УДК 519.688

Д.М. Чабаненко

АЛГОРИТМ ПРОГНОЗУВАННЯ ФІНАНСОВИХ ЧАСОВИХ РЯДІВ НА ОСНОВІ СКЛАДНИХ ЛАНЦЮГІВ МАРКОВА

Пропонується застосування технології складних ланцюгів Маркова (ланцюгів Маркова з пам'яттю) для прогнозування фінансових часових рядів. Головною відмінністю складних ланцюгів Маркова від простих є урахування післядії або пам'яті. Складний ланцюг Маркова може бути зведений до простого через поняття „узагальнений стан" як кортеж послідовних станів. Технологія передбачає прогнозування ряду за ієрархією інтервалів дискретизації часу та послідовного „склеювання" результатів прогнозів на різних частотних рівнях у один вихідний ряд прогнозу. Даний підхід дозволяє найбільш повно використати інформацію, яка міститься в часовому ряді і отримати найбільш адекватне його продовження. Здійснена експериментальна перевірка ефективності алгоритму на рядах фондових, валютних та спотових ринків.

Ключові слова: прогнозування, часові ряди, ланцюги Маркова, складні системи

Вступ

Прогнозування фінансово-економічних часових рядів є надзвичайно актуальною задачею. Сучасні підходи до даної задачі можна охарактеризувати наступними напрямками: 1) апроксимація часового ряду аналітичною функцією та екстраполяція знайденої функції у напрямку майбутнього – так звані трендові моделі [1]; 2) дослідження впливу усіх можливих факторів на показник, який прогнозується та побудова економетричних, або більш складних моделей за допомогою методу групового урахування аргументів (МГУА) [2]; 3) моделювання майбутніх цін як результатів прийняття рішень за допомогою нейронних мереж, генетичних алгоритмів, нечітких множин [3, 4]. На жаль, дані методики не демонструють стабільних прогнозів, що може бути пояснене складністю систем, динаміка яких прогнозується, постійною зміною їх структури. Ми намагаємось поєднати ці напрями в одному алгоритмі, але надаємо перевагу останньому, який полягає у побудові моделі, адекватної процесу, який породжує часовий ряд ціни [5]. Саме такий підхід дає можливість наблизитись до складності системи, яка генерує досліджуваний ряд, побудувати її модель та використовувати властивості моделі у якості прогнозу.

Мета статті

Метою роботи є опис застосування технології складних ланцюгів Маркова (ланцюгів Маркова з пам'яттю) для прогнозування фінансових часових рядів та дослідження ефективності описаної методики на рядах фондових, валютних та спотових ринків.

1. Технологія прогнозування

Припустимо, існує послідовність дискретних станів певної системи. З цієї послідовності можна визначити ймовірності переходу з одного стану в інший. Простим ланцюгом Маркова є випадковий процес, в якому ймовірність наступного стану залежить тільки від попереднього стану та не залежить від усіх інших станів. На відміну від простого, складним ланцюгом Маркова називають випадковий процес, в якому ймовірність наступного стану залежать не лише від наявного, а від послідовності декількох попередніх станів (передісторії). Кількість станів у передісторії є порядком ланцюга Маркова.

Теорія простих ланцюгів Маркова широко викладена в літературі, наприклад [6]. Що стосується ланцюгів Маркова вищих порядків, то в літературі [7] наведене лише означення.

Ланцюг Маркова порядку вище 1-го можна звести до простого ланцюга Маркова за допомогою введення поняття „узагальнений стан", включаючи в нього ряд послідовних станів системи. В цьому випадку апарат простих ланцюгів Маркова може бути застосований до складних.

Досліджуваний динамічний ряд є результатом певного процесу. Припускається, що цей процес є детермінованим, що означає існування причинно-наслідкової залежності наступних станів від передісторії. Неможливо зафіксувати та проаналізувати нескінченну передісторію, що заважає точному виявленню даного впливу та побудову абсолютно точних прогнозів. Поставлена задача полягає у максимальному використанні інформації, яка міститься у відомому відрізку ряду та моделювання на основі цього найбільш ймовірних сценаріїв продовження ряду в майбутньому.

Досліджуваний процес описується у вигляді часового ряду ціни $p(t)$ із заданим проміжком дискретизації Δt

$$p_{\mathrm{ti}}=p(t_0+\mathrm{i}\cdot\Delta t). \tag{1}$$

Дискретне представлення ряду є фактично способом існування даної системи. Формування ціни відбувається на основі угод, укладених на ринку в певні дискретні моменти часу, а часовий ряд ціни є рядом усереднених рівнів ціни за вибрані проміжки часу. Кожен трейдер, який є частиною системи ціноутворення, під час прийняття рішення працює з суто дискретними рядами на вибраному часовому інтервалі (наприклад, хвилинному, 5-хвилинному, годинному, денному тощо). При прямуванні $\Delta t \to 0$ точність представлення даних досягає певної межі, оскільки при достатньо малих Δt ціна змінюється стрибком в момент здійснення угоди, а протягом часу між угодами залишається незмінною і рівною останній угоді. Дані факти свідчать про те, що дискретність часових рядів необхідно розуміти як один із принципів організації складної фінансової системи, а не лише як обмежене представлення результатів її діяльності [5, 8].

Ряд вихідних значень необхідно перетворити у ряд дискретних станів. Позначимо кількість вибраних станів $\boldsymbol{s}$, кожен з яких пов'язаний зі зміною величини вихідного сигналу (прибутковістю). Наприклад, класифікація з двома станами, перший з яких відповідає додатній прибутковості при зростанні ціни, а другий – від'ємній при її спаданні. В загальному вигляді всі можливі прирости вихідного ряду класифікуємо на $\boldsymbol{s}$ груп. Способи розбиття будуть обговорюватись нижче.

Далі здійснюється прогнозування ряду дискретизованих станів. Для заданого порядку ланцюга Маркова та останнього узагальненого стану в якості наступного вибирається найбільш ймовірний стан. У випадках неоднозначності при визначенні стану з максимальною ймовірністю застосовується алгоритм, який дозволяє зменшити кількість можливих сценаріїв прогнозу. Таким чином, маємо ряд прогнозованих станів, які для відомого останнього значення ряду можуть бути перетворені на дискретизований ряд прогнозних значень.

Обчислення приростів, прогнозування та послідуюче відновлення здійснюється здійснити для заданої ієрархії приростів часу Δt. Для ефективного використання інформації, представленій в наявному часовому ряді, прогнозування здійснюється для приростів часу Δt=1,2,4,8..., або більш складної ієрархії приростів, яка розглянута в даній роботі в параграфі 1.4, та послідовного „склеювання" результатів отриманих на різних дискретизаціях прогнозів.

Процедура прогнозування та склеювання є ітераційною та проводиться, починаючи з менших приростів, додаючи на кожному кроці прогноз з більшим приростом часу.

При збільшенні кроку дискретизації часу Δt зменшується статистика для визначення ланцюгів Маркова, найбільший крок дискретизації, який приймає участь у прогнозуванні обмежується. Для доповнення прогнозу низькочастотною складовою використовується наближення нульового порядку у вигляді лінійного тренду , або комбінації лінійного тренду та гармонійних коливань [9].

2. Алгоритм побудови прогнозу

Розглянемо послідовність операцій, які необхідні для побудови прогнозного ряду. Для цього необхідно задати наступні параметри:

1) Вид ієрархії приростів часу (проста – степені двійки, складна - добуток степенів перших простих чисел)

2) Величини $\boldsymbol{s}$ – кількість станів та $\boldsymbol{r}$ - порядок ланцюга Маркова. Дані параметри можуть бути індивідуальними для кожного рівня дискретизації, знаходження оптимальних параметрів здійснюється експериментально.

3) Величина порогу δ, та мінімальна кількість переходів $N_{\min}$

Алгоритм побудови прогнозу включає наступні кроки:

1) Генерація ієрархії приростів часу – послідовності Δt, максимальний з яких повинен відповідати довжині прогнозного проміжку N_1.

2) Для кожного приросту часу Δt зі зростанням приростів, здійснюється прогнозування станів та відновлення ряду за прогнозними станами. Даний етап включає наступні дії:

2.1. Обчислення приростів (прибутковостей) ряду з дискретизацією Δt.

2.2. Перетворення ряду приростів у ряд номерів станів (1..$\boldsymbol{s}$).

2.3. Обчислення ймовірностей переходів для узагальнених станів.

2.4. Побудова ряду прогнозних станів, застосовуючи процедуру визначення найбільш ймовірного наступного стану.

2.5. Відновлення ряду значень з ряду станів з дискретизацією Δt.

2.6. Склеювання прогнозу з дискретизацією Δt з рядом, який отримався в результаті склеювання попередніх шарів (з меншим кроком Δt). У випадку якщо даний ряд є першим, в якості результату склеювання повертається ряд без змін.

3) Останній склеєний ряд склеїти з продовженням лінійного тренда, побудованого по усім попередньо відомим точкам.

Ряд, склеєний з лінійним трендом, є результатом прогнозування. Розглянемо етапи наведеного алгоритму більш детально.

3. Стани в складних ланцюгах Маркова та підходи до їх визначення

Стани в даній технології зв'язані с вимірюванням прогнозної величини. Пропонуються наступні способи класифікації прибутковостей у стани. Серед них – класифікація на основі принципу рівномірності за кількістю представників в класах; на основі принципу рівномірності за відхиленням, а також їх комбінації для різних модулів відхилень.

Основою для класифікації станів є приріст, або прибутковість ряду [10]. Розглядається абсолютні (2) та відносні (3) прирости ряду.

$$r_t = p_t - p_{t-\Delta t} \quad (2)$$

$$r_t = \frac{p_t - p_{t-\Delta t}}{p_t} \quad (3)$$

де p_t – вхідний ряд динаміки ціни, Δt – проміжок дискретизації, який вибрано для аналізу. Відомо, що математичне сподівання ряду прибутковостей дорівнює нулю, а дисперсія є фактично мірою волатильності ряду. На основі значень прибутковостей r_t здійснюється класифікація та перетворення значень ряду в ряд дискретних станів. Один з принципів проведення класифікації рівномірність за кількістю представників класів. Дана класифікація поділяє множину всіх приростів $\{r_t\}$ на $\boldsymbol{s}$ рівних по кількості груп. Обчислені прирости ряду з даною дискретизацією упорядковуємо за зростанням та здійснюємо поділ відсортованого масиву на рівні частини. Таким чином визначаються граничні значення $\{r_{lim,i}\}$, які використовуються при перетворенні прибутковостей на номери класів. Проблему може створити велика кількість однакових станів, що спричинює однакові межі декількох сусідніх станів. Це створює номера станів без жодного представника, що робить необхідним корекцію розбиття з ціллю досягнути найбільш можливої рівномірності в розподілу на стани. Класифікація відбувається за наступним алгоритмом [11]:

$$s_t = (i \mid r_{lim,i-1} > r_t > r_{lim,i}) \quad (4)$$

де s_t – номер стану, який відповідає моменту часу t, для якого обчислена прибутковість r_t; і – номер стану $[1\ldots \boldsymbol{s}]$, інтервал $[r_{lim,i-1}, r_{lim,i}]$ якого відповідає обчисленій прибутковості r_t.

Крім інтервалу прибутковостей, заданого зазначеними вище граничними значеннями $[r_{lim,i-1}, r_{lim,i}]$, для кожного стану вибирається середнє значення прибутковості $r_{avg,i}$, яке буде використовуватись при відновленні значень ряду за прогнозованими дискретними станами.

Другий спосіб розбиття на стани полягає в розбитті інтервалу значень прибутковостей на рівні частини, від мінімального до максимального відхилення на рівні частини. В цьому випадку рівномірність по кількості представників в станах не зберігається. Фактично даний спосіб відрізняється від попереднього алгоритмом визначення граничних значень $\{r_{lim,i}\}$ Можливі комбіновані способи розбиття, при яких, замість максимального і мінімального значення використовується граничне значення, залежне від середньоквадратичного відхилення, а розбиття здійснюється рівномірно за відхиленням.

Оскільки реальна причинно-наслідкова залежність у процесі невідома, для її виявлення необхідна адекватна класифікація станів, яка б не приховувала закономірності в ряді, а дозволяла їх виявити. Пропонується декілька способів розбиття на стани, при яких, по-перше, ймовірності переходів між станами були б адекватними, а по-друге, усереднення відхилень всередині станів не вплинуло б на точність одержаного прогнозу.

Для перевірки ефективності розбиття ми проводимо процедуру дискретизації та класифікації приростів на стани на кожній ієрархії, а потім по відомих станах для кожної ієрархії відновлюємо ряди та проводимо процедуру склеювання. Оскільки ряди станів повністю відповідають вихідному ряду, то отримується крива, відхилення якої від вихідної спричинені лише похибкою усереднення станів (похибка квантування). Таким чином, прийнявши певне значення кількості станів *s*, та провівши процедуру дискретизації, відновлення та склеювання (без прогнозування), одержуємо абсолютну похибку дискретизації (квантування).

Зі збільшенням кількості станів точність відновлення збільшується, але необхідно мати на увазі, що вибір кількості рівнів квантування обмежений тим, що необхідне визначення ймовірностей переходів з достатньою точністю, що підтверджується експериментами з прогнозуванням штучних тестових рядів.

4. Процедура покрокового прогнозування. Визначення найбільш ймовірного стану на наступному кроці, сценарії прогнозу

При прогнозуванні в якості наступного стану вибирається стан з найбільшою ймовірністю при даних умовах. Для цього ми використовуємо матрицю ймовірностей переходів із стану в стан. При цьому необхідно враховувати, що ймовірності обчислюються з похибкою. Ми не можемо точно обчислити ймовірності, оскільки принципово неможливо отримати нескінчений ряд, а відома тільки частина ряду – відомий відрізок, по якому обчислюються ймовірності. Другий важливий момент – випадок декількох станів з найбільшою ймовірністю.

Для того, щоб не допустити втрати станів, ймовірності яких обчислені з похибкою, необхідно до стану з максимальною ймовірністю додати стани, які знаходяться на відстані δ від максимального. Значення параметру δ залежить від похибки оцінювання ймовірностей та потребує експериментального уточнення.

Якщо прийняти $\delta > 0$, то кількість станів з найбільшою ймовірністю збільшиться в порівнянні зі значенням $\delta = 0$. Декілька сусідніх станів з максимальною ймовірністю назвемо кластером. Припускається, що стани кластера з середніми значеннями відхилення мають найбільшу ймовірність. Для прогнозування обмежимося одним або двома найбільш ймовірними станами. Для їх визначення пропонується наступний алгоритм:

1) Якщо рівні (квантовані прирости) з максимальною ймовірністю утворюють декілька кластерів (кластер - група з декількох сусідніх рівнів – елементів кластера, мінімальний кластер – один ізольований рівень), то вибираємо найбільший за розміром кластер.

2) Якщо число елементів кластера непарне, то вибираємо центральний елемент в якості $k_{\max}$.

3) Якщо число елементів в кластері парне, то розглядаємо два центральні елементи кластеру, та вибираємо в якості $k_{\max}$ той з них, який ближче до центра розподілу.

4) Якщо два центральних елементи кластера знаходяться на однаковій відстані від центра розподілу, то розглядуємо обидва як можливі варіанти 1 та 2 значень $k_{\max}$ (точка біфуркації).

5) Якщо кластерів з максимальним розміром декілька, то розглядаємо їх як нові елементи, які, в свою чергу, можуть формувати кластери, з якими ми вчинимо аналогічно пунктам 1)-4).

В цей принцип відбору закладені наступні міркування:

1) Якщо існують два однакових сусідні стани з максимальною ймовірністю, то краще взяти той, який ближче до центра розподілу, щоб звести до мінімуму ризик появи у прогнозі хибних лінійних трендів.

2) Якщо рівні з максимальною ймовірністю не сусідні (багатомодовий розподіл), то необхідно розглядувати як мінімум два варіанти, оскільки це може бути зв'язане з біфуркаціями, які теж хотілось не упустити з розгляду.

3) Реалізуючи прогноз по варіанту 1) (на усіх етапах ієрархії), ми отримуємо певне наближення нижньої границі прогнозу, а реалізуючи прогноз за варіантом 2) – отримуємо наближення верхньої границі.

Даний алгоритм адекватно відтворює випадок можливого бімодального розподілу ймовірностей, в результаті чого пропонується розглянути 2 сценарії прогнозу.

5. Складні ланцюги Маркова та вплив передісторії (пам'яті). Порядок ланцюга Маркова

У випадку ланцюга Маркова першого порядку на основі даного ряду станів обчислюються ймовірності переходів з кожного стану у всі інші. Ці ймовірності прийнято записувати у квадратну матрицю розмірністю ***s***. Рядки такої матриці відповідають номерам станів, з якого відбувається перехід, а стовпці (елементи цих рядків) – стани, ймовірності переходу в які визначається.

Наприклад, для трьох станів можна записати матрицю ймовірностей переходів P

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}, \; p_{ij} = p(s_{t-1} = i, s_t = j) \qquad (5)$$

в якій, наприклад, $P(1,2)$ – ймовірність переходу із стану 1 в стан 2.

У випадку складного ланцюга Маркова, ймовірність наступного стану залежить не тільки від попереднього стану, а й від послідовності r станів, які відбулись перед даним. У цьому випадку, необхідно обчислити ймовірності переходів з послідовності r станів у стан r+1. Формально ці ймовірності можна записати в прямокутну таблицю з розмірністю (r^s, s).

Можна звести ланцюги Маркова порядку r до ланцюга порядку 1, узагальнивши поняття "наявний стан", включивши в нього послідовність з r станів, які передують стану, ймовірність якого обчислюється. Таким чином, ймовірності переходів можна записати в квадратні матриці розмірністю (r^s,r^s), яка є матрицею ймовірностей переходів між узагальненими станами.

Процес прогнозування полягає в наступному: вибирається останній стан (у випадку ланцюга Маркова порядку r>1 береться послідовність r останніх станів). Визначається ймовірність переходу з даного стану у всі можливі. Із усіх можливих наступних станів вибирається стан з максимальною ймовірністю. Можливий випадок декількох станів з максимальною ймовірністю, який може бути пояснений бімодальним розподілом ймовірностей. Процес прийняття рішення в цьому випадку описаний нижче.

Вибраний найбільш ймовірний стан приймається як наступний прогнозований стан та процедура повторюється з наступним (доданим останнім) станом. Таким чином, ми отримуємо ряд прогнозних станів для даної величини дискретизації Δt.

Далі по одержаному ряду станів та відомому початковому значенню відбувається відновлення ряду для даної дискретизації Δt. При цьому кожний стан являє собою Δt точок ряду. На етапі класифікації станів з кожним станом був зв'язаний середній приріст $r_{avg,i}$, який додається до значення останньої точки ряду та обчислюється наступна дискретизована точка. Проміжні точки заповнюються як лінійна інтерполяція відомих 2-х сусідніх точок. Алгоритм відновлення значень ряду y_t на основі початкової ціни p_t та ряду середніх приростів, $r_{avg,ik}$відповідних прогнозним станам s_k можна задати послідовністю обчислень (6):

$$y_t=p_t$$

$$y_{t+1}=y_t+r_{avg,i1}/\Delta t= p_t+r_{avg,i1}/\Delta t$$

$$y_{t+2}=y_{t+1}+r_{avg,i1}/\Delta t=p_t+2r_{avg,i1}/\Delta t$$

$$\ldots$$

$$y_{t+\Delta t-1}=y_{t+\Delta t-2}+r_{avg,i1}/\Delta t=p_t+(\Delta t-1)r_{avg,i}/\Delta t$$

$$y_{t+\Delta t}=y_{t+\Delta t-1}+r_{avg,i1}/\Delta t=p_t+\Delta t\ r_{avg,i}/\Delta t= p_t+ r_{avg,i1}$$

$$y_{t+\Delta t+1}=y_{t+\Delta t}+r_{avg,i2}/\Delta t=p_t+ r_{avg,i1}+ r_{avg,i2}/\Delta t$$

$$\ldots$$

$$y_{t+n\Delta t-1} = y_{t+n\Delta t-2} + r_{avg,in} / \Delta t = p_t + \sum_{k=1}^{n-1} r_{avg,ik} + \frac{(\Delta t - 1)}{\Delta t} r_{avg,ik}$$

$$y_{t+n\Delta t} = y_{t+n\Delta t-1} + r_{avg,in} / \Delta t = p_t + \sum_{k=1}^{n} r_{avg,ik} \tag{6}$$

6. Ієрархія приростів часу та процедура склеювання

Прирости часового ряду ми будемо обчислювати з різними кроками. Наприклад, аналогічно з дискретним перетворенням Фур'є, розглядаємо прирости величиною степенів двійки. Спочатку обчислюємо прирости як різниця двох сусідніх значень ряду, потім через одне, з кроками 2, 4, 8, 16 і т.д. Позначимо цю різницю в часі через Δt.

Для кожного Δt здійснюємо перетворення ряду приростів у ряд станів, здійснюємо прогнозування майбутньої послідовності станів, потім відновлюємо ряд із заданою дискретизацією по спрогнозованому ряду станів.

Ряди, отримані при відновленні для різних Δt, проходять процедуру склеювання, в результаті якої отримується ряд, який і є прогнозний ряд.

Таким чином, вибирається ієрархія приростів, кожна з яких відповідає за свою частоту, на якій відбувається прогнозування та відновлення в процесі склеювання.

Суть процесу склеювання полягає в наступному. Процедура склеювання є ітераційною, ряд з кожною наступною (з більшим кроком) дискретизацією коректує, підтягуючи до свої точки прогноз, сформований з результатів при менших Δt. Перетворення, які виконуються в процесі склеювання, можна записати у вигляді наступних обчислень:

Нехай проведене склеювання для усіх приростів часу $\Delta t < \Delta t_i$, здійснено прогноз при дискретизації Δt_i та за формулами (6) отримано часовий ряд $\{y_i\}$ Розглянемо ітераційну процедуру склеювання отриманого ряду $\{y_i\}$ з рядом, отриманим в результаті усіх попередніх склеювань $\{g_i\}$.

Оскільки ряд $\{y_i\}$ містить системні точки лише в моменти часу, кратні Δt_i, а інші точки ряду є інтерпольованими, то процес склеювання полягає в заміні цих інтерпольованих точок значеннями системних точок з попередніх $\Delta t < \Delta t_i$, які містяться у ряді результатів попередніх склеювань $\{g_i\}$. Алгоритм склеювання може бути записаний у вигляді послідовності обчислень (7)

$$z_t=g_t=p_t$$
$$z_{t+1}=g_{t+1}+(y_{t+\Delta ti}-g_{t+\Delta ti})/\Delta t_i$$
$$z_{t+2}=g_{t+2}+2(y_{t+\Delta ti}-g_{t+\Delta ti})/\Delta t_i$$
$$\ldots$$
$$z_{t+\Delta t-1}=g_{t+\Delta t-1}+(\Delta t_i-1)(y_{t+\Delta ti}-g_{t+\Delta ti})/\Delta t_i$$
$$z_{t+\Delta t}=g_{t+\Delta t}+(\Delta t_i)(y_{t+\Delta ti}-g_{t+\Delta ti})/\Delta t_i=y_{t+\Delta ti}$$
$$z_{t+\Delta t+1}=g_{t+\Delta t+1}+((y_{t+2\Delta ti}-g_{t+2\Delta ti})-(y_{t+\Delta ti}-g_{t+\Delta ti}))/\Delta t_i$$
$$z_{t+\Delta t+2}=g_{t+\Delta t+2}+2((y_{t+2\Delta ti}-g_{t+2\Delta ti})-(y_{t+\Delta ti}-g_{t+\Delta ti}))/\Delta t_i$$
$$\ldots$$
$$z_{t+n\Delta t-1}=g_{t+n\Delta t-1}+\frac{(\Delta t-1)}{\Delta t}\left(\left(y_{t-n\Delta t}-g_{t-n\Delta t}\right)-\left(y_{t-(n-1)\Delta t}-g_{t-(n-1)\Delta t}\right)\right)$$
$$\begin{aligned} z_{t+n\Delta t}&=g_{t+n\Delta t}+\left(\left(y_{t-n\Delta t}-g_{t+n\Delta t}\right)-\left(y_{t+(n-1)\Delta t}-g_{t+(n-1)\Delta t}\right)\right)= \\ &=g_{t+(n-1)\Delta t}-y_{t+(n-1)\Delta t}-y_{t-n\Delta t} \end{aligned} \tag{7}$$

Розглянемо процедуру склеювання часового ряду на прикладі фрагменту індексу PFTS:

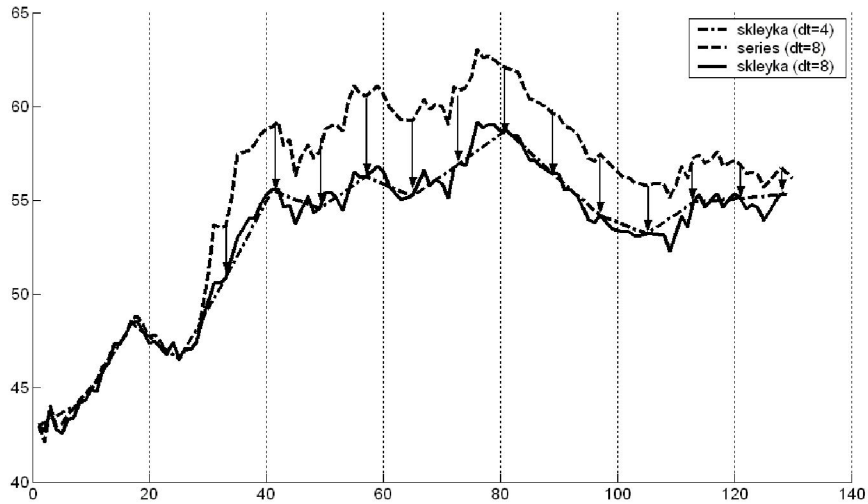

Рис. 1. Процес склеювання двох ієрархій часового ряду.

На рис. 1. зображено процес склеювання ряду, одержаного після склеювання прогнозів з дискретизаціями Δt=1,2,4 до нового ряду з дискретизацією Δt = 8. Як видно з рисунка, точки ряду, склеєного з Δt=1,2,4 (рис.1, пунктирна лінія) „підтягуються” до точок нового ряду з дискретизацією Δt=8 (рис. 1, штрихпунктирна лінія). В результаті отримується ряд, склеєний з Δt 1,2,4,8 (рис. 1, суцільна лінія).

Ієрархія приростів дискретизацій як степені числа 2 – не єдиний спосіб побудови ієрархії дискретизацій. Для покращення прогнозу можна використовувати більш часту ієрархію дискретизацій, наприклад, добуток степенів простих чисел.

Припустимо довжина ряду, який необхідно розкласти на ієрархії – N. Взявши послідовність перших простих чисел p_1, p_2, ..., p_n та послідовність їх степенів, при яких $p_1^{m1}\cdot p_2^{m2}\cdot\ldots\cdot p_n^{mn}=N_1<N$. Число N_1 повинне бути якомога ближчим до N, щоб збільшити кількість точок, які використовуються при прогнозуванні. Прості числа і їх степені слід вибирати таким чином, щоб кількість дільників цього числа було максимальним. Це забезпечило б максимальну щільність сітки точок дискретизацій. Оскільки N_1 є добутком заданих степенів простих чисел, то кількість дільників числа N_1 можна обчислить за наступною формулою: $(m_1+1)\cdot(m_2+1)\cdot\ldots\cdot(m_n+1)$. Це число ієрархій повинне бути максимізоване.

Задача полягає в тому, щоб знайти послідовність простих чисел та відповідні їх степені, для яких по-перше, добуток $p_1^{m1}\cdot p_2^{m2}\cdot\ldots\cdot p_n^{mn}=N_1$ було б близьке до N, а по-друге – кількість дільників цього числа було б максимальним. Для розв'язання цієї задачі ми використовуємо повний перебір варіантів, оскільки цей процес не займає значного обчислювального часу, а також проводиться лише один раз на початку прогнозування. Розглянемо приклад:

Для N=5000 необхідно підібрати такий набір простих чисел 2,3,5,7,…q_k, при якому число кроків дискретизації буде близьким до максимально можливого. Максимальне k та відповідне йому q_k визначається умовою:

$$\frac{\ln N}{k\ln q_k}\approx 1;\quad \Rightarrow\quad k\approx\frac{\ln N}{\ln q_k}.$$

$$N=5000;\ \ q_1=2;\ \ q_2=3;\ \ q_3=5;\ \ q_4=7;\ \ q_5=11;$$

$$k=3;\ \ \frac{\ln 5000}{\ln 5}=5,29;\ \ k=4;\ \ \frac{\ln 5000}{\ln 7}=4,37;\ \ k=5;\ \ \frac{\ln 5000}{\ln 11}=3,55.$$

Вибираємо $k=4$:

$$m_2=\frac{\ln 5000}{4\ln 2}=3,07;\ \ m_3=\frac{\ln 5000}{4\ln 3}=1,94;\ \ ;\ \ m_7=\frac{\ln 5000}{4\ln 7}=1,09;\ \ \Rightarrow$$

$$2^3 3^2 5^1 7^1=2520;\ \ \Rightarrow\ \ N=2^4 3^2 5^1 7^1=5040;\ \ \Rightarrow$$

$$K_{\max}=(m_2+1)(m_3+1)(m_5+1)(m_7+1)=5\cdot 3\cdot 2\cdot 2=60;\ \ \Rightarrow$$

$$\Delta t=1,2,3,4,5,6,7,8,9,10,12,14,15,16,18,20,21,24,28,30,35,36,40,42,45,$$
$$48,56,60,63,70,72,80,84,90,112,120,126,140,144,168,180,210,240,252,$$
$$280,315,336,360,420,504,560,630,720,840,1008,1260,1680,2520,5040.$$

Таким чином, замість 12 приростів у випадку простої ієрархії (2^{12}=4096), ми отримали 60 приростів, що дозволить створити більш густу сітку часових приростів, та за нашим припущенням, отримати більш точний прогноз.

7. Оптимізація алгоритму прогнозування, вибір оптимальних параметрів алгоритму

Описаний вище алгоритм має параметри, які необхідно підбирати експериментально. Для цього проводиться процедура оптимізації, яка полягає у тестовому прогнозуванні на відому частину ряду та оцінки результату.

Вихідний ряд розбивається на 2 частини, першу приймаємо як відому, а іншу як тестову. Будемо проводити тестові прогнози при різних параметрах та оцінимо відхилення прогнозу від реального значення.

Оскільки ми припускаємо, що для кожного приросту часу необхідно окремо підбирати параметри r та s, то пропонується наступний алгоритм оптимізації параметрів

1. Будуємо перше наближення – лінійний тренд;

2. Для кожного приросту часу Δt (дискретизації) в порядку спадання, проводимо підбір оптимальних параметрів. Для цього:

2.1. Для кожної пари параметрів (r, s), які ми будемо підбирати та для декількох початкових точок виконуємо наступні дії:

2.1.1. Здійснюємо прогноз на даній ієрархії та здійснюємо склеювання до всіх раніше побудованих прогнозів з більшим кроком Δt, використовуючи вже підібрані на минулих кроках оптимальні r та s.

2.1.2. Порівнюємо прогноз на величину Δt з реальними значеннями (тестовий проміжок). Обчислюємо похибку як сума квадратів відхилень відповідних точок прогнозу та реальних точок.

2.2. Таку похибку обчислюємо для декількох прогнозів з різних початкових точок. Знаходимо середнє значення похибок для даної пари r та s.

2.3. Із усіх пар r та s вибираємо ту, яка в середньому дає найменшу похибку.

2.4. Приймаємо кращу пару для даного кроку приросту часу Δt та переходимо до наступного рівню приросту в порядку спадання.

Таким чином, для кожного рівня приросту часу ми отримуємо оптимальні параметри, які будуть використовуватись в остаточному прогнозі. Припускається, що для певних рядів існує закономірність в оптимальних параметрах. Для виявлення цієї закономірності проводиться експериментальна робота, мета якої – отримати оптимальний набір параметрів для ряду без повного перебору варіантів.

8. Приклади апробації алгоритму прогнозування на штучних та реальних рядах

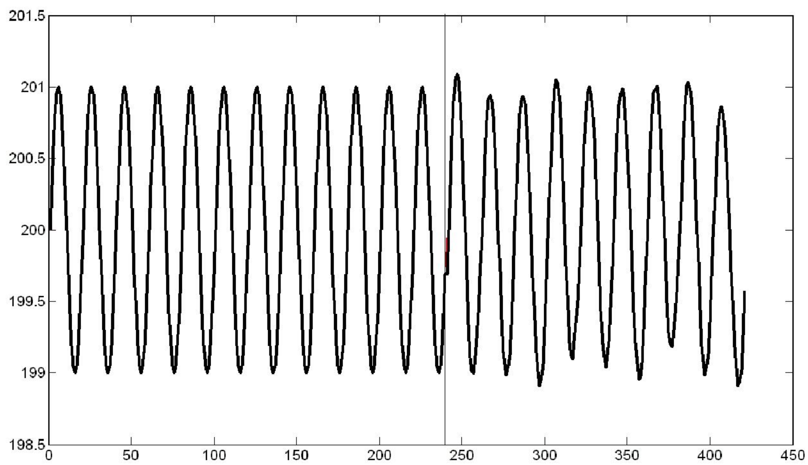

Рис. 2. Прогнозування функції $y=\sin\left(\frac{2\pi}{20}t\right)$ з параметрами k=5, s=9

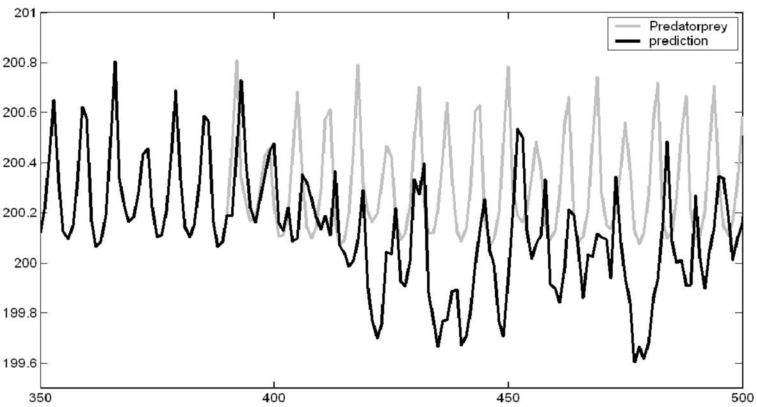

Рис. 3. Прогнозування ряду, згенерованого детермінованою моделлю «Хижак-жертва » при, *s*=5

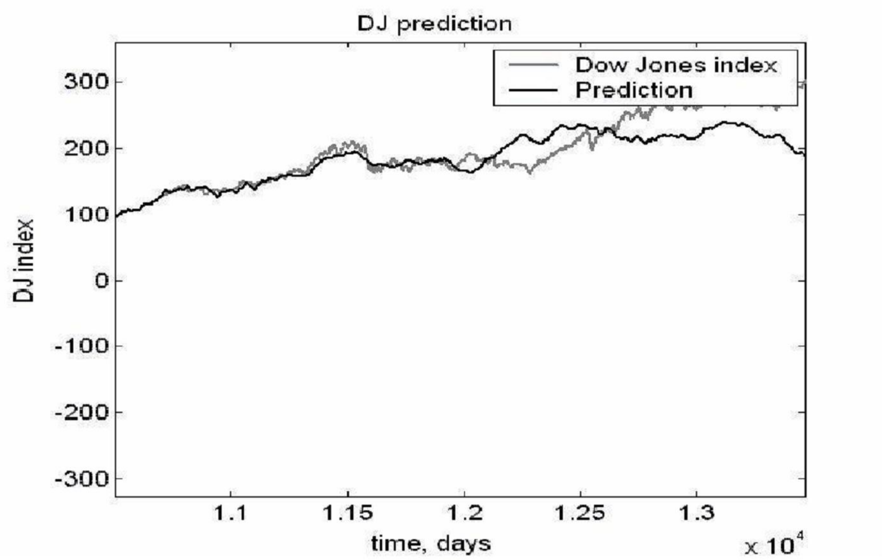

Рис 4. Прогнозування індексу Dow Jones (1940-1953 рр.)

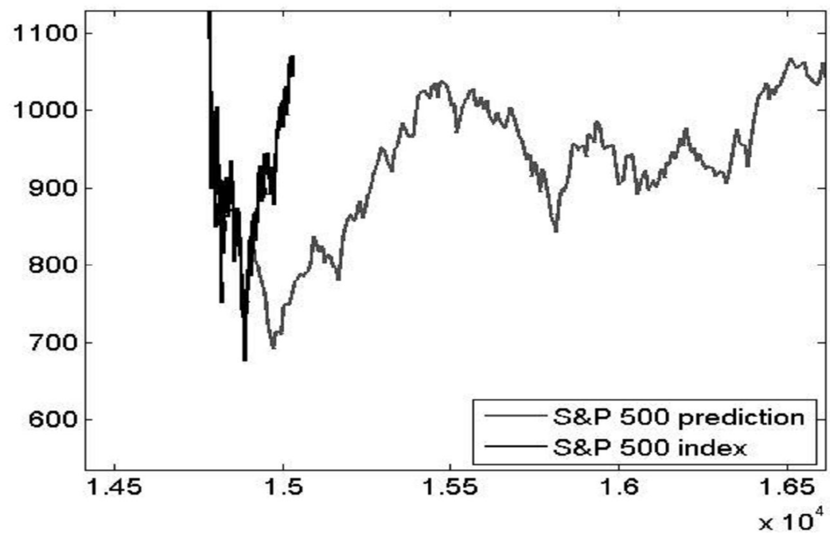

Рис 5. Прогнозування індексу S&P 500. Початок прогнозу – 15 травня 2009 р.

Висновки та подальша робота

В даній роботі запропоновано алгоритм прогнозування часових рядів на основі складних ланцюгів Маркова. Вперше запропоновано застосування складних ланцюгів Маркова, тобто ланцюгів Маркова з пам'яттю для аналізу фінансових часових рядів. Принцип ієрархії часових приростів дозволяє максимально повно використати інформацію, яка міститься у часовому ряді при побудові прогнозу. Експериментальна робота з прогнозування штучних сигналів та реальних фінансових рядів показує ефективність алгоритму та підтверджує актуальність подальших досліджень у даному напрямі. Дискусійні питання щодо доцільності використання терміну „ланцюги Маркова” до рядів з довгою та короткою пам'яттю не відносяться до мети даної роботи та будуть детально проаналізовані у найближчих публікаціях.

Автор висловлює щиру вдячність проф. Соловйову Володимиру Миколайовичу та доц. Сапціну Володимиру Михайловичу за керівництво роботою.

Література

1. Принципи моделювання та прогнозування в екології: Підручник. / [Богобоящий В.В., Курбанов К.Р., Палій П.Б., Шмандій В.М].- К.: Центр навчальної літератури, 2004. – 216 с.
2. Зайченко Ю. П. Нечеткие модели и методы в интелектуальных системах: учеб. пособие для иностр. студ. вузов, направления "Компьютерные науки" / Зайченко Ю. П.; [М.З. Згуровский (общ.ред.)]. – К.: Слово, 2008. — 344 с.
3. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применения в экономике и бизнесе. / Ежов А.А., Шумский С.А. – М ., 1998.
4. Заенцев И.В. Нейронные сети: основные модели. / Заенцев И.В. [електронний ресурс] :Учебное пособие к курсу «Нейронные сети» для студентов 5 курса магистратуры каф. электроники физического ф-та Воронежского государственного университета [эл. издание]. – Воронеж: Воронежский государственный университет, 1999.
5. Сапцин В.М. Опыт применения генетически сложных цепей Маркова для нейросетевой технологии прогнозирования. / Сапцин В.М. // Вісник Криворізького економічного інституту КНЕУ.- Кривий Ріг, КЕІ КНЕУ, 2009, вип. 2(18).- С.56-66.
6. Тихонов В.И., Миронов В.А. Марковские процессы. / Тихонов В.И., Миронов В.А. – М.: Сов. Радио, 1977. - 488 с.
7. Корн Г., Корн Т.. Справочник по математике для научных инженеров и работников. / Корн Г., Корн Т. – М.: Наука, 1973. - 832 с.
8. Курбанов К.Р., Сапцин В.М. Сложные цепи Маркова как технология прогнозирования социальных, экономических и экологических процессов. / Курбанов К.Р., Сапцин В.М. // МНПК аспір., мол. учених та науковців. «Проблеми та перспективи розвитку регіональної ринкової економіки» Кременчук, 11-13 травня 2007 р. С. 10-14.
9. Сапцін В. М., Чабаненко Д.М. Фур'є-продовження низькочастотних складових рядів економічної динаміки / Сапцін В. М., Чабаненко Д.М. // Проблеми економічної кібернетики: Тези доповідей XIV Всеукраїнської науково-методичної конференції (8-9 жовтня 2009 р., м. Харків). – Харків.: ХНУ імені В. Н. Каразіна, 2009. – С. 132-133.
10. Соловйов В.М. Математична економіка. [Навчально-методичний посібник для самостійного вивчення дисципліни] / Соловйов В.М. – Черкаси: ЧНУ, 2008. – 136 с.
11. Prediction of financial time series with the technology of high-order Markov chains [електронний ресурс] / Soloviev V., Saptsin V., Chabanenko D. // Working Group on Physics of Socio-economic Systems (AGSOE). – Drezden , 2009. – Режим доступу: http://www.dpg-verhandlungen.de/2009/dresden/agsoe.pdf
12. Блюмин С.Л. Дискретность против непрерывности в информационных технологиях: квантовое исчисление и его альтернативы / Блюмин С.Л. // Системы управления и информационные технологии. № 1,2(31), С. 217-221 (2004).

Прийнято до друку 15.10.2009

Аннотация

Д.Н. Чабаненко. Алгоритм прогнозирования финансовых временных рядов, основанный на сложных цепях Маркова

В работе предлагается применение технологи сложных цепей Маркова, т.е. цепей Маркова с памятью, для прогнозирования финансовых временных рядов. Главным отличием сложных цепей Маркова от простых является учет последействия, памяти. Сложная цепь Маркова может быть приведен к простому введением понятия „обобщенное состояние", как .кортеж последовательных дискретных состояний. Технология предполагает прогноз ряда, согласно иерархии интервалов дискретизации времени и последовательного „склеивания" результатов прогнозов на разных частотных уровнях. Предлагаемый подход позволяет наиболее полно использовать информацию, содержащуюся в ряде и получить наиболее адекватное его продолжение. Проведена экспериментальная проверка эффективности предлагаемого алгоритма на рядах фондовых, валютных и спотовых рынков.

Ключевые слова: прогнозирование, временные ряды, цепи Маркова, моделирование, сложные системы

Summary

D.N. Chabanenko. Financial time series prediction algorithm, based on complex Markov chains

In this research the technology of complex Markov chains, i.e. Markov chains with a memory is applied to forecast financial time-series. The main distinction of complex or high-order Markov Chains [1] and simple first-order ones is the existing of aftereffect or memory. The high-order Markov chains can be simplified to first-order ones by generalizing the states in Markov chains. Considering the "generalized state" as the sequence of states makes a possibility to model high-order Markov chains like first-order ones. The hierarchy of time discretizations gives a possibility to review long-memory properties of the series without increasing the order of the Markov chains, to make prediction on the different frequencies of the series. The technology is tested on several financial time-series: Forex, stock and spot markets.

Keywords: time series, prediction, Markov chains, complex systems.