

ОЦЕНИВАНИЕ ПАРАМЕТРА САМОПОДОБИЯ ДЛЯ СТАЦИОНАРНЫХ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ

Людмила Кириченко, Тамара Радивилова

На практике оценивание показателя Херста является достаточно сложной задачей, поскольку в реальных условиях для такого оценивания необходимо использовать конечные наборы данных, механизм генерации которых зачастую неизвестен. В данной статье приведены результаты сравнительного анализа статистических характеристик оценок, полученных разными методами по стационарным модельным данным малой длины.

В общем случае оценку показателя Херста можно представить следующим образом:

$$\hat{H}_{meth} = f(method, parameters, N),$$

где *method* – метод оценивания показателя Херста; *parameters* – параметры данного метода; *N* – длина временного ряда.

2.1 Метод нормированного размаха

Данный метод был предложен Г.Херстом для решения задач, связанных с накоплением водных ресурсов [Hurst, 1951; Hurst, 1955] и получил дальнейшее развитие в работах Б.Мандельброта [Mandelbrot, 1969; Mandelbrot, 1972; Mandelbrot, 1979], В.Феллера [Feller, 1951], Е.Федера [Feder, 1988], Э.Петерса [Петерс, 2004]. В настоящее время метод нормированного размаха является самым известным и популярным методом фрактального анализа [Clegg, 2006; Willinger 1996]. Он широко используется в телекоммуникационных технологиях, при исследовании самоподобных свойств информационных трафиков [Шелухин, 2008;

Шелухин, 2011], в изучении динамики финансовых рынков [Петерс, 2004; Ширяев, 1998], исследованиях геофизических и радио сигналов [Feder, 1988; Доля, 2006], и т.д..

Оценивание параметра Херста методом нормированного размаха основано на том, что для случайного процесса, обладающего долгосрочной зависимостью (1.7) при больших значениях τ выполняется следующее соотношение:

$$M \left[\frac{R(\tau)}{S(\tau)} \right] = (c \cdot \tau)^H, \quad (2.1)$$

где c – некоторая постоянная величина;

$R(\tau)$ – размах кумулятивного ряда $x^{cum}(t, \tau)$;

$S(\tau)$ – среднее квадратичное отклонение исходного ряда.

Построим пошаговый алгоритм нахождения H по временному ряду.

1. Для расчета необходимо преобразовать исходный ряд $X(t)$, $t = 1, \dots, N$, путем нахождения первой разности: $x(t) = X(t+1) - X(t)$. Если исходный ряд уже представляет собой ряд приращений, данный шаг пропускается.

2. Полученный ряд делится на A смежных подпериодов длины τ , так что $A \cdot \tau = 2^N$. Обозначим каждый подпериод I_a с учетом того, что $a = 1, 2, 3, \dots, A$. Для каждого I_a длины τ определяется среднее значение:

$$\bar{x}(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} x(t).$$

3. Временной ряд накопленных отклонений от среднего значения $x^{cum}(t, \tau)$ для каждого подпериода I_a определяется как:

$$x^{cum}(t, \tau) = \sum_{i=1}^t x(i) - \bar{x}(\tau), \quad t = \overline{1, \tau}.$$

4. Для каждого подпериода I_a определяется размах кумулятивного ряда $R(\tau) = \max(x^{cum}(t, \tau)) - \min(x^{cum}(t, \tau))$, $t = \overline{1, \tau}$, и выборочное стандартное

отклонение ряда приращений $S(\tau) = \sqrt{\frac{1}{\tau-1} \sum_{t=1}^{\tau} (x(t) - \bar{x})^2}$.

5. Нормированный размах в течение каждого I_a подпериода равен $\frac{R(\tau)}{S(\tau)}$.

Среднее значение R/S для длины τ определяется как

$$\frac{R}{S}(\tau) = (1/A) \sum_{a=1}^A \frac{R_a(\tau)}{S_a(\tau)}.$$

6. Длина τ увеличивается до следующего более высокого значения. Используются значения τ , включающие начальные и конечные точки временного ряда, и шаги 1-6 повторяются до $\tau = N$.

7. Строится линейная регрессия методом наименьших квадратов на $\log(\tau)$, как независимой переменной, и $\log(R/S)$, как зависимой переменной (рис. 2.1). Тангенс угла наклона прямой является оценкой показателя Херста H .

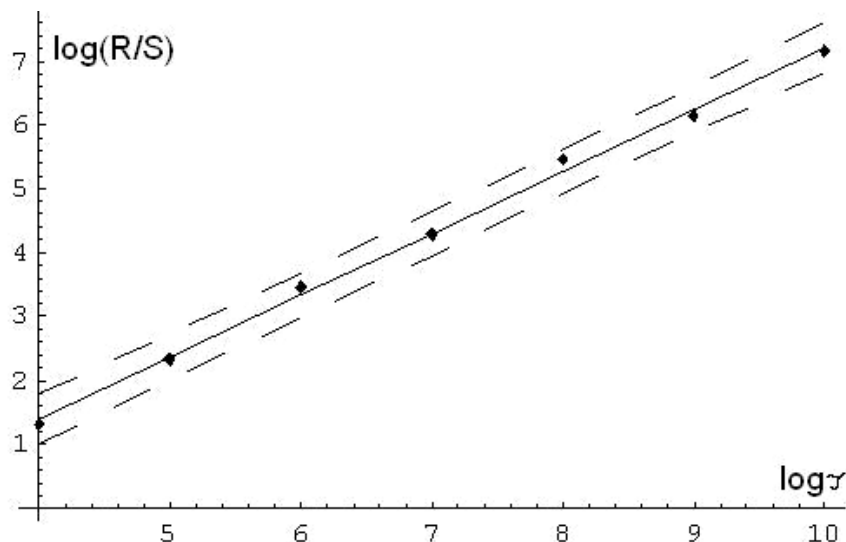


Рисунок 2.1 – Зависимость $\log(R/S)$ от $\log(\tau)$, полученная методом нормированного размаха

2.2 Метод изменения дисперсии

Метод изменения дисперсии агрегированного ряда непосредственно вытекает из определения и свойств самоподобного случайного процесса (1.4) и определения агрегированного временного ряда (1.8). Наибольшее распространение данный метод получил при анализе самоподобных свойств телекоммуникационного трафика [Шелухин, 2008; Шелухин, 2011; Willinger 1996].

Под агрегированием ряда $x(t)$ длиной τ по шкале времени с параметром

m понимается переход к такому процессу $x^{(m)}$, что $x_k^{(m)} = \frac{1}{m} \sum_{t=km-m+1}^{km} x(t)$, $k = \overline{1, \tau / m}$. Оценивание параметра Херста методом изменения дисперсии

основано на том, что для самоподобного процесса $X(t)$ дисперсия агрегированных временных серий x^m при больших значениях m подчиняется следующему соотношению:

$$Var[x^{(m)}] = \frac{Var[x]}{(c \cdot m)^\beta} \quad (2.2)$$

Параметр самоподобия $H = 1 - \beta / 2$.

Построим пошаговый алгоритм нахождения H по временному ряду.

1. Для расчета необходимо преобразовать исходный ряд $X(t)$, $t = 1, \dots, N$, путем нахождения первой разности: $x(t) = X(t+1) - X(t)$. Если исходный ряд уже представляет собой ряд приращений, данный шаг пропускается.

2. Полученный ряд агрегируется по минимальному уровню агрегации ΔT по формуле (1.8). Получается агрегированная эквидистантная реализация, содержащая m элементов, $x^{(m)} = \{x(\Delta T), x(2\Delta T), \dots, x(m\Delta T)\}$.

3. Для полученной выборки вычисляется выборочная дисперсия

$$\text{Var}[x^{(m)}] = \frac{1}{m-1} \sum_{t=1}^m (x(t) - \bar{x})^2 \quad (2.3)$$

4. Уровень агрегации увеличивается вдвое $\Delta T = 2\Delta T$. Для полученной выборки вычисляется выборочная дисперсия, и шаги 1–3 повторяются.

5. Строится линейная регрессия методом наименьших квадратов на $\log(m)$, как независимой переменной, и $\log[\text{Var}(x)]$, как зависимой переменной (рисунок 2.2). Тангенс угла наклона уравнения является

оценкой величины β , параметр Херста определяется как $H = 1 - \frac{\beta}{2}$

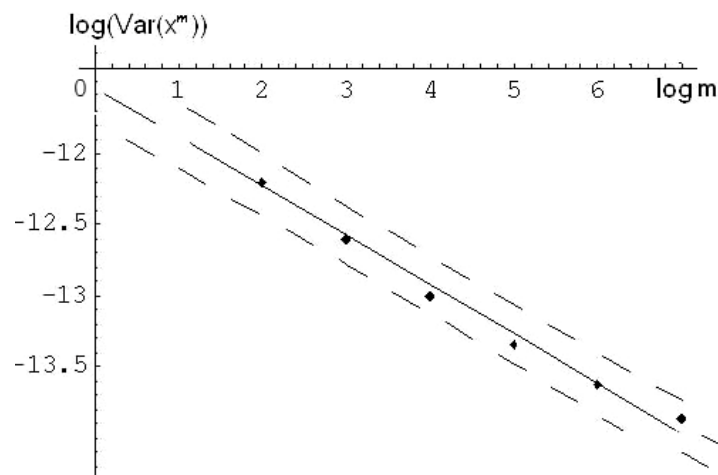


Рисунок 2.2 – Зависимость $\text{Var}[x^{(m)}]$ от $\log(m)$, полученная методом изменения дисперсии

2.3 Метод детрендированного флуктуационного анализа (ДФА)

Данный метод изначально был предложен для анализа длительных корреляций в структуре сердечного ритма [Peng, 1994; Penzel, 2007]. В настоящее время является одним из наиболее широко используемых методов исследования длительных корреляций в нестационарных

процессах [Kantelhardt, 2001; 2002; 2008]. В частности, в исследованиях динамики биологических систем по экспериментальным данным [Hoeyer, 1997; Gu, 2006], в последние годы ДФА активно используется во фрактальном анализе финансовых рядов для прогнозирования кризисных ситуаций [Cont, 2004; Ganchuk, 2006; Ramires, 2008].

Метод ДФА основан на скейлинговом соотношении флуктуационной функции $F(\tau)$ случайного процесса, обладающего долгосрочной зависимостью, и длины сегмента ряда τ :

$$F(\tau) = (c \cdot \tau)^H \quad (2.4)$$

Построим пошаговый алгоритм нахождения H по временному ряду.

1. Для расчета необходимо преобразовать исходный ряд $X(t)$, $t = 1, \dots, N$, путем нахождения первой разности: $x(t) = X(t+1) - X(t)$. Если исходный ряд уже представляет собой ряд приращений, данный шаг пропускается. Ряд $X(t)$ желательно привести к нулевому среднему: $X^*(t) = X(t) - \bar{X}(t)$.

2. Строится кумулятивный ряд $Y^{cum}(k) = \sum_{t=1}^k X^*(t)$.

3. Ряд $Y^{cum}(t)$ разбивается на N неперекрывающихся сегментов длиной τ и для каждого сегмента $Y^{cum}(t)$ вычисляется флуктуационная функция

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (Y^{cum}(t) - Y_m(t))^2, \quad (2.5)$$

где $Y_m(t)$ – локальный m -полиномиальный тренд в пределах данного сегмента.

4. Функция $F(\tau)$ усредняется по всему ряду $Y^{cum}(t)$:

$$F(\tau) = \sqrt{\frac{1}{N} \sum_{i=1}^N F_i^2(\tau)}. \quad (2.6)$$

6. Значение τ увеличивается и шаги 3–4 повторяются. Следует отметить, что максимальная длина сегмента не должна превышать значения $N/4$, иначе построение будет содержать большие погрешности [Kantelhardt, 2001; Павлов, 2007; Павлов, 2008].

7. Строится линейная регрессия методом наименьших квадратов на $\log(\tau)$, как независимой переменной, и $\log F(\tau)$, как зависимой переменной (рисунок 2.3). Тангенс угла наклона прямой является оценкой показателя Херста.

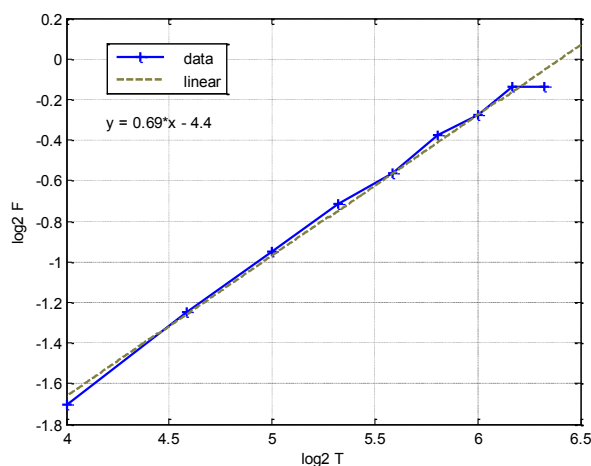


Рисунок 2.3 – Зависимость $\log F(\tau)$ от $\log(\tau)$, полученная методом ДФА

2.4 Метод с использованием дискретного вейвлет-преобразования

2.4.1 Дискретное вейвлет-преобразование одномерного сигнала

Вейвлет-преобразование одномерного сигнала – это его представление в виде обобщенного ряда или интеграла по системе базисных функций

$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$, полученных из материнского вейвлета $\psi(t)$, обладающего определенными свойствами за счет операций сдвига во времени b и изменения временного масштаба a . Дискретное вейвлет-преобразование (ДВП) строится с помощью кратномасштабного анализа, основная идея которого состоит в представлении сигнала в виде совокупности его последовательных приближений [Смоленцев, 2005; Малла, 2005].

Кратномасштабный анализ заключается в разбиении исследуемого сигнала $X(t)$ на две составляющие: аппроксимирующую и детализирующую, с последующим аналогичным дроблением аппроксимирующей до заданного уровня декомпозиции сигнала N . В результате декомпозиции сигнал $X(t)$ представляется в виде суммы аппроксимирующей компоненты $\text{approx}_N(t)$ и детализирующих компонент $\text{detail}_j(t)$:

$$X(t) = \text{approx}_N(t) + \sum_{j=1}^N \text{detail}_j(t) = \sum_{k=1}^{N_a} \text{apr}(N, k) \phi_{Nk}(t) + \sum_{j=1}^N \sum_{k=1}^{N_j} \text{det}(j, k) \psi_{jk}(t),$$

где ψ – материнский вейвлет, ϕ – соответствующая ему скейлинг-функция, N – выбранный максимальный уровень разложения,

$\text{apr}(N, k) = \int_{-\infty}^{\infty} X(t) \phi_{Nk}(t) dt$ – аппроксимирующие вейвлет-коэффициенты

уровня N , $\text{det}(j, k) = \int_{-\infty}^{\infty} X(t) \psi_{jk}(t) dt$ – детализирующие вейвлет-

коэффициенты уровня j , N_j – количество детализирующих коэффициентов на уровне j , N_a – количество аппроксимирующих коэффициентов на уровне N .

Величина вейвлет-энергии на заданном уровне вейвлет-разложения j определяется как $E_j = \frac{1}{N_j} \sum_{k=1}^{N_j} \det^2(j, k)$. Набор величин E_j для каждого уровня разложения составляет спектр вейвлет-энергии ряда. Полная вейвлет-энергия спектра представляет собой сумму энергий каждого уровня $E_{tot} = \sum_{j=1}^N E_j$. Относительная вейвлет-энергия показывает распределение энергии по уровням разложения: $p_j = \frac{E_j}{E_{tot}}$.

Одной из важных характеристик вейвлет-функции $\psi(t)$ является наличие нулевых моментов высокого порядка: $\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0$, $k = \overline{0, p-1}$. Число нулевых моментов p характеризует способность вейвлет-преобразования анализировать мелкомасштабные флуктуации и особенности высокого порядка, игнорируя регулярные полиномиальные составляющие степени $m < p$.

2.4.2. Свойства вейвлет-коэффициентов самоподобных процессов

Главным инструментом для исследования изучаемых процессов является обработка вейвлет-коэффициентов, полученных на разных масштабах. Полученный в результате ДВП ряд коэффициентов обладает определенными свойствами, позволяющими исследовать поведение стохастических процессов, обладающих свойствами самоподобия. Анализируя свойства вейвлет-коэффициентов для самоподобных процессов со стационарными приращениями, можно показать применение свойств вейвлет-коэффициентов для исследований и других процессов с долгосрочной зависимостью [Abry, 2009].

Свойство 1. Если случайный процесс $X(t)$ является самоподобным процессом со стационарными приращениями (1.5), то детализирующие

коэффициенты $d_X(j, k)$, $k = 1 \div N_j$ на каждом уровне разложения j самоподобны, что означает равенство законов распределения для ряда вейвлет-коэффициентов на каждом уровне разложения с некоторым масштабом:

$$\begin{aligned} \text{Law}\{(d_X(j, 0), d_X(j, 1), \dots, d_X(j, N_j - 1))\} = \\ = 2^{j\left(H + \frac{1}{2}\right)} \text{Law}\{(d_X(0, 0), d_X(0, 1), \dots, d_X(0, N_j - 1))\} \end{aligned} \quad (2.7)$$

где $d_X(j, k)$ – k -й детализирующий коэффициент уровня разложения j ; N_j – количество вейвлет-коэффициентов на соответствующем уровне разложения; H – показатель Херста.

Это свойство детализирующих коэффициентов вытекает из самоподобности процесса, определяемого свойством масштабирования (сжатия/растяжения) самих материнских вейвлетов. Если некоторый процесс является самоподобным, т.е. $X(2^j u) \doteq 2^{jH} X(u)$, то:

$$\begin{aligned} d_X(j, k) &= \int X(u) \psi(2^j u - k) 2^{j/2} du = \int 2^{-j/2} X(2^{-j} u) \psi(u - k) du \doteq \\ &\doteq 2^{-j\left(H + \frac{1}{2}\right)} \int X(u) \psi(u - k) du = 2^{-j\left(H + \frac{1}{2}\right)} d_X(0, k). \end{aligned}$$

Свойство 2. Вейвлет-коэффициенты, полученные в результате декомпозиции процесса со стационарными приращениями, являются стационарными на каждом масштабе 2^j . Это свойство вытекает из свойства вейвлет-функций $\int \psi(t) dt = 0$, которое гарантирует стационарность коэффициентов для процессов со стационарными приращениями.

Свойство 3. Если существуют моменты p -го порядка, то для вейвлет-коэффициентов, полученных в результате декомпозиции процесса $X(t)$, выполняется следующее равенство:

$$M|d_X(j, k)|^p = M|d_X(0, k)|^p 2^{-jp\left(H+\frac{1}{2}\right)} \quad (2.8)$$

Как следствие, для процессов с конечной дисперсией, что чаще всего встречается на практике, выражение (2.8) преобразуется в следующую форму:

$$M|d_X(j, k)|^2 = M|d_X(0, k)|^2 2^{-j(2H+1)} \quad (2.9)$$

Свойство 4. В случае, если $X(t)$ является самоподобным процессом со стационарными приращениями, корреляционная функция вейвлет-коэффициентов уровня j убывает в соответствии с соотношением:

$$M(d_X(j, k)d_X(j, k+n)) \sim n^{2(H-m)}, \quad n \rightarrow \infty,$$

где m – число нулевых моментов материнского вейвлета ψ .

Свойство 5. Для разных уровней разложения $j_1 \neq j_2$ корреляция детализирующих коэффициентов этих уровней $M(d_X(j_1, k)d_X(j_2, k+n)) = 0$ для всех n .

Свойство 6. Детализирующие коэффициенты ДВП на каждом уровне разложения j имеют нормальное распределение с нулевым математическим ожиданием.

Все перечисленные свойства вейвлет-разложения широко используются для анализа реальных временных рядов на практике.

2.4.3 Метод оценивания показателя Херста на основе дискретного вейвлет-преобразования

Метод оценивания показателя Херста на основе ДВП является относительно новым и не приобрел еще массового применения, поскольку для его корректного использования необходимо владеть математическим

аппаратом вейвлет-анализа. Основное развитие метод получил в работах французских исследователей P. Abry и P. Flandrin [Abry, 1998; Abry, 2000; Abry, 2009; Flandrin, 2002]. В настоящее время применяется при анализе самоподобного трафика [Шелухин, 2008; Шелухин, 2011].

Метод оценивания степени самоподобия H базируется на свойствах 1–6 детализирующих вейвлет-коэффициентов, полученных в результате декомпозиции временного ряда. Построим пошаговый алгоритм нахождения H по временному ряду.

1. Для исходного ряда $X(t)$, $t=1,2,\dots,n$ с помощью заданного набора вейвлет-функций выполняем вейвлет-декомпозицию с максимальным уровнем разложения $N_{\max} = \lceil \log_2 n \rceil$; где операция $\lceil \cdot \rceil$ обозначает взятие целой части.
2. На каждом уровне разложения j вычисляем сумму квадратов вейвлет-коэффициентов. Согласно формуле (2.17) показатель Херста находится из скейлингового соотношения

$$\frac{1}{N_j} \sum_{k=1}^{N_j} \det^2(j, k) \propto 2^{(2H+1)j} \quad (2.10)$$

где $\det(j, k)$ – детализирующие вейвлет-коэффициенты уровня j ; N_j – количество детализирующих коэффициентов на уровне j .

3. Каждому номеру уровня разложения j ставится в соответствие логарифм величины $\frac{1}{N_j} \sum_{k=1}^{N_j} \det^2(j, k)$. Аппроксимируя вычисленные

значения вейвлет-коэффициентов методом наименьших квадратов, получаем линейную зависимость, тангенс угла наклона соответствующей прямой определяет значение оценки показателя Херста.

На рисунке 2.4 представлен типичный спектр вейвлет-энергии самоподобного процесса (вверху) и зависимость $\log_2(E_j)$ от номера уровня разложения j (внизу).

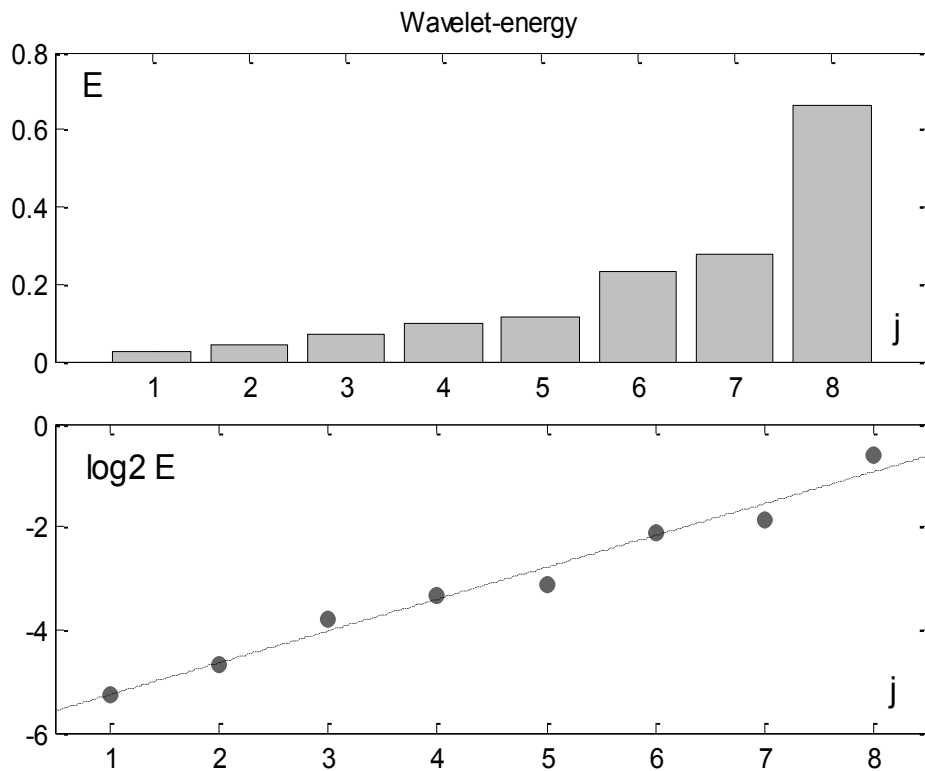


Рисунок 2.4 – Спектр вейвлет-энергии самоподобного процесса (вверху) и зависимость $\log_2(E_j)$ (внизу)

В виду нелинейности, полученной на этапе вычисления линии регрессии, которая образуется из-за функции логарифмирования, происходит смещение математического ожидания оценки показателя Херста. Свойства 5–6 детализирующих коэффициентов вейвлет-разложения, заключающиеся в некоррелируемости и нормальности, позволяют определить величину смещения аналитически и ввести поправочные коэффициенты для каждого уровня детализации j [Abry, 2009]:

$$w_j = (n \log^2 2) / 2^{j+1} \quad (2.11)$$

где w_j – поправочный коэффициент j -го уровня детализации;

n – число значений исследуемого временного ряда.

Тогда, следуя методу наименьших квадратов при линейной аппроксимации, получаем следующее:

$$\sum (w_j y_j - (b_0 + b_1 \cdot j))^2 \rightarrow \min,$$

где j – значение уровня детализации; y_j – логарифм величины

$\frac{1}{N_j} \sum_{k=1}^{N_j} \det^2(j, k)$ j -го уровня; b_0, b_1 – соответствующие коэффициенты в

уравнении аппроксимирующей прямой $b_0 + b_1 x_j$; w_j – поправочный коэффициент j -го уровня детализации. Несмещенная оценка показателя Херста определяется по следующей зависимости:

$$\hat{H}(j_1, j_2) \equiv \frac{1}{2} \left[\frac{\sum_{j=j_1}^{j_2} w_j j E_j - \sum_{j=j_1}^{j_2} w_j j \sum_{j=j_1}^{j_2} w_j E_j}{\sum_{j=j_1}^{j_2} w_j \sum_{j=j_1}^{j_2} w_j j^2 - \left(\sum_{j=j_1}^{j_2} w_j j \right)^2} + 1 \right] \quad (2.12)$$

где j_1, j_2 – уровни разложения, для которых рассчитывается линейная аппроксимация; w_j – поправочные коэффициенты; E_j – средние значения квадратов модулей вейвлет-коэффициентов; j – уровень вейвлет-разложения.

2.5 Статистические характеристики оценок показателя Херста

В данном подразделе приводятся результаты сравнительного анализа статистических характеристик оценок показателя Херста, полученных вышеописанными методами, по модельным фрактальным временным рядам небольшой длины [Kirichenko, 2010].

2.5.1 Описание численного эксперимента

В эксперименте моделировалось фрактальное броуновское движение с заданным показателем H . При этом значения H изменялись во всем возможном диапазоне $0 < H < 1$. Длина реализаций была выбрана равной от 256 до 4096 отсчетов.

Для каждого теоретического значения показателя Херста H были получены массивы оценок и исследованы их статистические характеристики, полученные по множеству различных реализаций. Для каждого сгенерированного временного ряда рассчитывалась оценки методом R/S -анализа ($\hat{H}_{R/S}$), изменения дисперсии агрегированного ряда ($\hat{H}_d$), детерендированного флуктуационного анализа ($\hat{H}_{fa}$) и вейвлет-оцениванием с помощью ДВП ($\hat{H}_w$). Показателями, по которым проводилось сравнение, были:

- среднее значение оценок показателя Херста $\bar{H}_{R/S}$, $\bar{H}_d$, $\bar{H}_{fa}$ и $\bar{H}_w$;
- среднее квадратическое отклонение оценок $S_{R/S}$, S_d , S_{fa} и S_w ;
- коэффициент корреляции между оценками, полученными разными методами;
- выборочный закон распределения оценок показателя Херста.

2.5.2 Сравнительный анализ средних значений и среднеквадратических отклонений оценок параметра Херста

На рисунке 2.5 показана зависимость средних значений оценок параметра Херста, полученных всеми исследуемыми методами, от его теоретического значения. По оси абсцисс отложены задаваемые теоретические значения показателя Херста, а по оси ординат – средние значения соответствующих оценок. Модельные реализации содержали 1024 отсчетов. Сплошная жирная прямая линия соответствует теоретическим значениям H . Очевидно, что средние значения оценок имеют смещение, которое зависит от теоретического значения показателя Херста. Оценки показателя Херста смещены, как в области персистентности, так и для антиперсистентного диапазона. Поскольку подавляющее большинство фрактальных процессов имеют долгосрочную зависимость, дальнейшие результаты, рассмотренные нами, будут приведены только для диапазона $0,5 < H < 1$. Из графика, представленного на рисунке 2.5, видно, что более смещенными являются оценки, полученные методами R/S -анализа и агрегированной дисперсии.

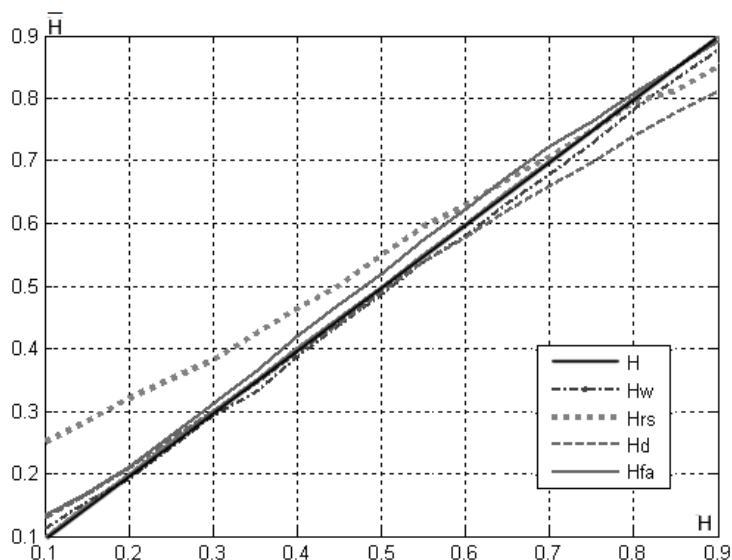


Рисунок 2.5 – Зависимости средних значений оценок показателя Херста, полученных разными методами, от теоретических значений H

Рассмотрим результаты оценивания параметра H методом R/S - анализа. На рисунке 2.6 (а) представлены зависимости средних значений оценок $\hat{H}_{rs}$, от теоретических значений показателя H для модельных рядов разной длины. Очевидно, что средние значения оценок достаточно хорошо можно аппроксимировать прямыми линиями $\hat{H}_N = k_N H + b_N$, где коэффициенты k_N и b_N зависят от длины реализации N , по которой проводится оценивание. Эти прямые пересекают линию теоретических значений параметра H в районе $H \approx 0,75$; и до этого значения оценки являются завышенными, а после – заниженными. Результаты проведенных исследований подтверждают данные, полученные при анализе оценивания R/S -методом по фрактальным реализациям других моделей, представленных в [Федер, 1991; Шелухин, 2005; Шелухин, 2011]. С возрастанием длины реализации N угол наклона аппроксимированной прямой k_N медленно увеличивается и стремится к теоретическому углу наклона $\pi/4$.

На рисунке 2.6 (б) представлены зависимости средних значений оценок $\hat{H}_d$, полученных методом изменения дисперсии. Эти зависимости также можно аппроксимировать прямыми $\hat{H}_N = k_N H + b_N$, где коэффициенты k_N и b_N зависят от длины реализации N . В данном случае аппроксимирующие прямые пересекают линию теоретических значений в районе $H \approx 0,5$ и, фактически, на всем диапазоне персистентности значения оценки показателя Херста являются заниженными. Смещение увеличивается с ростом показателя Херста, особенно для значений $H > 0,9$. Можно отметить, что смещения оценок $\hat{H}_d$ больше, чем соответствующие смещения $\hat{H}_{rs}$. С возрастанием длины реализации N смещение очень медленно уменьшается.

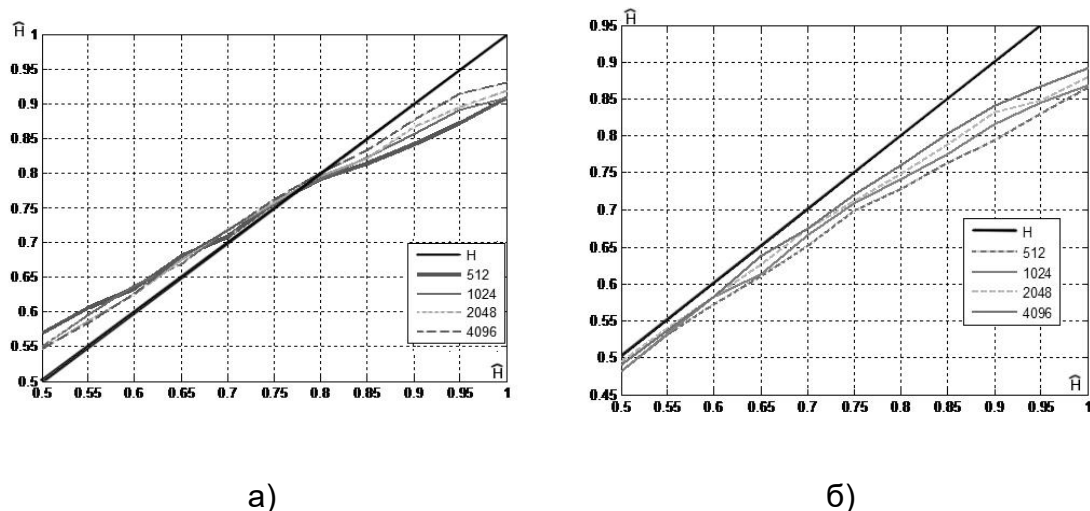


Рисунок 2.6 – Зависимости средних значений оценок от теоретического значения H :

а) метод нормированного размаха; б) метод изменения дисперсии

Метод изменения дисперсии агрегированного ряда в связи с простотой и прозрачной интерпретацией наиболее часто используется при оценивании самоподобия трафика в информационных сетях. Однако для процессов с долгосрочной зависимостью этот метод дает заниженные оценки [Шелухин, 2005; Шелухин, 2008; Шелухин, 2011]. Это может оказаться фатальным, например, при оценивании нагрузки сети при прохождении самоподобного трафика [Столинс, 2003].

Метод ДФА широко используется при анализе биоэлектрических сигналов. Для оценок $\hat{H}_{fa}$, полученных методом ДФА, характерно очень небольшое смещение (см. рисунок 2.7(а)) даже для реализаций небольшой длины. Это смещение меняет знак и увеличивается для значений $H > 0.9$. Стоит отметить, что большинство природных и информационных фрактальных процессов имеют степень самоподобия, меньшую 0.9.

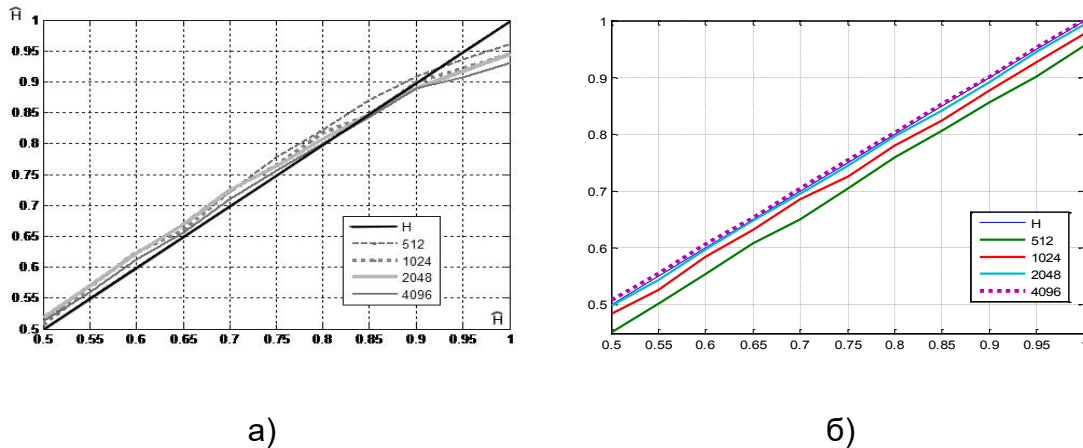


Рисунок 2.7 – Зависимости средних значений оценок от теоретического значения H :

а) метод флуктуационного анализа; б) метод вейвлет-оценивания

Вейвлет-оценивание проводилось на основе материнского вейвлета db4. Как показано в работе [Кириченко, 2009], использование этой функции дает наименьшее смещение и среднеквадратичное отклонение при оценивании показателя Херста. При сравнении основных методов оценивания параметра Херста с методом, основанным на ДВП, очевидны его преимущества по сравнению с другими методами оценки. При правильном выборе базовых вейвлетов оценки $\hat{H}_w$ являются асимптотически несмещенными, что демонстрирует рисунок 2.7 (б). При увеличении длины ряда N смещение уменьшается, и уже при $N \approx 4000$ отсчетов оно достаточно близко к 0.

Для количественного оценивания смещения оценки $\hat{H}_{meth}$, полученной с помощью определенного метода, от теоретического значения показателя H на всем диапазоне значений, была рассчитана величина смещения Δ^{meth} , которая рассчитывалась по формуле:

$$\Delta^{meth} = \frac{1}{k} \sum_{i=1}^k (\bar{H}_i^{meth} - H_i), \quad (2.13)$$

где $\bar{H}_i^{meth}$ – среднее значение оценки $\hat{H}_{meth}$, полученной для реализаций с теоретическим показателем H_i с помощью заданного метода расчета показателя Херста.

Для каждого значения H были получены массивы оценок показателя Херста $H_{R/S}$, H_d , H_{fa} и H_w , и рассчитаны средние значения оценок $\bar{H}_{R/S}$, $\bar{H}_d$, $\bar{H}_{fa}$ и $\bar{H}_w$, полученные по множеству различных реализаций.

На рисунке 2.8 представлена столбцовая диаграмма распределения смещения оценок показателя Херста для разных методов в зависимости от длины реализации. Очевидно, что наименьшую погрешность при различной длине временной реализации имеет метод, основанный на дискретном вейвлет-преобразовании.

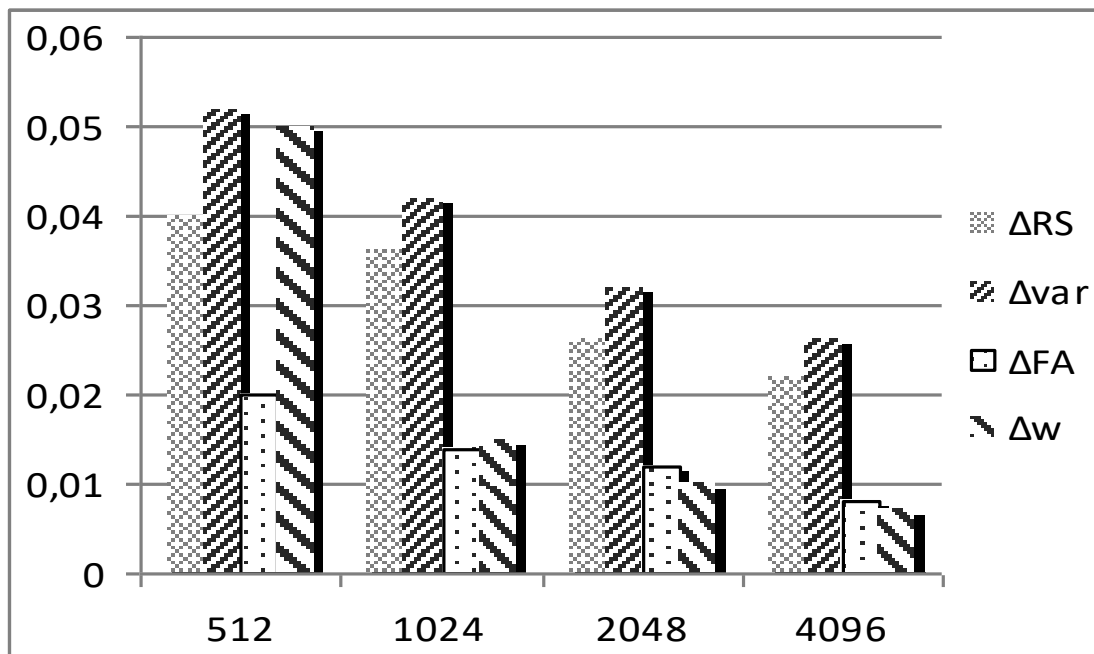


Рисунок 2.8 – Распределение смещения оценок показателя Херста для разных методов

В работах [Кириченко, 2009; Kirichenko, 2010] исследована зависимость средних квадратических отклонений оценок показателя Херста от заданных значений N и длины модельного фрактального ряда для каждого метода. Рисунок 2.9 иллюстрирует, как уменьшаются средние квадратические отклонения, полученные при расчете $\hat{H}$, при увеличении длины ряда. В данном случае теоретический показатель Херста $H = 0.8$.

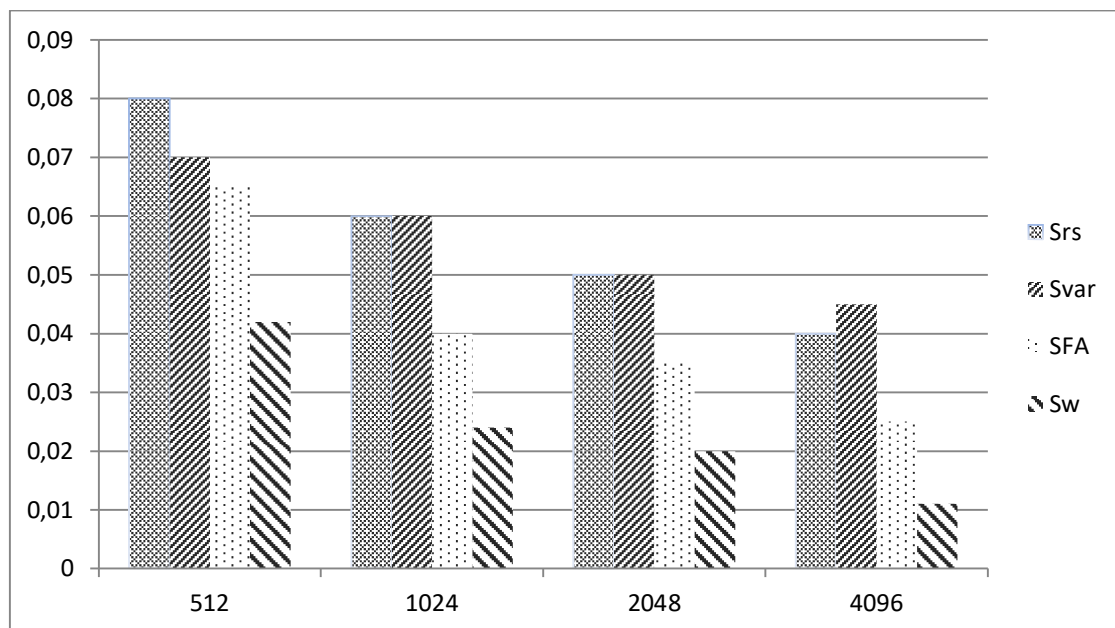


Рисунок 2.9 – Средние квадратические отклонения для разных методов

Исходя из данных столбцовой диаграммы очевидно, что оценки, полученные вейвлет-методом, имеют значительно меньший разброс, чем оценки, полученные другими основными методами оценивания показателя Херста.

2.5.3 Сравнительный анализ выборочных законов распределения и корреляции оценок параметра Херста

Вопрос о законе распределения оценок. был рассмотрен в ряде работ. где было показано, численно или аналитически, что оценки являются

нормальными для конкретного метода или конкретных значений показателя Херста. Аналитические и численные исследования, проведенные в работах [Feller, 1951; Mandelbrot, 1969; Mandelbrot, 1972; Mandelbrot, 1979], показали, что для случая $H=0,5$ оценка показателя Херста, полученная методом нормированного размаха, представляет собой случайную величину с нормальным распределением. В работах [Петерс, 2004] численно показано, что выборочный закон распределения $\hat{H}_{R/S}$ является нормальным. Вопрос о законе распределения оценок $\hat{H}_w$, полученных методом, основанным на ДВП, рассмотрен в ряде работ [Abry, 1998; Abry, 2009], где показано аналитически, что оценки являются нормально распределенными. Нормальное распределение оценок $\hat{H}_d$, полученных методом изменения дисперсии, численно продемонстрировано в работе [Кириченко, 2005]. Свойства оценок, в том числе выборочное нормальное распределение для некоторых значений показателя Херста, рассмотрены в работах [Kantelhardt, 2001; Kantelhardt, 2002; Kantelhardt, 2008].

В работах [Kirichenko, 2010; Кириченко, 2009] были исследованы законы распределения оценок $\hat{H}$ для каждого метода и разных значений параметра Херста. Для всех рассмотренных методов была выдвинута гипотеза о нормальном распределении выборочных значений оценок с параметрами $N(\bar{H}, S_{\hat{H}})$. Практически для всех выборочных данных гипотеза была принята с уровнем значимости $\alpha = 0.05$ по нескольким критериям согласия.

Поскольку выборочное распределение оценок показателя Херста является нормальным, то оценка показателя Херста H для исследуемого временного ряда длины N может быть представлена интервалом значений, внутри которого с заданной вероятностью находится истинное значение оцениваемого показателя H . Таким образом, истинное значение показателя Херста попадает в следующий доверительный интервал:

$$\hat{H} + \Delta - t_{\alpha} S < H < \hat{H} + \Delta + t_{\alpha} S, \quad (2.14)$$

где N – длина исследуемого временного ряда; $method$ – выбранный метод оценивания; $\hat{H} = \hat{H}(N, method)$ – полученное значение оценки показателя Херста по реализации длины N ; $\Delta = \Delta(N, method)$ – величина смещения оценки; $S = S(N, method)$ – среднеквадратическое отклонение, рассчитанное по модельным реализациям длины N ; α – требуемый уровень значимости; t_{α} – квантиль стандартного нормального распределения.

Такое интервальное оценивание особенно необходимо при малой длине временной реализации, когда точечная оценка малонадежна.

В работах [Kirichenko, 2010; Кириченко, 2009] был исследован вопрос о корреляционной связи между оценками, полученными разными методами. Численный анализ корреляционной зависимости между оценками параметра Херста показал, что для всех рассмотренных случаев, кроме корреляции с оценками, полученными с помощью вейвлет-преобразования, выборочные коэффициенты корреляции оказались значимыми и лежащими в диапазоне абсолютных значений в основном меньшем 0.5 (таблица 2.1).

Таким образом, исходя из результатов корреляционного, оценки показателя Херста, полученные с помощью разных методов, обладают слабой корреляционной зависимостью, коэффициенты корреляции находятся в диапазоне $[0, 0.5]$.

Таблица 2.1 – Выборочные коэффициенты корреляции для разных методов оценки показателя Херста

Методы, которыми получены оценки показателя Херста	Выборочный коэффициент корреляции r
R/S -анализ и изменение дисперсии	$0.3 \leq r \leq 0.6$
ДФА и R/S -анализ	$0.1 \leq r \leq 0.4$
ДФА и изменение дисперсии	$0.1 \leq r \leq 0.3$
Вейвлет-анализ и остальные методы	Является незначимым

Для увеличения точности оценки показателя Херста можно использовать среднее арифметическое оценок, полученных с помощью нескольких методов оценивания.

2.5.4 ВРЕМЕННЫЕ ЗАТРАТЫ ВЫПОЛНЕНИЯ МЕТОДОВ ОЦЕНИВАНИЯ

Сравнительная оценка эффективности рассмотренных в этой статьи методов производилась экспериментальным путем на основании оценок времени решения контрольных задач. В ходе исследования моделировалось фрактальное броуновское движение с заданным показателем Херста H .

Для каждого сгенерированного ряда соответствующим методом определялось время решения задачи в цикле. Временные затраты исследуемых методов для модельных реализаций длиной 4096 отсчетов для цикла числом 100 представлены в таблице 2.2.

Таблица 2.2 – Временные затраты методов оценивания показателя Херста

Метод	RS-анализ	Изменение дисперсии	ДФА	ДВП
Время	9.7969	10.6081	138.6693	3.6036

Данные показывают, что метод, основанный на применении ДВП, является самым быстродействующим. Это связано с тем, что в нем использован алгоритм быстрого вейвлет-преобразования. Наибольшее время счета алгоритма ДФА во многом связано с тем, что размер сегмента изменяется не как степень двойки, что используется в методе нормированного размаха и изменения дисперсии, а выбирается некоторое произвольное фиксированное значение. Увеличив шаг изменения длины сегмента, можно добиться существенного снижения времени счета, однако это влечет за собой снижение точности оценивания.

С увеличением длины модельной реализации время оценивания показателя Херста разными методами увеличивается. На рисунке 2.9 показано изменение времени расчета для разных методов в зависимости от длины фрактального временного ряда. Как видно из представленного графика для метода оценивания показателя Херста с помощью ДФА-метода время счета с увеличением длины реализации резко возрастает. Это связано с вышеуказанными причинами.

Используемый алгоритм быстрого вейвлет-преобразования позволяет с минимальными затратами по времени и объему информации применять метод фрактального анализа в различных технических и информационных системах в режиме реального времени.

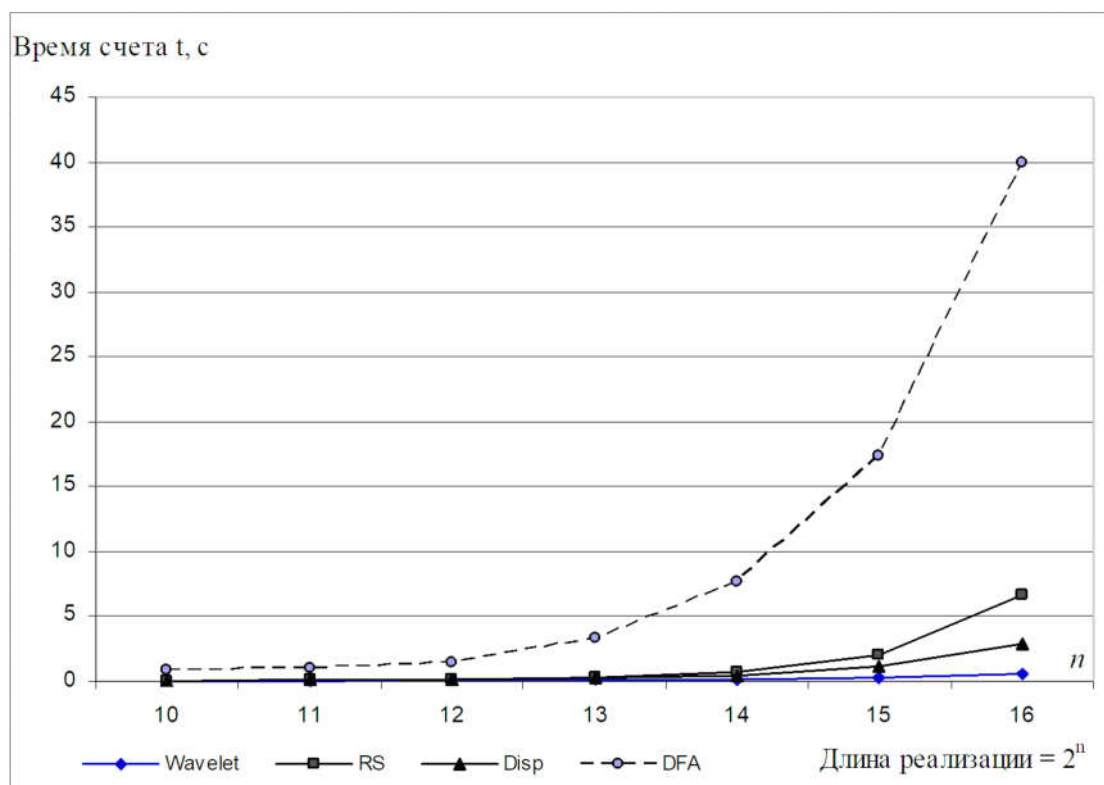


Рисунок 2.9 – Изменение временных затрат для методов оценки показателя Херста при увеличении длины выборки

Выводы

1. Приведены пошаговые алгоритмы основных методов оценивания показателя Херста: метод нормированного размаха, изменения дисперсии агрегированного ряда, детрендированного флуктуационного анализа и вейвлет-оценивания.
2. Проведен сравнительный анализ статистических характеристик оценок показателя Херста, полученных разными методами оценивания. Результаты исследования показали, что оценки являются смещенными. Для каждого метода смещение зависит от истинного значения степени самоподобия процесса и длины временного ряда. Минимальное смещение имеют оценки, полученные методом флуктуационного анализа и с помощью вейвлет-преобразования. Средние квадратические отклонения

оценок зависят от метода оценивания и уменьшаются с ростом длины ряда. Минимальные средние квадратические отклонения имеют вейвлет-оценки.

3. Результаты численного исследования показали, что оценки показателя Херста, полученные рассмотренными методами оценивания, имеют нормальное распределение и могут быть представлены доверительным интервалом значений, характеристики которого учитывают длину исследуемого временного ряда и метод оценивания.

4. Численный анализ корреляционной зависимости между вейвлет-оценками параметра Херста показал, что оценки показателя Херста, полученные с помощью разных методов, обладают слабой корреляционной зависимостью, а вейвлет-оценки не коррелируют с другими. Таким образом, для увеличения точности оценки показателя Херста можно использовать среднее арифметическое оценок, полученных с помощью нескольких методов оценивания.

5. Сравнительное оценивание быстродействия рассмотренных в этой статьи методов показало, что метод, основанный на применении вейвлет-преобразования, является самым быстродействующим.

Литература

[Abry, 1998] Abry, P., Veitch, D. Wavelet analysis of long-range dependent traffic. IEEE/ACM Transactions Information Theory, Vol. № 1 (44), 1998. pp. 2–15.

[Abry, 2000] Abry, P., Flandrin, P., Taqqu, M. S., Veitch, D. Self-Similar Network Traffic and Performance Evaluation. New-York: John Wiley & Sons, 2000. pp. 39–88.

[Abry, 2009] Abry, P., Goncalves P., Vehel, J. L. Scaling, Fractals and Wavelets. London: John Wiley & Sons, 2009. – 464 p.

- [Clegg, 2006] Clegg, R. G. A practical guide to measuring the hurst parameter. International Journal of Simulation. Systems, Science & Technology, 2006, Vol. 7, № 2. pp. 3–14.
- [Cont, 2004] Cont, R., Tankov, P. Financial modelling with jump processes. Chapman & Hall / CRC Press, 2004. - 527 p.
- [Feller, 1951] Feller, W. The asymptotic distribution of the range of sums of independence variables. Annals of mathematics and statistics, France, University of Nancy, 1951. – 306 p.
- [Feder, 1988] Feder, J. Fractals. Plenum, New York, 1988.
- [Flandrin, 2002] Flandrin P. Wavelet Tools for Scaling Processes. Turbulence: measurements and signals: Lecture given at the Summer School, Lyon, 2002. pp. 123–128.
- [Ganchuk, 2006] Ganchuk, A., Derbentsev, V., Soloviev, V. Multifractal Properties of the Ukraine Stock Market. [Электронный ресурс]: 2006. – Режим доступа: <http://arxiv.org/pdf/physics/0608009.pdf>.
- [Gu, 2006] Gu, G. F., Zhou, W. X. Detrended fluctuation analysis for fractals and multifractals in higher dimensions. Phys. Rev. E 74, 2006. pp.61-104.
- [Hoyer, 1997] Hoyer, D., Schmidt, K., Bauer, R., Zwiener, U., Kohler, M., Luthke, B., Eiselt, M. Nonlinear analysis of heart rate and respiratory dynamics. IEEE Engineering in Medicine and Biology Magazine, 1997, № 16. pp. 31–39.
- [Hurst, 1951] Hurst, H. E. The Long-Term Storage Capacity of Reservoirs. Transactions of the American Society of Civil Engineers, 1951, Vol. 116. pp. 770–799.
- [Hurst, 1955] Hurst, H. E. Methods of using long-term storage in reservoirs. Proceedings of the Institution of Civil Engineers, London, 1955, Part 1. pp. 519–577.
- [Kantelhardt, 2001] Kantelhardt, J. W., Koscielny-Bunde, E., Havlin, S. and et al. Detecting long-range correlations with detrended fluctuation analysis.

-
- Physica A., Statistical Mechanics and its Applications, 2001, Vol. 295, № 3–4. pp. 441–454.
- [Kantelhardt, 2002] Kantelhardt, J.W., Zschiegner, S.A., Bunde, A., Havlin, S., Koscielny-Bunde, E., Stanley, H.E. Multifractal detrended fluctuation analysis of non-stationary time series. Physica A., 2002, № 316. pp. 87–114.
- [Kantelhardt, 2008] Kantelhardt, J.W. Fractal and Multifractal Time Series. [Электронный ресурс]: 2008.– Режим доступа: <http://arxiv.org/abs/0804.0747>.
- [Kirichenko, 2010] Kirichenko, L., Radivilova, T. Comparative analysis of statistical properties of the Hurst exponent estimates obtained by different methods. Information Models of Knowledge / ed. K. Markov, V. Velychko, O. Voloshin. – Kiev–Sofia : ITHEA, 2010. pp. 451–459
- [Mandelbrot, 1969] Mandelbrot, B., Wallis, J. R. Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long-Run Statistical Dependence. Water Resources Research, 1969, Vol. 5. pp. 967–988.
- [Mandelbrot, 1972] Mandelbrot, B. Statistical methodology for non-periodic cycles: from the covariance to R/S analysis. Annals of economic social measurement, № 1, 1972. pp. 2–11.
- [Mandelbrot, 1979] Mandelbrot, B., Taqqu, M. S. Robust R/S analysis of long-run serial correlation. Bulletin of the International Statistical Institute. Proceedings of the 42nd Session of the International Statistical Institute, Manila. Book 2, Vol.48, 1979. pp. 69–104.
- [Peng, 1996] Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., Goldberger, A.L. Mosaic organization of DNA nucleotides. Phys. Rev. E., Vol. 49, 1994. pp. 1685-1689.
- [Penzel, 2007] Penzel, T., Kantelhardt, J.W., Becker, H.F., Peter, J.H., Bunde, A. Detrended Fluctuation Analysis and Spectral Analysis of Heart Rate Variability for Sleep Stage and Sleep Apnea Identification. Computers in Cardiology, 2007. pp. 307–310.
- [Ramires, 2008] Ramires, D., Ceballos, D., Cortez, K., Rodrigues, M. Financial crisis: Theory and Practice. Based on Workshop IAFI at Financial Crisis, 2008.

- [Willinger, 1996] Willinger, W., Taqqu, M. S., Erramilli, A.A. Bibliographical guide to self-similar traffic and performance modeling for modern high-speed network in «Stochastic networks: theory and applications». Oxford: Claredon Press, 1996. – 286 p.
- [Доля, 2006] Доля, Н. Г., Иванов, В. К., Пащенко, Р. Э., Стадник, А. М. и др. Фрактальный анализ процессов, структур и сигналов: Коллективная монография. под ред. Р.Э. Пащенко. Х. : ХООО “НЭО “ЭкоПерспектива”, 2006. – 348 с.
- [Кириченко, 2005] Кириченко, Л. О., Радивилова, Т. А., Синельникова, М. И. Сравнительный анализ методов оценки параметра Херста самоподобных процессов. Системи обробки інформації, Вип. 8(48), 2005. сс.36-40.
- [Кириченко, 2009] Кириченко, Л. О., Дейнеко, Ж. В. Оценивание самоподобия стохастического временного ряда методом вейвлет-анализа. Радіоелектронні і комп'ютерні системи, № 4 (38), 2009. сс. 99–105.
- [Малла, 2005] Малла, С. Вэйвлеты в обработке сигналов. М.: Мир, 2005. – 671 с.
- [Петерс, 2004] Петерс, Э. Фрактальный анализ финансовых рынков. Применение теории хаоса в инвестициях и экономике : пер. с англ. М : Интернет-трейдинг, 2004. – 304 с.
- [Смоленцев, 2005] Смоленцев, Н.К. Основы теории фракталов. Вейвлеты в MATLAB. М.: ДМК Пресс, 2005. – 304 с.
- [Столингс, 2003] Столингс В. Современный компьютерные сети. СПб.: Питер, 2003. – 783 с.
- [Шелухин, 2005] Шелухин, О. И., Тенякшев, А. В., Осин, А. В. Моделирование информационных систем. М. : Сайнс-Пресс, 2005. – 368 с.

- [Шелухин, 2008] Шелухин, О. И., Осин, А. В., Смольский, С. М. Самоподобие и фракталы. Телекоммуникационные приложения. М. : Физматлит, 2008. – 368 с.
- [Шелухин, 2011] Шелухин, О. И. Мультифракталы. Инфокоммуникационные приложения. М. : Горячая линия – Телеком, 2011. – 576 с.
- [Ширяев, 1998] Ширяев, А. Н. Основы стохастической финансовой математики. М. : Фазис, Т. 1: Факты. Модели, 1998. –512 с.

Информация об авторах



Людмила Кириченко – д.т.н., профессор Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: lyudmyla.kirichenko@nure.ua.

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, фрактальный анализ, вейвлет-анализ, детерминированные хаотические системы.



Тамара Радивилова – к.т.н., доцент Харьковского национального университета радиоэлектроники; пр. Науки 14, 61166, Харьков, Украина; e-mail: tamara.radivilova@gmail.com.

Основные области научных исследований: самоподобные и мультифрактальные временные ряды, телекоммуникационные системы, управление трафиком, информационная безопасность.