arXiv:0706.0153v3 [math.ST] 22 Sep 2008

The M-estimation in a multi-phase random nonlinear model

Gabriela CIUPERCA*

September 22, 2008

Abstract

This paper considers M-estimation of a nonlinear regression model with multiple change-points occuring at unknown times. The multi-phase random design regression model, discontinuous in each change-point, have an arbitrary error ε. In the case when the number of jumps is known, the M-estimator of locations of breaks and of regression parameters are studied. These estimators are consistent and the distribution of the regression parameter estimators is Gaussian. The estimator of each change-point converges, with the rate n^{-1}, to the smallest minimizer of the independent compound Poisson processes. The results are valid for a large class of error distributions.

Keywords: multiple change-points, M-estimator, random parametric regression, asymptotic properties

1 Introduction

Change-points are intrinsic features of signals that appear in economics, medicine and physical science. The statistics literature contains a vast amount of works on issues related to the estimation of the change-point for a parametric regression, most of it specifically designed for the case of a single break. The more used estimators are the maximum likelihood estimators, the least squares estimators or a wider class, the M-estimators. Statistical inference for a parametrical model is influenced by the continuity or by discontinuity of the regression function at the change-points, but also by the determinist character or not of the explicative variable. We give a non-exhaustive list with the recent papers. The area of research is so active that it is nearly impossible to list all the recent papers written.

For the least squares (LS) estimators we refer to Feder (1975a, 1975b) for continuous two-lines models, Lai et al. (1979), Yao and Au (1989) for a step function, Liu et al. (1997), Bai and Perron (1998) for multiple structural changes in a linear model.

For the maximum likelihood (ML) estimator, when the design is determinist, Bhattacharya

*Université de Lyon, Université Lyon 1, UMR 5208, blvd du 11 novembre 1918, F - 69622 Villeurbanne Cedex, France, E-mail:Gabriela.Ciuperca@univ-lyon1.fr

(1994) discusses his limiting behaviour for a discontinuous linear model. Gill (2004), Gill and Baron (2004) consider a model where the canonical parameter of an exponential family gradually begins to drift from its initial value at an unknown change-point. For a random design we refer to Koul and Qian (2002) for two lines model, Ciuperca (2004) for a single jump in a nonlinear model, Ciuperca and Dapzol (2008) for multiple change-points in linear and nonlinear model. If the model variance depends of the mean, the quasi-likelihood estimator can be considered. Braun et al. (2000) consider that the mean is constant between two change-points. Chiou and Muller (2004) propose a semi-parametric estimator in a generalized linear model with determinist design.

In the general case of M-estimators, Rukhin and Vajda (1997) consider the change-point estimation problem as a nonlinear regression problem, the model being continuous, with a single change-point and fixed design. Koul et al. (2003) study the M-estimators in two-phase linear regression with random design.

The present paper makes several contributions to the existing literature. The considered design is random, the regression function is nonlineary within the framework of a multi-regime and not lastly, a general method of estimation. We study the properties of the M-estimator in a multi-phase discontinuous nonlinear random regression model with a general error distribution. The class of the M-estimators was introduced by Huber (1964) and its principal properties are exposed in Huber (1981). We generalize among others, the results for the two-phase random linear model of Koul et al. (2003) obtained by M-estimation, the results obtained by the ML estimation of Ciuperca and Dapzol (2008) for a multiphase random nonlinear model and of Bai and Perron (1998) obtained by LS estimation in a multiple nonrandom linear regression. An important point of the proofs for the linear case is the relation between the regression function and its derivatives with respect to regression parameters. Thus we have to modify the approach for the non linear regression. Also, in the case of a single change-point, each of two regimes has one fixed boundary. For multiple breaks, each middle regime has boundaries completely unknown.

The paper is organized as follows. We give necessary notations and definitions in Section 2. In Section 3 we establish the estimators consistency and the convergence rate. Weak convergence results are also obtained: the asymptotic distribution of the regression parameters M-estimator is Gaussian. We also prove that $n(\hat{\theta}_{2n} - \theta_2^0)$ converges weakly to the smallest minimizer vector of the independent compound Poisson processes, where $\hat{\theta}_{2n}$ is the change-point estimator. Auxiliary results are given in Appendix.

2 Notations and model

Consider the step-function with K ($K \geq 1$) fixed change-points, for $x \in I\!R$:

$$f_\theta(x) = h_{\alpha_0}(x)1\!\!1_{x \leq \tau_1} + h_{\alpha_1}(x)1\!\!1_{\tau_1 < x \leq \tau_2} + ... + h_{\alpha_K}(x)1\!\!1_{\tau_K < x}$$

where $\theta_1 = (\alpha_0, \alpha_1, ..., \alpha_K)$ are the nonlinear regression parameters and $\theta_2 = (\tau_1, ..., \tau_K)$, $\tau_1 < \tau_2 < ... < \tau_K$ are the change-points. For all $k = 0, 1, ..., K$, we have the parameter α_k belongs to some compact $\Gamma \subseteq I\!R^d$. We consider that the vector $\theta_2 \in I\!R^K$ and we set $\theta = (\theta_1, \theta_2) \in \Omega = \Gamma^{K+1} \times I\!R^K$.
Consider the random design model:

$$Y_i = f_\theta(X_i) + \varepsilon_i, \qquad i = 1, ..., n$$

where (ε_i, X_i) is a sequence of continuous independent random variables with the same joint distribution as (ε, X). The parameter θ_1 and the change-points (or break points) are unknown. The purpose is to estimate $\theta = (\theta_1, \theta_2)$ when n observations of (Y, X) are available.
We denote the true value of a parameter with a 0 superscript. In particular, $\theta_1^0 = (\alpha_0^0, \alpha_1^0, ..., \alpha_K^0)$ and $\theta_2^0 = (\tau_1^0, ..., \tau_K^0)$ are used to denote, respectively, the true values of the regression parameters and the true change-points. Let be also $\theta^0 = (\theta_1^0, \theta_2^0)$. We suppose that θ_1^0 is an inner point of the set Γ^{K+1}.
The random variables X and ε satisfy the following assumptions:
(A1) X has a positive absolutely continuous Lebesgue density φ on $I\!R$. Moreover, $I\!E(X^2) < \infty$;
(A2) ε has a density absolutely continuous and positive everywhere on $I\!R$. Moreover, $I\!E(\varepsilon) = 0$, $I\!E(\varepsilon^2) < \infty$;
(A3) the random variables X_i and ε_i are independent.
In the case of linear model with a single change-point: $h_\alpha(x) = a + bx$, $\alpha = (a, b)$ and $K = 1$, assumptions (A1)-(A3) on X and ε are also considered by Koul et al (2003).

The nonlinear function h_α satisfies the conditions:
(B1) for all $x \in I\!R$, $h_\alpha(x)$ is three times differentiable with respect to α;
(B2) for all $x \in I\!R$, $\|\partial h_{\alpha_0}(x)/\partial\alpha\| \neq 0$;
(B3) the derivatives $\partial^3 h_\alpha(x)/\partial\alpha^3$, exist for $x \in I\!R$ and there exist functions $F_0, F_1, F_2 \in L^2(\varphi)$ such that:

$$\sup_{\alpha\in\Gamma} |h_\alpha(x)| \leq F_0(x), \qquad \sup_{\alpha\in\Gamma} \|\partial^j h_\alpha(x)/\partial\alpha^j\| \leq F_j(x), \quad j = 1, 2 \tag{1}$$

Obviously, in the case $h_\alpha(x) = a + bx$, the assumptions (B1), (B2) are verified and (B3) is transformed in (A1). If $h_\alpha(x)$ is a polynomial with degree p, assumption (B3) can be replaced by $I\!E(X^{p+1}) < \infty$.
Assumption (B2) is necessary for obtaining the convergence rate of regression parameters estimator.

Let us consider the functions: $d_{(\alpha_k,\alpha_j)}(x) := h_{\alpha_k}(x) - h_{\alpha_j}(x)$, $x \in I\!R$, $k, j \in \{1, ...K\}$ and the jump at the true break point: $d_k^0 := d_{(\alpha_k^0,\alpha_{k-1}^0)}(\tau_k^0)$. We make the identifiability assumption that the jump at each τ_k^0 is non-zero:

$$d_{(\alpha_k,\alpha_{k+1})}(\tau_k^0) \neq 0, \qquad \forall \alpha_k, \alpha_{k+1} \in \Gamma, \alpha_k \neq \alpha_{k+1} \tag{2}$$

a condition which implies that the function f_θ is not continuous in the true break points for all parameters in Γ. For $\theta^* = (\theta_1^*, \theta_2^*)$ and $\theta = (\theta_1, \theta_2)$, let us denote by $\delta_{(\theta,\theta^*)}(x) := f_\theta(x) - f_{\theta^*}(x)$ the difference between two models. Note also: $\dot{f}_\theta(x) = \partial f_\theta(x)/\partial\theta_1$.

In the following, we denote by C a generic positive finite constant not depending on n. For a vector, let us denote by $\|.\|$ the Euclidean norm and for a matrix $A = (a_{ij})$, $\|A\| = \sum_{i,j}|a_{ij}|$. For a vector $v = (v_1, ..., v_K)$ we make the convention that $|v| = (|v_1|, ..., |v_K|)$.

The most important method of constructing statistical estimators is to choose the estimator to maximize or minimize a certain criterion function. The such estimators are called the M-estimators. The maximum likelihood (ML), least squares, least absolute deviation estimators are particular cases. For a function $\rho : I\!R \to I\!R_+$, let the M-process be:

$$M_n(\theta) = \sum_{i=1}^{n} \rho\left(Y_i - f_\theta(X_i)\right)$$

The following assumptions are considered for the function ρ:
(C1) ρ is convex on $I\!R$ with right-continuous non-decreasing almost everywhere derivative ψ satisfying $I\!E_\varepsilon[\psi^2(\varepsilon + y)] < \infty$, $\forall y \in I\!R$. The function $\lambda(y) := I\!E_\varepsilon[\psi(\varepsilon + y)]$, $y \in I\!R$, is strictly increasing on $I\!R$ and λ is continuous at 0 with $\lambda(0) = 0$.
(C2) for all $c \in I\!R$, where $\bar{\Omega}$ is the closure of Ω.
(C3) the function $y \to I\!E[|\psi(\varepsilon + c + y) - \psi(\varepsilon)|]$ is continuous at 0, $\forall c \in I\!R$.
(C4) the function λ is differentiable in a neighborhood of 0, with derivative λ' satisfying $\lambda'(0) \neq 0$, and $\lim_{a\to 0} a^{-1} \int_0^a |\lambda'(s) - \lambda'(0)|ds = 0$.
(C5) the random variables $\rho(\varepsilon \pm d_k^0) - \rho(\varepsilon)$, $\forall k = 1, ..., K$, are continuous.

Assumptions (C1), (C2) are necessary for obtaining the consistency of the estimators, while (C1)-(C5) are used for obtaining the rate of convergence and the asymptotic distribution.
Notice that forthe two-phase linear regression function: $f_\theta(x) = (a_0 + b_0 x)1\!\!1_{x\leq\tau} + (a_1 + b_1 x)1\!\!1_{x>\tau}$, Koul et al. (2003) consider the same assumptions (C1)-(C5). Obviously, (C2) becomes: $I\!E_{(\varepsilon,X)}[\psi^2 + c_1 + c_2|X|] < \infty$, $\forall c_1, c_2 \in I\!R$ and (C5) becomes: $\rho(\varepsilon \pm d) - \rho(\varepsilon)$ continuous, with $d = (a_1^0 - a_0^0) + \tau^0(b_1^0 - b_0^0)$, $(a_0^0, b_0^0, a_1^0, b_1^0, \tau^0)$ the true value of $(a_0, b_0, a_1, b_1, \tau)$.

For each $\eta > 0$, denote the η-neighborhood of $\theta \in \Omega$ by:

$$\Omega_\eta(\theta) := \{\theta^* = (\theta_1^*, \theta_2^*) \in \Omega \ / \ \|\theta_1^* - \theta_1\| \leq \eta, \|\theta_2^* - \theta_2\| \leq \eta\}$$

The M-estimator is defined by:

$$\hat{\theta}_n := \left(\hat{\theta}_{1n}, \hat{\theta}_{2n}\right) = \arg\min_{\theta\in\bar{\Omega}} M_n(\theta) \quad \text{a.s.}$$

where $\bar{\Theta}$ is the close of Θ. Let $\bar{I\!R} = I\!R \cup \{-\infty, \infty\}$. The set $\bar{I\!R}$ is compact under the metric $m(x,y) = |arc\tan x - arc\tan y|$, $x, y \in I\!R$. For constructing the M-estimator, first we search the regression parameters estimator and then we localize the change-points. First, for a given $\theta_2 \in I\!R^K$, we set:

$$\tilde{\theta}_{1n}(\theta_2) := \arg \min_{\theta_1 \in \Gamma^{K+1}} M_n(\theta_1, \theta_2)$$

Since the number K of the change-points is fixed, the estimator $\tilde{\theta}_{1n}(\theta_2)$ is constant in θ_2 over any interval of two consecutive ordered X_i's. The M-process $M_n(\tilde{\theta}_{1n}(\theta_2), \theta_2)$ has only a finite number of possible values with change-points located at the ordered X_i's. Second, we find the minimizer $\tilde{\theta}_{2n}$ of $M_n(\tilde{\theta}_{1n}(\theta_2), \theta_2)$ with respect to θ_2 over the sample percentile $\{X_i, i = 1, ..., n\}$. This minimizer may be taken as the left end point of the interval over which it is obtained. Then $\tilde{\theta}_{2n} = \hat{\theta}_{2n}$ and the M-estimator is: $\hat{\theta}_n = (\tilde{\theta}_{1n}(\tilde{\theta}_{2n}), \tilde{\theta}_{2n})$.

Remark. The considered model and the estimator are very general. The class of M-estimators includes the least squares ($\rho(x) = x^2$), maximum likelihood ($\rho(.) = \log \varphi_\varepsilon(.)$, with φ_ε the density of ε) and least absolute deviations estimators ($\rho(x) = |x|$). Examples of distributions satisfying these conditions include Normal for X, double exponential or Normal for the errors ε if $\rho(x) = |x|^a$, $a \in \{1/2, 2\}$.
For the ML estimator in a multi-phase nonlinear random model, the conditions imposed on the random variables are (A1), (A2), (A3) and the density φ_ε of ε satisfies: $u(x) = \varphi'_\varepsilon(x)/\varphi_\varepsilon(x)$ the score function is Holder, is differentiable and u' is Holder also (see Ciuperca and Dapzol(2008)). The function h_α satisfy the condition (B2) and:

$$\sup_{\theta, \theta^* \in \Omega} I\!E_{(\varepsilon, X)}[u^2(\varepsilon + f_\theta(X) - f_{\theta^*}(X))] < \infty$$

3 Asymptotic properties

In this section we focus on study of the asymptotic properties of estimator. First, we study the convergence of the M-estimator and we find the rate of convergence.

3.1 Consistency and rate of convergence

For each change-point τ_k^0, since the density of X is absolutely continuous in $I\!R$, we have:

$$n^{-1} \sum_{i=1}^{n} 1\!\!1_{|X_i - \tau_k^0| \leq B/n} = O_{I\!P}(n^{-1}) \tag{3}$$

It is interesting to mention that in a identifiable regular model for a density with jumps, the ML estimator is of order n^{-1} (see Ibragimov and Has'minskii (1981)). For the multi-phase problem, we obtain that the M-estimator of the change-point has the same order of convergence. Always

in a regular model, van der Vaart and Wellner (1996) obtain the rate of convergence of the M-estimator.

The next theorem establishes the strong consistency of the M-estimator and shows that the rate of the convergence of $\hat{\theta}_{2n}$ to θ_2^0 is n^{-1} and $n^{-1/2}$ of $\hat{\theta}_{1n}$ to θ_1^0. The theorem includes the results derived by Koul et al. (2003) when h_α is linear for the M-estimator and by Ciuperca and Dapzol (2008) when h_α is nonlinear for the ML estimator. Remark that for the ML estimator in a nonlinear random model, the discontinuity of the function $f_\theta(x)$ in the change-points is not necessary to show the consistency of the estimators.
In order to simplify the study of the rate of convergence, three processes defined as the differences between two M-processes are considered. The first one is the difference between a M-process calculated in a some point θ and a M-process at the true point θ^0:

$$D_n(\theta_1, \theta_2) := M_n(\theta_1, \theta_2) - M_n(\theta_1^0, \theta_2^0) \tag{4}$$

For the second one, the regression parameters vary around θ_1^0, for $w_1 \in \Gamma^{K+1}$:

$$D_n^{(1)}(w_1) := M_n\left(\theta_1^0 + n^{-1/2} w_1, \theta_2^0\right) - M_n(\theta_1^0, \theta_2^0)$$

the coefficient of w_1 being the rate of convergence of the estimator $\hat{\theta}_{1n}$ and finally we make vary the change-points:

$$D_n^{(2)}(\theta_1, \theta_2) := M_n(\theta_1, \theta_2) - M_n(\theta_1, \theta_2^0)$$

The relation between these processes is given by the following decomposition:

$$D_n(\theta_1, \theta_2) = D_n^{(1)}\left(n^{1/2}(\theta_1 - \theta_1^0)\right) + D_n^{(2)}(\theta_1, \theta_2) \tag{5}$$

Theorem 3.1 *(i) Under assumptions (2), (A1), (A3), (B1), (B3), (C1) and (C2) we have:* $\hat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta^0$.
(ii) Under the assumptions (2), (A1)-(A3),(B1)-(B3), (C1)-(C5), we have

$$n\|\hat{\theta}_{2n} - \theta_2^0\| = O_{I\!P}(1), \qquad n^{1/2}\|\hat{\theta}_{1n} - \theta_1^0\| = O_{I\!P}(1) \tag{6}$$

Proof of Theorem 3.1 *(i)* To show the strong consistency of the M-estimator, we first prove that the function $e(\theta) := I\!E_{(\varepsilon,X)}\left[|\rho(Y - f_\theta(X)) - \rho(Y - f_{\theta^0}(X))|\right]$ is continuous. By the mean value theorem, we have:

$$|\rho(Y - f_\theta(X)) - \rho(Y - f_{\theta^0}(X))| \leq \int_0^{|\delta_{(\theta,\theta^0)}(X)|} |\psi(\varepsilon - v)| dv \tag{7}$$

Then $e(\theta) < \infty$, for all $\theta \in \Omega$, whence the function $e(\theta)$ is well defined. The result of uniform convergence of $e(\theta)$ given by Lemma 4.2 and $e(\theta^0) = 0$ imply that the function e is continuous on Ω. By (2), we have that $e(\theta) \neq 0$ for all $\theta \neq \theta^0$. Then, we can apply a similar method to that in Huber (1967) and we obtain the strong consistency of the M-estimator.
(ii) Since $\hat{\theta}_n$ is strongly consistent, it suffices to suppose θ in a ϱ-neighborhood of θ^0. For a positive constants b and ϱ, which will be later determined, let be the sets of parameters:

$$\mathcal{V}_{1b\varrho} := \left\{\theta \in \Omega_\varrho(\theta^0);\ \ n^{1/2}\|\theta_1 - \theta_1^0\| > b\right\}, \quad \mathcal{V}_{2b\varrho} := \left\{\theta \in \Omega_\varrho(\theta^0);\ \ n\|\theta_2 - \theta_2^0\| > b\right\}$$

The theorem is proved if we show that: for any $\gamma > 0$, $c \in (0, \infty)$ there exist $b \in (0, \infty)$ and $n_{c\gamma b} \in I\!N$ such that:

$$I\!P[\inf_{\theta \in \bar{\mathcal{V}}_{jb\varrho}} D_n(\theta_1, \theta_2) > c] > 1 - \gamma, \qquad \forall n \geq n_{c\gamma b}, \quad j = 1, 2 \tag{8}$$

where $\bar{\mathcal{V}}_{jb\varrho}$ is the close of $\mathcal{V}_{jb\varrho}$. By relation (5), we have, for $j = 1, 2$:

$$\inf_{\theta \in \bar{\mathcal{V}}_{jb\varrho}} D_n(\theta_1, \theta_2) \geq \inf_{\theta \in \bar{\mathcal{V}}_{jb\varrho}} D_n^{(1)}\left(n^{1/2}(\theta_1 - \theta_1^0)\right) + \inf_{\theta \in \bar{\mathcal{V}}_{jb\varrho}} D_n^{(2)}(\theta_1, \theta_2) \tag{9}$$

• The study of $D_n^{(1)}$ is simpler because it involves only the regression parameters:

$$\inf_{\theta \in \bar{\mathcal{V}}_{2b\varrho}} D_n^{(1)}\left(n^{1/2}(\theta_1 - \theta_1^0)\right) = \min\{\inf_{\|w_1\| \leq b} D_n^{(1)}(w_1), \inf_{\|w_1\| > b} D_n^{(1)}(w_1)\} \tag{10}$$

For $w_1 \in \Gamma^{K+1}$, since ρ is convexe, there are $\tilde{b} > 0$ such that $\inf D_n^{(1)}(w_1)$ is greater than $C \sum_{i=1}^n \rho(\delta_{((\theta_1^0 + n^{-1/2} w_1, \theta_2^0), \theta^0)}(X_i))$, for $\|w_1\| > \tilde{b}$. Assumption (B2) and the convexity of ρ imply that for all $\gamma, \gamma_1 > 0$, there are $b_1 > 0$, $(b_1 \geq \tilde{b})$ and $n_\gamma \in I\!N$ such that:

$$I\!P\left[\inf_{\|w_1\| > b_1} D_n^{(1)}(w_1) > \gamma_1\right] > 1 - \gamma/2, \qquad \forall n > n_\gamma \tag{11}$$

Using relation (11) and the approximation of $D_n^{(1)}$ given in Lemma 4.4 we obtain that the minimum of (10) is $O_{I\!P}(1)$.
On the other hand, for $\theta \in \bar{\mathcal{V}}_{1b\varrho}$ we have with an arbitrarily large probability, for n large: $\inf_{\theta \in \bar{\mathcal{V}}_{1b\varrho}} D_n^{(1)}\left(n^{1/2}(\theta_1 - \theta_1^0)\right) = \inf_{\|w_1\| > b_1} D_n^{(1)}(w_1)$. Relation (11) implies that:
$\inf_{\theta \in \bar{\mathcal{V}}_{1b\varrho}} D_n^{(1)}\left(n^{1/2}(\theta_1 - \theta_1^0)\right)$ is arbitrarily large and positive with a probability close to 1.
• We take charge now the study of $D_n^{(2)}$. For any positive numbers b and ϱ, we prove that $\inf_{\theta \in \bar{\mathcal{V}}_{1b\varrho}} D_n^{(2)}(\theta_1, \theta_2) = O_{I\!P}(1))$, using the decomposition:

$$\inf_{\theta \in \bar{\mathcal{V}}_{1b\varrho}} D_n^{(2)}(\theta_1, \theta_2) = \min\{\inf_{\theta \in \bar{\mathcal{V}}_{1b\varrho} \cap \mathcal{V}_{2b\varrho}} D_n^{(2)}(\theta_1, \theta_2), \inf_{\theta \in \bar{\mathcal{V}}_{1b\varrho} \cap \mathcal{V}_{2b\varrho}^c} D_n^{(2)}(\theta_1, \theta_2)\}$$

with $\mathcal{V}^c_{2b\varrho} = \{\theta \in \Omega_\varrho(\theta^0), n\|\theta_2 - \theta_2^0\| \leq b\}$. Taking into account the convexity of ρ and the approximation on $D_n^{(2)}$ given in Lemma 4.6, we obtain:

$$\inf_{\theta \in \bar{\mathcal{V}}_{1b\varrho} \cap \mathcal{V}^c_{2b\varrho}} D_n^{(2)}(\theta_1, \theta_2) = \inf_{\|t\| \leq b, \|w_1\| = b} D_n^{(2)}\left(\theta_1^0 + n^{-1/2} w_1, \theta_2^0 + n^{-1} t\right) = O_{I\!P}(1)$$

Consider $D_n^{(2)}$ for θ in $\mathcal{V}_{2b_2\varrho}$. By Lemma 4.5, for all positive numbers γ and c, it exist $\gamma_2, b_2 \in (0, \infty)$, $\varrho \in (0,1)$, and $n_2 \in I\!N$ such that: $\gamma_2 b_2 \inf_k \varphi(\tau_k^0) > 2c$ and that relation (22) is true. By (A1), we choose $\varrho \in (0,1)$ sufficiently small such as $\inf_{\tau_k^0 \leq x \leq \tau_k^0 + \varrho} g(x) \geq g(\tau_k^0)/2$ for all $k = 1, ..., K$. Then, for $n > b_2/\varrho$ we have: $\inf_{\tau_k^0 \leq x \leq \tau_k^0 + b_2 n^{-1}} g(x) \geq g(\tau_k^0)/2$, for all $k = 1, ..., K$ and:

$$I\!P\left[\inf_{\theta \in \bar{\mathcal{V}}_{2b_2\varrho}} D_n^{(2)}(\theta_1, \theta_2) > c\right] \geq I\!P\left[\inf_{\theta \in \bar{\mathcal{V}}_{2b_2\varrho}} \frac{D_n^{(2)}(\theta_1, \theta_2)}{nG(|\theta_2 - \theta_2^0|)} > \gamma_0\right] > 1 - \frac{\gamma}{2}, \quad \text{for } n > n_2 \tag{12}$$

Hence the second term of the right-hand side of inequality (9) is arbitrarily large with arbitrarily large probability for sufficiently large n.

• In conclusion, we showed that for every set $\mathcal{V}_{1b_2\varrho}$, $\mathcal{V}_{2b_2\varrho}$, the right-hand side of (9) is the sum of $O_{I\!P}(1)$ and of arbitrarily large random variables. This implies relation (8). $\Diamond$

3.2 Asymptotic distributions

We mean now giving the limiting distribution of the M-estimator and an asymptotic approximation for the M-process.

Let us consider $t \in I\!R^{*K}$ and $w_1 \in \Gamma^{K+1}$. For D_n defined by (4) as a process in the standardized parameters, we have the following decomposition:

$$D_n\left(\theta_1^0 + n^{-1/2} w_1, \theta_2^0 + n^{-1} t\right) = D_n^{(1)}(w_1) + D_n^{(2)}\left(\theta_1^0 + n^{-1/2} w_1, \theta_2^0 + n^{-1} t\right) \tag{13}$$

Let us denote $V_0 := I\!E_X\left[\dot{f}_{\theta^0}(X)\dot{f}_{\theta^0}(X)^t\right]$ the Fisher information matrix corresponding to the random model in X. We suppose that the matrix V_0 is inversible.

The M-process is rescaleted in D_n with regard to the rate of convergence. Let us consider the random vector:

$$\mathcal{Z}_n := n^{-1/2} \sum_{i=1}^n \dot{f}_{\theta^0}(X_i)\psi(\varepsilon_i)$$

Let $D(-\infty, \infty)$ be the set of all cadlag functions on $(-\infty, \infty)$ with the Skorokhod topology.

The next theorem gives the joint asymptotic distributions of the M-estimators. In the asymptotic behaviour of regression parameters estimator, the independence of error ε and of regressor

X intervenes in an essential way in variance formula. Also for $\hat{\theta}_{1n}$, the asymptotic approximation expression is similar to that of the M-estimator in a model without break. On the other hand, the asymptotic distribution of the change-points estimators depends only on the density of X in the true break points and on the difference $\rho(\varepsilon \pm d_k^0) - \rho(\varepsilon)$.

Theorem 3.2 *Under assumptions (2), (A1)-(A3), (B1)-(B3), (C1)-(C5), we have*

$$n^{1/2}(\hat{\theta}_{1n} - \theta_1^0) = \left[\lambda'(0)\right]^{-1} V_0^{-1} \mathcal{Z}_n + o_{I\!P}(1) \tag{14}$$

Moreover, $\left(n^{1/2}(\hat{\theta}_{1n} - \theta_1^0), n(\hat{\theta}_{2n} - \theta_2^0)\right) \xrightarrow[n\to\infty]{\mathcal{L}} (\mathcal{Z}, \Pi_-)$*, with*
$\mathcal{Z} \sim \mathcal{N}_{(K+1)d}\left(0, I\!E_\varepsilon\left[\psi^2(\varepsilon)\right]\lambda'(0)^{-2}V_0^{-1}\right)$ *a Gaussian random vector independent of* $\Pi_- = (\Pi_{1-}, ..., \Pi_{K-})$*,* $\Pi_{k-} = \arg\min_{t_k \in I\!R} \mathcal{P}_k(t_k)$*, where:*

$$\mathcal{P}_k(t_k) = \mathcal{P}_{k1}(t_k)1\!\!1_{t_k \geq 0} + \mathcal{P}_{k2}(-t_k)1\!\!1_{t_k \leq 0} \tag{15}$$

$\mathcal{P}_{k1}$ *and* $\mathcal{P}_{k2}$ *are two independent compound Poisson processes on* $[0, \infty)$ *with rate* $\varphi(\tau_k^0)$ *and* $\mathcal{P}_{k1}(0) = \mathcal{P}_{k2}(0) = 0$*. The distribution of jumps is given by:* $\rho(\varepsilon + d_k^0) - \rho(\varepsilon)$*, respectively* $\rho(\varepsilon - d_k^0) - \rho(\varepsilon)$*.*

Proof of Theorem 3.2 Using the approximation results obtained in Lemmas 4.4 and 4.6 and also the decomposition: $M_n\left(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t\right) = M_n(\theta_1^0, \theta_2^0) + D_n\left(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t\right)$, we obtain an asymptotic approximation for the standardized M-process as the sum of two processes. The first is quadratic form $Q_n(w_1)$ in the standardized regression parameters, the second is a empirical process in the standardized change-point parameters:

$$M_n\left(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t\right) = Q_n(w_1) + D_n^{(2)}(\theta_1^0, \theta_2^0 + n^{-1}t) + o_{I\!P}(1) \tag{16}$$

where:

$$Q_n(w_1) = M_n(\theta_1^0, \theta_2^0) - n^{-1/2}w_1^t \sum_{i=1}^n \dot{f}_{\theta^0}(X_i)\psi(\varepsilon_i) + \frac{\lambda'(0)}{2} w_1^t V_0 w_1 \tag{17}$$

Let us remind that ψ is the derivative of the function ρ.
For $t = (t_1, ..., t_K) \in I\!R^{*K}$, $w_1 \in \Gamma^{K+1}$, by relation (16) we have that the minimum of $M_n\left(\theta_1^0 + n^{-1/2}w_1, r + n^{-1}t\right)$ with respect to (w_1, t) is equivalent with the minimum with respect to w_1 of $Q_n(w_1)$ and with respect to t of $D_n^{(2)}\left(\theta_1^0, \theta_2^0 + n^{-1}t\right)$. Then relation (14) results from (17). Relation (14) implies that the study of the distribution limit of $n^{1/2}(\hat{\theta}_{1n} - \theta_1^0)$ amount to study the law limit of $\mathcal{Z}_n$. But taking into account (C1), by a Central Limit Theorem, $\mathcal{Z}_n$ converges in distribution to the gaussian distribution: $\mathcal{N}_{(K+1)d}\left(0, V_0 I\!E_\varepsilon[\psi^2(\varepsilon)]\right)$.
In view of Theorem 3.1 (*ii*) for the change-point estimator, we have:

$$n(\hat{\theta}_{2n} - \theta_2^0) = \arg\min_{t \in I\!R^K} D_n^{(2)}\left(\theta_1^0, \theta_2^0 + n^{-1}t\right) + o_{I\!P}(1)$$

For study jointly the distribution of $\mathcal{Z}_n$ and of $D_n^{(2)}$ we apply Theorem 4.2 of Koul et al. (2003) for $f_n(X,\varepsilon) := \dot{f}_{\theta_1^0}(X)\psi(\varepsilon)$ and $h_n(X,\varepsilon) := \rho(\varepsilon + d_{(\alpha_k^0,\alpha_{k-1}^0)}(X)) - \rho(\varepsilon)$. Note that: $D_n^{(2)}\left(\theta_1^0, \theta_2^0 + n^{-1}t\right) = \sum_{i=1}^n \sum_{k=1}^K h_n(X_i,\varepsilon_i)\mathbb{1}_{\min\{\tau_k^0,\tau_k^0+t_k/n\}<X_i\leq\max\{\tau_k^0,\tau_k^0+t_k/n\}}$. On the other hand, for $\xi_n(x,z) := I\!E_\varepsilon\left[\exp\left(in^{-1/2}z^t f_n(X,\varepsilon)\right)|X=x\right]$ we have:

$$|n\left(1-\xi_n(x,z)\right)| \leq \frac{I\!E_\varepsilon\left[\psi^2(\varepsilon)\right]}{2}\left[z^t\dot{f}_{\theta_1^0}(x)\right]^2 \leq CI\!E_\varepsilon\left[\psi^2(\varepsilon)\right]\|z\|^2\sup_\alpha\left\|\frac{\partial h_\alpha(x)}{\partial\alpha}\right\|^2$$

By assumptions (1) and (A1), we obtain that $n\left(1-\xi_n(x,z)\right)$ is uniformly integrable with respect to $dH(x)$, where H is the distribution function of X. Thus $n\left(1-\xi_n(x,z)\right) \to I\!E_\varepsilon\left[\psi^2(\varepsilon)\right]z^t\Lambda(x)z$, with: $\Lambda(x) := \dot{f}_{\theta_1^0}(x)\dot{f}_{\theta_1^0}^t(x)$ and $\Lambda := V_0 = I\!E_X[\Lambda(X)]$. Whence:

$$\left(\mathcal{Z}_n, D_n^{(2)}(\theta_1^0,\theta_2^0+n^{-1}t)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \left(\mathcal{N}_{(K+1)d}\left(0, V_0 I\!E_\varepsilon[\psi^2(\varepsilon)]\right), \mathcal{P}(t)\right)$$

in $I\!R^{(K+1)d}\times D(-\infty,\infty)^K$ with $\mathcal{P}(t) := \sum_{k=1}^K \mathcal{P}_k(t_k)$. The random vector $\mathcal{N}_{(K+1)d}\left(0, V_0 I\!E_\varepsilon[\psi^2(\varepsilon)]\right)$ is independent of $\mathcal{P}_k$, $k=1,...,K$.

We prove now that $n(\hat{\theta}_{2n}-\theta_2^0)$ converges weakly to the smallest minimizer Π_- of the process $\mathcal{P}$ and show then that the components of this vector coincide with the minimizer of $\mathcal{P}_k(t_k)$, with the probability 1. Seen the Skorokhod space definition, $D(-\infty,\infty)$, we consider that change-points vary in a compact of $I\!R^K$.

We consider the M-estimator of the change-points: $\hat{\theta}_{2n}^b := \arg\min_{t\in[-b,b]^K} M_n\left(\hat{\theta}_{1n}(t),t\right)$ and the minimizer of $\mathcal{P}(t)$: $\Pi_-^b := \arg\min_{t\in[-b,b]^K}\mathcal{P}(t)$, for a fixed $b>0$. By Theorem 3.1, there is a real number $b<\infty$ such that $\hat{\theta}_{2n}-\hat{\theta}_{2n}^b \to 0$ a.s. for $n\to\infty$. More, it also exists a real $b<\infty$ such that $\Pi_- = \Pi_-^b$ with a probability arbitrarily large.

Then, we shall first prove that for all $b>0$:

$$n(\hat{\theta}_{2n}^b-\theta_2^0) \xrightarrow[n\to\infty]{\mathcal{L}} \Pi_-^b \tag{18}$$

For $t\in[-b,b]^K$, $\tilde{b}=(b,...,b)$ a K-vector, we consider the random process $\mathcal{P}^b(t) := \mathcal{P}(t)\mathbb{1}_{|t|<\tilde{b}}$ and: $\hat{M}_n^b(t) := \left[M_n\left(\hat{\theta}_{1n}(\theta_2^0+n^{-1}t),\theta_2^0+n^{-1}t\right) - M_n\left(\hat{\theta}_{1n}(\theta_2^0),\theta_2^0\right)\right]\mathbb{1}_{|t|<\tilde{b}}$. Let also, for $v\in I\!R$, the random process:

$$H_n^k(v) = \sum_{i=1}^n\left[\rho(\varepsilon_i + sign(v)d_k^0) - \rho(\varepsilon_i)\right]\mathbb{1}_{\min\left(\tau_k^0,\tau_k^0+n^{-1}v\right)<X_i\leq\max\left(\tau_k^0,\tau_k^0+n^{-1}v\right)}$$

and theirs sum: $H_n(t) = \sum_{k=1}^K H_n^k(t_k)$. So by (C3), $I\!E_{(\varepsilon,X)}\left[\sup_{\|t\|\leq b}\left|D_n^{(2)}(\theta_2^0+n^{-1}t) - H_n(t)\right|\right]$ is bounded to upper by

$$n\sum_{k=1}^K\int_{|x-\tau_k^0|\leq n^{-1}b}\varphi(x)I\!E_\varepsilon\left[\left|\rho\left(\varepsilon + d_{(\alpha_k^0,\alpha_{k-1}^0)}(x)\right) - \rho\left(\varepsilon + sign(t_k)d_k^0\right)\right|\right]dx$$

$$= n\sum_{k=1}^{K}\int_{|x-\tau_k^0|\le n^{-1}b}\varphi(x)\mathbb{E}_\varepsilon\left[|d_{(\alpha_k^0,\alpha_{k-1}^0)}(x)-d_k^0|\,|\psi(\varepsilon+y_x)|\right]dx, \quad \text{with } y_x\to 0, \text{ for } x\to 0$$

$$\le Cn\sum_{k=1}^{K}\int_{|x-\tau_k^0|\le n^{-1}b}|x-\tau_k^0|\left(\sup_\alpha\left\|\frac{\partial h_\alpha(x)}{\partial\alpha}\right\|\right)\varphi(x)\mathbb{E}_\varepsilon\left[\psi(\varepsilon+y_x)\right]dx$$

But $\varphi(x)<C$ and $\mathbb{E}_\varepsilon\left[\psi(\varepsilon+y_x)\right]<C$ as a continuously function on a compact. Then, by the Cauchy-Schwarz inequality:

$$\mathbb{E}_{(\varepsilon,X)}\left[\sup_{\|t\|\le b}\left|D_n^{(2)}(\theta_2^0+n^{-1}t)-H_n(t)\right|\right]\le Cn\sum_{k=1}^{K}\left[\int_{|x-\tau_k^0|\le n^{-1}b}(x-\tau_k^0)^2dx\right]^{1/2}=o(1)$$

Hence: $\sup_{\|t\|\le b}\left|\hat{M}_n^b(t)-H_n(t)\right|=o_{\mathbb{P}}(1)$. Let us consider: $\Pi_n^b=\arg\min_{t\in[-b,b]^K}H_n(t)$. By Lemmas 4.3 and 4.4 of Koul et al. (2003) we obtain:
$n(\hat{\theta}_{2n}^b-\theta_2^0)-\Pi_n^b\xrightarrow[n\to\infty]{\mathcal{L}}0$, $\Pi_n^b\xrightarrow[n\to\infty]{\mathcal{L}}\Pi_-^b$. Then relation (18) follows. Because for two different change-points we have to make of two independent sets of random variables we have that: $\arg\min_{t\in[-b,b]^K}H_n(t))=\sum_{k=1}^K\arg\min_{t_k\in[-b,b]}H_n^k(t_k))$. The last relation, with (18) and $\Pi_n^b\xrightarrow[n\to\infty]{\mathcal{L}}\Pi_-^b$, imply that the asymptotic distribution of $n(\hat{\theta}_{2n}-\theta_2^0)$ is Π_-. $\Diamond$

Remarks. 1. For $K=1$, we find the results that the empirical processes $D_n^{(2)}\left(\theta_1^0,\theta_2^0+n^{-1}t\right)$ converges to a compound Poisson process. We also find all the asymptotic distributions for particular estimators ML, LS, LAD. Particularly, the asymptotic variance of the ML estimator of the regression parameters is: $\mathbb{E}_\varepsilon[(\psi_\varepsilon'(\varepsilon)/\psi_\varepsilon(\varepsilon))^2]V_0$, with ψ_ε the density of ε.
2. Consequence of Theorem 3.2, we can find the confidence interval or make hypothesis test for the parameter θ.
3. The discontinuity in the change-points of the regression functions influences the rate of convergence of the change-point estimator. The proved results are differently from those in the continuous or discontinuous in the change-points for non-random design cases. For example, Van der Geer (1988) prove that in the uniform non-random design two-phase, discontinuous, the limiting distribution of the change-point estimator is determined by a Brownian motion with a linear drift. Rukhin and Vajda (1997) for a continuous model prove that the M-estimator of the change-point is asymptotically normal.

4 Appendix: Lemmas

To begin, we state a elementary lemma.

Lemma 4.1 *For any k random variables $Z_1, ..., Z_k$ the following inequalities are valid:*

$$\sum_{i=1}^{k} I\!P\left[Z_i < 0\right] - (k-1) \leq I\!P\left[\left(\sum_{i=1}^{k} Z_i\right) < 0\right] \leq \sum_{i=1}^{k} I\!P\left[Z_i < 0\right]$$

The following lemma of uniform convergence will be useful in the proofs of the main theorems.

Lemma 4.2 *Under assumptions (A1), (B1), (B3) and (C2), we have*

$$\lim_{\eta \searrow 0} I\!E_{(\varepsilon,X)}\left[\sup_{\theta^* \in \Omega_\eta(\theta)} \left|\rho\left(Y - f_\theta(X)\right) - \rho\left(Y - f_{\theta^*}(X)\right)\right|\right] = 0$$

Proof of Lemma 4.2 We apply a version of the mean value theorem:

$$\rho\left(Y - f_\theta(X)\right) - \rho\left(Y - f_{\theta^*}(X)\right) = \delta_{(\theta,\theta^*)}(X) \int_0^1 \psi\left(Y - f_\theta(X) + v\delta_{(\theta,\theta^*)}(X)\right) dv \tag{19}$$

We begin by showing that:

$$I\!E_X\left[\sup_{(\theta,\theta^*) \in \Omega_\eta(\theta)} \delta^2_{(\theta,\theta^*)}(X)\right] \underset{\eta \to 0}{\longrightarrow} 0 \tag{20}$$

Regarding the change-points, there are two possible cases.
Case 1. $\tau_k \in I\!R$, $\forall k = 1, ..., K$. We have:

$$\sup_{\theta^* \in \Omega_\eta(\theta)} \left|\delta_{(\theta,\theta^*)}(X)\right| \leq C\left[\eta \left\|\sup_{\alpha \in \Gamma} \frac{\partial h_\alpha(X)}{\partial \alpha}\right\| + 2\sup_{\alpha \in \Gamma} |h_\alpha(X)| \sum_{k=1}^{K} 1\!\!1_{|X - \tau_k| \leq \eta}\right]$$

Furthermore $I\!P\left[|X - \tau_k| \leq \eta\right] \to 0$ for $\eta \to 0$. Then, with condition (2) we obtain (20).
Case 2. $\tau_1 = -\infty$ or $\tau_K = \infty$. Without loss of generality, we consider $\tau_1 = -\infty$. Obviously $\tau_1^* \geq \tau_1$. We have $|\tau_1^* - \tau_1| \leq \eta$. Then:

$$\sup_{\theta^* \in \Omega_\eta(\theta)} |\delta_{(\theta,\theta^*)}(X)| \leq C\left[\eta \left\|\sup_{\alpha \in \Gamma} \frac{\partial h_\alpha(X)}{\partial \alpha}\right\| + 2\sup_{\alpha \in \Gamma} |h_\alpha(X)| \left(1\!\!1_{X < \tau_1^*} + \sum_{k=2}^{K} 1\!\!1_{|X - \tau_k| \leq \eta}\right)\right]$$

But $I\!P\left[X < \tau_1^*\right] \to 0$ for $\eta \to 0$. Using assumption (B3) and the Cauchy-Schwarz inequality, we obtain relation (20).
On the other hand, using the inequality: $\forall x \in I\!R$, $|\psi(x + \varepsilon)| \leq |\psi(\varepsilon + |x|)| + |\psi(\varepsilon - |x|)|$, we have, by the Cauchy-Schwarz inequality:

$$I\!E_{(\varepsilon,X)}[\sup_{\theta^* \in \bar{\Omega}_\eta(\theta)} \left|\delta_{(\theta,\theta^*)}(X)\right| \int_0^1 |\psi(Y - f_\theta(X) + v\delta_{(\theta,\theta^*)}(X))| dv]$$

$$\leq C\mathbb{E}^{1/2}_{(\varepsilon,X)}[\sup_{\theta^*\in\bar{\Omega}_\eta(\theta)}\delta^2_{(\theta,\theta^*)}(X)]\mathbb{E}^{1/2}_{(\varepsilon,X)}[\sup_{\theta^*\in\bar{\Omega}_\eta(\theta)}\psi^2(\varepsilon+1/2|\delta_{(\theta,\theta^*)}(X)|+|\delta_{(\theta,\theta^0)}(X)|)]$$

$$+C\mathbb{E}^{1/2}_{(\varepsilon,X)}[\sup_{\theta^*\in\bar{\Omega}_\eta(\theta)}\delta^2_{(\theta,\theta^*)}(X)]\mathbb{E}^{1/2}_{(\varepsilon,X)}[\sup_{\theta^*\in\bar{\Omega}_\eta(\theta)}\psi^2(\varepsilon-1/2|\delta_{(\theta,\theta^*)}(X)|-|\delta_{(\theta,\theta^0)}(X)|)]$$

The conclusion results from relations (19), (20) and from assumption (C2). $\Diamond$

For $x, z \in \mathbb{R}$, $\tau \in \mathbb{R}^K$, for each $k = 1, ..., K$, let be function: $\nu_k(x, z) := \rho(z + sgn(\tau_k - \tau_k^0)d_{(\alpha_k^0,\alpha_{k-1}^0)}(x)) - \rho(z)$. Let be the function: $p_k(x) := \mathbb{E}_\varepsilon[\nu_k(x, \varepsilon)]$. For each break point τ_k^0 we count the number of X_i which fall into the interval $(\tau_k^0, \tau_k^0 + |u_k|)$, with $u = (u_1, ..., u_K) \in \mathbb{R}^K$. Let us consider the functions $G_k, G_{k,n} : \mathbb{R}^* \to (0, 1]$, where $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$:

$$G_{k,n}(u_k) := n^{-1}\sum_{i=1}^{n}\mathbb{1}_{\min(\tau_k^0,\tau_k^0+u_k)<X_i\leq\max(\tau_k^0,\tau_k^0+u_k)}$$

and its expectation: $G_k(u_k) := \mathbb{E}_X[\mathbb{1}_{\min(\tau_k^0,\tau_k^0+u_k)<X_i\leq\max(\tau_k^0,\tau_k^0+u_k)}]$. For all K change-points we define the functions $G, G_n : \mathbb{R}^{*K} \to \mathbb{R}_+$,

$$G(u) := \sum_{k=1}^{K} G_k(u_k), \qquad G_n(u) := \sum_{k=1}^{K} G_{k,n}(u_k)$$

We present a lemma that states an important property for this functions.

Lemma 4.3 *Under (A1), for each $\gamma > 0$, $\eta > 0$, there exists a constant $0 < B < \infty$, such that for all $b \in (0, 1)$, and $n \geq [B/b] + 1$,*

$$\mathbb{P}[\sup_{B/n<\|u\|\leq b}|\frac{G_n(u)}{G(u)} - 1| < \eta] > 1 - \gamma$$

$$\mathbb{P}[\sup_{B/n<\|u\|\leq b}|\frac{\sum_{k=1}^{K} Z_{k,n}^{(1)}(u_k)}{G(u)} - 1| < \eta] > 1 - \gamma$$

where $Z_{k,n}^{(1)} : \mathbb{R}^* \to \mathbb{R}$, $k = 1, ..., K$ is defined by:
$Z_{k,n}^{(1)}(u_k) := n^{-1}\sum_{i=1}^{n}[p_k(X_i) - \nu_k(X_i, \varepsilon_i)]\,\mathbb{1}_{\min(\tau_k^0,\tau_k^0+u_k)<X_i\leq\max(\tau_k^0,\tau_k^0+u_k)}$

Proof of Lemma 4.3 The proof is similar to that of Lemma 5.1 of Ciuperca and Dapzol (2008) using the results for a single change-point (see Lemma 3.2 of Koul and Qian (2002)). $\Diamond$

Let us now given an approximation of the M-process in θ^0 in the direction of the parameters of regression.

Lemma 4.4 *Under the assumptions (A1)-(A3), (B1), (B3), (C1), (C4) and (C5), for each $b \in (0, \infty)$, we have:*

$$\sup_{\|w_1\|<b} |D_n^{(1)}(w_1) + n^{-1/2} w_1^t \sum_{i=1}^n [\dot{f}_{\theta^0}(X_i)\psi(\varepsilon_i)] - \frac{\lambda'(0)}{2} w_1^t V_0 w_1| = o_{I\!P}(1) \tag{21}$$

Proof of Lemma 4.4 Using (C1), (C4) and (C5), we have that: $D_n^{(1)}(w_1)$ is equal to:

$$\sum_{i=1}^n \left[f_{(\theta_1^0, \theta_2^0)}(X_i) - f_{(\theta_1^0 + n^{-1/2} w_1, \theta_2^0)}(X_i) \right] \psi(\varepsilon_i)$$

$$-\frac{1}{2} \sum_{i=1}^n \left[f_{(\theta_1^0, \theta_2^0)}(X_i) - f_{(\theta_1^0 + n^{-1/2} w_1, \theta_2^0)}(X_i) \right]^2 \psi'(\varepsilon_i)(1 + o_{I\!P}(1))$$

Since $\lambda(0) = 0$ and by the assumption (1) we obtain:

$$D_n^{(1)}(w_1) = \left\{ -n^{-1/2} w_1^t \sum_{i=1}^n \dot{f}_{\theta^0}(X_i)\psi(\varepsilon_i) - \frac{1}{2} n^{-1} w_1^t \sum_{i=1}^n \dot{f}_{\theta^0}(X_i) \dot{f}_{\theta^0}^t(X_i) w_1 \psi'(\varepsilon_i) \right\} (1 + o_{I\!P}(1))$$

with $o_{I\!P}(1))$ uniformly in w_1 and n. Thus, using (A1)-(A3) and (C4), by the strong law of large numbers for $\sum_{i=1}^n \dot{f}_{\theta^0}(X_i) \dot{f}_{\theta^0}^t(X_i)\psi'(\varepsilon_i)$ and by the assumption (1) for h_α, we get:
$D_n^{(1)}(w_1) = \{-n^{-1/2} w_1^t \dot{f}_{\theta^0}(X_i)\psi(\varepsilon_i) - \frac{\lambda'(0)}{2} w_1 V_0 w_1^t\}(1 + o_{I\!P}(1))$. Thus the proof is complete. $\Diamond$

In the following lemma, the set $\mathcal{V}_{2b_2\varrho}$ is defined in the proof of Theorem 3.1. The proof of the similar result for two-phase linear model of Koul et al. (2003) is facilitated by the existence of a single change-point and especially by the linearity in x of $h_\alpha(x)$.

Lemma 4.5 *Under the assumptions (2), (A1)-(A3), (B1), (B3), (C1)-(C5), for all positive numbers γ and c, it exist $\gamma_2, b_2 \in (0, \infty)$, $\varrho \in (0, 1)$, and $n_2 \in I\!N$ such that: $\gamma_2 b_2 \inf_k \varphi(\tau_k^0) > 2c$ and that:*

$$I\!P \left[\inf_{\theta \in \bar{\mathcal{V}}_{2b_2\varrho}} \frac{D_n^{(2)}(\theta_1, \theta_2)}{nG(|\theta_2 - \theta_2^0|)} > \gamma_2 \right] > 1 - \gamma/2, \qquad \forall n > n_2 \tag{22}$$

Proof of Lemma 4.5 Let us introduce some notations for ease of exposition. For each change-point τ_k^0, consider the processes:
$S_{k,n}^{(1)}(\theta_1, u_k) := n^{-1} \sum_{i=1}^n \left[\rho(\varepsilon_i + d_{(\alpha_k^0, \alpha_{k-1})}(X_i)) - \rho(\varepsilon_i + d_{(\alpha_k^0, \alpha_{k-1}^0)}(X_i)) \right] 1\!\!1_{\min(\tau_k^0, \tau_k^0 + u_k) < X_i \le \max(\tau_k^0, \tau_k^0 + u_k)}$
$S_{k,n}^{(2)}(\theta_1, u_k) := n^{-1} \sum_{i=1}^n [\rho(\varepsilon_i) - \rho(\varepsilon_i + d_{(\alpha_k^0, \alpha_k)}(X_i))] 1\!\!1_{\min(\tau_k^0, \tau_k^0 + u_k) < X_i \le \max(\tau_k^0, \tau_k^0 + u_k)}$
and the functions: $Z_{k,n}^{(2)} : I\!R^* \to I\!R$, $k = 1, ..., K$:

$Z^{(2)}_{k,n}(u_k) := n^{-1}\sum_{i=1}^n \left[p_k(X_i) - \tau^0_k(r_k)\right] \mathbb{1}_{\min(\tau^0_k,\tau^0_k+u_k)<X_i\leq\max(\tau^0_k,\tau^0_k+u_k)}$
Let us consider $\theta_2 = \theta^0_2+u$ with $u = (u_1, ..., u_K)$. Given these notations, we see that $n^{-1}D^{(2)}_n(\theta_1, \theta_2)$ can be written as:

$$\begin{aligned} n^{-1}D^{(2)}_n(\theta_1,\theta_2) &= \textstyle\sum_{k=1}^K p_k(\tau^0_k)G_k(u_k) + \sum_{k=1}^K p_k(\tau^0_k)\left[G_{k,n}(u_k) - G_k(u_k)\right] \\ &+ \textstyle\sum_{k=1}^K \left[Z^{(1)}_{k,n}(u_k) + Z^{(2)}_{k,n}(u_k) + S^{(1)}_{k,n}(\theta_1,u_k) + S^{(2)}_{k,n}(\theta_1,u_k)\right] \end{aligned} \tag{23}$$

We shall prove that the supremum on the set $\bar{\mathcal{V}}_{2b_2\varrho}$ of all terms on the right-hand side of (23), except the first, divided by $G(|\theta_2-\theta^0_2|)$ is $o_{I\!P}(1)$. On the other hand, we prove that the first term is strictly positive with the probability 1. Remember that $\lambda(y) = I\!E_\varepsilon[\rho'(\varepsilon + y)]$. By Fubini's lemma, by (A2), (C1) and (2), we obtain that:

$$p_k(\tau^0_k) = I\!E_\varepsilon\left[\rho(\varepsilon + d^0_k) - \rho(\varepsilon)\right] = \int_{\min(0,d^0_k)}^{\max(0,d^0_k)} \lambda(z)dz$$

Since the function λ is strictly increasing and $\lambda(0) = 0$, we obtain: $p_k(\tau^0_k) > 0$, for each $k = 1, ..., K$. For all $u_k \leq \varrho$, $k = 1, ..., K$ we have: $|Z^{(2)}_{k,n}(u_k)| \leq \sup_{0\leq\upsilon\leq\varrho}\left|p_k(\tau^0_k + \upsilon) - p_k(\tau^0_k)\right| G_{k,n}(u_k)$. By assumption (C1), for $n \to \infty$ and $\varrho \to 0$, we have: $\sup_{B/n<\upsilon\leq\varrho}\left|p_k(\tau^0_k + \upsilon) - p_k(\tau^0_k)\right| = o_{I\!P}(1)$ and with Lemma 4.3 for all $\varrho > 0$, there is a $B_1 > 0$ such that for $u_k \in (B/n, \varrho)$, $k = 1, ..., K$, and for $n > B_1/\varrho$, we have: $0 \leq G_{k,n}(u_k)/G(u) \leq G_n(u)/G(u) = 1 + o_{I\!P}(1)$. Hence:

$$\sup_{B_1/n<\|u\|\leq\varrho} \frac{\sum_{k=1}^K \left|Z^{(2)}_{k,n}(u_k)\right|}{G(u)} = o_{I\!P}(1) \qquad \text{for } n \to \infty, \quad \varrho \searrow 0 \tag{24}$$

We have a similar relation for $Z^{(1)}_{k,n}$, for a $B_2 > 0$ and $n > B_2/\varrho$.
For $S^{(1)}_{k,n}$, by Theorem 3.1(i), for all $x \in I\!R$ and for all $\theta \in \bar{\mathcal{V}}_{2b\varrho}$ we have: $|d_{(\alpha_{k-1},\alpha^0_{k-1})}(x)| < C\varrho$, whence:

$$|S^{(1)}_{k,n}(\theta_1, u_k)| \leq n^{-1}\sum_{i=1}^n \int_{-C\varrho}^{C\varrho} |\psi(\varepsilon_i + d_{(\alpha^0_k,\alpha^0_{k-1})}(X_i) + v)|dv \mathbb{1}_{\min(\tau^0_k,\tau^0_k+u_k)<X_i\leq\max(\tau^0_k,\tau^0_k+u_k)}$$

Applying Lemma 3.2 of Koul et al. (2003) for: $J(x, z) = \int_{-C\varrho}^{C\varrho} |\psi\left(z + d_{(\alpha^0_k,\alpha^0_{k-1})}(x) + v\right)|dv$ we obtain that there exists a $B_3 \in (0, \infty)$ such that for $n \to \infty$, $\varrho \searrow 0$:

$$\sup_{\theta\in\bar{\mathcal{V}}_{2B_3\varrho}} \frac{|S^{(1)}_{k,n}(\theta_1, u_k)|}{G_k(u_k)} = o_{I\!P}(1)$$

Then, for all $\gamma > 0$, $\eta > 0$:

$$I\!P[\sup_{\theta\in\bar{\mathcal{V}}_{2B_3\varrho}} \frac{\sum_{k=1}^{K}|S_{k,n}^{(1)}(\theta_1,u_k)|}{G(u)} < \eta] \geq I\!P[\sup_{\theta\in\bar{\mathcal{V}}_{2B_3\varrho}} \sum_{k=1}^{K}\frac{|S_{k,n}^{(1)}(\theta_1,u_k)|}{G_k(u_k)} < \eta] \geq 1-\gamma$$

A similar relation holds for $\sum_{k=1}^{K} S_{k,n}^{(2)}(\theta_1,u_k)$ for a $B_4 > 0$. Then, for $l = 1, 2$:

$$\sup_{\theta\in\bar{\mathcal{V}}_{2B_{l+2}\varrho}} \frac{\sum_{k=1}^{K} S_{k,n}^{(l)}(\theta_1,u_k)}{G(u)} = o_{I\!P}(1), \quad \text{for } \gamma > 0, \varrho \searrow 0 \tag{25}$$

On the other hand, for each $\eta > 0$:

$$I\!P[\frac{|\sum_{k=1}^{K} p_k(\tau_k^0)[G_{k,n}(u_k) - G_k(u_k)]|}{G(u)} < \eta] \geq I\!P[\sum_{k=1}^{K}\frac{|G_{k,n}(u_k) - G_k(u_k)|}{G_k(u_k)} < \frac{\eta}{\max_k p_k(\tau_k^0)}]$$

$$\geq \sum_{k=1}^{K} I\!P[\frac{|G_{k,n}(u_k) - G_k(u_k)|}{G_k(u_k)} < \frac{\eta}{\max_k p_k(\tau_k^0)}] - (K-1)$$

the last inequality is obtained by Lemma 4.1. By Lemma 4.3, for each $\eta > 0$ and $\tilde{\gamma} > 0$ there exists a $B_5 > 0$ such that the probability which intervenes in the last inequality is bigger than $1 - \tilde{\gamma}$. Choose $\tilde{\gamma} = \gamma/(4K)$ and $\eta < \max_k p_k(\tau_k^0)[4 + 1/(K \max_k p_k(\tau_k^0))]^{-1}$. We obtain inequality (22) for $\gamma_2 = [\max_k p_k(\tau_k^0) - \eta(4 + 1/(K \max_k p_k(\tau_k^0)))]/2$, $b_2 = \max\{B_1, ..., B_5\}$ and $n_2 = b_2/\varrho + 1$. $\Diamond$

Following result gives the behaviour of $D_n^{(2)}$ in a $n^{-1/2}$-neighborhood of θ_1^0.

Lemma 4.6 *Under assumptions (2), (A1)-(A3), (B1), (B3), (C2), if we define* $A_n(w_1,t) := D_n^{(2)}\left(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t\right) - D_n^{(2)}\left(\theta_1^0, \theta_2^0 + n^{-1}t\right)$*, we have, for every* $b \in (0,\infty)$*:*

$$\sup_{\|(w_1,t)\|\leq b} |A_n(w_1,t)| = o_{I\!P}(1) \tag{26}$$

Proof of Lemma 4.6 Without loss of generality, we consider the vector $t = (t_1, ..., t_K) \in I\!R_+^{*K}$. The general case is obtained by very similar arguments. Let us note $w_1 = (w_{1,0}, w_{1,1}, ..., w_{1,K}) \in \Gamma^{K+1}$. First, observe that we can write

$$A_n(w_1,t) = \sum_{i=1}^{n}\sum_{k=1}^{K+1}\int_0^{d_{(\alpha_{k-1}^0,\alpha_{k-1}^0+n^{-1/2}w_{1,k-1})}(X_i)} \psi(\varepsilon_i + h_{\alpha_k^0}(X_i) + v)dv 1\!\!1_{\tau_k^0 < X_i \leq \tau_k^0 + n^{-1}t_k}$$

$$-\sum_{i=1}^{n}\sum_{k=1}^{K+1}\int_0^{d_{(\alpha^0_{k-1},\alpha^0_k+n^{-1/2}w_{1,k})}(X_i)}\psi(\varepsilon_i+v)dv\mathbb{1}_{\tau^0_k<X_i\leq\tau^0_k+n^{-1}t_k}$$

Let us denote:

$$B(w_1,t):=A_n(w_1,t)+B_1(w_1,t)-B_2(w_1,t) \tag{27}$$

with

$$B_1(w_1,t):=\sum_{i=1}^{n}\sum_{k=1}^{K}d_{(\alpha^0_{k-1}+n^{-1/2}w_{1,k-1},\alpha^0_{k-1})}(X_i)\psi(\varepsilon_i)\mathbb{1}_{\tau^0_k<X_i\leq\tau^0_k+n^{-1}t_k}$$

and

$$B_2(w_1,t):=\sum_{i=1}^{n}\sum_{k=1}^{K+1}d_{(\alpha^0_k+n^{-1/2}w_{1,k},\alpha^0_k)}(X_i)\psi(\varepsilon_i)\mathbb{1}_{\tau^0_k<X_i\leq\tau^0_k+n^{-1}t_k}$$

By assumption (1) for $\partial h_\alpha/\partial\alpha$ we have that: $|d_{(\alpha^0_{k-1}+n^{-1/2}w_{1,k-1},\alpha^0_{k-1})}(X_i)|\leq n^{-1/2}U_i$ with U_i a random variable such as $\mathbb{E}_X(U_i)<\infty$ and there exists a real $c_{1,b}>0$ such that $\mathbb{P}\left[U_i<c_{1,b}\right]>1-\gamma$, for all $\gamma>0$.
Since $|h_{\alpha^0_{k-1}}(X_i)|\mathbb{1}_{\tau^0_k<X_i\leq\tau^0_k+n^{-1}t_k}<C$ with the probability 1, we have that $\mathbb{E}_{(\varepsilon,X)}[|B_1(w_1,t)|]$ is bounded to upper by :

$$n\int_0^{n^{-1/2}c_{1,b}}\left\{\mathbb{E}_\varepsilon[|\psi(\varepsilon+C+v)-\psi(\varepsilon)|]+\mathbb{E}_\varepsilon[|\psi(\varepsilon-C+v)-\psi(\varepsilon)|]\right\}dv\sum_{k=1}^{K}\mathbb{E}_X\left[\mathbb{1}_{\tau^0_k<X_i\leq\tau^0_k+n^{-1}t_k}\right]$$

Since ε and X are independently and by the relation (3):
$\mathbb{E}_{(\varepsilon,X)}[|B_1(w_1,t)|]\leq o(1)n\sum_{k=1}^{K}G_k(n^{-1}t_k)=o(1)$. Alike:

$$\mathbb{E}_{(\varepsilon,X)}[|B_2(w_1,t)|]\leq n\int_0^{n^{-1/2}c_{1,b}}\mathbb{E}_\varepsilon\left[|\psi(\varepsilon+v)-\psi(\varepsilon)|\right]dv\sum_{k=1}^{K}G_k(n^{-1}t_k)=o(1)$$

Therefore, we have

$$\mathbb{E}_{(\varepsilon,X)}[|B_1(w_1,t)|]=o(1),\qquad \mathbb{E}_{(\varepsilon,X)}[B_2(w_1,t)]=o(1) \tag{28}$$

Let us note the random process: $D_1(w_1,t):=B(w_1,t)-A_n(w_1,t)=B_1(w_1,t)-B_2(w_1,t)$. Then, we can write:

$$D_1(w_1,t)=\sum_{i=1}^{n}\sum_{k=1}^{K+1}\{d_{(\alpha^0_{k-1}+n^{-1/2}w_{1,k-1},\alpha^0_{k-1})}(X_i)-d_{(\alpha^0_k+n^{-1/2}w_{1,k},\alpha^0_k)}(X_i)\}\psi(\varepsilon_i)\mathbb{1}_{\tau^0_k<X_i\leq\tau^0_k+n^{-1}t_k}$$

Using assumption (1) for $\partial^2h_\alpha/\partial\alpha^2$ we obtain:

$$D_1(w_1,t)=n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K+1}[w_{1,k-1}\frac{\partial h_{\alpha^0_{k-1}}(X_i)}{\partial\alpha}-w_{1,k}\frac{\partial h_{\alpha^0_k}(X_i)}{\partial\alpha}]\psi(\varepsilon_i)\mathbb{1}_{\tau^0_k<X_i\leq\tau^0_k+n^{-1}t_k}+o_{\mathbb{P}}(1) \tag{29}$$

Finally, by assumption (C2) for $c = 0$ and by (3), we obtain that $\forall b \in (0, \infty), \forall w_1 \in I\!R^{K+1}$:

$$\sup_{0 \le \|t\| \le b} \| \sum_{i=1}^{n} \sum_{k=1}^{K+1} [w_{1,k-1} \frac{\partial h_{\alpha_{k-1}^0}(X_i)}{\partial \alpha} - w_{1,k} \frac{\partial h_{\alpha_k^0}(X_i)}{\partial \alpha}] \psi(\varepsilon_i) 1\!\!1_{\tau_k^0 < X_i \le \tau_k^0 + n^{-1} t_k} \| = O_{I\!P}(1) \quad (30)$$

The conclusion follows from the relations (27), (28), (29) and (30) . $\Diamond$

References

[1] Bai, J., Perron P. (1998), Estimating and testing linear models with multiple structural changes, *Econometrica* **66**(1), 47-78.

[2] Bhattacharya, P.K.(1994), *Some aspects of change-point analysis.* IMS Lecture Notes-Monograph Series, Vol. 23, Hayward, CA, 28-56.

[3] Braun, J.V., Braun, R.K., Muller, H.G.(2000), Multiple changepoint fitting via likelihood, with application to DNA sequence segmentation. *Biometrika* **87**(2), 301-314.

[4] Chiou, J.M., Muller, H.G.(2004), Quasi-likelihood regression with multiple indices and smooth link and variance functions. *Scandinavian Journal of Statistics* **31**, 367-386.

[5] Ciuperca, G.(2004), Maximum likelihood estimator in a two-phase nonlinear regression model. *Statitic and Decision* **22**(4), 335-349.

[6] Ciuperca, G., Dapzol, N., (2008) Maximum likelihood estimator in a multi-phase random regression model. *Statistics* **42**(4), 363-381.

[7] Feder, P.I.(1975a), On asymptotic distribution theory in segmented regression problems-identified case. *Ann. Statist.* **3**, 49-83.

[8] Feder, P.I.(1975b), The log likelihood ratio in segmented regression. *Ann. Statist.* **3**, 84-97.

[9] Gill, R.(2004), Maximum likelihood estimation in generalized broken-line regression. *Canadian Journal of Statistics* **3**, 227-238.

[10] Gill, R., Baron, M.(2004), Consistent estimation in generalized broken-line regression. *J. Journal of Statitical Planning and Inference* **126**, 441-460.

[11] Huber, P.J.(1964), Robust estimation of a location parameter. *Ann. Math. Statist.* **36**, 73-101.

[12] Huber, P. J.(1967), The behaviour of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematics Statistic and Probability, Vol 1, University California Press, Berkeley, 221-234.

[13] Huber P.J.(1981), *Robust Statistics.* Wiley, New York.

[14] Ibragimov I.A., Has'minskii R.Z.(1981), *Statistical Estimation: Asymptotic Theory.* Springer, New York.

[15] Koul, H.L., Qian, L.(2002), Asymptotics of maximum likelihood estimator in a two-phase linear regression model. *Journal of Statitical Planning and Inference* **108**, 99-119.

[16] 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.

[17] Lai, T.L., Robbins, H., Wei, C. Z.(1979) , Strong consistency of least squares estimates in multiple regression. *J. Multivariate Anal.* **9**, 343-361.

[18] Liu, J., Wu, S., Zidek, J.V.(1997), On segmented multivariate regressions. *Statistica Sinica* **7**, 497-525.

[19] Rukhin, A.L., Vajda, I.(1997), Change-point estimation as a nonlinear regression problem. *Statistics* **30**(3), 181-200.

[20] Van der Geer, S.(1988), *Regression analysis and empirical processes.* CWI Tract., Vol. 45, Amsterdam, NL.

[21] Van der Vaart, A.W., Wellner J.A.(1996), *Weak convergence and empirical processes.* Springer, New York.

[22] Yao, Y.C., Au, S.T.(1989), Least-squares estimation of a step function. *Sankhya (Series A)* **51**(3), 370-381.