

# DETECTION AND ESTIMATION OF STRUCTURAL CHANGES IN DYNAMICAL REGRESSION MODELS

Brodsky B.E. <sup>\*,1</sup> Darkhovsky B.S. <sup>\*\*\*</sup>

*\* Central Insitute for Mathematics and Economics RAS  
Moscow, Russia*

*\*\*\* Institute for Systems Analysis RAS  
Moscow, Russia  
e-mail : bbrodsky@yandex.ru*

Abstract: The problem of retrospective detection of structural changes (change-points) in regression and cointegration models is considered. The asymptotic low bound for the quality functional of change-point detection is proved. The asymptotically optimal methods for detection of structural changes in regression and cointegration models are proposed. *Copyright*© 2005 IFAC

Keywords: structural changes, change-point problems, regression, cointegration models.

## 1. INTRODUCTION

The past four decades have seen considerable theoretical and empirical research on the detection of structural changes in regression relationships, dynamical systems, and econometric models. The vast amount of statistical literature on this topic nowadays represents the specific chapter in the body of *Change-Point Analysis* – the rapidly developing branch of the mathematical statistics.

For detecting of change-points in regression models the cumulative sums of regression residuals were used (see Brown et al., 1975). Regression models with multiply structural breaks, as well as point and interval estimates of endogenous breaks in multivariate time series were considered in (Bay and Perron, 1998; Bay et al., 1998). Many results in detection and estimation of structural changes in regression models including the limit distributions of the maximum likelihood statistics, the sums of regression residuals, and the sums

of recursive regression residuals are studied are systematized in (Csörgő and Horvath, 1997).

In spite of the vast econometric literature on detection and estimation of structural breaks in dynamic systems and regression models, most of proposed methods suffer from certain substantial drawbacks. First, the majority of methods utilize the linear regression estimates constructed by all decompositions of an obtained sample into some quasi-stationary subsamples. It leads to a large number of false structural breaks, because the LSE's of regression coefficients obtained by small subsamples are, as a rule, rather bad, and the asymptotic consistency demonstrated by many authors is achieved only for tremendous data volumes. Second, there is no research on the asymptotic optimality of the proposed methods of structural breaks detection and estimation.

In (Brodsky and Darkhovsky, 2000) nonparametric approach to retrospective change-point estimation in some regression models were considered.

---

<sup>1</sup> Corresponding author

In this paper some generalizations of the results from (Brodsky and Darkhovsky, 2000) are given, namely:

a) the a priori theoretical low bound for the error probability in change-point estimation for regression and cointegration models. This bound provides the theoretical basis for the proofs of the asymptotic optimality of change-point estimates and for the comparative analysis of these estimates;

2) the asymptotically optimal method of change-point estimation for regression models with deterministic predictors without the assumption of orthogonality of the design matrix;

3) the asymptotically optimal method of change-point estimation for regression and cointegration models with stochastic nonstationary predictors.

## 2. A PRIORI INEQUALITIES

On the probability space  $(\Omega, \mathcal{F}, \mathbf{P}_\theta)$  let us consider a sequence of i.r.v.'s  $x_1, x_2, \dots, x_N$  with the following density function (d.f)

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

where  $1 > \theta > 0$  is an unknown change-point parameter ( $[a]$  here and below is integer part of number  $a$ ) and the density function  $f_0(x, t) \neq f_1(x, t)$  in some neighborhood  $t \in T(\theta)$  of the change-point parameter.

Consider the class  $\mathcal{M}_N$  of all estimates  $\hat{\theta}_N$  of the change-point parameter  $\theta$  constructed by the sample  $X^N = \{x_1, \dots, x_N\}$ .

*Theorem 2.1.* For any  $0 < \epsilon < 1$ ,

$$\begin{aligned} & \liminf_N N^{-1} \\ & \ln \inf_{\hat{\theta}_N \in \mathcal{M}_N} \sup_{0 < \theta < 1} \mathbf{P}_\theta \{ |\hat{\theta}_N - \theta| > \epsilon \} \\ & \geq - \min \left( \int_\theta^{\theta+\epsilon} J_0(t) dt, \int_{\theta-\epsilon}^\theta J_1(t) dt \right) \end{aligned}$$

where

$$J_0(t) = \mathbf{E}_0 \ln \frac{f_0(x, t)}{f_1(x, t)}, \quad J_1(t) = \mathbf{E}_1 \ln \frac{f_1(x, t)}{f_0(x, t)}$$

Some particular cases of model (1) are given below.

### 2.1 A broken trend in the mathematical expectation $\phi(t)$ of Gaussian observations

In this case

$$\begin{aligned} f_0(x, t) &= h(x) \exp(\phi_0(t)x - \frac{1}{2}\phi_0^2(t)), \quad t \leq \theta \\ f_1(x, t) &= h(x) \exp(\phi_1(t)x - \frac{1}{2}\phi_1^2(t)), \quad t > \theta, \end{aligned}$$

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

It follows from Theorem 1 the following low bound for the error probability of any change-point estimate  $\hat{\theta}_N$ :

$$\begin{aligned} & \mathbf{P}_\theta \{ |\hat{\theta}_N - \theta| > \epsilon \} \geq (1 - o(1)) \\ & \exp \left( -N/2 \min \left( \int_\theta^{\theta+\epsilon} J(t) dt, \int_{\theta-\epsilon}^\theta J(t) dt \right) \right) \end{aligned}$$

where  $J(t) = (\phi_0(t) - \phi_1(t))^2$ .

### 2.2 A linear deterministic regression with Gaussian errors

Consider the following regression model

$$y_n = c_1(n)x_{1n} + \dots + c_k(n)x_{kn} + \xi_n, \quad (2)$$

where  $\{\xi_n\}$  is the sequence of independent Gaussian random variables,  $\xi_n \sim \mathcal{N}(0, \sigma^2)$ ,  $\mathbf{c}(n) = (c_1(n), \dots, c_k(n))^* = \mathbf{a}I(n \leq [\theta N]) + \mathbf{b}I(n > [\theta N])$ ,  $\mathbf{a} \neq \mathbf{b}$ ,  $I(A)$  is an indicator function of set  $A$ ,  $*$  is the transposition symbol.

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

In this case we obtain from Theorem 1 for all estimates  $\hat{\theta}_N$  of the change-point parameter:

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

where  $J_0(t) = (\sum_{i=1}^k f_i(t)(a_i - b_i))^2$ ,  $J_1(t) = (\sum_{i=1}^k f_i(t)(a_i - b_i))^2$ .

### 2.3 A linear stochastic regression model with independent Gaussian predictors

Here we consider model (2) where  $\xi_n \equiv 0$ , vectors  $\mathbf{x}_n = (x_{1n}, \dots, x_{kn})^*$  are random and independent and there exist continuous functions  $f_i(\cdot), \sigma_i(\cdot)$ ,  $i = 1, \dots, k$ , such that  $x_{in}$  are Gaussian random variables,  $x_{in} \sim \mathcal{N}(f_i(n/N), \sigma_i^2(n/N))$ ,  $n = 1, \dots, N$ , the components  $x_{in}, x_{jn}$  are independent if  $i \neq j$  and  $\mathbf{c}(n) = (c_1(n), \dots, c_k(n))^* = \mathbf{a}I(n \leq [\theta N]) + \mathbf{b}I(n > [\theta N])$ ,  $\mathbf{a} \neq \mathbf{b}$ .

In this case we obtain from Theorem 1 for all estimates  $\hat{\theta}_N$  of the change-point parameter:

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

where

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

and  $\phi_0(t) = a_1 f_1(t) + \dots + a_k f_k(t)$ ,  $\Delta_0^2(t) = a_1^2 \sigma_1^2(t) + \dots + a_k^2 \sigma_k^2(t)$ , with analogous expressions for  $\phi_1(t)$ ,  $\Delta_1(t)$ ,  $J_1(t)$ .

### 3. ASYMPTOTICALLY OPTIMAL METHODS

From Theorem 1 it follows that the optimal speed of convergence of change-point estimates to the true values of change-point parameters is exponential w.r.t.  $N$ . So the problem arises to find methods for which this optimal exponential order of convergence is attained asymptotically as  $N \rightarrow \infty$ . Below we propose such asymptotically optimal methods of change-point estimation for regression models with deterministic and stochastic predictors. These methods can be considered as generalisations of statistics proposed earlier by Brodsky and Darkhovsky (2000) for regression models. Note that Deshayes and Picard (1985) considered the problem of detection of change-points in the spectrum function of observations and, in particular, in autoregressive models.

#### 3.1 Non-random designs

Consider the regression model (2) where now  $\{\xi_i\}$  is a sequence of centered and, in general, dependent random variables. Remind that  $x_{in} = f_i(n/N)$ ,  $n = 1, \dots, N$  and put  $F(t) = (f_1(t), \dots, f_k(t))^*$ .

Let us formulate the assumptions for a collection  $\{f_i\}$  in this problem:

a) the functions  $\{f_i\}$  are almost everywhere (w.r.t. Lebesgue measure) continuous and bounded on  $[0, 1]$ ;

b) for any  $0 \leq t_1 < t_2 \leq 1$  the matrix

$$\int_{t_1}^{t_2} F(s) F^*(s) ds$$

is positively definite.

Put

$$\int_0^1 F(s) F^*(s) ds = R$$

Due to the assumptions, the matrix  $R$  is symmetric and positively definite. Therefore there exists the matrix  $R^{-1/2}$ .

Define for any  $t \in [0, 1]$

$$\begin{aligned} A_t &= \int_t^1 F(s) F^*(s) ds, \quad B_t = R - A_t \\ P_t &= R^{-1/2} A_t R^{-1/2}, \\ Q_t &= R^{-1/2} B_t R^{-1/2} \end{aligned}$$

Due to the assumptions, the matrices  $P_t$ ,  $Q_t$  are positively definite for any  $0 < t < 1$  and

$$P_t + Q_t \equiv \mathbb{E}$$

where  $\mathbb{E}$  is the identity matrix in  $\mathbb{R}^k$ .

Put

$$\tilde{\mathbf{a}} = R^{1/2} \mathbf{a}, \quad \tilde{\mathbf{b}} = R^{1/2} \mathbf{b},$$

Then the regression model can be rewritten as follows under consideration

$$y_i = F^*(i/N) R^{-1/2} \tilde{\mathbf{c}}(i) + \xi_i$$

where  $\tilde{\mathbf{c}}(i) = \tilde{\mathbf{a}} I(i \leq [\theta N]) + \tilde{\mathbf{b}} I(i > [\theta N])$ .

Denote

$$R_N = N^{-1} \sum_{i=1}^N F(i/N) F^*(i/N).$$

Due to the assumptions  $R_N \rightarrow R$  as  $N \rightarrow \infty$ .

Define the vector

$$z(n_1, n_2) = R_N^{-1/2} \sum_{i=n_1}^{n_2} F(i/N) y_i,$$

and the matrix ( $1 \leq n_1 < n_2 \leq N$ ):

$$\mathcal{P}_{n_1}^{n_2} = R_N^{-1/2} \left\{ \sum_{k=n_1}^{n_2} F(k/N) F^*(k/N) \right\} R_N^{-1/2},$$

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

$$\begin{aligned} Y_N(n) &= N^{-1} \left( (\mathcal{P}_{n+1}^N)^{1/2} (\mathcal{P}_1^n)^{-1/2} z(1, n) \right. \\ & \quad \left. - (\mathcal{P}_1^n)^{1/2} (\mathcal{P}_{n+1}^N)^{-1/2} z(n+1, N) \right). \end{aligned}$$

An arbitrary point  $\hat{n}$  of the set  $\arg \max_{1 \leq n \leq N} \|Y_N(n)\|$  ( $\|\cdot\|$  is an Euclidean norm) is assumed to be the change-point estimate. Define also the value  $\hat{\theta}_N = \hat{n}/N$  as the estimate of the change-point parameter  $\theta$ .

Put  $\Theta = [\beta, \gamma]$ ,  $0 < \beta < \gamma < 1$ ,  $\beta$  and  $\gamma$  are known numbers.

*Theorem 3.1.* Suppose the sequence  $\xi_i$ ,  $i = 1, \dots, N$  in (2) satisfies Cramer's and  $\psi$ -mixing conditions and assumptions a), b) hold. Then the estimate

$\hat{\theta}_N$  converges to  $\theta$   $\mathbf{P}_\theta$ -a.s. and for any  $\delta > 0$  there exist constants  $N(\delta), A(\delta) > 0, B(\delta) > 0$  such that for  $N > N(\delta)$  the following inequality holds:

$$\begin{aligned} & \sup_{\theta \in \Theta} \mathbf{P}_\theta \{ |\hat{\theta}_N - \theta| > \delta \} \\ & \leq A(\delta) \exp(-B(\delta)N). \end{aligned}$$

### 3.2 Random designs

In this paragraph we assume that predictors  $x_{ji}$  in model (2) are random. On the probability space  $(\Omega, \mathcal{F}, \mathbf{P}_\theta)$  consider an increasing sequence of  $\sigma$ -subfields  $\{\mathcal{F}_n\}$ ,  $n = 1, \dots, n$  of  $\mathcal{F}$ , where  $\mathcal{F}_n$  represents all available information up to the instant  $n$ .

Important examples of models with random predictors include the autoregressive model and the autoregressive model with exogenous inputs.

Cointegration models are also a particular case of (2) with random predictors: if the d.f.'s of predictors  $x_{ji}, j = 1, \dots, k$ , change at any time  $i$ , e.g., in the case of unit root stochastic trends or even "explosive" trends, then the problem is to detect and estimate instants of structural changes in coefficients of model (2).

Put  $\mathbf{x}(n) = (x_1(n), \dots, x_k(n))^*$ .

In this section we assume that the following assumptions hold:

i) there exists continuous matrix function  $V(t), t \in [0, 1]$  such that for any  $0 \leq t_1 < t_2 \leq 1$

$$\begin{aligned} & \lim_{N \rightarrow \infty} \mathbf{E}_\theta N^{-1} \sum_{j=[t_1 N]}^{[t_2 N]} \mathbf{x}(j) \mathbf{x}^*(j) \\ & = \int_{t_1}^{t_2} V(t) dt \end{aligned}$$

where  $\int_{t_1}^{t_2} V(t) dt$  is a positively definite matrix;

ii) the random vector sequence  $\{(\mathbf{x}(n), \xi_n)\}$  satisfies  $\psi$ -mixing and Cramer's conditions;

iii) the random sequence  $\{\xi_n\}$  is a martingale-difference sequence w.r.t. the flow  $\{\mathcal{F}_n\}$ .

Put

$$N^{-1} \sum_{k=1}^N \mathbf{x}(k) \mathbf{x}^*(k) = \mathcal{V}_N$$

and

$$\tilde{\mathbf{a}} = \mathcal{V}_N^{1/2} \mathbf{a}, \quad \tilde{\mathbf{b}} = \mathcal{V}_N^{1/2} \mathbf{b},$$

Then the model under consideration model can be rewritten as follows

$$y_i = \mathbf{x}^*(i) \mathcal{V}_N^{-1/2} \tilde{\mathbf{c}}(i) + \xi_i$$

where  $\tilde{\mathbf{c}}(i) = \tilde{\mathbf{a}} I(i \leq [\theta N]) + \tilde{\mathbf{b}} I(i > [\theta N])$ .

Define the vector

$$u(n_1, n_2) = \mathcal{V}_N^{-1/2} \sum_{i=n_1}^{n_2} \mathbf{x}(i) y_i,$$

and the matrix

$$\begin{aligned} \mathcal{T}_{n_1}^{n_2} &= \mathcal{V}_N^{-1/2} \left\{ \sum_{k=n_1}^{n_2} \mathbf{x}(k) \mathbf{x}^*(k) \right\} \mathcal{V}_N^{-1/2} \\ &1 \leq n_1 < n_2 \leq N \end{aligned}$$

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

$$\begin{aligned} & \mathbb{Y}_N(n) \\ &= N^{-1} \left( (\mathcal{T}_{n+1}^N)^{1/2} (\mathcal{T}_1^n)^{-1/2} u(1, n) \right. \\ & \left. - (\mathcal{T}_1^n)^{1/2} (\mathcal{T}_{n+1}^N)^{-1/2} u(n+1, N) \right) \end{aligned}$$

An arbitrary point  $\hat{n}$  of the set  $\arg \max_{1 \leq n \leq N} \|\mathbb{Y}_N(n)\|$  is assumed to be the change-point estimate. Define also the value  $\hat{\theta}_N = \hat{n}/N$  as the estimate of the change-point parameter  $\theta$ .

*Theorem 3.2.* Suppose conditions i)–iii) are satisfied. Then the estimate  $\hat{\theta}_N$  converges to  $\theta$ ,  $\mathbf{P}_\theta$ -a.s. and for any  $\delta > 0$  there exist constants  $\tilde{N}(\delta), \tilde{A}(\delta) > 0, \tilde{B}(\delta) > 0$  such that for  $N > \tilde{N}(\delta)$  the following inequality holds:

$$\begin{aligned} & \sup_{\theta \in \Theta} \mathbf{P}_\theta \{ |\hat{\theta}_N - \theta| > \delta \} \\ & \leq \tilde{A}(\delta) \exp(-\tilde{B}(\delta)N) \end{aligned}$$

## 4. EXPERIMENTAL RESULTS

For testing the efficiency of the proposed method for finite samples, some Monte-Carlo tests were performed.

The following model of observations was considered:

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

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

We note that the vector  $(x_1, \dots, x_N)^*$  was taken as one of dynamic series from the Russian economic statistics (daily interbank credit rate from 01.01.95 to 31.12.2003, see Internet site of the Central Bank of Russia).

The number of independent Monte-Carlo trials for each test was equal to 2000. The threshold estimates were obtained as follows. For each sample without change-points, the maximum of the module of the decision statistic was computed, then in 2000 trials the variation series from these

values was constructed and the 95-percent and 99-percent quantile of this series were computed. These quantiles were assumed to be the decision thresholds with error levels 5-percent and 1-percent respectively.

The values of the threshold  $C$  for the confidence probability of 95 percent were used below as decision boundaries in Monte-Carlo tests for samples with changes in regression coefficients. The following coefficients were considered:

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

The parameter  $\delta$  and the sample volume  $N$  were varied in tests with the following quality characteristics:

- empirical average of the maxima of the statistic  $\|Y_N(\cdot)\| : C$ ;
- empirical estimate of the 2nd type error probability:  $\hat{w}_N$ ;
- empirical estimate of the change-point parameter  $\hat{\theta}_N$ .

Table 1. Estimation of the change-point parameter  $\theta = 0.30, \delta = 0.3; 0.4$

$N$		300	400	500	700	1000
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
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 2. Estimation of the change-point parameter  $\theta = 0.50; \delta = 0.3; 0.4$

$N$		300	400	500	700	1000
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
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 in Tables 2 and 3 we conclude that the qualitative characteristics of change-point detection depend of the change-point parameter  $\theta$ : detection of change-points at the both ends of a sample is a more complicated task.

## 5. CONCLUSION

A new method of estimation of structural changes in deterministic and stochastic regression models is proposed. Unlike many earlier methods for

detection of structural breaks in regression and cointegration models, this method is not based upon LSE of regression parameters and therefore is more robust to different errors in specification of a model.

The a priori information-theoretical low bounds for the error probability of change-point estimation are proved that enable us to conclude that the asymptotically optimal speed of convergence of a change-point estimate to its true value is exponential with respect to the sample volume  $N$ . The same a priori low bounds can be used for determination of the asymptotically optimal confidence intervals in change-point estimation.

The method of change-point estimation proposed in this paper is valid both for deterministic and stochastic regressions and is asymptotically optimal in the above sense: the speed of convergence of the change-point estimates constructed according to this method is exponential w.r.t. the sample volume  $N$ . This property is of substantial practical importance, since in many situations in econometrics and biomedecine it is hard to discern between deterministic and stochastic predictors.

The above considered method of estimation of unique change-point in regression and cointegration models can be generalized to the case of the finite number of change-points. This generalization is based upon some ideas from (Brodsky and Darkhovsky, 2000).

## References

- Bai, J., R. Lumsdaine and J. Stock (1998). Testing for and Dating Common Breaks in Multivariate Time Series. *Review of Economic Studies*, **65**, 395–432.
- Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 1, 47–78.
- Brodsky, B.E. and B.S. Darkhovsky (2000). *Non-Parametric Statistical Diagnosis. Problems and Methods*. Kluwer Academic Publishers, Dordrecht.
- Brown, R.L., J. Durbin and J.M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of Royal Statistical Society, Seria B*, **37**, 149–192.
- Csörgő, M. and L. Horvath (1997). *Limit theorems in change-point analysis*. Wiley, Chichester.
- Deshayes, J. and D. Picard (1985). Off-line statistical analysis of change-point models using non-parametric and likelihood methods. In: *Detection of Abrupt Changes in Signals and Dynamical Systems. Lecture Notes in Control and Information Sciences* ( M.Basseville and A.Benveniste, Ed), Vol. 77, pp. 70–130. Springer-Verlag, Berlin, Heidelberg, New York.