Sequential Detection of Change-Points in Linear Models

Boris Brodsky
(CEMI RAS, Moscow, Russia)

In this report the problem of sequential detection of change-points in linear models is considered. Suppose the multivariate system with structural changes is described by the following model:

$$Y(n) = \Pi X(n) + \nu_n, \quad n = 1, 2, \dots, \tag{1}$$

where $Y(n) = (y_{1n}, \dots, y_{Mn})'$ is the vector of endogenous variables; $X(n) = (x_{1n}, \dots, x_{Kn})'$ is the vector of pre-determined variables; $\nu_n = (\nu_{1n}, \dots, \nu_{Mn})'$ is the vector of errors. $'$ is the transposition symbol.

Remind that the class of pre-determined variables $(X(n))$ includes all lagged endogenous variables $(Y(n-1), Y(n-2), \dots)$, as well as all exogenous variables (predictors) for this system.

The $M \times K$ matrix Π changes abruptly at some unknown change-point m, i.e.

$$\Pi = \Pi(n) = \mathbf{a}I(n \le m) + \mathbf{b}I(n > m), \quad n = N, N+1, \dots \tag{2}$$

where $\|\mathbf{a} - \mathbf{b}\| > 0$.

Now let us formulate assumptions about the random noise process ν_n and predictors $X(n)$ defined on the probability space $(\Omega, \mathfrak{F}, \mathbf{P})$. Consider a filtration $\{\mathcal{F}_n\}$, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \mathcal{F}_n \subset \mathfrak{F}$, where $\mathcal{F}_n$ is the volume of information available at the instant n.

Suppose that predictors $X(n)$ and noises ν_n are continuously distributed and strictly stationary and the following conditions are satisfied:

1) the vector $X(n) = (x_{1n}, \dots, x_{Kn})'$ is $\mathcal{F}_{n-1}$ measurable.

2) there exists a continuous matrix function $V(t)$, $t \in [0, 1]$ such that for any $0 \le t_1 \le t_2 \le 1$

$$D_N(t_1, t_2) = \frac{1}{N} \sum_{j=[t_1 N]}^{[t_2 N]} X(j)X'(j) \to \int_{t_1}^{t_2} V(t)dt, \quad P - \text{a.s. as } N \to \infty,$$

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

3) the random vector sequence $\{(X(n), \nu_n)\}$ satisfies ψ-mixing and the unified Cramer condition.

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

The idea of our method is based upon the "moving window" statistic for sequential detection of a change-point. Suppose the size of this window is defined by a certain large parameter N. For any $n = N, N+1, \dots$ consider N last vectors of observations $Y(i), X(i), i = n - N + 1, \dots, n$.

The method of detection is constructed as follows. First, consider the $K \times K$ matrices:

$$\mathcal{T}^n(1,l) = \sum_{i=1}^{l} X(i+n-N)X^{'}(i+n-N), \quad l = 1,\dots,N, \tag{3}$$

second, the $K \times M$ matrices:

$$z^n(1,l) = \sum_{i=1}^{l} X(i+n-N)Y^{'}(i+n-N), \quad l = 1,\dots,N, \tag{4}$$

and third, the decision statistic

$$Y_N^n(l) = \frac{1}{N}(z^n(1,l) - \mathcal{T}^n(1,l)(\mathcal{T}^n(1,N))^{-1} z^n(1,N)). \tag{5}$$

where $l = 1,\dots,N$, $Y_N^n(N) = 0$ and by definition, $Y_N^n(0) = 0$.

Fix the number $0 < \beta < 1/2$. For detection of the change-point $m > N$, we define the stopping time

$$\tau_N = \inf\{n : \max_{[\beta N] \le l \le N} \|Y_N^n(l)\| > C\} \tag{6}$$

where C is a certain decision threshold, $\|A\|$ is the Euclidean norm of the matrix A.

1) Probability of type 1 error ("false decision"):

$$\alpha_N = \sup_n P_0\{\max_{[\beta N] \le l \le N} \|Y_N^n(l)\| > C\}, \tag{7}$$

2) Probability of type 2 error ("missed goal"):

$$\delta_N = P_m\{\max_{m \le n \le m+N} \max_{[\beta N] \le l \le N} \|Y_N^n(l)\| \le C\}.$$

This characteristic describes the situation when the decision statistic does not exceed the boundary C for a sample with a change-point, i.e. for $m \le n \le m+N$.

3) The normalized delay time in change-point detection:

$$\gamma_N = (\tau_N - m)^+/N, \tag{8}$$

where $a^+ = \max(0, a)$.

Theorem 1.

Suppose the above assumptions 1),3),4) are satisfied. Then for any $C > 0$ the following exponential upper estimate for the “false alarm" probability holds:

$$\alpha_N \le \phi_0(C_1) \begin{cases} \exp(-\dfrac{TNC_1\beta}{4\phi_0(C_1)}), & C_1 > hT \\ \exp(-\dfrac{NC_1^2\beta}{4h\phi_0(C_1)}), & C_1 \le hT, \end{cases} \tag{9}$$

where the constants h, T and $\phi_0(C_1) \geq 1$ are taken from Cramer's and ψ-mixing condition, respectively, $C_1 = C/(1+\sqrt{K})$.

Consider the $K \times K$ matrix

$$A(t) = \int_0^t V(\tau) d\tau, \quad 0 \leq t \leq 1.$$

Define $I = A(1)$. For any $0 < t \leq 1$ the matrix $A(t)$ is positive definite.

For any $0 \leq \theta \leq 1$, consider the function

$$g(\theta) = \|A(\theta)(E - I^{-1}A(\theta))(\mathbf{a} - \mathbf{b})^{'}\|,$$

where E is the unit matrix $K \times K$.

Theorem 2.

Suppose the above conditions 1)-4) are satisfied and $rank(D) = M$, where $D = (E - I^{-1}A(\theta))(\mathbf{a} - \mathbf{b})^{'}$. Assume also that the sequence $D_N(0,1)$ from condition 2) is uniformely bounded for any $\omega \in \Omega$. Denote $d = (g(\tilde{\theta}) - C)/(1+\sqrt{K})$. Then the following exponential upper estimate holds for type 2 error:

$$\delta_N \leq L_1 \exp(-L_2 d N), \tag{10}$$

where constants $L_1 > 0, L_2 > 0$ do not depend on N.

The relative delay time γ_N tends almost surely to a deterministic limit as $N \to \infty$:

$$\gamma_N = \frac{(\tau_N - m)^+}{N} \to \gamma^* \quad P_m - \text{a.s. as } N \to \infty, \tag{11}$$

where γ^* is the minimal root of the equation $g(t) = C$, $0 < \gamma^* < 1$.

Moreover, for any finite N and $0 < \epsilon < 1$ the following exponential inequality holds ($v = \epsilon/(1+\sqrt{K})$):

$$P_m\{|\gamma_N - \gamma^*| > \epsilon\} \leq \mathcal{L}_1 \exp(-\mathcal{L}_2 v N) \tag{12}$$

where constants $\mathcal{L}_1 > 0, \mathcal{L}_2 > 0$ do not depend on N.

Experiments

In this section we present results of a simulation study of the proposed method in comparison with other well known tests for detecting structural changes in model coefficients, i.e.

- Fluctuation test (Chu, et al. (1996))

- CUSUM test based on 'historical' OLS residuals (Ploberger, Kramer (1992))

- CUSUM test based on recursive residuals (Horvath, et al. (2004))

The following *regression model* was considered:

$$y_i = c_0 + c_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots,$$

where $x_i = 2 + \xi_i$ and ϵ_i, $\xi_i \sim \mathcal{N}(0,1)$ are independent Gaussian random sequences.

In order to estimate the false alarm rate, the regression model without structural changes was considered with $c_0 = 0$, $c_1 = 1$. Then models with a change-point in the coefficient c_1 were considered.

Method

a) CUSUM test based on 'historical' OLS residuals

Parameter $c_\alpha(\gamma) = 2.2365$ of this test was chosen to ensure the false alarm rate $pr = 0.05$.

Table 1. Performance characteristics of CUSUM test based on 'historical' residuals (5000 replications, pr - empirical false alarm rate, w_2 - type 2 error, $E\tau$ - average delay time)

n		25	50	100	200
pr		0.02	0.02	0.015	0.02
$c_1 = 1.5$	w_2	0.004	0	0	0
	$E\tau$	23.9	25.3	29.9	38.4
$c_1 = 1.3$	w_2	0.32	0.04	0.002	0
	$E\tau$	59.0	71.1	65.4	74.3
$c_1 = 1.2$	w_2	0.65	0.36	0.07	0.0
	$E\tau$	68.6	131.4	150.9	159.9

b) CUSUM test based on recursive residuals

In Table 2 we demonstrate the corresponding results for the CUSUM test based on recursive residuals. The parameter $a = 1.5$ of this test was chosen in order to ensure the empirial false alarm rate $pr = 0.02$.

Table 2. Performance characteristics of CUSUM test based on recursive residuals

n		25	50	100	200
pr		0.02	0.02	0.02	0.02
$c_1 = 1.5$	w_2	0.02	0	0	0
	$E\tau$	14.48	15.2	19.34	25.17
$c_1 = 1.3$	w_2	0.40	0.08	0.002	0
	$E\tau$	26.55	37.91	40.87	46.75
$c_1 = 1.2$	w_2	0.71	0.42	0.13	0.0
	$E\tau$	32.4	61.3	83.67	85.08

c) Fluctuation test

Table 3 below contains the corresponding results of Monte Carlo tests for the fluctuation test based on 'historical' regression estimates. The parameter $\lambda = 7.0$ of this test was chosen to ensure the empirical false alarm rate $pr = 0.02$.

Table 3. Performance characteristics of the fluctuation test

n		25	50	100	200
pr		0.02	0.02	0.02	0.02
$c_1 = 1.5$	w_2	0.32	0.25	0.004	0
	$E\tau$	21.5	28.4	29.5	31.5
$c_1 = 1.3$	w_2	0.47	0.43	0.40	0.04
	$E\tau$	157.3	182.7	201.26	207.71
$c_1 = 1.2$	w_2	0.93	0.89	0.80	0.55
	$E\tau$	202.2	278.7	345.6	389.7

d) Nonparametric test

$$C = \frac{\sigma(\max_i Ex_i^2)^{1/2}}{\sqrt{N}}\,\lambda,$$

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

We obtain the following formula for computation of $\lambda = th\sqrt{N}/2.2361$. The obtained results are reported in Table 4.

Table 4. Decision bounds for the nonparametric test

N	20	50	100	200	300	400	500
$p = 0.95$	0.65	0.51	0.32	0.24	0.18	0.16	0.14
$p = 0.99$	0.85	0.65	0.45	0.33	0.27	0.23	0.20
λ	1.7	2.05	2.01	2.08	2.09	2.05	2.00

Table 5. Performance characteristics of the nonparametric test (5000 replications, pr - empirical false alarm rate, w_2 - type 2 error, $E\tau$ - average delay time)

N		100	200	300	400
th		0.45	0.33	0.25	0.21
pr		0.021	0.025	0.015	0.025
$c_1 = 1.5$	w_2	0.05	0	0	0
	$E\tau$	18.04	28.4	32.3	35.5
$c_1 = 1.3$	w_2	0.13	0.05	0	0
	$E\tau$	29.0	50.1	53.3	62.1
$c_1 = 1.2$	w_2	0.43	0.36	0.06	0.01
	$E\tau$	44.4	65.6	85.9	90.5

2) System of simultaneous equations (SSE)

The following system of simultaneous econometric equations was considered:

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

where $\xi_i, \nu_i, \eta_i,\ i = 1, 2, \ldots$ are independent $\mathcal{N}(0, 1)$ r.v.'s.

So $(y_i, z_i)'$ is the vector of endogenous variables, x_i is the exogenous variable, and $(1, y_{i-1}, z_{i-1}, x_i)'$ is the vector of predetermined variables of this system.

The 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 stationary dynamics is characterized by the coefficients [0.1 0.5 0.3 0.7 0.2 0.4 0.6].

Table 8. Decision bounds of the nonparametric test (SSE model)

N	20	50	100	200	300	400
$p = 0.95$	0.99	0.67	0.49	0.39	0.30	0.25
$p = 0.99$	1.50	0.85	0.65	0.47	0.38	0.32
th	1.45	0.91	0.65	0.46	0.37	0.32

Table 9. Performance characteristics of the nonparametric test (SSE model, 5000 replications, pr - empirical false alarm rate, w_2 - type 2 error, $E\tau$ - average delay time)

N		20	50	100	200
th		1.50	0.85	0.65	0.47
pr		0.02	0.03	0.02	0.03
$c(6) = 0.95$	w_2	0.09	0	0	0
	$E\tau$	3.80	1.71	1.21	1.01
$c(6) = 0.9$	w_2	0.19	0.02	0	0
	$E\tau$	4.83	2.46	1.04	1.10
$c(6) = 0.8$	w_2	0.45	0.15	0.04	0
	$E\tau$	6.52	9.20	13.2	11.2

References

[1] Chu C., Stinchcombe M. and White H. (1996). Monitoring structural change, *Econometrica* 64: 1045–1065.

[2] Horvath L., Huskova M., Kokoszka P., and Steinebach J. (2004). Monitoring changes in linear models, *Journal of Statistical Planning and Inference* 126: 225-251.