

УДК 519.24; 004.8; 519.6
doi: 10.21685/2072-3059-2024-1-4

Быстрый алгоритм робастных оценок показателя Херста при анализе малых выборок биометрических данных и данных рынка

В. И. Волчихин¹, А. И. Иванов², В. А. Тихомиров³, Д. В. Тарасов⁴

^{1,4}Пензенский государственный университет, Пенза, Россия

²Пензенский научно-исследовательский электротехнический институт, Пенза, Россия

³Научно-производственное объединение

«Развитие инновационных технологий», Тверь, Россия

¹cnit@pnzgu.ru, ²ivan@pniei.penza.ru, ³info@npo-rit.ru, ⁴tarasovdv@mail.ru

Аннотация. *Актуальность и цели.* В настоящее время показатель Херста достаточно легко интерпретируется применительно к биометрическим, медицинским и экономическим данным, однако его принято оценивать на больших выборках. Целью работы является снижение выборки реальных данных, на которых достаточно надежно может быть вычислен показатель Херста. *Материалы и методы.* Используется связь показателя Херста с автокорреляционными функционалами. Предложено свести задачу к оценке автокорреляционных свойств исследуемой последовательности реальных данных. *Результаты и выводы.* Автокорреляционные функционалы относятся к задачам квадратичной вычислительной сложности, тогда как показатель Херста имеет существенно большую вычислительную сложность и менее устойчив. Это позволяет снизить требования к объему используемой для вычислений выборки реальных данных.

Ключевые слова: автокорреляционный функционал, показатель Херста, малые выборки, биометрические данные

Для цитирования: Волчихин В. И., Иванов А. И., Тихомиров В. А., Тарасов Д. В. Быстрый алгоритм робастных оценок показателя Херста при анализе малых выборок биометрических данных и данных рынка // Известия высших учебных заведений. Поволжский регион. Технические науки. 2024. № 1. С. 48–55. doi: 10.21685/2072-3059-2024-1-4

A fast algorithm for robust estimates of the Hurst exponent when analyzing small samples of biometric and market data

V.I. Volchikhin¹, A.I. Ivanov², V.A. Tikhomirov³, D.V. Tarasov⁴

^{1,4}Penza State University, Penza, Russia

²Penza Scientific Research Electrotechnical Institute, Penza, Russia

³Scientific-production association “Razvitiye innovatsionnukh tekhnologiy” (Development of innovative technologies), Tver, Russia

¹cnit@pnzgu.ru, ²ivan@pniei.penza.ru, ³info@npo-rit.ru, ⁴tarasovdv@mail.ru

Abstract. *Background.* Currently, the Hurst exponent is quite easily interpreted in relation to biometric, medical and economic data, but it is customary to evaluate it on large samples. The purpose of the work is to reduce the sample of real data on which the Hurst exponent can be

calculated quite reliably. *Materials and methods.* The connection between the Hurst exponent and autocorrelation functionals is used. It is proposed to reduce the problem to assessing the autocorrelation properties of the studied sequence of real data. *Results and conclusions.* Autocorrelation functionals relate to problems of quadratic computational complexity, while the Hurst exponent has a significantly higher computational complexity and is less stable. This makes it possible to reduce the requirements for the size of the real data sample used for calculations.

Keywords: autocorrelation functional, Hurst exponent, small samples, biometric data

For citation: Volchikhin V.I., Ivanov A.I., Tikhomirov V.A., Tarasov D.V. A fast algorithm for robust estimates of the Hurst exponent when analyzing small samples of biometric and market data. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = University proceedings. Volga region. Engineering sciences.* 2024;(1):48–55. (In Russ.). doi: 10.21685/2072-3059-2024-1-4

Введение

Показатель Херста является одним из активно используемых параметров при анализе рынка [1–4] и коллективной биометрии [5]. К сожалению, эмпирический показатель Херста по умолчанию предполагает использование больших выборок из-за того, что он является степенным:

$$\frac{R}{\sigma} = \left\{ \frac{N}{2} \right\}^H, \quad (1)$$

где N – размер выборки; R – размах выборки; σ – стандартное отклонение выборки; H – степенной показатель Херста, изменяющийся в интервале от 0,5 до 1,0 для предсказуемых персистентных систем и в интервале от 0,0 до 0,5 для антиперсистентных систем с периодическим переключением рынка из состояния «медведи» в состояние «быки».

Переходя в логарифмическую форму уравнения (1), получим еще один вариант записи показателя Херста:

$$H = \log(R/\sigma)/\log(N/2). \quad (2)$$

Вторая форма удобна для пояснения причин, по которым при анализе данных рынков и данных коллективной биометрии необходимы большие выборки. Если предположить, что логарифм нормированного размаха данных является константой, то ошибка оценки показателя будет оцениваться следующим соотношением:

$$\Delta H \approx \text{const}/\log(N/2). \quad (3)$$

То есть ошибки из-за сокращения размеров выборки реальных данных должны расти обратно пропорционально логарифму объема выборки.

Кажется, что оценка значения ошибок (3) крайне пессимистична и обрекает исследователей на накопление больших объемов данных. Возможен обходной путь решения проблемы через создание быстрых алгоритмов оценки показателя Херста [6, 7]. Этот путь состоит в отказе от попыток наблюдения нормированного размаха экспериментальных данных (1) с замещением на наблюдение поведения автокорреляционных функционалов [8].

Теоретическое обоснование такой возможности строится на том, что высоко персистентные временные ряды с высоким значением показателя

Херста $H \approx 1,0$ должны давать высокие значения классических автокорреляционных функционалов $r(x(t), x(t-\tau)) \approx 1,0$ при любом сдвиге данных. В противоположном случае, когда показатель Херста минимален $H \approx 0,0$, значение автокорреляционных функционалов должно иметь случайный знак при высоком значении его модуля $r(x(t), x(t-\tau)) \approx \pm 1,0$ (предельно высокая антиперсистентность). В случае, когда показатель $H \approx 0,5$, автокорреляционные функционалы продолжают иметь случайный знак при малых значениях их модулей $r(x(t), x(t-\tau)) \approx \pm 0,0$.

Критерий подсчета смены знаков откликов автокорреляционных функционалов

Для высокого значения показателя Херста $H \approx 1,0$ должно наблюдаться малое значение вычисленного числа смены знака откликов автокорреляционных функционалов. Соответственно для нейросетевого распознавания больших значений показателя мы можем воспользоваться критерием подсчета числа наблюдаемых смен знаков у соседних отсчетов автокорреляционных функционалов.

Пример программной реализации критерия на языке программирования MathCAD для данных малых выборок в 32 опыта приведен в левой части рис. 1. В правой части рисунка приведены дискретные спектры нового критерия при воздействии на него слабо коррелированными и сильно коррелированными данными.

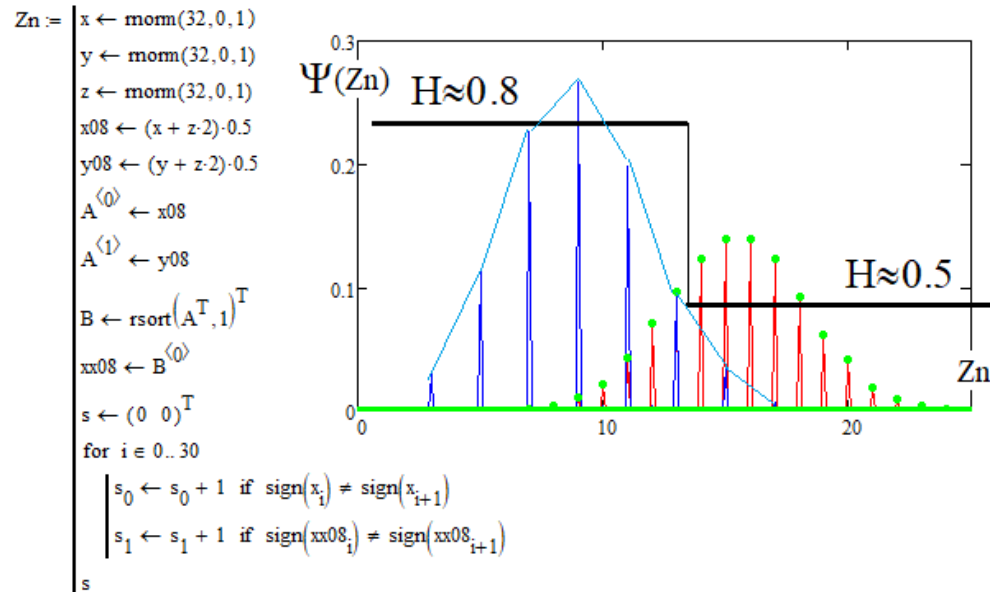


Рис. 1. Разделение сильно зависимых данных и данных со слабой корреляционной сцепленностью при анализе дискретных спектров подсчета числа смен знаков анализируемой последовательности

Формально из рассматриваемого критерия может быть создан искусственный нейрон, если воспользоваться выходным квантователем. Выходной квантователь должен давать состояние «1» при обнаружении сильно коррелированных данных $H \approx 0,8$. Квантователь сравнивает выходные данные критерия с порогом $k = 13,5$. Порог расположен между спектральными линиями анализируемых данных.

Если обнаружение сильно коррелированных данных $H \approx 0,8$ рассматривать как основную задачу, то искусственный нейрон будет давать значение вероятности ошибок второго рода $P_2 \approx 0,146$. В этом случае вероятность ошибок первого рода (ошибочное принятие решения «0») составит $P_1 \approx 0,231$. Среднее значение вероятностей ошибок первого и второго рода $P_{EE} \approx 0,189$.

Критерий оценки суммы амплитуд выходных состояний автокорреляционных функционалов

Кроме критерия подсчета числа смены знаков значений автокорреляционных функционалов, возможно использование критерия суммирования амплитуд значений функционалов. На рис. 2,*а* приведена программная реализация нового критерия, на рис. 2,*б* – результаты численного эксперимента.

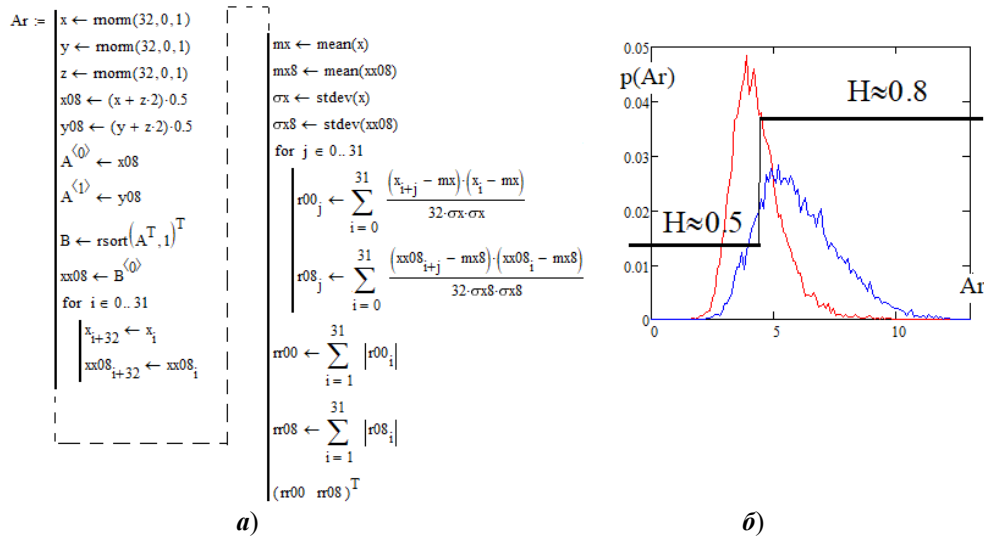


Рис. 2. Разделение данных с высоким и средним значением показателя Херста при анализе непрерывных спектров состояний автокорреляционного функционала

Критерий построен на сдвиге данных, перемножении данных и накоплении нормированных произведений. Критерий дает непрерывные выходные данные, обладающие достаточно высокой разделимостью $P_1 = P_2 \approx P_{EE} \approx 0,261$. То есть непрерывный критерий слабее предыдущего дискретного критерия примерно на 42 %. Тем не менее эти два критерия хорошо дополняют друг друга из-за того, что их отклики слабо коррелированы между собой $\text{corr}(\text{Zn}, \text{Ar}) \approx -0,06$.

Моделирование персистентных и антиперсистентных хаотических систем

Оказалось, что модель значений показателя Херста хорошо описывается теорией хаоса (теорией двумерного броуновского движения [2, 3]). Естественно, что малый объем выборки всегда приводит к размыванию результатов оценок. На рис. 3 приведены распределения данных, полученные при моделировании трех вариантов броуновских блужданий:

- с малым показателем Херста $H \approx 0,2$ при $\text{corr}(x, y) \approx -0,8$;
- со средним показателем $H \approx 0,5$ при $\text{corr}(x, y) \approx \pm 0,0$;
- с большим показателем Херста $H \approx 0,8$ при $\text{corr}(x, y) \approx +0,8$.

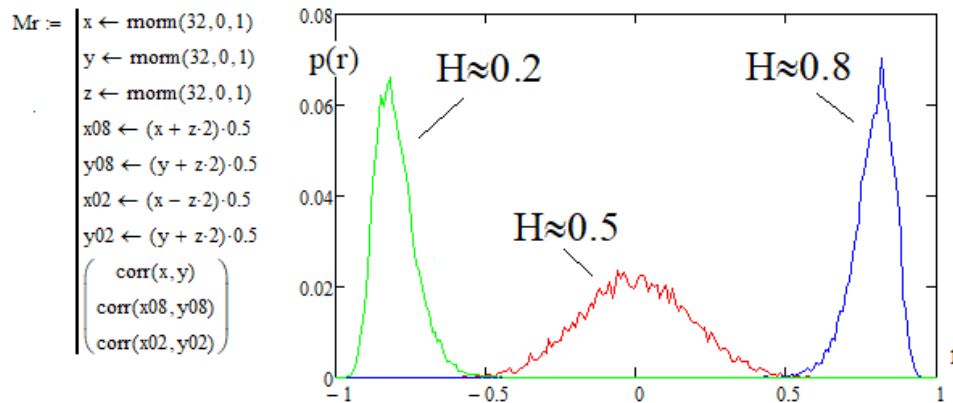


Рис. 3. Распределение данных при моделировании персистентного и антиперсистентного хаоса

Следует отметить, что рассмотренный выше искусственный нейрон инвариантен по отношению к положительно коррелированным данным $H \approx 0,8$ и отрицательно коррелированным данным $H \approx 0,2$. Это является одной из технических проблем синтеза быстрых алгоритмов вычисления показателей Херста на малых выборках. На данный момент очевидным является то, что автокорреляционные функционалы позволяют решить задачу в интервале показателей от $H \approx 0,5$ до $H \approx 1,0$ для большинства фрактальных объектов реального мира с длинной памятью и монотонной импульсной переходной функцией.

Для расширения диапазона оцениваемых показателей от $H \approx 0,0$ до $H \approx 1,0$ быстрыми алгоритмами на малых выборках необходимо автокорреляционные функционалы (автокорреляционные нейроны) дополнять функционалами-нейронами иного типа. Предположительно, они могут быть построены на теории броуновского движения, восстанавливающей по косвенным признакам периодичности смены знаков корреляционной сцепленности [9, 10] (периодического переключения системы в два разных состояния «медведи» и «быки»).

Еще одним решением проблемы является принудительное (внешнее) принятие решения о характере наблюдаемых пользователем процессов [11] на основании его интуитивной классификации.

Заключение

Было показано, что два простейших автокорреляционных критерия уже позволяют давать достаточно точные оценки значений показателя Херста на малых выборках в 32 опыта. Так как на сегодня известно порядка 20 автокорреляционных критериев [7], точность оценок значений показателей Херста может быть существенно увеличена. По крайней мере сегодня мы можем создавать нейросетевые предсказатели с 22-кратной выходной кодовой избыточностью. Наличие хорошо разработанной двухмерной теории хаоса (броуновского движения на плоскости), видимо, даст возможность надежно протестировать нейросетевые конструкции. Вполне возможно, что нейросетевые конструкции можно будет аттестовать по уровню их погрешностей (по уровням вероятности появления ошибок предсказаний).

Список литературы

1. Калуж Ю. А., Логинов В. М. Показатель Херста и его скрытые свойства // Сибирский журнал индустриальной математики. 2002. Т. 5, № 4. С. 29–37.
2. Петерс. Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка / пер. с англ. В. И. Гусевой. М. : МИР, 2000. 333 с.
3. Мандельброт Б., Хадсон З. Л. (НЕ)послушные рынки. Фрактальная революция в финансах / пер. с англ. и ред. А. Ю. Заякина. М. ; СПб. ; Киев : Вильямс, 2006. 408 с.
4. Найман Э. Как покупать дешево и продавать дорого : пособие для разумного инвестора. М. : Альпина Паблишерз, 2011. 552 с.
5. Иванов А. И. Высокоразмерная коллективная биометрия подсознательного поведения людей на рынке и производстве : препринт. Пенза : Из-во ПГУ, 2021. 60 с.
6. Иванов А. И., Егорова Ю. Ю. Корреляционный метод быстрой оценки текущего значения показателя Херста биометрических данных и данных рынка // Нейрокомпьютеры: разработка, применение. 2012. № 3. С. 26–27.
7. Иванов А. И., Золотарева Т. А. Искусственный интеллект в защищенном исполнении: синтез статистико-нейросетевых автоматов многокритериальной проверки гипотезы независимости малых выборок биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 105 с.
8. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок : справочник. Пенза : Изд-во ПГУ, 2022. 218 с.
9. Федер Е. Фракталы. М., 2014. № 69. 267 с. (Книжная серия «Синергетика: от прошлого к будущему»).
10. Орлов Ю. Н., Осминин К. П. Нестационарные временные ряды: Методы прогнозирования с примерами анализа финансовых и сырьевых рынков. М. : УРСС, 2011. 384 с.
11. Фейс К. Трейдинг, основанный на интуиции. СПб. : Питер, 2011. 240 с.

References

1. Kalush Yu.A., Loginov V.M. Hurst exponent and its hidden properties. *Sibirskiy zhurnal industrial'noy matematiki = Siberian journal of industrial mathematics*. 2002;5(4):29–37. (In Russ.)
2. Peters. E. *Khaos i poryadok na rynkakh kapitala. Novyy analiticheskiy vzglyad na tsikly, tseny i izmenchivost' rynka = Chaos and order in capital markets. A new analytical look at cycles, prices and market volatility*. Transl. from Eng. by V.I. Guseva. Moscow: MIR, 2000:333. (In Russ.)

3. Mandel'brot B., Khadson Z.L. *(NE)poslushnye rynki. Fraktal'naya revolyutsiya v finansakh = (NOT)obedient markets. Fractal revolution in finance*. Transl. from Eng. and ed. by A.Yu. Zayakin. Moscow; Saint Petersburg; Kiev: Vil'yams, 2006:408. (In Russ.)
4. Nayman E. *Kak pokupat' deshevo i prodavat' dorogo: posobie dlya razumnogo investora = How to buy low and sell high: guidance for the smart investor*. Moscow: Al'pina Publisherz, 2011:552. (In Russ.)
5. Ivanov A.I. *Vysokorazmernaya kollektivnaya biometriya podsoznatel'nogo povedeniya lyudey na rynke i proizvodstve: preprint = High-dimensional collective biometrics of subconscious behavior of people in the market and production: preprint*. Penza: Iz-vo PGU, 2021:60. (In Russ.)
6. Ivanov A.I., Egorova Yu.Yu. Correlation method for quickly assessing the current value of the Hurst exponent of biometric data and market data. *Neyro-komp'yutery: razrabotka, primeneniye = Neuro-computers: development, application*. 2012;(3):26–27. (In Russ.)
7. Ivanov A.I., Zolotareva T.A. *Iskusstvennyy intellekt v zashchishchenom ispolnenii: sintez statistiko-neyrosetevykh avtomatov mnogokriterial'noy proverki gipotezy nezavisimosti malykh vyborok biometricheskikh dannykh: preprint = Artificial intelligence in a secure design: synthesis of statistical-neural network automata for multi-criteria testing of the hypothesis of independence of small samples of biometric data: preprint*. Penza: Izd-vo PGU, 2020:105. (In Russ.)
8. Ivanov A.I. *Neyrosetevoy mnogokriterial'nyy statisticheskiy analiz malykh vyborok: spravochnik = Neural network multicriteria statistical analysis of small samples: reference book*. Penza: Izd-vo PGU, 2022:218. (In Russ.)
9. Feder E. *Fraktaly (Knizhnaya seriya «Sinergetika: ot proshlogo k budushchemu») = Fractals (Book series “Synergetics: from past to future”)*. Moscow, 2014;(69):267. (In Russ.)
10. Orlov Yu.N., Osminin K.P. *Nestatsionarnye vremennyye ryady: Metody prognozirovaniya s primerami analiza finansovykh i syr'evykh rynkov = Non-stationary time series: Forecasting methods with examples of analysis of financial and commodity markets*. Moscow: URSS, 2011:384. (In Russ.)
11. Feys K. *Treyding, osnovanny na intuitsii = Trading based on intuition*. Saint Petersburg: Piter, 2011:240. (In Russ.)

Информация об авторах / Information about the authors

Владимир Иванович Волчихин

доктор технических наук, профессор,
президент Пензенского государственного
университета (Россия, г. Пенза,
ул. Красная, 40)

E-mail: cnit@pnzgu.ru

Vladimir I. Volchikhin

Doctor of engineering sciences, professor,
president of Penza State University
(40 Krasnaya street, Penza, Russia)

Александр Иванович Иванов

доктор технических наук, профессор,
научный консультант, Пензенский
научно-исследовательский
электротехнический институт (Россия,
г. Пенза, ул. Советская, 9)

E-mail: ivan@pniei.penza.ru

Aleksandr I. Ivanov

Doctor of engineering sciences, professor,
scientific adviser, Penza Scientific
Research Electrotechnical Institute
(9 Sovetskaya street, Penza, Russia)

Валерий Александрович Тихомиров

доктор технических наук, профессор,
генеральный директор научно-
производственного объединения
«Развитие инновационных технологий»
(Россия, г. Тверь, ул. Озерная, 14,
корп. 1)

E-mail: info@npo-rit.ru

Valeriy A. Tikhomirov

Doctor of engineering sciences, professor,
CEO of Scientific-production association
“Razvitiye innovatsionnukh tekhnologiy”
(Development of innovative technologies)
(building 1, 147 Ozernaya street,
Tver, Russia)

Дмитрий Викторович Тарасов

кандидат технических наук, доцент
кафедры высшей и прикладной
математики, Пензенский
государственный университет
(Россия, г. Пенза, ул. Красная, 40)

E-mail: tarasovdv@mail.ru

Dmitriy V. Tarasov

Candidate of engineering sciences,
associate professor of the sub-department
of higher and applied mathematics,
Penza State University (40 Krasnaya
street, Penza, Russia)

Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflicts of interests.

Поступила в редакцию / Received 10.11.2023

Поступила после рецензирования и доработки / Revised 18.12.2023

Принята к публикации / Accepted 16.01.2024