УДК 621.37 DOI: 10.17587/it.25.300-312

Н. Т. Абдуллаев[1], канд. техн. наук, зав. кафедрой, e-mail: a.namik46@mail.ru,
О. А. Дышин[2], канд. физ.-мат. наук, доц., **И. Д. Ибрагимова**[1], ассистент, e-mail: irada432@gmail.com,
Х. Р. Ахмедова[1], канд. физ.-мат. наук, e-mail: yubaba66@hotmail.com,
[1]Азербайджанский технический университет,
[2]Азербайджанский государственный университет нефти и промышленности

Сегментирование нестационарных физиологических сигналов с фрактальными свойствами

Для сегментирования нестационарного фрактального временного ряда сначала с помощью вейвлет-разложения определяется уровень разрешения, определяющий интервал самоподобия (фрактальности) исходного ряда и находится оценка трендовой составляющей. С использованием регрессионной модели временного ряда с волатильностью, представляющей детрендированный ряд, находим точки изменения волатильности, которые разделяют исходный ряд на квазистационарные участки (сегменты), соответствующие скачкам функции волатильности. Для оценивания точек изменения волатильности разработана вычислительная процедура, обобщающая итеративный метод центрированных кумулятивных сумм квадратов (ICSS).

Ключевые слова: *фракталы, вейвлет-разложение, нестационарность, временной ряд, волатильность*

Введение

Основной проблемой современной физиологии является нестационарность временных рядов, образующаяся под воздействием внешних условий. Поэтому многие исследователи концентрируют внимание на присущих физиологическим сигналам эффектах нестационарности [1, 2]. Отметим, что нестационарность проявляется не только под влиянием окружающей среды, но также является важным естественным свойством физиологических временных рядов, которое может привести к патологогическим последствиям.

Нестационарность временного ряда проявляется в том, что его статистические свойства изменяются с течением времени. В результате порождается неоднородность данных в том смысле, что среднее значение, стандартное отклонение и более высокие моменты (асимметрия, эксцесс) становятся зависимыми от интервала времени, на котором они вычисляются. Это приводит к тому, что анализ таких рядов усложняется, поскольку справедливость многих статистических процедур основана на предположении о стационарности исследуемых данных.

Кроме того, необходимо иметь в виду, что для многих процессов различной физической природы неотъемлемым является свойство статистической самоподобности, или фрактальности, выражающееся в масштабной инвариантности по отношению к тем или иным статистическим характеристикам. Так, например, наличие в динамическом ряде долговременных фрактальных корреляций, которые известны как корреляции типа $1/f$, приводит к неоднородностям на всех масштабах в форме самоподобия [3—5], затрудняя тем самым процедуру сегментирования исходного ряда, т. е. разбиения его на непересекающиеся смежные фрагменты, которые будут статистически однородны или, по крайней мере, будут обладать этим свойством в бóльшей степени, чем исходные данные. Эта проблема хорошо изучена и известна как задача нахождения точек изменения (change-point problem) [6].

Методы оценки самоподобности динамических рядов приведены в работе [7]. В значительной степени свободным от ограничений является методика оценки самоподобности, основанная на вейвлет-анализе динамического ряда [5]. Наличие самоподобности в некоторых компонентах исследуемого динамического ряда означает неинформативность этих компонентов для оценки происходящих в изучаемой системе изменений. В рамках мультиразрешающего анализа [8], который проводится последовательным применением дискретного вейвлет-преобразования к исходной реализации $x(t)$, приближение $\tilde{x}_j(t_i)$ к сигналу $x(t_i)$ на уровне разрешения j разбивается на две составляющие — грубую (аппроксимирующую) $\tilde{x}_j^a(t_i)$ и уточненную (детализирующую) высокочастотную составляющую $\tilde{x}^d(t_i)$ с последующим их уточнением итерационным методом. Используя логарифмическую характеристику самоподобности [7], по линейному возрастающему участку этой характеристики с показателем Херста $0{,}5 \leqslant H \leqslant 1$ можно определить масштабный диапазон, в котором проявляется самоподобность изучаемой реализации. Вейвлет-анализ дает возможность установить обрыв линейного участка и характерный спад вышеуказанной логарифмической характеристики, определяющей предельную глубину j_f масштабной инвариантности. В ходе обработки реальных электрокардиограмм (ЭКГ) удалось установить [5], что $j_f = 4$ характеризует предельную глубину масштабной инвариантности сердечного ритма. Алгоритм быстрого вейвлет-преобразования позволяет вычислять коэффициенты вейвлет-разложения итерационным методом [9].

Многократное применение прямого и обратного вейвлет-разложения при $j > j_f$ позволяет получить оценку $\hat{\mu}(t)$ трендовой (регулярной) составляющей $\mu(t)$ процесса $x(t)$ и выделить нерегулярную статистическую составляющую $z(t) = x(t) - \hat{\mu}(t)$, изменение статистических характеристик (средней и/или дисперсии) которой позволяет сегментировать исходный временной ряд.

Сегментирование нестационарных временных рядов, представляемых ЭКГ, позволяет установить границы, разделяющие качественно различные по своим статистическим свойствам составляющие ритмограмм, что согласуется с физиологическими представлениями, на которых основан классический анализ вариабельности сердечного ритма [10].

Для нахождения точек изменения (change-points) временного ряда нами использован метод непараметрического оценивания точек структурного изменения в моделях временных рядов с волатильностью [11], являющийся обобщением итеративного алгоритма кумулятивных сумм квадратов (iterated cumulative sums of squares — ICSS-alqorithm), предложенного в работе [12]. Для временного ряда нами использована модель [11] с неизвестной регрессивной функцией (условное среднее) и неизвестной функцией волатильности (условная вариация) с оценкой каждой из этих функций с помощью ядерных функций и последующего непараметрического оценивания точек структурного изменения волатильности. Для отыскания таких точек в работе [11] разработан алгоритм, основанный на статистике V_k^ν, $0 \leqslant \nu \leqslant 1/2$, определяемый по k ($k = 1, ..., N$) значениям временного ряда $\{y_i\}$, $i = 1, ..., N$ (N — общая длина ряда). При $\nu = 0$ данный алгоритм является модификацией, обобщающей ICSS-algorithm [12].

1. Оценка самоподобности случайного процесса

Самоподобность, или фрактальность, для детерминированных фракталов означает масштабную инвариантность (неизменяемость) их геометрической конфигурации в определенном диапазоне масштабов (интервале самоподобия). В отличие от детерминированных фракталов, для которых возможно точное воспроизведение свойств при масштабировании в интервале самоподобия, для случайных фрактальных процессов можно говорить о статистической самоподобности, выражающейся в масштабной инвариантности по отношению к тем или иным характеристическим свойствам.

По определению [7], случайный процесс называется фрактальным, если некоторые из его важных статистических характеристик проявляют свойство масштабированной инвариантности. Так как такое свойство масштабирования математически выражается степенными зависимостями, часто степенной характер статистических характеристик случайного процесса является признаком принадлежности его к классу фрактальных процессов.

Одним из определяющих свойств фрактальных процессов является самоподобность [13]. Случайный процесс X_t называется статистически самоподобным, если сам этот процесс и процесс $a^{-H}X_{at}$, полученный из X_t с учетом измененного временного масштаба at, имеют одинаковые конечномерные плотности распределения вероятностей для всех положительных целых n:

$$W\{X_1, X_2, \ldots, X_n\} = \\ = W\{a^{-H}X_{1a}, a^{-H}X_{2a}, \ldots, a^{-H}X_{na}\}, \tag{1}$$

где a — коэффициент расширения; $0{,}5 \leqslant H \leqslant 1$ — показатель Херста [7].

При исследовании процессов по конечным выборкам в большинстве случаев распределение вероятностей неизвестно, поэтому на практике чаще всего используют понятие самоподобности N-го порядка, когда процессы X_t и $a^{-H}X_{at}$ имеют одинаковые характеристики, по крайней мере, до N-го порядка включительно. Также используется понятие асимптотически подобных процессов, когда свойство (1) выполняется при $a \to \infty$.

К основным признакам самоподобных процессов второго порядка ($N = 2$) относятся:

1) гиперболический вид корреляционной функции

$$R(k) \cong k^{(2H-2)}L(t_k),\ k \to \infty, \tag{2}$$

где k — номер отсчета в дискретном временном ряде t_k; $L(t)$ — медленно меняющаяся функция при $t \to \infty$: $\lim\limits_{k\to\infty}[L(tx)/L(t)] = 1, \forall x > 0$;

2) медленно затухающая дисперсия

$$D[X_m] \sim m^{(2H-2)},\ m \to \infty, \tag{3}$$

где D — символ выборочной дисперсии; X_m — временная последовательность, полученная усреднением исходной последовательности $X_k = X(t_k)$ по непересекающимся последовательным блокам с размером m;

3) степенной характер спектральной плотности мощности (СПМ) $S(\omega)$ вблизи нуля: $S(\omega) \sim \omega^{-\gamma}L_2(\omega)$, $\omega \to \infty$, $0 < \gamma < 1$, $L_2(\omega)$ — медленно меняющаяся функция частоты [7].

Из соотношения (2) видно, что $\sum\limits_k R(k) \to \infty$ при $0{,}5 \leqslant H \leqslant 1$, что характеризует еще одно важное свойство многих фрактальных процессов — наличие долговременной корреляционной зависимости или, в данном случае, ее линейной составляющей. Однако подобный вид корреляционной функции не является необходимым условием фрактальности процесса.

Для многих процессов с фрактальной структурой характерным признаком является наличие распределения величин с "тяжелыми хвостами" [7]. Говорят, что случайная величина X имеет распределение $F(x)$ с "тяжелыми хвостами", если вероятность

$$P\{X > x\} = 1 - F(x) = x^{-\alpha}L(x),\ x \to \infty, \tag{4}$$

где $0 < \alpha < 2$ — индекс хвоста распределения, или параметр формы.

Как видно из соотношения (4), к распределениям с "тяжелыми" хвостами относятся распределения вероятностей, затухающие при $x \to \infty$ медленнее экспоненты. В то же время данное свойство также не является необходимым условием наличия фрактальной структуры процесса. Оценка параметра формы распределения α может быть выплнена на основании имеющейся в распоряжении реализации изучаемого процесса, но требует чрезвычайно большого объема статистических данных. Тем не менее, существует ряд тестов, позволяющих провести оценку параметра α по реализации в несколько тысяч отсчетов. Среди них следует отметить оценку Хилла, оценку по модифицированному QQ-графику и оценку момента DeHaan [7].

Как следует из определения (1), показатель Херста H является основной величиной, характеризующей самоподобность случайного процесса. При этом следует иметь в виду, что вывод о самоподобном характере случайного процесса на основании оценки, проведенной по одной реализации, может быть сделан, строго говоря, только при установленном свойстве его эргодичности, когда среднее по времени, установленное по одной реализации процесса, равно среднему по ансамблю всех реализаций (траекторий) процесса. В других случаях следует говорить о наличии самоподобной структуры процесса в заданном масштабном диапазоне для заданного набора данных.

К основным методам оценки показателя Херста для временных рядов относятся: анализ нормированного размаха, предложенный самим Херстом; анализ изменения выборочного значения дисперсии в зависимости от объема выборки; анализ СПМ с применением оценки Виттла и вейвлет-анализ [7].

В значительной мере свободной от ограничений и независимой от формы распределения (т. е. непараметрической) является методика оценки самоподобности на основе вейвлет-анализа динамического ряда, которая будет использована нами в расчетах. В основу данной методики положен мультиразрешающий анализ [8], который может быть проведен последовательным применением дискретного преобразования к исходной реализации. Любую функцию $x(t)$ из $L^2(R)$ ($L^2(R)$-пространство квадратично интегрируемых функций на евклидовом пространстве $R = (-\infty, \infty)$) можно разложить на некотором заданном уровне разрежения J в дискретный неизбыточный вейвлет ряд вида [15]

$$\tilde{x}(t) = \sum_{k\in Z} a_{J,k}\varphi_{J,k}(t) + \sum_{j=1}^{n_0}\sum_{k\in Z} d_{j,k}\psi_{j,k(t)}, \qquad (5)$$

где функции $\varphi_{j,k}(t) = 2^{-j|2}\varphi_0(2^{-j}t - k)$ и $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$; $\varphi_R = \{\varphi_{j,k}\}, \psi_k = \{\psi_{j,k}\}, k \in Z$ (Z — множество всех целых чисел), образуют ортонормированный базис в $L^2(R)$ из сдвинутых и расширенных реализаций масштабирующей функции $\varphi_0(t)$ и вейвлет-функции $\psi_0(t)$, называемой иногда "материнским вейвлетом".

Дискретное вейвлет-преобразование состоит в отыскании коэффициентов $a_{j,k}$ и $d_{j,k}$, определяемых как

$$a_{j,k} = \int x(t)\varphi_{j,k}(t)dt,\ d_{j,k} = \int x(t)\psi_{j,k}(t)dt. \qquad (6)$$

Итерационные формулы быстрого вейвлет-преобразования (БВП) имеют вид:

$$a_{j+1,k} = \sum_m h_m a_{j,2k+m},\ d_{j+1,k} = \sum_m g_m a_{j,2k+m} \qquad (7)$$

$$\text{с } a_{ok} = \int f(t)\varphi_0(t-k)dt. \qquad (8)$$

Коэффициенты g_m выражения (7) связаны через h_m соотношением

$$g_k = (-1)^k h_{2M-1-k};\ h_k = \sqrt{2}\int\varphi(t)\varphi(2t-k)dk;$$

целое число M определяет число ненулевых h_k и длину области задания вейвлета.

Основополагающая идея вейвлет-представления сигнала $x(t)$ заключается в разбиении приближения $\tilde{x}_j(t_i)$ к сигналу $x(t_i)$ на две составляющие — грубую (аппроксимирующую $\tilde{x}_{j-1}(t_j)_i$ и уточненную (детализирующую) $\tilde{x}^d_{j-1}(t_i)$ с последующим их уточнением итерационным методом с помощью формулы БВП (7):

$$\begin{aligned}\tilde{x}_j(t_i) &= \tilde{x}_{j-1}(t_i) + \tilde{x}^d_{j-1}(t_i) = \\ &= \sum_{k\in Z} a_{j-1,k},\varphi_{j-1,k}(t_i) + \sum_{k\in Z} d_{j-1,k}\psi_{j-1,k}(t_i),\end{aligned} \qquad (9)$$

где j — уровень разрешения; $a_j = \{a_{j-1,k}\}$, $d_j = \{d_{j-1,k}\}$ — наборы аппроксимирующих и детализирующих коэффициентов разложения ($j - 1$)-го уровня разрешения. Обычно в качестве $a_0 = \{a_{0,k}\}$ выбирается массив значений сигнала $x(t)$, $a_{0,i} = x\{t_i\}$.

Повторяя процедуру (9) m раз, $m = 1, ..., m_0$, разлагая каждый раз сглаженную функцию $\tilde{x}_{j-m}(t_i)$ на еще более сглаженную часть $\tilde{x}_{j-m-1}(t_i)$ и детализирующую часть $\tilde{x}^d_{j-m-1}(t_i)$, получаем вейвлет-разложение аппроксимации j-го уровня разрешения $\tilde{x}(t)$ для глубины разложения m.

С учетом соотношения [7]

$$D\left[\sum_k w_{j,k}\psi_{j,k}(t)\right] = \sum_k w^2_{j,k} \sim 2^{j(2H-2)} \qquad (10)$$

можно провести оценку самоподобности, выделив линейный участок на графике зависимости логарифмической характеристики

$$\log_2[(2^j/n_0)]\sum_k d^2_{j,k} \text{ от } \log_2 j. \qquad (11)$$

Наклон этого графика определяется как $2\widehat{H} - 1$ ($\widehat{H}$ — оценка показателя Херста H); n_0 — объем данных.

Аналитически доказано [14], что полученная таким образом оценка является несмещенной в достаточно произвольных условиях, а также эффективной при предположении гауссовской структуры данных.

В соответствии с выражением (10) линейно возрастающий участок зависимости (11) от $\log_2 j$ с показателем Херста $0{,}5 \leqslant H \leqslant 1$ соответствует масштабному диапазону, в котором проявляется самоподобность изучаемой реализации. На основании этого в работе [5] предложено использовать результаты вейвлет-анализа для выделения фрактального компонента исследуемой реализации как суммы восстановленных из вейвлет-разложения компонентов исходного динамического ряда для уровней разрешения j, соответствующих линейному участку зависимости (11). Обрыв линейного участка логарифмической характеристики (11) и характерный ее спад при некотором значении $j = j_f$ характеризуют предельную глубину самоподобности и связанной с ней масштабной инвариантности. В ходе обработки реальных данных ЭКГ с помощью вейвлет-анализа установлено, что для вариабельности сердечного ритма $j_f = 4$, что, по-видимому, является характерным для данного класса физиологических процессов.

Для повышения достоверности указанной выше процедуры выделения фрактального компонента рекомендуется при наличии достаточного объема исходных данных проводить оценку на различных пространственных масштабах k. Глубина самоподобности для любого исследуемого случайного процесса может быть установлена путем проведения подобной процедуры для ряда реализаций (теоретически по всему ансамблю реализаций), хотя для многих практических приложений можно ограничиться несколькими достаточно длинными реализациями при вариации всей совокупности воздействующих факторов.

Итерационная процедура (9) позволяет получить оценку (5) трендовой составляющей $\mu(t)$ при $J > j_f$ (обозначим ее $\hat{\mu}(t)$) и выделить нерегулярную стохастическую составляющую $z(t) = x(t) - \hat{\mu}(t)$, обладающую свойством стационарности (в широком смысле), т. е. с постоянной средней m_z и ковариационной функцией $K_z(t_1, t_2) = E[z(t_1)z(t_2)]$, зависящей только от разности аргументов:

$$K_z(t_1, t_2) = K_z(t_1 - t_2) = K_z(\tau).$$

В этом случае и корреляционная функция $R_z(t_1, t_2) = E[z(t_1) - m_z)(z(t_2) - m_z]$ зависит только от разности $\tau = t_2 - t_1$, $R_z = R_z(\tau)$; E — знак математического ожидания.

Для проверки процесса $z(t)$ на стационарность строится график выборочной автокорреляционной функции-коррелограммы [15]:

$$r_k = \frac{\sum_{t=k+1}^{n}(z(t) - m_z)(z(t-k) - m_z)}{\sum_{t=1}^{n}(z_t - m_z)^2}.$$

Коррелограмма стационарного временного ряда быстро убывает с ростом k.

2. Оценка точки изменения волатильности временного ряда

Рассмотрим модель временного ряда

$$Y_i = \mu(t_i) + \sigma(t_i)\varepsilon_i,\ i = 1, 2, ..., T, \tag{12}$$

где $\{Y_i\}$ — последовательность случайных величин и $\{\varepsilon_i\}$ — последовательность стационарных ошибок наблюдений величины Y в моменты времени t_i с $E(\varepsilon_i) = 0$ и $\mathrm{Var}(\varepsilon_i) = 1$; $\mu(t)$ и $\sigma(t)$ — регрессионная функция (условное среднее) и функция волатильности (условная вариация), соответственно. Модель (12) является частным случаем непараметрической модели [11]

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{13}$$

когда $X = t$ и, следовательно, $X_i = t_i$.

Следуя процедуре нахождения одной точки изменения (single change-point) волатильности [12], положим

$$\sigma(t) = \begin{Bmatrix} \tau_1\sigma_0(t), & \text{если } t \leqslant k_0 \\ \tau_2\sigma_0(t), & \text{если } t > k_0 \end{Bmatrix}, \tag{14}$$

где τ_1, τ_2 и k_0 — неизвестные параметры. Для упрощения положим $k_0 = [T\theta_0]$, $0 < \theta_0 < 1$, где $[a]$ означает наибольшее целое число, не превосходящее a.

Пусть гипотеза H_0 означает, что волатильность не имеет точек изменения. При выполнении этой гипотезы из (12) следует, что

$$\mathrm{E}(Y_i - \mu(t_i))^2 = \sigma^2(t_i).$$

Введем обозначения

$$Z_i = (Y_i - \mu(t_i)/\sigma_0(t_i),$$
$$S_T = \sum_{i=1}^{T} Z_i^2,\ S_k = \sum_{i=1}^{k} Z_i^2,\ S_{T-k} = \sum_{i=k+1}^{T} Z_i^2 \tag{15}$$

и определим

$$V_k = \left(\left(\frac{k(T-k)}{T^2}\right)\right)^{1|2}\left(\frac{1}{T-k}S_{T-k} - \frac{1}{k}S_k\right). \tag{16}$$

При известных функциях $\mu(t)$ и $\sigma_0(t)$ методом наименьших квадратов (МНК) для точки скачка волатильности k_0 можно получить следующую оценку [11]:

$$\hat{k} = \arg\max\left|V_k\right|. \tag{17}$$

При выполнении гипотезы H_0 выражения $S_{T-k}\sigma_0(t)/(T - k)$ и $S_k\sigma_0(t)/k$ будут несмещенными оценками для общей волатильности. Разность $(1/(T-k))S_{T-k} - (1|k)S_k$ (и, стало быть, V_k) будет близка к 0 при выполнении гипотезы H_0 и отлична от нуля, если волатильность изменяется. Простые вычисления приводят к равенству

$$\begin{aligned} V_k &= \left(\frac{(T-k)k}{T^2}\right)^{1/2}\left(\frac{1}{T-k}S_{T-k} - \frac{1}{k}S_k\right) = \\ &= \left(\frac{1}{k(T-k)}\right)^{1/2} S_T D_R, \end{aligned} \tag{18}$$

где

$$D_k = \frac{k}{T} - \frac{S_k}{S_T}. \tag{19}$$

Из соотношения (18) следует, что D_k также можно использовать в качестве оценки точки изменения k_0:

$$\hat{k} = \arg\max_k\left|D_k\right| = \arg\max_k(k(T-k))^{1/2}V_k. \tag{20}$$

Вообще говоря, можно ввести следующую статистику обнаружения изменения волатильности [11]:

$$V_k^{\nu} = \left(\frac{k}{T}\left(1 - \frac{k}{T}\right)\right)^{\frac{1}{2}-\nu} V_k, \; 0 \leqslant \nu \leqslant 1/2. \tag{21}$$

Статистика D_k, определяемая формулой (19), лишь по знаку отличается от статистики $D_k = \frac{S_k}{S_T} - \frac{k}{T}$, используемой в работе [12].

Из выражения (21) можно получить оценку для точки изменения k_0:

$$\hat{k}(\nu) = \arg\max_k \left|V_k^{\nu}\right|. \tag{22}$$

При проверке гипотезы $H_0{:}\lambda_T = 0$ против гипотезы H_1, где $\lambda_T = \tau_2^2 - \tau_1^2$, множитель $\left(\frac{k}{T}\left(1 - \frac{k}{T}\right)\right)^{\frac{1}{2}}$ при $0 \leqslant \nu \leqslant \frac{1}{2}$ обеспечивает явное преимущество статистике V_k, поскольку она становится очень близкой к смежной альтернативе $\lambda_T \sim T^{1/2}$ из работы [16].

Задавая $\hat{k}(\nu)$, можно получить следующие оценки величин τ_1^2 и τ_2^2:

$$\hat{\tau}_1^2(\nu) = \frac{1}{\hat{k}(\nu)} S_{\hat{k}(\nu)}, \; \hat{\tau}_2^2(\nu) = \frac{1}{T - \hat{k}(\nu)} S_{T - \hat{k}(\nu)}.$$

Для анализа асимптотических свойств оценок точек изменения волатильности вводятся следующие условия [11].

Условие J_1. Скачок функции условной вариации $\lambda_T = \tau_2^2 - \tau_1^2$ есть постоянная величина.

Условие J_2. Значение скачка стремится к нулю при неограниченном росте размера выборки, т. е. $\lambda_T \to 0$ и $T\lambda_T^2/\log T \to \infty$ при $T \to \infty$.

Условия (AS.1). Последовательность $\{t_i, \; \varepsilon_i\}$ удовлетворяет одному из двух альтернативных условий:

а) пусть $\mathcal{F}_i = \{t_1, t_2, \ldots, t_i, \varepsilon_1, \varepsilon_2 \ldots \varepsilon_{i-1}\}$. Предположим, что $\{\varepsilon_i\}$ — последовательность мартингал-разностей (с дискретным временем) относительно $\{\mathcal{F}_i\}$, т. е. $E|\varepsilon_i| < \infty$ для всех $i \geqslant 0$ и $E(\varepsilon_{i+1}/\mathcal{F}_i) = 0$ и $\sup_i E|\varepsilon_i|^{4+\delta} < \infty$ для некоторого $\delta > 0$.

б) $\{\varepsilon_i\}$ — строго стационарная последовательность и удовлетворяет условию перемешивания с коэффициентами перемешивания $\alpha(n)$, для которых $\sum_{n=1}^{\infty}(\alpha(n))^{\delta/2+\delta}\langle\infty, \delta\rangle 0$. К тому же, $E|\varepsilon_i|^{4+\delta} < \infty$.

Условие перемешивания для стационарных последовательностей определяется следующим образом [17].

Пусть $\{X_n, \; n \geqslant 1\}$ — стационарная последовательность случайных величин на вероятностном пространстве (с координатами Ω, $\mathcal{F}$, P). Определим для любых множеств $\mathcal{A}$, $\mathcal{B} \subset \mathcal{F}$ величины

$$\begin{aligned} &\rho(\mathcal{A}, \mathcal{B}) = \sup\left\{\left|\operatorname{corr}(X, Y)\right|, X \in L_2(\mathcal{A}), Y \in L_2(\mathcal{B})\right\}; \\ &\rho(n) = \sup_k \rho(X_i 1 \leqslant i \leqslant k), \sigma(X_i, i \geqslant k + n)). \end{aligned} \tag{23}$$

Последовательность $\{X_n, \; n \geqslant 1\}$ называется ρ-перемешанной, если $\{\rho(n) \to 0\}$ при $n \to \infty$.

При условиях J_1, J_2 и (AS1) верно асимптотическое предельное соотношение по вероятности [11]:

$$\hat{k}(\nu) - k_0 = O_P(1|\lambda_T^2) \tag{24}$$

для $0 \leqslant \nu \leqslant 1/2, \; k_0 = [\theta_0, T]$ для некоторого $0 < \theta_0 < 1$. Равенство (24) означает что при $T \to \infty$ $\hat{k}(\nu)$ сходится к k_0 по вероятности, т.е. $P\{|\hat{k}(\nu) - k_0| > \varepsilon\} \to 0$ при $T \to \infty$ со скоростью $1/\lambda_T^2$. При $0 \leqslant \nu \leqslant 1/2$ условие J_2 может быть ослаблено так, что $\lambda_T \to 0$ и $T\lambda_T^2 \to \infty$ при $T \to \infty$. Оценка точки изменения $\theta_\nu = \hat{k}(\nu)/T$ для $0 \leqslant \nu \leqslant 1/2$ при тех же условиях, что и в (24), сходится по вероятности к точной точке изменения $\theta_0 = k_0/T$ со скоростью $(T\lambda_T^2)^{-1}$, т. е.

$$\left|\theta_\nu - \theta_0\right| = 0_P((T\lambda_T^2)^{-1}). \tag{25}$$

Для установления асимптотического распределения статистики V_k^{ν} при известных функциях регрессии $\mu(t)$ и условной вариации $\sigma(t)$ вводится обозначение

$$V_T^{\nu}(t) = \\ = \begin{cases} 0, \text{ если } 0 \leqslant t < 1/(T+1); \\ V^{\nu}_{[(T+1)t]}, \text{ если } 1/(T+1) \leqslant t \leqslant T/(T+1); \\ 0, \text{ если } T/(T+1) \leqslant t \leqslant 1. \end{cases} \tag{26}$$

Если $\tau_1 = \tau_2 = 1$, т. е. верна гипотеза H_0, то при условии (AS.1) выполняется предельное соотношение [11]

$$\begin{aligned} &\lim_{T\to\infty} \frac{\sqrt{T}\sigma}{\sigma_w} \sup_{\delta < t < 1-\delta} V_T^{\nu}(t) \xrightarrow{d} \\ &\xrightarrow{d} \sup_{\delta < t < 1-\delta} (t(1-t))^{-\nu} \left|B(t)\right|, \end{aligned} \tag{27}$$

где $\xrightarrow{d}$ означает сходимость по распределению; δ — любое число из интервала (0,1); $\{B(t), 0 \leqslant \nu \leqslant 1\}$ — броуновское движение на [0,1]; $\sigma = \lim_{T\to\infty} S_T/T$ и

$$\sigma_w^2 = E(Z_1^2 - EZ_1^2)^2 + \\ + 2\sum_{i=2}^{T} E((Z_1^2 - EZ_1^2)(Z_i^2 - EZ_i^2)). \tag{28}$$

Аппроксимация распределения случайной величины $\left(\frac{1}{x^4}\right)$

$$\sup_{\delta<t<1-\delta} (t(1-t))^{-1/2} |B(t)|$$

из правой части (27) с $\nu = 1/2$ получается с помощью формулы [16]

$$P\left\{\sup_{h<t<1-l}\left(\frac{B^2(t)}{t(1-t)}\right)^{1/2} \geqslant x\right\} = \\ = \frac{x\exp\{-x^2/2\}}{(2\pi)^{1/2}}\left\{\log\frac{(1-h)(1-l)}{hl} - \right. \\ \left. -\frac{1}{x^2}\log\frac{(1-h)(1-l)}{hl} + \frac{4}{x^2} + 0\left(\frac{1}{x^4}\right)\right\}, \tag{29}$$

если положить в ней $h = l = \delta$. Это дает возможность рассчитать асимптотические критические значения правой части (27).

В случае наличия скачка волатильной функции, т. е. в случае гетероскедастичной регрессионной модели с $\sigma^2(t) \neq \text{const}$ (при $\sigma^2(t) = \text{const} = \sigma^2$ регрессионная модель называется гомоскедастичной), оценка точки изменения определяется как [11]

$$\hat{k} = \arg\max_k \left|\tilde{V}_k^\nu\right|, \tag{30}$$

где

$$\tilde{V}_k^\nu = \left(\frac{k(T-k)}{T^2}\right)^{1-\nu}\left(\frac{1}{T-k}\tilde{R}_{T-k} - \frac{1}{k}\tilde{R}_k\right), \tag{31}$$

$$\tilde{R}_k = \sum_{i=1}^{k}\tilde{W}_i^2,\ \tilde{R}_{T-k} = \sum_{i=k+1}^{T}\tilde{W}_i^2,\ \tilde{R}_T = \sum_{i=1}^{T}\tilde{W}_i^2; \tag{32}$$

$$\tilde{W}_i = Y_i - \hat{\mu}(t_i), \tag{33}$$

$\hat{\mu}(t)$ — оценка регрессионной функции $\mu(t)$. Предельные соотношения (24), (25) и (27) остаются справедливыми при замене $\bar{k}(\nu)$ и $\bar{V}_k$ на $\tilde{k}(\nu)$ и $\tilde{V}_k$, соответственно, если выполнены условия J_1, J_2, а вместо (AS.1) — следующие условия (AS.2)—(AS8) из работы [11]:

(AS2): $\mu(t)$ и $\sigma_0(t)$ непрерывны на $[a_1, a_2]$ вместе с производными до третьего порядка включительно;

(AS.3): плотность распределения $f(t)$ величин $\{t_i\}$ ограничена с $M < f(t) < M'$ (M, M' — положительные числа) и непрерывна на $[a_1, a_2]$ вместе с производными до второго порядка включительно;

(AS.4): функции условных плотностей $f_{t|y_i}$ и $f_{(t_1,t_l)|(y_1,y_l)}(t_1, t_l \mid (Y_1, Y_e))$ ограничены при всех $l > 0$;

(AS.5): $E|t|^l < \infty$ и $E|Y|^l < \infty$ при достаточно большом $l = 0$;

(AS.6): ядерная функция $K(\cdot)$, используемая для оценки $\sigma_0^2(x)$, имеет симметрическую функцию плотности с носителем из интервала $[-c_e, c_0]$ с ограниченной производной, и преобразование Фурье-функций $K(\cdot)$ абсолютно интегрируемо;

(AS.7): ширина b_n полос ядерной функции $K(\cdot)$ удовлетворяет неравенству

$$c_1 n^{-1/5} \leqslant b_n \leqslant c_2 n^{-1/5},$$

где c_1, c_2 — некоторые положительные постоянные (в расчетах нами принято $b_n = n^{-1/5}$);

(AS.8): (t_i, ε_i), $i = 1, 2, \ldots$ — строго стационарная последовательность и удовлетворяет условию перемешивания с коэффициентом $\alpha(n) = O(c^n)$ для некоторого $0 < c < 1$.

Условие (AS.8) более сильное, чем условие (AS.1) Поскольку последовательность времен $\{t_i\}$ состоит из независимых и одинаковых распределенных (с равномерным распределением) величин, условия (AS.3) и (AS.4) выполняются автоматически.

При наличии точки изменения k_0 у волатильности $\sigma(t)$ оценка для $\sigma_0(t)$ при условии $E(\varepsilon^2|t) = 1$ определяется обычно формулой [20]

$$\hat{\sigma}_s^2(t) = \frac{\sum_{i=1}^{n}\mathcal{K}_{n,b}(t)(t_i - t)(Y_i - \hat{\mu}(t_i))^2}{\sum_{i=1}^{n}\mathcal{K}_{n,b}(t_i - t)}, \tag{34}$$

где $1 \leqslant n \leqslant k_0$; $n \to \infty$ при $T \to \infty$ в случае выполнения альтернативной гипотезы H_1 ($\sigma(t)$ имеет на $[0, T]$ точку изменения k_0) и $n = T$ в случае выполнения нулевой гипотезы H_0 ($\sigma(t)$ не имеет на $[0, T]$ точку изменения).

В формуле (34)

$$K_{n,b}(t) = \frac{1}{b}K\left(\frac{t_i - t}{b}\right), \tag{35}$$

$K(\cdot)$ — ядерная функция и $b = b_n$ — последовательность ширин полос. Выбор ядерной функции $K(\cdot)$ и последовательности b_n обсуждается ниже.

Как известно, сглаживание полинома второго порядка $Y^2(t)$ или $(Y(t) - \hat{\mu}(t))^2$ чувствительно к выбросам. В частности, оценка (34) имеет очень большое смещение, когда имеется некоторое число выбросов или функция распределения наблюдений имеет "тяжелые хвосты". В связи с этим целесообразно использовать для оценки $\sigma_0(t)$ оценку абсолютного отклонения [21]:

$$\hat{\sigma}_d^2(t) = \frac{\sum_{i=1}^{n} y\mathcal{K}_{i_{n,b}}(t_i - t)|Y_i - \hat{\mu}(t_i)|}{\sum_{i=1}^{n} \mathcal{K}_{n,b}(t_i - t)}. \quad (36)$$

В качестве оценки величины σ в (27) принимается $\hat{\sigma} = s_T / T$. Состоятельную оценку для величины σ_w^2, являющейся дисперсией случайной величины $Z_t = (Y_t - \mu(t))/\sigma_0(t) = \varepsilon_t$, получим на основе анализа [17] выборочной стационарной последовательности $\{\widehat{Z}_i, i = 1, \ldots, T\}$, $\widehat{Z}_i = (Y_i - \hat{\mu}(t_i))/\hat{\sigma}_0(t_i)$ с $\hat{\sigma}_0(t_i) = \hat{\sigma}_d(t_i)$, где $\hat{\sigma}_d^2(t)$ определяется формулой (36), а $\hat{\mu}(t)$ — оценка $\overline{x}(t)$, получаемая вейвлет-разложением (5) с $J > j_f$.

Будем полагать, что $\widehat{Z}_n, n \geqslant 1$, есть стационарная последовательность случайных величин, обладающая свойством ρ-перемешивания при $n \to \infty$, т. е. согласно (23)

$$\rho(n) = \sup_k \rho(\sigma(\widehat{Z}_i, 1 \leqslant i \leqslant k), \sigma(\widehat{Z}_i, 1 \geqslant k + n)).$$

Пусть

$$E\widehat{Z}_1 = \mu_1,\ E\widehat{Z}_1^2 < \infty,\ S_n = S(n) = \sum_{i=1}^{n} \widehat{Z}_i. \quad (37)$$

Предположим, что $\sigma_n^2 = \mathrm{Var}\widehat{S}_n \to \infty$ и $\sum_{n=1}^{\infty} \rho(2^n) < \infty.$ Тогда

$$\sigma_n^2 / n \to \sigma^2 \text{ при } n \to \infty, \quad (38)$$

где σ — некоторое положительное число; Var — вариабельность значения [18].

Пусть далее $\{l_n, n \geqslant 1\}$ — последовательность целых положительных чисел с $1 \leqslant l_n \leqslant n$. Обозначим

$$S_j(k) = \sum_{i=j+1}^{j+k} \widehat{Z}_i,\ \overline{\widehat{Z}}_n = \frac{1}{n}\sum_{i=1}^{n} \widehat{Z}_i,$$
$$C_p = 2^{-p/2}\sqrt{\pi}\,\Gamma\left(\frac{p+1}{2}\right),\ p \geqslant 1, \quad (39)$$

где Γ — гамма-функция.

Введем следующие статистики:

$$B_{n,p} = \left\{ \frac{C_p}{n - l_n 1} \sum_{j=0}^{n-l_n} \left(\frac{\left| S_j(l_n) - l_n \overline{\widehat{Z}}_n \right|}{\sqrt{l_n}} \right)^p \right\}^{1|p}; \quad (40)$$

$$\widehat{B}_{n,p} = \left\{ \frac{C_p}{n - \hat{l}_n + 1} \sum_{j=0}^{n-l_n} \left(\frac{\left| S_j(l_n) - l_n \mu_1 \right|}{\sqrt{l_n}} \right)^p \right\}^{1|p}. \quad (41)$$

Если $l_n \to \infty$ и $l_n/n \to 0$ при $n \to \infty$, тогда при $n \to \infty$

$$B_{n,p} \to \sigma; \quad (42)$$

$$\frac{\widehat{S}_n - n\mu_1}{B_{n,p}\sqrt{n}} \xrightarrow{d} N(0,1), \quad (43)$$

где σ — то же, что и в выражении (38), при этом (43) остается справедливым при замене $B_{n,p}$ на $\widehat{B}_{n,p}$; $\xrightarrow{d}$ означает сходимость по распределению, что эквивалентно сходимости функции распределения левой части $F_n(x)$ к функции распределения правой части $F(x)$ в каждой точке x, где $F(x)$ непрерывна [19].

Таким образом, при вышеуказанных условиях на последовательность $\left\{\widehat{Z}_n, n \geqslant 1\right\}$ в качестве оценки величины σ_w можно принять $B_{n,p}$ или $\widehat{B}_{n,p}$ (в расчетах принималось $p = 1$ с $C_1 = \Gamma(1)$ и $l_n = \ln(n)$).

3. Выбор ядерной оценки и ширин полосы в оценках волатильности

В случае гладкой функции волатильности $\sigma(x)$ (т. е. функция $\sigma(x)$ непрерывна вместе с производными до некоторого кончного порядка $s \geqslant 1$) в ядерных оценках (34), (36), где $\mathcal{K}_{n,b}(t)$ представляется в виде (35), в качестве функции $\mathcal{K}(t)$ можно принять стандартизованное ядро Епанечникова

$$K(t) = 3(1 - t^2/5)\,|\,(t^2 \leqslant 5)/(4\sqrt{5}) \quad (44)$$

или ядро Гаусса

$$K(t) = \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\}. \quad (45)$$

Для оценки ширин полос b_n в выражении (35) воспользуемся методом, использующим χ^2-статистику $R^2\left(y, \mu, \hat{\sigma}_{b_n}\right)$, которая определяется как [22]

$$R^2(y, \hat{\mu}, \sigma_{b_n}) = \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}(t_i))^2}{\hat{\sigma}^2_{b_n}(t_i)}, \quad (46)$$

где $y = (Y_1, \ldots, Y_n)$; $\hat{\sigma}^2_{b_n}$ — оценка $\hat{\sigma}^2_d(\cdot)$, определяемая формулой (36) с шириной полосы $b = b_n$; ER^2 — ожидаемое значение χ^2-статистики Пирсона (46), приближенно равное числу степеней свободы n [22]. Так как R^2 зависит от оцениваемой (условной) вариации, то равенство (46) можно использовать для оптимального выбора ширины полосы b_n следующим образом:

$$\hat{b}_{n,\mathrm{opt}} = \arg\min_{b_n} \left| R^2(y, \hat{\mu}, \sigma_{b_n}) - n \right|. \quad (47)$$

4. Обнаружение нескольких точек изменения функции волатильности

Для обнаружения точек изменения (multiple change points) функции волатильности в работе [12] был предложен итеративный метод центрированных кумулятивных сумм квадратов (iterative cumulative sums of squares, ICSS). Суть метода заключается в следующем.

Рассмотрим стационарный временной ряд $\{a_t\}$ с нулевым средним значением и дисперсиями $\sigma_t^2, t = 1, \ldots, T$. Пусть $C_k = \sum_{k=1}^{n} a_k$ — кумулятивная (накопленная) сумма ряда $\{a_t\}$ и

$$D_k = \frac{C_k}{C_T} - \frac{k}{T}, k = 1, \ldots, T, \quad D_0 = D_T = 0 \quad (48)$$

— центрированная (и нормализованная) кумулятивная сумма квадратов.

Для постоянной дисперсии $\sigma_t = \sigma$ последовательность $\{D_k\}$ осциллирует относительно нуля. Если возможны внезапные изменения дисперсии σ_t, то график зависимости D_k от k с высокой вероятностью будет заключен в некоторых границах. Эти границы вычисляются из анализа асимптотического распределения D_k относительно некоторой постоянной дисперсии.

Пусть

$$k^* = \arg\max_k |D_k|. \quad (49)$$

Если этот максимум превышает упомянутые выше граничные уровни $|D_k|$ с уровнем значимости α, то k^* можно принять за оценку точки изменения дисперсии.

В работе [11] предложен модифицированный алгоритм ICSS (назовем его MICSS), в котором статистика D_k заменяется на V_k^{ν}. Критическое значение $c_{\alpha}^{\nu}(T)$ при $\nu = 0$ заданы в работе [12], а при $\nu = 1/2$ — в работе [16]. Значения $c_{\alpha}^{\nu}(T)$ при $\nu = 1/2$ нетрудно подсчитать, как будет показано ниже, с помощью неравенства (29). Обычно за уровень значимости применяется $\alpha = 0{,}05$.

В методе MICSS на шаге 1

$$M(t_i : t_2) = \max_{t_1 \leqslant k \leqslant t_2} \sqrt{(t_2 - t_1 + 1)/2} \left| D_k(a[t_i : t_2]) \right| \quad (50)$$

заменяется статистикой

$$M(t_1 : t_2) = \max_{t_1 \leqslant k \leqslant t_2} \frac{\hat{\sigma}\sqrt{t_2 - t_1 + 1}}{\hat{\sigma}_w} \left| \tilde{V}_k^{\nu}(a[t_1 : t_2]) \right|, \quad (51)$$

где $a[t_1 : t_2]$ означает, что выборка значений a_t берется при $t_1 \leqslant t \leqslant t_2$, где $t_1 < t_2$, а $\tilde{V}_k^{\nu}(a[t_1 : t_2])$ — статистика (31), определенная на интервале $[t_1, t_2]$.

В равенстве (51) оценка $\hat{\sigma}$ определяется по формуле

$$\hat{\sigma} = \frac{1}{T} \tilde{R}_T = \frac{1}{T} \sum_{i=1}^{T} (Y_i - \hat{\mu}(t_i)), \quad (52)$$

а $\hat{\sigma}_w$ — по формуле (40) с $p = 1$, где

$$\hat{Z}_i = \frac{(Y_i - \hat{\mu}(t_i))}{\hat{\sigma}_0(t_i)}, \hat{S}_j(l_n) = \sum_{i=j+1}^{j+l_n} \hat{Z}_i, \; \overline{\hat{Z}}_i = \sum_{i=j+1}^{n} \hat{Z}_i,$$

а $\hat{\sigma}_0(t_i) = \hat{\sigma}_d(t_i)$, где $\hat{\sigma}_d(t)$ определяется в (36). Оценки $\hat{\sigma}$ и $\hat{\sigma}_0(t_i)$, зависящие от n ($n = T$), являются состоятельными в том смысле, что при $n \to \infty$ они сходятся по вероятности к своим предельным значениям, т. е. вероятность того, что их отклонения от предельных значений (σ и $\sigma_0(t_i)$ соответственно) превышают любое заданное $\varepsilon > 0$, стремится к нулю при $n \to \infty$ [19].

Численная реализация алгоритма обнаружения изменения волатильности нестационарного временного ряда проводится в два этапа. На первом этапе по заданным значениям

$Y_1,\ldots,Y_{n_0}$ временного ряда вычисляются коэффициенты $w_{j,k}$ $k \in Z$, $j = 1, ..., J$, ($J \gg 4$) в вейвлет-разложении (5) с использованием пакета Wavelet Toolbox MATLAB [15]. Строим график зависимости логарифмической характеристики (11) от $\log_2 j$. Наклон линейно возрастающего участка этой зависимости определяет показатель Херста H, а его спад — границу j_f интервала самоподобности временного ряда. Задаваясь интервалом дискретизации Δt по времени t, вычислим вейвлет-разложение (5) при уровне разрешения $J > j_f$ и примем это разложение за оценку $\hat{\mu}(t)$ тренда $\mu(t)$ исходного ряда. Рассчитаем значения $\mu(t_i)$ во всех точках интервала наблюдения [1, T] с шагом Δt (в расчетах принималось $\Delta t = 0{,}01$).

На втором этапе для вычисления статистики $\tilde{V}_k^{\nu}$, определяемой формулой (31), по формулам (32) проводится расчет величин $\tilde{W}_i, \tilde{R}_k, \tilde{R}_{T-k}, \tilde{R}_T,$ по формулам (36) — величины $\hat{\sigma}_d^2(t_i)$, по которым получаются оценки $\hat{\sigma}_0(t_i) = \hat{\sigma}_{b_n}(t_i)$, и оптимальная ширина l_n полосы b_n оценки $\hat{\sigma}_d(t_i) = \hat{\sigma}_{b_n}(t_i)$ — по формуле (47). Оценка $\hat{\sigma}_w$ принимается равной величине $B_{n,p}$ с $p = 1$, вычисляемой по формуле (40). За ядерную функцию $K(t)$ при расчете $\hat{\sigma}_d^2(t_i)$ по формуле (46) принимается функция Епанечникова (44). Величина V_k^{ν} рассчитывается при $\nu = 0$ и $\nu = 1/2$ по формуле (31). Для обнаружения точек изменения волатильности $\sigma(t)$ модели временного ряда (12) будем использовать алгоритм MICSS.

Критические значения статистики (50) при различных $T^* = t_2 - t_1 + 1$ определены в таблице для разных уровней значимости α ($\alpha = 1 - p$).

T^*	100	200	300	400	500	∞
p	SE	SE	SE	SE	SE	
0,05	0,003	0,003	0,003	0,003	0,003	**0,520**
0,10	0,003	0,003	0,003	0,003	0,003	0,571
0,25	0,004	0,003	0,003	0,003	0,003	0,677
0,50	0,004	0,003	0,003	0,003	0,003	0,828
0,75	0,004	0,004	0,004	0,004	0,004	10,019
0,90	0,006	0,006	0,007	0,006	0,006	1,224
0,95	**0,009**	**0,004**	**0,008**	**0,010**	**0,009**	**1,358**
0,99	0,004	0,012	0,028	0,020	0,018	1,628

Стандартные ошибки в этой таблице получены как $\mathrm{SE} = \sqrt{p(1-p)/10{,}000}$, где p — доля серий с $\max\limits_k \sqrt{T/2}\,|D_k| < 1{,}358$. При $\alpha = 0{,}05$ (именно такой уровень значимости принимался нами в расчетах) асимптотические значения (при больших T) определяются как $c_{\alpha,0}(T^*) = 1{,}27;\ 1{,}30;\ 1{,}31;\ 1{,}31;\ 1{,}33;\ 1{,}358$ при $T^* = 100;\ 200;\ 300;\ 400;\ 500;\ \infty$, соответственно, будем предполагать, что

$$c_{\alpha,0}(T^*) = c_{\alpha,0}(100i) + \left(\frac{T^*}{100} - i\right) \times$$
$$\times (c_{\alpha,0}(100(i+1) - c_{\alpha,0}(100i)),$$

если

$$100i < T^* < 100(i+1) \text{ при } i = 1, 2, 4$$
$$\text{и } c_{\alpha,0}(T^*) = 1{,}31 \text{ при } 300 \leqslant T^* \leqslant 400.$$

Из соотношений (18), (22) следует

$$\left|V_k^0\right| = \left(\frac{k}{T}\left(1 - \frac{k}{T}\right)\right)^{1/2} \frac{1}{(k(T-k))^{1/2}} S_T\,|D_k| = \frac{S_T}{T}|D_k|.$$

С учетом оценки $\hat{\sigma} = S_T/T$ статистику (51) при $\nu = 1/2$ можно записать в виде

$$M(t_1{:}t_2) = \max_{k} \ _{t_1 \leqslant k \leqslant t_2} \frac{\hat{\sigma}^2\sqrt{t_2 - t_1 + 1}}{\hat{\sigma}_w}|D_k|. \quad (53)$$

Следовательно, критические значения статистики (51) при $\nu = 1/2$ (обозначим их $\tilde{c}_{\alpha,0}(T^*)$) отличаются от $c_{\alpha,0}(T^*)$ множителем $\frac{\hat{\sigma}\sqrt{2}}{\hat{\sigma}_w}$, т. е. $\hat{c}_{\alpha,0}(T^*) = \frac{\hat{\sigma}^2\sqrt{2}}{\hat{\sigma}_w} c_{\alpha,0}(T^*)$, где $T^* = t_2 - t_1 + 1$.

Для определения критических значений статистики (51) при $\nu = C$ (обозначим их $\hat{c}_{\alpha,1/2}(T^*)$) можно воспользоваться асимптотическим соотношением (27). С этой целью осуществим замену $\tilde{t} = t/T$, при которой $t = t_1 \to \tau = 1/T$ и $t = t_2 \to \tau = t_2/T$, при этом интервал $t_1 \leqslant t \leqslant t_2$ отображается в интервал $\tilde{t}_1 \leqslant \tilde{t} \leqslant \tilde{t}_2$.

Пусть $\delta = \frac{1}{T(T+1)}$. Тогда отрезки [1, T] и $\left[\frac{1}{T+1}, \frac{T}{T+1}\right]$ оси t после замены $\tilde{t} = t/T$ отображаются в отрезки $\left[\frac{1}{T}, 1\right]$ и $\left[\frac{1}{T(T+1)}, \frac{1}{T+1}\right]$ оси $\tilde{t}$ соответственно, при этом отрезок $\frac{1}{T(T+1)} \leqslant \tilde{t} \leqslant \frac{1}{T+1}$ находится внутри отрезка

$\delta \leqslant \tilde{t} \leqslant 1 - \delta$. Поэтому равенство (26) на оси $\tilde{t}$ запишется в виде $V_T^{\nu}(\tilde{t}) = V^{\nu}_{\left[\frac{T+1}{T}t\right]}$, а соотношение (27) можно записать на оси как

$$\begin{aligned} &\lim_{T\to\infty} \frac{\sqrt{T}\hat{\sigma}}{\hat{\sigma}_w} \sup_{\tilde{\delta}<\tilde{t}<1-\tilde{\delta}} \left|\tilde{V}_T^{\nu}(\tilde{t})\right| \xrightarrow{d} \\ &\xrightarrow{d} \sup_{\tilde{\delta}<\tilde{t}<1-\tilde{\delta}} (\tilde{t}(1-\tilde{t}))^{-\nu} \left|\tilde{B}(\tilde{t})\right|, \end{aligned} \qquad (54)$$

где $\tilde{\delta} = \delta/T$, $\tilde{B}(\tilde{t}) = B(t)|_{t=T\tilde{t}}$.

Обозначим α правую часть уравнения (29) без слагаемого $O\left(\frac{1}{x^4}\right)$. Решим относительно x нелинейное уравнение

$$\begin{aligned} &x\exp\{-x^2/2\}/\sqrt{2\pi} \times \\ &\times\left\{2\log\left(\frac{1}{\delta}-1\right)\left(1-\frac{1}{x^2}\right)+\frac{4}{x^2}\right\} = \alpha, \end{aligned} \qquad (55)$$

полученное из правой части (29) при $h = l = \tilde{\delta}$. При начальном условии $x = x_0$ (например, $x_0 = 100$) найдем решение x^* уравнения (55) при фиксированном α в системе MATLAB. Тогда величину $\tilde{c}_{\alpha,1|2}(T^*) = \sqrt{\frac{T^*}{T}}x^*$ можно принять за критическое значение статистики (51) при $\nu = \frac{1}{2}$.

Заключение

1. Под сегментированием обычно понимается задача разделения имеющегося временного ряда на сегменты (периоды) с разной динамикой.

2. Если представить поведение сложной системы с помощью адекватной модели, то переходному процессу будет соответствовать переход параметров модели в пространстве состояний от одной устойчивой фазовой траектории к другой.

3. В случае достаточно большого числа отсчетов в выборке и числа реализаций динамических рядов процедура сегментации сигнала на стационарные фрагменты должна давать однозначные результаты. Если сравнивать между собой распределения и дисперсии в двух последовательных фрагментах временного ряда, то в точке их стыка можно констатировать нарушение стационарности процесса в узком смысле этого слова.

4. Оценка статистических свойств физиологического сигнала возможна, однако, только на некотором конечном временном интервале. При этом сократить оцениваемый интервал путем увеличения числа отсчетов в единицу времени за счет большей частоты оцифровки изучаемого непрерывного физиологического процесса можно только до определенного предела: до того момента, когда соседние отсчеты не станут сильно коррелированными. Именно этим обстоятельством ограничивается возможность распространения теоретического определения стационарности на реальные физические процессы, поэтому было введено в практику понятие "квазистационарность", выделяющее понимание стационарности в таком ограничительно оценочном аспекте.

5. Наличие самоподобности, т. е. масштабной инвариантности по отношению к основным статистическим характеристикам, приводит к непрерывному повторению в интервале самоподобия свойств некоторых компонентов исследуемого динамического ряда, означая неинформативность этих компонентов для оценки происходящих изменений.

6. Наиболее свободной от различных ограничений является методика оценки самоподобности, основанная на дискретном вейвлет-преобразовании динамического ряда в рамках мультиразрешающего анализа. С помощью суммы восстановленных из вейвлет-разложения компонентов исходного динамического ряда выделяется фрактальная компонента с установлением глубины самоподобности изучаемого случайного процесса и находится оценка его трендовой составляющей.

7. С использованием регрессионной модели временного ряда с волатильностью, представляемой динамическим рядом после режектирования (удаления) его трендовой составляющий, можно определить точки изменения (change-points) волатильности, которые разбивают исходный ряд на квазистационарные сегменты, соответствующие скачкам функции волатильности. С этой целью применяется специальная вычислительная процедура, обобщающая итеративный метод центрированных кумулятивных сумм квадратов.

Список литературы

1. **Kantz H., Schreiber T.** Nonlinear Time series Analysis. Cambridge, Cambridge Univ. Press, 2003. 388 p.

2. **Hegger R., Kantz H., Matassini L.** Coping with Nonstationarity by Oerembedding // Physical Review Letters. 2000. Vol. 84. P. 3197.

3. **Theory** and Application of long-range dependence / Edited by P. Donkhan, G. Oppenhein, M. S. Taqqu. Boston: Biqrhauser, 2002. P. 355—367.

4. **Ivanov P. Ch., Amaral L. A. N., Goldberger A. L., Stanley H. E.** Stochastic feedback and regulation of biological rhythms // Europhys, Lett. 1998, 43(4). P. 363—368
5. **Богачев М. И.** Исследования влияний фрактальных свойств динамических рядов на оценку параметров их нестационарных фрагментов // Изв. вузов России. Радиоэлектроника. 2006. Вып. 3. С. 3—12.
6. **Change-point** problems. Lecture notes and monograph series / Edited by E. Carlstein, H. G. Muller, D. Siegmunal. Hayward, CA: Institute of Mathematical Statistics, 1994. Vol. 23. 385 p.
7. **Peintgen H. O., Jurgens H., Saupe D.** Chaos and fractals, New frontiers of science. New York, Springer-Verlaq, 1992. 984 p.
8. **Малла С.** Вейвлеты в обработке сигналов. М.: Мир, 2005. 671 с.
9. **Дремин И. М., Иванов О. В., Нечитайло В. А.** Вейвлеты и их наименования // УФН. 2001. Т. 171, № 5. С. 465—501.
10. **Marek M.** Heart Rate Variability: Standards of measurement, Physiological interpretation and clinical use // Circulation. 1996. Vol. 93(5). P. 1043—1065.
11. **Chen G., Choi Y., Zhou Y.** Nonparametric estimation of structural change points in volatility models for time series // Journal of Econometries. 2005. Vol. 12. P. 79—114.
12. **Inclan C., Tiao G.** Use of Cumulative Sums of square for retrospective detection of changes of variance // Journal of American Statistical Association. 1994. Vol. 89, N. 427. P. 913—923.
13. **Mandelbrot B. B., Van Ness J. W.** Fractional Brownian motions, fractional noises and applications // SIAM Rev. 1968. Vol. 10. P. 422—437.
14. **Abry P., Veitch D.** Wavelet analysis of long rage dependence traffic // IEEE Trans. Of Inf. Th. 1998. Vol. 44, N. 1. P. 2—15.
15. **Смоленцев Н. К.** Основы теории вейвлетов. Вейвлеты в МАТЛАБ. М.: ДМК пресс, 2008. 448 с.
16. **Csorgo M., Horvath L.** Limit Theorems in Change-point analysis. New York: Wiley, 1997. 438 p.
17. **Poligrad M., Shao Q. M.** Estimation of the Variance of partil Sums for ρ Mixing random variables // Journal of multivariate analysis. 1995. Vol. 52. P. 140—157.
18. **Poligrad M., Shao Q. M.** Self-normalizing central limit teorem for suns of weakly dependent random variables // J. Theoret. Probab. 1994. Vol. 7. P. 309—338.
19. **Ширяев А. Н.** Вероятность. М.: Наука, 1989. 640 с.
20. **Fan J., Yao Q.** Effiant estimation of conditional variance functions in stoachastic regression // Biometrika. 1998. Vol. 85. P. 645—660.
21. **Xia Y., Tong H., Li W. K.** Abstitude deviation estimation of volatility in nonparametric models // Research report No. 177. Departament of Statistics and actuarial science. The University of Hong Kong, Hong Kong, 1988. P. 133—173.
22. **Choin J. M., Muller H. G.** Nonparametrics quasilikelihood // The annual of statistics. 1999. Vol. 27. P. 36—64.

N. T. Abdullaev[1], Ph. D., Head of Department, e-mail: a.namik46@mail.ru,
O. A. Dyshin[2], Ph. D., Associate Professor, **I. D. Ibrahimova**[1], Assistant, e-mail: irada432@gmail.com,
Kh. R. Ahmadova[1], Ph. D., e-mail: yubaba66@hotmail.com,
[1] Azerbaijan Technical University,
[2]Azerbaijan State Oil and Industrial University

Segmentation of Non-Stationary Physiological Signals with Fractal Properties

Segmentation is usually understood as the task of separation an available time series into segments (periods) with different dynamics.

If we present the behavior of a complex system with an adequate model, then the transition process will correspond to the transition of model parameters in the state space from one stable phase trajectory to another.

In the case of a sufficiently large number of countdown in the sample and the number of realizations of the dynamic series, the procedure of segmentation of a signal into stationary fragments should give unambiguous results. If compare the distributions and dispersions in two successive fragments of a time series, then at the point of their junction, can state a violation of the stationarity of the process in the narrow sense.

Evaluation of the statistical properties of the physiological signal is possible, however, only on a certain finite time interval. At the same time, the estimated interval can be reduced by increasing the number of samples per unit of time due to the higher frequency of digitization of the studied continuous physiological process only up to a certain limit: until the moment when neighboring counts become strongly correlated. It is this circumstance that limits the possibility of distribution the theoretical definition of stationarity to real physical processes, therefore the concept of "quasi-stationarity" was introduced into practice, highlighting the understanding of stationarity in such a restrictively evaluative aspect.

Presence of self-similarity, i.e. scale invariance in relation to the main statistical characteristics leads to continuous repetition in the self-similarity interval of the properties of some components of the studied dynamic series, meaning that these components are not informative for assessing the changes.

The most free from various restrictions is the method of self-similarity assessment, based on discrete wavelet transform of the dynamic series in the framework of multiresolution analysis. Using the sum of the components of the original time series reconstructed from the wavelet decomposition, the fractal component is selected with the determination of the self-similarity depth of the random process under study and an estimate of its trend component is found.

Using the regression model of a time series with volatility represented by a dynamic series after rejection (removal) of its trend component, it is possible to determine change-points of volatility, which divide the initial series into quasistationary segments corresponding to the volatility function jumps. For this purpose, a special computational procedure is applied, which generalizes the iterative method of centered cumulative sums of squares.

Keywords: *fractals, wavelet expansion, nonstationarity, time series, volatility*

DOI: 10.17587/it.25.300-312

References

1. **Kantz H., Schreiber T.** Nonlinear Time series Analysis, Cambridge, Cambridge Univ. Press, 2003, 388 p.
2. **Hegger R., Kantz H. and Matassini L.** Coping with Nonstationarity by Oerembedding, *Physical Review Letters*, 2000, vol. 84, p. 3197.
3. **Donkhan P., Oppenhein G., Taqqu M. S.** ed. Theory and Application of long-range dependence, Boston, Biqrhauser, 2002, pp. 355—367.
4. **Ivanov P. Ch., Amaral L. A. N., Goldberger A. L., Stanley H. E.** Stochastic feedback and regulation of biological rhythms, *Europhys, Lett.,* 1998, 43(4), pp. 363—368.
5. **Bogachev M. I.** *Issledovaniya vliyatel'nykh fraktal'nykh svoystv dinamicheskikh ryadov na otsenku parametrov ikh nestatsionarnykh fragmentov* (Studies of the effects of fractal properties of dynamical series on the estimation of the parameters of their nonstationary fragments), *Izv. Vuzov Rossii. Radioelektronika,* 2006, iss. 3, pp. 3—12 (in Russian).
6. **Carlstein E., Muller H. G., Siegmunal D.** ed. Changepointproblems. Lecture notes and monograph series, Hayward, CA, Institute of Mathematical Statistics, 1994, VD 23, 385 p.
7. **Peintgen H. O., Jurgens H., Saupe D.** Chaos and fractals, New frontiers of science, New York, Springer-Verlaq, 1992, 984 p.
8. **Malla S.** *Veyvlety v obrabotke signalov* (Wavelets in signal processing), Moscow, Mir, 2005, 671 p. (in Russian).
9. **Dremin I. M., Ivanov O. V., Nechitaylo V. A.** *Veyvlety i ikh naimenovaniya* (Wavelets and their names), *UFN*, 2001, vol. 171, no. 5, pp. 465—501 (in Russian).
10. **Marek M.** Heart Rate Variability: Standards of measurement, Physiological interpretation and clinical use, *Circulation*, 1996, vol. 93(5), pp. 1043—1065.
11. **Chen G., Choi Y., Zhou Y.** Nonparametric estimation of structural change points in volatility models for time series, *Journal of Econometries,* 2005, vol. 12, pp. 79—114.
12. **Inclan C., Tiao G.** Use of Cumulative Sums of square for retrospective detection of changes of variance, *Journal of American Statistical Association*, 1994, vol. 89, no. 427, pp. 913—923.
13. **Mandelbrot B. B., Van Ness J. W.** Fractional Brownian motions, fractional noises and applications, *SIAM Rev.*, 1968, vol. 10, pp. 422—437.
14. **Abry P., Veitch D.** Wavelet analysis of long rage dependence traffic, *IEEE Trans. Of Inf. Th.,* 1998, vol. 44, no. 1, pp. 2—15.
15. **Smolentsev N. K.** *Osnovy teorii veyvletov* (Fundamentals of the theory of wavelets), *Veyvlet in MATLAB*, Moscow, DMKA press, 2008, 448 p. (in Russian).
16. **Csorgo M., Horvath L.** Limit Theorems in Change-point analysis, New York, Wiley, 1997, 438 p.
17. **Poligrad M., Shao Q. M.** Estimation of the Variance of partil Sums for ρ Mixing random variables, *Journal of Multivariate Analysis,* 1995, vol. 52, pp. 140—157.
18. **Poligrad M., Shao Q. M.** Self-normalizing central limit teorem for suns of weakly dependent random variables, *J. Theoret. Probab.,* 1994, vol. 7, pp. 309—338.
19. **Shiryaev A. N.** *Veroyatnost'* (Probability), Moscow, Science, 1989, 640 p. (in Russian).
20. **Fan J., Yao Q.** Effiant estimation of conditional variance functions in stoachastic regression, *Biometrika*, 1998, vol. 85, pp. 645—660.
21. **Xia Y., Tong H., Li W. K.** Abstitude deviation estimation of volatility in nonparametric models, Research report No. 177. Departament of Statistics and actuarial science, The University of Hong Kong, Hong Kong, 1988, pp. 133—173.
22. **Choin J. M., Muller H. G.** Nonparametrics quasilikelihood, *The annual of statistics*, 1999, vol. 27, pp. 36—64.