

УДК 519.24; 004.8; 519.6

## Исключение методической погрешности оценки показателя Хёрста, возникающей из-за малого объема выборок

Иванов А.И.<sup>1</sup>, Тарасов Д.В.<sup>2</sup>

АО «Пензенский научно-исследовательский электротехнический институт»<sup>1</sup>,  
ФГБОУ ВО Пензенский государственный университет<sup>2</sup>

*Аннотация:* В настоящее время показатель Хёрста достаточно легко интерпретируется применительно к биометрическим, медицинским и экономическим данным, однако его принято оценивать на больших выборках. Целью работы является нейтрализация методической ошибки, возникающей из-за недостаточного объема исходных данных. В работе используется связь показателя Хёрста с автокорреляционными функционалами. Предложено средствами имитационного моделирования заранее вычислять аддитивную методическую погрешность и устранять ее для каждой выборки. Получена связь значения аддитивной методической погрешности оценки показателя Хёрста с размерами малых выборок в пределах от 20 до 80 опытов. Это позволяет повысить точность оценок, за счет корректировки аддитивной методической погрешности.

*Ключевые слова:* белый шум, показатель Хёрста, малые выборки, биометрические данные, методическая погрешность.

### 1. Введение

При анализе параметров рынка [1–3] и коллективной биометрии [9] одним из активно используемых инструментов является показатель Хёрста. К сожалению, эмпирический показатель Хёрста по умолчанию предполагает использование больших выборок из-за того, что он является степенным:

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

где  $N$  – размер выборки;  $R$  – размах выборки;  $\sigma$  – стандартное отклонение выборки;  $H$  – степенной показатель Хёрста, 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».

Другой вариант записи показателя Хёрста получается путем перехода от уравнения (1) в его логарифмическую форму:

$$H(N) = \frac{\log_2 \frac{R(N)}{\sigma(N)}}{\log_2 \left( \frac{\pi N}{2} \right)}, \quad (2)$$

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:

$$\Delta H(N) \approx \frac{\text{const}}{\log_2\left(\frac{\pi N}{2}\right)}, \quad (3)$$

Из этого следует, что ошибки из-за сокращения размеров выборки реальных данных должны расти обратно пропорционально логарифму объема выборки.

Проведенные численные эксперименты показали, что вместе со случайной составляющей погрешности оценки (2) содержится методическая составляющая погрешности, которую можно устранить. Одна из первых работ, отмечающая наличие мультипликативной составляющей методической погрешности, является работа Е. Федера [5] при слишком низких и слишком больших значениях показателя Хёрста. Одним из первых попытался устранить мультипликативную погрешность Э. Найман [6], однако работ о том, насколько эффективен мультипликативный корректор Э. Наймана нами не обнаружено. Это и послужило толчком к выполнению данной работы по корректировке только аддитивной составляющей методической погрешности.

## 2. Численное моделирование значений показателя Хёрста для малых выборок на эталонном «белом шуме»

Первоначально в рамках численного эксперимента построим порядка 10000 выборок имеющих нормальное распределение с нулевым математическим ожиданием и с величиной  $\sigma$  близкой к единице (здесь будем подавать незначительный шум как равномерно распределенную на интервале  $(0; 0,01)$  величину). Результаты численного моделирования значений показателя Хёрста для выборок объема  $N = 50$  показывают, что выборочное среднее значение показателя Хёрста здесь составляет 0,3470 (рис. 1), хотя его величина должна быть равна 0,5.

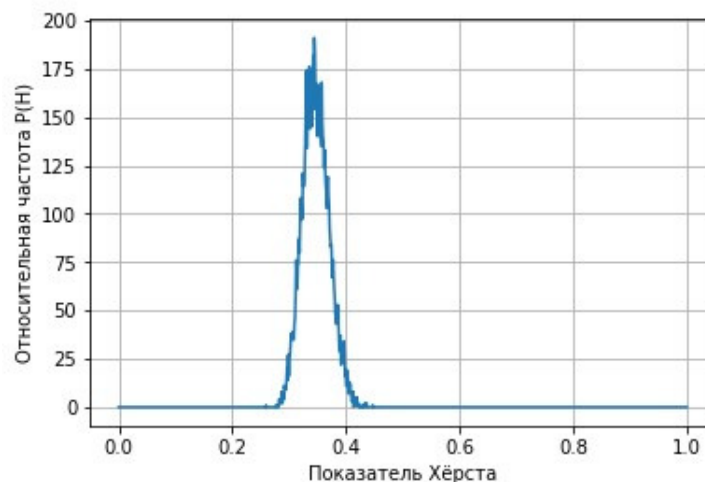


Рис. 1. Результаты численного эксперимента по оценке показателя Хёрста для независимых данных. Объем выборки  $N = 50$ , выборочное среднее значение показателя составляет 0,3470.

Напомним, что если бы формулы (1)-(2) оценки показателя Хёрста верно отражали действительность, то с ростом размеров выборки мы наблюдали бы постоянное значение математического ожидания распределений. Однако случайная составляющая погрешности с ростом объема выборки падает и, кроме того, существует значи-

тельная методическая составляющая погрешности оценок [7,8]. Таким образом, методическая погрешности для нашего примера составляет  $\Delta H \approx 0,5 - 0,3470 = 0,1530$ .

В табл. 1 представлены значения соответствующих поправок для 12-ти выборок объемом от 10 до 320 значений.

**Таблица 1.** Методическая погрешность как функция объема выборки

$N$	10	20	30	40	50	60	80	100	120	140	160	320
$\Delta H$	0,054	0,108	0,13	0,143	0,153	0,161	0,172	0,181	0,187	0,192	0,197	0,218

Устранение методической погрешности позволяет поднять точность оценок примерно в 2 раза на выборках в 30 опытов и более, однако при этом остается не скомпенсированной случайная составляющая погрешности.

### 3. Влияние размеров выборки на относительную погрешность оценки показателя Хёрста в эталонной выборке «белого шума»

Если методическую составляющую погрешности полностью устранить из оценок (2), то мы будем наблюдать идеальную ситуацию, когда математическое ожидание показателя Хёрста уже не будет зависеть от размеров выборки:

$$E(H(N)) \approx \text{const} = 0,5, \quad (4)$$

Для того, что бы наблюдать соотношение (4) придется многократно повторять вычисления с заданным объемом выборки. При этом случайную составляющую погрешности в первом приближении можно считать равной величине стандартного отклонения. В этом случае мы будем наблюдать монотонное снижение случайной составляющей по мере роста объема выборки (табл. 2). Отметим, что для нашего первоначально примера (рис. 1) случайная составляющая ошибки составит величину 0,0232.

**Таблица 2.** Случайная составляющая ошибки как функция объема выборки

$N$	10	20	30	40	50	60	80	100	120	140	160	320
$\Delta H$	0,035	0,031	0,027	0,025	0,023	0,022	0,02	0,019	0,018	0,017	0,017	0,013

Зависимость, определяющая случайную составляющую ошибки (табл. 2), в логарифмической системе координат хорошо описывается линейным приближением. Т.е., оценивая случайную составляющую, мы можем воспользоваться следующим приближением:

$$\Delta H(N) \approx \sigma(N) \approx 0,036 - 0,09 \cdot \lg N, \quad (5)$$

Последнее означает, что для выборки в 40 опытов значение относительной ошибки показателя Хёрста составит порядка 5%. Если же мы будем оценивать ошибку по

более жесткому правилу трех стандартных отклонений («трех сигм»), то на выборке в 40 опытов мы получим относительную ошибку в 15%. Крайне важным моментом является так же то, что относительная случайная ошибка очень медленно снижается. Так при увеличении выборки с 40 до 320 опытов происходит всего лишь двукратное снижение относительной составляющей случайной ошибки.

## 4. Заключение

Компенсация методической погрешности является принципиально важным технологическим элементом оценки значений показателя Хёрста как на малых, так и на больших выборках. Если методическая погрешность оценок не устраняется, то увеличение объема исследуемых выборок не приводит к снижению ошибок.

## Литература

1. Калуш Ю.А., Логинов В.М. Показатель Хёрста 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., Герасин В. Ю. Перспективы использования автокорреляционных статистических критериев как аналогов показателя Хёрста при обработке малых выборок данных экономики и биометрии // Труды международного симпозиума «Надёжность и качество». Пенза, 2023. Т. 2, С. 59–61.
8. Иванов А. И., Тарасов Д. В. Дополнительная коррекция аддитивной методической погрешности оценок показателя Хёрста при анализе эмпирических данных // Проблемы информатики в образовании, управлении, экономике и технике : сб. ст. XXIII Междунар. науч.-техн. конф., посвящ. 80-летию Пензенского государственного университета (г. Пенза, 24–25 ноября 2023 г.) / под ред. А. В. Кузьмина, В. В. Дрождина. – Пенза : Изд-во ПГУ, 2023. С. 14–18.

MSC 91B82

## **Elimination of the methodological error in estimating the Hurst index, which occurs due to the small volume of samples**

A.I. Ivanov<sup>1</sup>, D.V. Tarasov<sup>2</sup>

Penza Scientific Research Electrotechnical Institute<sup>1</sup>, Penza State University<sup>2</sup>

*Abstract:* Currently, the Hurst indicator 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 neutralize the methodological error that occurs due to insufficient amount of initial data. The paper uses the connection of the Hurst exponent with autocorrelation functionals. It is proposed to calculate the additive methodological error in advance by means of simulation modeling and eliminate it for each sample. The relationship between the value of the additive methodological error in estimating the Hurst index and the size of small samples in the range from 20 to 80 experiments was obtained. This makes it possible to increase the accuracy of estimates by correcting the additive methodological error.

*Keywords:* white noise, Hurst index, small samples, biometric data, methodological error.

### **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). P. 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 p. (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 p. (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: Iz-vo PGU, 2021. 60 p. (In Russ.)
5. Feder E. *Fraktaly [Fractals]*. Moscow: MIR, 1991. 254 p. (In Russ.)
6. 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 Pablishez, 2011. 552 p. (In Russ.)
7. Ivanov A.I., Gerasin V.Yu. Prospects for the use of autocorrelation statistical criteria as analogues of the Hurst index in processing small samples of economic and

- biometrics data. Trudy mezhdunarodnogo simpoziuma «Nadezhnost' i kachestvo» [Proceedings of the International Symposium "Reliability and Quality"] . Penza, 2023. 2. P. 59–61.
8. Ivanov A.I., Tarasov D.V. Additional correction of the additive methodological error of estimates of the Hurst index in the analysis of empirical data. Problemy informatiki v obrazovanii, upravlenii, ekonomike i tekhnike : sb. st. XXIII Mezhdunar. nauch.-tekhn. konf., posvyashch. 80-letiyu Penzenskogo gosudarstvennogo universiteta (g. Penza, 24–25 noyabrya 2023 g.) [Problems of computer science in education, management, economics and technology : collection of Articles XXIII International Scientific and Technical conf., dedicated. The 80th anniversary of Penza State University (Penza, November 24-25, 2023)]. Penza: Iz-vo PGU, 2023. P. 14-18.