



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

Gabriela Ciuperca

► To cite this version:

Gabriela Ciuperca. The M-estimation in a multi-phase random nonlinear model. Statistics and Probability Letters, 2009, 79 (5), pp.573. <10.1016/j.spl.2008.10.003>. <hal-00508914>

HAL Id: hal-00508914

<https://hal.science/hal-00508914v1>

Submitted on 7 Aug 2010

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

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

Gabriela Ciuperca

PII: S0167-7152(08)00468-9
DOI: [10.1016/j.spl.2008.10.003](https://doi.org/10.1016/j.spl.2008.10.003)
Reference: STAPRO 5234

To appear in: *Statistics and Probability Letters*

Received date: 4 July 2008
Revised date: 1 October 2008
Accepted date: 1 October 2008

Please cite this article as: Ciuperca, G., The M-estimation in a multi-phase random nonlinear model. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.10.003

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.



* Manuscript

[Click here to view linked References](#)

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

Gabriela CIUPERCA¹

Université Lyon 1, France

Abstract

This paper extends the results of M-estimation in Koul et al. (2003) to a general nonlinear random design regression model with multiple change-points at unknown times. The M-estimator of locations of breaks and of regression parameters are consistent. Convergence rate and asymptotic distribution are obtained.

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

1 Introduction

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 (ML) estimators, the least squares (LS) estimators or a wider class, the M-estimators. For the LS estimators we refer to Feder (1975a, 1975b) for continuous two-lines models, Lai et al. (1979),

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

Yao and Au (1989) for a step function, Bai and Perron (1998) for multiple structural changes in a linear model. For the ML estimator, when the design is random 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. 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 lines model with random design.

The present paper makes several contributions to the existing literature. The considered design is random, the regression function is nonlinearity 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. 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 ML estimation of Ciuperca and Dapzol (2008) for a multiphase random nonlinear model and the results obtained by LS estimation in a multiple nonrandom linear regression of Bai and Perron (1998). 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, due to the fact that these derivatives do not depend on parameters.

2 Notations and model

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

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

where $\theta_1 = (\alpha_0, \alpha_1, \dots, \alpha_K)$ are the nonlinear regression parameters and $\theta_2 = (\tau_1, \dots, \tau_K)$, $\tau_1 < \tau_2 < \dots < \tau_K$ are the change-points. We suppose that, for all $k = 0, 1, \dots, K$, $\alpha_k \in \Gamma \subseteq \mathbb{R}^d$, with Γ compact. We consider $\theta_2 \in \mathbb{R}^K$ and we set $\theta = (\theta_1, \theta_2) \in \Omega = \Gamma^{K+1} \times \mathbb{R}^K$.

Consider the random design model: $Y_i = f_\theta(X_i) + \varepsilon_i$, for $i = 1, \dots, n$, 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 θ when n observations of (Y, X) are available. We denote by $\theta_1^0 = (\alpha_0^0, \alpha_1^0, \dots, \alpha_K^0)$ and $\theta_2^0 = (\tau_1^0, \dots, \tau_K^0)$, 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 assume 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 \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.

Remark 1 In the case $h_\alpha(x) = a + bx$, $\alpha = (a, b)$ and $K = 1$, assumptions (A1)-(A3) on X and ε are also considered by Koul et al (2003).

In the following, we denote by C a generic positive finite constant not depend-

ing on n . For a vector $v = (v_1, \dots, v_K)$ we denote by $\|v\|$ its Euclidean norm and we make the convention that $|v| = (|v_1|, \dots, |v_K|)$. For a three dimension matrix $R = (r_{ijk})$ we can consider the norm $\|R\|_1 = \sum_{i,j,k} |r_{ijk}|$.

The nonlinear function h_α satisfies the conditions:

- (B1) for all $x \in \mathbb{R}$, $h_\alpha(x)$ is three times differentiable with respect to α ;
- (B2) for all $x \in \mathbb{R}$, $k \in \{0, 1, \dots, K\}$, $\|\partial h_{\alpha_k^0}(x)/\partial \alpha\| \neq 0$;
- (B3) the derivatives $\partial^3 h_\alpha(x)/\partial \alpha^3$, exist for $x \in \mathbb{R}$ and there exists the functions $F_0, F_1, F_2 \in L^2(\varphi)$ such that:

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

The derivative $\partial^3 h_\alpha(x)/\partial \alpha^3$ is a three dimension matrix .

Obviously, in the case $h_\alpha(x) = a + bx$, 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 $\mathbb{E}(X^{p+1}) < \infty$. The assumption (B2) is necessary for obtaining the convergence rate of regression parameters estimator, by considering α in a neighborhood of α_k^0 .

Let us consider the functions: $d_{(\alpha_k, \alpha_j)}(x) := h_{\alpha_k}(x) - h_{\alpha_j}(x)$, $x \in \mathbb{R}$, $k, j \in \{0, 1, \dots, K\}$ 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:

$$d_{(\alpha_k, \alpha_{k-1})}(\tau_k^0) \neq 0, \quad \forall \alpha_k, \alpha_{k-1} \in \Gamma, \alpha_k \neq \alpha_{k-1}, k \in \{1, \dots, K\} \quad (2)$$

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

For a function $\rho : \mathbb{R} \rightarrow \mathbb{R}_+$, let the M-process be:

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

The following assumptions are considered for the 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 all $c \in \mathbb{R}$, $\mathbb{E}_{(\varepsilon, X)}[\psi^2(\varepsilon + c \sup_{\theta, \theta^* \in \bar{\Omega}} \delta_{(\theta, \theta^*)}(X))] < \infty$, 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 $\rho(\varepsilon \pm d_k^0) - \rho(\varepsilon), \forall k = 1, \dots, K$, are continuous.

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

Assumptions (C1), (C2) are necessary for obtaining the consistency of the estimators, while (C1)-(C6) are used for obtaining the rate of convergence and the asymptotic distribution.

Notice that for the two-phase linear regression function: $f_\theta(x) = (a_0 + b_0 x) \mathbb{1}_{x \leq \tau} + (a_1 + b_1 x) \mathbb{1}_{x > \tau}$, Koul et al. (2003) consider assumptions (C1)-(C5). Obviously, (C2) becomes: $\mathbb{E}_{(\varepsilon, X)}[\psi^2 + c_1 + c_2 |X|] < \infty$.

The M-estimator is defined by:

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

For constructing the M-estimator, first we search the regression parameters estimator and then we localize the change-points. For a given $\theta_2 \in \mathbb{R}^K$, 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 $M_n(\tilde{\theta}_{1n}(\theta_2), \theta_2)$. 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: $\hat{\theta}_n = (\tilde{\theta}_{1n}(\tilde{\theta}_{2n}), \tilde{\theta}_{2n})$. Obviously, $\hat{\theta}_{1n} = \tilde{\theta}_{1n}(\tilde{\theta}_{2n})$.

Remark 2 *The considered model and the estimator are very general. The class of M-estimators includes the LS ($\rho(x) = x^2$), ML ($\rho(\cdot) = \log \varphi_\varepsilon(\cdot)$, with φ_ε the density of ε) and least absolute deviations (LAD) estimators ($\rho(x) = |x|$). Examples of distributions satisfying these conditions include Normal for X , double exponential or Normal for the errors ε if $\rho(x) = |x|^a$, $a \in \{1/2, 2\}$.*

3 Asymptotic properties

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 :

$$D_n(\theta_1, \theta_2) := M_n(\theta_1, \theta_2) - M_n(\theta_1^0, \theta_2^0) \quad (3)$$

For the second one, the regression parameters vary around θ_1^0 , for $w_1 \in \Gamma^{K+1}$: $D_n^{(1)}(w_1) := M_n(\theta_1^0 + n^{-1/2}w_1, \theta_2^0) - M_n(\theta_1^0, \theta_2^0)$, the coefficient of w_1 being the rate of convergence of the estimator $\hat{\theta}_{1n}$, and finally the change-points vary:

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

For certain results of this section the proofs are omitted or we give only their outline. The full versions can be found in the preprint version of this paper (Ciuperca (2008)).

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

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

The following lemma of uniform convergence will be useful in the consistency proof.

Lemma 3.1 *Under assumptions (A1), (B1), (B3) and (C2), we have*

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

Proof of Lemma 3.1 We apply a version of the mean value theorem:

$$\rho(Y - f_\theta(X)) - \rho(Y - f_{\theta^*}(X)) = \delta_{(\theta, \theta^*)}(X) \int_0^1 \psi(Y - f_\theta(X) + v\delta_{(\theta, \theta^*)}(X)) dv \quad (4)$$

We begin by showing that:

$$\mathbb{E}_X \left[\sup_{(\theta, \theta^*) \in \Omega_\eta(\theta)} \delta_{(\theta, \theta^*)}^2(X) \right] \xrightarrow{\eta \rightarrow 0} 0 \quad (5)$$

Case 1. $\tau_k \in \mathbb{R}, \forall k = 1, \dots, K$. We have:

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

Furthermore $\mathbb{P}[|X - \tau_k| \leq \eta] \rightarrow 0$ for $\eta \rightarrow 0$. Then, using (B3) we obtain (5).

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:

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

But $\mathbb{P}[X < \tau_1^*] \rightarrow 0$ for $\eta \rightarrow 0$. Using assumption (B3) and the Cauchy-Schwarz inequality, we obtain relation (5).

On the other hand, using the inequality: $\forall x \in \mathbb{R}, |\psi(x + \varepsilon)| \leq |\psi(\varepsilon + |x|)| + |\psi(\varepsilon - |x|)|$, by the Cauchy-Schwarz inequality, (C2) and relations (4), (5) we obtain the conclusion. ■

The next theorem establishes the strong consistency of the M-estimator and shows that the convergence rate of $\hat{\theta}_{2n}$ to θ_2^0 is n^{-1} and $n^{-1/2}$ of $\hat{\theta}_{1n}$ to θ_1^0 .

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

(ii) Under assumptions (2), (A1)-(A3), (B1)-(B3), (C1)-(C6), we have

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

Proof of Theorem 3.1 Lemma 3.1 is used to show (i). For (ii), Lemmas 3.3, 3.5, 3.6 are needed. The full version of the proof is given in the preprint version of this paper (Ciuperca (2008)). ■

Remark 3 Assumptions (2) is necessary to prove that: $e(\theta) \neq 0$ for all $\theta \neq \theta^0$ and then we can apply the Huber (1967) method to prove the convergence.

For $x, z \in \mathbb{R}$, $\tau \in \mathbb{R}^K$, for each $k = 1, \dots, K$, let be functions: $\nu_k(x, z) := \rho(z + \text{sgn}(\tau_k - \tau_k^0)d_{(\alpha_k^0, \alpha_{k-1}^0)}(x)) - \rho(z)$ and $p_k(x) := \mathbb{E}_\varepsilon[\nu_k(x, \varepsilon)]$. Let us consider the functions $G_k, G_{k,n} : \mathbb{R}^* \rightarrow (0, 1]$, 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 $u = (u_1, \dots, u_K)$, we define also the functions $G, G_n : \mathbb{R}^{*K} \rightarrow \mathbb{R}_+$,

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

Lemma 3.2 *Under (A1), for each $\gamma > 0$, $\eta > 0$, there exists a constant $0 < B < \infty$, such that for all $b \in (0, 1)$, and $n \geq [B/b] + 1$,*

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

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

where $Z_{k,n}^{(1)} : \mathbb{R}^* \rightarrow \mathbb{R}$, $k = 1, \dots, K$ 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 3.2 See proof of Lemma 4.3 in Ciuperca(2008). ■

Let us now given an approximation of the M-process in θ^0 in the direction of the parameters of regression. Let us denote $V_0 := \mathbb{E}_X [\dot{f}_{\theta^0}(X) \dot{f}_{\theta^0}(X)^t]$ the Fisher information matrix corresponding to the random model in X . We suppose that the matrix V_0 is invertible.

Lemma 3.3 *Under assumptions (A1)-(A3), (B1), (B3), (C1), (C4)-(C6), for each $b \in (0, \infty)$:*

$$\sup_{\|w_1\| < b} |D_n^{(1)}(w_1) + n^{-1/2} w_1^t \sum_{i=1}^n [\dot{f}_{\theta^0}(X_i) \psi(\varepsilon_i)] - \frac{\lambda'(0)}{2} w_1^t V_0 w_1| = o_{\mathbb{P}}(1) \quad (7)$$

Proof of Lemma 3.3 Using (C1), (C4) and (C5), we have that:

$$\begin{aligned} D_n^{(1)}(w_1) &= \sum_{i=1}^n [f_{(\theta_1^0, \theta_2^0)}(X_i) - f_{(\theta_1^0 + n^{-1/2} w_1, \theta_2^0)}(X_i)] \psi(\varepsilon_i) \\ &\quad - \frac{1}{2} \sum_{i=1}^n [f_{(\theta_1^0, \theta_2^0)}(X_i) - f_{(\theta_1^0 + n^{-1/2} w_1, \theta_2^0)}(X_i)]^2 \psi'(\varepsilon_i) (1 + o_{\mathbb{P}}(1)) \end{aligned}$$

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

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

with $o_P(1)$ uniformly in w_1 and n . In final, taking into account (A1)-(A3) and (C4), the strong law of large numbers is applied for $\sum_{i=1}^n \dot{f}_{\theta^0}(X_i) \dot{f}_{\theta^0}^t(X_i) \psi'(\varepsilon_i)$. \blacksquare

Lemma 3.4 *For any k random variables Z_1, \dots, Z_k we have:*

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

Lemma 3.5 *Under assumptions (2), (A1)-(A3), (B1), (B3), (C1)-(C5), for all positive numbers γ and c , there exist $\gamma_2, b_2 \in (0, \infty)$, $\varrho \in (0, 1)$, and $n_2 \in \mathbb{N}$ such that: $\gamma_2 b_2 \inf_k \varphi(\tau_k^0) > 2c$ and that:*

$$\mathbb{P}\left[\inf_{\theta \in \mathcal{V}_{2b_2\varrho}} \frac{D_n^{(2)}(\theta_1, \theta_2)}{nG(|\theta_2 - \theta_2^0|)} > \gamma_2\right] > 1 - \gamma/2, \quad \forall n > n_2 \quad (8)$$

where $\mathcal{V}_{2b_2\varrho} = \{\theta \in \Omega_\varrho(\theta^0); n\|\theta_2 - \theta_2^0\| > b_2\}$ for some $b_2 > 0$.

Proof of Lemma 3.5 For each change-point τ_k^0 , consider the processes:

$$S_{k,n}^{(1)}(\theta_1, u_k) := n^{-1} \sum_{i=1}^n \left[\rho(\varepsilon_i + d_{(\alpha_k^0, \alpha_{k-1}^0)}(X_i)) - \rho(\varepsilon_i + d_{(\alpha_k^0, \alpha_{k-1}^0)}(X_i)) \right] \mathbb{1}_{\min(\tau_k^0, \tau_k^0 + u_k) < X_i \leq \max(\tau_k^0, \tau_k^0 + u_k)}$$

$$S_{k,n}^{(2)}(\theta_1, u_k) := n^{-1} \sum_{i=1}^n [\rho(\varepsilon_i) - \rho(\varepsilon_i + d_{(\alpha_k^0, \alpha_k)}(X_i))] \mathbb{1}_{\min(\tau_k^0, \tau_k^0 + u_k) < X_i \leq \max(\tau_k^0, \tau_k^0 + u_k)}$$

and the functions: $Z_{k,n}^{(2)}: \mathbb{R}^* \rightarrow \mathbb{R}$, $k = 1, \dots, K$:

$$Z_{k,n}^{(2)}(u_k) := n^{-1} \sum_{i=1}^n [p_k(X_i) - p_k(\tau_k^0)] \mathbb{1}_{\min(\tau_k^0, \tau_k^0 + u_k) < X_i \leq \max(\tau_k^0, \tau_k^0 + u_k)}$$

Let us consider $\theta_2 = \theta_2^0 + u$ with $u = (u_1, \dots, u_K)$. Then:

$$\begin{aligned} n^{-1} D_n^{(2)}(\theta_1, \theta_2) &= \sum_{k=1}^K p_k(\tau_k^0) G_k(u_k) + \sum_{k=1}^K p_k(\tau_k^0) [G_{k,n}(u_k) - G_k(u_k)] \\ &+ \sum_{k=1}^K [Z_{k,n}^{(1)}(u_k) + Z_{k,n}^{(2)}(u_k) + S_{k,n}^{(1)}(\theta_1, u_k) + S_{k,n}^{(2)}(\theta_1, u_k)] \end{aligned}$$

The lemma will result by showing that the supremum on the set $\bar{\mathcal{V}}_{2b_2\varrho}$ of all terms on the right-hand side of the last relation, except the first, divided by

$G(|\theta_2 - \theta_2^0|)$ is $o_{\mathbb{P}}(1)$ that the first term is strictly positive with the probability 1. By Fubini's lemma, (2), (A2), (C1) and since the function λ is strictly increasing and $\lambda(0) = 0$, we have: $p_k(\tau_k^0) > 0$, for each $k = 1, \dots, K$.

For all $u_k \leq \varrho$: $|Z_{k,n}^{(2)}(u_k)| \leq \sup_{0 \leq v \leq \varrho} |p_k(\tau_k^0 + v) - p_k(\tau_k^0)| G_{k,n}(u_k)$. By (C1), Lemma 3.2, $\forall \varrho > 0$, $\exists B_1 > 0$ such that:

$$\sup_{B_1/n < \|u\| \leq \varrho} \frac{\sum_{k=1}^K |Z_{k,n}^{(2)}(u_k)|}{G(u)} = o_{\mathbb{P}}(1) \quad \text{for } n \rightarrow \infty, \varrho \searrow 0$$

We have a similar relation for $Z_{k,n}^{(1)}$, for a $B_2 > 0$ and $n > B_2/\varrho$.

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

$$\begin{aligned} \mathbb{P} \left[\frac{\left| \sum_{k=1}^K p_k(\tau_k^0) [G_{k,n}(u_k) - G_k(u_k)] \right|}{G(u)} < \eta \right] &\geq \mathbb{P} \left[\sum_{k=1}^K \frac{|G_{k,n}(u_k) - G_k(u_k)|}{G_k(u_k)} < \frac{\eta}{\max_k p_k(\tau_k^0)} \right] \\ &\geq \sum_{k=1}^K \mathbb{P} \left[\frac{|G_{k,n}(u_k) - G_k(u_k)|}{G_k(u_k)} < \frac{\eta}{\max_k p_k(\tau_k^0)} \right] - (K-1) \end{aligned}$$

the last inequality is obtained by Lemma 3.4. By Lemma 3.2, $\forall \eta, \tilde{\gamma} > 0$, $\exists B_5 > 0$ such that the probability which intervenes in the last inequality is bigger than $1 - \tilde{\gamma}$. The nonlinearity not occuring in an essential way in the rest of proof, we prefer omit it (see proof of Lemma 4.5 in Ciuperca(2008)). ■

Lemma 3.6 *Under assumptions (2), (A1)-(A3), (B1), (B3), (C2), if we define $A_n(w_1, t) := D_n^{(2)}(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t) - D_n^{(2)}(\theta_1^0, \theta_2^0 + n^{-1}t)$, we have, for every $b \in (0, \infty)$: $\sup_{\|(w_1, t)\| \leq b} |A_n(w_1, t)| = o_{\mathbb{P}}(1)$*

Proof of Lemma 3.6 See proof of Lemma 4.6 of Ciuperca(2008). ■

Let us consider $t \in \mathbb{R}^{*K}$ and $w_1 \in \Gamma^{K+1}$. For D_n defined by (3) as a process in the standardized parameters, we have the following decomposition:

$$D_n(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t) = D_n^{(1)}(w_1) + D_n^{(2)}(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t) \quad (9)$$

Let us consider the random vector:

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

The next theorem gives the joint asymptotic distributions of the M-estimators. For $\hat{\theta}_{1n}$, the asymptotic approximation expression is similar to that of the M-estimator in a model without break. 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)-(C6), we have $(n^{1/2}(\hat{\theta}_{1n} - \theta_1^0), n(\hat{\theta}_{2n} - \theta_2^0)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} (\mathcal{Z}, \Pi_-)$, with $\mathcal{Z} \sim \mathcal{N}_{(K+1)d}(0, \mathbb{E}_\varepsilon[\psi^2(\varepsilon)] \lambda'(0)^{-2} V_0^{-1})$ independent of $\Pi_- = (\Pi_{1-}, \dots, \Pi_{K-})$, $\Pi_{k-} = \arg \min_{t_k \in \mathbb{R}} \mathcal{P}_k(t_k)$, where, for $k = 1, \dots, K$:*

$$\mathcal{P}_k(t_k) = \mathcal{P}_{k1}(t_k) \mathbb{1}_{t_k \geq 0} + \mathcal{P}_{k2}(-t_k) \mathbb{1}_{t_k \leq 0} \quad (10)$$

\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 We give the outline of the proof.

Using the approximation results obtained in Lemmas 3.3, 3.6 and also the decomposition: $M_n(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t) = M_n(\theta_1^0, \theta_2^0) + D_n(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t)$ we have an asymptotic approximation for the standardized M-process as the sum of two processes. The first is the quadratic form $Q_n(w_1)$ in the standardized regression parameters, the second is a empirical process in the standardized change-point parameters:

$$M_n(\theta_1^0 + n^{-1/2}w_1, \theta_2^0 + n^{-1}t) = Q_n(w_1) + D_n^{(2)}(\theta_1^0, \theta_2^0 + n^{-1}t) + o_P(1) \quad (11)$$

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

Then $n^{1/2}(\hat{\theta}_{1n} - \theta_1^0) = [\lambda'(0)]^{-1} V_0^{-1} \mathcal{Z}_n + o_P(1)$. Taking into account (C1),

$$\mathcal{Z}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_{(K+1)d}(0, V_0 \mathbb{E}_\varepsilon[\psi^2(\varepsilon)]).$$

In view of Theorem 3.1 (ii) for the change-point estimator, we have:

$$n(\hat{\theta}_{2n} - \theta_2^0) = \arg \min_{t \in \mathbb{R}^K} D_n^{(2)}(\theta_1^0, \theta_2^0 + n^{-1}t) + o_P(1)$$

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 $f_n(X, \varepsilon) := \dot{f}_{\theta^0}(X) \psi(\varepsilon)$ and $h_n(X, \varepsilon) := \rho(\varepsilon + d_{(\alpha_k^0, \alpha_{k-1}^0)}(X)) - \rho(\varepsilon)$. Note that:

$$D_n^{(2)}(\theta_1^0, \theta_2^0 + n^{-1}t) = \sum_{i=1}^n \sum_{k=1}^K h_n(X_i, \varepsilon_i) \mathbb{I}_{\min\{\tau_k^0, \tau_k^0 + t_k/n\} < X_i \leq \max\{\tau_k^0, \tau_k^0 + t_k/n\}}.$$

On the other hand, for $\xi_n(x, z) := \mathbb{E}_\varepsilon \left[\exp \left(i n^{-1/2} z^t f_n(X, \varepsilon) \right) | X = x \right]$, we have:

$$|n(1 - \xi_n(x, z))| \leq \frac{\mathbb{E}_\varepsilon[\psi^2(\varepsilon)]}{2} \left[z^t \dot{f}_{\theta^0}(x) \right]^2 \leq C \mathbb{E}_\varepsilon[\psi^2(\varepsilon)] \|z\|^2 \sup_\alpha \left\| \frac{\partial h_\alpha(x)}{\partial \alpha} \right\|^2$$

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

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

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

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 \mathbb{R}^K .

We consider the M-estimator of the change-points: $\hat{\theta}_{2n}^b := \arg \min_{t \in [-b, b]^K} M_n(\hat{\theta}_{1n}(t), t)$ 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 $\hat{\theta}_{2n} - \hat{\theta}_{2n}^b \rightarrow 0$ a.s. for $n \rightarrow \infty$. 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 $b > 0$:

$$n(\hat{\theta}_{2n}^b - \theta_2^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \Pi_-^b \quad (12)$$

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

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

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

$$\begin{aligned} & n \sum_{k=1}^K \int_{|x - \tau_k^0| \leq n^{-1}b} \varphi(x) \mathbb{E}_\varepsilon [\rho(\varepsilon + d_{(\alpha_k^0, \alpha_{k-1}^0)}(x)) - \rho(\varepsilon + \text{sign}(t_k)d_k^0)] dx \\ &= n \sum_{k=1}^K \int_{|x - \tau_k^0| \leq n^{-1}b} \varphi(x) \mathbb{E}_\varepsilon [d_{(\alpha_k^0, \alpha_{k-1}^0)}(x) - d_k^0 |\psi(\varepsilon + y_x)|] dx, \quad \text{with } y_x \rightarrow 0, \text{ for } x \rightarrow 0 \\ &\leq Cn \sum_{k=1}^K \int_{|x - \tau_k^0| \leq n^{-1}b} |x - \tau_k^0| \left(\sup_\alpha \left\| \frac{\partial h_\alpha(x)}{\partial \alpha} \right\| \right) \varphi(x) \mathbb{E}_\varepsilon [\psi(\varepsilon + y_x)] dx \end{aligned}$$

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

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

Hence: $\sup_{\|t\| \leq b} |\hat{M}_n^b(t) - H_n(t)| = o_{\mathbb{P}}(1)$. 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:

$$n(\hat{\theta}_{2n}^b - \theta_2^0) - \Pi_n^b \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0, \quad \Pi_n^b \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \Pi_-^b. \quad \text{Then relation (12) follows. Because for}$$

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 (12) and $\Pi_n^b \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \Pi_-^b$, imply that the asymptotic distribution of $n(\hat{\theta}_{2n} - \theta_2^0)$ is Π_- . ■

Remark 4 *Consequence of Theorem 3.2, we can find the confidence interval or make hypothesis test for the parameter θ .*

Remark 5 *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 change-point M-estimator is asymptotically normal.*

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

References

- [1] Bai, J., Perron P., 1998, Estimating and testing linear models with multiple structural changes, *Econometrica* **66**(1), 47-78.
- [2] Ciuperca, G., 2004, Maximum likelihood estimator in a two-phase nonlinear regression model. *Statistic and Decision* **22**(4), 335-349.
- [3] Ciuperca, G., Dapzol, N., 2008, Maximum likelihood estimator in a multi-phase random regression model. *Statistics* **42**(4), 363-381.

- [4] Ciuperca, G., 2008, The M-estimation in a multi-phase random nonlinear model
(Longer version is available at <http://arXiv.org/pdf/0706.0153>)
- [5] Feder, P.I., 1975a, On asymptotic distribution theory in segmented regression problems-identified case. *Ann. Statist.* **3**, 49-83.
- [6] Feder, P.I., 1975b, The log likelihood ratio in segmented regression. *Ann. Statist.* **3**, 84-97.
- [7] 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.
- [8] Koul, H.L., Qian, L., 2002,, Asymptotics of maximum likelihood estimator in a two-phase linear regression model. *Journal of Statistical Planning and Inference* **108**, 99-119.
- [9] 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.
- [10] Lai, T.L., Robbins, H., Wei, C. Z., 1979 , Strong consistency of least squares estimates in multiple regression. *J. Multivariate Anal.* **9**, 343-361.
- [11] Rukhin, A.L., Vajda, I., 1997, Change-point estimation as a nonlinear regression problem. *Statistics* **30**(3), 181-200.
- [12] Van der Geer, S., 1988, *Regression analysis and empirical processes*. CWI Tract., Vol. 45, Amsterdam, NL.
- [13] Yao, Y.C., Au, S.T., 1989, Least-squares estimation of a step function. *Sankhya (Series A)* **51**(3), 370-381.