

Л.О. КИРИЧЕНКО, канд. техн. наук, доц. ХНУРЭ (г. Харьков)

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТАТИСТИЧЕСКИХ СВОЙСТВ ОЦЕНОК ПОКАЗАТЕЛЯ ХЕРСТА

Проведен численный анализ статистических характеристик оценок показателя Херста для самоподобных временных рядов, полученных различными методами. Показано, что оценки имеют нормальное распределение, средние значения которого значительно смещены относительно оцениваемого параметра. Средние значения и средние квадратические отклонения оценок зависят от значения параметра и метода оценивания. Ил.: 1. Табл.: 3. Библиогр.: 12 назв.

Ключевые слова: показатель Херста, оценка показателя Херста, самоподобный временной ряд.

Постановка проблемы и анализ литературы. Задачи современной нелинейной физики, радиоэлектроники, теории управления, обработки изображений, требуют разработки и применения новых математических моделей, методов и алгоритмического обеспечения анализа данных. В настоящее время стало общепризнанным, что многие стохастические процессы в природе и технике обладают долгосрочной зависимостью и фрактальной структурой. Наиболее адекватным математическим аппаратом для исследования динамики и структуры таких рядов является фрактальный анализ, особое значение которого состоит в том, что он учитывает поведение системы не только в период измерений, но и его предьсторию.

Стохастический процесс $X(t)$ является статистически самоподобным, если процесс $a^{-H}X(at)$ обладает теми же статистическими характеристиками второго порядка, что и $X(t)$. Долгосрочная зависимость означает медленное (гиперболическое) убывание во времени автокорреляционной функции случайного процесса. Параметр H , называемый параметром Херста, представляет собой меру самоподобия или меру длительности долгосрочной зависимости стохастического процесса. Значение $H = 0,5$ указывает на отсутствие долгосрочной зависимости. Чем ближе значение H к 1, тем выше степень устойчивости долгосрочной зависимости.

При значениях $0,5 < H < 1$ временной ряд демонстрирует персистентное (трендоустойчивое) поведение. Т.е., если ряд возрастает (убывает) в предыдущий период, то с вероятностью, тем большей, чем показатель Херста больше 0,5, он будет сохранять эту тенденцию такое же время в будущем. Значение $H = 0,5$ указывает на независимость (отсутствие какой-либо памяти о прошлом) значений временного ряда. Чем ближе H к 0,5, тем более зашумлен ряд и тем менее выражен его тренд. Диапазон $0 < H < 0,5$ соответствует антиперсистентным рядам: если система демонстрирует рост в предыдущий период, то с вероятностью, тем большей, чем показатель Херста меньше 0,5, в следующем периоде начнется спад.

Таким образом очевидно, что оценивание показателя Херста по экспериментальным данным играет важнейшую роль в изучении процессов, обладающих свойствами самоподобия. Для оценки параметра Херста по временному ряду существует множество методов, достаточно полный обзор которых представлен в [1, 2]. Все рассмотренные методы обладают определенными недостатками. В частности, все оценки показателя Херста являются смещенными случайными величинами (см., например [3 – 5]). В некоторых работах рассмотрен вопрос о законе распределения оценок H и показано, численно или аналитически, что они являются нормальными для конкретного метода или конкретных значений показателя Херста [3, 5, 6]. Однако, в настоящее время практически нет обзорного исследования, где были бы обобщены результаты оценивания показателя Херста H разными методами и дан сравнительный анализ статистических свойств оценок. Представленная работа является попыткой такого исследования для наиболее популярных методов оценивания.

Целью данной работы является вычисление показателя Херста H по модельным фрактальным временным рядам и сравнительный анализ статистических характеристик оценок, полученных различными методами. В работе исследованы следующие методы определения параметра Херста: R/S -анализ (метод нормированного размаха) [3, 6, 7], изменение во времени дисперсии агрегированного ряда [4, 7], расчет с помощью вейвлет-анализа [5, 8, 9] и определение индекса фрактальности [10]. В качестве модельного случайного процесса, обладающего фрактальными свойствами, был выбран процесс фрактального броуновского движения.

Метод нормированного размаха. При исследовании R/S -статистики временного ряда $x(t)$ длиной τ определяется отношение $\frac{R(\tau)}{S(\tau)}$, где $R(\tau)$ – размах кумулятивного ряда $x^{\text{cum}}(t, \tau)$, $S(\tau)$ – среднее квадратическое отклонение исходного ряда:

$$R/S = \frac{\max(x^{\text{cum}}(t, \tau)) - \min(x^{\text{cum}}(t, \tau))}{\sqrt{\frac{1}{\tau-1} \sum_{i=1}^{\tau} (x(i) - \bar{x})^2}}, \quad t = \overline{1, \tau}, \quad (1)$$

где $\bar{x}(\tau) = \frac{1}{\tau} \sum_{i=1}^{\tau} x(i)$, $x^{\text{cum}}(t, \tau) = \sum_{i=1}^t x(i) - \bar{x}(\tau)$.

Для самоподобного процесса это отношение при больших значениях τ обладает следующей характеристикой:

$$M \left[\frac{R}{S} \right] \sim (c \cdot \tau)^H, \quad (2)$$

где c – некоторая постоянная величина.

График зависимости $\frac{R(\tau)}{S(\tau)}$ от τ в логарифмическом масштабе будет представлять собой прямую линию, аппроксимированную методом наименьших квадратов. Оценка показателя H вычисляется как тангенс угла наклона прямой зависимости $\log \frac{R(\tau)}{S(\tau)}$ от $\log(\tau)$.

Метод изменения дисперсии агрегированного ряда. Под агрегацией по шкале времени с параметром m понимается переход к такому процессу $x^{(m)}$, что $x_k^{(m)} = \frac{1}{m} \sum_{i=km-m+1}^{km} x_i$. Для агрегированных временных серий $x^{(m)}$ самоподобного процесса дисперсия при больших значениях m подчиняется следующей формуле:

$$\text{Var}(x^{(m)}) \sim \frac{\text{Var}(x)}{m^\beta}. \quad (3)$$

В этом случае параметр самоподобия $H = 1 - \frac{\beta}{2}$ можно определить, если сгенерировать агрегированный процесс на разных уровнях агрегации m и вычислить дисперсию для каждого уровня. График зависимости $\log(\text{Var}(x^{(m)}))$ от $\log(m)$ будет представлять собой прямую линию с наклоном, равным $-\beta$.

Метод с использованием дискретного вейвлет-преобразования. В этом случае оценку показателя Херста можно найти по скорости роста средних значений квадратов модулей вейвлет-коэффициентов:

$$W(k) = \frac{1}{N(k)} \sum_{j=1}^{N(k)} |c_j^{(k)}|^2. \quad (4)$$

Здесь $c_j^{(k)}$ – коэффициенты ортогонального дискретного вейвлет-преобразования выборочного самоподобного временного ряда; k – номер уровня детальности вейвлет-разложения; $N(k)$ – число вейвлет-коэффициентов на уровне детальности k . Тогда, выполняется

$$W(k) \sim s_k^{2H+1}, \quad (5)$$

где s_k – характерный временной масштаб уровня детальности k . Таким образом, значение коэффициента наклона прямой графика зависимости $W(k)$ от s_k в логарифмическом масштабе дает оценку для величины $2H + 1$.

Метод расчета индекса фрактальности временного ряда. В данном методе вводится равномерное разбиение временного ряда на интервалы $[t_{i-1}, t_i]$

и рассматривается минимальное по площади покрытие графика ряда прямоугольниками с основанием δ и высотой $A_i(\delta)$, которая равняется разности между максимальным и минимальным значениями ряда на интервале $[t_{i-1}, t_i]$.

Для величины $V(\delta) = \sum_{i=1}^m A_i(\delta)$ при $\delta \rightarrow 0$ выполняется соотношение

$$V(\delta) \sim \delta^{-\mu}, \quad (6)$$

где $\mu = D - 1$.

Показатель μ называется индексом фрактальности и связан с показателем Херста соотношением $H = 1 - \mu$. Для численного определения H необходимо построить зависимость $\log V(\delta)$ от $\log \delta$ и определить μ как тангенс угла наклона прямой.

Фрактальное броуновское движение. В качестве случайного процесса, обладающего фрактальными свойствами, часто рассматривается фрактальное броуновское движение (ФБД), которое нашло широкое применение в физике, химии, биологии, экономике и теории сетевого трафика.

Гауссовский процесс $X(t)$ называется фрактальным броуновским движением с параметром H , $0 < H < 1$, если приращения случайного процесса $\Delta X(\tau) = X(t + \tau) - X(t)$ имеют гауссовское распределение вида

$$P(\Delta X < x) = \frac{1}{\sqrt{2\pi\sigma_0\tau^H}} \cdot \int_{-\infty}^x \text{Exp}\left[-\frac{z^2}{2\sigma_0^2\tau^{2H}}\right] dz, \quad (7)$$

где σ_0 – коэффициент диффузии.

ФБД с параметром $H = 0,5$ совпадает с классическим броуновским движением. Приращения ФБД называются фрактальным гауссовским шумом, дисперсия которого подчиняется соотношению $D[X(t + \tau) - X(t)] = s_0^2 t^{2H}$.

Существует несколько методов построения ФБД для случая дискретного времени. Одним из наиболее используемых на практике является метод последовательного случайного сложения Фосса. Метод включает в себя следующий пошаговый алгоритм [3]. Исходные значения координат $X(t_i)$ в моменты времени $t_i = 0, 1/2, 1$ равны нулю. На первом шаге к значениям координат $X(t_1), X(t_2), X(t_3)$ прибавляются случайные числа, выбранные из нормального распределения с нулевым средним и начальной дисперсией σ_1^2 .

Средние значения времени на каждом интервале рассматриваются как дополнительные узлы на оси времени; значения координат в них оцениваются интерполяцией. На следующем шаге ко всем координатам $X(t_i)$, ($t_i = 0, 1/4, 1/2, 3/4, 1$) прибавляются случайные числа с нулевым средним значением и

уменьшенной дисперсией $\sigma_2^2 = \frac{\sigma_1^2}{2^{2H}}$. На n -ом шаге алгоритма мы получаем значения реализации ФБД для $1 + 2^n$ значений времени t_i . Дисперсия

слагаемых n -го поколения равна $\sigma_n^2 = \frac{\sigma_{n-1}^2}{2^{2H}} = \frac{\sigma_0^2}{2^{2Hn}}$. Процесс, предложенный Фоссом, приводит к обобщенному броуновскому движению при любом разрешении.

Принципиально другим подходом является построение ФБД с помощью биортогональных вейвлетов [11]. В этом случае построение реализации ФБД осуществляется с помощью дискретного вейвлет-преобразования, где детализирующие вейвлет-коэффициенты на каждом уровне детальности являются независимыми нормально распределенными случайными величинами, а аппроксимирующие вейвлет-коэффициенты получены с помощью фрактального процесса авторегрессии и скользящего среднего FARIMA. Первоначально предложенный алгоритм, в котором искомая реализация содержала слишком много высокочастотных компонент, был усовершенствован в 2003 г. [12]. Программная реализация данного алгоритма доступна в математическом пакете MathLab, начиная с 7 версии.

Результаты исследований. В работе представлены результаты численного эксперимента, в ходе которого моделировалось фрактальное броуновское движение с заданным показателем H двумя описанными методами. Значения H для модельного ряда изменялись во всем возможном диапазоне $0 < H < 1$. Для полученного временного ряда рассчитывалась оценка H вышеописанными методами: R/S -анализа (\hat{H}_{rs}), дисперсии агрегированного ряда (\hat{H}_d), с помощью дискретного вейвлет-преобразования (\hat{H}_w) и индекса фрактальности (\hat{H}_{mu}). Для каждого значения показателя H были получены выборки его оценок и исследованы их статистические характеристики.

На рисунке показана зависимость средних значений оценки показателя Херста от его истинного значения. Вычисление средних значений H было проведено для модельных фрактальных рядов различной длины. Численное моделирование показало, что смещение не зависит от длины временного ряда. Сплошная прямая соответствует заданным значениям H . Очевидно, что оценки, полученные каждым из методов, являются сдвинутыми в зависимости от действительного значения показателя Херста. Средние значения оценок достаточно хорошо можно аппроксимировать прямыми линиями $\hat{H}_m = k_m H + b_m$, где коэффициенты k_m и b_m можно легко определить для каждого выбранного метода m .

В работе исследована зависимость средних квадратических отклонений оценок показателя Херста от заданных значений H и длины модельного фрактального ряда для каждого метода. В табл. 1 представлены значения средних квадратических отклонений оценок показателя Херста, которые были получены для рядов данных длиной 1024 значения. Значения H для модельного ряда изменялись во всем возможном диапазоне $0 < H < 1$.

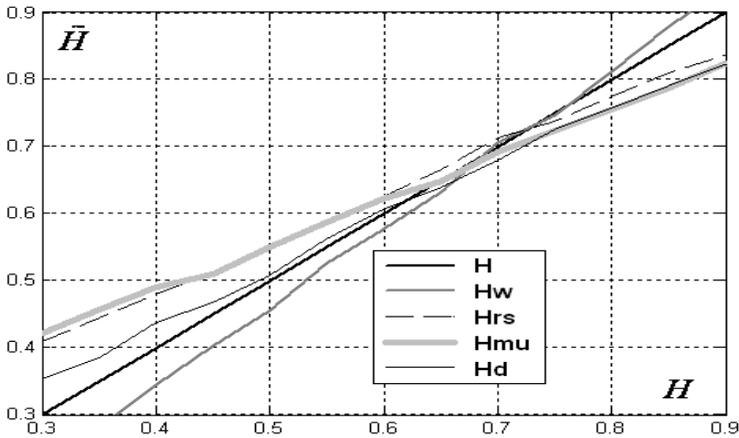


Рис. Зависимость средних значений оценок от действительных значений H

Таблица 1

Средние квадратические отклонения оценок H

Метод оценивания	Диапазон $S_{\hat{H}}$	Зависимость от H
R/S -анализ	$0,03 \leq S_{\hat{H}} \leq 0,08$	Возрастает вместе с H
Изменение дисперсии	$S_{\hat{H}} \approx 0,06$	Не зависит от H
Индекс фрактальности	$0,02 \leq S_{\hat{H}} \leq 0,08$	Возрастает вместе с H
Вейвлет-анализ	$S_{\hat{H}} \approx 0,06$	Не зависит от H

Табл. 2 иллюстрирует, как уменьшаются средние квадратические отклонения, полученные при расчете \hat{H} при увеличении длины ряда. Длина ряда, показанная в верхней строке таблицы, изменялась как 2^k , где $7 \leq k \leq 11$. В данном случае модельный показатель Херста $H = 0,8$.

Таблица 2

Зависимость средних квадратических отклонений \hat{H} от длины ряда

$S_{\hat{H}}$	128	256	512	1024	2048
S_{Hw}	0,19	0,14	0,10	0,08	0,05
S_{Hrs}	0,13	0,10	0,08	0,06	0,05
S_{Hmu}	0,11	0,09	0,08	0,07	0,05
S_{Hd}	0,09	0,07	0,06	0,05	0,04

В работе были исследованы законы распределения оценок \hat{H} . 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.

Одним из интересных, но слабоизученных вопросов оценивания параметра Херста, является 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. В табл. 3 представлены диапазоны значений выборочного коэффициента корреляции.

Таблица 3

Выборочный коэффициент корреляции между оценками \hat{H}

Методы, которыми получены оценки	Выборочный коэффициент корреляции r
R/S –анализ и изменение дисперсии	$0,3 \leq r \leq 0,7$
R/S –анализ и индекс фрактальности	$-0,6 \leq r \leq -0,3$
Изм. дисперсии и индекс фрактальности	$-0,7 \leq r \leq -0,3$
Вейвлет–анализ и другие методы	Является незначимым

Выводы. Обобщая результаты численного исследования, можно сделать вывод, что оценки показателя Херста, которые получены рассмотренными методами, являются нормально распределенными случайными величинами. Средние значения оценок значительно сдвинуты относительно оцениваемого параметра Херста. Сдвиг среднего значения зависит от действительного значения H и метода оценивания; эту зависимость можно аппроксимировать прямой. Средние квадратические отклонения оценок также зависят от истинного значения параметра Херста оцениваемого временного ряда и метода оценивания и уменьшаются с ростом длины ряда. Таким образом, для каждого метода можно найти доверительные интервалы оцениваемого значения показателя Херста. Исходя из результатов корреляционного анализа, для увеличения точности оценки H надо использовать среднее арифметическое исправленных несмещенных оценок, полученных несколькими методами, одним из которых желателен должен быть метод вейвлет–преобразования.

Список литературы. 1. Willinger W. Bibliographical guide to self-similar traffic and performance modeling for modern high-speed network in "Stochastic networks: theory and applications" / W. Willinger, M.S. Taqqu, A.A. Erramilli: Editors F.P. Kelly, S. Zachary, I. Ziedins. – Oxford: Clarendon Press (Oxford University Press), 1996. – 384 p. 2. Clegg R.G. A practical guide to measuring the hurst parameter / R.G. Clegg // Computing science technical report. – 2005. – № CS–TR–916. – P. 125–138. 3. Федер Е.

Фракталы / *Е. Федер.* – М.: Мир, 1991. – 254 с. **4. Шелухин О.И.** Самоподобие и фракталы / *О.И. Шелухин, А.В. Осин, С.М. Смольский* // Под ред. О.И. Шелухина. – М.: ФИЗМАТЛИТ 2008. – 368 с. **5. Abry P.** Wavelet analysis of long-range dependent traffic / *P. Abry, D. Veitch* // IEEE/ACM Transactions Information Theory. – 1998. – № 1 (44). – P. 2–15. **6. Петерс Э.** Фрактальный анализ финансовых рынков. Применение теории хаоса в инвестициях и экономике / *Э. Петерс.* – М.: Интернет-трейдинг, 2004. – 304 с. **7. Столлингс В.** Современные компьютерные сети / *В. Столлингс.* – С.Пб.: Питер, 2003. – 784 с. **8. Малла С.** Вэйвлеты в обработке сигналов / *С. Малла.* – М.: Мир, 2005. – 671 с. **9. Abry P.** Self-similarity and long-range dependence through the wavelet lens / *P. Abry, P. Flandrin, M.S. Taqqu, D. Veitch* // Theory and applications of long-range dependence: Birkhäuser, 2003. – P. 527-556. **10. Дубовиков М.М.** Размерность минимального покрытия и локальный анализ фрактальных временных рядов / *М.М. Дубовиков, А.В. Крянев, Н.В. Старченко* // Вестник РУДН. – 2004. – Т. 3. – № 1. – С. 81-95. **11. Abry P.** The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation / *P. Abry, F. Sellan* // Appl. and Comp. Harmonic Anal. – 1996. – V. 3 (4). – P. 377-383. **12. Bardet J.-M.** Generators of long-range dependence processes: a survey, "Theory and applications of long-range dependence" / *J.-M. Bardet, G. Lang, G. Oppenheim, A. Philippe, S. Stoev, M.S. Taqqu* // Theory and applications of long-range dependence: Birkhäuser, 2003. – P. 579-623.

УДК 519.2

Порівняльний аналіз статистичних властивостей оцінок показника Херста / Кіриченко Л.О. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2010. – № 21. – С. 88 – 95.

Проведено чисельний аналіз статистичних характеристик оцінок показника Херста для самоподібних часових рядів, отриманих різними методами. Показано, що оцінки мають нормальний розподіл, середні значення якого значно зміщені щодо оцінюваного параметра. Зсув залежить від дійсного значення параметра й методу оцінювання; цю залежність можна апроксимувати прямою. Середні квадратическі відхилення оцінок також залежать від значення параметра й методу оцінювання. Іл.: 1. Табл.: 3. Библиогр.: 12 назв.

Ключові слова: показник Херста, оцінка показника Херста, самоподібний часовий ряд.

UDC 519.2

Comparative analysis of the statistical properties of Hurst parameter estimators / Kirichenko L.O. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2010. – № 21. – P. 88 – 95.

The aim of this paper is estimation of Hurst parameter by different methods and comparative analysis of the statistical properties of estimators. Based on the results of numerical experiment, we can conclude that all estimating Hurst have a normal distribution, the mean is significantly shifted on the estimated parameter. The shift of mean depends on the value of Hurst parameter and the method of estimation. The standard deviations of the estimators also depend on the value of Hurst parameter and the method of estimation. Figs: 1. Tabs: 3. Refs: 12 titles.

Key words: Hurst parameter, estimator of Hurst parameter, self-similar time series.

Поступила в редакцію 27.10.2009