

УДК 519.6; 004.8; 519.2

doi: 10.21685/2072-3040-2024-2-6

Численный алгоритм нейросетевого распознавания персистентного и антиперсистентного состояний рынка

А. И. Иванов¹, Д. В. Тарасов²

¹Пензенский научно-исследовательский электротехнический институт, Пенза, Россия

²Пензенский государственный университет, Пенза, Россия

¹ivan@pniei.penza.ru, ²tarasovdv@mail.ru

Аннотация. *Актуальность и цели.* В прошлом веке Хёрст предложил свой статистический показатель, исследуя историю разливов Нила. К сожалению, показатель Хёрста нечувствителен к знаку коэффициентов корреляции между анализируемыми данными и шкалой времени. Как результат только показателя Хёрста недостаточно для достоверного статистического анализа: данных рынка, данных сложных социальных систем, биометрических данных людей. Целью работы является синтез нейросетевого обобщения показателя Хёрста (нейрон Хёрста усиливается двумя дополнительными нейронами). Это позволяет различать персистентное и антиперсистентное состояния рынка, а также оценивать показатель когерентности переходных процессов в той или иной сложной системе. *Материалы и методы.* Предложено программно размечать данные рынка на участки, соответствующие «медвежьему» тренду падения цен или «бычьему» тренду повышения цен. Через численный эксперимент показано, что положительно коррелированные данные персистентного рынка всегда имеют более вероятное состояние «бычьего» повышения цен. Для антиперсистентного рынка наблюдается обратная ситуация преобладания «медвежьих» трендов падения цен. *Результаты.* Предложено оценивать показатель когерентности рынка, обусловленный его нестационарной компонентой, через учет статистик случайной смены персистентного состояния рынка на антиперсистентное. *Выводы.* Сделано предположение, что персистентные и антиперсистентные состояния и персистентное распределение эмпирического показателя Хёрста должны описываться в двух разных шкалах, каждая из которых получена зеркальным отображением от точки идеального белого шума.

Ключевые слова: показатель Хёрста, персистентные состояния рынка, антиперсистентные состояния рынка, малые выборки, оценка коэффициентов корреляции

Для цитирования: Иванов А. И., Тарасов Д. В. Численный алгоритм нейросетевого распознавания персистентного и антиперсистентного состояний рынка // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2024. № 2. С. 63–73. doi: 10.21685/2072-3040-2024-2-6

Numerical algorithm for neural network recognition of persistent and antipersistent market conditions

A.I. Ivanov¹, D.V. Tarasov²

¹Penza Scientific Research Electrotechnical Institute, Penza, Russia

²Penza State University, Penza, Russia

¹ivan@pniei.penza.ru, ²tarasovdv@mail.ru

Abstract. *Background.* In the last century, Hearst proposed his own statistical indicator, exploring the history of the Nile floods. Unfortunately, the Hurst index is insensitive to the sign of the correlation coefficients between the analyzed data and the time scale. As a re-

sult, the Hurst indicator alone is not enough for reliable statistical analysis: market data, data from complex social systems, biometric data of people. The purpose of the work is to synthesize a neural network generalization of the Hurst index (the Hurst neuron is amplified by two additional neurons). This makes it possible to distinguish between persistent and anti-persistent market conditions, as well as to evaluate the coherence indicator of transients in a particular complex system. *Materials and methods.* It is proposed to programmatically mark up market data into sections corresponding to a “bearish” trend of falling prices or a “bullish” trend of rising prices. Through numerical experiment, it is shown that positively correlated persistent market data always have a more likely state of “bullish” price increases. For the anti-persistent market, the opposite situation is observed with the predominance of “bearish” trends in falling prices. *Results.* It is proposed to evaluate the indicator of market coherence due to its non-stationary component by taking into account the statistics of the random change of the persistent state of the market to the anti-persistent one. *Conclusions.* It is assumed that the persistent and antipersistent states and the persistent distribution of the empirical Hurst index should be described in two different scales, each of which is obtained by a mirror image from the point of ideal white noise.

Keywords: Hurst index, persistent market conditions, antipersistent market conditions, small samples, estimation of correlation coefficients

For citation: Ivanov A.I., Tarasov D.V. Numerical algorithm for neural network recognition of persistent and antipersistent market conditions. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Fiziko-matematicheskie nauki = University proceedings. Volga region. Physical and mathematical sciences.* 2024;(2):63–73. (In Russ.). doi: 10.21685/2072-3040-2024-2-6

Введение

В прошлом веке английский исследователь Хёрст, занимаясь предсказанием разливов Нила, предложил использовать показатель [1–3], являющийся отношением размаха данных к их стандартному отклонению:

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

где N – размер выборки; R – размах выборки; σ – стандартное отклонение выборки; H – эмпирический степенной показатель Хёрста, который должен изменяться в интервале от 0,5 до 1,0 для предсказуемых персистентных систем и в интервале от 0,0 до 0,5 для антиперсистентных систем.

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

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

Одним из важных свойств показателя является то, что он оказался применим к большому числу фрактальных объектов реальной природы [4]. В том числе показатель оказался востребованным биржевыми брокерами [5].

Принципиальным недостатком показателя Хёрста является то, что он нечувствителен к знаку корреляционной связанности, анализируемых данных. Пользуясь им, можно уверенно оценивать то, насколько данные модули коррелированы, или то, насколько они по модулю корреляционной сцепленности отличаются от «белого» шума.

Для «белого» шума эмпирический и теоретический показатели Хёрста $H \approx 0,5$ [6] легко воспроизводимы численно, для этого достаточно восполь-

зваться любым программным генератором псевдослучайных чисел. Однако для отрицательно и положительно коррелированных данных малых выборок (например, $r \approx -0,5$ и $r \approx +0,5$) значения эмпирического показателя Хёрста (2) совпадают между собой [6].

Последнее является существенным недостатком эмпирического показателя Хёрста, данная работа является попыткой его устранения.

Взгляд на проблему со стороны биржевого брокера

Еще одной дополнительной технической проблемой анализа биржевых данных является то, что показатель Хёрста обычно оценивают на больших выборках. Ждать, когда накопится большой (представительный) объем реальных данных, не всегда возможно. В связи с этим актуальной оказывается задача моделирования данных на малых выборках с последующей компенсацией методической составляющей погрешности [7] из-за малого объема выборки.

Следует также отметить, что торговля биржевых брокеров [5] построена на интуитивном выделении «медвежьих» фрагментов снижения цен и выделении «бычьих» трендов роста цен. Реализация численного эксперимента может быть построена с использованием двух выборок x, y (объема $n = 21$), имеющих нормальное распределение $N(0,1)$. Для задания корреляционной сцепленности данных выполнялось последующее присваивание: $x := (x \mp 0,6y) \cdot 0,6$, где знак « $\mp$ » определяет отрицательную, а « $+$ » – положительную сцепленность данных. Следом строилась матрица из столбцов x, y и выполнялась сортировка матрицы по первому столбцу. Второй столбец в этом случае и представлял собой итоговую выборку для последующего анализа.

Результаты моделирования малых выборок в 21 опыт с отрицательной коррелированностью отображены на рис. 1.

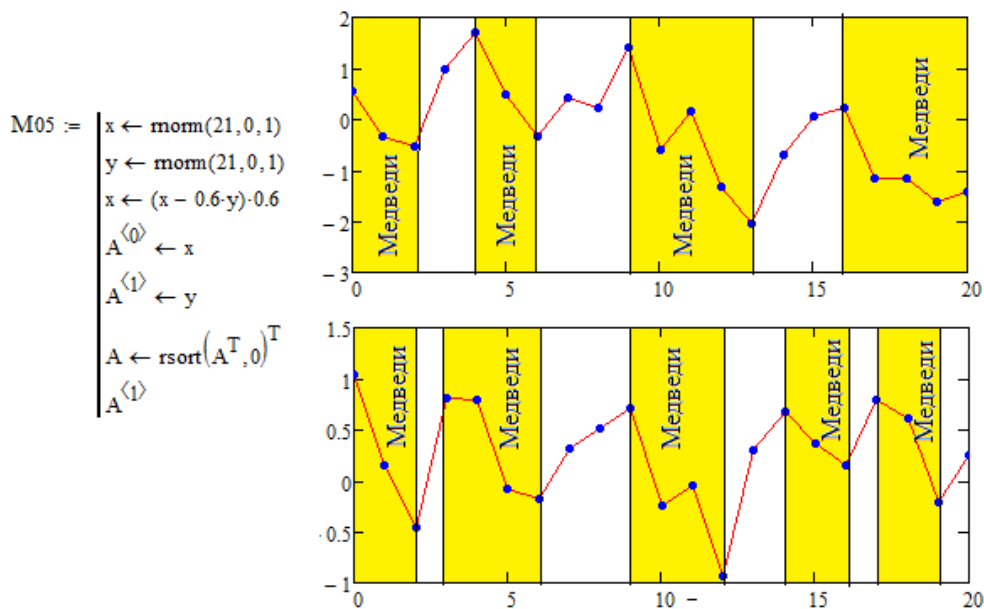


Рис. 1. Пример малых выборок с отрицательно коррелированными данными («медведи» преобладают)

В левой части рис. 1 размещена программная реализация численного эксперимента, написанная на языке MathCAD. В правой части рис. 1 представлены две реализации численного эксперимента. Из данных правой части рисунка видно, что для отрицательно коррелированных данных «медвежьих» участков по площади больше, чем «бычьих».

Ситуация кардинально меняется, если моделировать малые выборки с положительной корреляционной сцепленностью. Программная реализация этого варианта численного эксперимента представлена в левой части рис. 2.

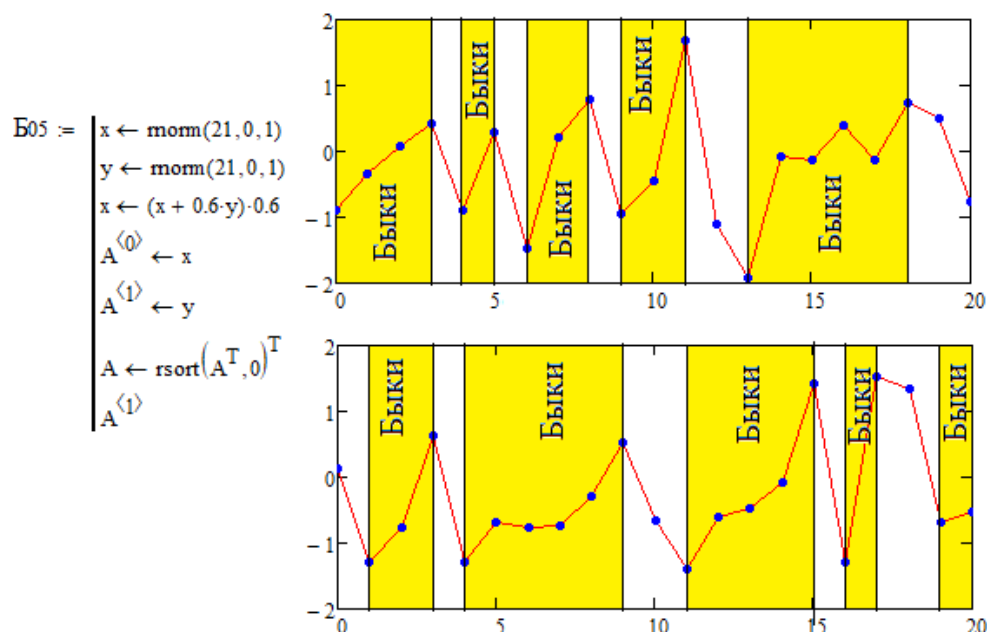


Рис. 2. Пример малых выборок с положительно коррелированными данными («быки» преобладают)

В правой части рис. 2 даны две реализации численного эксперимента. Из этих данных следует, что площадь участков роста цен «быки» оказывается существенно больше участков падения цен «медведи».

Новый дискретный корреляционный критерий быки/медведи для проверки гипотезы $r \approx -0,5$

В силу того, что статистики длин участков рынка «быки» и «медведи» существенно различаются при разном показателе корреляционной сцепленности данных, появляется реальная возможность использовать классический статистический критерий [8, 9]. Например, можно воспользоваться классической формулой конца XIX в. для оценки коэффициента корреляции на малой выборке в 21 опыт:

$$r(x, y) = \sum_{i=1}^{21} \frac{(x_i - E(x)) \cdot (y_i - E(y))}{\sigma(x) \cdot \sigma(y)}, \quad (3)$$

где $E(\cdot)$ – функционал вычисления математического ожидания; $\sigma(\cdot)$ – функционал вычисления стандартного отклонения малой выборки.

Результаты имитационного моделирования оценок (3) для малых выборок в 21 опыт приведены на рис. 3.

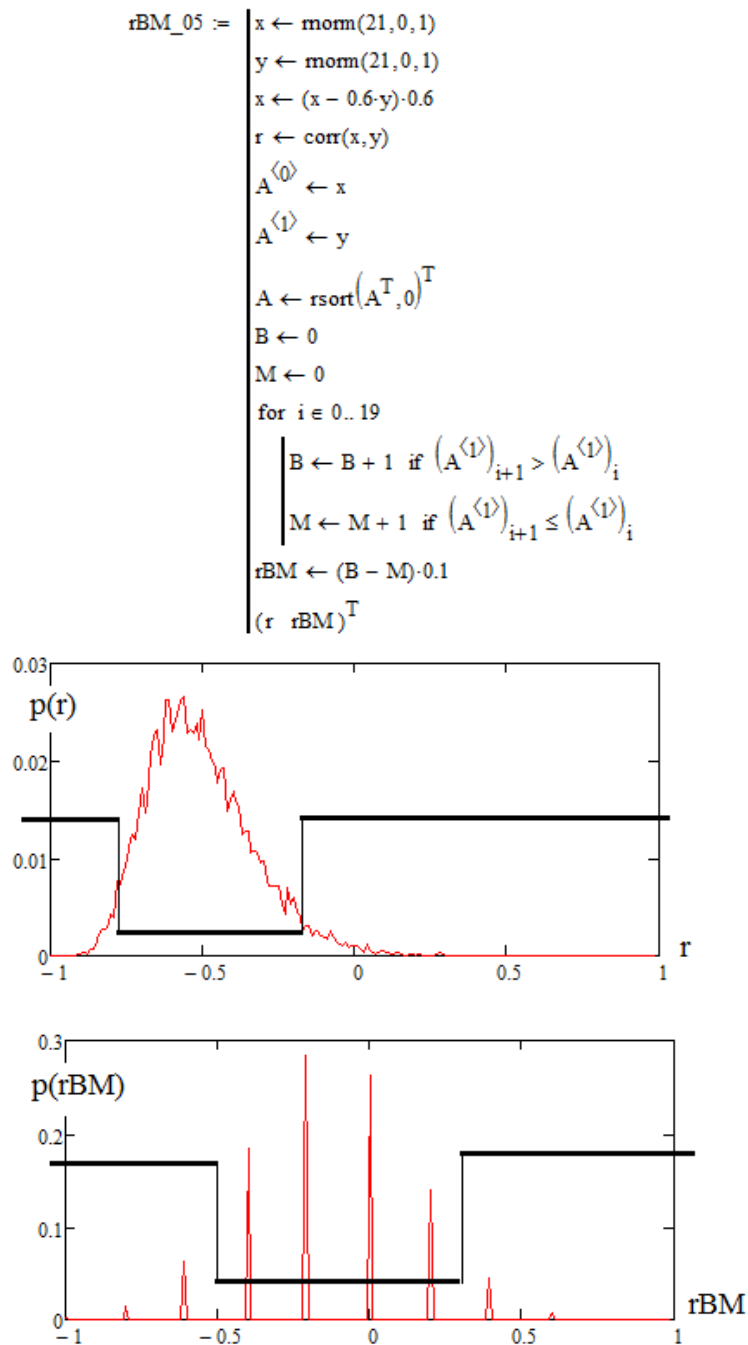


Рис. 3. Результаты имитационного моделирования классического непрерывного корреляционного функционала (3) и нового дискретного функционала разности площадей двух трендов «быки/медведи»

В силу того, что выборка мала, классический корреляционный функционал (3) дает значительный разброс откликов. Формально мы можем, опираясь на классический функционал (3), создать искусственный нейрон, который будет откликаться состоянием «0» при попадании значений в интервал от $-0,81$ до $-0,24$. При этом появление состояния «0» будет возникать с вероятностью $(1 - 0,131) = 0,869$. Последнее может быть интерпретировано как обнаружение значений корреляции, близкое к $r \approx -0,5$. Очевидно, что появление выходных состояний «1» с вероятностью $0,131$ следует рассматривать как ошибочное значение отклика синтезированного нейрона.

Определим количество трендов «быки» B как количество пар выборки y_i , отвечающих условию $y_{i+1} > y_i$, количество трендов «медведи» M – условию $y_{i+1} \leq y_i$. Тогда произведение количества трендов («быки» и «медведи») на шаг численно будет определять значение площади области, занимаемой соответствующими трендами.

Если мы будем рассматривать разность площадей, занимаемых трендами «быки» и «медведи», то мы получим еще один новый (ранее не используемый) корреляционный функционал [9]. В левой части рис. 3 дана программная реализация нового функционала. Результаты имитационного моделирования нового корреляционного функционала приведены в нижней правой части рис. 3. Как и в предыдущем случае, мы можем модифицировать новый функционал выходным бинарным квантователем. В итоге мы получим искусственный нейрон, подтверждающий состоянием «0» гипотезу значения коэффициента корреляции « $r \approx -0,5$ » с доверительной вероятностью $(1 - 0,133) = 0,867$.

Имитационное моделирование третьего нейрона Хёрста

В силу того, что показатель Хёрста (2) популярен при анализе статистик колебания цен на рынке, его целесообразно доработать до нейрона, применив к его данным бинарный квантователь. Результаты имитационного моделирования такой математической конструкции приведены на рис. 4.

Программная реализация численного эксперимента приведена в верхней части рис. 4. Она ориентирована на воспроизведение данных с отрицательной корреляционной сцепленностью $E(r) = -0,5$. Это обусловлено использованием разности в третьей строке программной реализации. Если в третьей строке программной реализации заменить знак минуса на плюс, то корреляционная сцепленность данных окажется положительной $E(r) = +0,5$. Это отражено в нижней части рис. 4.

Независимо от использованного знака в третьей строке программной реализации, распределение откликов функционала Хёрста не изменяется. Функционалы Хёрста нечувствительны к знаку корреляционной сцепленности анализируемых данных. Последнее означает, что мы можем добавить к функционалу Хёрста бинарный квантователь, который даст нам искусственный нейрон Хёрста. Этот нейрон может быть использован для проверки гипотезы « $H \approx 0,33$ » или антиперсистентного состояния рынка на малой выборке. Эта гипотеза будет подтверждаться выходным состоянием «0» с доверительной вероятностью $(1 - 0,138) = 0,862$. Очевидно, что такой доверительной вероятности недостаточно для практического использования.

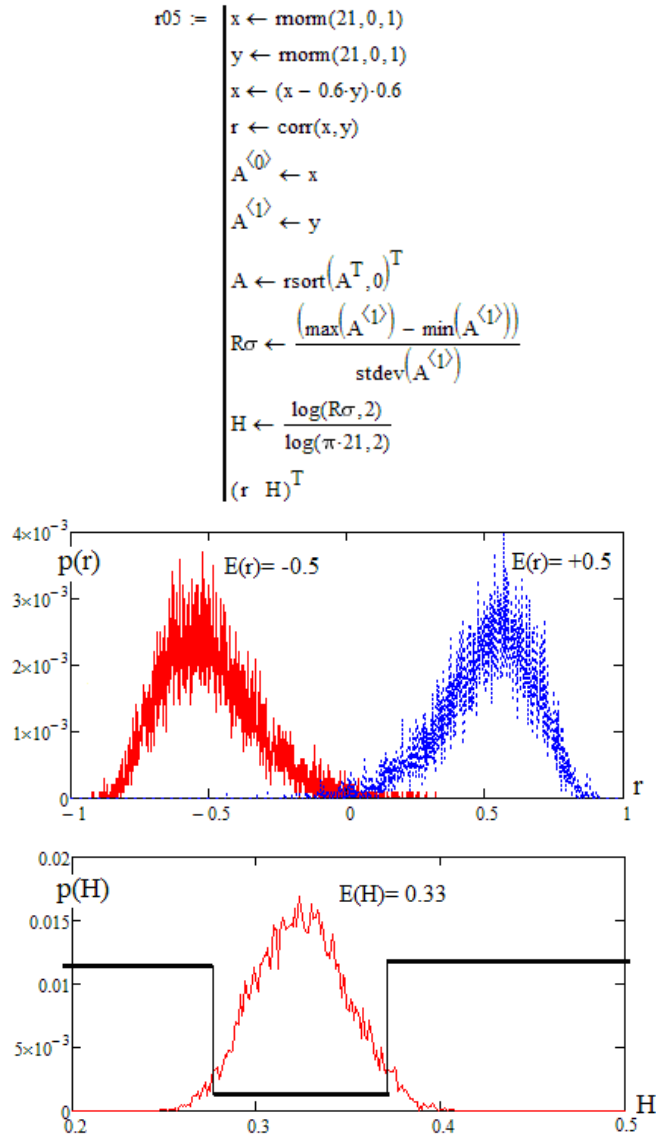


Рис. 4. Имитационное моделирование нейрона Хёрста для малых выборок в 21 опыт

Свертывание трех бинарных откликов искусственных нейронов

Каждый из рассмотренных выше искусственных нейронов имеет существенные вероятности ошибочных выходных состояний. Важно то, что появления в их откликах ошибочных состояний «1» почти независимы. Это позволяет объединить рассмотренные нейроны в сеть, которая будет откликаться избыточным кодом. Так, если нейросеть откликнется состоянием «000», то ее отклик можно рассматривать как тройное подтверждение проверяемой гипотезы « $H \approx 0,33$ ». Появление одного неверного состояния {«001», «010», «100»} также следует рассматривать как частичное подтверждение проверяемой гипотезы [9].

Фактически мы используем некоторый простейший механизм обнаружения и исправления ошибок [10]. Выигрыш от использования нейронной се-

ти в нашем случае легко оценивается в рамках гипотезы независимости данных. Численный эксперимент дал следующие оценки коэффициентов корреляции: $\text{corr}(r, rBM) \approx -0,124$; $\text{corr}(H, rBM) \approx 0,02$; $\text{corr}(r, H) \approx 0,05$. В таком случае для оценки выигрыша необходимо найти среднее геометрическое вероятности ошибок $\sqrt[3]{0,131 \cdot 0,133 \cdot 0,138} = 0,134$. Если корреляции между откликами полностью отсутствуют, то свертывание трехкратной избыточности приведет к существенному снижению вероятности появления ошибок $0,134 \cdot 0,134 = 0,018$. Как итог – наблюдаем существенный рост доверительной вероятности к нейросетевому решению $(1 - 0,018) = 0,982$. Начальная недостаточно высокая доверительная вероятность $0,862$ одиночного нейрона Хёрста существенно увеличивается до значения $0,982$ при проверке гипотезы сетью из трех искусственных нейронов.

Следует обратить внимание на то, что столь значительное повышение доверительной вероятности возникает только при полной независимости откликов искусственных нейронов. На практике это условие редко выполняется. Реальные отклики искусственных нейронов, как правило, имеют существенные корреляционные связи. Чем выше корреляционные связи, тем меньше будет выигрыш от свертывания кодовой избыточности [11]. При полной зависимости откликов сети искусственных нейронов улучшение вероятностей ошибок первого и второго рода возникнуть не может.

Две разные шкалы пересчета эмпирического показателя Хёрста для персистентных и антиперсистентных состояний рынка

В связи с тем, что эмпирический показатель Хёрста нечувствителен к знаку корреляционной сцепленности данных, должны существовать две разные шкалы оценок. Одна шкала показателя должна отражать антиперсистентные состояния рынка, а вторая шкала должна отражать персистентные состояния рынка. При этом обе шкалы должны быть зеркальными, как это показано на рис. 5.

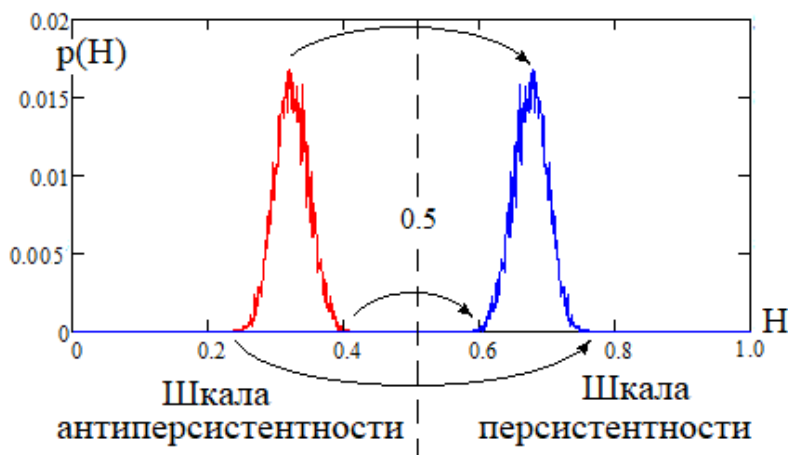


Рис. 5. Две шкалы распределения значений эмпирического показателя Хёрста

Из рис. 5 видно, что для значения показателя $H = 0,4$ для преимущественно отрицательно сцепленных данных должно преобразовываться в зна-

чение показателя $H = 0,6$ для преимущественно положительно корреляционно сцепленных данных. Для значения показателя $H = 0,24$ при отрицательной коррелированности по второй шкале должно быть значение $H = 0,76$.

Соответственно связь между двумя шкалами может быть определена следующей формулой:

$$\tilde{H} = |H - 0,5| + 0,5, \quad (4)$$

где H – первая шкала антиперсистентных распределений расстояний Хемминга со значениями менее 0,5; $\tilde{H}$ – вторая шкала для устойчивых персистентных распределений со значениями более 0,5.

Таким образом, введение двух шкал для пересчета эмпирического показателя Хёрста снимает проблему значительной методической ошибки при анализе данных рынка и иных биометрических данных [9]. При этом использование только показателя Хёрста, скорее всего, не дает возможности оценить, какой из двух шкал следует воспользоваться.

Список литературы

1. Калуш Ю. А., Логинов В. М. Показатель Хёрста 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. М. ; СПб. ; Киев : Вильямс, 2006. 408 с.
4. Федер Е. Фракталы. М. : Мир, 1991. 254 с.
5. Найман Э. Как покупать дешево и продавать дорого : пособие для разумного инвестора. М. : Альпина Паблишерз, 2011. 552 с.
6. Тарасов Д. В. Программное формирование одномерных эталонных данных малых выборок с заранее заданным показателем Хёрста // Безопасность информационных технологий : сб. науч. ст. по материалам V Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2023. Т. 1. С. 147–152.
7. Кузнецов В. Е., Иванов А. И., Герасин В. Ю. Устранение эффекта ошибочного наблюдения антиперсистентных состояний при оценках показателя Хёрста на малых выборках за счет подбора показателя логарифмирования // Безопасность информационных технологий : сб. науч. ст. по материалам V Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2023. Т. 1. С. 74–77.
8. Орлов Ю. Н., Осминин К. П. Нестационарные временные ряды. Методы прогнозирования с применением анализа финансовых и сырьевых рынков. М. : ЛЕНЛАНД, 2023. 384 с.
9. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок. Проверка гипотезы независимости : справочник. Пенза : Изд-во ПГУ, 2022. 218 с.
10. Морелос-Сарагоса Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение / пер. с англ. В. Б. Афанасьева. М. : Техносфера, 2006. 320 с.
11. Иванов А. И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD) : препринт. Пенза : Изд-во ПГУ, 2020. 36 с.

References

1. Kalush Yu.A., Loginov V.M. Hurst exponent and its hidden properties. *Sibirskiy zhurnal industrial'noy matematiki* = *Siberian Journal 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. 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*. Moscow; Saint Petersburg; Kiev: Izd-vo Vil'yams, 2006:408. (In Russ.)
4. Feder E. *Fraktaly* = *Fractals*. Moscow: Mir, 1991:254. (In Russ.)
5. Nayman E. *Kak pokupat' deshevo i prodavat' dorogo: posobie dlya razumnogo investora* = *How to buy low and sell high: a guide for the smart investor*. Moscow: Al'pina Publisherz, 2011:552. (In Russ.)
6. Tarasov D.V. Software generation of one-dimensional reference data of small samples with a predefined Hurst exponent. *Bezopasnost' informatsionnykh tekhnologiy: sb. nauch. st. po materialam V Vseros. nauch.-tekhn. konf.* = *Information technology security: proceedings of the 5th All-Russian scientific and engineering conference*. Penza: Izd-vo PGU, 2023;1:147–152. (In Russ.)
7. Kuznetsov V.E., Ivanov A.I., Gerasin V.Yu. Elimination of the effect of erroneous observation of antipersistent states when estimating the Hurst exponent in small samples by selecting a logarithm index. *Bezopasnost' informatsionnykh tekhnologiy: sb. nauch. st. po materialam V Vseros. nauch.-tekhn. konf.* = *Information technology security: proceedings of the 5th All-Russian scientific and engineering conference*. Penza: Izd-vo PGU, 2023;1:74–77. (In Russ.)
8. Orlov Yu.N., Osminin K.P. *Nestatsionarnye vremennye ryady. Metody prognozirovaniya s primeneniem analiza finansovykh i syr'evykh rynkov* = *Non-stationary time series. Forecasting methods using analysis of financial and commodity markets*. Moscow: LENLAND, 2023:384. (In Russ.)
9. 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: guidance*. Penza: Izd-vo PGU, 2022:218. (In Russ.)
10. Morelos-Saragosa R. *Iskusstvo pomekhoustoychivogo kodirovaniya. Metody, algoritmy, primeneniye* = *The art of noise-resistant coding. Methods, algorithms, application*. Transl. from Eng. V.B. Afanas'ev. Moscow: Tekhnosfera, 2006:320. (In Russ.)
11. Ivanov A.I. *Iskusstvennye matematicheskie molekuly: povyshenie tochnosti statisticheskikh otsenok na malykh vyborkakh (programmy na yazyke MathCAD): preprint* = *Artificial mathematical molecules: increasing the accuracy of statistical estimates on small samples (MathCAD programmes): preprint*. Penza: Izd-vo PGU, 2020:36. (In Russ.)

Информация об авторах / Information about the authors

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)

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 18.02.2024

Поступила после рецензирования и доработки / Revised 04.04.2024

Принята к публикации / Accepted 15.05.2024