УДК 519.24; 53; 57.017
doi: 10.21685/2307-4205-2023-3-6

ОЦЕНКИ ПОКАЗАТЕЛЯ ХЕРСТА НА МАЛЫХ ВЫБОРКАХ: ПРОСТЕЙШИЙ ВАРИАНТ НЕЛИНЕЙНОГО КОМПЕНСАТОРА МЕТОДИЧЕСКИХ ОШИБОК ФЕДЕРА ПРИ МОДЕЛИРОВАНИИ ДАННЫХ ЭКОНОМИКИ И БИОМЕТРИИ

А. И. Иванов[1], Д. В. Тарасов[2], К. А. Горбунов[3]

[1] Пензенский научно-исследовательский электротехнический институт, Пенза, Россия
[2, 3] Пензенский государственный университет, Пенза, Россия
[1] ivan@pniei.penza.ru, [2] tarasovdv@mail.ru, [3] kirill.gorbunov@gmail.com

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 устранение методической ошибки, возникающей из-за малых выборок реальных данных. *Материалы и методы.* Используется моделирование двухмерного броуновского движения, порождающее возможность вычисления эталонных показателей Херста. Предложено средствами имитационного моделирования заранее строить нелинейный корректор методических ошибок, обнаруженных ранее Е. Федером. *Результаты и выводы.* Получена связь значения методических погрешностей оценки показателя Херста, позволяющая корректировать оценки при малых значениях показателя $H < 0{,}35$ и больших значениях показателя $H > 0{,}65$. Необходимость в корректировке методических ошибок растет по мере снижения объемов малых выборок реальных данных экономики и биометрии.

Ключевые слова: автокорреляционный функционал, показатель Херста, малые выборки, биометрические данные, методическая погрешность, корректор погрешности

Для цитирования: Иванов А. И., Тарасов Д. В., Горбунов К. А. Оценки показателя херста на малых выборках: простейший вариант нелинейного компенсатора методических ошибок федера при моделировании данных экономики и биометрии // Надежность и качество сложных систем. 2023. № 3. С. 50–54. doi: 10.21685/2307-4205-2023-3-6

HURST EXPONENT ESTIMATES ON SMALL SAMPLES: THE SIMPLEST VERSION OF FEDER'S NON-LINEAR METHOD ERROR COMPENSATOR FOR MODELING ECONOMIC AND BIOMETRIC DATA

A.I. Ivanov[1], D.V. Tarasov[2], K.A. Gorbunov[3]

[1] Penza Research Electrotechnical Institute, Penza, Russia
[2, 3] Penza State University, Penza, Russia
[1] ivan@pniei.penza.ru, [2] tarasovdv@mail.ru, [3] kirill.gorbunov@gmail.com

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 aim of the work is to eliminate the methodological error that occurs due to small samples of real data. *Materials and methods.* The simulation of two-dimensional Brownian motion is used, which gives rise to the possibility of calculating the Hurst exponents. It is proposed by means of simulation modeling to build in advance a nonlinear corrector of methodological errors discovered earlier by E. Feder. *Results and conclusions.* A relation has been obtained for the value of methodological errors in estimating the Hurst exponent, which makes it possible to correct estimates for small values of the exponent H<0.35 and large values of the exponent $H > 0.65$. The need to correct methodological errors is growing as the size of small samples of real economic and biometric data decreases.

Keywords: autocorrelation functional, Hurst exponent, small samples, biometric data, methodological error, error corrector

For citation: Ivanov A.I., Tarasov D.V., Gorbunov K.A. Hurst exponent estimates on small samples: the simplest version of Feder's non-linear method error compensator for modeling economic and biometric data. *Nadezhnost' i kachestvo slozhnykh sistem = Reliability and quality of complex systems.* 2023;(3):50–54. (In Russ.). doi: 10.21685/2307-4205-2023-3-6

Введение

Показатель Херста является одним из активно используемых параметров при анализе рынка [1–3] и коллективной биометрии [4]. К сожалению, эмпирический показатель Херста по умолчанию предполагает использование больших выборок из-за того, что он является степенным:

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

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 изменяющийся в интервале от 0,0 до 0,5 для антиперсистентных систем.

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

$$H(N) = \log_2(R(N)/\sigma(N))/\log_2(\pi \cdot N/2). \tag{2}$$

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:

$$\Delta H(N) \approx \mathrm{const}/\log_2(\pi \cdot N/2), \tag{3}$$

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

Проведенные численные эксперименты показали, что наряду со случайной составляющей погрешности $\Delta H(N)$ оценки вида (2) содержат методическую составляющую погрешности, которая может быть устранена. Одной из первых работ, отмечающей наличие мультиприкативной составляющей методической погрешности, является работа Е. Федера [5] при слишком низких и слишком больших значениях показателя Херста. Одним из первых попытался устранить мультипликативно погрешность Э. Нейман [6]. Работ о том, на сколько эффективен мультипликативный корректор Э. Неймана [6], нами не обнаружено. Это и послужило толчком к выполнению данной работы по корректировке методической погрешности, выявленной Е. Федером [5].

Численное моделирование значений показателя Херста для малых выборок на эталонных шумах броуновского движения

Теория броуновских блужданий [2, 3, 5] позволяет связать коэффициент парной корреляции со значениями показателя Херста. Примеры случайных броуновских блужданий, соответствующих разным значениям показателя Херста, приведены на рис. 1.

Из теории следует, что при независимых данных, использованных для формирования броуновского блуждания, показатель Херста должен составлять $H \approx 0{,}5$ при $r(x, y) \approx 0{,}0$. В нашем случае для выборок объемом 32 опыта коэффициент парной корреляции должен вычисляться по следующей формуле:

$$r(x, y) = \frac{1}{32}\sum_{i=1}^{32}\frac{(x_i - E(x))\cdot(y_i - E(y))}{\sigma(x)\cdot\sigma(y)}. \tag{4}$$

Кроме того, известно, что для $r(x, y) \approx -1{,}0$ показатель Херста должен составить минимальное значение $H \approx 0{,}0$. Напротив, для сильно коррелированных данных $r(x, y) \approx +\,1{,}0$ показатель должен принимать максимальное значение Н ≈ 1,0.

Моделирование предельных случаев минимального и максимального значения показателей Херста через использования программных генераторов псевдостучайных чисел невозможно. Однако возможно моделирование близких к предельным случайных блужданий для очень низких и очень больших значений показателей Херста. Возможность подобного моделирования иллюстрирует рис. 2.

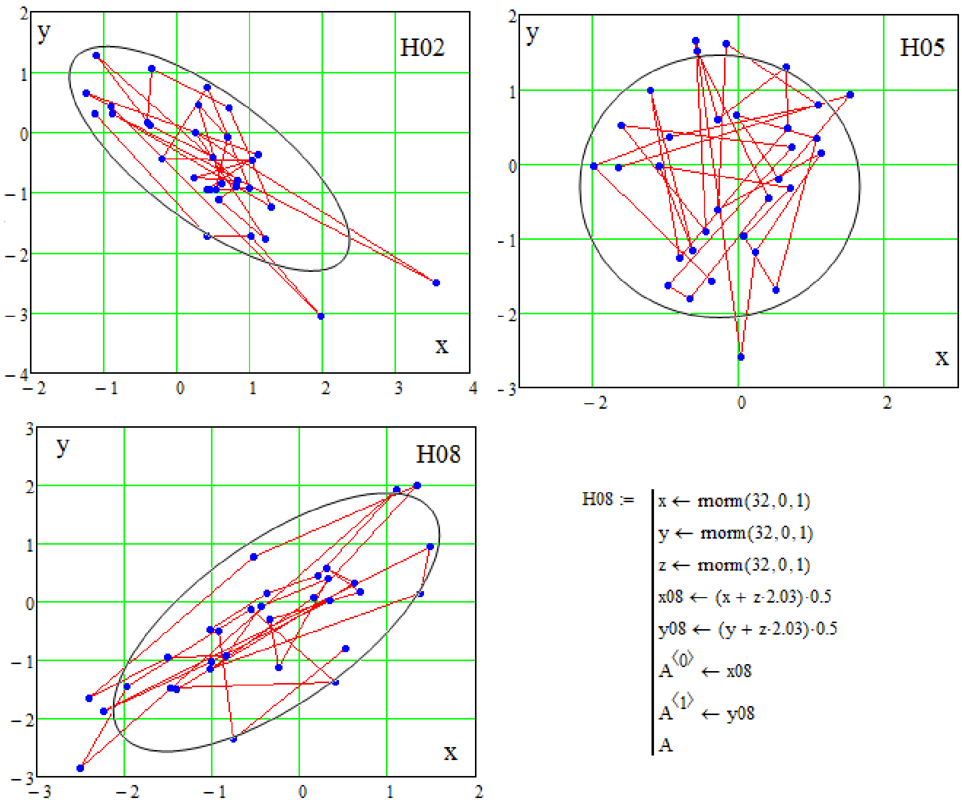


Рис. 1. Случайные броуновские блуждания, соответствующие разным значениям показателя Херста $H \approx 0{,}2$, $H \approx 0{,}5$, $H \approx 0{,}8$

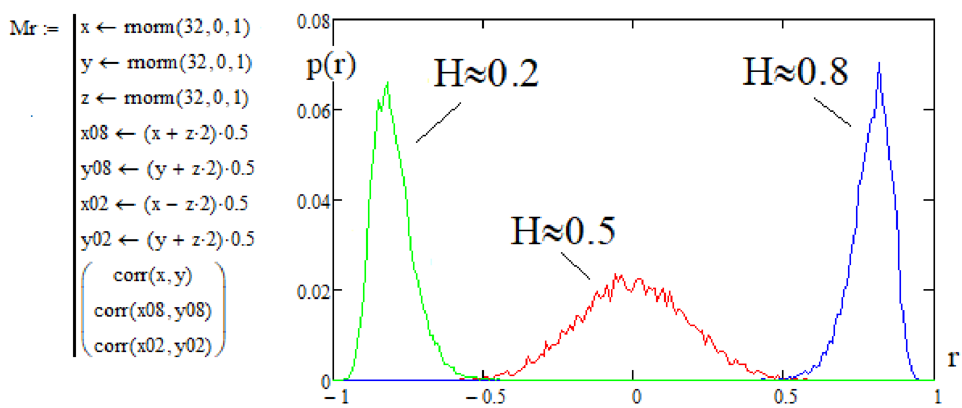
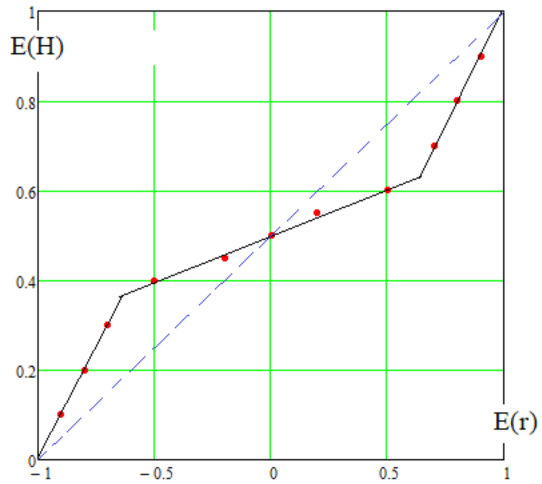


Рис. 2. Распределение данных при моделировании персистентного и антиперсистентного броуновского хаоса

Кусочно-линейная сшивка шкалы показателя Херста и шкалы двухмерных коэффициентов корреляции $r(x, y)$

В силу того, что моделирование броуновских движений возможно в широких пределах, мы можем выполнить подобное моделирование и получить 11 точек взаимосвязи значений математических ожиданий показателя Херста и значений математических ожиданий коэффициентов парной корреляции. Это необходимо из-за того, чтобы устранить неопределенность, обусловленную случайными данными. На рис. 3 приведены точки, полученные численным моделированием.

Из рис. 3 видно, что Е. Федер [5] верно указал на необходимость корректировки связки шкал Херста с парным коэффициентом корреляции (4) на краях шкал. Формально сшивка шкал может быть выполнена кусочно-линейной аппроксимацией. В простейшем случае может быть использована аппроксимация, состоящая всего из трех участков линейного приближения:



	0	1	2	3	4	5	6	7	8	9	10
E(H)	0.1	0.2	0.3	0.4	0.45	0.5	0.55	0.6	0.7	0.8	0.9
E(r)	-0.9	-0.8	-0.7	-0.5	-0.2	0	0.2	0.5	0.7	0.8	0.9

Рис. 3. Кусочно-линейное описание связи показателя Херста
с коэффициентом парной корреляции броуновского движения

$$\begin{cases} H(r) = (1-r), \text{ если } r \le -0{,}65, \\ H(r) = \dfrac{(2-r)}{4}, \text{ если } r \le -0{,}65 \vee r < 0{,}65, \\ H(r) = r, \text{ если } r \ge 0{,}65. \end{cases} \tag{5}$$

Эта аппроксимация построена на двух точках $r(x, y) = \pm 0{,}65$, где происходит переход на крайние участки аномалий Е. Федера с двойной крутизной наклона. Между этими крайними точками связь показателя с корреляцией в четыре раза меньшей крутизны, чем на краях диапазона. Простейшая (самая примитивная) линейная связь с неизменяемым наклоном двух сшиваемых показателей на рис. 3 отображена пунктиром.

Заключение

Таким образом, кусочно-линейная связь показателя Херста с классическим показателем двухмерной корреляции оказывается очень простой в случае использования приближения тремя участками. При этом программная реализация приближения так же оказывается достаточно простой (5).

В статье рассматриваются два типа функционалов. Функционал Херста (2) и классический корреляционный функционал (4), который легко наблюдаем только при моделировании зависимых данных броуновскими блужданиями. При этом функционал Херста не чувствителен к знаку корреляционной связи. Если пользоваться только функционалом Херста, то персистентные и антиперсистентные временные ряды неразличимы $H \approx 0{,}2 \approx 0{,}8$; $H \approx 0{,}1 \approx 0{,}9$. В связи с этим актуальной становится задача восстановления знака корреляционной сцепленности эквивалентных броуновских блужданий по последовательности реально наблюдаемых данных экономики, медицины и биометрии. Особую остроту проблема оценки знака корреляционной сцепленности реальных данных, видимо, будет приобретать при оценках показателя Херста и иных автокорреляционных функционалов на малых выборках [7]. Чем меньше выборка, тем больше внимания нужно уделять компенсации методических погрешностей.

Список литературы

1. Калуш Ю. А., Логинов В. М. Показатель Херста и его скрытые свойства // Сибирский журнал индустриальной математики. 2002. Т. 5, вып. 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. М. ; СПб. ; Киев : Вильямс, 2006. 408 с.
4. Иванов А. И. Высокоразмерная коллективная биометрия подсознательного поведения людей на рынке и производстве : препринт. Пенза : Изд-во ПГУ, 2021. 60 с.
5. Федер Е. Фракталы. М. : Мир, 1991. 254 с.
6. Найман Э. Как покупать дешево и продавать дорого : пособие для разумного инвестора. М. : Альпина Паблишерз, 2011. 552 с.
7. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок. Проверка гипотезы независимости : справочник. Пенза : Изд-во ПГУ, 2022. 218 с.

References

1. Kalush Yu.A., Loginov V.M. The Hurst index 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 the capital markets. A new analytical view on cycles, prices and market variability*. Transl. from the English by V.I. Gusev. 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*. Moscow; Saint Petersburg; Kiev: Vil'yams, 2006:408. (In Russ.)
4. 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: Izd-vo PGU, 2021:60. (In Russ.)
5. Feder E. *Fraktaly = Fractals*. Moscow: Mir, 1991:254. (In Russ.)
6. Nayman E. *Kak pokupat' deshevo i prodavat' dorogo: posobie dlya razumnogo investora = How to buy cheap and sell expensive : a guide for a reasonable investor*. Moscow: Al'pina Pablisherz, 2011:552. (In Russ.)
7. Ivanov A.I. *Neyrosetevoy mnogokriterial'nyy statisticheskiy analiz malykh vyborok. Proverka gipotezy nezavisimosti: spravochnik = Neural network multicriteria statistical analysis of small samples. Testing the independence hypothesis : a reference book.* Penza: Izd-vo PGU, 2022:218. (In Russ.)

Информация об авторах / Information about the authors

Александр Иванович Иванов
доктор технических наук, доцент,
научный консультант,
Пензенский научно-исследовательский электротехнический институт
(Россия, г. Пенза, ул. Советская, 9)
E-mail: ivan@pniei.penza.ru

Aleksandr I. Ivanov
Doctor of technical sciences, associate professor,
scientific consultant,
Penza Research Electrotechnical Institute
(9 Sovetskaya street, Penza, 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 technical sciences,
associate professor of the sub-department of higher and applied mathematics,
Penza State University
(40 Krasnaya street, Penza, Russia)

Кирилл Андреевич Горбунов
аспирант,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: kirill.gorbunov@gmail.com

Kirill A. Gorbunov
Postgraduate student,
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 16.04.2023
Поступила после рецензирования/Revised 15.05.2023
Принята к публикации/Accepted 13.06.2023