / \mathbb{E}[\rho(\varepsilon)] \geq CA_n. \end{aligned}$$

Since $\lim_{n \rightarrow \infty} (C_n / A_n) = 0$, we have for all $K < K_0$, $B(K) - B(K_0) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \infty$.

Then $\mathbb{P}[\hat{K}_n < K_0] \rightarrow 0$ for $n \rightarrow \infty$.

Case $K > K_0$. In this case we have, with probability one:

$$\begin{aligned} S_0 &\geq S(\tau_1^0, \dots, \tau_{K_0}^0) \geq S(\hat{\tau}_{1,K_0}, \dots, \hat{\tau}_{1,K_0}) \\ &\geq S(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}) \geq S(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}, \tau_1^0, \dots, \tau_{K_0}^0). \end{aligned} \quad (16)$$

The last sum can be written: $S(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}, \tau_1^0, \dots, \tau_{K_0}^0) = \sum_{r=1}^{K_0+1} T_r$
 $= \min_{\tilde{\theta}_1} \sum_{i=1}^n \sum_{r=1}^{K_0+1} \rho(\varepsilon_i - f_{(\tilde{\theta}_{1,K+K_0}, \tilde{\theta}_{2,K+K_0})}(X_i) + f_{\theta^0}(X_i)) \mathbb{1}_{\tau_{r-1}^0 < X_i \leq \tau_r^0}$, with
 $\tilde{\theta}_{2,K+K_0} = (\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}, \tau_1^0, \dots, \tau_K^0)$, $\tilde{\theta}_{1,K+K_0} \in \Gamma^{K+K_0+1}$ and T_r contains
all observations between τ_{r-1}^0 and τ_r^0 . For $r = 1, \dots, K_0 + 1$ let us consider:
 $k_{1,r} < \dots < k_{J(r),r}$ the elements of the set $\{\hat{\tau}_1, \dots, \hat{\tau}_{K,K}\}$ which are strictly
between $\tau_r^0 + 1$ and τ_{r+1}^0 , the points where function $f_{(\tilde{\theta}_1, \tilde{\theta}_2)}(X_i)$ changes. Ob-
viously $0 \leq J(r) \leq K_0$. If $J(r) = 0$, then, using (2), $E[\psi(\varepsilon)] = 0$, as-
sumption (B3), we have: $\min_{\beta \in \Gamma} \sum_{i=1}^n [\rho(\varepsilon_i + d_{(\beta_r^0, \beta)}(X_i)) - \rho(\varepsilon_i)] \mathbb{1}_{\tau_{r-1}^0 < X_i \leq \tau_r^0} =$
 $(\tilde{\beta}_{r,K+K_0} - \beta_r^0) \sum_{i=1}^n \dot{h}_{\beta_r^0}(X_i) \psi(\varepsilon_i) \mathbb{1}_{\tau_{r-1}^0 < X_i \leq \tau_r^0} (1 + o_{\mathbb{P}}(1)) = O_{\mathbb{P}}(1)$.
If $J(r) > 0$, then $T_r = \min \sum_{j=1}^{J(r)+1} \sum_{i=k_{j-1,r}+1}^{k_{j,r}} \rho(\varepsilon_i + d_{(\beta_r^0, \beta_j)}(X_i))$. Let us con-
sider the interval $[k_{j-1,r}, k_{j,r}]$ for $j = 2, \dots, J(r) - 1$. In this case $k_{j,r} - k_{j-1,r} \geq$
 CA_n and similar that for $J(r) = 0$ we have:

$$\min_{\beta \in \Gamma} \sum_{i=k_{j-1,r}}^{k_{j,r}} [\rho(\varepsilon_i + d_{(\beta_r^0, \beta)}(X_i)) - \rho(\varepsilon_i)] = O_{\mathbb{P}}(1). \quad (17)$$

Let us consider now both intervals of the end $[L_n, k_{1,r}]$ and $[k_{J(r),r}, U_n]$, with
 $L_n = \sum_{i=1}^n \mathbb{1}_{\tau_{r-1}^0 \geq X_i}$, $U_n = \sum_{i=1}^n \mathbb{1}_{\tau_r^0 \geq X_i}$. Let $B_n = \min(k_{1,r} - L_n, U_n - k_{J(r),r})$
be. If $B_n \rightarrow \infty$ in probability, for $n \rightarrow \infty$, by similar arguments we have a
relation as (17). If there exists $c > 0$ such that $B_n < c$ with probability 1, for
all n , then, since $\rho(\varepsilon)$ is bounded with probability close to one, we have:

$$-O_{\mathbb{P}}(1) = - \sum_{i=L_n}^{k_{1,r}} \rho(\varepsilon_i) \leq \min_{\beta \in \Gamma} \sum_{i=L_n}^{k_{1,r}} [\rho(\varepsilon_i + d_{(\beta_r^0, \beta)}(X_i)) - \rho(\varepsilon_i)] \leq 0.$$

Thus, in all cases $T_r = O_{\mathcal{P}}(1)$, for all $r = 1, \dots, K_0 + 1$. Then:

$$0 \geq S(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}, \tau_1^0, \dots, \tau_{K_0}^0) - S_0 = O_{\mathcal{P}}(1). \quad (18)$$

On the other hand, by relation (15) we have that \hat{s}_{K_0} is bounded and strictly positive with arbitrarily large probability. Taking into account inequalities (16) and (18), we obtain: $0 \leq S(\hat{\tau}_{1,K_0}, \dots, \hat{\tau}_{K_0,K_0}) - S(\hat{\tau}_{1,K}, \dots, \hat{\tau}_{K,K}) = O_{\mathcal{P}}(1)$. Then $0 \leq \hat{s}_{K_0} - \hat{s}_K = O_{\mathcal{P}}(n^{-1})$ and together with $0 \neq \hat{s}_{K_0} = O_{\mathcal{P}}(1)$, we get: $n \log \hat{s}_{K_0} - n \log \hat{s}_K = n(\hat{s}_{K_0} - \hat{s}_K)(1 + o_{\mathcal{P}}(1)) = O_{\mathcal{P}}(1)$. Then, since $G(K, p) > G(K_0, p)$ for $K > K_0$, $C_n \rightarrow \infty$, we have $B(K) - B(K_0) = O_{\mathcal{P}}(1) + [G(K, p) - G(K_0, p)]C_n \rightarrow \infty$ for $n \rightarrow \infty$. Hence $B(K) > B(K_0)$ for sufficiently large n . Therefore, $\mathbb{P}[\hat{K}_n > K_0] \rightarrow 0$ when $n \rightarrow \infty$. ■

Acknowledgements The author would like to thank the referee for carefully reading the paper and for his comments which greatly improved the paper.

References

- Bai J.,1999. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, **91**, 299-323.
- Chen, J., Gupta, A.K., Pan, J.,2006. Information criterion and change point problem for regular models. *Sankhya* **68**(2), 252-282.
- Ciuperca G.,2009. The M-estimation in a multi-phase random nonlinear model. *Statistics and Probability Letters*, **75**(5), 573-580.
- Ciuperca, G.,2011a. Estimating nonlinear regression with and without change-points by the LAD-method. *Annals of the Institute of Statistical Mathematics*, DOI: 10.1007/s10463-009-0256-y.

- Ciuperca, G.,2011b. Penalized least absolute deviations estimation for nonlinear model with change-points. *Statistical Papers*,10.1007/s00362-009-0236-6.
- Koul, H.L., Qian, L., Surgailis, D.,2003. Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Processes and their Applications*, **103**, 123-154.
- Nosek, K.,2010. Schwarz information criterion based tests for a change-point in regression models. *Statistics and Probability Letters*, **75**(5), 573-580.
- Osorio, F., Galea, M.,2005. Detection of a change-point in student-t linear regression models. *Statistical Papers*, **45**, 31-48.
- Pan, J., Chen, J.,2006. Application of modified information criterion to multiple change point problems. *Journal of Multivariate Analysis* **97**(10), 2221-2241.
- Seber, G.A.F., Wild, C.J.,2003. *Nonlinear regression*. Wiley Series in Probability and Statistics, New Jersey.
- Yao, Y.C., Au, S.T.,1988. Least-squares estimation of a step function. *Sankhya*, **51**, pp. 370-381.
- Wu, Y.,2008. Simultaneous change point analysis and variable selection in a regression problem. *Journal of Multivariate Analysis* **99**, 2154-2171.