

ПРОГНОЗ ВРЕМЕННОГО РЯДА С ПОМОЩЬЮ АППРОКСИМАЦИИ ФРАКТАЛЬНЫМ БРОУНОВСКИМ ДВИЖЕНИЕМ

В.В. БОНДАРЕНКО

Рассмотрена задача экстраполяции (прогноза) наблюдаемого временного ряда. Схема решения данной задачи — двухэтапная операция (определение тренда и функциональное преобразование) над исходным рядом, сводящая его к последовательности, параметры которой совпадают с параметрами гауссовых данных. Выполненные действия позволяют применить к преобразованному временному ряду процедуру линейного прогнозирования. Приведены результаты численного эксперимента, подтверждающие эффективность предложенного алгоритма, иллюстрирующего качество функционирования предложенных алгоритмов, в частности, компьютерное имитационное моделирование базового случайного процесса — фрактального броуновского движения. Подробно рассмотрено фрактальное броуновское движение и на конкретных четырех примерах построен удовлетворительный прогноз и оценены его параметры.

ВВЕДЕНИЕ

Анализ временного ряда x_1, \dots, x_n предполагает наличие математической модели, т.е. последовательности случайных величин ξ_1, \dots, ξ_n , для которой наблюдаемое значение x_k есть реализация ξ_k . Дискретная модель определяется формулой

$$\xi_k = \Phi(\xi_{k-p}, \dots, \xi_{k-1}, \varepsilon_1, \dots, \varepsilon_r),$$

где $\varepsilon_1, \dots, \varepsilon_r$ — «источники случайности». Непрерывная модель определяется как случайный процесс $X(t)$ с известными характеристиками, для которого x_k есть значение случайной величины $X\left(\frac{k}{n}\right)$, $k = 1, \dots, n$.

В данной работе для преобразованных исходных данных в качестве модели выбрано фрактальное броуновское движение, что позволило для реальных примеров построить оптимальный прогноз и оценить его качество.

ПОСТАНОВКА ЗАДАЧИ

Одной из основных проблем анализа наблюдаемого временного ряда x_1, x_2, \dots, x_n является его прогноз, т.е. построение оценки \hat{x}_{n+r} . Если известна математическая модель этого ряда — случайный процесс $X(t)$, $x_k = X\left(\frac{k}{n}\right)$, $EX(t) = 0$, то наилучший прогноз (экстраполяция) является условным средним

$$\hat{X}(t+s) = E(X(t+s)|X(\tau), 0 \leq \tau \leq t).$$

Для гауссовой последовательности этот прогноз совпадает с линейным ([1], с. 324): если обозначить

$$\xi = (\xi_1, \dots, \xi_m), \quad \eta = (\xi_{m+1}, \dots, \xi_{m+r}),$$

то

$$\hat{\eta} = DS^{-1}\xi, \quad (1)$$

где S, D — блоки корреляционной матрицы случайного вектора $(\xi; \eta)$:

$$s_{jk} = E\xi_j\xi_k; \quad 1 \leq j, k \leq m;$$

$$d_{jk} = E(\xi_{m+j}\xi_k), \quad 1 \leq j \leq r, \quad 1 \leq k \leq m.$$

Если предположение о гауссовости исходного временного ряда (точнее, его приращений $y_k = x_{k+1} - x_k$) не имеет места, то можно перейти к новой последовательности

$$z_k = f(y_k),$$

где функция f подбирается так, чтобы некоторые характеристики $\{z_k\}$ совпадали с гауссовыми. Схема такой конструкции (где f — степенная функция) предложена в работе [2] и подробно рассмотрена ниже. Полагая z_k приращениями фрактального броуновского движения

$$z_k = \sigma \left(B_H \left(\frac{k+1}{n} \right) - B_H \left(\frac{k}{n} \right) \right)$$

оценим параметры H и σ по алгоритму, приведенному в [3], что позволит построить прогноз по формуле (1), где

$$\xi = (z_1, \dots, z_m), \quad \eta = (z_{m+1}, \dots, z_{m+r}),$$

$$s_{jk} = \frac{|j-k+1|^{2H} + |j-k-1|^{2H}}{2} - |j-k|^{2H},$$

$$d_{jk} = \frac{|m+j-k+1|^{2H} + |m+j-k-1|^{2H}}{2} - |m+j-k|^{2H}.$$

Окончательно, прогнозируемые приращения исходного временного ряда находятся по формуле

$$y_{m+j} = f^{-1}(z_{m+j}), \quad j = 1, 2, \dots, r.$$

Цель работы — построение прогноза для временного ряда. Метод прогнозирования состоит в следующем: наблюдаемый временной ряд преобразуется в случайную последовательность, соответствующую фрактальному броуновскому движению. В свою очередь, для этой последовательности, распределение которой предполагается гауссовым, строится оптимальная экстраполяция. Реализация предложенного алгоритма прогнозирования для реальных временных рядов свидетельствует о его эффективности.

**ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ — КАЧЕСТВО ПРОГНОЗА
ДЛЯ ФРАКТАЛЬНОГО БРОУНОВСКОГО ДВИЖЕНИЯ**

Применим формулу (1) для экстраполяции приращений фрактального броуновского движения (fBm) с известным показателем Харста H . Данные получены имитационным моделированием fBm, основанным на его представлении стохастическим интегралом [4].

$$B_H(t) = c_H \left(\int_{-\infty}^0 ((t-s)^\alpha - (-s)^\alpha) dw(s) + \int_0^t (t-s)^\alpha dw(s) \right),$$

где $\alpha = H - \frac{1}{2}$, а нормирующая константа

$$c_H = \frac{\sqrt{2H\Gamma(2H) \sin \pi H}}{\Gamma\left(H + \frac{1}{2}\right)}.$$

В процессе моделирования интеграл заменяется суммой. Исследования по точности дискретизации, изложенные в [3], [5], приводят к формуле для $t = \frac{k}{n}$:

$$B_H\left(\frac{k}{n}\right) = \frac{c}{n^H} \left(\sum_{j=1}^{n^{5/2}} ((k+j)^\alpha - j^\alpha) \varepsilon_j + \sum_{i=0}^{k-1} (k-i)^\alpha \varepsilon_{n^2+i+1} \right), \quad k = 0, 1, 2, \dots, n.$$

Пусть $y_k = B_H\left(\frac{k+1}{n}\right) - B_H\left(\frac{k}{n}\right)$, $k = 1, 2, \dots, n-1$. Полагая обучающей выборкой вектор $\xi = (y_1, \dots, y_m)$, т.е. считая неизвестным вектор $\eta = (y_{m+1}, \dots, y_{m+r})$ построим прогноз $\hat{\eta}$.

Вычислительный эксперимент выполнен для трех значений параметра H : 0,3; 0,6; 0,8. Объем выборки $n=1023$. Объем обучающей выборки $m=1000$. В табл. 1 приведены наблюдаемые 8 значений приращений y_{m+1}, \dots, y_{m+8} .

Таблица 1. Наблюдаемые значения полученных приращений для размерности $n=1008$

| H | y_{1001} | y_{1002} | y_{1003} | y_{1004} | y_{1005} | y_{1006} | y_{1007} | y_{1008} |
|-----|------------|------------|------------|------------|------------|------------|------------|------------|
| 0,3 | 1,715 | -0,687 | 0,486 | -0,604 | -0,896 | -0,194 | 1,349 | -1,094 |
| 0,6 | 1,783 | -2,321 | 0,792 | -0,241 | -0,262 | -0,138 | -0,289 | 1,467 |
| 0,8 | 1,004 | -1,599 | 0,406 | 1,969 | -1,098 | 0,035 | 0,735 | -0,221 |

Результаты прогноза приращений fBm-значения $\hat{y}_{m+1}, \dots, \hat{y}_{m+r}$ — вычислены согласно формуле (1) и сведены в таблицы 2–4. Размерность матрицы S равна $m \times m$, матрица D состоит из m столбцов и r строк.

Таблица 2. Прогнозируемые значения полученных приращений при $H = 0,3$

| H | \hat{y}_{1001} | \hat{y}_{1002} | \hat{y}_{1003} | \hat{y}_{1004} | \hat{y}_{1005} | \hat{y}_{1006} | \hat{y}_{1007} | \hat{y}_{1008} | \hat{y}_{1009} | \hat{y}_{1010} |
|---------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| $r = 1$ | -2,519 | | | | | | | | | |
| $r = 2$ | -0,458 | 0,125 | | | | | | | | |
| $r = 4$ | -0,147 | 0,008 | 1,25 | 1,671 | | | | | | |
| $r = 8$ | 1,089 | -2,543 | 0,993 | 1,147 | -2,781 | 0,978 | 1,016 | -0,471 | -1,279 | 0,252 |

Таблица 3. Прогнозируемые значения полученных приращений при $H = 0,6$

| H | \hat{y}_{1001} | \hat{y}_{1002} | \hat{y}_{1003} | \hat{y}_{1004} | \hat{y}_{1005} | \hat{y}_{1006} | \hat{y}_{1007} | \hat{y}_{1008} |
|---------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| $r = 1$ | 1,523 | | | | | | | |
| $r = 2$ | 1,598 | -2,102 | | | | | | |
| $r = 4$ | 1,624 | -2,447 | 0,647 | -0,351 | | | | |
| $r = 8$ | 1,689 | -2,205 | 0,429 | -0,197 | -0,314 | -0,367 | -0,302 | 1,299 |

Таблица 4. Прогнозируемые значения полученных приращений при $H = 0,8$

| H | \hat{y}_{1001} | \hat{y}_{1002} | \hat{y}_{1003} | \hat{y}_{1004} | \hat{y}_{1005} | \hat{y}_{1006} | \hat{y}_{1007} | \hat{y}_{1008} |
|---------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| $r = 1$ | 1,025 | | | | | | | |
| $r = 2$ | 0,89 | -1,502 | | | | | | |
| $r = 4$ | 0,95 | -1,578 | 0,372 | 1,756 | | | | |
| $r = 8$ | 1,247 | -1,632 | 0,221 | 2,05 | -1,165 | 0,158 | 0,901 | -0,182 |

Представленные результаты приводят к следующим выводам.

- Антиперсистентный ($H < 0,5$) процесс fBm не допускает удовлетворительного прогноза. Результат ожидаемый, и качественные свойства такого «очень зашумленного» процесса рассмотрены в монографиях [6, 7].
- Для персистентного ($H > 0,5$) процесса fBm имеет место удовлетворительный результат прогнозирования на 1, 2, 4 шага и с большей погрешностью для 8 шагов.

ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ — КАЧЕСТВО ПРОГНОЗА ДЛЯ ГАУССОВОЙ АППРОКСИМАЦИИ

Для экстраполяции реально наблюдаемого временного ряда x_1, \dots, x_n необходимо, в первую очередь, построить адекватную математическую модель. Предложенный в работе [2] алгоритм такого построения состоит в следующем. Рассмотрим приращения $y_k = x_{k+1} - x_k$ и обозначим:

$$R_{jn} = \frac{1}{n-1} \sum_{k=1}^{n-1} |y_k|^j, \quad j = 1, 2. \quad (2)$$

Предположим, что выполнены следующие условия:

- последовательность $\{y_k\}$ стационарна в узком смысле, т.е. $E y_k = 0$, $D y_k = \text{const}$, $P\{y_j < u, y_k < v\} = P\{y_{j+m} < u, y_{k+m} < v\}$;

- с вероятностью 1 существуют пределы

$$\lim_{n \rightarrow \infty} \frac{R_{1n}^2}{R_{2n}} \rightarrow d = \frac{(E|y_k|)^2}{(Ey_k^2)}$$

Тогда последовательность

$$d_n = \frac{R_{1n}^2}{R_{2n}} \rightarrow d = \frac{(E|y_k|)^2}{(Ey_k^2)}, P=1$$

и для гауссовых приращений $d = \frac{2}{\pi}$. Критерием гауссовости может служить значение выборочного «коэффициента эксцесса» d_n . Если d_n значимо отличается от $\frac{2}{\pi}$, аппроксимируем временной ряд y_1, \dots, y_n гауссовой последовательностью z_1, \dots, z_n согласно следующему алгоритму. Положим

$$y_k = \text{sgn } y_k |z_k|^\lambda, \lambda > 0. \tag{3}$$

Тогда $d_n = \frac{\left(\frac{1}{n} \sum_{k=1}^n |z_k|^\lambda\right)^2}{\frac{1}{n} \sum_{k=1}^n |z_k|^{2\lambda}}$ и $d = \lim_{n \rightarrow \infty} d_n$ равен отношению соответствующих

математических ожиданий. Так как для гауссовой случайной величины $\zeta \in N(0; \sigma^2)$ выполняется равенство

$$E|\zeta|^\alpha = \frac{2^{\frac{\alpha}{2}}}{\sqrt{\pi}} \sigma^\alpha \Gamma\left(\frac{\alpha+1}{2}\right),$$

получаем

$$d = \frac{1}{\sqrt{\pi}} \frac{\Gamma^2\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\lambda + \frac{1}{2}\right)}.$$

Приравняв правую часть к d_n , найдем соответствующее λ (единственное, в силу убывания d как функции λ). Таким образом, предложенная аппроксимация приводит к следующей модели исходного временного ряда:

$$y_k = \sum_{j=1}^k \text{sgn } y_j \cdot |z_j|^\lambda.$$

Из предположения о стационарности последовательности $\{y_k\}$ следует стационарность $\{z_k\}$, т.е. $Ez_k = 0$, $Dz_k = \text{const}$, $E(z_j z_k) = Dz_j \cdot f\left(\frac{k-j}{n}\right)$, $f(0) = 1$, $f(s)$ — убывающая функция.

Аппроксимируя f степенной функцией

$$\frac{\left(s + \frac{1}{n}\right)^{2H} + \left(s - \frac{1}{n}\right)^{2H}}{2} - s^{2H},$$

приходим к аппроксимации $\{x_k\}$ фрактальным броуновским движением

$$z_k = \sigma \left(B_H \left(\frac{k}{n} \right) - B_H \left(\frac{k-1}{n} \right) \right), \quad x_k = \sigma^\lambda \sum_{j=1}^k \operatorname{sgn} y_j \left| B_H \left(\frac{j}{n} \right) - B_H \left(\frac{j-1}{n} \right) \right|^\lambda,$$

что позволяет исследовать данный временной ряд методами, разработанными для fBm.

В качестве примеров рассмотрим следующие реальные данные. Первый пример — ежедневные данные солнечной активности за 2008 год (366 данных, [8] (рис. 1)). Первичная обработка исходного временного ряда состоит в выделении его тренда. Разобьем данные S_1, \dots, S_{366} на 6 временных окон $\Delta_1, \dots, \Delta_6$ с числом элементов $n = 61$ в каждом окне. Выборочное среднее в каждом окне Δ_r :

$$m_r = \frac{1}{n} \sum_{k=1}^n S_{k+n(r-1)}, \quad r = 1, \dots, 6.$$

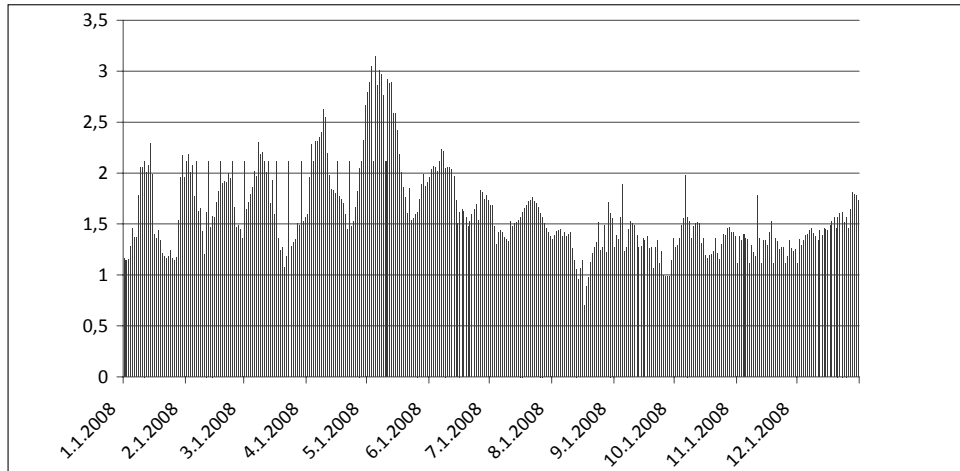


Рис. 1. Данные солнечной активности за 2008 год, указаны в W (число Вольфа)

Аппроксимируем тренд в Δ_r линейной функцией, полагая

$$M_r(k) = a_r(k - n(r-1)) + b_r, \quad n(r-1) < k \leq nr,$$

где коэффициенты a_r, b_r удовлетворяют системе 11 уравнений

$$a_r \left(\frac{n+1}{2} \right) + b_r = m_r, \quad a_r n + b_r = b_{r+1}, \quad (4)$$

т.е. ордината прямой в центре Δ_r равна m_r и отрезки прямых стыкуются в точках $k = nr$ (общий график тренда-ломаная без точек разрыва). Дополнив (4) уравнением $a_6 n + b_6 = S_{366}$ получим замкнутую систему, решаемую рекуррентно.

Результаты вычислений приведены в табл. 5.

Таблица 5. Значения коэффициентов линейного тренда

| r | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|--------|-------|---------|--------|---------|-------|
| m_r | 1,66 | 1,88 | 1,99 | 1,41 | 1,36 | 1,39 |
| a_r | -0,023 | 0,029 | -0,0245 | 0,0051 | -0,0066 | 0,073 |
| b_r | 2,36 | 0,98 | 2,75 | 1,255 | 1,563 | 1,163 |

Для преобразованного временного ряда

$$x_k = S_k - M_r(k), \quad \frac{1}{366} \sum_{k=1}^{366} x_k = -0,013$$

рассмотрим приращения $y_k = x_{k+1} - x_k$. Применяя алгоритм аппроксимации, получим

$$d = 0,5, \quad \lambda = 1,4,$$

то есть

$$z_k = \operatorname{sgn} y_k \cdot |y_k|^{0,7}, \quad k = 1, \dots, 365.$$

Оценивание параметра H приводит к значению $\hat{H} = 0,6$. Для прогнозирования выберем объём обучающей выборки $m=355$. Результаты прогноза для $r = 1; 2; 4$ приведены в табл. 6.

Таблица 6. Сравнение реальных и прогнозируемых значений

| | | | | |
|-------------------------------|-----------------|-----------------|-----------------|-----------------|
| Наблюдаемые $u_k = 100z_k$ | -5,8 | -10,0 | 9,1 | 14,3 |
| Прогноз | \hat{u}_{356} | \hat{u}_{357} | \hat{u}_{358} | \hat{u}_{359} |
| $r = 1$ | -5,91 | | | |
| $r = 2$ | -5,57 | -9,52 | | |
| $r = 4$ | -6,1 | -9,73 | 8,8 | 14,6 |

Определим ошибку δ_r прогноза на r шагов формулой

$$\delta_r = \frac{1}{r} \sum_{j=1}^r \left| \frac{u_j - \hat{u}_j}{u_j} \right|,$$

откуда

$$\delta_1 = 0,019; \quad \delta_2 = 0,044; \quad \delta_4 = 0,033,$$

то есть прогноз на 4 шага является удовлетворительным.

Второй пример — финансовые данные — значения банковской процентной ставки $S(t)$, за 2010–2012 годы, всего 337 данных [9]. При известной процентной ставке a НБУ ($a < S(t)$) определим процесс $X(t)$ формулой

$$S(t) = a + b \exp\{X(t)\}, \quad \text{т.е.} \quad X(t) = \ln \frac{S(t) - a}{S(0) - a}, \quad a = 0,775,$$

где коэффициент b выбран так, чтобы $S(0) = 0$. Для дискретного времени

$$t = \frac{k-1}{n}$$

$$x_k = X\left(\frac{k-1}{n}\right) = \ln \frac{S_k - a}{S_1 - a}, \quad x_1 = 0.$$

Структура данных (рис. 2) предполагает разбиение этих данных на два временных окна Δ_1, Δ_2 . В первом окне ($1 \leq k \leq 165$) $m_1 \approx 0$, во втором ($1 \leq k \leq 337$) аппроксимируем тренд кривой второго порядка $M_r(k)$.

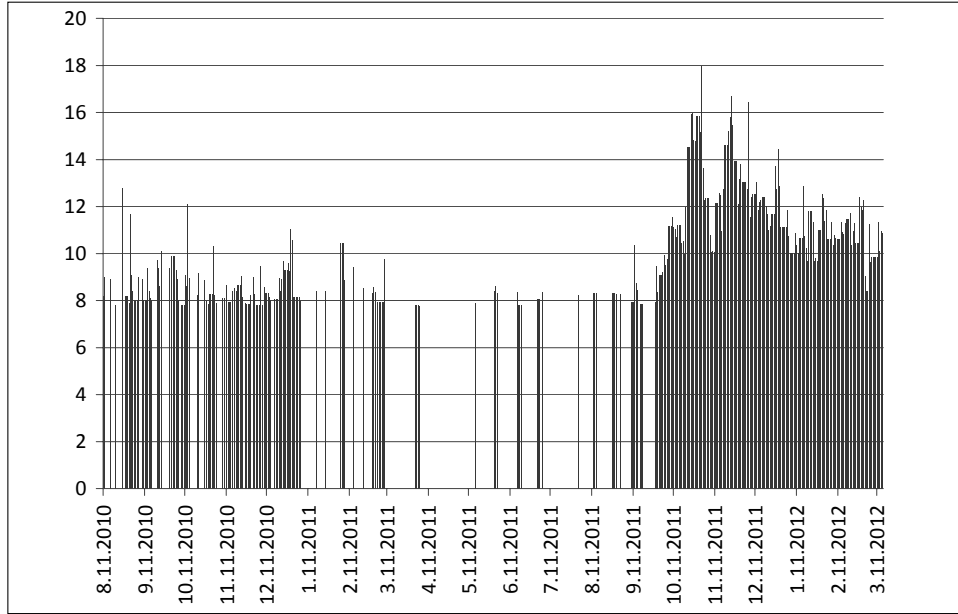


Рис. 2. Значения среднебанковской процентной ставки за 2010-2012 года, %

Полагая $u_k = x_k$ в Δ_1 ; $u_k = x_k - M_r(k)$ в Δ_2 , рассмотрим приращения $y_k = u_{k+1} - u_k$.

Вычисляя статистики, определенные формулой (2), получим

$$R_{1n} = 0,49; \quad R_{2n} = 0,72; \quad d_n = \frac{R_{1n}^2}{R_{2n}} = 0,334.$$

Так как d_n значительно отличается от $\frac{2}{\pi} \approx 0,637$, перейдем к новой последовательности $\{z_k\}$, определенной формулой (3), где λ определяется из уравнения

$$\frac{1}{\sqrt{\pi}} \frac{\Gamma^2\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\lambda + \frac{1}{2}\right)} = 0,334; \quad \lambda = 2,$$

то есть $y_k = \text{sgn } y_k |z_k|^2$, $z_k = \text{sgn } y_k \sqrt{|y_k|}$.

Предполагая, что преобразованные данные z_1, \dots, z_{336} образуют приращения fBm, оценим параметр H . Применяя алгоритм оценивания, предложенный в работе [3], получим $\hat{H} = 0,7$.

В табл. 7 приведены наблюдаемые значения преобразованных выборочных значений z_{329} , z_{330} , z_{331} , z_{332} и результаты прогноза по формуле (1) для объема обучающей выборки $m = 328$ на 1, 2, 4 шага.

Таблица 7. Результаты прогноза для обучающей выборки

| | | | | |
|----------------------------|-----------------|-----------------|-----------------|-----------------|
| Наблюдаемое значение z_k | 0,68 | 1,52 | -1,1 | 0,89 |
| Прогноз | \hat{z}_{329} | \hat{z}_{330} | \hat{z}_{331} | \hat{z}_{332} |
| $r = 1$ | 0,62 | | | |
| $r = 2$ | 0,76 | 1,46 | | |
| $r = 4$ | 0,72 | 1,47 | -0,95 | 1,12 |

Из полученных значений $\delta_1 = 0,09$, $\delta_2 = 0,079$ и $\delta_4 = 0,122$ следует, что удовлетворительным является только прогноз на два шага.

ВЫВОДЫ

В данной работе рассмотрен метод экстраполяции временного ряда, базирующийся на его функциональном преобразовании. Это преобразование сводит исходный временной ряд к новой последовательности, значения которой предполагаются случайными величинами, соответствующими фрактальному броуновскому движению. Предложенный алгоритм преобразования позволяет оценить параметр Харста и построить оптимальный прогноз для преобразованной последовательности, что в силу обратимости преобразования, приводит к прогнозу исходного временного ряда. Вычислительный эксперимент позволил сравнить качество линейных прогнозов для гауссового временного ряда с известными параметрами и аппроксимирующей последовательности, у которой лишь одна характеристика (d_n) совпадает с гауссовой. Результат сравнения свидетельствует, что предложенная аппроксимация временного ряда фрактальным броуновским движением применима для краткосрочного прогнозирования.

ЛИТЕРАТУРА

1. Ширяев А.Н. Вероятность. — М.: Наука, 1989. — 640 с.
2. Бондаренко В.В. Аппроксимация временного ряда степенной функцией фрактального броуновского движения // Проблемы управления и информатики. — 2013. — № 3. — С. 113–116.
3. Бондаренко В.В. Итерационный алгоритм оценивания параметров фрактального броуновского движения // Проблемы управления и информатики. — № 4. — 2012. — С. 28–33.
4. Mandelbrot B.B., van Ness I.W. The Fractional Brownian motion, fractional noises and applications. — Philadelphia: SIAM Review, 1968. — **10**, № 4. — 1325 p.
5. Coeurjolly J.-F. Simulation and identification of the fractional Brownian motion: A bibliographical and comparative study // Journal of statistical software. — 2000. — **5**. — issue 7. — P. 1–52.
6. Федер Е. Фракталы. — М.: Мир, 1991. — 258 с.
7. Mandelbrot B.B., Hudson R.L. The Misbehavior of Market / A Fractal View of Risk, Ruin and Reward. — NY: Basic Books, 2006. — 368 p.
8. The National Geophysical Data Center. — <ftp.ngdc.noaa.gov>.
9. Сайт Национального банку України. — <http://bank.gov.ua/control/uk/index>.

Поступила 12.01.2013