Tatra Mt. Math. Publ. **22** (2001), 197–208

TATRA MOUNTAINS
Mathematical Publications

M-ESTIMATORS OF STRUCTURAL CHANGES IN REGRESSION MODELS

Marie Hušková — Jaromír Antoch

Dedicated to Lubomír Kubáček on the occasion of his 70th birthday

ABSTRACT. A class of M-estimators of change point(s) in linear regression model is introduced and their properties studied. Selected conclusions of a simulation study are presented.

1. Introduction

We consider the regression model with a change after an unknown time point m_n:

$$Y_{in} = \boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{x}_i^T \boldsymbol{\delta}_n I\{i > m_n\} + e_i\,, \qquad i = 1\ldots, n\,, \tag{1}$$

where $m_n\,(\le n)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ and $\boldsymbol{\delta}_n = (\delta_{n1}, \ldots, \delta_{np})^T \neq \mathbf{0}$ are unknown parameters, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$, $x_{i1} = 1$, $i = 1, \ldots, n$, are known design points and $e_1, \ldots, e_n$ are iid random variables with common distribution F that fulfill regularity conditions specified below. Function $I\{A\}$ in (1) denotes the indicator of the set A. We write the index n with the observations Y_{in}, the parameters m_n and $\boldsymbol{\delta}_n$ because we study the limit properties as $n \to \infty$ and we assume that both m_n and $\boldsymbol{\delta}_n$ are changing when n is increasing. More precisely, we assume that, as $n \to \infty$,

$$m_n = \lfloor \lambda n \rfloor\,, \qquad \text{for some } \lambda \in (0,1]\,, \tag{2}$$

and

$$\sqrt{\log n}\; \|\boldsymbol{\delta}_n\| \to 0 \qquad \text{and} \qquad \frac{\|\boldsymbol{\delta}_n\|\sqrt{n}}{\sqrt{\log n}} \to \infty\,. \tag{3}$$

2000 Mathematics Subject Classification: 62G20, 62E20, 60F17.
Keywords: linear regression, structural changes, change point problem, L_1-, L_2- and M-estimators, Huber estimator.

The work was supported by grants GAČR 201/00/0769 and MSM 113200008.

Here $\|\cdot\|$ denotes the Euclidean norm and $\lfloor a \rfloor$ the integer part of a. It is clear that the parameter $\boldsymbol{\delta}_n$ reflects the amount of the change (size of the jump) and the assumption (1.3) describes a class of local changes.

Model (1) describes the situation where the first m_n observations follow the linear model with the parameter $\boldsymbol{\beta}$ and the remaining $n-m_n$ observations follow the linear regression model with the parameter $\boldsymbol{\beta}+\boldsymbol{\delta}_n$. The parameter m_n is usually called the *change point*. To estimate it and study its limit properties is of prime interest of the present paper.

In practice one typically observes a sequence of variables and might be interested to know whether the possible statistical model remains the same during the whole observational period or whether the model changes at some unknown time point. If the observations indicate a change then an estimator of its location is of interest. Such problems occur in various situations — the changes in hydrological, meteorological or econometric time series. For recent references, see, e.g., Csörgő and Horváth (1997).

There are papers on estimators of changes in linear models based either on the L_2-estimators, see, e.g., Hušková (1997) or Horváth, Hušková and Serbinowska (1997). Bai (1995, 1998) introduced and studied L_1-estimators. In this paper we consider a more general type of M-estimators corresponding to the Huber score generating function.

Set

$$\boldsymbol{C}_k = \sum_{i=1}^{k} \boldsymbol{x}_i \boldsymbol{x}_i^T \quad \text{and} \quad \boldsymbol{C}_k^o = \boldsymbol{C}_n - \boldsymbol{C}_k, \qquad k = 1, \dots, n. \tag{4}$$

Let us remind the definition of the L_2-estimator $\widetilde{m}_n$ of m_n which is of the form

$$\widetilde{m}_n = \min\Big\{k\colon p<k<n-p,\ \boldsymbol{S}_k^T \boldsymbol{C}_k^{-1} \boldsymbol{C}_n \boldsymbol{C}_k^{o-1} \boldsymbol{S}_k \\ = \max_{p<j<n-p} \big\{\boldsymbol{S}_j^T \boldsymbol{C}_j^{-1} \boldsymbol{C}_n \boldsymbol{C}_j^{o-1} \boldsymbol{S}_j\big\}\Big\}, \tag{5}$$

where

$$\boldsymbol{S}_k = \sum_{i=1}^{k} \boldsymbol{x}_i \big(Y_{in} - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_n\big), \qquad k = 1, \dots, n,$$

with $\widehat{\boldsymbol{\beta}}_n$ being the least squares estimator of $\boldsymbol{\beta}$ based on $Y_1, \dots, Y_n$, i.e., $\widehat{\boldsymbol{\beta}}_n = \boldsymbol{C}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i Y_i$. Replacing $\widehat{\boldsymbol{\beta}}_n$ and the residuals $Y_{in} - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_n$, $i = 1, \dots, n$, by their M-type counterparts, we get the M-estimator $\widehat{m}_n(\psi)$ of the change point m_n. More precisely, the M-estimator $\widehat{m}_n(\psi)$ of the change point m_n, generated by a score function ψ, is defined as

$$\widehat{m}_n(\psi) = \min\Big\{k\colon p<k<n-p,\ \boldsymbol{S}_k^T(\psi) \boldsymbol{C}_k^{-1} \boldsymbol{C}_n \boldsymbol{C}_k^{o-1} \boldsymbol{S}_k(\psi) \\ = \max_{p<j<n-p} \big\{\boldsymbol{S}_j^T(\psi) \boldsymbol{C}_j^{-1} \boldsymbol{C}_n \boldsymbol{C}_j^{o-1} \boldsymbol{S}_j(\psi)\big\}\Big\}, \tag{6}$$

where

$$\boldsymbol{S}_k(\psi) = \sum_{i=1}^{k} \boldsymbol{x}_i \psi\big(Y_{in} - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_n(\psi)\big)\,, \qquad k = 1, \ldots, n\,. \tag{7}$$

At last we remind the definition of the Huber score function $\psi_K(\cdot)$, $K > 0$,

$$\psi_K(x) = \begin{cases} x\,, & |x| \le K\,, \\ K \ \mathrm{sign}(x)\,, & |x| \ge K\,. \end{cases} \tag{8}$$

For simplicity we will write ψ instead of ψ_K whenever possible.

The rest of the paper is structured as follows. The main results together with some remarks are formulated in Section 2. Section 3 contains selected conclusions of a simulation study. The proof of the main results can be found in Section 4.

2. Main result

In this section we formulate and discuss the main assertion on the limit distribution of $\widehat{m}_n(\psi)$.

• We assume that the design points $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$, $i = 1, \ldots, n$, satisfy:

(A.1) $x_{i1} = 1$, $i = 1, \ldots, n$.

(A.2) There exists a positively definite $p \times p$ matrix $\boldsymbol{C}$ such that $\lim_{n\to\infty} n^{-1}\boldsymbol{C}_{[nt]} = \boldsymbol{C}(t)$, $t \in [0,1]$, where $\boldsymbol{C}(t)$ is continuous function in t and is a positive definite matrix for each $t \in (0,1]$.

(A.3) There exist $\varepsilon \in (0, 1/2)$ and $\gamma > 0$ such that, as $n \to \infty$,

$$\left\| \frac{1}{m_n - k}\big(\boldsymbol{C}_{m_n} - \boldsymbol{C}_k\big) - \boldsymbol{H} \right\| = O\big(|k - m_n|^{-\gamma}\big)$$

uniformly for $|k - m_m| \le \varepsilon$ for some positive definite matrix $\boldsymbol{H}$ and some $\varepsilon > 0$.

(A.4) It holds, as $n \to \infty$,

$$\max_{1 \le k \le n} \left\{ \frac{1}{k} \sum_{i=1}^{k} \|\boldsymbol{x}_i\|^4 + \frac{1}{n-k} \sum_{i=k+1}^{n} \|\boldsymbol{x}_i\|^4 \right\} = O(1)\,.$$

• The distribution function F of the error terms e_i's satisfies the following set of assumptions:

(B.1) F has absolutely continuous density f and finite nonzero Fisher's information $I(f)$.

(B.2) $f'(x) = df(x)/dx$.

(B.3) $f(-x) = f(x)\ \forall x \in \mathbb{R}^1$.

(B.4) $0 < f(x) < \infty$ and $f'(x)$ is bounded in a neighborhood of $K > 0$ (which will be specified later).

The assumptions imposed on the design points $\boldsymbol{x}_i$'s are slightly stronger than those considered in standard linear regression problems. However, the formulated assumptions are fulfilled for a broad spectrum of situations as, e.g., when

$$\boldsymbol{x}_i = \boldsymbol{g}(i/n)\,, \qquad i = 1\ldots, n\,,$$

where $\boldsymbol{g}$ is a vector of continuously differentiable functions on $[0,1]$. Another possibility is to choose p linearly independent p-dimensional vectors $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*$ and repeat "almost regularly", in this case $\boldsymbol{C}(t) = t\boldsymbol{C}$, $t \in (0,1)$, for some $\boldsymbol{C} > 0$.

Now, we formulate the main result.

Theorem 2.1. *Let $(Y_{1n}, \boldsymbol{x}_1), \ldots, (Y_{nn}, \boldsymbol{x}_n)$ follow the model* (1)*, let assumption* (2) *with $\lambda \in (0,1)$ be satisfied. Moreover, let assumptions* (3), (A.1)–(A.4) *and* (B.1)–(B.4) *be satisfied. Then, as $n \to \infty$,*

$$\kappa^2(\psi, F)\, \boldsymbol{\delta}_n^T \boldsymbol{H} \boldsymbol{\delta}_n (\widehat{m}_n - m_n) \xrightarrow{\mathcal{D}} V, \tag{9}$$

where

$$\kappa^2(\psi, F) = \frac{\left(\int \psi'(x)\, dF(x)\right)^2}{\int \psi^2(x)\, dF(x)} \tag{10}$$

and

$$V = \operatorname{argmax}\{W(t) - |t|/2\colon t \in \mathbb{R}^1\} \tag{11}$$

with $\{W(t)\colon t \in \mathbb{R}^1\}$ being a two-sided Wiener process.

Remark 2.1. The assertion of the theorem remains valid if $\kappa^2(\psi, F)$ is replaced by a suitable estimator, e.g., $\int \psi^2(x) dF(x)$ can be estimated by

$$\begin{aligned} \widehat{\sigma}_n^2(\psi) = \sum_{i=1}^n \psi^2\left(Y_{in} - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_n(\psi)\right) \\ - \max_{1 \le k < n} \frac{1}{k(n-k)} \left(\sum_{i=1}^k \psi\left(Y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_n(\psi)\right)\right)^2. \end{aligned} \tag{12}$$

Since ψ is considered in the form (8), we have

$$\int \psi'(x) dF(x) = F(K) - F(-K) = 2F(K) - 1\,,$$

where F is the distribution of the error terms e_i and therefore one needs only an estimator of $F(K)$.

Remark 2.2. The assertion of the theorem can be extended to other score generating functions, however, the proof of the technical part is less transparent.

Remark 2.3. The distribution of V is known and can be be found, e.g., in Horváth, Hušková and Serbinowska (1997).

Remark 2.4. The limit distribution of the estimator $\widehat{m}$ depends on the error distribution function F and on the score function ψ only through $\kappa^2(\psi, F)$. If F is known then the score function can be chosen in such a way that the asymptotic variance of $\widehat{m}$ is minimal within a class of considered score functions.

Remark 2.5. It can be shown that the main assertion holds also in the case when $\boldsymbol{x}_i$ are iid random variables with finite moment of the order $2+\kappa$, $\kappa > 0$.

3. Simulation study

It is clear from above, in this paper we were interested in two main features of the procedure under the study, i.e.:

(B.1) What is the difference between the behaviour of the considered change point estimators, i.e., the M-estimator (represented here by the Huber's estimator) and classical least squares estimator, when the errors are either normally or not normally distributed.

(B.2) What is the difference between the empirical distribution of the change point estimator under the study and the limit, i.e., asymptotic, distribution.

In order to get an idea about that, we conducted a small simulation study. More precisely, we considered several (more or less complicated) models. Here, for greater clarity (and also due to the space limitations) only selected conclusions will be presented for the following model:

$$T_i = \beta_1 + \beta_2 x_i + \big(\delta_1 + \delta_2 x_i\big)\, I[i > n\lambda] + e_i\,, \qquad i = 1, \ldots, n\,, \tag{13}$$

where $\lambda \in (0,1]$. In the simulations we put

- $n = 80$, 120 and 200;
- $\lambda = 1/4$, $1/2$, $3/4$ and 1;
- $\delta_1 = 0$, 1, 2, 3, 5 and 10;
- $\delta_2 = 0$, 1, 2, 3, 5 and 10;
- normally, Laplace and t_4 distributed errors;
- x_i uniformly distributed in $[0, 10]$.
- The number of the simulations for each setup was set to 50.000.

Note, that the choice $\lambda = 1$ corresponds to the situation of no change.

TABLE 1. Results for classical estimator $\sim m_n$ defined by (5).

n	δ_1	δ_2	err	λ	mean	std	λ	mean	std
80	0	0	N	0.25	38.87	20.39	0.75	40.80	18.38
80	1	0	N	0.25	30.54	15.41	0.75	48.98	16.80
80	2	0	N	0.25	22.21	7.12	0.75	57.06	9.46
80	3	0	N	0.25	20.20	2.21	0.75	59.91	0.73
80	0	0	t_4	0.25	42.22	19.45	0.75	39.82	20.74
80	1	0	t_4	0.25	36.39	18.43	0.75	44.75	18.38
80	2	0	t_4	0.25	26.01	13.71	0.75	53.39	13.18
80	3	0	t_4	0.25	22.19	7.83	0.75	57.13	9.07
80	0	0	N	0.50	40.51	19.67	1.00	39.28	18.58
80	1	0	N	0.50	40.93	11.90	1.00	39.39	18.49
80	2	0	N	0.50	40.08	2.82	1.00	38.82	20.45
80	3	0	N	0.50	40.03	0.36	1.00	39.98	19.51
80	0	0	t_4	0.50	41.06	20.54	1.00	39.81	19.21
80	1	0	t_4	0.50	39.03	17.13	1.00	44.78	18.96
80	2	0	t_4	0.50	41.17	8.51	1.00	40.39	19.28
80	3	0	t_4	0.50	40.23	2.51	1.00	36.98	18.75

TABLE 2. Results for robust estimator $\widehat{m}_n(\psi)$ defined by (6).

n	δ_1	δ_2	err	λ	mean	std	λ	mean	std
80	0	0	N	0.25	40.17	19.72	0.75	40.63	18.41
80	1	0	N	0.25	31.11	16.72	0.75	48.86	16.78
80	2	0	N	0.25	24.05	9.31	0.75	57.15	9.37
80	3	0	N	0.25	20.17	2.18	0.75	59.54	0.93
80	0	0	t_4	0.25	40.78	20.11	0.75	39.80	19.11
80	1	0	t_4	0.25	32.61	17.11	0.75	47.63	17.24
80	2	0	t_4	0.25	23.87	10.42	0.75	55.89	10.96
80	3	0	t_4	0.25	21.03	3.98	0.75	58.37	3.02
80	0	0	N	0.50	40.45	19.34	1.00	39.03	19.17
80	1	0	N	0.50	41.07	11.78	1.00	39.51	18.12
80	2	0	N	0.50	40.11	1.72	1.00	39.05	20.11
80	3	0	N	0.50	40.05	0.54	1.00	40.12	19.41
80	0	0	t_4	0.50	40.93	20.05	1.00	39.97	18.94
80	1	0	t_4	0.50	39.42	14.25	1.00	43.18	18.54
80	2	0	t_4	0.50	40.78	4.03	1.00	40.35	19.07
80	3	0	t_4	0.50	40.10	1.15	1.00	37.15	18.12

TABLE 3. Results for classical estimator $\widetilde{m}_n$ defined by (5).

n	δ_1	δ_2	err	λ	mean	std	λ	mean	std
80	0	0	N	0.25	38.85	19.18	0.75	40.72	20.09
80	0	1	N	0.25	35.89	18.48	0.75	46.34	17.02
80	0	2	N	0.25	37.15	19.88	0.75	55.79	10.09
80	0	3	N	0.25	36.07	19.55	0.75	57.16	8.15
80	0	0	t_4	0.25	41.12	19,43	0.75	38.01	19.17
80	0	1	t_4	0.25	39.65	20.62	0.75	43.79	18.99
80	0	2	t_4	0.25	38.66	19.78	0.75	50.84	14.68
80	0	3	t_4	0.25	36.25	20.13	0.75	54.37	11.44
80	0	0	N	0.50	39.42	19.08	1.00	38.22	18.99
80	0	1	N	0.50	38.72	18.07	1.00	39.66	19.14
80	0	2	N	0.50	37.37	10.11	1.00	38.33	19.09
80	0	3	N	0.50	38.10	9.10	1.00	38.57	19.37
80	0	0	t_4	0.50	40.84	19.69	1.00	40.78	19.20
80	0	1	t_4	0.50	39.03	19.44	1.00	41.11	20.91
80	0	2	t_4	0.50	39.08	17.13	1.00	39.34	19.69
80	0	3	t_4	0.50	39.58	12.65	1.00	39.19	19.68

TABLE 4. Results for robust estimator $\widehat{m}_n(\psi)$ defined by (6).

n	δ_1	δ_2	err	λ	mean	std	λ	mean	std
80	0	0	N	0.25	38.77	19.02	0.75	40.29	19.88
80	0	1	N	0.25	36.10	18.44	0.75	45.18	17.69
80	0	2	N	0.25	37.03	20.07	0.75	54.81	10.97
80	0	3	N	0.25	36.17	19.41	0.75	57.12	8.93
80	0	0	t_4	0.25	40.98	19.29	0.75	39.25	19.72
80	0	1	t_4	0.25	39.52	20.35	0.75	45.18	18.01
80	0	2	t_4	0.25	38.41	20.11	0.75	54.02	11.17
80	0	3	t_4	0.25	36.34	20.03	0.75	55.17	9.78
80	0	0	N	0.50	39.52	19.25	1.00	38.34	18.54
80	0	1	N	0.50	39.12	19.51	1.00	39.54	19.11
80	0	2	N	0.50	37.63	10.28	1.00	38.29	19.32
80	0	3	N	0.50	38.71	10.53	1.00	39.05	19.41
80	0	0	t_4	0.50	40.15	19.39	1.00	41.02	19.15
80	0	1	t_4	0.50	39.13	19.01	1.00	40.93	20.72
80	0	2	t_4	0.50	39.11	12.34	1.00	39.18	19.63
80	0	3	t_4	0.50	39.83	10.57	1.00	39.21	19.78

Moreover, it is clear that the proper choice of the parameters δ_1 and δ_2 covers the following situations:

- the change in the slope of the considered regression model $(\delta_1 = 0)$;
- the change in intercept among otherwise parallel regression lines $(\delta_2 = 0)$;
- there is as change in both intercept and slope $(\delta_1 \neq 0, \delta_2 \neq 0)$.

This setup resulted in a lot of variants of the considered model. To get a flavour about the behavior of our estimators and especially about the variability describing the preciseness of the estimation, Tables 1–4 and selected conclusions are included.

Selected conclusions

I) It appears that robust estimators should be used when we are not sure about the normality of the errors. While the use of the Huber's estimator (properly tuned) practically does not influence the quality of the estimation procedure, the use of the classical least squares estimator in the presence of the errors from the heavy tailed distribution (e.g., from t_4) can lead to the "disaster" even under the hypothesis. By "disaster" we mean the fact that the least squares estimator (5) is not able to estimate m_n correctly. Roughly said, it is more dispersed over all its possible values.

Moreover, it should be pointed out that this is true even in the case when the errors follow Laplace (i.e., doubly exponential) distribution. Indeed, when we have a suspicion that the errors might follow a Laplace distribution, the L_1-estimator should be used, i.e., the M-estimator with the score function $\psi(x) = \operatorname{sign}(x)$. While this conclusion can be a bit surprising for some readers, this phenomenon confirms the previous results of our studies and simulations we conducted for the case of the detection of the change point in the location case.

II) Evidently, it is more easy to detect correctly the change point (in our model) provided both intercept and slope changed than in the situation when there is a change in the intercept or slope only.

III) It is much more difficult to detect a change in slope than the change in intercept.

IV) The convergence to the limit distribution as described in Theorem 2.1 is extremely slow.

4. Proofs

To prove Theorem 2.1 we have to use a number of results proved elsewhere, mostly from Hušková (2001), or their modifications. The proofs are quite technical and we will give only their essence and avoid too many technicalities. We start with various asymptotic linearity results.

Lemma 4.1. *Let assumptions of Theorem 2.1 be satisfied and let* $q_n \to 0$. *Then for any* $c \in (0,1)$, *as* $n \to \infty$,

$$\sup_{\|\boldsymbol{t}\|\le D} \frac{1}{\sqrt{n}} \Big\| \sum_{i=1}^{n} \boldsymbol{x}_i \Big(\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t})) - \psi(e_i) - E\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\delta}_{ni}(\boldsymbol{t}))\Big) \Big\| = O_P(\|\boldsymbol{\delta}_n\| + |q_n|)\,, \tag{14}$$

$$\begin{aligned}&\max_{1\le k\le cn} \sup_{\|\boldsymbol{t}\|\le D} \frac{1}{\sqrt{k}} \Big\| \sum_{i=1}^{k} \boldsymbol{x}_i \Big(\psi(e_i+\boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t})) - \psi(e_i) - E\psi(e_i+\boldsymbol{x}_i^T \boldsymbol{\delta}_{ni}(\boldsymbol{t}))\Big) \Big\| \\ &\quad + \max_{cn\le k\le n} \sup_{\|\boldsymbol{t}\|\le D} \frac{1}{\sqrt{n-k}} \Big\| \sum_{i=k+1}^{n} \boldsymbol{x}_i \\ &\qquad \times \Big(\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t})) - \psi(e_i) - E\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\delta}_{ni}(\boldsymbol{t}))\Big) \Big\| \\ &= O_P\Big((\|\boldsymbol{\delta}_n\| + |q_n|)\sqrt{\log n}\Big)\,,\end{aligned} \tag{15}$$

$$\begin{aligned}&\max_{1\le k\le k_n} \sup_{\|\boldsymbol{t}\|\le D} \frac{1}{\sqrt{k}} \Big\{ \Big\| \sum_{i=m_n-k+1}^{m_n} \boldsymbol{x}_i \\ &\qquad \times \Big(\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t})) - \psi(e_i) - E\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t}))\Big) \Big\| \\ &\quad + \Big\| \sum_{i=m_n+1}^{k} \boldsymbol{x}_i \Big(\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t})) - \psi(e_i) - E\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t}))\Big) \Big\| \Big\} \\ &= O_P\Big((\|\boldsymbol{\delta}_n\| + |q_n|)\sqrt{\log n}\Big)\,,\end{aligned} \tag{16}$$

uniformly in $0 \le k \le k_n$, $k_n \le \min(n-m_n, m_n)$. *Moreover, for the expectations we have*

$$\begin{aligned}&\max_{1\le k< k_n} \sup_{\|\boldsymbol{t}\|\le D} \frac{1}{k} \Big\{ \Big\| E \sum_{i=m_n-k+1}^{m_n} \boldsymbol{x}_i \Big(\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t})) - \psi(e_i)\Big) \\ &\qquad + (\boldsymbol{C}_{m_n} - \boldsymbol{C}_k)(\boldsymbol{C}_n^{-1} \boldsymbol{C}_{m_n}^o \boldsymbol{\delta}_n - q_n \boldsymbol{t}) \int \psi'(x)\, dF(x) \Big\| \\ &\quad + \Big\| E \sum_{i=m_n+1}^{m_n+k} \boldsymbol{x}_i \Big(\psi(e_i + \boldsymbol{x}_i^T \boldsymbol{\Delta}_{ni}(\boldsymbol{t})) - \psi(e_i)\Big)\end{aligned}$$

$$- (\boldsymbol{C}_{m_n} - \boldsymbol{C}_k)(\boldsymbol{C}_n^{-1}\boldsymbol{C}_{m_n}\boldsymbol{\delta}_n - q_n\boldsymbol{t}) \int \psi'(x)\,dF(x)\Big\|\Big\}$$

$$= O_P\big(\|\boldsymbol{\delta}_n\|^3 + |q_n|^3\big) \tag{17}$$

uniformly in $1 \le k \le k_n$, $1 \le k_n \le \min(m_n, n - m_n)$ *and*

$$\max_{1\le k<m_n} \sup_{\|\boldsymbol{t}\|\le D} \frac{1}{k}\Big\| E\sum_{i=1}^{k} \boldsymbol{x}_i\Big(\psi\big(e_i + \boldsymbol{x}_i^T\boldsymbol{\Delta}_{ni}(\boldsymbol{t})\big) - \psi(e_i)\Big)$$
$$+ \Big(\boldsymbol{C}_k\boldsymbol{C}_n^{-1}\boldsymbol{C}_{m_n}^o\boldsymbol{\delta}_n + q_n\boldsymbol{C}_k\boldsymbol{t}\Big)\int \psi'(x)\,dF(x)\Big\|$$
$$+ \max_{m_n+1\le k<n} \sup_{\|\boldsymbol{t}\|\le D} \frac{1}{n-k}\Big\| E\sum_{i=k+1}^{n} \boldsymbol{x}_i\Big(\psi\big(e_i + \boldsymbol{x}_i^T\boldsymbol{\Delta}_{ni}(\boldsymbol{t})\big) - \psi(e_i)\Big)$$
$$- \Big(\boldsymbol{C}_k^o\boldsymbol{C}_n^{-1}\boldsymbol{C}_{m_n}\boldsymbol{\delta}_n - q_n\boldsymbol{C}_k^o\boldsymbol{t}\Big)\int \psi'(x)\,dF(x)\Big\|$$
$$= O_P\big(\|\boldsymbol{\delta}_n\|^3 + |q_n|^3\big) \tag{18}$$

for any $D > 0$, *where*

$$\boldsymbol{\Delta}_{ni}(\boldsymbol{t}) = \boldsymbol{\delta}_n I\{i > m_n\} - \boldsymbol{C}_n^{-1}\boldsymbol{C}_{m_n}^o\delta_n - q_n\boldsymbol{t}\,, \qquad \boldsymbol{t} \in \mathbb{R}^p.$$

P r o o f . See Lemma 3.1 in H u š k o v á (2001). □

LEMMA 4.2. *Let the assumptions of Lemma 4.1 be satisfied. Then, as* $n \to \infty$,

$$\widehat{\boldsymbol{\beta}}_n(\psi) - \boldsymbol{\beta} = \boldsymbol{C}_n^{-1}\boldsymbol{C}_{m_n}^o\boldsymbol{\delta}_n - \boldsymbol{C}_n^{-1}\frac{1}{\int\psi'(x)\,dF(x)}\sum_{i=1}^{n}\boldsymbol{x}_i\psi(e_i)$$
$$+ O_P\big(\|\boldsymbol{\delta}_n\|^3 + n^{-1/2}\|\boldsymbol{\delta}_n\|\big) \tag{19}$$

and

$$\widehat{\sigma}_n^2(\psi) - \int \psi^2(x)\,dF(x) = O_P\big(\|\boldsymbol{\delta}_n\|\big)\,, \tag{20}$$

where $\widehat{\sigma}_n^2(\psi)$ *is defined in* (12).

P r o o f . See Lemma 3.2 and Lemma 3.3 in H u š k o v á (2001). □

LEMMA 4.3. *Let the assumptions of Lemma 4.1 be satisfied. Then, as* $n \to \infty$,

$$\sum_{i=1}^{k}\boldsymbol{x}_i\psi\big(Y_i - \boldsymbol{x}_i^T\widehat{\beta}_n(\psi)\big) = \boldsymbol{\xi}_k(\psi) - \boldsymbol{C}_k\boldsymbol{C}_n^{-1}\boldsymbol{C}_{m_n}^0\boldsymbol{\delta}_n\int\psi'(x)\,dF(x)$$
$$+ O_P\Big(\sqrt{k\log n}\,\|\boldsymbol{\delta}_n\| + k\|\boldsymbol{\delta}_n\|^3\Big) \tag{21}$$

uniformly for $1 \le k \le m_n$, *and*

$$\sum_{i=k+1}^{n} \boldsymbol{x}_i \psi\big(Y_i - \boldsymbol{x}_i^T \widehat{\beta}_n(\psi)\big) = \boldsymbol{\xi}_k(\psi) - \boldsymbol{C}_k^o \boldsymbol{C}_n^{-1} \boldsymbol{C}_{m_n} \boldsymbol{\delta}_n \int \psi'(x)\, dF(x) + O_P\Big(\sqrt{(n-k)\log n}\, \|\boldsymbol{\delta}_n\| + (n-k)\, \|\boldsymbol{\delta}_n\|^3\Big) \tag{22}$$

uniformly for $m_n < k < n$, *where*

$$\boldsymbol{\xi}_k(\psi) = \sum_{i=1}^{k} \boldsymbol{x}_i \psi(e_i) - \boldsymbol{C}_k \boldsymbol{C}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \psi(e_i), \quad k = 1, \ldots, n. \tag{23}$$

P r o o f . It is a straightforward consequence of Lemma 4.1 and Lemma 4.2. □

P r o o f o f t h e T h e o r e m 2.1. Proceeding as in the proof of Theorem 2.1 in H u š k o v á (2001) we find that, as $n \to \infty$,

$$P\Big(\max_{1 \le k \le n} \boldsymbol{S}_k^T(\psi) \boldsymbol{C}_k^{-1} \boldsymbol{C}_n (\boldsymbol{C}_k^o)^{-1} \boldsymbol{S}_k(\psi) = \max_{m_n(1-\varepsilon_n) \le k \le m_n(1+\varepsilon_n)} \boldsymbol{S}_k^T(\psi) \boldsymbol{C}_k^{-1} \boldsymbol{C}_n (\boldsymbol{C}_k^o)^{-1} \boldsymbol{S}_k(\psi)\Big) \to 1$$

for some $\varepsilon_n \to 0$, in other words, as $n \to \infty$,

$$\frac{\widehat{m}_n - m_n}{n} \xrightarrow{\mathcal{P}} 0\,.$$

By Lemma 4.2 and after long and tedious calculations we find that, as $n \to \infty$,

$$\begin{aligned} &\boldsymbol{S}_{m_n-k}^T(\psi) \boldsymbol{C}_{m_n-k}^{-1} \boldsymbol{C}_n (\boldsymbol{C}_{m_n-k}^o)^{-1} \boldsymbol{S}_{m_n-k}(\psi) - \boldsymbol{S}_{m_n}^T(\psi) \boldsymbol{C}_{m_n}^{-1} \boldsymbol{C}_n (\boldsymbol{C}_{m_n}^o)^{-1} \boldsymbol{S}_{m_n}(\psi) \\ &\quad = \sum_{i=m_n-n+1}^{m_n} \boldsymbol{x}_\psi(e_i) \boldsymbol{\delta}_n \int \psi'(x)\, dF(x) \\ &\qquad - |m_n - k|\, \boldsymbol{\delta}_n^T \boldsymbol{H} \boldsymbol{\delta}_n \Big(\int \psi'(x)\, dF(x)\Big)^2 + o_P\Big(\sqrt{|m_n - k|} \cdot \|\boldsymbol{\delta}_n\|^2\Big) \end{aligned}$$

uniformly in $|k - m_n| \le n\varepsilon_n$ for any $\varepsilon_n \to 0$. This further implies, as $n \to \infty$, that

$$\big\|\boldsymbol{\delta}_n\big\|^2 (\widehat{m}_n - m_n) = O_P(1)\,.$$

The proof is finished similarly as the proof of Theorem A in H u š k o v á (1996). □

REFERENCES

[1] ANTOCH, J.—HUŠKOVÁ, M.: *Estimators of changes*, in: Asymptotics, Nonparametrics and Time Series (Subir Ghosh, ed.), Marcel Dekker, New York, 1999, pp. 533–578.

[2] ANTOCH, J.—HUŠKOVÁ, M.—JARUŠKOVÁ, D.: *Off-line quality control*, in: Multivariate Total Quality Control: Foundation and Recent Advances (Lauro N. C. et al., eds.), Springer-Verlag, Heidelberg, 2001, pp. 1–86.

[3] BAI, J.: *Least absolute deviation estimation of a shift*, Econometric Theory **11** (1995), 403–436.

[4] BAI, J.: *Estimation of multiple-regime regressions with least absolute deviation*, J. Statist. Planning and inference **74** (1998), 103–134.

[5] CSÖRGŐ, M.—HORVÁTH, L.: *Limit Theorems in Change-point Analysis*, J. Wiley, New York, 1997.

[6] HUŠKOVÁ, M.: *L_1-test procedures for detection of change*, in: L_1-Statistical Procedures and Related Topics (Y. Dodge, ed.), IMS Lecture Notes-Monograph Series, Vol. **31**, pp. 56–70, 1997.

[7] HUŠKOVÁ, M.: *Some invariant test procedures for detection of structural changes*, Kybernetika **36** (2000), 401–414.

[8] HUŠKOVÁ, M.: *Some invariant test procedures for detection of structural changes. Behavior under alternatives.* Kybernetika **37** (2002), 669–684.

[9] HORVÁTH, L.—HUŠKOVÁ, M.—SERBINOWSKA, M.: *Estimators for the time of changes in linear models*, Statist. **29** (1997), 109–130.

Received May 14, 2001

Charles University of Prague
Department of Statistics, Sokolovská 83
CZ – 186 75 Praha 8
CZECH REPUBLIC

E-mail: marie.huskova@karlin.mff.cuni.cz
jaromir.antoch@karlin.mff.cuni.cz