Пропонується метод селективного зіставлення зі зразком для прогнозування знаків приростів фінансових часових рядів. Підхід базується на індексації часових рядів для знаходження в їх динаміці подібних ділянок на основі методу К-найближчих сусідів, а також селективного групування цих ділянок за знаками приростів, які спостерігаються після їх завершення

Ключові слова: часовий ряд, прогнозування, індексація, знак приросту, зіставлення зі зразком

Предлагается метод селективного сопоставления с образцом для прогнозирования знаков приростов финансовых временных рядов. Подход базируется на индексации временных рядов для определения в их динамике подобных участков на основе метода К-ближайших соседей, а также селективной группировки этих участков в зависимости от знаков приростов, которые наблюдаются после их завершения

Ключевые слова: временной ряд, прогнозирование, индексация, знак прироста, сопоставление с образцом

УДК 004:519.2

DOI: 10.15587/1729-4061.2015.54812

ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ МЕТОДОМ СЕЛЕКТИВНОГО ЗІСТАВЛЕННЯ ЗІ ЗРАЗКОМ

О. Ю. Кучанський
Кандидат технічних наук, доцент*
E-mail: kuczanski@gmail.com

А. О. Білощицький
Доктор технічних наук,
професор, завідувач кафедри*
E-mail: bao1978@gmail.com

*Кафедра інформаційних технологій
Київський національний
університет будівництва і архітектури
пр. Повітрофлотський, 31, м. Київ, Україна, 03680

1. Вступ

У фінансовій сфері часто розглядається проблема прогнозування короткочасної тенденції часового ряду без розрахунку безпосередньо прогнозних значень. Це пов'язано з тим, що фінансовим часовим рядам часто притаманний нестійкий характер коливань. Такі ряди можуть бути близькими до випадкових. В цьому випадку будують моделі, які б дозволяли розрахувати прогнози знаків приростів значень часового ряду на одну точку вперед з необхідною максимальною точністю. Особливістю таких моделей і методів є можливість знаходження достатньо стійкої залежності поточного спостереження від попередніх. На підставі цієї залежності розраховується прогноз знаку приросту на наступну точку, тобто з періодом 1. Моделі такого типу зазвичай застосовуються для визначення напрямку руху часових рядів валютних пар і можуть використовуватися для визначення точок зміни тенденцій. Для прогнозування знаків приростів застосовують специфічні моделі та методи. Слід зазначити, що перед реалізацією таких моделей необхідно дослідити вхідний фінансовий часовий ряд на абсолютну випадковість. Це може бути реалізовано на основі фрактального аналізу або з використанням критеріїв поворотних точок, розподілу довжини фази тощо.

Основними етапами при побудові прогнозу в цьому випадку є:

– прогнозна ретроспекція (тобто встановлення об'єкту прогнозування, що складається з передпрогнозного аналізу об'єкту та оцінки його параметрів);

– вибір методів прогнозування та побудова моделі прогнозування і її формалізація;

– побудова проспекції (розрахунок прогнозу на визначений період);

– оцінка прогнозу і його верифікація.

Традиційні математичні моделі не дозволяють отримати необхідну точність прогнозування фінансових часових рядів, які близькі до випадкових. У зв'язку із цим, розробка моделей та методів, які призначені для прогнозування знаків приростів таких часових рядів має важливе наукове й практичне значення. Дослідження спрямоване на розробку такого методу, який дозволяв би виконувати прогноз знаку приросту часового ряду з максимальною точністю. При цьому може використовуватися апарат фрактального аналізу інформації, підходи інтелектуального аналізу часових рядів, зокрема індексації за методами найближчого сусіда і К-найближчих сусідів тощо.

2. Аналіз літературних даних і постановка проблеми

Прогнозування фінансових часових рядів є класичною задачею інтелектуального аналізу даних. На сьогоднішній день існує багато досліджень в сфері Time-Series Data Mining і прогнозування на основі нейронних мереж [1], генетичних алгоритмів [2, 3], нечітких методів [4], класичних економетричних підходів [5, 6]. Також існує багато моделей і відповідних методів, які дозволяють розраховувати оцінки прогнозів із заданим горизонтом як стаціонарних, так і нестаціонарних часових рядів [7–9]. Серед них традиційно застосовуються: AR-, MA- та ARMA-моделі [10]. Проте проблема побудови проспекції достатньої точності для нестаціонарних часових рядів, які близькі до випадкових (курсові пари і т. д.), залишається вивченою не повною мірою. Для цих цілей можуть використовуватися такі моделі: адаптивна поліноміальна модель прогнозування часових рядів з нестабільним характером коливань [6], комбінована модель прогнозування знаків приростів на основі плинних середніх [11, 12], наївні моделі тощо. Крім вказаних моделей, для цілей короткострокового прогнозування часових рядів такого типу можуть застосовуватися моделі, які базуються на аналізі подібностей в поведінці часових рядів, кластеризації, індексації тощо [13–15]. Такі моделі та методи можуть використовуватися для побудови проспекцій фінансових часових рядів, оскільки їх поведінка, як правило, характеризується наявністю квазіциклів. Для побудови моделей прогнозування за даними методами необхідно робити припущення, що поведінка часового ряду повторюється. Тобто якщо знайти і зафіксувати деяку ділянку часового ряду, яка подібна до ділянки, яка передує прогнозу, то поведінка вхідного часового ряду після цих ділянок буде наближено повторюватись. Подібність часових рядів визначається на основі деякої міри подібності. Як правило, ця міра метрична, проте є спроби використання і неметричних відстаней. Крім даної гіпотези, необхідно враховувати, що часовий ряд повинен мати пам'ять про свою ретроспекцію. Слід зазначити, що побудувати оцінку прогнозу для рядів, які близькі до випадкових, означає передусім знайти знак приросту часового ряду на одну точку вперед. Для виконання біржових операцій на валютному ринку цього може бути достатньо. В цьому дослідженні саме реалізована задача знаходження знаку приросту фінансового часового ряду на основі індексації ряду і аналізу подібностей, які ідентифікуються.

3. Ціль і задачі дослідження

Ціллю дослідження є побудова моделі і методу для прогнозування знаків приростів фінансових часових рядів, що демонструють поведінку, яка близька до випадкової на основі аналізу подібних ділянок цих рядів або зіставлення зі зразком.

Для досягнення мети були поставлені такі завдання:

– проведення процедури фрактального аналізу вхідного фінансового часового ряду для визначення характеристик часового ряду та можливості застосування моделі прогнозування знаків приростів часових рядів;

– проведення індексації часового ряду на основі методу К-найближчих сусідів з мірою подібності Евкліда та Хемінга (залежно від представлення часового ряду);

– формалізація моделі для розрахунку прогнозу знаку приросту часового ряду з періодом 1 на основі результатів фрактального аналізу та за результатами індексації.

4. Формальна постановка проблеми прогнозування знаків приростів часових рядів

Нехай $\{z_i\}_{i=1}^{n}$ – деякий фінансовий дискретний часовий ряд без пропусків довжини n. Значення часового ряду фіксуються в дискретні моменти часу $t_i \in S$, $i=\overline{1,n}$, S – деяка дискретна множина, t_1 – початковий момент часу:

$$\{z_i\}_{i=1}^{n}=\{z_1,z_2,\ldots,z_n\}=\{z(t_1),z(t_2),\ldots,z(t_n)\}. \quad (1)$$

На основі даного ряду побудуємо часовий ряд, який складається з перших різниць $\{\Delta z_i\}_{i=1}^{n}$, де $\Delta z_i=z_i-z_{i-1}$, $i=\overline{2,n}$. Знаковий ряд, який утворений з вхідного часового ряду $\{z_i\}_{i=1}^{n}$ позначимо через $\{\chi_i\}_{i=2}^{n}$ або

$$\{\chi_i\}_{i=2}^{n}=\{\chi_2,\chi_3,\ldots,\chi_n\}, \quad (2)$$

де $\chi_i=\mathrm{sgn}(\Delta z_i)$.

Задача прогнозування знаків приростів полягає у побудові такої моделі F, яка дозволяє знайти оцінку знаку приросту $\hat{\chi}_1(n)$, що представляє собою прогноз, який виконується в точці n з періодом 1, причому розрахована на основі результатів прогнозування похибка має бути мінімальною. Особливості розрахунку похибок прогнозування у випадку знаків приростів часових рядів розглянуто в роботі [11]. Формально таку модель можна записати у вигляді (3). Якщо одразу перейти від часового ряду $\{z_i\}_{i=1}^{n}$ до знакового ряду $\{\chi_i\}_{i=2}^{n}$, то модель прогнозування матиме аналогічний вигляд (4). Позначимо цю модель через M. Тоді моделі прогнозування матимуть вигляд:

$$\hat{\chi}_1(n)=F(z_{n-m+1},z_{n-m+2},\ldots,z_n), \quad (3)$$

$$\hat{\chi}_1(n)=M(\chi_{n-m+1},\chi_{n-m+2},\ldots,\chi_n), \quad (4)$$

де m – довжина ділянки часового ряду, яка безпосередньо передує точці, в якій розраховується прогноз або довжина передісторії. Всі значення часового ряду з цієї ділянки використовуються для побудови проспекції.

5. Прогнозування знаків приростів часових рядів

Метод, який описано в цьому розділі, може використовуватися для прогнозування знаків приростів як фінансових часових рядів, так і часових рядів іншої природи, які близькі до випадкових.

Для побудови прогнозу за цим методом необхідно виконати ряд етапів:

Етап I. Передпрогнозний фрактальний аналіз фінансового часового ряду для встановлення рівня персистентності, антиперсистентності і випадковості часового ряду, перевірки на наявність довготривалої пам'яті, а також визначення значення m, тобто ідентифікації тієї ділянки часового ряду, яка буде безпосередньо використовуватися для розрахунку прогнозу. Ця ділянка довжини m називається опорною історією і позначається:

$$\chi_{(m)}^{n-m+1}=\{\chi_{n-m+1},\chi_{n-m+2},\ldots,\chi_n\}=\{\chi_j\}_{j=n-m+1}^{n}, \quad (5)$$

де в позначенні верхній індекс визначає точку, з якої починається побудова історії, а нижній індекс в дужках – довжину історії. Всі інші історії називаються неопорними і позначаються:

$$\chi_{(m)}^{N}=\{\chi_N,\chi_{N+1},\ldots,\chi_{N+m-1}\}=\{\chi_{N+j}\}_{j=0}^{m-1}, \quad (6)$$

де $N=\overline{1,n-m}$.

Процедура фрактального аналізу складається з таких алгоритмів:

1. Фрактальний R/S-аналіз для розрахунку показника Херста H вхідного часового ряду. Детально цей алгоритм описано в роботах [16, 17]. Для цих цілей також може бути використано метод детрендованого флуктуаційного аналізу.

2. Оцінка стійкості розрахованого показника H та дослідження його поведінки в динаміці [18]. У випадку застосування фрактального R/S-аналізу для оцінки значимості показника для вхідного часового ряду використовують функцію Еніса-Ллойда [19]. Розраховується величина $E \in [0,1]$, яка дозволяє оцінити персистентність:

– якщо $H > E$, то часовий ряд персистентний (трендостійкий) і для його прогнозування можна використати звичайні економетричні моделі: моделі експоненціального згладжування Брауна, Хольта, Хольта-Вінтерса, Трігга-Ліча, MA-моделі тощо;

– якщо $0.5 < H < E$, то часовий ряд випадковий і будь-яка модель прогнозування, яка буде використана в цьому випадку, буде неадекватною;

– якщо $0.5 < H$, то часовий ряд антиперсистентний або ергодичний, тобто він реверсує частіше, ніж випадковий часовий ряд;

– якщо $H - E < \delta$ і $0 \le \delta < 0.05$, то часовий ряд є близьким до випадкового. Застосування традиційних моделей прогнозування в цьому випадку не дозволяє отримати необхідну точність прогнозу і для побудови проспекції або знаходження оцінки знаку приросту з періодом 1 потрібно застосовувати модифіковані моделі, які базуються, зокрема, на основі попередньої індексації або кластеризації часового ряду. До таких часових рядів у фінансовій сфері належать ряди курсових пар, показник Херста яких в більшості випадків перевищує розрахункове значення Еніса-Ллойда E менш, ніж на δ. Слід зазначити, що в періоди невизначеності на ринку часові ряди курсових пар можуть «втрачати пам'ять» і ставати на короткий період випадковими або навіть антиперсистентними. Такі явища спостерігаються здебільшого під час кризових явищ. Тому для отримання повного розуміння поведінки часового ряду і його властивостей необхідне дослідження його показника Херста в динаміці та ідентифікації наявності пам'яті. В цьому дослідженні розглядається випадок часових рядів, які близькі до випадкових. Детальніше про дослідження показника H в динаміці розглянуто в роботі [18].

3. Розрахунок середньої довжини квазіциклів. Підхід визначення довжини квазіциклів на основі побудови сім'ї часових рядів і згладжуванні їх V-статистик за допомогою індексу Кауфмана описано в роботі [18]. Величина середньої довжини квазіциклу для вхідного часового ряду буде використовуватися для побудови опорної та неопорних історій як їх довжина m.

Етап II. Індексація часового ряду. Задача індексації часового ряду є класичною задачею Time-Series Data Mining і полягає у визначенні на основі деякої міри подібності або відмінності того, які з часових рядів деякої бази B подібні до вхідного часового ряду Z [13, 20]. База B включає часові ряди такої ж довжини, що й вхідний часовий ряд.

Неопорна історія називається подібною до опорної історії деякого дискретного часового ряду, якщо міра подібності d між ними не перевищує мір подібності d між опорною та всіма іншими неопорними історіями. У випадку задання часового ряду у вигляді (1) мірою подібності може бути відстань Евкліда, манхетенська метрика, відстані Махаланобіса, Мінковського, Журавлева тощо. У випадку задання часового ряду у вигляді (2) зручно використовувати відстані Хемінга або Роджерса-Танімото [21, 22]. Якщо розглядати опорну та неопорні історії як двійкові вектори в m-вимірному просторі, то відстань Хеммінга між ними довжини m представляє собою число позицій, в яких вони приймають різні значення. Загальна схема відбору подібної історії показана на рис. 1. Вибір опорної та неопорних історій на часовому ряді показано на рис. 2.

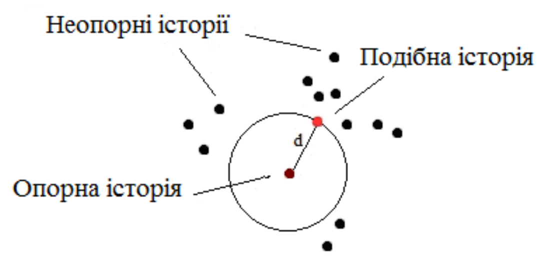

Рис. 1. Схема відбору подібної історії

Існує дві основні постановки задачі індексації:

1. Знайти з бази часових рядів B такі $k > 0$ рядів, міра подібності d яких до вхідного часового ряду не перевищує фіксованого, наперед заданого числа ε, $d \le \varepsilon$.

2. Знайти з бази B такі $k > 0$ часових рядів, міра подібності яких d до вхідного часового ряду мінімальна.

Слід зазначити, що перед проведенням індексації часового ряду необхідно переконатися, що обрана міра подібності між двома часовими рядами задовольняє вимогам симетричності, сталості при самоподібності, невід'ємності та правило трикутника.

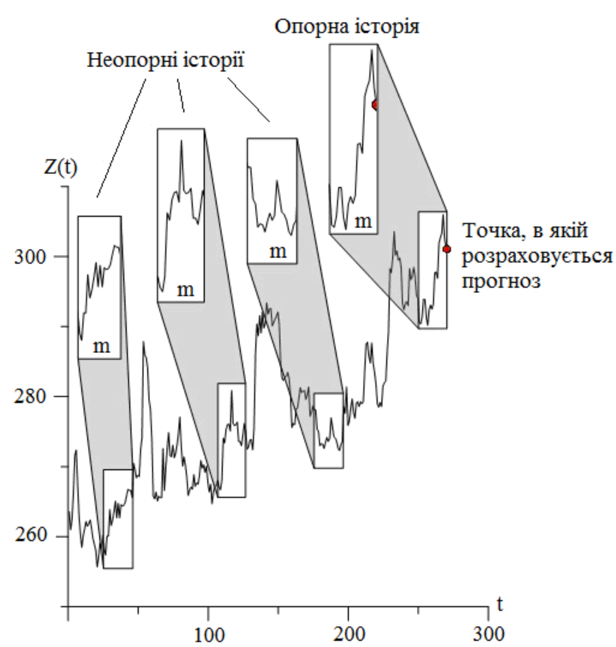

Рис. 2. Вибір історій на вхідному фінансовому часовому ряді

Є два підходи до реалізації індексації для нашої задачі, залежно від представлення вхідного часового ряду (у вигляді (1) та у вигляді (2)). Якщо ряд представлено у вигляді (1), то перед проведенням індексації його потрібно нормувати за формулою [23]:

$$\tilde{z}_{N+j} = \frac{z_{N+j} - \mu\left(z_{(m)}^{N}\right)}{\sigma\left(z_{(m)}^{N}\right)}, \tag{7}$$

де

$$\mu\left(z_{(m)}^{N}\right) = \frac{1}{m}\sum_{i=0}^{m-1} z_{N+i}, \tag{8}$$

– середнє арифметичне кожної історії часового ряду,

$$\sigma\left(z_{(m)}^{N}\right) = \sqrt{\frac{1}{m}\sum_{i=0}^{m-1}\left(z_{N+i} - \mu\left(z_{(m)}^{N}\right)\right)^2}, \tag{9}$$

– середньоквадратичне відхилення історій.

Далі необхідно виділити з нормованого часового ряду $\{\tilde{z}_i\}_{i=1}^{n}$ опорну історію:

$$\tilde{z}_{(m)}^{n-m+1} = \left\{\tilde{z}_{n-m+1}, \tilde{z}_{n-m+2}, \ldots, \tilde{z}_n\right\} = \left\{\tilde{z}_j\right\}_{j=n-m+1}^{n} \tag{10}$$

і неопорну історію:

$$\tilde{z}_{(m)}^{N} = \left\{\tilde{z}_N, \tilde{z}_{N+1}, \ldots, \tilde{z}_{N+m-1}\right\} = \left\{\tilde{z}_{N+j}\right\}_{j=0}^{m-1}, \ N = \overline{1, n-m}. \tag{11}$$

Знайдемо всі відстані Евкліда w між опорною та всіма неопорними історіями за формулою для $N = \overline{1, n-m}$:

$$w\left(\tilde{z}_{(m)}^{n-m+1}, \tilde{z}_{(m)}^{N}\right) = \sqrt{\sum_{j=1}^{m}\left(\tilde{z}_{n-m+j} - \tilde{z}_{N+j-1}\right)^2}. \tag{12}$$

Серед розрахованих значень відберемо такі $k \in \mathbb{N}$ відстаней, які є мінімальними:

$$w\left(\tilde{z}_{(m)}^{n-m+1}, \tilde{z}_{(m)}^{N}\right) \xrightarrow[N=\overline{1,n-m}]{} \min. \tag{13}$$

Позначимо неопорні історії таких мінімальних відстаней через $\overline{z}_{(m)}^{K}$, де $K \in [1, n-m]$. Розіб'ємо ці історії на дві групи:

– група таких історій $\overline{z}_{(m)}^{K}$, для яких $\overline{z}_{(m)}^{K+1} - \overline{z}_{(m)}^{K} > 0$. Кількість таких історій позначимо через η_m^{+1}, $\eta_m^{+1} \le k$;

– група таких історій $\overline{z}_{(m)}^{K}$, для яких $\overline{z}_{(m)}^{K+1} - \overline{z}_{(m)}^{K} < 0$. Кількість таких історій позначимо через η_m^{-1}, $\eta_m^{-1} \le k$.

Для того, щоб сформувати такі групи у випадку задання фінансового часового ряду у вигляді (2), необхідно розрахувати всі міри Хемінга $\overline{w}$ між опорною та всіма неопорними історіями для $N = \overline{1, n-m}$ за формулою:

$$\overline{w}\left(\chi_{(m)}^{n-m+1}, \chi_{(m)}^{N}\right) = \sum_{j=1}^{m}\left|\chi_{n-m+j} - \chi_{N+j-1}\right|. \tag{14}$$

Аналогічно, формуємо такі $k \in \mathbb{N}$ відстаней, для яких виконується умова мінімума:

$$\overline{w}\left(\tilde{z}_{(m)}^{n-m+1}, \tilde{z}_{(m)}^{N}\right) \xrightarrow[N=\overline{1,n-m}]{} \min. \tag{15}$$

Неопорні історії, які визначають такі мінімальні відстані позначимо через $\overline{\chi}_{(m)}^{K}$, $K \in [1, n-m]$ і розіб'ємо їх на дві групи:

– група таких історій $\overline{\chi}_{(m)}^{K}$, для яких $\overline{\chi}_{(m)}^{K+1} = 1$. Кількість таких історій позначимо через ζ_m^{+1}, $\zeta_m^{+1} \le k$;

– група таких історій $\overline{\chi}_{(m)}^{K}$, для яких $\overline{\chi}_{(m)}^{K+1} = 0$. Кількість таких історій позначимо через ζ_m^{-1}, $\zeta_m^{-1} \le k$.

Етап III. Розрахунок прогнозу знаку приросту фінансового часового ряду. Позначимо через $Q\left(\hat{\chi}_1(n) = \delta\right)$ оцінку того, що $\hat{\chi}_{n+1} = \delta$, при $\delta \in \{-1, 1\}$. Знайти ці оцінки у випадку задання вхідного часового ряду у вигляді (1) можна за формулами [24]:

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \frac{\eta_m^{\delta}}{\eta_m^{-1} + \eta_m^{+1}}, \tag{16}$$

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \frac{\sum_{m=1}^{n-1}\eta_m^{\delta}}{\sum_{m=1}^{n-1}\left(\eta_m^{-1} + \eta_m^{+1}\right)}, \tag{17}$$

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \frac{\sum_{m=1}^{n-1} m\,\eta_m^{\delta}}{\sum_{m=1}^{n-1} m\left(\eta_m^{-1} + \eta_m^{+1}\right)}, \tag{18}$$

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \prod_{m=1}^{n-1}\frac{\eta_m^{\delta}}{\eta_m^{-1} + \eta_m^{+1}}. \tag{19}$$

У випадку задання вхідного фінансового часового ряду у вигляді (2) оцінки розраховуються за аналогічними формулами:

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \frac{\zeta_m^{\delta}}{\zeta_m^{-1} + \zeta_m^{+1}}, \tag{20}$$

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \frac{\sum_{m=1}^{n-1}\zeta_m^{\delta}}{\sum_{m=1}^{n-1}\left(\zeta_m^{-1} + \zeta_m^{+1}\right)}, \tag{21}$$

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \frac{\sum_{m=1}^{n-1} m\,\zeta_m^{\delta}}{\sum_{m=1}^{n-1} m\left(\zeta_m^{-1} + \zeta_m^{+1}\right)}, \tag{22}$$

$$Q\left(\hat{\chi}_1(n) = \delta\right) = \prod_{m=1}^{n-1}\frac{\zeta_m^{\delta}}{\zeta_m^{-1} + \zeta_m^{+1}}. \tag{23}$$

Для розрахунку прогнозу знаку приросту за даним методом, введемо порогове значення $\phi \approx 0.5$, а також значення $\lambda \approx 0.1$ для ідентифікації проміжку, на якому спрямований приріст не спостерігається. Ці значення підбираються на початку побудови проспекції і коригуються з кожним новим розрахованим прогнозом, враховуючи похибку прогнозування. Таким чином, якщо вхідний ряд має вигляд (1), то прогноз його знаку приросту з періодом 1 розраховується за формулою:

$$\hat{\delta}_{n+1} = \begin{cases} \delta, \mu > \phi + \lambda \\ 0, \mu \in \left[\phi - \lambda, \phi + \lambda\right] \\ -\delta, \mu < \phi - \lambda \end{cases}, \tag{24}$$

де $\delta \in \{-1,1\}$ і відношення оцінок

$$\mu = \frac{Q\left(\hat{\chi}_1(n) = \delta\right)}{Q\left(\hat{\chi}_1(n) = -\delta\right)}. \tag{25}$$

Алгоритм розрахунку прогнозу знаку приросту для часового ряду, який задано у вигляді (1) має вигляд:

1. Завантаження вхідного часового ряду $\{z_i\}_{i=1}^n$ (1).
2. Нормалізація часового ряду $\{z_i\}_{i=1}^n$ за формулами (7)–(9), тобто формування часового ряду $\{\tilde{z}_i\}_{i=1}^n$.
3. Проведення процедури фрактального аналізу (R/S-аналіз або детрендований флуктуаційний аналіз). Встановлення параметрів, які необхідні для побудови моделі прогнозування.
4. Розбиття часового ряду $\{\tilde{z}_i\}_{i=1}^n$ на ділянки довжиною m методом плинного вікна. Виділення опорної та неопорних історій.
5. Індексація часового ряду для ідентифікації подібних ділянок за формулою Евкліда (12) на основі методу К-найближчих сусідів. За основу обирається ділянка, що передує точці, в якій розраховується прогноз, тобто опорна історія. В результаті отримаємо ділянки, які подібні до опорної.
6. Підрахунок кількості подібних ділянок, після яких спостерігається тенденція до зростання та спадання.
7. Розрахунок оцінок за формулами (16)–(19).
8. Розрахунок прогнозу з періодом 1 за формулами (24), (25).

Алгоритм розрахунку прогнозу знаку приросту з періодом 1 для знакового часового ряду (2) має вигляд:

1. Завантаження вхідного часового ряду $\{z_i\}_{i=1}^n$ (1).
2. Перетворення часового ряду $\{z_i\}_{i=1}^n$ на знаковий ряд $\{\chi_i\}_{i=2}^n$ (2) або безпосереднє завантаження знакового ряду.
3. Проведення процедури фрактального аналізу.
4. Розбиття знакового часового ряду $\{\chi_i\}_{i=2}^n$ на ділянки довжиною m. Виділення опорної та неопорних історій.
5. Індексація часового ряду для ідентифікації подібних ділянок за формулою Хемінга (14) на основі методу К-найближчих сусідів. За основу обирається ділянка, що передує точці, в якій розраховується прогноз, тобто опорна історія. В результаті отримаємо ділянки, які подібні до опорної.
6. Підрахунок кількості подібних ділянок, після яких спостерігається тенденція до зростання та спадання.
7. Розрахунок оцінок за формулами (20)–(23).
8. Розрахунок прогнозу з періодом 1 за формулами (24), (25).

6. Висновки

В результаті проведених досліджень було встановлено:

1. Якщо за результатами передпрогнозного фрактального аналізу отримано, що фінансовий часовий ряд є персистентним, то можна застосувати традиційні економетричні методи прогнозування. Якщо часовий ряд є персистентним, але близьким до випадкового ($H \approx 0.5$, $H > 0.5$), то можна використовувати запропонований метод прогнозування знаків приростів. В інших випадках, якщо часовий ряд випадковий або антиперсистентний, то використовувати будь-які методи прогнозування не має змісту. За результатами фрактального аналізу також встановлено, що застосування запропонованого методу не має змісту, якщо у часовому ряді не ідентифіковано наявність довготривалої пам'яті. Ідентифікована за результатами аналізу середня довжину квазіциклу часового ряду може бути використана в запропонованому методі прогнозування як граничне максимальне значення для довжини передісторії m.
2. На наступному етапі дослідження на основі мір Хемінга і Евкліда за методом К-найближчих сусідів була проведена процедура індексації часового ряду або зіставлення зі зразком для визначення тих ділянок ряду, які подібні до зразка (до опорної історії).
3. В результаті індексації отримані історії селективно групувались на дві групи, залежно від знаку приросту, який по їх завершенні спостерігався. На основі формул (16)–(23) залежно від вхідного представлення фінансового часового ряду розраховувались оцінки того, що ряд буде зростати або спадати. Далі знаходились відношення цих оцінок, що визначало силу спрямованості приросту до зростання або спадання (25). Прогноз розраховувався за формулою (24). Описаний метод використовувався для прогнозування часових рядів курсових пар (EUR-USD, EUR-GDP тощо), які є близькими до випадкових. Точність прогнозування за описаним методом для даних часових рядів склала >58 %, що для рядів такого типу є непоганим результатом.

В результаті дослідження описано і формалізовано метод селективного зіставлення зі зразком для розрахунку прогнозу знаку приросту з періодом 1 для фінансових часових рядів, які близькі до випадкових. Цей метод, що є альтернативою для таких методів, які описані в роботах [6, 13], може бути використано в системах підтримки прийняття рішень і прогнозування у фінансовій сфері для прогнозування курсових пар, приросту індексів тощо [25]. Описаний метод прогнозування може бути складовою частиною експертної системи, яка може використовуватись для прийняття рішень і планування в певній сфері діяльності.

Література

1. Azoff, M. E. Neural Network Time Series Forecasting of Financial Markets [Text] / M. E. Azoff. – Cornwall : John Wiley and Sons, 1994. – 212 p.

2. Ferreira, C. Gene Expression Programming. A new adaptive Algorithm for solving Problems [Text] / C. Fereira // Complex Systems. – 2001. – Vol. 13, Issue 2. – P. 87–129.
3. Barbulescu, A. Time Series Modeling Using an Adaptive Gene Expression Programming Algorithm Barbulescu, A. Time Series Modeling Using an Adaptive Gene Expression Programming Algorithm [Text] / A. Barbulesku, E. Bautu // International Journal of Mathematics and Computation. – 2009. – Vol. 3, Issue 2. – P. 85–93.
4. Матвійчук, А. В. Моделювання економічних процесів із застосуванням методів нечіткої логіки [Текст] / А. В. Матвійчук. – К. : КНЕУ, 2007. – 264 с.
5. Снитюк, В. Є. Прогнозування. Моделі. Методи. Алгоритми [Текст]: навч. посібник / В. Є. Снитюк. – К. : Маклаут, 2008. – 364 с.
6. Лукашин, Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов [Текст]: учеб. пособие / Ю. П. Лукашин. – М. : Финансы и статистика, 2003. – 416 с.
7. Box, G. E. P. Time series analysis: forecasting and control [Text] / G. E. P. Box, G. M. Jenkins. – San Francisco : Holden-Day, 1976. – 575 p.
8. Brown, R. G. Statistical forecasting for inventory control [Text] / R. G. Brown. – US : McGraw-Hill Inc., 1959. – 223 p.
9. Holt, C. C. Forecasting trends and seasonal by exponentially weighted averages [Text] / C. C. Holt // International Journal of Forecasting. – 1957. – Vol. 20, Issue 1. – P. 5–10.
10. Vercellis, C. Business intelligence: data mining and optimization for decision making [Text] / C. Vercellis. – Cornwall : John Wiley & Sons Ltd. Publication, 2009. – 417 p. doi: 10.1002/9780470753866
11. Берзлев, О. Ю. Метод прогнозування знаків приростів часових рядів [Текст] / О. Ю. Берзлев // Східно-Європейський журнал передових технологій. – 2013. – Т. 2, № 4 (62). – С. 8–11. – Режим доступу: http://journals.uran.ua/eejet/article/view/12362/10250
12. Берзлев, А. Ю. Разработка комбинированных моделей прогнозирования с кластеризацией временных рядов по методу ближайшего соседа [Текст]: сб. науч. тр. / А. Ю. Берзлев // Автоматизированные системы управления и приборы автоматики. – 2012. – Вып. 161. – С. 51–59.
13. Singh, S. Pattern Modeling in Time-Series Forecasting [Text] / S. Singh // Cybernetics and Systems. An International Journal. – 2000. – Vol. 31, Issue 1. – P. 49–65. doi: 10.1080/019697200124919
14. Perlin, M. S. Nearest neighbor method [Text] / M. S. Perlin // Revista Eletrônica de Administração. – 2007. – Vol. 13, Issue 2. – P. 15.
15. Fernández-Rodríguez, F. Nearest-Neighbour Predictions in Foreign Exchange Markets [Text] / F. Fernández-Rodríguez, S. Sosvilla-Rivero, J. Andrada-Félix // Fundacion de Estudios de Economia Aplicada. – 2002. – Vol. 5. – P. 36.
16. Peters, E. E. Fractal market analysis: applying chaos theory to investment and economics [Текст] / E. Peters. – Cornwall : John Wiley & Sons Inc, 1994. – 336 p.
17. Hurst, H. E. Long-term storage capacity of reservoirs [Text] / H. Hurst // Transactions of the American Society of Civil Engineers. – 1951. – Vol. 116. – P. 770–799.
18. Берзлев, О. Ю. Методика передпрогнозного фрактального аналізу часових рядів [Текст] / О. Ю. Берзлев // Управління розвитком складних систем. – 2013. – Вип. 16. – С. 76–81.
19. Anis, A. The expected value of the adjusted rescaled Hurst Range of independent normal summands [Text] / A. Anis, E. Lloyd // Biometrika. – 1976. – Vol. 63, Issue 1. – P. 111–116. doi: 10.2307/2335090
20. Kahveci, T. Variable length queries for time series data [Text] / T. Kahveci, A. Singh // Proceedings of the 17th International Conference on Data Engineering. – Heidelberg, Germany, 2001. – P. 273–282. doi: 10.1109/icde.2001.914838
21. Chang, C. L. E. Clustering for approximate similarity search in high-dimensional spaces [Text] / C. L. E. Chang, H. Garcia-Molina, G. Wiederhold // IEEE Transactions on Knowledge and Data Engineering. – 2002. – Vol. 14, Issue 4. – P. 792–808. doi: 10.1109/tkde.2002.1019214
22. Зайченко, Ю. П. Нечеткие модели и методы в интеллектуальных системах [Текст] : учеб. пособие / Ю. П. Зайченко. – К. : Слово, 2008. – 344 с.
23. Goldin, D. Q. On Similarity Queries for Time-Series Data: Constraint Specification and Implementation [Text] / D. Q. Goldin, P. C. Kanellakis // Proceedings of the 1-st International Conference on the Principles and Practice of Constraint Programming. – Cassis, France, 1995. – P. 137–153. doi: 10.1007/3-540-60299-2_9
24. Кучанський, О. Ю. Прогнозування часових рядів методом зівставлення зі зразком [Текст] / О. Ю. Кучанський, В. В. Ніколенко // Управління розвитком складних систем. – 2015. – Вип. 22. – С. 101–106.
25. Берзлев, О. Ю. Сучасний стан інформаційних систем прогнозування часових рядів [Текст] / О. Ю. Берзлев // Управління розвитком складних систем. – 2013. – Вип. 13. – С. 78–82.