

АЛГОРИТМЫ ОБНАРУЖЕНИЯ ИЗМЕНЕНИЙ СВОЙСТВ НЕСТАЦИОНАРНЫХ ПРОЦЕССОВ В РЕЖИМЕ МОНИТОРИНГА⁴²

Гребенюк Е. А.

*Институт проблем управления им. В. А. Трапезникова РАН,
Россия, г. Москва, ул. Профсоюзная д.65*

lgreben@ipu.ru

Аннотация: В докладе рассматриваются задачи обнаружения и идентификации типов изменений нестационарных временных рядов, приводящих к необходимости перестройки моделей, описывающих процесс. Сравнивается два подхода к обнаружению и идентификации изменений в режиме мониторинга: флуктуационные тесты для обнаружения структурных сдвигов в режиме мониторинга и алгоритмы последовательного обнаружения, обладающие при известных условиях оптимальными свойствами в смысле критерия наискорейшего обнаружения и минимизации ошибок идентификации произошедших изменений.

Ключевые слова: Последовательный анализ, флуктуационные тесты, мониторинг, нестационарные процессы

Введение

Мониторинг нестационарных процессов, проводимый с целью обнаружения изменений их свойств, является необходимым элементом формирования достоверных прогнозов, выполняемых с использованием описывающих эти ряды моделей.

События, происходящие во внешней среде, такие как природные и экономические катаклизмы, политические события, смена стратегий участников финансовых и товарных рынков и пр. приводят к возникновению структурных сдвигов. Структурным сдвигом мы будем называть неожиданное изменение во времени параметров наблюдаемого временного ряда: статистических характеристик, коэффициентов описывающих его моделей, взаимосвязей между рядами.

Широкий класс методов прогнозирования макроэкономических и финансовых индексов основан на применении регрессионных и авторегрессионных моделей. В результате структурных сдвигов текущие наблюдения перестают соответствовать модели: она становится непригодной для описания и прогнозирования процесса. Качественные прогнозы мы можем получить только в том случае, если свойства прогнозируемого процесса и других процессов, включенных в модель, постоянны. Для получения качественных прогнозов в условиях изменяющейся внешней среды необходимо выполнение следующих условий: 1) своевременное обнаружение изменений, приводящих к необходимости смены или коррекции модели, 2) определение типа отклонения процесса от модели в результате структурного сдвига (изменение скорости или направления тренда, увеличение, уменьшение волатильности процесса, изменение причинных связей между различными процессами, возникновение или нарушение коинтеграции и пр.).

Алгоритмы мониторинга в режиме поступления данных выполняют контроль соответствия прогнозируемого процесса используемой модели. Если обнаружено изменение свойств процесса и удастся идентифицировать его тип, то прогнозирование может выполняться по модели, включающей информацию об обнаруженном изменении. Задача обнаружения и идентификации изменений нестационарных процессов в случае изменений тренда, волатильности, перехода процесса от нестационарного к стационарному рассматривалась в [1]. Для обнаружения и идентификации изменений использовались алгоритмы последовательного анализа.

Используемый в [2] подход А. Вальда основан на последовательной схеме обработки наблюдений с использованием статистики отношения правдоподобия. В качестве критерия качества обнаружения используется минимизация запаздывания при ограничениях на частоту ложных обнаружений и вероятность ошибочной идентификации. Последовательные алгоритмы ранее использовались, преимущественно, в технических приложениях. Однако в настоящее время появились работы, показывающие эффективность их применения для экономических данных [3,4].

Другим известным подходом к обнаружению изменений в режиме получения текущих наблюдений являются флуктуационные тесты для мониторинга, представляющие собой модификацию ретроспективного теста рекурсивных оценок момента изменения свойств [5], границы которого расширяются с ростом объема выборки [6]. Эти тесты просты и удобны в применении и реализованы в статистическом пакете *strucchange* 1.5-2 [7], в системе R для статистических вычислений, что значительно облегчает их использование для решения практических задач.

⁴² Работа выполнена в рамках комплексного проекта Минобрнауки России №13331-2 «Моделирование и интеллектуальное управление производственными процессами в промышленности и электроэнергетике».

В данной статье рассматривается и сравнивается качество обнаружения и идентификации изменений в случае нестационарных процессов для двух основных подходов: 1) флуктуационных тестов, модифицированных для применения в режиме мониторинга; 2) алгоритмов последовательного обнаружения и идентификации изменений.

1 Алгоритмы мониторинга структурных изменений с применением флуктуационных тестов

Постановка задачи. Рассмотрим модель множественной регрессии, которая построена по историческим данным на интервале $t = 1, 2, \dots, T$ [8]:

$$y_t = x_t' \beta + \varepsilon_t, \quad (1)$$

где x_t – вектор $r \times 1$ объясняющих переменных, $\beta = (\beta_1, \beta_2, \dots, \beta_r)$ – вектор коэффициентов, свойства случайных возмущений $\{\varepsilon_t\}$ и значения вектора $\beta = \beta_0$ на интервале $[1, T]$ являются постоянными. Проверяется гипотеза $H_0: \beta = \beta_0$ против гипотезы $H_1: \beta \neq \beta_0$ при $t = T, T + 1, \dots$, при условии, что выполняются следующие предположения:

$\{\varepsilon_t\}$ образует мартингальную разностную последовательность с дисперсией σ_ε^2 .

$\{x_t\}$ удовлетворяет условию $\lim_{T \rightarrow \infty} \sup \frac{1}{T} \sum_{t=1}^T \|\varepsilon_t\|^{2+\delta} < \infty$ для некоторого $\delta > 0$, откуда следует, что для некоторой нерегулярной не стохастической матрицы Q выполняется условие:

$$\frac{1}{T} \sum_{i=1}^T x_i x_i' \rightarrow Q.$$

Требуется построить алгоритм, обнаруживающий изменения в модели (1) в режиме поступления данных.

Структура алгоритма. Алгоритм мониторинга, основанный на применении обобщенного флуктуационного теста, включает следующие шаги:

- построение модели на основе исторических данных, на интервале, не содержащем изменений свойств;
- построение эмпирической функции отклонений модели от фактических значений при $t > T$, либо в «остатках» - отклонениях модели от фактических значений, либо в оценках вектора коэффициентов β , включающей вычисление кумулятивной функции оценок, для которой выполняется функциональная центральная предельная теорема (FCLT);
- формирование границ, при выходе за которые алгоритм фиксирует отклонение процесса от устойчивого режима, описываемого моделью (1).

Наиболее известны следующие алгоритмы мониторинга отклонений: по рекурсивным остаткам, по остаткам, рассчитанным по методу наименьших квадратов (МНК), по отклонениям в оценках вектора коэффициентов β [7]. При нулевой гипотезе вероятность пересечения эмпирической функцией граничных значений равна выбранному уровню достоверности α . Если эмпирическая функция пересекает эти границы, то нулевая гипотеза отклоняется. В качестве эмпирических функций отклонений авторы рассматривали кумулятивную функцию рекурсивных остатков (CUSUM-RE) и кумулятивную функцию OLS-остатков (CUSUM-OLS).

Алгоритм кумулятивных сумм рекурсивных остатков (CUSUM-RE). Эмпирическая функция отклонений определяется по формуле:

$$W_{n-k}(t) = \frac{1}{\sigma \sqrt{n}} \sum_{j=1}^{k+(n-k)t} w_j$$

где $\hat{\sigma}^2 = S_T / (T - k)$, S_T – сумма квадратов остатков построенной модели,

$$w_j = \frac{y_j - x_j' \hat{\beta}_j}{\sqrt{1 + x_j' (X_{j-1}' X_{j-1})^{-1} x_j}}$$

а- рекурсивные остатки, $\hat{\beta}_j$ - оценки β_j . При нулевой гипотезе w_1, w_2, \dots, w_T – независимые нормально распределенные значения. Процесс $W_{n-k}(t)$ аппроксимируется непрерывным гауссовским процессом Z_t , $k \ll t \ll T$, поскольку для процесса W_T выполняется функциональная центральная предельная теорема [8].

Алгоритм кумулятивных сумм МНК остатков CUSUM-OLS. Вместо рекурсивных остатков используются остатки, полученные методом наименьших квадратов, которые вычисляются как

$$\tilde{w}_j = y_j - x_j' \hat{\beta}_j$$

где $\hat{\beta}_j$ – стандартные оценки наименьших квадратов. CUSUM-OLS задается формулой:

$$\tilde{W}_n^0(t) = \frac{1}{\hat{\sigma}\sqrt{n}} \sum_{j=1}^{[nt]} \tilde{w}_j,$$

$\tilde{W}_n^0(t)$ сходится к стандартному Броуновскому мосту. Предельный процесс в момент времени $t = 1$ принимает нулевое значение. Если существует единственный структурный сдвиг в момент t_0 , то $\tilde{W}_n^0(t)$ имеет единственный максимум в окрестности точки t_0 . При работе в режиме мониторинга построенная эмпирическая функция сравнивается с новыми наблюдениями, а значения границ изменяются во времени.

Алгоритм мониторинга изменений в коэффициентах β процесса (ME). Алгоритм мониторинга оценок был впервые предложен в [6]. Как и в алгоритмах на основе остатков, авторы рассматривали отклонения оценок вектора β , полученных по исторической выборке, от оценок, полученных в скользящем окне вне исторической выборки [7], предельные процессы представляют собой k -мерный броуновский мост. Нулевая гипотеза отвергается, если эти отклонения выходят за заданные границы.

Последовательные тесты для обнаружения структурных изменений, пригодные для мониторинга формируют статистику и граничную функцию, такие что в отсутствие изменений значения этой статистики не превышают значений граничной функции. Для выполнения этого условия длительность мониторинга в ряде алгоритмов ограничивается известным заранее моментом времени M . При фиксированной длительности мониторинг начинается в момент $T + 1$ и заканчивается в M , при этом граничная функция зависит от отношения $m = M/T$. Процедуры мониторинга для фиксированной длительности предложены в работах: [9-11]. Алгоритмы с бесконечным горизонтом мониторинга рассматривались в работах [6,12-13].

2 Методы последовательного анализа для наискорейшего обнаружения изменений и идентификации типа изменения

Алгоритм наискорейшего обнаружения изменений свойств в режиме получения текущих наблюдений для независимой случайной последовательности был предложен Пейджем [14]. Исследуемый сигнал описывается последовательностью независимых случайных величин

$$y_t \in R, t = 1, 2, \dots, \quad (2)$$

распределенных по вероятностному закону:

$$F(y_t, \theta) = \begin{cases} F(y_t, \theta_0), & \text{if } t < t_a \\ F(y_t, \theta_1), & \text{if } t \geq t_a \end{cases}$$

где $F(y_t, \theta_i)$ – функции распределения, $\theta_i, i = 0, 1$ – векторы параметров законов распределения. Требуется обнаружить изменение свойств сигнала y_t в режиме получения наблюдений. Статистика алгоритма, вычисленная при получении очередного наблюдения, представляет собой логарифм отношения правдоподобия плотностей распределения сигналов $f_{\theta_0}, f_{\theta_1}$ до и после изменения свойств, соответственно: $s_t = \log \frac{f_{\theta_1}(y_t)}{f_{\theta_0}(y_t)}$. При получении следующего наблюдения алгоритм вычисляет решающую функцию $g(t)$ по формуле:

$$g(t) = (g(t-1) + s_t)^+, g(0) = 0, \quad (3)$$

где $x^+ = \max(x, 0)$, и сравнивает его с заданным порогом. Если решающая функция превышает пороговое значение, алгоритм генерирует сигнал об изменении свойств последовательности.

Как было показано в ряде исследований, алгоритм является оптимальным по критерию минимизации максимальной средней задержки обнаружения при заданных пределах частоты ложных срабатываний в случае точно известных значений параметра θ до и после изменения свойств. Рекурсивное вычисление значения $g(t)$ на каждом шаге алгоритма расширяет диапазон его

возможных применений в задачах мониторинга за счет уменьшения объема необходимых вычислений.

На основе алгоритма Пейджа в [15-16] был предложен Рекурсивный алгоритм диагностики резких изменений (Simple Recursive Algorithm for Diagnosis of Abrupt Changes - RADAC), на который в дальнейшем изложении мы будем ссылаться как на RADAC.

Алгоритм RADAC одновременно решает задачу наискорейшего обнаружения изменений (разладки) и диагностики ее типа, в случае нескольких гипотез. Решающая функция алгоритма вычисляется рекурсивно, что значительно повышает его вычислительную эффективность.

Постановка задачи. Пусть $Y_t \in R, t = 1, 2, \dots$, - независимая последовательность наблюдений с известным законом распределения $F(X_t, \theta_0)$ с параметром θ_0 , который в неизвестный и неслучайный момент времени t_α изменяется на закон распределения $F(X_t, \theta_m)$ с параметром θ_m :

$$F(Y_t, \theta) = \begin{cases} F(Y_t, \theta_0), & \text{если } t < t_\alpha \\ F(Y_t, \theta_m), & \text{если } t \geq t_\alpha \end{cases} \quad (4)$$

где $m = 1, 2, \dots, r, \{f_{\theta_m}(Y_t)\}$ - плотность распределения $\{F(Y_t, \theta_m)\}$. Требуется наискорейшим образом обнаружить факт изменения параметра θ и определить тип изменения m .

В каждый момент времени алгоритм RADAC вычисляет по наблюдениям $Y_t \in R, t = 1, 2, \dots$ пару (N, v) , где N - момент, в который алгоритм сигнализирует о наличии изменений, v - тип изменения. В [16] доказано, что при точно известных параметрах распределения сигналов до и после возможных изменений алгоритм оптимален по критерию минимума максимальной средней задержки обнаружения с ограничениями на частоту ложных срабатываний и вероятность ложной диагностики.

Пусть $E_{t_\alpha}^l, \bar{E}(N)$ - математические ожидания средней задержки обнаружения алгоритмом изменения параметра θ_l и максимальной средней задержки обнаружения для всех возможных параметров, соответственно, $E_0(N)$ - математическое ожидание средней длины интервала между ложными срабатываниями, $f_{t_\alpha}^l (v = j \neq l | N \geq t_\alpha)$ - вероятность ложной диагностики. Критерий качества работы алгоритма определяется как

$$\bar{E}(N) = \sup_{t_\alpha \gg 1, 1 \leq l \leq m} E_{t_\alpha}^l (N - t_\alpha + 1 | N \geq t_\alpha) \Rightarrow \min$$

где $f_{t_\alpha}^l$ равно f_0 , если $t < t_\alpha$, и f_l , если $t \gg t_\alpha, 1 \ll l \ll m$, при ограничениях:

$$E_0(N) \geq \gamma, \sup_{t_\alpha \geq 1} f_{t_\alpha}^l (v = j \neq l | N \geq t_\alpha) \leq \beta$$

где $1 \leq l, j \neq l \leq m, 0 < \beta \ll 1$.

В основе алгоритма лежит статистика отношения правдоподобия для гипотез $H_l: \theta = \theta_l$ и $H_j: \theta = \theta_j$:

$$s(l, j) = \log \frac{f_{\theta_l}(X_t)}{f_{\theta_j}(X_t)}$$

где $f_{\theta_l}(X_t), f_{\theta_j}(X_t)$ - плотности распределения вероятностей. Момент N обнаружения/идентификации алгоритма автор определяет как [15]:

$$N_r^l = \inf \left\{ t \gg 1: \min_{0 < j \neq l < m-1} [g_t(l, 0) - g_t(j, 0) - h_{l,j}] \gg 0 \right\},$$

где рекурсивная решающая функция $g_t(l, 0)$ имеет вид:

$$g_t(l, 0) = (g_{t-1}(l, 0) + s_t(l, 0))^+, x^+ = \max(x, 0), g_0(l, 0) = 0, l = 1, 2, \dots, m-1, g_0(0, 0) = 0,$$

порог $h_{l,j}$ вычисляется по формуле:

$$h_{l,j} = \begin{cases} h_d, & \text{if } l = 1, 2, \dots, m-1 \text{ and } j = 0 \\ h_i, & \text{if } l, j = 1, 2, \dots, m-1 \text{ and } j \neq l \end{cases}$$

где h_d, h_i - пороги обнаружения и идентификации, соответственно.

Методы последовательного анализа имеют широкий спектр применения. Разработаны алгоритмы, в которых входные сигналы описываются моделями временных рядов типа *ARIMA* (p, d, q), регрессионными моделями, моделями пространства состояний, *GARCH* и т. д. Ниже мы рассмотрим предложенную нами методологию применения последовательного анализа для мониторинга разностно - стационарных процессов, описывающих финансовые и макроэкономические показатели.

3 Применение флуктуационных тестов и последовательного анализа для обнаружения и идентификации изменений в режиме мониторинга

Для демонстрации свойств и сравнения алгоритмов мониторинга рассмотрим задачу обнаружения структурных сдвигов различного типа на модельных примерах.

Пусть исследуемый сигнал (2) представляет собой нестационарный процесс, который является стационарным относительно операции взятия первых разностей и описывается моделью авторегрессии с дрейфом:

$$y_t = \mu + \sum_{i=1}^k \alpha_i y_{t-i} + \varepsilon_t, \quad (5)$$

где $\alpha_i: \sum_{i=1}^k \alpha_i = 1$, - коэффициенты модели, μ - величина дрейфа, $\varepsilon_t \sim N(0, \sigma^2)$. Модель (4) тождественными преобразованиями может быть приведена к виду:

$$\Delta y_t = \mu + \sum_{i=1}^{k-1} \beta_i \Delta y_{t-i} + \varepsilon_t \quad (5a)$$

где μ - дрейф процесса, β_i - коэффициенты, полученные в результате преобразования (4) к (4a), $\Delta y_{t-i} = y_{t-i} - y_{t-i+1}$, $i = 1, 2, \dots, k-1$, $\{\varepsilon_t\}$, - та же, что и в (4).

В некоторый неизвестный момент времени t_a в процессе возникают изменения одного из следующих видов:

I. Изменяется дрейф процесса:

$$y_t = \mu + \sum_{i=1}^k \alpha_i y_{t-i} + DT_t \mu_2 + \varepsilon_t,$$

где μ_2 - величина изменения дрейфа, DT_t определяется формулой:

$$DT_t = \begin{cases} 0, & t < t_0 \\ 1, & t \gg t_0 \end{cases}$$

II. Изменяется дисперсия процесса:

$$y_t = \mu + \sum_{i=1}^k \alpha_i y_{t-i} + v_t,$$

где

$$v_t = \begin{cases} \varepsilon_t, & t < t_0 \\ \eta_t, & t \geq t_0 \end{cases}, \quad \eta_t \sim N(0, \sigma_2^2), \sigma_2^2 \neq \sigma_1^2.$$

III. Изменение типа процесса. Нестационарный процесс, интегрированный первого порядка, изменяется на стационарный:

$$y_t = \begin{cases} \mu + \sum_{i=1}^k \alpha_i y_{t-i} + \varepsilon_t, & t < t_0 \\ \mu + \sum_{i=1}^k \alpha'_i y_{t-i} + \varepsilon_t, & t_0 \ll t \end{cases}$$

где $\sum_{i=1}^k \alpha'_i < 1$, $\sum_{i=1}^k \alpha_i = 1$.

В [1] было показано, что для ряда разностей процесса (2) последовательность $Res_k, Res_{k+1}, \dots, Res_t$, которая представляет собой результат фильтрации ряда разностей $\Delta y_k, \Delta y_{k+1}, \dots, \Delta y_t$ моделью (4a), имеет вид:

$$Res_t = \Delta y_t - \tilde{\mu} - \sum_{i=1}^{k-1} \tilde{\beta}_i \Delta y_{t-i}, \quad (6)$$

где $\tilde{\mu}$, $\tilde{\beta}_i$ - оценки модели (4a), и после структурных сдвигов I-III ее статистические характеристики, изменяются в соответствии с Таблицей 1.

Таблица 1. Изменение статистических характеристик последовательности (5)

Тип изменения	Условное математическое ожидание последовательности	Условная дисперсия
Начальное состояние на участке без изменений	0	σ_0^2
Изменение коэффициентов и тренда	$\mu_2 \neq 0$	σ_0^2
Изменение дисперсии	0	$\sigma_2^2 \neq \sigma_0^2$
Переход к стационарному процессу	$-\mu$ – краткосрочное изменение в направлении противоположном направлению дрейфа	$(1 + \gamma)\sigma_0^2$, где $\gamma > 0$ – увеличение дисперсии

Таким образом обнаружение изменения среднего при постоянной дисперсии указывает на структурный сдвиг типа I, изменение дисперсии на структурный сдвиг типа II, а структурный сдвиг III мы обнаружим, если одновременно произойдут следующие события: 1) кратковременное изменение среднего последовательности (5) в направлении, противоположном оцененному по исторической выборке дрейфу μ , 2) увеличение ее дисперсии.

Для обнаружения и идентификации структурных сдвигов мы применяем последовательный алгоритм обнаружения и диагностики RADAC, предложенный в [15], алгоритм мониторинга CUSUM-OLS, и алгоритм мониторинга оценок ME. Последние два алгоритма запрограммированы в пакете strucchange, и процедуры их применения подробно описаны в работах [7-9].

На реализации алгоритма RADAC остановимся несколько подробнее. Процедура обнаружения и идентификации структурных сдвигов с использованием алгоритма, предложенного в [15], включает следующие шаги.

Шаг 1. По историческим данным на участке, не содержащем структурных сдвигов, по ряду разностей $\Delta y_1, \Delta y_2, \dots, \Delta y_t$, строим модель (4а) и формируем последовательность $Res_k, Res_{k+1}, \dots, Res_t$, с нулевым математическим ожиданием и дисперсией σ_0^2 , определяем диапазоны изменения дисперсии и математического ожидания для каждого из структурных сдвигов.

Шаг 2. Формируем 4 алгоритма последовательного обнаружения и идентификации (3):

а) для проверки гипотезы $H_0: \mu = \mu_0$ против набора гипотез $H_{m1}: \mu = \mu_0 + \Delta_1$ для обнаружения изменений среднего;

б) для проверки гипотезы $H_0: \sigma^2 = \sigma_0^2$ против набора гипотез $H_{\sigma 1}: \sigma^2 = \sigma_0^2 + \Upsilon_1$, где $\Delta_1, \Upsilon_1, l = 1, 2, \dots$, выбираются в соответствии с диапазонами изменения параметров при структурных сдвигах I и II, соответственно, по историческим данным.

При превышении алгоритмом одного из порогов формируется сигнал о наличии структурного сдвига, тип которого определяется выбранными гипотезами а) или б).

Для обнаружения структурного сдвига III применяется пара алгоритмов (3), один из которых проверяет кратковременное изменение среднего, другой – увеличение дисперсии. Диапазоны, в которых изменяются $\Delta_1, \Upsilon_1, l = 1, 2, \dots$, в каждом из алгоритмов выбираются на шаге 1 по историческим данным.

Шаг.3. Выбор порогов для каждого из алгоритмов выполняется по историческим данным для каждой из альтернативных гипотез H_1 или $H_{\sigma 1}$. В случае структурных сдвигов I и II обнаружение структурного сдвига фиксируется при превышении алгоритмом а) или б) одного из порогов, обнаружение структурного сдвига III фиксируется при одновременном превышении алгоритмом обоих порогов, выбираемых для проверки гипотез H_{m1} и $H_{\sigma 1}$.

4 Описание эксперимента

Эксперимент выполнялся с использованием моделирования по методу Монте – Карло, здесь приведены результаты обнаружения и идентификации одиночного структурного сдвига. Сравнивались алгоритмы: последовательного анализа и диагностики (3), кумулятивных сумм МНК

остатков CUSUM-OLS и мониторинга изменений в коэффициентах процесса ME. Для обнаружения и идентификации рассмотренных выше структурных сдвигов I, II, III анализировались следующие характеристики: среднее запаздывание в обнаружении, стандартное отклонение запаздывания, число ложных обнаружений, число пропущенных изменений.

Для каждого из структурных сдвигов моделировалась последовательность с изменениями свойств длиной 300 наблюдений, момент изменения – 150 - е наблюдение, число повторений $N=2000$. Параметры процесса до изменения свойств: среднее $\mu = 0.4$, стандартное отклонение $\sigma_0 = 1$. Результаты приведены в Таблице 2. Применение ансамбля алгоритмов

Таблица 2. Значения характеристик алгоритмов мониторинга

	Среднее запаздывание	Стандартное отклонение запаздывания	Ложные обнаружения	Ложная диагностика	Пропущенные
Структурный сдвиг I, параметры: среднее процесса до изменения $\mu = 0.4$, после $\mu = 1$					
CUSUM-OLS	12	8	0	не проводится	нет
ME	28.2	16.1	4.8%	не проводится	нет
RADAC	7.9	12	2%	Нет	нет
Структурный сдвиг II, параметры: стандартное отклонение до изменения $\sigma_0 = 1$, после $\sigma_0 = 2$					
CUSUM-OLS	37	27	20%	не проводится	85%
ME	28	19.5	30%	не проводится	87%
RADAC	7	8	4%	Нет	1%
Структурный сдвиг II, параметры: стандартное отклонение до изменения $\sigma_0 = 1$, после $\sigma_0 = 3$					
CUSUM-OLS	46	40	9%	не проводится	67%
ME	17	27.6	7.5	не проводится	20.5%
RADAC	3.4	3.2	0.25%	Нет	Нет
Структурный сдвиг III, параметры: среднее процесса $\mu = 0.4$, стандартное отклонение $\sigma_0 = 1$, авторегрессия первого порядка $a_1 = 1$ до изменения, после $a_1 = 0.95$					
CUSUM-OLS	4	6.7	0	не проводится	2%
ME	5.6	8.4	0	не проводится	1%
RADAC	0.3.	0.02	8	20%	2%

На Рис.1 приведены временные ряды со структурными сдвигами I, II и III, соответственно, вертикальной линией обозначен момент начала структурного изменения.

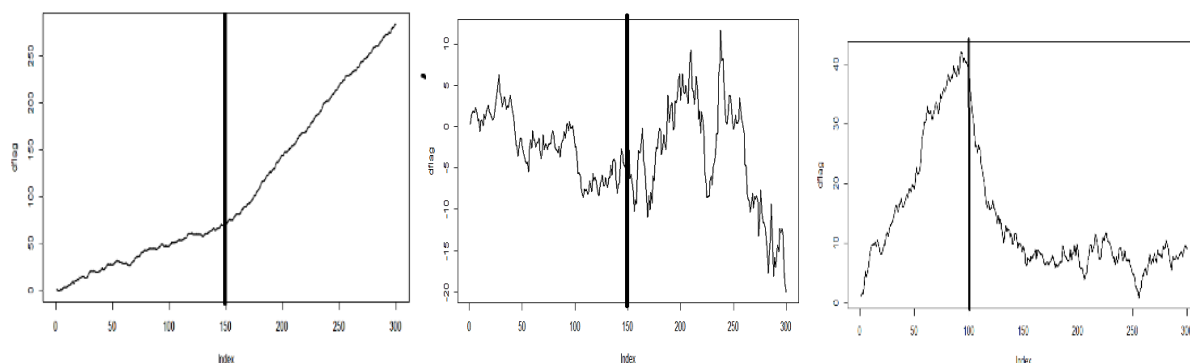


Рис. 1. Структурные сдвиги: увеличение среднего, увеличение дисперсии, изменение типа процесса

На Рис 2 представлены решающие функции алгоритмов обнаружения для изменений, показанных на Рис.1.

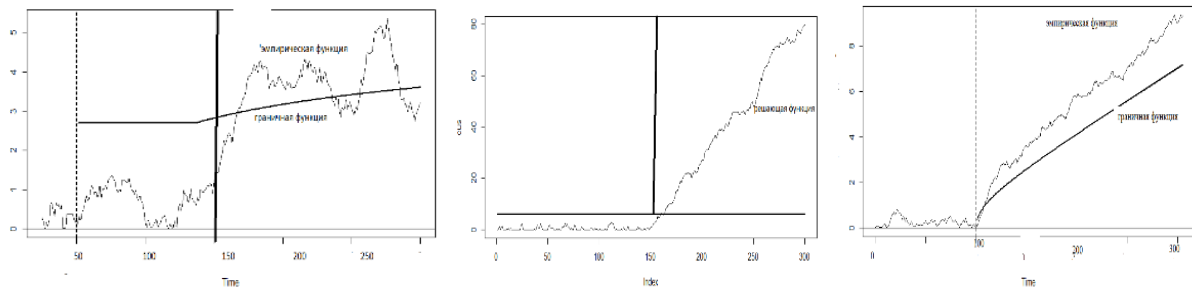


Рис. 2. Обнаружение структурных сдвигов: увеличения среднего, увеличения дисперсии, изменения типа процесса

На рисунке слева представлено изменение среднего, выполненное алгоритмом ME, на центральном рисунке – работа алгоритма RADAC с настроечными параметрами для обнаружения изменения дисперсии, на правом рисунке – обнаружение изменения типа процесса алгоритмом CUSUM-OLS. Вертикальной чертой обозначен момент изменения свойств, горизонтальные линии – пороги.

Заключение

Как следует из результатов эксперимента, алгоритм RADAC имеет минимальную задержку обнаружения для всех рассматриваемых режимов мониторинга, кроме обнаружения структурного сдвига I и идентифицирует тип разрыва. Алгоритм CUSUM-OLS имеет минимальную задержку при обнаружении дрейфа. Характеристики алгоритма ME для обнаружения дисперсии улучшаются с увеличением масштаба изменения дисперсии. Оба алгоритма, основанные на флуктуационных тестах, не имеют ложных тревог при обнаружении структурного сдвига III при незначительном увеличении задержки и разброса.

Результаты экспериментов показали, что каждый из рассмотренных нами алгоритмов имеет преимущества перед другими по отдельным критериям, что позволяет сделать вывод о возможности улучшения результатов мониторинга за счет использования ансамбля алгоритмов.

Литература

1. E. A. Grebenyuk. Monitoring and identification of structural shifts in processes with a unit root / Proceedings of the 13th International Conference "Management of Large-Scale System Development" (MLSD). Moscow: IEEE, 2020. С. <https://ieeexplore.ieee.org/document/9247829>.
2. Wald A., Sequential Analysis. John Wiley & Sons, Inc., New York. 1947.
3. A Tartakovsky, I Nikiforov, Mi Basseville Sequential analysis: Hypothesis Testing and Changepoint Detection 2014 P 603.
4. A. Pepelyshev and A. S. Polunchenko Real-time financial surveillance via quickest change-point detection methods, Statistics and Its Interface, v0, 1-14, 2015.
5. Ploberger W., Kramer W., Kontrus K, A new test for structural stability in the linear regression model, Journal of Econometrics, vol. 40, 1989, №2, pp. 307-318.
6. Chu C.S.J., Stinchcombe M. and White H., Monitoring Structural Change, Econometrica, 1996, Vol. 64, No.5, pp. 1045-1065.
7. Zeileis A, Leisch F. Hornik, K. Kleiber, strucchange: An R Package for Testing for Structural Change in Linear Regression Models. Journal of Statistical Software, vol. 7. 2002, №2, pp.1–38.
8. A. Zeileis, F. Leisch, C. Kleiber, and K. Hornik, Monitoring structural change in dynamic econometric models. Journal of Applied Econometrics, vol 20. 2005, № 1 pp. 99-121.
9. F. Leisch, K. Hornik, and C.-M. Kuan, Monitoring structural changes with the generalized fluctuation test. Econometric Theory, vol. 16. 2000, No. 6, pp. 835-85.
10. D. Wied, and P. Galeano, Monitoring correlation change in a sequence of random variables. Journal of Statistical Planning and Inference, 143. 2013, pp. 186-196.
11. H. Dette, J. and Gosmann, A likelihood ratio approach to sequential change point detection for a general class of parameters. Journal of the American Statistical Association, vol 115 (531), 2019, pp.1361-1277.

12. *L. Horvath, M. Huskova, P. Kokoszka, J. and Steinebach*, Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126: 2004, pp. 225-251.
13. *J. Gosmann, T. Kley, and H. Dette*, (2019). A new approach for open-end sequential change point monitoring. 2019, <https://arxiv.org/abs/1906.03225>.
14. *Page E.S.* Continuous inspection schemes, *Biometrika*. vol. 41. 1954, pp. 100–115.
15. *I.V. Nikiforov*, A simple recursive algorithm for diagnosis of abrupt changes in random signals, *IEEE Transactions on Information Theory*, vol.46, 2000, №7, pp. 2740–2746.
16. *I.V. Nikiforov*, Optimal sequential change detection and isolation, 15th Triennial World Congress, Barcelona, Spain, vol.35, 2002,7, pp. 29-34.