Monitoring Structural Changes in Econometric Models

Boris Brodsky

1. Introduction

The problem of sequential detection and diagnosis of structural changes in stochastic multivariate systems on the basis of sequential observations has many applications, including detection of changes in parameters of regression equations, testing adequacy of econometric models, fault detection and isolation in complex dynamical systems. There is an extensive statistical and econometric literature dealing with methods of solving these problems.

Page (1954) considered the cumulative sums (CUSUM) test for detection of possible changes in the distribution function (d.f.) of a sequence of independent observations. Girshick and Rubin (1952) proposed a quasi-Bayesian test for solving the same problem.

In 1959 Kolmogorov and Shiryaev proposed the formal statement of the problem of "quickiest detection of spontaneous effects" which was later called the "disorder problem". Shiryaev in 1959-1965 found the optimal solution of this problem for the situation of full a priori information on the distribution function (d.f.) of observations and a change-point. This test coincides with the Girshick-Rubin test is therefore called the GRSh (Girshick-Rubin-Shiryaev) test.

In the situation when there is no a priori information on a change-point, Lorden (1971), Pollack (1985), and Moustakides (1986) proved that CUSUM and GRSh tests are asymptotically optimal in the problem of sequential detection of an abrupt change in the one-dimensional d.f. of independent observations.

Willsky (1976), Willsky and Jones (1976) pioneered research into sequential detection of abrupt changes in stochastic dynamical systems. Stochastic "noises" in such systems were assumed to be Gaussian and structural changes were interpreted as spontaneously emerging additive terms in equations of considered systems. For detection of these structural changes, the innovation process of the Kalman filter was used. Different methods generalising these ideas for sequential change-point detection in stochastic dynamical systems were proposed by Basseville and Benveniste (1983), Basseville and Nikiforov (1993), Nikiforov (1995), Bansal and Papantoni-Kazakos (1983).

In works of Lai (1995, 1998) the problem of sequential change-point detection in dynamical systems was generalized to the non-i.i.d. case. Lai considers the window-limited generalized likelihood ratio (GLR) schemes and proves their asymptotical optimality in different problems of sequential change-point detection in dynamical systems.

In spite of extensive research into the problem of sequential change-point detection in stochastic dynamical systems, several open problems still exist, and in particular, the problem of a priori information on observations.

Between the poles of the *full knowledge* (both the probabilistic mechanism of data generation and the specification of a system are known) and the *full ignorance* (neither the probabilistic mechanism of data generation nor the specification of a system is known) there exists the most practically relevant field of *semi-parametric* model description in which we know the specification of a stochastic system but the d.f. of observations is unknown to us. Some important examples include:

1) the multiple regression models and the systems of simultaneous equations in econometrics. As usual, we know the specification of a model (e.g., the linear regression or the autoregression model) but the d.f.'s of "noise" sequences are unknown to us. The problem consists in sequential detection of structural changes in these models. These structural changes include both abrupt changes in coefficients of equations and new terms in their specification (e.g., new additive factors in econometric equations).

2) conventional input-output dynamical systems in engineering and control science (the 'transfer function' of this system is known but the d.f. of the 'noise' sequence is unknown). The problem consists in sequential detection of spontaneous changes in the transfer function.

3) the general multivariate state-space model in which equations of the 'state' and 'observation' vector processes are known but the d.f. of 'noise' processes are unknown. Again the problem consists in sequential detection of possible changes in matrices of coefficients of the 'state' and 'observation' vector processes. The additional difficulty here is the fact that the state vector is unobservable, so possible changes in its statistical characteristics can be detected only via the analysis of the vector of observations.

The problem of monitoring structural changes in multivariate models attracted research interest in econometrics only at the end of 1990s. Chu, Stinchcombe, White (1996) considered the problem of monitoring structural changes in coefficients of a linear regression. They used the fluctuation test for sequential detection and diagnosis of abrupt changes in coefficients. These results were followed by Leisch, Hornik, Kuan (2000). Sequential tests based upon sums of regression residuals were used by Horvath, Huskova, Kokoszka, Steinebach (2004) for monitoring structural changes. Dynamic econometric models with structural changes were considered by Zeileis, Leisch, Kleiber, Hornik (2005).

The common drawback of these works is as follows: the quality of proposed tests is analyzed only from perspective of their limit distributions as the sample volume tends to infinity. Properties of these tests for finite sample volumes are investigated only empirically. Moreover, there is completely no research on optimality and asymptotic optimality of these methods.

In this paper a new method for monitoring structural changes in econometric models is proposed. The main performance characteristics of this method are studied for finite sample volumes. The a priori theoretical lower bounds for these performance characteristics are proved which enable us to analyze the asymptotic optimality of the proposed method. Monte Carlo study of the proposed method for static and dynamic econometric models is performed. Practical applications for the analysis of stability of the German quarterly model of demand for money (1961-1995) and the Russian monthly model of inflation (1994-2005) are considered.

2. Method of detection

Model

Consider the following basic specification of the multivariate system with structural changes:

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

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.

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

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

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

Model (1) generalizes many widely used econometric models, i.e.

- static and dynamic regression models with multiple predictors
- systems of simultaneous econometric equations.

Assumptions

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 $\{\mathcal{F}_n\}$, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \mathcal{F}_n \dots \subset \mathfrak{F}$, where $\mathcal{F}_n$ is the volume of information available at the instant n.

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two σ-algebras contained in $\mathfrak{F}$. Let $L_p(\mathcal{H})$ be a collection of L_p-integrated random variables measurable with respect to some σ-algebra $\mathcal{H} \subseteq \mathfrak{F}$. Define the following measure of dependence between $\mathcal{H}_1$ and $\mathcal{H}_2$:

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

Let $(\xi_i, i \geq 1)$ be a sequence of real random vectors on $(\Omega, \mathfrak{F}, \mathbf{P})$. Let $\mathfrak{F}_s^t = \sigma\{\xi_i : s \leq i \leq t\}, 1 \leq s \leq t < \infty$, be the minimal σ-algebra generated by random vectors $\xi_i, s \leq i \leq t$. Put

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

Definition 1.

A sequence $(\xi_i; i \geq 1)$ is said to be a sequence with *ψ-mixing* if the function $\psi(n)$ (which is also called the *coefficient of ψ-mixing*), tends to zero as n tends to infinity.

For any $\epsilon > 0$ define the number $\phi_0(\epsilon) \geq 1$ from the following condition: $\psi(l) \leq \epsilon$ for $l \geq \phi_0(\epsilon)$.

Definition 2.

A sequence $\{\zeta(n)\}$ of real random vectors $\zeta(n) \stackrel{\text{def}}{=} (\zeta_1(n), \dots, \zeta_k(n))$, satisfies the uniform Cramer condition if there exists a constant $H > 0$ such that

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

for any $i, j = 1, \dots, k$ and $|t| < H$.

In particular, for a centered random sequence ξ_n, the unified Cramer condition is equivalent to the following (see Petrov, 1987): there exist numbers $h > 0$, $T > 0$ such that for $0 < t < T$:

$$Ee^{t\xi_n} \leq exp(\frac{1}{2}t^2 h), \quad \forall n \geq 1.$$

Let us formulate assumptions about predictors $X(n)$ and noises ν_n. Suppose that predictors $X(n)$ and noises ν_n are random and strictly stationary and the following conditions are satisfied:

1) $X(n)$ is $\mathcal{F}_{n-1}$ measurable;

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

$$\lim_{N \to \infty} \mathbf{E} N^{-1} \sum_{j=[t_1 N]}^{[t_2 N]} X(j)X^{'}(j) = \int_{t_1}^{t_2} V(t)dt,$$

where $\int_{t_1}^{t_2} V(t)dt$ is the positively defined 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 $\{\mathcal{F}_n\}$.

These assumptions are satisfied in most practical problems of econometric analysis, and in particular, for multifactor regression models and systems of simultaneous econometric equations.

Method

The idea of our method is based upon the "moving window" statistic for sequential detection of a structural change. 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 matrices $K \times K$:

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

second, the matrices $K \times M$:

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

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, $Y_N^n(0) = 0$. The existence of the inverse matrix $\mathcal{T}^n(1,N))^{-1}$ for a large N follows from condition 2).

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

$$\tau_N = \inf\{n : \max_l \|Y_N^n(l)\| > C\}, \tag{6}$$

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

In the sequel we denote by $P_0(E_0)$ the measure (mathematical expectation) corresponding to the observed sequence without change-points and by

$P_m(E_m)$ - to the sequence with the change-point m. H_0 denotes the hypothesis of statistical homogeneity of observations (no structural changes); H_1 - the hypothesis about the presence of a change-point in the sample.

The proposed method has the following performance characteristics:

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

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

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

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

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 $m \le n \le m+N$. Remark that this definition formally corresponds to the following characteristic

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

However, $\delta_N^* \le \delta_N$ and therefore the exponential upper estimate for δ_N obtained in Theorem 2 is valid for δ_N^* as well.

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

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

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

In the following theorem the asymptotical behavior of the "false alarm" probability is studied.

Theorem 1.

Suppose the random process ν_n satisfies the uniform Cramer and ψ-mixing condition.

For any $C > 0$ the following exponential upper estimate for the "false alarm" probability holds:

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

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

In the following theorem we study type 2 error δ_N and the normalized delay time γ_N in sequential change-point detection.

Consider the matrix $K \times K$

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

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

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

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

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

Evidently, $g(0) = g(1) = 0$. Consider the point θ^* of the global maximum of $g(\theta)$ on the segment $[0, 1]$ - the root of the equation $E = I^{-1}A(\theta) + A(\theta)I^{-1}$, i.e. $A(\theta) = I/2$. In virtue of the above assumptions, the root of this equation exists and is unique. The function $g(\theta)$ is continuously differentiable by $\theta \in (0, 1)$.

Choose the decision threshold $0 < C < g(\theta^*)$. The following theorem holds.

Theorem 2.

Suppose the above assumptions 1)-4) are satisfied and $rank(D) = M$, where $D = (E - I^{-1}A(\theta))(\mathbf{a} - \mathbf{b})'$. Denote $d_1 = (g(\theta^*) - C)/(1 + K)$. Then for the probability of type 2 error the following upper estimate holds:

$$\beta_N \le \phi_0(d_1) \begin{cases} \exp(-\dfrac{N d_1 \beta}{4\phi_0(d_1)}), & d_1 > hT \\ \exp(-\dfrac{N d_1^2}{4h\phi_0(d_1)}), & d_1 \le hT, \end{cases} \tag{10}$$

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 $g(t) = C$, $0 < \gamma^* < 1$. Denote $G = \dfrac{dg(t)}{dt}|_{\gamma^*}$.

Moreover, for any finite N and $0 < \epsilon < 1$ the following inequality holds:

$$P_m\{|\gamma_N - \gamma^*| > \epsilon\} \le \phi_0(v) \begin{cases} \exp(-\dfrac{N v \beta}{4\phi_0(v)}), & v > hT \\ \exp(-\dfrac{N v^2 \beta}{4h\phi_0(v)}), & v \le hT. \end{cases} \tag{12}$$

where $v = \dfrac{\epsilon G}{1+K}$.

In theorems 1 and 2 the main performance characteristics of the proposed method were considered. These characteristics can be used for the study of asymptotic optimality of our method. This study is based upon the information-theoretical lower bounds for the main performance measures of any method of sequential detection of structural changes. These information bounds are proved in the next section.

3. Asymptotic optimality

In this section we prove that the proposed method is asymptotically (as $N \to \infty$) optimal w.r.t. certain criteria. The following model of observations is considered. Let $(\mathbf{z}_1, \mathbf{z}_2, \dots)$ be a random sequence of independent vector-valued observations defined on the probability space $(\Omega,\ \mathcal{F},\ P)$.

Suppose that the one-dimensional distribution density function $f(\cdot)$ of observations in this sequence changes at some unknown instant $m > 0$:

$$f(\mathbf{z}_n) = \begin{cases} f_0(\mathbf{z}_n, n/N), & \text{if } 1 \le n < m \text{ or if } m = \infty \text{ (no change)} \\ f_1(\mathbf{z}_n, n/N), & \text{if } m \le n \end{cases} \tag{13}$$

We assume here that the one-dimensional density functions $f_o(\cdot)$, $f_1(\cdot)$ of observations $\mathbf{z}_n$ depend on the time n. For formal convenience, we introduce here the normalized time $t = n/N$, where N is some "large parameter" of a method. In particular, for the method proposed in the previous section, N is the volume of the window-limited sample of the last N observations. Then the one-dimensional density function of observations changes at the time m/N, i.e. $f_0(\mathbf{z}, t) \neq f_1(\mathbf{z}, t)$ in some neighborhood of the change-point parameter m/N.

Define the stopping time:

$$\tau_N = \inf\{n :\ d_N(n) = 1\} \tag{14}$$

and the probability of the 1st type error (“false alarm"):

$$\alpha_N = \sup_k P_\infty\{d_N(k) = 1\}. \tag{15}$$

Define also the following value:

$$\gamma_N = (\tau_N - m)^+/N. \tag{16}$$

The following Theorem holds true.

Theorem 3.

For any $m \geq 1$ and sufficiently large N:

$$\mathbf{E}_m \int\limits_0^{\gamma_N} J(s)\, ds \geq \frac{|\ln (N\alpha_N)|}{N} + O(\frac{1}{N}), \tag{17}$$

where $J(t)$ is the Kullback information between distributions $f_0(\cdot, t)$, $f_1(\cdot, t)$:

$$J(t) = \int (\ln \frac{f_1(\mathbf{z}, t)}{f_0(\mathbf{z}, t)}) f_1(\mathbf{z}, t) dz. \tag{18}$$

In the sequel we assume that the function $J(t)$ is continuous.

The proof of this Theorem is given in the Appendix.

A method of sequential detection of structural changes is called *asymptotically optimal* if the lower bound in inequality (17) is attained as $N \to \infty$.

Now we prove another a priori inequality which gives the optimal rate of convergence of the normalized delay time to its limit value as $N \to \infty$.

Again we consider the sequence of *independent vector-value random variables* $\{x(n)\}$ with the one-dimensional density function depending on a certain parameter $\theta \in \Theta$ and satisfying the following relationship

$$\frac{d}{dz}\mathbf{P}\{x(n) \leq z\} = \begin{cases} f_0(z, \theta, n/N), & \text{if } n < m \text{ or if } m = \infty \text{ (no changes)} \\ f_1(z, \theta, n/N), & \text{if } m \leq n. \end{cases} \tag{19}$$

By $\mathbf{E}_{\theta m}$, ($\mathbf{P}_{\theta m}$) we denote the mathematical expectation (measure) corresponding to the sequence with the change-point at the instant m and the parameter of $\theta \in \Theta$ of the density function.

Below in this section we suppose that the following conditions are satisfied:

1) $\theta \in \Theta$, where Θ is an interval of the real axis;

2) $\frac{\partial^i}{\partial \theta^i} f_1(x, \theta, t)$ exists and is finite $P_{\theta m}$ almost surely for all $\theta \in \Theta$ and $i = 1, 2$.

3) $\int |\frac{\partial^i}{\partial \theta^i} f_1(x, \theta, t)|\, \mu(dx) < \infty$ for all $\theta \in \Theta$ and $i = 1, 2$.

4) $E_{\theta m}[\frac{\partial}{\partial \theta} \log f_1(x, \theta, t)]^2 < \infty$ for all $\theta \in \Theta$.

5) the sequence of the normalized delay times γ_N converges for every m $P_{\theta m}$-a.s. to some deterministic value $\gamma(\theta)$. The function $\gamma(\theta)$ is assumed to be continuously differentiable with respect to $\theta \in \Theta$ and $\gamma^{'}(\theta) \neq 0$. Without loss of generality we assume $\gamma'(\theta) > 0$.

6) The Fisher information

$$\mathbf{I}(\theta, t) = \int \frac{\left(f_1'(z, \theta, t)\right)^2}{f_1(z, \theta, t)} dz \tag{20}$$

is Riemann integrable with respect to t on $[0, \gamma(\theta)]$ for every θ.

Theorem 4.

Let $\epsilon > 0$ be fixed. Then under above assumptions 1)-6) the following inequality holds:

$$\liminf_{N} N^{-1} \ln P_{\theta m}\{|\gamma_N - \gamma(\theta)| > \epsilon\} \geq -\frac{\epsilon^2}{2[\gamma'(\theta)]^2} \int_0^{\gamma(\theta)} \mathbf{I}(\theta, t) dt. \qquad (21)$$

In theorems 3 and 4 the a priori theoretical lower bounds for the main performance characteristics in sequential change-point detection problems are established. These lower bounds are precise: the equality signs in (18) and (21) are attained for the nonparametric CUSUM test in the problem of sequential detection of abrupt changes in the mathematical expectation of an observed random sequence y_n, $n = 1, 2, \dots$ of the following type:

$$y_n = -a + hI(n \geq m) + \xi_n, \qquad (22)$$

where m is an unknown change-point, $a > 0$, $h > a$ and ξ_n is the centered independent sequence of "noises".

The CUSUM test has the following form: $\tau_c = \inf\{n : Z_n > C\}$,

$$Z_n = (Z_{n-1} + y_n)^+, \qquad (23)$$

where C is a certain bound.

It should be noted that the CUSUM test is essentially based upon the a priori information on the distribution function of observations before and after the change-point or on the supposed "direction" of this structural change (e.g., a shift "up" or "down" by the mathematical expectation).

However, this information, as a rule, is not available in considered problems of sequential detection of structural changes in multivariate stochastic systems (including multivariate econometric models). The proposed test enables us to effectively detect spontaneous structural changes in multivariate systems without this a priori information. However, we should pay some price for this generality: the proposed test is asymptotically optimal only by the order of convergence in (18) and (21).

Let us illustrate it by the example of a change in the mathematical expectation of observations. The decision statistic of our test in this case turns into

$$Z(n) = (\sum_{i=1}^{n} y_i - \frac{n}{N} \sum_{i=1}^{N} y_i)/N.$$

Under the null hypothesis the process $\sqrt{N}Z(n)$ tends to the Brownian bridge $\sigma(W(t) - tW(1))$ as $N \to \infty$, where $W(t)$, $0 \le t \le 1$ is the standard Wiener process. Therefore, for the type 1 error probability, the following asymptotic equality holds: $\alpha_N = (1 + o(1)) \exp(-2NC^2/\sigma^2)$. On the other hand, the normed by N delay time tends as $N \to \infty$ to the deterministic limit γ^* - the minimal root of the equation $\gamma^*(1-\gamma^*) = C/h$. Therefore inequality (18) in this case turns into $1 - (1 - 4\frac{C}{h})^{1/2} > 8C^2/h^2$. Remark that this inequality is strict for all $0 < C \le h/4$.

Now let us compare upper estimate (12) obtained in theorem 2 for the proposed method with a priori estimate from theorem 4. We easily conclude that the proposed method i asymptotically optimal by the order of parameters N and ϵ entering these inequalities. Moreover, for a Gaussian sequence of independent observations, $\phi_0 = 1$, $h = \sigma^2$, $I(\theta, t) = 1/\sigma^2$ and $g'(\gamma) = 1/\gamma'(\theta)$. Therefore, upper (12) and lower (21) estimate differ only by a constant. Remark, however, that these upper and lower estimate are precisely equal for the above CUSUM procedure in the one-dimensional case (see Brodsky and Darkhovsky (2000)).

4. Experiments

In this section some results of a small simulation study of the proposed method are given. This study was performed in order to evaluate the effectiveness of this method in different situations including sequential detection of structural changes in multiple regressions and systems of simultaneous equations.

1) Regression models

The following regression model was considered:

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

where $x_i = 2 + \xi_i$ and $\xi_i \sim \mathcal{N}(0, 1)$.

First, the regression model without structural changes was considered with $c_0 = 0$, $c_1 = 1$ and the maximums of decision statistic ( ) were computed in k=2000 trials of each experiment for different values of the sample volume N. Then the variation series of these maxima was constructed and the 95 percent and 99 percent quantiles computed. The values of 99 percent quantiles for each value of N were assumed to be the decision thresholds th. The obtained results are reported in Table 1.

Table 1.

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	0.40	0.33	0.27	0.23	0.20

In the following series of experiments the regression models with changes in the coefficient c_1 were considered. For each sample volume N and the chosen values of the decision threshold th, the estimates of the 1st ('false alarm') and the 2nd type error probabilities were computed, as well as the average delay time in change-point detection in $k = 2000$ independent trials. The results are reported in Table 2.

Table 2.

N		20	50	100	200	300	400
th		0.85	0.65	0.40	0.33	0.25	0.21
pr		0.03	0.015	0.07	0.025	0.015	0.025
$\|\ln(pr)\|$		3.50	4.20	2.50	3.69	4.20	3.69
$c_1 = 2.0$	w_2	0	0	0			
	$E\tau$	3.96	7.73	8.04			
$c_1 = 1.5$	w_2		0.47	0.05	0		
	$E\tau$		18.02	18.04	28.4		
$c_1 = 1.3$	w_2			0.13	0.05	0	
	$E\tau$			29.0	50.1	53.3	
$c_1 = 1.2$	w_2				0.36	0.06	0.01
	$E\tau$				65.6	85.9	90.5

From these results we conclude that efficient detection of smaller structural changes requires larger sample volumes N.

Now let us consider the following dynamic regression:

$$y_i = 2 + \rho y_{i-1} + u_i, \quad y_0 = 0,\ i = 1, 2, \dots$$

where $u_i \sim \mathcal{N}(0, 1)$.

Here we consider the problem of sequential detection of unknown changes in the coefficient ρ.

In the first series of tests the model without structural changes and the coefficient $\rho = 0.3$ was considered. The maximums of decision statistic ( ) were computed in k=2000 trials of each experiment for different values of the sample volume N. Then the variation series of these maxima was constructed and the 95 percent and 99 percent quantiles computed. The values of 99 percent quantiles for each value of N were assumed to be the decision thresholds th. The obtained results are reported in Table 3.

Table 3.

N	20	50	100	200	300	400	500
$p = 0.95$	0.73	0.52	0.38	0.28	0.24	0.20	0.18
$p = 0.99$	1.30	0.90	0.63	0.41	0.38	0.32	0.25

Then we consider the regression model with changes in the coefficient ρ. For each sample volume N and the chosen values of the decision threshold th, the estimates of the 1st ('false alarm') and the 2nd type error probabilities were computed, as well as the average delay time in change-point detection in $k = 2000$ independent trials. The results are reported in Table 4.

Table 4.

N		20	50	100	200	300
th		1.30	0.90	0.63	0.41	0.38
pr		0.03	0.02	0.04	0.04	0.02
$\lvert \ln(pr) \rvert$		3.50	3.91	3.21	3.21	3.91
$\rho = 0.7$	w_2	0	0	0		
	$E\tau$	3.39	3.06	2.76		
$\rho = 0.5$	w_2	0.35	0.18	0.04	0	
	$E\tau$	9.6	20.5	34.2	18.3	
$\rho = 0.4$	w_2			0.74	0.20	0.07
	$E\tau$			39.3	80.5	60.5

2) System of simultaneous equations

The following system of simultaneous econometric equations was considered:

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

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].

In the first series of tests the decision threshold was estimated. For this purpose, the model with the initial set of coefficients $\mathbf{u}$ and without structural changes was used. In 2000 independent trials the maximums of the decision statistic were computed and the variation series of these numbers constructed. The 95 and 99 percent quantiles are reported in the following table. The 99 percent quantiles were assumed to be the decision thresholds for the corresponding sample volumes. The obtained results are presented in Table 3.

Table 5.

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

In the following series of experiments the models with changes in the coefficient c_2 were considered. For each sample volume N and the chosen values of the decision threshold th, the estimates of the 1st ('false alarm') and the 2nd type error probabilities were computed, as well as the average delay time in change-point detection in $k = 2000$ independent trials. The results are reported in Table 6.

Table 6.

N		20	50	100	200
th		1.50	0.85	0.65	0.47
pr		0.04	0.06	0.05	0.06
$\|\ln(pr)\|$		3.21	2.81	2.99	2.81
$c(6) = 0.95$	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

From these results we again conclude that the smaller is the structural change in the coefficients of the considered SSE model, the larger must be the sample volume for effective detection of this structural change.

5. Practical applications

5.1. Demand for money in Germany

Lütkepol, Terasvirta, and Wolters (1999) analyzed stability of demand for money in Germany of 1961-1995s. The quarterly data on money aggregate M1, index of implicit price deflator of gross national product, real GDP, long-run interest rate of 1960(1)-1995(4) were used. The cointegration and ECM (error correction model) relationships for the demand for money function in the period 1961(1)-1990(2) were constructed which include the following variables:

$m = \log(M1/PN)$ - the logarithm of real M1 per capita;

$p = \log(P)$ - the logarithm of implicit price deflator;

$y = \log(Y/PN)$ - the logarithm of the real GDP per capita;
R - the nominal interest rate;
N - the size of the population;
$Q1, Q2, Q3$ - quarterly seasonal dummy variables.
The ECM model constructed by the authors has the following form:

$$\begin{aligned}\Delta m_t = \; & -0.30\Delta y_{t-2} - 0.67\Delta R_t - 1.00\Delta R_{t-1} - 0.53\Delta p_t \\ & -0.12 m_{t-1} + 0.13 y_{t-1} - 0.62 R_{t-1} \\ & -0.05 - 0.13Q1 - 0.016Q2 - 0.11Q3 + \hat{u}_t.\end{aligned}$$

All regression coefficients in this relationship except the intercept are statistically relevant at the error level 1 percent and the determination coefficient is $R^2 = 0.943$.

This model includes the series of residuals of the long-term cointegration relationship

$$e_{t-1} = -0.12 m_{t-1} + 0.13 y_{t-1} - 0.62 R_{t-1},$$

which is stationary by MacKinnon-Davidson criterion (see Mac-Kinnon (1993)).

This model was used by Lütkepol et al.(1999), Zeileis et al. (2005) for testing structural changes in the full sample of 1960(1)-1995(4).The OLE-based CUSUM test was used and a structural change detected at the point 1990(3).

The goal was to compare results of sequential detection of structural changes in the same volume of data using the method proposed in this paper with the above results of Lütkepol et al. In our tests the volume of the moving window $N = 70$ was chosen and the decision threshold $C = 0.011$ was computed by the sample 1961(1)-1980(4). Two points of structural changes $n1 = 52$, $n2 = 61$ corresponding to the periods 1990(2) and 1992(3) were detected in the whole sample using our method.

Comparing these results with Zeileis et al. (2005) we conclude that the method proposed in this paper is quite well adapted to sequential detection of structural changes in the real econometric data.

5.2. Russian inflation in 1994-2005s

First, we give the regression model for the rate of CPI inflation (pi=CPI/100-1) computed for the period 1994(1)-2004(12) (monthly data) with the following set of predictors:

- inflation expectations (pi(-1));

-the rate of money growth: mu=M2/M2(-1)-1, where M2 is the monetary aggregate M2;

- the rate of growth of the nominal exchange rate of dollar: eps=E/E(-1)-1;

- the rate of growth of electric energy tariffs for population: piel;
- seasonal dummy: Seas.

$$\begin{array}{llllllll} pi = & \underset{(0.214)}{0.0022} & + & \underset{(7.781)}{0.2734} & pi(-1)+ & \underset{(5.353)}{0.2105} & piel+ & \underset{(24.852)}{0.3547} \; eps+ \\ & \underset{(4.877)}{0.1639} & mu(-6)+ & \underset{(2.312)}{0.012} & Seas- & \underset{(-3.515)}{0.017} & Seas(-7) & \end{array}$$

The main quality characteristics of this model are as follows: R2 = 0.887; approximation error $\sigma = 0.015$; Breusch-Godfrey statistic for higher order residuals autocorrelation AR 1-7 F(7,111)=2.697. All these characterstics are quite good.

The proposed method of monitoring structural changes is based upon the chosen set of predictors for this model (i.e. specification of the model) but the concrete values of regression coefficients are not essential. The choice of the decision bound C (threshold) is also important. For this purpose the quasi-stationary subsample 1995(7)-1998(1) of observations was used. The decision bound C computed by this subsample equals $C = 0.002$. The volume of the "moving window" $n = 30$.

Two structural changes at the instants $n1 = 11$ and $n2 = 40$ were detected in the whole sample. These changes corespond to two important events in the Russian macroeconomic policy of 1990-2000s: introducing the " currency corridor" in June 1995 and the financial crisis of September 1998. So the proposed method enables us to detect substantial structural changes in the real econometric data.

Conclusion

A new method of monitoring structural changes in multivariate stochastic systems is proposed which enables us to effectively detect changes in parameters of econometric models by sequential observations. The a priori theoretical lower bounds for the main performance characteristics of sequential tests are proved including inequalities for the average delay time in detection of a structural change and the estimation error probability. The asymptotic optimality of the proposed method of monitoring structural changes is proved.

References

[1] **Banzal R.K., Papantoni-Kazakos P.** *An algorithm for detecting a change in a stochastic process* // IEEE Trans. on Inform. Theory. – 1983. – **20, 5. – 709–723.**

[2] Basseville M., Benveniste A. *Detection of abrupt changes in signals and dynamic systems. Springer. N.Y., 1986.*

[3] **Basseville M., Nikiforov I.** *Detection of abrupt changes: Theory and applications.* Prentice-Hall, N.Y., **1993.**

[4] Brodsky B., Darkhovsky B. *Non-Parametric Statistical Diagnosis: Problems and Methods.* Kluwer **2000**

[5] Chu C., Stinchcombe M., White H. *Monitoring structural change* // Econometrica. – 1996. – **64, 5, pp. 1045–1065.**

[6] Girshick M.A., Rubin H. *A Bayes approach to a quality control model* // Annals of Mathematical Statistics. – 1952. – **23, 1, 114–125.**

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

[8] Ibragimov I.A., Linnik Yu.V. *Independent and stationary dependent variables - M., Nauka, 1965.*

[9] **Kolmogorov A.N., Prokhorov Yu.V., Shiryaev A.N.***Probabilistical and statistical methods for detection of spontaneous effects - Proceedings of MIAN* **182.– 1988, pp.4-23.**

[10] Lai T.L. *Information bounds and quick detection of parameter changes in stochastic systems* // IEEE Transactions on Information Theory. – 1998. – **44, 2917–2929.**

[11] Lai T.L. *Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems* // IEEE Transactions on Information Theory. – 2000. – **46, pp. 595–608.**

[12] Leisch F., Hornik K., Kuan CM *Monitoring structural changes with the generalized fluctuation test* // Econometric Theory. - **2000, 16, pp.835-854.**

[13] Lorden G. *Procedures for reacting to a change in distribution* // Annals of Mathematical Statistics. – 1971. – **42, 6, pp. 1897–1908.**

[14] Moustakides G.V. *Optimal stopping times for detecting changes in distribution* // Ann.Stat. – 1986. – **14.** – 1379–1387.

[15] **Nikiforov I.** *A generalized change detection problem* // IEEE Transactions on Information Theory. – 1995. – **41, 1. – 171–187.**

[16] Nikiforov I. *Two strategies in the problem of change detection and isolation* // IEEE Transactions on Information Theory. – 1997. – **43, 2. – 770–776.**

[17] Page E.S. *A test for a change in a parameter occuring at an Unknown Point* // Biometrika. – 1955. – **42.** – 523–526.

[18] **Perron P.** *The Great Crash, the oil price shock and the unit root hypothesis* // Econometrica. – 1989. – **57.** – 1361–1401.

[19] **Petrov V.V.** *Sums of independent random variables. M. Nauka, 1987.*

[20] **Pollak M.** *Optimal detection of a change in distribution* // Annals of Statistics. – 1985. – **13.** – 206–227.

[21] **Shiryaev A.N.** *Optimality methods sequential change-point detection problems* // Probability Theory and Applications- 1963.– **8, 1, pp.26-51.**

[22] Willsky A.S. *A survey of design methods for failure detection in dynamic systems* // Automatica. – 1976. – **12. – 601–611.**

[23] Willsky A.S., Jones H.L. *A generalised likelihood ratio approach to detection and estimation of jumps in linear systems* // IEEE Trans. on Autom. Control. – 1976. – **AC–21, 1. – 108–112.**

[24] Zeileis A., Leisch F., Kleiber C., Hornik K. *Monitoring structural change in dynamic econometric models* // Journal of Applied Econometrics. - **2005, 20, pp.99-121.**

[25] Lütkepohl H., Terasvirta T., Wolters J. *Investigating stability and linearity of a German money demand function* // Journal of Applied Econometrics. - **1999, 14, pp.511-525.**