Retrospective change-point detection and estimation in multivariate linear models

Boris Brodsky, Boris Darkhovsky

Abstract

In this paper the problem of retrospective change-point detection and estimation in multivariate linear models is considered. The lower bounds for the error of change-point estimation are proved in different cases (one change-point: deterministic and stochastic predictors, multiple change-points). A new method for retrospective change-point detection and estimation is proposed and its main performance characteristics (type 1 and type 2 errors, the error of estimation) are studied for dependent observations in situations of deterministic and stochastic predictors and unknown change-points. We prove that this method is asymptotically optimal by the order of convergence of change-point estimates to their true values as the sample size tends to infinity. Results of a simulation study of the main performance characteristics of proposed method in comparison with other well known methods of retrospective change-point detection and estimation are presented.

Keywords: change-point; retrospective detection and estimation; performance measure; asymptotic optimality

1 Introduction

This paper deals with change-point problems for multivariate linear models. We begin with a short review of this field.

The change-point problem for regression models was first considered by Quandt (1958, 1960). Using econometric examples Quandt proposed a method for estimation of a change-point in a sequence of independent observations based upon the likelihood ratio test.

Let us describe the change-point problem for the linear regression models considered in the literature. Let $y_1, y_2, \ldots, y_n$ be independent random variables (i.r.v.'s). Under the null hypothesis $\mathbf{H}_0$ the linear model is

$$y_i = \mathbf{x}_i^* \beta + \epsilon_i, \quad 1 \le i \le n,$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_d)^*$ is an unknown vector of coefficients, $\mathbf{x}_i^* = (1, x_{2i}, \ldots, x_{di})$ are known predictors (here and below $*$ is the transposition symbol).

The errors ϵ_i are supposed to be independent identically distributed random variables (i.i.d.r.v.'s) with $\mathbf{E}\epsilon_i = 0, \quad 0 < \sigma^2 = var\, \epsilon_i < \infty$.

Under the alternative hypothesis $\mathbf{H}_1$ a change at the instant k^* occurs, i.e.

$$y_i = \begin{cases} \mathbf{x}_i^* \beta + \epsilon_i, & 1 \le i \le k^* \\ \mathbf{x}_i^* \gamma + \epsilon_i, & k^* < i \le n, \end{cases}$$

where k^* and $\gamma \in \mathbb{R}^d$ are unknown parameters, and $\beta \ne \gamma$.

Denote

$$\bar{y}_k = \frac{1}{k} \sum_{1 \le i \le k} y_i, \bar{\mathbf{x}}_k = \frac{1}{k} \sum_{1 \le i \le k} \mathbf{x}_i,$$
$$Q_n = \sum_{1 \le i \le n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^*$$

and $\mathbf{X}_n = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^*$, $Y_n = (y_1, y_2, \ldots, y_n)^*$.

The least square estimate of β is:

$$\hat{\beta}_{\mathbf{n}} = (\mathbf{X}_n^* \mathbf{X}_n)^{-1} \mathbf{X}_n^* Y_n.$$

Siegmund with co-authhours (James, James, Siegmund (1989)) proposed to reject $\mathbf{H}_0$ for the large values of $\max_{1 \le k \le n} |U_n(k)|$, where

$$U_n(k) = \left(\frac{k}{1 - k/n}\right)^{1/2} \frac{\bar{y}_k - \bar{y}_n - \hat{\beta}_n(\bar{\mathbf{x}}_{\mathbf{k}} - \bar{\mathbf{x}}_n)^*}{(1 - k(\bar{x}_k - \bar{x}_n)(\bar{x}_k - \bar{x}_n)^*/(Q_n(1 - k/n)))^{1/2}}.$$

Earlier, Brown, Durbin, and Evans (1975) used the cumulative sums of regression residuals

$$\sum_{1 \le i \le k} (y_i - \bar{y}_n - \hat{\beta}_n(\mathbf{x}_i - \bar{\mathbf{x}}_n)^*), \quad 1 \le k \le n.$$

It is easy to see that

$$U_n(k) = w_n(k)\, R_n(k)$$
$$R_n(k) = \left(\frac{n}{k(n-k)}\right)^{1/2} \sum_{1 \le i \le k} (y_i - \bar{y}_n - \hat{\beta}_n(\mathbf{x}_i - \bar{\mathbf{x}}_n)^*)$$
$$w_n(k) = 1 - k(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)^*/(Q_n(1 - k/n)))^{-1/2}.$$

The functionals of $U_n(k)$ and $R_n(k)$ were used as the test statistics for detection of change-points in regression relashionships.

Kim and Siegmund (1989) obtained the limit distribution of $\max\limits_{1 \leq k < n} |U_n(k)|$. Alternatively, Maronna and Yohay (1978), and Worsley (1986) used the maximum likelihood method for testing $\mathbf{H}_0$ against $\mathbf{H}_1$ for Gaussian errors. Later Gombay and Horvath (1994) studied the limit distributions of statistics $Z_n(i,j) = \max\limits_{i \leq k < j} |U_n(k)|$, $T_n(i,j) = \max\limits_{i \leq k < j} |R_n(k)|$ for deterministic and stochastic regression plans. The monograph by Csorgo and Horvath (1997) puts together various results in detection of structural changes in regression models.

Besides change-point detection problems, results in change-point estimation for regressions are of especial practical importance. This theme is considered in papers by Darkhovsky (1995), Huskova (1996), Horvath, Huskova, and Serbinovska (1997). In two last papers the asymptotical characteristics of change-point estimates based upon the maximum likelihood statistics are studied. For the case of contiguous alternatives, the limit distribution of the change-point estimates is obtained and weak and strong consistency of these estimates is proved. The paper by Darkhovsky (1995) develops the nonparametric approach to retrospective change-point estimation. Here the limit characteristics of change-point estimates in the functional regression model are studied without the contiguity assumption, and the rate of convergence of these estimates to the 'true' change-point parameters is estimated. Some generalizations of these results can be found in the monograph by Brodsky and Darkhovsky (2000).

A new wave of research interest to change-point problems in regressions was formed in 2000s. Different generalizations to change-point problems for autoregressive time series (Huskova, Praskova, Steinebach (2007, 2008), Gombay (2008)), for multiple change-point estimation in non-stationary time series (Davis, Lee, Rodriguez-Yam (2006)), for testing change-points in covariance structure of linear processes (Berkes, Gombay, Horvath (2009)) were studied.

However, as a result we see the multitude of methods proposed for solving different change-point problems in linear relationships and almost no theoretical approaches to their *comparative analysis*. We cannot even estimate the asymptotic efficiency of these methods. All that is empirically observed for 'structural breaks' tests in statistics and econometrics can be reduced to the following 'vague' statement: the power of these methods is rather low. Let us agree that this 'practical conclusion' requires a more serious verification.

In this paper, we pursue the following main goals:

1) To prove the prior theoretical lower bounds for the error probability in change-point estimation in multivariate models. These bounds provide the theoretical basis for the proofs of the asymptotic optimality of change-point estimates and for the comparative analysis of these estimates;

2) To propose a new nonparametric method for the problem of retrospective change-point detection and estimation in multivariate linear systems. Then we study the main performance characteristics of this method: type 1 and type 2 errors, the error of change-point estimation.

3) For the problem of multiple change-point detection and estimation, to propose a general statement in which both *the number of change-points and their coordinates in the sample are unknown.* For this problem statement, to propose a new asymptotically optimal method which gives consistent estimates of an unknown number of change-points and their coordinates.

The structure of this paper is as follows. In Section 2 the general change-point problem for multivariate linear systems is formulated and general assumptions are given. In Section 3 we prove the prior informational inequalities for the main performance characteristic of the retrospective change-point problem, namely, the error of change-point estimation. The lower bounds for the error of estimation are found in different situations of change-point detection (deterministic and stochastic regression plan, multiple change-points). In Section 4 we propose a new method for the retrospective change-point detection and estimation in multivariate linear models and study its main performance characteristics (type 1 and type 2 errors, the error of estimation) in different situations of change-point detection and estimation (dependent observations, deterministic and stochastic regression plan, multiple change-points). We prove that this method is asymptotically optimal by the order of convergence of change-point estimates to their true values as the sample size tends to infinity. In Section 5 a variant of the functional limit theorem in the case of absence of change-points is given. In Section 6 a simulation study of characteristics of the proposed method for finite sample sizes is performed. The main goals of this study are as follows: to compare performance characteristics of the proposed method with characteristics of other well known methods of change-point detection in linear regression models, to consider more general multivariate linear models and performance characteristics of the proposed method in these multivariate models. Section 7 contains main conclusions.

2 Problem statement and general assumptions

2.1 General model

The following basic specification of the multivariate system with structural changes is considered:

$$\mathbf{Y}(n) = \mathbf{\Pi X}(n) + \nu_n, \quad n = 1, \ldots, N \tag{1}$$

where $\mathbf{Y}(n) = (y_{1n}, \ldots, y_{Mn})^*$ is the vector of endogenous variables, $\mathbf{X}(n) = (x_{1n}, \ldots, x_{Kn})^*$ is the vector of pre-determined variables, Π is $M \times K$ matrix, $\nu_n = (\nu_{1n}, \ldots, \nu_{Mn})^*$ is the vector of random errors.

The matrix $\Pi = \Pi(\vartheta, n)$, $\vartheta = (\theta_1, \ldots, \theta_k)$ can change abruptly at some unknown change-points $m_i = [\theta_i N]$, $i = 1, \ldots, k$ (here and below $[a]$ denote the integer part of number a), i.e.,

$$\mathbf{\Pi}(\vartheta, n) = \sum_{i=1}^{k+1} \mathbf{a}_i \, \mathbb{I}([\theta_{i-1}N] < n \le [\theta_i N]),$$

where θ_i are unknown change-point parameters such that $0 \equiv \theta_0 < \theta_1 < \ldots \theta_k < \theta_{k+1} \equiv 1$, $\mathbf{a}_i \ne \mathbf{a}_{i+1}$, $i = 1, \ldots, k$ are unknown matrices (here and below $\mathbb{I}(A)$ is the indicator of the set A).

The problem is to estimate the unknown parameters θ_i (and therefore, the change-points m_i) by observations $\mathbf{Y}(i), \mathbf{X}(i)$, $i = 1, \ldots, N$ (the case $\theta_i \equiv 1, i = 1, \ldots, k$ corresponds to the model without change-points).

Therefore, first, we need to test an obtained dataset of observations for the presence of change-points. Second, in the case of a rejected stationarity hypothesis, we wish to estimate all detected change-points.

Model (1) generalizes many widely used regression models, namely:

a) *autoregression model (AR)*

$$y_n = c_0 + c_1 y_{n-1} + \cdots + c_m y_{n-m} + \nu_n,$$

Here $\mathbf{X}(n) = (1, y_{n-1}, \ldots, y_{n-m})^*$, $\mathbf{\Pi} = (c_0, c_1, \ldots, c_m)$.

b) *autorgression-moving average (ARMA) model*

$$y_n = c_1 y_{n-1} + \cdots + c_k y_{n-k} + d_1 u_{n-\Delta} + \cdots + d_m u_{n-\Delta-m} + \nu_n,$$

where u_n is the input variable, y_n is the output variable at the instant n, Δ is the delay time. Here $\mathbf{X}(n) = (y_{n-1}, \ldots, y_{n-m}, u_{n-\Delta}, \ldots, u_{n-\Delta-m})^*$, $\mathbf{\Pi} = (c_1, \ldots, c_k, d_1, \ldots, d_m)$.

c)*multi-factor regression model*

$$y_n = c_1 y_{n-1} + \cdots + c_k y_{n-m} + \sum_{i=1}^{r} \sum_{j=1}^{l_i} d_{ij} x_i(n-j) + \nu_n,$$

where $r, m, l_i \geq 1$. Here $\mathbf{X}(n) = (y_{n-1}, \ldots, y_{n-m}, x_1(n-1), \ldots, x_1(n-l_1), x_2(n-1), \ldots, x_2(n-l_2), \ldots, x_r(n-1), \ldots, x_r(n-l_r))^*$, $\mathbf{\Pi} = (c_1, \ldots, c_k, d_{11}, \ldots, d_{rl_r})$.

d)*simultaneous equation systems (SES)*

$$B\mathbf{Y}(n) + \Gamma\mathbf{X}(n) = \epsilon_n,$$

where $\mathbf{Y}(n) = (y_{1n}, y_{2n}, \ldots, y_{Mn})^*$ is the vector of endogenous variables, $\mathbf{X}(n) = (x_{1n}, x_{2n}, \ldots, x_{Kn})^*$ is the vector of pre-determined variables (all exogenous variables plus lagged endogenous variables), $\epsilon_n = (\epsilon_{1n}, \epsilon_{2n}, \ldots, \epsilon_{Mn})^*$ is the vector of random errors, B is a $M \times M$ non-degenerate matrix ($\det B \neq 0$), Γ is a $M \times K$ matrix.

This general structural form of the SES can be written in the following *reduced form*:

$$\mathbf{Y}(n) = -B^{-1}\Gamma\mathbf{X}(n) + B^{-1}\epsilon_n = \mathbf{\Pi}\mathbf{X}(n) + \nu_n$$

This system is usually used for the analysis of change-points (structural changes) in multivariate linear models (see, e.g., Bai, Lumsdaine, Stock (1998)).

2.2 General assumptions

In this subsection we formulate general assumptions which will be used in our main theorems 3-5. Some specific assumptions will be formulated together with the corresponding theorems.

Let us start from the following definitions. Consider the probability space $(\Omega, \mathfrak{F}, \mathbf{P})$. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two σ-algebras from $\mathfrak{F}$. Consider the following measure of dependence between $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\psi(\mathcal{H}_1, \mathcal{H}_2) = \sup_{A \in \mathcal{H}_1, B \in \mathcal{H}_2, \mathbf{P}(A)\mathbf{P}(B) \neq 0} \left| \frac{\mathbf{P}(AB)}{\mathbf{P}(A)\mathbf{P}(B)} - 1 \right|$$

Suppose $(X_i, i \geq 1)$ is a sequence of random vectors defined on $(\Omega, \mathfrak{F}, \mathbf{P})$. Denote by $\mathfrak{F}_s^t = \sigma\{X_i : s \leq i \leq t\}, 1 \leq s \leq t < \infty$ the minimal σ-algebra generated by random vectors $X_i, s \leq i \leq t$. Define

$$\psi(n) = \sup_{t \geq 1} \psi(\mathfrak{F}_1^t, \mathfrak{F}_{t+n}^\infty)$$

A) *Mixing condition*

We say that scalar random sequence $\{x_n\}$ satisfies the ψ-*mixing condition* if the function $\psi(n)$ (which is also called the ψ-*mixing coefficient*) tends to zero as n goes to infinity.

We say that vector random sequence $\{X(n)\}$, $X(n) = (x_1(n), \ldots, x_k(n))^*$ satisfies the *uniform ψ-mixing condition* if $\max\limits_{i,j} \psi_{ij}(n)$ tends to zero as n goes to infinity, where $\psi_{ij}(n)$ is the ψ-mixing coefficient for the sequence $\{x_i(n)x_j(n)\}$.

The ψ-mixing condition is satisfied in most practical situations of change-point detection. In particular, for a Markov chain (not necessarily stationary), if $\psi(n) < 1$ for a certain n, then $\psi(k)$ goes to zero at least exponentially as $k \to \infty$ (see Bradley, 2005, theorem 3.3).

B) *Cramer condition*

Let $\{\zeta(n)\}$, $\zeta(n) = (\zeta_1(n), \ldots, \zeta_k(n))^*$ be a vector random sequence. We say that the *uniform Cramer condition* is satisfied if there exists a constant $L > 0$ such that

$$\sup_n \mathbf{E} \exp\left(t\zeta_i(n)\zeta_j(n)\right) < \infty$$

for every $i, j = 1, \ldots, k$ and $|t| < L$.

For a centered random sequence ξ_n this condition is equivalent to the following: there exist constants $g > 0$, $T > 0$ such that for each $|t| < T$:

$$\sup_n \mathbf{E} e^{t\xi_n} \leq \exp\left(\frac{t^2 g}{2}\right).$$

3 Preliminary results: prior inequalities

3.1 Unique change-point

On a probability space $(\Omega, \mathcal{F}, \mathbf{P}_\theta)$ consider a sequence of i.r.v.'s $x_1, \ldots, x_N$ with the following density function (w.r.t. some σ-finite measure μ)

$$f(x_n) = \begin{cases} f_0(x_n, n/N), & 1 \leq n \leq [\theta N], \\ f_1(x_n, n/N), & [\theta N] < n \leq N. \end{cases} \tag{2}$$

Here $0 < \theta < 1$ is an *unknown change-point parameter*.

Define the following objects:

$$T_N(\Delta) : \mathbb{R}^N \longrightarrow \Delta \subset \mathbb{R}^1 \tag{3}$$

is the Borel function on $\mathbb{R}^N$ with the values in the set Δ;

$$\mathcal{M}_N(\Delta) = \{T_N(\Delta)\} \tag{4}$$

is the collection of all Borel functions T_N.

Theorem 1. *Suppose the following assumption is satisfied:*
the functions $J_0(t) \overset{def}{=} \mathbf{E}_0 \ln \dfrac{f_0(x,t)}{f_1(x,t)}$ *and* $J_1(t) \overset{def}{=} \mathbf{E}_1 \ln \dfrac{f_1(x,t)}{f_0(x,t)}$ *are continuous at*
$[0,1]$ *and such that*

$$J_0(t) \geq \delta > 0, \ J_1(t) \geq \delta > 0.$$

Then for any fixed $0 < \theta < 1$, $0 < \epsilon < \theta \wedge (1 - \theta)$ the following inequality holds:

$$\liminf_{N \to \infty} N^{-1} \ln \inf_{\hat{\theta}_N \in \mathcal{M}_N((0,1))} \mathbf{P}_\theta\{|\hat{\theta}_N - \theta| > \epsilon\} \geq -\min\left(\int\limits_\theta^{\theta+\epsilon} J_0(t)dt, \int\limits_{\theta-\epsilon}^\theta J_1(t)dt\right).$$

Remark 1. *The lower bound in Theorem 1 can not be improved essentially. It follows from the results of Korostelev (1997). In this work the exact lower bound for the change-point estimate in continuous time model for the Wiener process was given. The exact lower bound in Korostelev (1997) differs from our bound only by a constant factor.*

Consider the following particular cases of model (2).

1. A break in the trend function $\phi(t)$ of the mathematical expectation of Gaussian observations

Let
$$
\begin{aligned}
f_0(x,t) &= h(x)\exp\left(\phi_0(t)x - \phi_0^2(t)/2\right), \quad t \leq \theta \\
f_1(x,t) &= h(x)\exp\left(\phi_1(t)x - \phi_1^2(t)/2\right), \quad t > \theta,
\end{aligned}
$$

where $h(x) = \dfrac{1}{\sqrt{2\pi}}\exp(-x^2/2)$, $\phi_0(\cdot) \neq \phi_1(\cdot)$.

In this case from Theorem 1 we obtain the following lower bound for the error probability:

$$\mathbf{P}_\theta\{|\hat{\theta}_N - \theta| > \epsilon\} \geq (1 - o(1)) \cdot$$

$$\cdot \exp\left(-\frac{N}{2}\min\left(\int\limits_\theta^{\theta+\epsilon}(\phi_0(t) - \phi_1(t))^2 dt, \int\limits_{\theta-\epsilon}^\theta(\phi_0(t) - \phi_1(t))^2\, dt\right)\right).$$

2. Linear regression with deterministic predictors and Gaussian errors

8

Let

$$y_n = c_1(n)x_{1n} + \cdots + c_k(n)x_{kn} + \xi_n, \quad n = 1, \ldots, N, \tag{5}$$

where $\{\xi_n\}$ is a sequence of independent Gaussian r.v.'s with zero mean, $\xi_n \sim \mathcal{N}(0, \sigma^2)$, $\mathbf{c}(n) \stackrel{\text{def}}{=} (c_1(n), \ldots, c_k(n))^* = \mathbf{a}\mathbb{I}(n \leq [\theta N]) + \mathbf{b}\mathbb{I}(n > [\theta N])$, $\mathbf{a} = (a_1, \ldots, a_k)^* \neq \mathbf{b} = (b_1, \ldots, b_k)^*$, $x_{in} = f_i(n/N)$, $n = 1, \ldots, N$, and $f_i(\cdot) \in C[0, 1]$, $i = 1, \ldots, k$.

In this case from Theorem 1 applied to the sequence of observations $y_1, \ldots, y_N$ we obtain:

$$\mathbf{P}_\theta\{|\hat{\theta}_N - \theta| > \epsilon\} \geq (1 - o(1)) \cdot$$
$$\cdot \exp\left(-\frac{N}{2\sigma^2} \min\left(\int_\theta^{\theta+\epsilon} (\sum_{i=1}^k f_i(t)(a_i - b_i))^2 dt, \int_{\theta-\epsilon}^\theta (\sum_{i=1}^k f_i(t)(a_i - b_i))^2 dt\right)\right).$$

3. Linear stochastic regression model with Gaussian predictors

Consider model (5) with $\xi_n \equiv 0$. Suppose that there exist continuous functions $f_i(\cdot), \sigma_i(\cdot)$, $i = 1, \ldots, k$ such that x_{in} are Gaussian i.r.v.'s, $x_{in} \sim \mathcal{N}\left(f_i(n/N), \sigma_i^2(n/N)\right)$, $n = 1, \ldots, N$. Suppose also that x_{in} and x_{jn} are independent for $i \neq j$ and $\mathbf{c}(n)$ is the same as in model (5).

Then from Theorem 1 we obtain:

$$\mathbf{P}_\theta\{|\hat{\theta}_N - \theta| > \epsilon\} \geq (1 - o(1)) \exp\left(-\frac{N}{2} \min\left(\int_\theta^{\theta+\epsilon} J_0(t)dt, \int_{\theta-\epsilon}^\theta J_1(t)dt\right)\right),$$

where

$$J_0(t) = \left(\frac{\phi_0(t)}{\Delta_0(t)} - \frac{\phi_1(t)}{\Delta_1(t)}\right)^2 + 2\frac{\phi_0(t)}{\Delta_0(t)}\frac{\phi_1(t)}{\Delta_1(t)}\left(1 - \frac{\Delta_0(t)}{\Delta_1(t)}\right) +$$
$$2\ln\frac{\Delta_1(t)}{\Delta_0(t)} + \left(1 + \frac{\phi_0^2(t)}{\Delta_0^2(t)}\right)\left(\frac{\Delta_0(t)}{\Delta_1(t)} - 1\right),$$

and

$$\phi_0(t) = a_1 f_1(t) + \cdots + a_k f_k(t), \ \Delta_0^2(t) = a_1^2\sigma_1^2(t) + \cdots + a_k^2\sigma_k^2(t),$$
$$\phi_1(t) = b_1 f_1(t) + \cdots + b_k f_k(t), \ \Delta_1^2(t) = b_1^2\sigma_1^2(t) + \cdots + b_k^2\sigma_k^2(t).$$

3.2 Multiple change-points

Theorem 1 can be generalized to the case of several change-points in the sequence of independent r.v.'s with the following density function:

$$f(x_n) = f_i(x_n, n/N)\, \mathbb{I}([\theta_{i-1}N] < n \leq [\theta_i N]), \quad n = 1, \ldots, N,$$

where $i = 1, \ldots, k+1$ and $0 \equiv \theta_0 < \theta_1 < \cdots < \theta_k < \theta_{k+1} \equiv 1$.

Suppose the following assumptions are satisfied:

i) change-points θ_i are such that $\min\limits_{1 \leq i \leq k+1} (\theta_i - \theta_{i-1}) \geq \delta > 0$.

ii) the functions $J_i(t) = \mathbf{E}_i \ln \dfrac{f_i(x,t)}{f_{i-1}(x,t)}$ and $J^{i-1}(t) = \mathbf{E}_{i-1} \ln \dfrac{f_{i-1}(x,t)}{f_i(x,t)}, i = 1, \ldots, k$
are continuous at $[0,1]$ and such that

$$J_i(t) \geq \Delta > 0, \ i = 1, \ldots, k$$

For the multiple change-point problem we estimate both the number k and the vector $\vartheta \overset{\text{def}}{=} (\theta_1, \ldots, \theta_k)$ of change-points' coordinates. Let $s^* \overset{\text{def}}{=} [1/\delta]$ and denote $Q = \{1, 2, \ldots, s^*\}$.

For any $s \in Q$ define

$$
\begin{aligned}
\mathcal{D}_s &= \{x \in \mathbb{R}^s : \delta \leq x_i \leq 1 - \delta, \ x_{i+1} - x_i \geq \delta, x_0 \equiv 0, x_{s+1} \equiv 1\} \\
\mathcal{D}^\star &= \textstyle\bigcup_{i=1}^{s^*} \mathcal{D}_i, \mathcal{D}^\star \subset \mathbb{R}^{s^*} \equiv \mathbb{R}^\star
\end{aligned}
\tag{6}
$$

By the construction, an unknown vector ϑ is an arbitrary point of the set $\mathcal{D}_k$ and an unknown number of the change-points k is an arbitrary point of the set Q.

As before, it is reasonable to consider objects (3)-(4). In this notation $\mathcal{M}_N(\mathcal{D}^*)$ is the set of all arbitrary estimates of the parameter ϑ and $\mathcal{M}_N(Q)$ is the set of all arbitrary estimates of the parameter k on the basis of observations with the sample size N.

Let $\hat{k} \in \mathcal{M}_N(Q)$ is an estimate of an unknown number of change-points k and $\hat{\vartheta} \in \mathcal{M}_N(\mathcal{D}_k)$ is an estimate of unknown change-point coordinates on condition that the number of the coordinates was estimated correctly.

Theorem 2. *Suppose assumptions i) and ii) are satisfied. Then for any fixed $0 < \epsilon < \delta$ the following inequality holds:*

$$
\liminf_{N \to \infty} N^{-1} \ln \inf_{\hat{\vartheta} \in \mathcal{M}_N(\mathcal{D}_k)} \inf_{\hat{k} \in \mathcal{M}_N(Q)} \sup_{\vartheta \in \mathcal{D}_k} \sup_{k \in Q} \mathbf{P}_\theta \{\{\hat{k} \neq k\} \cup \{(\hat{k} = k) \cap
$$

$$
\cap (\max_{1 \leq i \leq k} |\hat{\theta}_i - \theta_i| > \epsilon)\} \geq - \min_{1 \leq i \leq k} \min\left(\int_{\theta_i}^{\theta_i + \epsilon} J^{i-1}(\tau) d\tau, \int_{\theta_i - \epsilon}^{\theta_i} J_i(\tau) d\tau \right).
$$

4 Main results

Now consider model (1). In this Section we assume that the uniform mixing condition (A) and the uniform Cramer condition (B) (see Section 2) are satisfied, and an unknown

vector of change-point parameters $\vartheta = (\theta_1, \ldots, \theta_k)$ is such that $0 < \beta \leq \theta_1 < \theta_2 < \cdots < \theta_k \leq \alpha < 1$, where β, α are known numbers. Everywhere below the measure $\mathbf{P}_\vartheta$ corresponds to a sample with the change-point ϑ ($\mathbf{P}_0$ corresponds to a sample without change-points).

4.1   Unique change-point

In this subsection model (1) with unique change-point $0 < \beta \leq \theta \leq \alpha < 1$ is considered.

4.1.1   Deterministic predictors

Let us formulate assumptions for model (1) in the case of a unique change-point (remind that in model (1) the vector $\mathbf{X}(n)$ has the dimension K and the vector $\mathbf{Y}(n)$ has the dimension M):

a) the vector random sequence $\{\nu_n\}$ satisfies conditions (A) and (B) (see section 2).

b) there exist functions $f_i(\cdot) \in C[0, 1]$, $i = 1, \ldots, K$ such that $x_{in} = f_i(n/N)$, $n = 1, \ldots, N$.

Denote $F(t) = (f_1(t), \ldots, f_K(t))^*$, $t \in [0, 1]$.

c) for arbitrary $0 \leq t_1 < t_2 \leq 1$, the matrix

$$A(t_1, t_2) \overset{\text{def}}{=} \int_{t_1}^{t_2} F(s) F^*(s) \, ds$$

is positive definite (below we denote $A(t) \overset{\text{def}}{=} A(0, t)$, $A(1) \overset{\text{def}}{=} I$).

In virtue of our assumptions, the matrix I is symmetric and positive definite.

Define $K \times M$ matrix

$$Z(n_1, n_2) = \sum_{i=n_1}^{n_2} F(i/N) \mathbf{Y}^*(i)$$

and $K \times K$ matrix

$$\mathcal{P}_{n_1}^{n_2} \overset{\text{def}}{=} \sum_{k=n_1}^{n_2} F(k/N) F^*(k/N), \quad 1 \leq n_1 < n_2 \leq N.$$

The following matrix statistic is used for estimation of an unknown change-point:

$$\mathcal{Z}_N(n) = N^{-1} \left(Z(1, n) - \mathcal{P}_1^n (\mathcal{P}_1^N)^{-1} Z(1, N) \right). \tag{7}$$

An arbitrary point $\hat{n}$ of the set $arg \max_{[\beta N] \leq n \leq [\alpha N]} \|\mathcal{Z}_N(n)\|^2$ is assumed to be the estimate of an unknown change-point (here and below $\|C\|$ denotes the Gilbert norm of a quadratic matrix C, namely $\|C\| = \sqrt{tr(CC^*)}$).

We define also the value $\hat{\theta}_N = \hat{n}/N$ - the estimate of the change-point parameter θ. Denote $B \overset{\text{def}}{=} B(\theta) = (E - I^{-1}A(\theta))(\mathbf{a} - \mathbf{b})^*$.

Theorem 3. *Suppose assumptions a)–c) are satisfied and rank$(B) = M$ if $\theta \in [\beta, \alpha]$.*

Then the estimate $\hat{\theta}_N$ converges to the change-point parameter θ $\mathbf{P}_\theta$-almost surely as $N \to \infty$.

Besides, for any fixed $(\alpha - \beta) > \epsilon > 0$ the following inequality is satisfied for $N > N_0(F)$:

$$
\sup_{\beta \leq \theta \leq \alpha} \mathbf{P}_\theta\{|\hat{\theta}_N - \theta| > \epsilon\} \leq m_0\left(C(\epsilon, N)/\mathcal{R}\right)
\begin{cases}
\exp\left(-\dfrac{N\beta\left(C(\epsilon, N)/\mathcal{R}\right)^2}{4gm_0\left(C(\epsilon, N)/\mathcal{R}\right)}\right), \\
\quad \text{if } C(\epsilon, N) \leq \mathcal{R}gT \\
\exp\left(-\dfrac{TN\beta\left(C(\epsilon, N)/\mathcal{R}\right)}{4m_0\left(C(\epsilon, N)/\mathcal{R}\right)}\right), \\
\quad \text{if } C(\epsilon, N) > \mathcal{R}gT.
\end{cases}
\tag{8}
$$

where the constants g, T, $m_0(\cdot) \geq 1$ are taken from the uniform Cramer's and ψ-mixing conditions, respectively, $C(\epsilon, N) = \left[\dfrac{\epsilon\lambda_F}{4\mathcal{M}}\|\mathbf{a} - \mathbf{b}\|^2 - L_F/N\right]$, $N_0(F)$, λ_F, L_F, $\mathcal{R}$ are constants which can be exactly calculated for any given family of functions $F(t)$, and the constant $\mathcal{M}$ is given in the proof.

Remark 2. *The assumption rank$B = M$ yields $K \geq M$, i.e., the number M of endogenous variables in (1) cannot exceed the number K of pre-determined variables. Note that for one regression equation this assumption is always satisfied.*

Remark 3. *For independent random errors $m_0(\epsilon) = 1$.*

Remark 4. *Comparing theorems 1 and 3, we conclude that the order of convergence of the proposed estimate of the change-point parameter to its true value is asymptotically optimal as $N \to \infty$.*

Remark 5. *For any given family of functions $F(t)$ one can calculate the function $f(t) = \|m(t)\|^2$, $m(t) = \lim_{N \to \infty} \mathbf{E}_\theta \mathcal{Z}_N([Nt])$ (see the proof) and investigate this function on the square $(\theta, t) \in [\beta, \alpha] \times [\beta, \alpha]$. Such investigation gives the opportunity to calculate all constants from the formulation.*

From the proof we obtain the following

Corollary 1. *Let $C > 0$ be the decision threshold and $\mathbb{C} \overset{def}{=} C - \dfrac{L_F}{N}$. Then:*

- for type 1 error the following inequality is satisfied:

$$\mathbf{P}_0\{\max_{[\beta N] \leq n \leq [\alpha N]} \|\mathcal{Z}_N(n)\|^2 > C\} \leq m_0\left(\mathbb{C}/\mathcal{R}\right) \begin{cases} \exp\left(-\dfrac{TN\mathbb{C}\beta}{4\mathcal{R}m_0\left(\mathbb{C}/\mathcal{R}\right)}\right), \\ \text{if } \mathbb{C} > \mathcal{R}gT \\ \exp\left(-\dfrac{N\beta\mathbb{C}^2}{4\mathcal{R}^2gm_0\left(\mathbb{C}/\mathcal{R}\right)}\right), \\ \text{if } \mathbb{C} \leq \mathcal{R}gT, \end{cases} \tag{9}$$

- for type 2 error the following inequality is satisfied:

$$\mathbf{P}_\theta\{\max_{[\beta N] \leq n \leq [\alpha N]} \|\mathcal{Z}_N(n)\|^2 \leq C\} \leq m_0(d) \begin{cases} \exp\left(-\dfrac{TN\beta d}{4m_0(d)}\right), & d > gT \\ \exp\left(-\dfrac{N\beta d^2}{4gm_0(d)}\right), & d \leq gT, \end{cases}$$

where $d = \mathcal{R}^{-1}\left(\|m(\theta)\| - C - \dfrac{L_F}{N}\right) > 0$, $\|m(\theta)\|^2 = tr(B^*A^2(\theta)B)$.

4.1.2 Stochastic predictors

In this subsection we suppose that predictors x_{ji} in (1) are random. On the probability space $(\Omega, \mathcal{F}, \mathbf{P}_\theta)$ consider filtration $\{\mathcal{F}_n\}$, $n = 1, \ldots, n$, where $\{\mathcal{F}_n\} \in \mathcal{F}$, $\mathcal{F}_n$ can be interpreted as all available information up to the instant n.

Put $\mathbf{X}(n) \overset{def}{=} (x_{1n}, \ldots, x_{Kn})^*$.

Suppose that the following conditions are satisfied:

a) there exists a continuous symmetric matrix function $V(t), t \in [0, 1]$ such that the matrix $\int_{t_1}^{t_2} V(s)ds$ is positive definite for any $0 \leq t_1 < t_2 \leq 1$, and $\mathbf{E}_\theta \mathbf{X}(n)\mathbf{X}^*(n) = V(n/N)$;

b) the sequence of random vectors $\{(\mathbf{X}(n), \nu_n)\}$ satisfies the uniform Cramer's and ψ-mixing conditions;

c) the random sequence $\{\nu_n\}$ is a martingale-difference sequence w.r.t. the filtration $\{\mathcal{F}_n\}$;

d) the vector of predictors $\mathbf{X}(n) \overset{def}{=} (x_{1n}, \ldots, x_{Kn})^*$ is $\mathcal{F}_{n-1}$-measurable.

On the segment $[0, 1]$ define the $K \times M$ matrix process

$$u_N(t) \overset{def}{=} \sum_{i=1}^{[Nt]} \mathbf{X}(i)\mathbf{Y}^*(i),$$

13

and the $K \times K$ matrix process

$$\mathcal{T}_N(t) \overset{\text{def}}{=} \sum_{k=1}^{[Nt]} \mathbf{X}(k)\mathbf{X}^*(k).$$

In virtue of conditions a), b), c), the matrix process $N^{-1}\mathcal{T}_N(t)$ weakly converges (in the Skorokhod space) to a positive definite symmetric matrix function $\mathbb{R}(t) \overset{\text{def}}{=} \int_0^t V(s)ds$, and the rate of convergence is exponential. Below we denote $\mathbb{R}(1) \overset{\text{def}}{=} \mathbb{R}$.

Analogously, due to conditions a)-d), the matrix process $N^{-1}\sum_{k=1}^{[Nt]} \mathbf{X}(k)\nu^*(k)$ weakly converges to zero with the exponential rate. Both conclusions follow from the fact that the random processes

$$N^{-1} \sum_{n=1}^{[Nt]} \left(x_{in}x_{jn} - \mathbf{E}_\theta x_{in}x_{jn} \right),$$

$$N^{-1} \sum_{n=1}^{[Nt]} \left(x_{in}\nu_n \right), \quad i,j = 1,\ldots,k$$

weakly converge to zero (as $N \to \infty$) with the exponential rate (see Brodsky, Darkhovsky (2000)).

For estimation of an unknown change-point, the following statistic is used:

$$\mathbb{Z}_N(n) = N^{-1}\left(u_N(n/N) - \mathcal{T}_N(n/N)(\mathcal{T}_N(1))^{-1} u_N(1) \right), \quad n = 1, 2, \ldots, N. \qquad (10)$$

An arbitrary point $\hat{n}$ of the set $\operatorname{Arg} \max_{[\beta N] \leq n \leq [\alpha N]} \|\mathbb{Z}_N(n)\|^2$ is assumed to be the estimate of an unknown change-point. Again we define $\hat{\theta}_N = \hat{n}/N$ as the estimate of the change-point parameter θ.

Statistic (10) generalizes statistic (7) to the situation of stochastic predictors. Assumptions a)-d) guarantee the analogous properties of this statistic. In particular, the limit value (as $N \to \infty$) of the mathematical expectation of the statistic $\mathbb{Z}_N([Nt])$ attains its unique global maximum on the segment $[0,1]$ at the point $t^* = \theta$.

Assumptions a)-d) guarantee convergence in probability of an arbitrary point of $\operatorname{Arg} \max_{[\beta N] \leq n \leq [\alpha N]} \|\mathbb{Z}_N(n)\|^2$ to the point θ with the exponential rate. Hence the $\mathbf{P}_\theta$-a.s. convergence of the proposed estimate to θ follows.

Theorem 4. *Suppose that the conditions a)-d) are satisfied and* $\operatorname{rank}(\mathbb{B}) = M$ *if* $\theta \in [\beta, \alpha]$, *where* $\mathbb{B} \overset{\text{def}}{=} \mathbb{B}(\theta) = \left(E - \mathbb{R}^{-1}\mathbb{R}(\theta) \right)(\mathbf{a} - \mathbf{b})^*$.

Then the estimate $\hat{\theta}_N$ of the change-point parameter θ converges to θ $\mathbf{P}_\theta$-a.s. as $N \to \infty$.

14

Besides, there exists the number $N_1 = N_1(\{\mathbf{X}(n)\})$ such that for $N > N_1$ and any fixed ϵ, $(\min((\alpha - \beta), \|\mathbb{R}\|/2) > \epsilon > 0)$, the following inequality holds:

$$\sup_{\beta \leq \theta \leq \alpha} \mathbf{P}_\theta\{|\hat{\theta}_N - \theta| > \epsilon\} \leq \delta_N(\epsilon) +$$

$$m_0\left(\mathbb{C}(\epsilon, N)/\mathbf{R}\right) \begin{cases} \exp\left(-\dfrac{N\beta\Big(\mathbb{C}(\epsilon, N)/\mathbf{R}\Big)^2}{4gm_0\left(\mathbb{C}(\epsilon, N)/\mathbf{R}\right)}\right), & \text{if } \mathbb{C}(\epsilon, N) \leq \mathbf{R}gT \\[4ex] \exp\left(-\dfrac{TN\beta\Big(\mathbb{C}(\epsilon, N)/\mathbf{R}\Big)}{4m_0\left(\mathbb{C}(\epsilon, N)/\mathbf{R}\right)}\right), & \text{if } \mathbb{C}(\epsilon, N) > \mathbf{R}gT, \end{cases}$$

where $\mathbb{C}(\epsilon, N) = \left[\dfrac{\epsilon\lambda_V}{4\mathbb{M}}\|\mathbf{a} - \mathbf{b}\|^2 - \dfrac{L_V}{N}\right]$, $\mathbb{M} = \max\limits_{\beta \leq t \leq \alpha} \|M(t)\|$, the constants $g, T, m_0(\cdot)$ are taken from the uniform Cramer's and ψ-mixing conditions, and $M(t)$, λ_V, L_V, δ_N, $\mathbf{R}$ are described in the proof.

In particular, for independent observations $m_0(\cdot) = 1$.

Comparing Theorems 1 and 3, we conclude that the order of convergence of the proposed estimate of the change-point parameter to its true value is asymptotically optimal as $N \to \infty$.

From the proof we obtain the following

Corollary 2. Let $S > 0$ be the decision threshold and $\mathbb{S} \overset{def}{=} S - \dfrac{L_V}{N}$. Then:
- for type 1 error the following inequality is satisfied:

$$\mathbf{P}_0\{\max_{[\beta N] \leq n \leq [\alpha N]} \|\mathbb{Z}_N(n)\|^2 > S\} \leq \delta_N(\mathbb{S}) + m_0\left(\mathbb{S}/\mathbf{R}\right) \begin{cases} \exp\left(-\dfrac{TN\mathbb{S}\beta}{4\mathbf{R}m_0\left(\mathbb{S}/\mathbf{R}\right)}\right), \\[1ex] \mathbb{S} > \mathbf{R}gT \\[2ex] \exp\left(-\dfrac{N\beta\mathbb{S}^2}{4\mathbf{R}^2gm_0\left(\mathbb{S}/\mathbf{R}\right)}\right), \\[1ex] \mathbb{S} \leq \mathbf{R}gT. \end{cases}$$

- for type 2 error the following inequality holds:

$$\mathbf{P}_\theta\{\max_{[\beta N] \leq n \leq [\alpha N]} \|\mathbb{Z}_N(n)\|^2 \leq S\} \leq \delta_N(\mathbb{S}) + m_0(r) \begin{cases} \exp\left(-\dfrac{TN\beta r}{4Rm_0(r)}\right), \\[1ex] r > RgT \\[2ex] \exp\left(-\dfrac{N\beta r^2}{4R^2gm_0(d)}\right), \\[1ex] r \leq RgT, \end{cases}$$

where $r = \mathbf{R}^{-1}\left(\|M(\theta)\| - S - L_V\right) > 0$; $\|M(\theta)\|^2 = \text{tr}(\mathbb{B}^*\mathbb{R}^2(\theta)\mathbb{B})$.

4.2 Multiple change-points

The proposed method can be generalized to problems of detection and estimation of multiple change-points in regression models. A widespread approach to solving these problems (see, e.g., Bai, Lumsdaine, Stock (1998)) consists in decomposition of the whole obtained sample to all possible subsamples and construction of regression estimates for each of these subsamples. The decomposition for which the minimum of the general sum of regression residuals is attained, is assumed to be the estimate of a true decomposition of the whole samples of obtained observations into subsamples with different regression regimes.

These methods turn out to be rather time consuming and have a low power. For example, if there are only two regression regimes in an obtained sample but we do not know this fact and are obliged to try all possible subsamples up to the order 20, then many false structural changes will be obtained.

In this paper we propose a new method of detection and estimation of multiple change-points which is not based upon LSE of regression parameters and computation of corresponding residuals. This method is more effective and robust to possible inaccuracies in specification of regression models.

Let us explain the idea of this method by the following example of a multiple regression model (1) with deterministic predictors and the row-matrix $\Pi(\vartheta, n)$. In other words, let $\vartheta = (\theta_1, \theta_2, \ldots, \theta_k)$, $k \geq 1$ is an unknown vector of change-point parameters such that $0 \equiv \theta_0 < \beta \leq \theta_1 < \cdots < \theta_k \leq \alpha < \theta_{k+1} \equiv 1$, where, as before, β, α are known numbers, and the observations has the form

$$y_n = \Pi^*(\vartheta, n)F(n/N) + \nu_n. \tag{11}$$

Here

$$\Pi(\vartheta, n) = \sum_{i=1}^{k+1} a_i \, \mathbb{I}([\theta_{i-1}N] < n \leq [\theta_i N]),$$

where $a_i \neq a_{i+1}, i = 1, 2, \ldots, k$ are unknown *vectors*, $F(t)$ is a given vector-function (all assumptions and notations see in Subsection 4.1.1).

Consider our main statistic (7). The mathematical expectation of this statistic converges as $N \to \infty$ to the function

$$m(t) = \int\limits_0^t F(s)F^*(s)\Pi(\vartheta, s)ds - A(t)I^{-1} \int\limits_0^1 F(s)F^*(s)\Pi(\vartheta, s)ds.$$

16

In the situation when there is no change-points, i.e., the vector of regression coefficients is constant on $[0,1]$, the vector function $m(t)$ equals to zero for each $t \in [0,1]$. This property of $m(t)$ makes it possible to effectively reject the null hypothesis about the absence of change-points when they are really present in an obtained sample.

Consider the following method of detection and estimation of multiple change-points. Fix a small parameter ϵ, $\min(\beta, 1-\alpha) > \epsilon > 0$. The proposed method consists of the following steps:

1. Compute statistic (7) by the data in the diapason of arguments $\mathcal{N} \overset{\text{def}}{=} ([\beta N], \ldots, [\alpha N])$. If $\max\limits_{n \in \mathcal{N}} \|\mathcal{Z}_N(n)\|^2 > C$, where $C = C(N)$ is the decision threshold, then compute $nmax = \text{argmax}\|\mathcal{Z}_N(n)\|^2$, otherwise the sample is assumed to be stationary (without change-points).

2. Put $N' = nmax - [\epsilon N]$ and compute statistic (7) by the data in the diapason of arguments $\mathcal{N}' \overset{\text{def}}{=} ([\beta N], \ldots, N')$ according to step 1. This cycle is repeated until:

1) we obtain a stationary sub-sample in the diapason of data with arguments $([\beta N], \ldots, N')$, i.e. $\max\limits_{n \in \mathcal{N}'} \|Z_{N'}(n)\|^2 \leq C(N')$. Then we put $n(1) = N' + [\epsilon N]$ as the estimate of the first change-point and go to step 3.

or

2) we obtain a sample of the size $N' \leq [2\epsilon N]$. Then we put $n(1) = N' + [\epsilon N]$ as the estimate of the first change-point and go to step 3.

3. Put $n' = n(1) + [\epsilon N]$ and compute statistic (7) by the data in the diapason of arguments $(n', \ldots, [\alpha N])$ (i.e. with the relative arguments $[1, \ldots, [\alpha N] - n' + 1]$) and do according to steps 1 and 2. The cycle is repeated until we obtain a stationary sub-sample in the diapason of data with arguments $[n', \ldots, nmax]$ or $nmax - n' \leq [2\epsilon N]$. Then we put $n(2) = nmax$ as the estimate of the next change-point. If $N - n(2) < [2\epsilon N]$ then stop, otherwise repeat step 3 by the data in the diapason of arguments $(n(2), \ldots, [\alpha N])$.

In this way we continue to compute the estimates $n(3), \ldots$ of change-points. As a result we obtain the series of estimates $n(1), n(2), \ldots$ of the true change-points $[\theta_1 N], \ldots, [\theta_k N]$. The number $\hat{k}_N$ of these estimates is determined by the quantity of stationary sub-samples

$$[1, \ldots, n(1)], \ldots, [n(i), \ldots, n(i+1)], \ldots, [n(\hat{k}_N), \ldots, N]$$

.

The proposed method is based upon reduction to the case of only one change-point

and the properties of the matrix $m(t)$. The crucial point of this method is the choice of the decision threshold $C(N)$ which depends on the sample size N. Below we give an explicit formula for computation of $C(N)$.

Let $\hat{k}_N$ be the estimate of the number of change-points in the sample and $\hat{\vartheta}_N = (\theta_{N1}, \ldots, \theta_{N\hat{k}_N})^*$ be the vector of estimated coordinates of change-point parameters. The following theorem holds for model (11).

Theorem 5. *Suppose assumptions of Theorem 3 are satisfied. Moreover, assume that there exist $h > 0$, $B > 0$ such that for all $i = 2, \ldots, k+1$:*

$$0 < \|A(\theta_{i-1}, \theta_i)A^{-1}(\theta_{i-2}, \theta_{i-1})\| \leq h$$
$$\|A(\theta_{i-1}, \theta_i)(a_i - a_{i-1})\| \geq B > 0,$$

Then for sufficiently small $\delta > 0$:

$$\mathbf{P}\{(\hat{k}_N \neq k) \cup \{(\hat{k}_N = k) \cap (\max_{1 \leq i \leq k} |\hat{\theta}_{Ni} - \theta_i| > \delta)\}\} \leq C(\delta)\exp(-D(\delta)N),$$

where constants $C(\delta) > 0, D(\delta) > 0$ do not depend on N.

Analogous theorem can be proved also for stochastic predictors.

From theorem 5 it follows that the estimated number of change-points converges almost surely to its unknown true value, as well as estimated coordinates of unknown change-points converge exponentially to their true values as the sample size tends to infinity. Moreover, comparing results of theorem 2 and theorem 5 we conclude that the proposed method of detection and estimation of multiple change-points is asymptotically optimal by the order of convergence of estimated change-point parameters to their true values.

4.3 A variant of the limit distribution theorem for the decision statistic under the null hypothesis

For practical applications of the proposed method and, in particular, for the rational choice of the decision threshold $C(N)$, we need to study the limit distribution of the decision statistic under the null hypothesis.

Let us formulate a variant of the limit theorem for the simple case of unique change-point, deterministic predictors, statistically independent noises ν_n, and the one-dimensional dependent variable y_n.

Suppose there exists a continuous function $g(t)$, $0 \leq t \leq 1$ such that $\mathbf{E}_\theta \, \nu_n^2 = g^2(n/N)$.

Put

$$\sigma_i^2 = \frac{1}{t} \int\limits_0^t f_i^2(s) g^2(s) ds, \; i = 1, \ldots, K$$

$$G(t) = (\sigma_1(t), \ldots, \sigma_K(t))^*, \; \mathbf{Z}(t) = G(t) W(t), \; U(t) = \mathbf{Z}(t) - A(t) I^{-1} \mathbf{Z}(1),$$

where $W(t)$ is the standard Wiener process, $A(t)$, I are the above defined matrices (see Subsection 4.1.1).

Consider our main statistic, the vector process $\mathcal{Z}_N(t) = \mathcal{Z}_N([Nt])$ (see (7)). Then for any $\theta \in [\beta, \alpha]$, the vector process $\sqrt{N}(\mathcal{Z}_N(t) - \mathbf{E}_\theta \, \mathcal{Z}_N(t))$ weakly converges to the vector process $U(t)$ in the Skorokhod space $D^K[\beta, \alpha]$ (see Brodsky, Darkhovsky (2000)). In particular, under the null hypothesis, the weak convergence is valid at $[0, 1]$.

Therefore, we have the following

Theorem 6.

$$\lim_{N \to \infty} \mathbf{P}_0 \{ \sqrt{N} \max_{t \in [0,1]} \| \mathcal{Z}_N(t) \| > C \} = \mathbf{P}_0 \{ \max_{t \in [0,1]} \| U(t) \| > C \} \tag{12}$$

(here we use the Euclidean norm for vectors).

The vector $U(t)$ is Gaussian with zero mean and the following $K \times K$ correlation matrix $D(t)$:

$$D(t) = t \left[G(t) G^*(t) - G(t) G^*(1) I^{-1} A(t) - A(t) G(1) G^*(t) \right] + A(t) I^{-1} G(1) G^*(1) I^{-1} A(t).$$

Therefore, we have the following equality by distribution

$$U(t) = \sqrt{D(t)} \zeta \tag{13}$$

where $\zeta = (\zeta_1, \ldots, \zeta_K)^*$ is the standard Gaussian vector.

Taking (13) into account, we get

$$\max_{0 \leq t \leq 1} \| U(t) \| = \max_{0 \leq t \leq 1} \sqrt{\sum_{i=1}^K d_i^2(t) \zeta_i^2} \overset{\text{def}}{=} \rho(\zeta), \tag{14}$$

where $d_i^2(t)$ are eigenvalues of the matrix $D(t)$. The function $\rho(\zeta)$ can be explicitly calculated for any given family of functions $F(t), g(t)$.

19

Therefore, from (14) we have

$$\mathbf{P}_0\{\max_{0 \leq t \leq 1} \|U(t)\| > C\} = \int\limits_{\{u:\rho(u)>C\}} \varphi(u)du, \tag{15}$$

where $\varphi(u)$ is the density of the standard Gaussian distribution.

From (12) and (15) we can conclude that type 1 error goes to zero as $\exp(-const\,NC^2)$ for the proposed method. This fact allows us to choose the decision threshold. Note that the same asymptotical order can be obtained from corollary 2 (see Subsection 4.1.1). For independent noises we have

$$\mathbf{P}_0\{\max_{[\beta N] \leq n \leq N} \|\mathcal{Z}_N(n)\|^2 > C\} \leq \begin{cases} \exp\left(-\dfrac{TN\mathbb{C}\beta}{4R}\right), & \mathbb{C} > gT \\[3mm] \exp\left(-\dfrac{N\beta\mathbb{C}^2}{4R^2gm_0(\mathbb{C})}\right), & \mathbb{C} \leq gT, \end{cases}$$

(the notations see in Subsection 4.1.1).

Therefore, we conclude that type 1 error α_N goes to zero exponentially as $N \to \infty$ for the proposed method.

So, the threshold can be calculated from the relation

$$C = C(N) = \frac{1}{\sqrt{N}} |\ln \alpha_N| \lambda,$$

where λ is a certain calibration parameter which depends on variations of predictors, dispersions of noises and characteristics of their statistical dependence.

A more close study allows us to obtain the following practical formula for the decision threshold $C = C(N)$:

$$C(N) = \frac{\left(\max_i \sigma_i^2 \cdot \max_i \max_{0 \leq t \leq 1} f_i^2(t)\right)^{1/2}}{\sqrt{N}} \lambda,$$

where σ_i^2 is the dispersion of ν_i and $\lambda > 0$ is the calibration parameter.

5   Experiments

In this section we present results of a simulation study of the proposed method in comparison with other well known tests. The following methods are most often used for detection of structural changes in regression models:

- The Chow test most often used in econometric packages;

- The CUSUM (cumulative sums) test based upon recursive regression residuals (Brown, Durbin, Evans, 1975);

- The CUSUM test based upon residuals of ordinary least squares method (OLS CUSUM test, Ploberger, Kramer, 1992);

- Fluctuation test (Ploberger, Kramer, Kontrus, 1989)

- Wald test (Andrews, 1993, Andrews, Ploberger, 1994)

- LM тест (Lagrange Multilpier test, Andrews, 1993).

However, it is well known (see, e.g., Maddala and Kim (1998)) that the Wald test (together with the QMLE - quasi-maximum likelihood estimation test) is the best and most often used for detection of changes in regression models because it has the best characteristics of power and accuracy of change-point estimation.

The Wald test statistic is defined as follows:

$$SupW = \max_{1 \leq m \leq N} N\Big[\frac{S(N) - S_1(m) - S_2(N-m)}{S_1(m) + S_2(N-m)}\Big],$$

where $S(N)$ is the sum of regression residuals constructed by the whole sample of the size N; $S_1(m)$ is the sum of regression residuals constructed by the sub-sample of the first m observations; $S_2(N-m)$ is the sum of residuals of the regression model constructed by the last $N-m$ observations.

It is natural to define the estimate of the change point as $n_0 \in arg \sup W$, and the corresponding estimate of the change-point parameter $\hat{\theta}_N = n_0/N$.

Comparison of characteristics of different methods is carried out in the following way. First, methods are 'equalized' by the value of type 1 error by means of choice of the corresponding decision thresholds. In practice, for this purpose we use experiments with stationary samples (without structural changes) in which the 95-percent quantiles of the variation series of the decision statistics are computed (see below, table 1). Second, for the chosen sample sizes and decision thresholds, experiments with non-stationary samples are performed in which we compute estimates of the type 2 error probability and instants of change-points (see tables 2 and 4). The method of change-point detection 'a' is preferable w.r.t. the method "b" if for the same values of the type 1 error, it gives lower estimates of the type 2 error and the error of change-point estimation.

5.1 Deterministic regression plan

We compared characteristics of our method with those of the Wald test using the following regression model with deterministic predictors:

$$y_i = c_0 + c_1\,x_i + \xi_i, \quad i = 1, \ldots, N \tag{16}$$

where $(x_1, \ldots, x_N)^*$ is the vector of deterministic predictors; $\{\xi_i\}$ is the Gaussian noise sequence with zero mean and unit variance; c_0, c_1 are regresson coefficients which change at the instant $n_0 = [\theta\,N]$, $0 < \theta < 1$.

The number of independent trials of each experiment was equal to k=2000. The estimates of decision thresholds were obtained as follows. For each stationary sample, the 95-percent and 99-percent quantiles of the variation series of maximums of the decision statistic were computed in 2000 trials. These quantiles were then assumed to be estimates of the decision thresholds for 5-percent and 1-percent error level, respectively.

The values of the threshold C given in table 1, were used as decision bounds for the confidence probability 95 percent in experiments with non-stationary regression models. The following cases were considered:

- before the change-point: $c_0 = 0$, $c_1 = 1$
- after the change-point: $c_0 = \delta$, $c_1 = 1$.

In experiments the parameter δ and the sample size N were changed. The following characteristics of the proposed method were estimated:

- The empirical estimate of decision threshold C (more exactly, the empirical estimate of $\max\limits_{n} \|\mathcal{Z}_N(n)\|$);
- The empirical estimate of type 2 error probability $\hat{w}_N$;
- The empirical estimate of the change-point parameter $\hat{\theta}_N$.

Results obtained for the Wald test are given in the following tables.

Table 1. Estimation of the decision thresholds for the Wald test for different sample sizes

N	100	200	300	400	500	700	1000	1200
$p = 0.95$	10.10	8.09	9.59	8.66	8.12	7.62	7.51	7.43
$p = 0.99$	12.60	10.88	14.14	12.10	12.20	9.97	11.68	10.02

Table 2. Estimation of the change-point parameter $\theta = 0.30$ by the Wald test

N		300	400	500	700	1000
$\delta = 0.3$	C	5.63	6.76	8.24	9.77	12.09
	$\hat{w}_N$	0.83	0.71	0.59	0.46	0.32
	$\hat{\theta}_N$	0.29	0.25	0.22	0.19	0.20
$\delta = 0.4$	C	9.65	10.20	11.88	15.27	19.32
	$\hat{w}_N$	0.56	0.47	0.34	0.23	0.18
	$\hat{\theta}_N$	0.28	0.25	0.22	0.20	0.23

The same model was studied with the help of the method proposed in this paper.

1) Decision thresholds

In the first series of experiments, model (16) with constant coefficients $c_0 = 0$, $c_1 = 1$ was used. The following results were obtained.

Table 3. Estimation of the decision thresholds

N	100	200	300	400	500	700	1000	1200
$p = 0.95$	0.401	0.257	0.202	0.182	0.150	0.125	0.103	0.081
$p = 0.99$	0.450	0.300	0.247	0.211	0.187	0.162	0.138	0.102

2) The estimates of the change-point parameter

Table 4. Results of estimation of the change-point parameter $\theta = 0.30$

N		300	400	500	700	1000
$\delta = 0.3$	C	0.179	0.177	0.168	0.157	0.151
	$\hat{w}_N$	0.64	0.55	0.33	0.13	0.03
	$\hat{\theta}_N$	0.340	0.322	0.332	0.324	0.307
$\delta = 0.4$	C	0.220	0.211	0.208	0.195	0.192
	$\hat{w}_N$	0.28	0.24	0.11	0.02	0.005
	$\hat{\theta}_N$	0.315	0.312	0.308	0.305	0.304

Table 5. Results of estimation of the change-point parameter $\theta = 0.50$

N		300	400	500	700	1000
$\delta = 0.3$	C	0.194	0.184	0.175	0.168	0.164
	$\hat{w}_N$	0.62	0.50	0.25	0.05	0.01
	$\hat{\theta}_N$	0.456	0.485	0.501	0.502	0.499
$\delta = 0.4$	C	0.231	0.221	0.215	0.214	0.211
	$\hat{w}_N$	0.26	0.22	0.003	0.02	0
	$\hat{\theta}_N$	0.495	0.495	0.489	0.501	0.499

Comparing results from tables 2 and 4, we conclude that type 2 error estimates for our method are lower than for the Wald test, and the error of estimation for our method is much lower than for the Wald test. Therefore, we conclude that our method is essentially better by the main performance characteristics of change-point detection than the Wald test, and so, we conclude that the proposed method is one of the most effective among all known tests for detection and estimation of structural changes in regression models.

Comparing results from table 4 and 5, we can conclude that the quality of estimation of the change-point parameter θ depends on its location on the segment $[0, 1]$: estimation of θ which is closer to the bounds of the segment $[0, 1]$ is more difficult.

In next two subsections we investigate our methods.

5.2 Stochastic regression plan

In this series of experiments the following model of observations was used:

$$y_i = c_0 + c_1 x_i + \xi_i, \quad i = 1, \ldots, N$$

where $(x_1, \ldots, x_N)^*$ is a stationary random sequence of the following type:

$$x_i = \rho x_{i-1} + \eta_i, \quad i = 1, \ldots, N, \ x_0 \equiv 0,$$

$\{\xi_i, \eta_i\}$ is the sequence of independent Gaussian r.v.'s with zero mean and unit dispersion; c_0, c_1 are regression coefficients which change at the instant $n_0 = [\theta N]$, $0 < \theta < 1$; $|\rho| < 1$.

1) Estimation of decision thresholds

In the first series of tests decision thresholds were estimated. For this purpose, stationary sequences (without change-points) were used: $c_0 = 0, c_1 = 1, \rho = 0.3$. The following results were obtained.

Table 6. Estimation of decision thresholds (the case of stochastic predictors)

N	100	200	300	400	500	700	1000	1200
$p = 0.95$	0.355	0.291	0.230	0.188	0.150	0.132	0.103	0.082
$p = 0.99$	0.401	0.332	0.273	0.218	0.192	0.171	0.141	0.100

2) Estimation of the change-point parameter

24

In the following series of experiments a model with a structural change in the regression coefficients was used:

   - before the change-point: $c_0 = 0$, $c_1 = 1$
   - after the change-point: $c_0 = 0$, $c_1 = 1.3$.

Results obtained are presented in table 7.

Table 7. Estimation of change-point parameters (the case of stochastic predictors)

N		500	700	1000	1200
$\theta = 0.5$	C	0.167	0.157	0.152	0.152
	$\hat{w}_N$	0.32	0.21	0.02	0
	$\hat{\theta}_N$	0.481	0.495	0.498	0.499
$\theta = 0.3$	C	0.156	0.148	0.142	0.140
	$\hat{w}_N$	0.45	0.30	0.03	0
	$\hat{\theta}_N$	0.312	0.310	0.308	0.301

5.3   Multiple structural changes in multivariate systems

The following multivariate system was used:

$$y_i = c_0 + c_1 y_{i-1} + c_2 z_{i-1} + c_3 x_i + \epsilon_i$$
$$z_i = d_0 + d_1 y_i + d_2 x_i + \xi_i$$
$$x_i = 0.5 x_{i-1} + \nu_i$$
$$\epsilon_i = 0.3 \epsilon_{i-1} + \eta_i,$$

where $\xi_i, \nu_i, \eta_i,\ i = 1, 2, \ldots$ are independent standard Gaussian random variables.

Here $(y_i, z_i)^*$ is the vector of endogenous variables, x_i is the vector of exogenous variables, $(y_{i-1}, z_{i-1}, x_i)^*$ - the vector of pre-determined variables of the considered system.

Dynamics of this system is characterized by the following vector of coefficients: $\mathbf{u} = [c_0\ c_1\ c_2\ c_3\ d_0\ d_1\ d_2]$. The initial vector of coefficients is $[0.1\ 0.5\ 0.3\ 0.7\ 0.2\ 0.4\ 0.6]$. The first structural change occurs at the instant $\theta_1 = 0.3$. The vector of coefficients $\mathbf{u}$ changes into $[0.1\ 0.5\ 0\ 0.7\ 0.2\ 0.4\ 0.6]$. The second structural change occurs at the instant $\theta_2 = 0.7$. Then the vector $\mathbf{u}$ changes into $[0.1\ 0.5\ 0\ 0.7\ 0.2\ 0.4\ 0.9]$.

In the first series of tests the decision threshold C was estimated. For this purpose, the model with the initial vector of coefficients $\mathbf{u}$ and without change-points was used.

In 2000 independent trials the maximums of the decision statistic were computed and the variation series of these maximum was constructed. Then the 95-percent and the 99-percent quantiles of this series were computed. These values are presented in table 8.

Table 8. Estimation of decision thresholds (the case of a multivariate system)

N	200	400	500	700	900	1000	1200	1500
$p = 0.95$	0.28	0.20	0.19	0.18	0.16	0.15	0.145	0.14
$p = 0.99$	0.36	0.33	0.28	0.24	0.23	0.21	0.19	0.17

The computed 95-percent quantiles were assumed to be the decision thresholds for the corresponding sample volumes.

In the next series of tests non-stationary samples with multiple change-points were used. The true number of change-points was equal to $p = 2$, the coordinates of these change-points were $\theta_1 = 0.3$ and $\theta_2 = 0.7$. In table 9 the following performance characteristics are given:

- w is the estimate of the probability $\mathbf{P}_\theta\{\hat{p}_N \neq p\}$ in 2000 independent trials, where $\hat{p}_{scriptscriptstyle N}$ is the estimate of the number of change-points in the data.

- Δ is the estimation error on condition that $\hat{p}_N = p$, i.e. $\Delta = \sqrt{\sum_{i=1}^{p}(\hat{\theta}_i - \theta_i)^2}$.

Table 9. Estimation of change-point parameters (the case of a multivariate system)

N	200	400	500	700	900	1000	1200	1500
w	0.96	0.54	0.39	0.21	0.04	0.03	0.02	0.01
Δ	0.02	0.05	0.04	0.02	0.03	0.02	0.01	0.005

6 Conclusions

In this paper the following main results were obtained:

1. The general statement of the retrospective change-point detection and estimation problem in multivariate linear systems is given (both one change-point and multiple change-point problems, both independent and dependent sequences of observations)

2. The prior lower bounds are proved for the main performance characteristic in retrospective change-point detection and estimation: *the probability of the error of change-point estimation*, in different contexts of change-point estimation: from one

change-point in multi-factor linear regressions with deterministic and stochastic regression plans, to multiple change-point problems in multivariate linear models.

3. A new method is proposed for the problem of retrospective change-point detection and estimation in multivariate linear systems. The main performance characteristics of this method: type 1 and type 2 errors, the error of change-point estimation, are studied theoretically. We prove that the proposed method is *asymptotically optimal* by the order of convergence of the change-point estimate to its true value as the sample size tends to infinity.

4. For the problem of multiple change-point detection and estimation, we propose a general setup in which both *the number of change-points and their coordinates in the sample are unknown.* For this problem statement, a new method is proposed which gives consistent estimates of an unknown number of change-points and their coordinates. This method is also asymptotically optimal by the order of convergence of these estimates to true change-point parameters.

5. A simulation study of characteristics of the proposed method for finite sample sizes is performed. The main goals of this study are as follows: to compare performance characteristics of the proposed method with characteristics of other well known methods of change-point detection in linear regression models: the Wald test, the Chow test, the CUSUM tests with ordinary and recursive regression residuals, the fluctuation test; to consider more general multivariate linear models and performance characteristics of the proposed method in these multivariate models. The main conclusion: performance characteristics of the proposed method are no worse but often even better than those of well known change-point tests.

References

[1] Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica, 61, 821-856.

[2] Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisanse parameter is present only under the alternative. Econometrica, 62, 1383-1414.

[3] Bai, J., Lumsdaine R., Stock J, 1998. Testing for and Dating Common Breaks in Multivariate Time Series. Review of Economic Studies, 65, 395-432.

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

[5] Berkes I., Gombay E., Horvath L., 2009. Testing for changes in the covariance structure of linear processes, Journal of Statistical Planning and Inference, vol.139, is.6.1, pp.2044-2063.

[6] Bradley, R., 2005. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. Probability Surveys, 2, 107-144.

[7] Brodsky, B., Darkhovsky, B., 1993. Non-parametric Methods in Change-Point Problems. Dordrecht: Kluwer Academic Publishers.

[8] Brodsky, B., Darkhovsky, B., 2000. Non-Parametric Statistical Diagnosis: Problems and Methods. Dordreht: Kluwer Academic Publishers.

[9] Brown, R.L., Durbin, J., Evans, J.M., (1975). Techniques for testing the constancy of regression relationships over time. Journal of Royal Statistical Society, Series B, 37, 149-192.

[10] Csörgő, M., Horvath, L.,(1997). Limit theorems in change-point analysis. Chichester: Wiley.

[11] Chow, G.C., (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28, 591-605.

[12] Christiano, L., (1992). Searching for a break in GDP. Journal of Business and Economic Statistics, 10, 237-250.

[13] Darkhovsky, B., (1995). Retrospective change-point detection in some regression models. Theory of Probability and Applications, 40, 4, 898-903.

[14] Davis R.A., Lee C.M.T., Rodriguez-Yam G.A., 2006. Structural break estimation for nonstationary time series models. Journal of American Statistical Association, vol.101, pp.223-239.

[15] Gombay, E., Horvath, L., (1994). Limit theorems for changes in linear regression. Journal of Multivariate Analysis, 48, 43-69.

[16] Gombay E., 2008. Change detection in autoregressive time series. Journal of Multivariate Analysis, vol.99, pp.451-464.

[17] Horváth, L., Huskova, M., Serbinowska, M., (1997). Estimators for the time of change in linear models. Statistics, 29, 109-130.

[18] Huskova, M., (1996). Estimation of a change in linear models Statist. Probab. Letters, 26, 13-24.

[19] Huskova M., Praskova Z., Steinebach J., 2007. On the detection of changes in autoregressive time series. I. Asymptotics. Journal of Statistical Planning and Inference, vol. 137, is.2, pp.1243-1259.

[20] Huskova M., Kirch C., Praskova Z., Steinebach J., 2007. On the detection of changes in autoregressive time series. II. Resampling procedures. Journal of Statistical Planning and Inference, vol. 138, is.6,1, pp.1697-1721.

[21] Ibragimov, I. A., Linnik, Yu., V., (1971). Independent and stationary sequences of random variables. Wolters-Noordhoff Publishing, Groningen.

[22] James, B., James, K.J., Siegmund, D., (1987). Tests for a change-point. Biometrika, 74. 71-83.

[23] Kim, H.-Ju., Siegmund, D., (1989). The likelihood ratio test for a change-point in simple linear regression. Biometrika, 76, 3, 409-423.

[24] Kim, H. J., (1994). Tests for a change-point in linear regression. IMS Lecture Notes, Monograph Series, 23, 170-176.

[25] Korostelev, A., (1997). Minimax large deviations risk in change-point problems. Mathematical Methods in Statistics, v. 6, no. 3, 365-374.

[26] Krämer, W., Ploberger, W., Alt, R., (1988). Testing for structural change in dynamic models. Econometrica, 56, 6, 1355-1369.

[27] Maddala, G., Kim, I., (1998). Unit roots, cointegration, and structural change. Cambridge: Cambridge Univ. Press.

[28] Maronna, R., Yohai, V., (1978). A bivariate test for the detection of a systematic change in mean. Journal of American Statistical Association, 73, 640-645.

[29] Nelson, C., Plosser, C., (1982). Trends and random walks in macroeconomic time series: some evidence and implications. Journal of Monetary Economics, 10, 130-162.

[30] Perron, P., (1989). The Great crash, the oil price shock, and the Unit root hypothesis. Econometrica, 57, 1361-1401.

[31] Perron, P., Vogelsang, T., (1992). Nonstationarity and level shifts with an application to purchasing power parity. Journal of Business and Economic Statistics, 10, 3, 301-320.

[32] Ploberger, W., Kramer, W., (1992). The CUSUM test with OLS residuals. Econometrica, 60, 271-285.

[33] Ploberger, W., Kramer, W., Kontrus, K., (1989). A new test for structural stability in the linear regression model. Journal of Econometrics, 40, 307-318.

[34] Quandt, R.E., (1958). The estimation of parameters of a linear regression system obeying two separate regimes. Journal American Statistical Association, 50, 873-880.

[35] Quandt, R.E., (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal American Statistical Association, 55, 324-330.

[36] Worsley, K. J., (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika, 73, 91-104.

[37] Zivot, E., Andrews, D., (1992). Further evidence on the Great crash, the oil price shock and the Unit root hypothesis. Journal of Business and Economic Statistics, 10, 251-287.