

ЙМОВІРНІСНЕ ПРОГНОЗУВАННЯ ПРОЦЕСІВ ЦІНОУТВОРЕННЯ НА ФОНДОВИХ РИНКАХ

П.І. БІДЮК, А.В. ФЕДОРОВ

Запропоновано два типи математичних моделей для прогнозування процесів ціноутворення на біржі. Ймовірнісна модель у вигляді динамічної мережі Байеса та авторегресійна модель є взаємно доповнюючими, що сприяє підвищенню якості прогнозу і рішень щодо торгових операцій на біржі. Побудовано модель для прогнозування нестандартних ситуацій.

ВСТУП

Збільшення числа операцій з біржовими та банківськими активами вимагає підвищення якості оцінок прогнозів їх вартості з метою зменшення втрат та підвищення об'ємів прибутків. Задача прогнозування стосується як рівнів змінних, що характеризують ціни активів, так і множини відповідних біржових індексів, трендів процесів ціноутворення та дисперсії цін.

Задачі фінансового моделювання та прогнозування розглядаються, зокрема, в роботах [1 – 11].

Робота [1] поєднує основи теорії фінансів та математичної теорії їх моделювання з метою поглибленого дослідження фінансових процесів, оцінювання прогнозів та управління.

Роботу [2] присвячено аналізу сучасних адаптивних методів короткострокового прогнозування, які стали підручним інструментом для прогнозування фондового ринку, фінансових потоків різного призначення, техніко-економічних характеристик технологічних процесів і технічних систем, аналізу якості на виробництві.

У [3] наведено модифіковану методику Бокса–Дженкінса побудови регресійних моделей часових рядів, яку успішно застосовано до моделювання та прогнозування фінансово-економічних процесів, показників якості технологічних процесів і т. ін.

У [4, 5] розглядаються можливості прогнозування нелінійних динамічних процесів, у тому числі процесів із змінною дисперсією (гетероскедастичні процеси). Для цієї мети застосовуються регресійні методи, нейронні мережі, метод подібних траєкторій.

У [6] досліджується можливість прогнозування нестационарних процесів за допомогою алгоритму клонального відбору, який практично не залежить від типу нестационарності та збурень, що діють на процес.

У [7] аналізується застосування оберненого відображення Кастельжо в прогнозуючих нечітких нейронних моделях та наведено приклади успішного застосування даного методу до короткострокового прогнозування нелінійних відносно змінних нестационарних процесів різної природи.

У [8, 9] запропоновано методику побудови ймовірнісних байєсівських мереж (МБ) та аналізуються можливості її застосування до прогнозування станів процесів у бізнесі та фінансах. Прогноз, отриманий за таким методом, є ймовірнісним, тобто визначає ймовірність попадання значення змінної у деякий інтервал.

У [10] розглянуто можливості використання теорії екстремальних значень до розв'язання задачі їх прогнозування. Це також ймовірнісний метод, який ґрунтується на оцінюванні параметрів (моментів) відповідних розподілів.

Роботу [11] присвячено аналізу розподілів, відмінних від нормальних, та можливостей їх застосування до оцінювання та прогнозування фінансових ризиків з використанням методик VaR та CVaR. Наведено приклади оцінювання характеристик еліптичних розподілів, прийнятних для опису інвестиційних процесів.

ПОСТАНОВКА ЗАДАЧІ

Для експериментальних даних, представлених випадковою послідовністю $\{y(k)\}$, $k=0,1,\dots,N$ із довільним розподілом, побудувати структуру динамічної ймовірнісної МБ у вигляді спрямованого ациклічного графа. Розробити алгоритми навчання мережі та формування статистичного висновку на його основі. Запропонувати методику обчислення короткострокового ймовірнісного прогнозу рівня і дисперсії випадкової змінної з прийнятними статистичними характеристиками якості та встановити можливості його використання при прийнятті рішень щодо торгівлі активами на біржі. Виконати апробацію розробленої процедури прогнозування на реальних статистичних даних фондових бірж.

ПОБУДОВА МОДЕЛІ ПРОЦЕСУ

Модель випадкового процесу будемо у вигляді спрямованого ациклічного графа. Для формування причинно-наслідкових зв'язків між змінними можна скористатися взаємною інформацією або вибірковими кореляційними функціями, які легко обчислити при наявності статистичних даних.

Так, для визначення необхідності включення у рівняння регресії авторегресійної складової необхідно обчислити та дослідити автокореляційну функцію змінної $y(k)$. Порядок авторегресії визначається за допомогою автокореляційної функції, число коефіцієнтів якої, що відмінні від нуля в статистичному розумінні, і буде порядком авторегресії.

Коефіцієнти вибіркової автокореляційної функції (АКФ)

$$r_y(s) = r_{y(k)y(k-s)} = \frac{1}{N-1} \frac{\sum_{k=s+1}^N \{[y(k) - \bar{y}][y(k-s) - \bar{y}]\}}{\sigma_y^2}, \quad s=1,2,3,\dots,$$

де σ_y^2 — вибіркова дисперсія змінної $y(k)$; \bar{y} — вибіркове середнє; N — довжина вибірки. Число коефіцієнтів АКФ, відмінних від нуля в статистичному розумінні, вказує на оцінку порядку авторегресійної частини моделі.

Уточнити порядок авторегресійної складової можна за допомогою часткової автокореляційної функції (ЧАКФ)

$$\Phi_{11} = r(1); \quad \Phi_{22} = \frac{r_2 - r_1^2}{1 - r_1^2}; \quad \Phi_{ss} = \frac{r_s - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_{s-j}}{1 - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_j}.$$

ЧАКФ чіткіше відображає порядок АР-моделі завдяки відсутності впливу проміжних коефіцієнтів кореляції на вибрані значення змінної, які знаходяться на відстані більше одного періоду дискретизації вимірів. Тобто, коефіцієнт Φ_{11} характеризує степінь взаємозв'язку між сусідніми (в часі) значеннями змінної, а Φ_{22} — взаємозв'язок тільки між значеннями змінної, що розділені в часі двома періодами дискретизації.

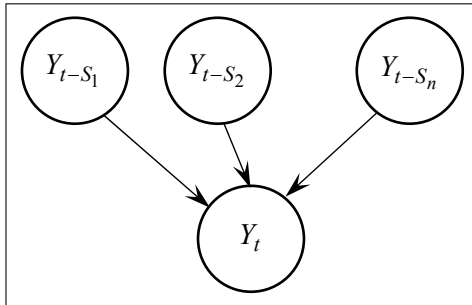


Рис. 1. Проста мережа, побудована за допомогою ЧАКФ

Для визначення структури мережі розраховуємо ЧАКФ і вибираємо змінні, степінь взаємозв'язку між якими найбільша. Далі формуємо просту динамічну мережу (рис. 1)

Змінна Y приймає одне із значень множини станів $M = [m_1, m_2, \dots, m_k]^T$, де m_1, m_2, \dots, m_k — стани змінної, що представляють собою числові інтервали виду $[a, b]$, де $a, b \in R$.

Для того щоб знайти оцінку прогнозу, необхідно визначити ймовірність

$$\begin{aligned} p[y(k) = m_{i_0}, y(k - s_1) = m_{i_1}, \dots, y(k - s) = m_{i_s}] = \\ = p[y(k) = m_{i_0} | y(k - s_1) = m_{i_1}, \dots, y(k - s) = m_{i_s}]. \end{aligned}$$

Маємо вектор ймовірностей \mathbf{p} . При цьому прогнозом буде значення $m(i)$, яке відповідає найбільшому елементу в \mathbf{p} .

Множина M формується наступним чином:

1. З вхідної послідовності даних X утворюється новий часовий ряд різниць

$$y(k) = x(k) - x(k - 1). \quad (1)$$

2. Розраховується крок розбиття на інтервали

$$\text{step} = \frac{\sum_{i=1}^q |y(i)|}{k_f (q - 1)}, \quad (2)$$

де q — довжина ряду; k_f — коефіцієнт розбиття, який впливає на довжину інтервалу, а, отже, і на кількість станів. Значення k_f вибирається аналітиком у відповідності до конкретної задачі.

3. Здійснюється розбиття на інтервали m_i

$$[\min y(k) + i \text{ step}, \min y(k) + (i + 1) \text{ step}].$$

Після побудови мережі та визначення множини станів виконується навчання мережі та прогнозування за однією із схем, показаних на рис. 2 і 3.

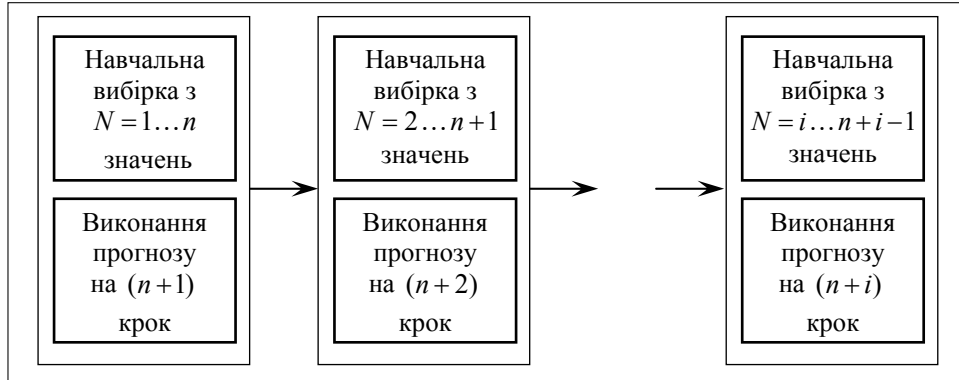


Рис. 2. Структурна схема прогнозування методом «вікна»

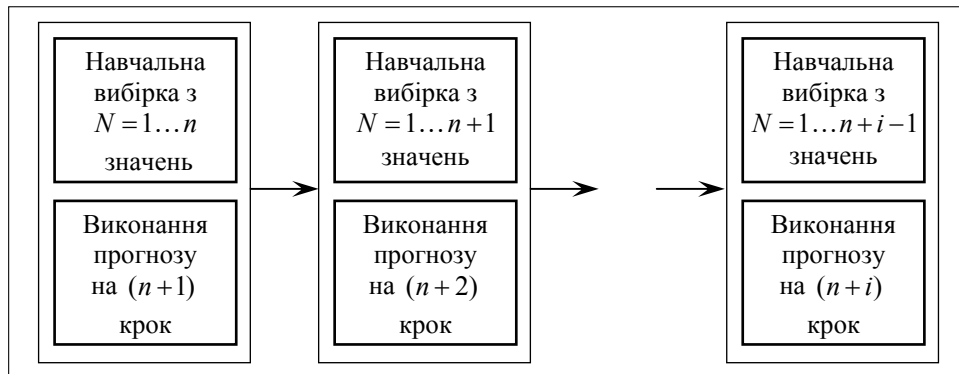


Рис. 3. Структурна схема прогнозування методом «накопичення»

Приклад № 1 (рис. 4). Розглянемо часовий ряд із 1667 значень, елементами якого є значення індексу ПФТС

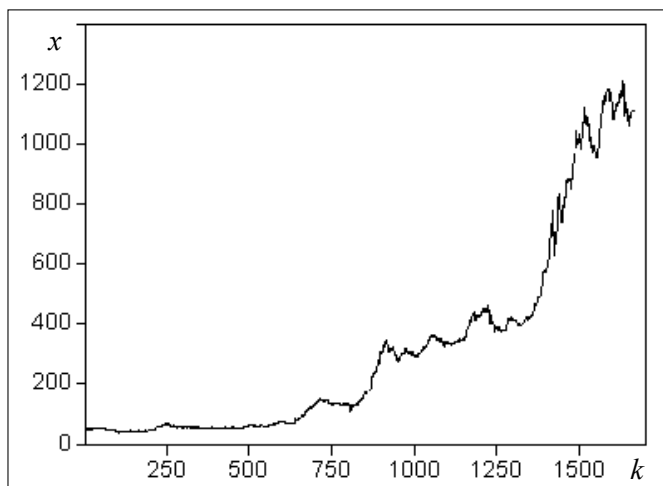


Рис. 4. Зміна в часі індексу ПФТС України

(Перша Фондова Торговельна Система) — найбільша біржа цінних паперів України.

Обробивши статистичні дані індикатором ренко (Renko) (висота блоку в ренко дорівнює 5), отримаємо вибірку $r(k)$ із значень локальних максимумів та мінімумів (78 елементів). Застосуємо для про-

гнозування спочатку Байєсів підхід. На основі аналізу ЧАКФ включаємо у мережу 1-й елемент. Навчальна вибірка (розмір ковзного вікна) складається з 50 значень, тестова — з 28, коефіцієнт розбиття — 1. Після розбиття отримуємо таку множину станів:

0	[- 12	- 9,1578947368421]
1	[- 9,1578947368421	- 6,3157894736842]
2	[- 6,3157894736842	- 3,4736842105263]
3	[- 3,4736842105263	- 0,6315789473682]
4	[- 0,6315789473684	2,21052631578947]
5	[2,21052631578947	5,05263157894737]
6	[5,05263157894737	7,89473684210526]
7	[7,89473684210526	10,7368421052632]
8	[10,7368421052632	13,5789473684211]

Точний прогноз стану системи отримали у 71,43% випадків.

Тепер спрогнозуємо значення індексу, використовуючи моделі авторегресії. ЧАКФ ряду даних для $y(k)$ показує, що модель повинна мати другий порядок по авторегресії

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k),$$

де $\varepsilon(k)$ — випадковий процес, який враховує вплив випадкових збурень на $y(k)$, невраховані змінні та інші можливі негативні впливи. Побудована модель має такі основні статистичні характеристики якості:

$$R^2 = 0,96; \text{ SSE} = 304906,5; \text{ AIC} = 11,21; \text{ DW} = 1,74,$$

де R^2 — коефіцієнт детермінації; SSE — сума квадратів похибок моделі; AIC — інформаційний критерій Акайке; DW — статистика Дарбіна-Уотсона. Всі характеристики задовільні. Статистичні характеристики якості однокрокового (історичного) прогнозу

$$\text{СКП} = 116,43; \text{ САП} = 93,58; \text{ САПП} = 18,39\%; U = 0,07,$$

де СКП — середньоквадратична похибка; САП — середня абсолютна похибка; САПП — середня абсолютна похибка в процентах; U — коефіцієнт Тейла.

Якщо включити в рівняння AP(2) трендову складову другого порядку, то модель буде мати такі характеристики:

$$R^2 = 0,96; \text{ SSE} = 304396,2; \text{ AIC} = 11,2; \text{ DW} = 1,74,$$

а однокроковий прогноз —

$$\text{СКП} = 109,66; \text{ САП} = 86,36; \text{ САПП} = 17,74\%; U = 0,066.$$

Включення у рівняння AP(2) трендової складової третього порядку дає

$$R^2 = 0,97; \text{ SSE} = 279654,2; \text{ AIC} = 11,18; \text{ DW} = 1,72.$$

Характеристики однокрокового прогнозу для цієї моделі

СКП = 81,74; САП = 62,25; САПП = 12,84%; $U = 0,048$.

Виділимо похибки останньої моделі в окремий ряд і включимо в модель авторегресії з ковзним середнім першого порядку нового ряду. Отримаємо характеристики

$R^2 = 0,97$; $SSE = 275883,9$; $AIC = 11,2$; $DW = 1,73$.

Характеристики однокрокового прогнозу моделі, описані вище,

СКП = 74,17; САП = 57,44; САПП = 11,9%; $U = 0,044$.

Таким чином, введення в модель ковзного середнього дало можливість покращити якість прогнозу.

Приклад № 2. Розглянемо процес формування промислового індексу Доу Джонса. Це індикатор фондового ринку, що представляє собою середній показник зміни курсу акцій групи, яка містить 30 найбільших американських промислових компаній. А саме: значення цін закриття до відомих подій в США у вересні 2001 р. (рис. 5).

Після обробки індикатором ZigZag (параметри обробки: $ExtDepth=20$, $ExtDeviation=5$, $ExtBackstep=3$) отримали вибірки: $z(k)$ — 271 (значення локальних мінімумів і максимумів індикатора) та $zd(k)$ — 270 (кількість днів між значеннями індикатора). Наша мета в даному випадку — спрогнозувати початок (область) глобального падіння індексу до 11 вересня 2001 р., використовуючи для навчання дані з 1971.08.10 до 2001.03.22. Включаємо у мережу перший елемент (рис. 6).

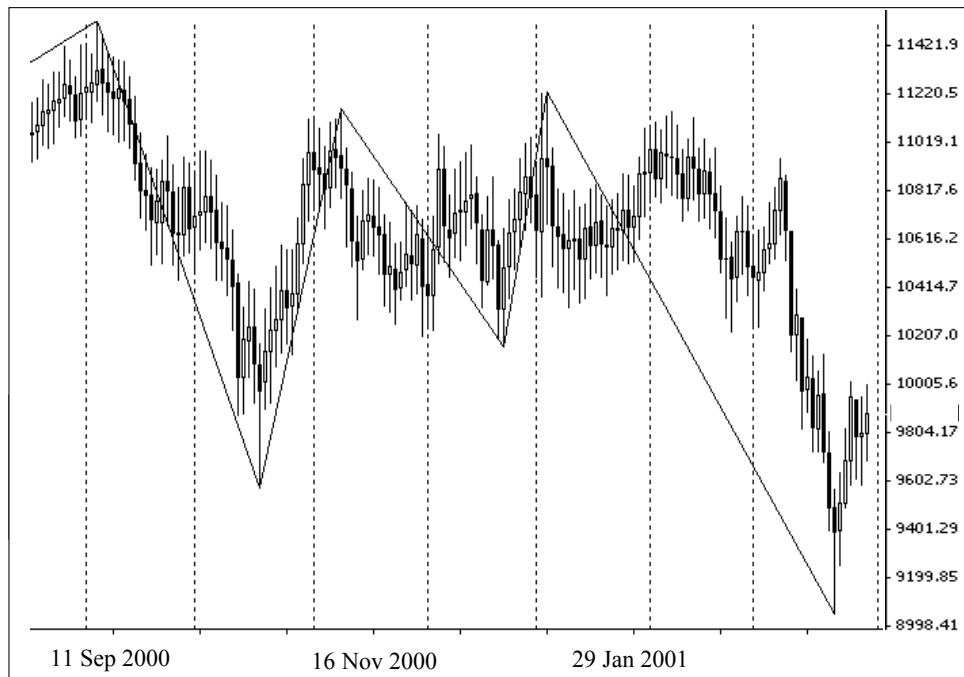


Рис. 5. Формування індексу Доу Джонса у 2000–2001 рр.

Навчальна вибірка для прогнозу $z(k)$ складає 100 значень, тестова — 171, коефіцієнт розбиття — 0,2.

Після розбиття маємо таку множину станів:

0	[-2296,7	-656,17962962963]
1	[-656,17962962963	984,34074074074]
2	[984,34074074074	2624,86111111111]
3	[2624,86111111111	4265,38148148148]

Точні оцінки прогнозу стану системи отримано у 81,66% випадків. Прогноз на наступний крок: значення індикатора буде знаходитися в інтервалі [10031,34; 11672,46] з ймовірністю 68,42%. Тобто окреслено область, в якій із зазначеною ймовірністю відбудеться падіння. Визначимо кількість днів, коли відбудеться формування локального максимуму індикатора з певною ймовірністю.

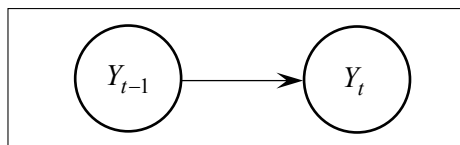


Рис. 6. Мережа, яка описує динаміку значень індикаторів

Рис. 6. Мережа, яка описує динаміку значень індикаторів

Навчальна вибірка для прогнозу $zd(k)$ складає 100 значень, тестова — 170, коефіцієнт розбиття — 0,4. Після розбиття маємо таку множину станів:

0	[-123	-60,973977695167]
1	[-60,973977695167	1,05204460966542]
2	[1,05204460966542	63,0780669144981]
3	[63,0780669144981	125,104089219331]
4	[125,104089219331	187,130111524164]

У 65,08% точно спрогнозовано стан системи, тобто інтервал, в який попадає значення змінної стану. Прогноз на наступний крок: кількість днів (інтервал, протягом якого відбудеться зміна напрямку тренду) буде знаходитися в проміжку [3; 53] з ймовірністю 78,57% (рис. 7).



Рис. 7. Прогнозування інтервалу, на протязі якого відбудеться зміна напрямку тренду

Прямокутник — область зміни тренду. Таким чином, нам вдалося спрогнозувати область (початок) падіння індексу (рис. 8).



Рис. 8. Спрогнозована область падіння індексу

Тобто з досить високою ймовірністю ($p \approx 68,42\%$) вдалося спрогнозувати інтервал зміни тренду та кількість днів, на протязі яких зміниться тренд ($p \approx 78,57\%$) напередодні теракту 11 вересня 2001 р.

ВИСНОВКИ

Побудовано ймовірнісні математичні моделі процесу ціноутворення для індексів ПФТС України та Доу Джонса на вибраних часових інтервалах. Кращою моделлю для короткострокового прогнозування виявилася ймовірнісна модель на основі динамічної мережі Байєса та модель авторегресії з ковзним середнім першого порядку (САПП=11,9%). Ці моделі є взаємно доповнючими, оскільки кожна з них несе різну інформацію щодо оцінки прогнозу процесу.

Встановлено, що застосування ймовірнісної моделі дає можливість підвищити якість короткострокового прогнозу складного стохастичного процесу ціноутворення. Це можна пояснити можливістю враховувати невизначеності різного характеру.

Застосування ймовірнісного підходу до аналізу індексу Доу Джонса дало можливість спрогнозувати падіння значення цього індексу напередодні подій 11 вересня 2001 р. в США. Спрогнозовано область його падіння. Це свідчить про можливість застосування ймовірнісних підходів до прогнозування погано прогнозованих нестационарних процесів на кілька періодів наперед.

У подальших дослідженнях всі отримані оцінки прогнозів будуть використані при побудові правил торгівлі на біржі, а також прийнятті рішень щодо менеджменту нестандартних ситуацій. Передбачається комбінування оцінок, отриманих за різними методами, що сприяє підвищенню якості прогнозу.

ЛІТЕРАТУРА

1. Уотшем Т.Дж., Паррамоу К. Количественные методы в финансах. — М.: ЮНИТИ, 1999. — 527 с.
2. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. — М.: Финансы и статистика, 2003. — 416 с.
3. Бідюк П.І., Баклан І.В. Системний підхід до прогнозування на основі моделей часових рядів // Системні дослідження та інформаційні технології. — 2003. — № 3. — С. 88–110.
4. Бідюк П.І., Баклан І.В., Литвиненко В.І. Моделювання та прогнозування нелінійних динамічних процесів. — Київ: ЕКМО, 2004. — 121 с.
5. Бідюк П.І., Литвиненко В.І., Слободенюк О.В. Моделювання і прогнозування гетероскедастичних процесів // Наук. праці Миколаївського держ. гуманіт. ун-ту ім. Петра Могили. — 2004. — Вип. 22. — С. 24–39.
6. Алгоритм клонального отбора для прогнозирования нестационарных динамических систем / П.И. Бидюк, В.И. Литвиненко, И.В. Баклан, А.А. Фефелов // Искусственный интеллект. — 2004. — № 4. — С. 89–99.
7. Бідюк П.І., Митник О.Ю. Обернене відображення Кастельжо в нечітких нейронних моделях // Системні дослідження та інформаційні технології. — 2004. — № 2. — С. 24–34.
8. Бідюк П.І., Терентьев А.Н., Гасанов А.С. Построение и методы обучения байесовских сетей // Кибернетика и системный анализ. — 2005. — № 4. — С. 24–34.
9. Бідюк П.І. Оцінювання і прогнозування стану малого підприємства за допомогою мережі Байеса // Наук. праці Миколаївського держ. гуманіт. ун-ту ім. Петра Могили. — 2006. — Вип. 44. — С. 7–29.
10. Бідюк П.І., Кроптя А.В. Аналіз і методи розв'язання задачі оцінювання екстремальних значень // Наук. вісті НТУУ «КПІ». — 2005. — № 4. — С. 34–47.
11. Бідюк П.І., Литинська А.Ю., Кравчук Ю.В. Оцінювання VaR та CVaR для квадратичного портфеля цінних паперів з факторами ризику, що розподілені за еліптичним законом // Наук. праці Миколаївського держ. гуманіт. ун-ту ім. Петра Могили. — 2007. — Вип. 55. — С. 8–18.

Надійшла 19.10.2007