

УДК 62-50

БІДЮК П.І., ДЕМКІВСЬКИЙ Є.О.

СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

Пропонується СППР при прогнозуванні динаміки стаціонарних і нестаціонарних процесів довільної природи. Розроблено дерево прийняття рішень при аналізі часових рядів, методи прогнозування на основі різницевих рівнянь для динаміки змінних та їх дисперсії. Наведені приклади застосування СППР до моделювання і прогнозування нестаціонарних процесів.

A decision support system (DSS) is proposed to be used for solving the problem of dynamics forecasting for stationary and nonstationary processes of various nature. A decision making tree is developed for time series analysis as well as forecasting techniques based on difference equations. Examples of DSS application are given to modeling and forecasting of nonstationary series and their variance.

Вступ

В останні десятиліття аналітична обробка даних привертає до себе велику увагу як в усьому світі, так і в Україні, адже в умовах ринкової економіки якість інформаційної підтримки діяльності керівників і аналітиків є одним з факторів досягнення успіху підприємства [1]. Керівникам вищого рангу притаманна робота в умовах неструктурованих або слабо структурованих проблем, в яких не завжди задані кількісні залежності між важливими характеристиками, а прийняття рішень значною мірою базується на творчому підході, інформованості, кваліфікації, таланті, інтуїції та інших рисах керівника. Для ефективного аналізу і прогнозування динамічних процесів довільної природи пропонується система підтримки прийняття рішень (СППР), яка б допомагала керівникам та менеджерам приймати ефективні управлінські рішення [2].

Постановка задачі

Метою даної роботи є створення технології розробки інформаційної системи підтримки прийняття рішень для розв'язку задач математичного моделювання і прогнозування нестаціонарних процесів. Система повинна забезпечувати виконання наступних задач:

1. *Моделювання процесу.* Використовуючи існуючі моделі реальних процесів (або створюючи нові), створювати підсистеми прогнозування їхнього протікання і підсистеми синтезу оптимальних рішень на основі поточних даних.

2. *Моделювання критеріїв.* За допомогою математичних методів знаходити математичне описання або правила для автоматичного об'єднання атрибутів, що характеризують різні варіанти рішень, що знімає когнітивні обмеження ОПР.

3. *Інформаційний менеджмент.* Для збереження, читання й обробки інформації, даних, знань використовувати сучасні комп'ютерні інформаційні технології та гнучкий і комфортний інтерфейс.

4. *Автоматизований і напівавтоматизований аналіз і логічний висновок.* Для часткової або повної автоматизації процесу логічного висновку використовувати методи штучного інтелекту і чисельні методи. Це дає можливість підвищити якість результату і зменшити час на розв'язок задачі.

5. *Способи підтримки представлення результатів.* Для того щоб реалізувати функції доступу до інших СППР, баз даних і знань, застосовувати засоби комп'ютерної графіки й інструментарій для обробки мов.

6. *Підвищення якості суджень.* З метою усунення систематичних помилок, що впливають з деяких кількісних евристичних суджень людини, впроваджуються статистичні й інші методи корекції результатів.

СППР забезпечує можливість вибору конкретних обчислювальних процедур для реалізації кожної функції моделювання та прогнозування поведінки часового ряду при умові наявності групи експертів, що впливають на прийняття рішень.

Структурна схема запропонованої СППР зображена на рис. 1.

Програмне забезпечення інтерфейсного модуля забезпечує наступні можливості: уведення запитів до бази знань на внутрішній формальній мові системи; редагування запитів користувача; виконання запитів до бази експертних знань і вивід результатів на екран; перегляд проміжних результатів роботи; поповнення і коректування бази експертних знань у режимі діалогу; верифікацію відповідей на запити.

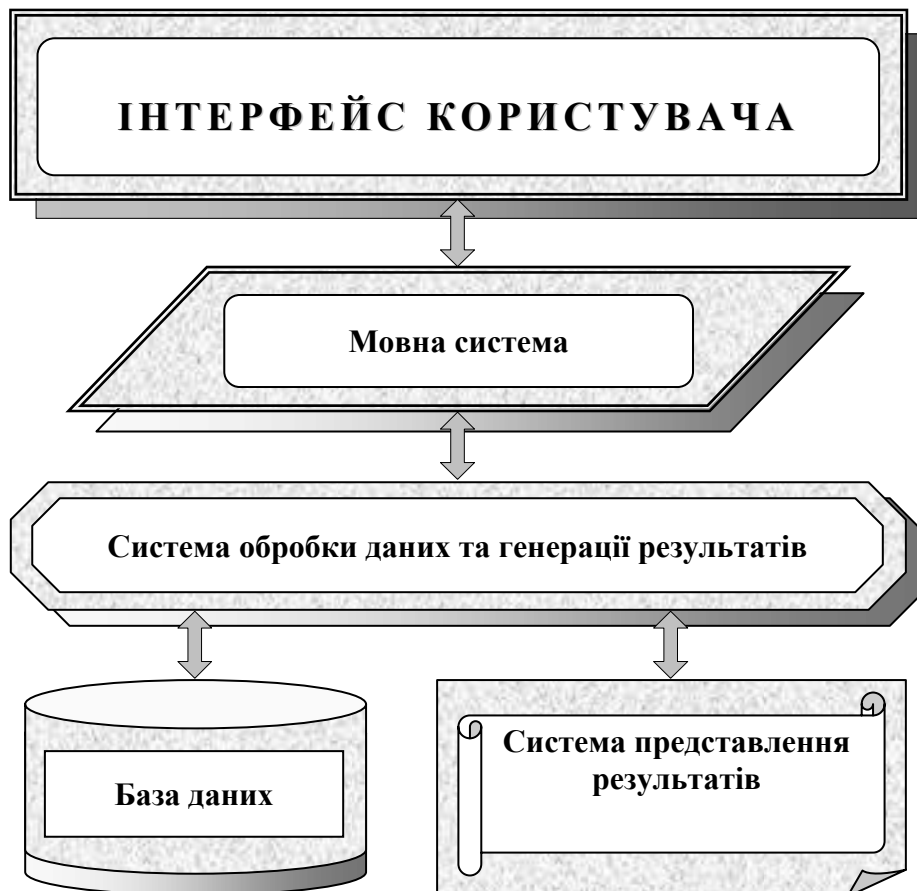


Рис. 1. Структурна схема СППР

Мовна система формує запити, які необхідно для організації діалогу користувач – система. Основними запитами мовної системи, є: запити на модифікацію та доповнення бази даних і знань; ведення нових алгоритмів оцінювання параметрів математичних моделей; розширення системи за рахунок нових алгоритмів прогнозування; запити на формування конкретних процедур обробки даних та прогнозування (формулювання вимог); запити на вибір та формулювання критеріїв розв'язку задачі; запити на виконання задач моделювання і прогнозування; запити на форму представлення результатів; запит на ретроспективний аналіз розв'язування подібних задач; перевірка запитів на коректність та генерація підказок користувачу; запит щодо поточного стану системи.

Системи обробки даних та генерації результатів (СОДГР) є головною системою (ядром) СППР з точки зору її основного призначення – підтримки прийняття рішень при розв'язку задач.

СОДГР приймає коректні запити від мовної системи і виконує наступні дії (які задаються запитами): поповнює (при необхідності) базу даних і знань; вибирає алгоритм (чи алгоритми) обробки даних з бази даних і знань; застосовує вибраний алгоритм до даних, що містяться в базі даних, з метою побудови математичної моделі, обчислення прогнозу або виконання інших функцій СППР; використовує критерії адекватності моделі з метою визначення ступеня адекватності побудованої моделі; застосовує критерії якості прогнозу для визначення кращого прогнозованого значення; при необхідності реалізує функцію ретроспективного аналізу результатів з метою порівняння останнього отриманого результату з результатами, які були отримані раніше; таке порівняння дає змогу визначити існування подібних ситуацій в минулому і теперішньому часі; передає отримані результати обчислень в систему представлення результатів, яка представляє отриманий результат у формі, заданій користувачем; зберігає, при необхідності, частину отриманих результатів в короткостроковій пам'яті для подальшого використання в поточній сесії застосування СППР.

Оскільки СППР призначена для прогнозування, то це такі алгоритми: алгоритми попередньої обробки даних (логарифмування, нормування, фільтрація, заповнення пропусків, обробка великих імпульсних значень); оцінювання параметрів (коефіцієнтів) математичних та/або статистичних моделей; алгоритми обчислення статистичних параметрів, які характеризують якість (адекватність) побудованої моделі; алгоритми обчислення прогнозованих значень (на основі рівнянь АР, АРКС, АРІКС, АРУГ, УАРУГ множинної регресії; за допомогою МГУА; за допомогою фільтра Калмана; за методом подібних траєкторій; методами експоненціального згладжування; нечітких множин; нейронних мереж і т.ін.); алгоритми обчислення показників якості прогнозів; алгоритми вибору кращої моделі та кращого прогнозу на основі розрахованих показників адекватності моделі та якості прогнозу; алгоритми формування бази результатів для проведення ретроспективного аналізу; алгоритми поповнення бази даних і знань.

База даних призначена для зберігання таких елементів: вихідні часові ряди, що характеризують динаміку процесів різної природи; тип джерела інформації та формати даних; модифіковані часові ряди (перші різниці та різниці вищих порядків, логарифмовані та нормовані дані); результати попереднього прогнозування, якщо таке було, для обраного часового ряду; типи моделей; коефіцієнти (параметри) математичних моделей; алгоритми оцінювання параметрів моделей; алгоритми прогнозування; критерії оцінювання якості прогнозу; інформаційно-довідкові матеріали.

Система представлення результатів приймає результати обробки даних та знань від СОДГР і надає їх користувачу в зручній для сприйняття формі.

Можливі форми представлення результатів роботи СОДГР наступні: гістограми, графічне зображення функцій розподілу ймовірностей отриманих величин; графічне представлення результатів у вигляді двовимірних та тривимірних графіків; застосування ліній різних типів для різних змінних на графіках; одночасне зображення поточних результатів роботи СППР та результатів, отриманих раніше – *ретроспективне порівняння* результатів; представлення результатів у вигляді таблиць зручного формату; поєднання тексту з цифровим матеріалом, таблицями і графіками; кругові та стовпчикові діаграми; тривимірні стовпчикові діаграми; використання можливостей гортання сторінок (*пейджинг*) та переміщення змісту екрану на один рядок за допомогою функції (*скролінг*).

Пропонується дерево прийняття рішень при аналізі часових рядів (рис. 2).

Технологія застосування дерева рішень складається з наступних кроків:

1. Попередня обробка та аналіз даних, що включає в себе: нормування, логарифмування, заповнення пропусків, згладжування екстремальних адитивних значень, оптимальне оцінювання змінних. Перехід на крок 2.
2. Перевірка наявності нелінійностей визначається за допомогою тесту Фішера або кореляційних функцій вищих порядків. Якщо процес містить нелінійності, то переходимо до кроку 3, інакше до кроку 4.
3. Визначення типу та порядку нелінійності, побудова моделі за МГВА і лінійна апроксимація процесу з метою порівняння з нелінійною моделлю. Перехід на крок 9.
4. Перевірка процесу на стаціонарність за допомогою тесту Дікі-Фуллера. Якщо процес стаціонарний переходимо до кроку 9, інакше до кроку 5.
5. Для перевірки наявності гетероскедастичності застосовуються тести: Уайта, Бройша-Пагана/Годфрі, Голдфельда-Квандта. Наявність тренду визначається за допомогою тесту Дікі-Фуллера. Аналіз коінтегрованості процесів проводиться за методикою Ігла-Грейнджера або Йохансена. Якщо процес гетероскедастичний, то переходимо до пункту 6, якщо містить тренд, то до кроку 7, інакше до кроку 8.
6. Визначення типу моделі для описання гетероскедастичності: УАРУГ(p, q), експоненційна УАРУГ(p, q), УАРУГ-М або інша. Після вибору найбільш підходящої моделі переходимо до кроку 9.
7. Визначення способу вилучення або моделювання тренду включає в себе визначення порядку інтегрованості процесу або можливістю описання тренду поліноміальною функцією, експоненціальною, логарифмічно або іншими функціями. Після вилучення тренду переходимо до кроку 9.
8. Будується модель часового ряду та обчислюються критерії адекватності отриманої моделі і переходимо до кроку 10.
9. На основі обраної моделі, будується функція прогнозування. Обчислюється прогноз поведінки ряду та визначаються оцінки точності прогнозу.
10. Якщо точність прогнозу не задовольняє особу, що приймає рішення, то уточнюються вхідні дані і переходимо до кроку 1. Інакше кінець.
11. Для коінтегрованих процесів будується модель корекції похибок і переходимо до кроку 1

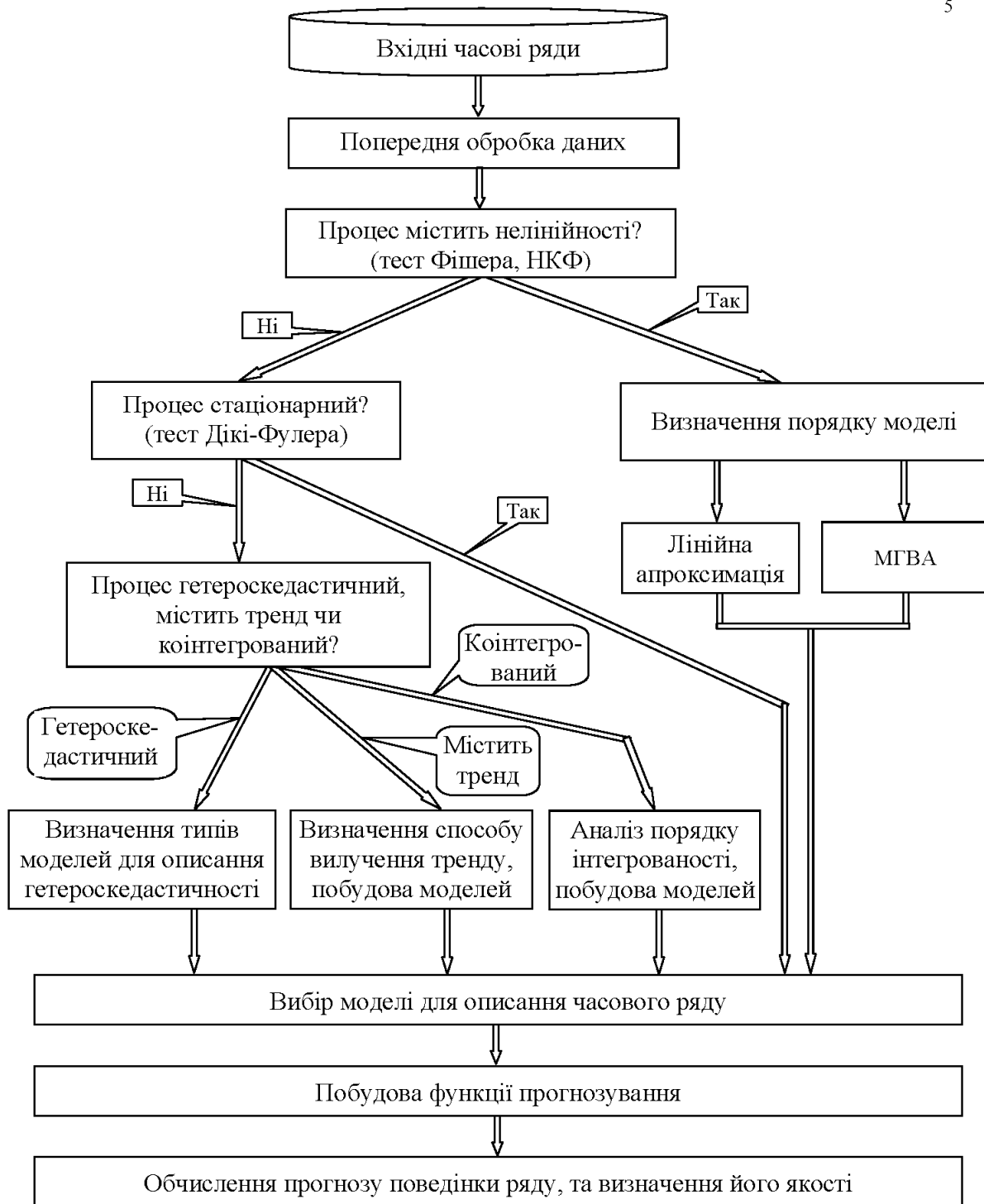


Рис. 2. Дерево прийняття рішень при аналізі часових рядів

Методи прогнозування в СППР

1. Прогнозування на основі різницевих рівнянь

Структура різницевого рівняння така, що дозволяє виконувати прогнозування на один крок (один період дискретизації вимірів) без додаткових перетворень. Тобто в праву частину необхідно підставити минулі значення змінних і обчислити оцінку прогнозу головної змінної в лівій частині. Але для того щоб знайти оцінку прогнозу на

більше число кроків, необхідно застосувати деякі попередні перетворення. Розглянемо деякі можливі підходи до обчислення прогнозованих значень.

Для рівняння AP(1)

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad E[\varepsilon(k)] = 0. \quad (1)$$

Функція прогнозування на s кроків можна записати у вигляді [8]

$$\hat{y}(k+s) = E_s[y(k+s)] = a_0 \left(\sum_{i=0}^{s-1} a_1^i \right) + a_1^s y(k) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k). \quad (2)$$

Отримане рівняння називають функцією прогнозування на довільне число кроків. Прогноз представляє собою збіжний процес, якщо $|a_1| < 1$, тобто

$$\lim_{s \rightarrow \infty} E_k[y(k+s)] = \frac{a_0}{1-a_1}, \quad (3)$$

де a_1 – знаменник геометричної прогресії. Вираз (3) свідчить про те, що для будь-якого стаціонарного процесу AP чи APKC оцінка умовного прогнозу асимптотично ($s \rightarrow \infty$) збігається до безумовного середнього.

Знайдемо похибку прогнозування $f_k(s) = y(k+s) - E_k[y(k+s)]$.

Похибка прогнозу на один крок

$$f_k(1) = y(k+1) - E_k[y(k+1)] = a_0 + a_1 y(k) + \varepsilon(k+1) - a_0 - a_1 y(k) = \varepsilon(k+1).$$

Похибка прогнозу на два кроки

$$\begin{aligned} f_k(2) &= y(k+2) - E_k[y(k+2)] = \\ &= a_0 + a_1[a_0 + a_1 y(k) + \varepsilon(k+1)] + \varepsilon(k+2) - E_k[y(k+2)] = \\ &= a_0 + a_0 a_1 + a_1^2 y(k) + a_1 \varepsilon(k+1) + \varepsilon(k+2) - a_0 - a_0 a_1 - a_1^2 y(k) = \\ &= \varepsilon(k+2) + a_1 \varepsilon(k+1). \end{aligned}$$

Таким чином, можемо записати вираз для похибки для довільного числа кроків прогнозування як

$$f_k(s) = \varepsilon(k+s) + a_1 \varepsilon(k+s-1) + a_1^2 \varepsilon(k+s-2) + \dots + a_1^{s-1} \varepsilon(k+1).$$

Враховуючи те, що $E[f_k(s)] = 0$, то оцінка прогнозу є незміщеною. Дисперсія похибки прогнозування

$$Var[f_k(s)] = \sigma^2 [1 + a_1^2 + a_1^4 + a_1^6 + \dots + a_1^{2(s-1)}],$$

тобто дисперсія є функцією s . Асимптотичне значення дисперсії похибки прогнозу для стаціонарного процесу $\lim_{s \rightarrow \infty} Var[f_k(s)] = \frac{\sigma^2}{1-a_1^2}$, де a_1^2 – знаменник геометричної прогресії.

Узагальнення функції прогнозування для процесу APKC (2, 1), який описується наступним рівнянням:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) + \beta_1 \varepsilon(k-1), \quad (4)$$

Рекурсивна формула для прогнозу на довільне число s кроків може бути записана тільки у такому вигляді [8]:

$$y(k+s) E_k[y(k+s)] = a_0 + a_1 E_k[y(k+s-1)] + a_2 E_k[y(k+s-2)]. \quad (5)$$

Якщо корені характеристичного рівняння, записаного для (5), знаходяться всередині одиничного кола, то оцінка прогнозу асимптотично збігається до безумовного середнього значенням

$$\lim_{S \rightarrow \infty} E[y(k+s)] = \frac{a_0}{1-a_1-a_2}, \quad \text{а для довільного процесу}$$

АРКС(p, q) оцінку умовного прогнозу можна записати як

$$y(k+s) = E_k[y(k+s)] = a_0 + \sum_{i=1}^p a_i E_k[y(k+s-i)].$$

Розглянемо побудову функції прогнозування на основі розв'язку рівняння АРКС (1, 1) $y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-1)$, $|a_1| < 1$, де $\varepsilon(k)$ – білий шум з нульовим середнім; $y(0) = y_0$ – відома початкова умова.

Розв'язок із урахуванням отриманого значення довільної константи:

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i-1) + \left[y_0 - \frac{a_0}{1-a_1} - \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(-i-1) \right] a_1^k$$

або

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i-1) + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k.$$

Рівняння для оцінки прогнозу відносно нульового моменту часу із врахуванням того, що на момент $k=0$ відоме значення збурення $E_0[\varepsilon(0)] = \varepsilon_0$ має такий вигляд:

$$y(0+k) = E_0[y(k)] = \frac{a_0}{1-a_1} + \beta_1 a_1^{k-1} \varepsilon_0 + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k. \quad (6)$$

Рівняння (6) можна розглядати як функцію прогнозування на k кроків наперед на основі інформації, яка є в наявності на момент $k=0$.

Функція прогнозування на s кроків наперед на основі інформації, яка є в наявності на момент k :

$$E_k[y(k+s)] = \left(\frac{a_0}{1-a_1} \right) (1-a_1^s) + \beta_1 a_1^{s-1} \varepsilon(k) + y(k) a_1^s.$$

Отримане рівняння являє собою функцію прогнозування на основі відомої інформації про процес на момент k , включно. Використовуючи наведені вище викладки, можна записати функції прогнозування для різного числа кроків таким чином:

$$\begin{aligned} s+1: & \quad E_k[y(k+1)] = a_0 + \beta_1 \varepsilon(k) + a_1 y(k); \\ s=2: & \quad E_k[y(k+2)] = \left(\frac{a_0}{1-a_1} \right) (1-a_1^2) + \beta_1 a_1 \varepsilon(k) + a_1^2 y(k); \\ s=3: & \quad E_k[y(k+3)] = \left(\frac{a_0}{1-a_1} \right) (1-a_1^3) + \beta_1 a_1^2 \varepsilon(k) + a_1^3 y(k). \end{aligned}$$

$$\text{При цьому } \lim_{s \rightarrow \infty} E_k[y(k+s)] = \frac{a_0}{1-a_1}.$$

Доволі просто можна перейти від моделі АРКС(1,1) до моделі АР(1), якщо покласти $\beta_1 = 0$. Для АР(1) отримаємо функцію прогнозування у вигляді:

$$E_k[y(k+s)] = \left(\frac{a_0}{1-a_1} \right) (1-a_1^s) + a_1^s y(k).$$

По аналогії можна знайти функції прогнозування для моделей іншої структури.

Якщо модель побудована на основі вибірки даних довжиною N , то для рівняння АРКС(2,1) $y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) + \beta_1 \varepsilon(k-1)$, функцію прогнозу можна записати наступним чином:

$$E_N[y(N+1)] = a_0 + a_1 y(N) + a_2 y(N-1) + \beta_1 \varepsilon(k);$$

$$E_N[y(N+2)] = a_0 + a_1 E_N[y(N+1)] + a_2 y(N);$$

$$E_N[y(N+s)] = a_0 + a_1 E_N[y(N+s-1)] + a_2 E_N[y(N+s-2)], \quad s \geq 2.$$

2. Прогнозування за допомогою методу подібних траєкторій

Суть методу полягає в такому: нехай даний ряд спостережень $\{y(1); y(2); \dots; y(n)\}$ графік якого поданий на рис. 3.

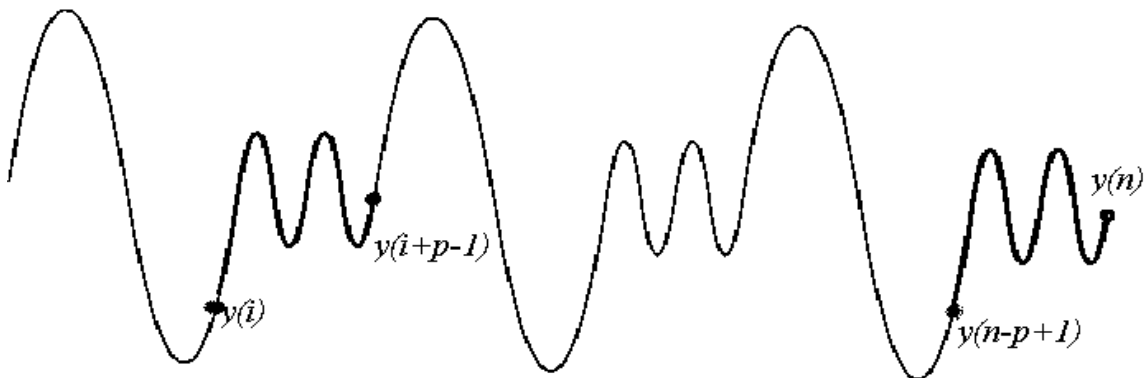


Рис. 3. Динаміка ряду спостережень

За обраним критерієм вибирається ділянка траєкторії “найближчу” до ділянки, що передуює прогнозній точці. Потім робиться прогноз за формулою: $\hat{y}(k+s) = y(I+J+s-1)$, де

$$I = \min_j \left\{ \sum_{i=1}^p |y(j+i-1) - y(n-p+i)| \right\} \quad j = 1, 2, \dots, n-p;$$

$$J = \min_i |y(I+i-1) - y(n)| \quad i = I, I+1, \dots, I+p-1.$$

Формалізація методу

Нехай Y_1, Y_2, \dots, Y_n наступні вектори

$$Y_1 = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} \quad Y_2 = \begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ y_{p+1} \end{pmatrix} \quad \dots \quad Y_k = \begin{pmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+p-1} \end{pmatrix} \quad \dots \quad Y_n = \begin{pmatrix} y_{n-p+1} \\ y_{n-p+2} \\ \vdots \\ y_n \end{pmatrix}.$$

Знаходимо найближчу точку із умови мінімальної відстані:

$$Y_k = \arg \min_j d(Y_n, Y_j).$$

3. Метод групового врахування аргументів (МГВА)

Основні ідеї методу групового врахування аргументів. Метод групового врахування аргументів (МГВА) базується на заданні правил ускладнення моделі, системі опорних функцій, критерії селекції та методі регуляризації, згідно із зовнішніми критеріями. ЕОМ проводить генерацію моделей-претендентів, селекцію, згідно з зовнішніми критеріями, та відсів моделей, що не пройшли селекцію. В зв'язку з чим основну структуру алгоритму самоорганізації можна навести у такому вигляді [7]:

- 1) попередня обробка спостережень з урахуванням системи обраних опорних функцій (скорочується кількість претендентів);
- 2) генерація множини моделей-претендентів;
- 3) обчислення критеріїв селекції, що є зовнішніми доповненнями, та пошук моделі оптимальної складності.

Головна ідея МГВА полягає в наступному: стверджується, що для задачі однокрокового прогнозу доцільно знизити точність визначення оцінок коефіцієнтів рівняння регресії, але за рахунок цього придати йому більшу регулярність. Тобто метою в цій задачі є не мінімізація помилок на вже відомих вузлах інтерполяції, а мінімізація помилок на нових точках, які в момент синтезу рівняння регресії ще не мали.

Розглянемо поліноміальні алгоритми методу групового урахування аргументу. Послідовність вихідних даних ділиться на перевірочну та навчальну. M – кількість вузлів інтерполяції; m — кількість членів повного поліному регресії. При $m > M$ розв'язок можливо отримати тільки за допомогою МГВА.

Нехай повний опис об'єкту задається деякою залежністю $y = \varphi(x_1, x_2, \dots, x_n)$. Замінімо цей вираз кількома рядами часткових описів. Перший ряд селекції:

$$y_1 = f(x_1, x_2), y_2 = f(x_1, x_3), \dots, y_s = f(x_{n-1}, x_n), s = C_n^2.$$

Другий ряд селекції:

$$z_1 = f(y_1, y_2), z_2 = f(y_1, y_3), \dots, z_p = f(z_{n-1}, z_n), p = C_s^2 \dots$$

Різноманітні алгоритми МГВА розрізняються по виду функції часткового опису $f(\cdot)$. Розглядаються три види часткових описів:

лінійний частковий опис – $f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j$

неповний квадратичний частковий опис –

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j$$

квадратичний частковий опис –

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2, \text{ де } A_i - \text{ нечітке число.}$$

З ряду в ряд селекції за допомогою порогових відборів передається тільки деяка кількість самих регулярних або незміщених змінних. Звичайно ця кількість однакова на всіх рядах селекції та дорівнює F .

Правило зупинки селекції: ряди селекції нарощуються до тих пір, доки критерій незміщеності розв'язків зменшується, така кількість називається свободою вибору. Для запобігання перенавчання, селекцію при досягненні мінімуму необхідно зупинити.

Зовнішні критерії оптимальності

При селекції моделей використовуються критерії регулярності та незміщеності. Загальним для сформульованих нижче критеріїв є те, що вони використовуються як зовнішні доповнення, тобто визначаються на основі перевірочної вибірки та в значній мірі відповідають вимозі відтворюваності (можливості повторення) результату.

Реалізація вибірки N ділиться на реалізацію навчальної вибірки N_A , за допомогою якої оцінюються параметри моделі, та реалізацію перевірочної вибірки N_B , за допомогою якої відбувається вибір придатної моделі.

Критерій регулярності визначає середньоквадратичне відхилення моделі на

первірочній вибірці:
$$\Delta^2(B) = \frac{\sum_{t \in N_B} (y_t^M - y_t)^2}{N_B} \rightarrow \min.$$

Якщо виходити з того, що при постійному комплексі умов добра апроксимація в минулому гарантує добру апроксимацію в найскорішому майбутньому, то критерій регулярності можна рекомендувати для короткострокового прогнозу, тому що розв'язок, отриманий на нових реалізаціях, дає лише невелике відхилення вихідних даних. При цьому в процесі селекції можуть бути загубленими важливі змінні, вплив яких, однак, непрямо буде урахований через інші змінні.

Критерій незміщеності (несуперечності) моделі. В його основу закладений той факт, що для одного об'єкту досліджень по різних вибірках даних, отриманих від нього при інших рівних умовах, повинні бути отримані близькі моделі, що визначають поведінку об'єкту.

$$\text{Критерій можна записати наступним чином: } n_{cm} = \frac{1}{R_1 + R_2} \cdot \sum_{r=1}^{R_1+R_2} (z_r^* - z_r^{**})^2,$$

де R_1, R_2 – розміри першої та другої підвбірок даних відповідно, z_r^*, z_r^{**} — значення прогнозу першої та другої моделі відповідно до всіх точок вибірок.

4. Метод експоненційного згладжування

Методи експоненційного згладжування почали бурхливо розвиватися в середині 60-х років. Вони виникли для одночасного прогнозування руху великої кількості товарів на складах. Завдяки своїй простоті та ефективності методи експоненційного згладжування стали одними з найпопулярніших методів прогнозування. Найбільш поширеними з них є методи простого, подвійного та потрійного згладжування, методи Холта та Холта-Вінтерса [12]. Метод простого експоненційного згладжування може бути застосований при явній відсутності тенденції та сезонності в часовому ряді. Метод подвійного експоненційного згладжування та метод Холта можуть застосовуватись при наявності лінійної тенденції, метод потрійного експоненційного згладжування застосовується при наявності нелінійної тенденції, а методом Холта-Вінера краще скористатися при наявності як тенденції, так і сезонності, які комбінуються або адитивно, або мультиплікативно.

Ідея методу простого експоненційного згладжування полягає в згладжуванні часового ряду ковзною середньою з експоненційними вагами. Така середня більше характеризує значення процесу на кінці інтервалу згладжування, ніж на початку. Назва методу впливає з того, що дані згладжуються за допомогою зваженої середньої, в якій ваги змінюються за експоненційним законом. Формалізовано метод простого експоненційного середнього можна подати таким чином. Нехай ми маємо ряд спостережень за змінною e : y_1, y_2, \dots, y_n . Експоненційне середнє першого порядку розраховується за формулою:

$$S_n^1 = \sum_{j=0}^{n-1} \alpha(1-\alpha)^j y_{n-j}, \text{ де } S_n^1 - \text{експоненційне середнє першого порядку для } n\text{-го}$$

періоду, α – коефіцієнт згладжування.

$$0 < \alpha < 1; \alpha(1-\alpha)^j \geq 0; \quad \sum_{j=0}^{n-1} \alpha(1-\alpha)^j \Rightarrow 1.$$

Експоненційне середнє першого порядку є прогнозом досліджуваного показника в $(n+1)$ періоді, тобто: $y_{n+1} = S_n^1$. Для експоненційного середнього дійсна така рекурентна формула: $S_n^1 = \alpha y_n + (1-\alpha)S_{n-1}^1$. Використання рекурентної формули потребує задавання початкового значення. Як його задати буде розглянуто нижче. Слід відзначити, що недоліком методу є відсутність формалізації при виборі параметра згладжування α .

Методи вибору α

Значення параметру α впливає на результати прогнозу. Якщо значення α близьке до одиниці, то при прогнозуванні враховується в основному вплив останніх спостережень; якщо α близьке до нуля, то ваги, за якими зважуються рівні ряду, спадають повільно, що дозволяє врахувати минулі значення. Для розв'язання практичних завдань Браун рекомендував вибирати параметр α в межах від 0.1 до 0.3, але його рекомендація нічим особливо не обгрунтована. Інколи α вибирають за формулою: $\alpha = \frac{2}{m+1}$, де m – кількість спостережень, які входять в інтервал згладжування.

Можна вибирати значення α , використовуючи таку ітеративну процедуру [9]:

1. Вибрати один із критеріїв оцінки якості прогнозу, наприклад СКП, СПП, АСПП та інші.

2. Розробити область зміни параметра α на значення, які змінюються з певним кроком, наприклад, з кроком 0.1. Тоді отримаємо підмножину значень α , яка дорівнює: [0; 0.1; 0.2; ...; 0.9; 1].

3. Вибрати початкове наближення, наприклад $S_0^1(y) = y_1$.

4. Для кожного значення α з побудованої підмножини провести розрахунок методом експоненційного згладжування.

5. Розрахувати значення одного з обраних критеріїв: MSE, MAE, MAPE...

6. Вибрати α , для якого ми маємо найменше значення критерію.

Звичайно, такий спосіб полегшує завдання вибору параметра згладжування, але має деякі незручності: по-перше, замість того, щоб зберігати лише попереднє значення прогнозу, необхідно зберігати весь ряд спостережень; по-друге, метод вимагає великої кількості операцій, крім того, необхідно зберігати прогнозні значення та помилки прогнозу для усіх значень α .

Вибір початкового наближення

Слід відзначити, що важливу роль для методів простого, подвійного та потрійного експоненційного згладжування відіграють експоненційні середні різних порядків. Ми їх знаходимо, використовуючи рекурентну формулу. При цьому постає питання про початкові наближення. Єдиного підходу до задання початкових наближень немає, часто їх задають, виходячи з умов економічного дослідження. Наприклад, як початкове наближення вибирають $S_0^1(y) = y_1$, особливо коли маємо ряд спостережень з флуктуаціями (біржовий курс і т. ін.). Інколи вибирають середнє значення, особливо коли даний ряд коливається навколо свого середнього (продаж товару, попит на який

стабільний у часі, і т. ін.): $S_0^1(y) = \frac{\sum_{i=1}^n y_i}{n}$.

Крім того, як початкові наближення можна вибирати середні декількох початкових значень; можна взяти наближення у вигляді параметра. Модифікації методу експоненційного згладжування можна знайти в [13].

5. Моделювання трендів

Детерміновані тренди описуються поліномом від часу, експонентою або лінійною комбінацією гармонічних процесів.

Випадковий крок плюс шум плюс дрейф

Модель випадкового кроку з шумом та дрейфом має вигляд:

$$y(k) = \mu(k) + \eta(k),$$

$$\mu(k) = a_0 + \mu(k-1) + \varepsilon(k), \quad (7)$$

де a_0 – константа; $\{\varepsilon(k)\}$, $\{\eta(k)\}$ – незалежні процеси білого шуму. В даному випадку тренд $\mu(k)$ містить випадкову складову $\varepsilon(k)$ та детерміновану a_0 . Розв'язок

для $\mu(k)$: $\mu(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i)$, що приводить до виразу для $y(k)$ у вигляді:

$$y(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k). \quad (8)$$

Враховуючи початкову умову $y_0 = \mu_0 + \eta_0$, рівняння для $y(k)$:

$$y(k) = y_0 - \eta_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k) \quad (9)$$

Таким чином, $y(k)$ являє собою суму детермінованого і стохастичного трендів, а також чисто випадкової складової $\eta(k)$. Шумова складова не обов'язково повинна бути білим шумом. Рівняння (9) переписано у вигляді:

$$y(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + A(L)\eta(k), \quad (10)$$

де $A(L)$ – поліном відносно оператора затримки L ; $A(L)\eta(k)$ – стаціонарний шумовий процес. Рівняння (10) називають *моделлю узагальненого тренду з нерегулярною складовою*.

Локальний лінійний тренд

Модель локального лінійного тренду (ЛЛТ) поєднує в собі кілька процесів випадкового кроку з шумом:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= \mu(k-1) + \lambda(k) + \varepsilon(k), \\ \lambda(k) &= \lambda(k-1) + v(k), \end{aligned} \quad (11)$$

де $\{\eta(k)\}$, $\{\varepsilon(k)\}$, $\{v(k)\}$ – незалежні процеси білого шуму. В даному випадку зміни тренду спричиняються процесом випадкового кроку та шумовою складовою, тобто $\Delta\mu(k)$ складається з процесу випадкового кроку $\lambda(k)$ та білого шуму $\varepsilon(k)$. Можна легко показати, що розглянуті вище процеси (випадковий крок плюс шум та випадковий крок плюс дрейф і шум) являють собою окремі випадки ЛЛТ.

Розв'язок для $y(k)$ має вигляд [12]:

$$y(k) = y_0 + [\eta(k) - \eta_0] + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + (k-1)v(2) + \dots + v(k).$$

Кожний елемент послідовності $\{y(k)\}$ містить детермінований тренд $(k(\lambda_0 + v(1)) + (k-1)v(2) + \dots + v(k))$, стохастичний тренд $(\sum \varepsilon(i))$ та нерегулярну компоненту $\eta(k)$. Звичайно, що в узагальненій формі моделі ЛЛТ нерегулярна компонента визначатиметься членом $A(L)\eta(k)$. Детермінований тренд залежить у даному випадку від поточних та минулих значень послідовності $\{v(k)\}$. Якщо в момент k сума $(\lambda_0 + v(1) + \dots + v(k))$ буде додатною, то коефіцієнт при k буде додатним. Очевидно, що в загальному випадку ця сума може бути додатною для деяких k , а для інших від'ємною, а тому тренд може мати ділянки з додатним та від'ємним нахилом.

Функція прогнозування на s кроків для ЛЛТ:

$$y(k+s) = y_0 + [\eta(k+s) - \eta_0] + \sum_{i=1}^{k+s} \varepsilon(i) + (k+s)(\lambda_0 + v(1)) + \\ + (k+s-1)v(2) + (k+s-2)v(3) + \dots + v(k+s)$$

або

$$y(k+s) = y(k) + [\eta(k+s) - \eta(k)] + \sum_{i=1}^k \varepsilon(i) + s[\lambda_0 + v(1) + v(2) + \dots + v(k)] + \\ + \sum_{i=1}^s (s+1-i)v(k+i).$$

Умовне математичне сподівання:

$$E_k [y(k+s)] = [y(k) - \eta(k)] + s[\lambda_0 + v(1) + v(2) + \dots + v(k)].$$

Нахил тренду визначається сумою $s[\lambda_0 + v(1) + v(2) + v(3) + \dots + v(k)]$.

Стохастичні тренди описуються на обмежених часових інтервалах такими моделями.

Стохастична модель волатильності

Стохастична модель волатильності, запропонована в роботі [83], може бути представлена як

$$x(k) = \psi_0 + \psi_1 x(k-1) + y(k), \quad (12)$$

$$y(k) = \sqrt{h(k)}u(k), \quad u(k) \approx N(0,1), \quad (13)$$

$$\log h(k+1) = \alpha + \phi \log h(k) + \eta(k), \quad \eta(k) \approx N(0, \sigma^2), \quad (14)$$

де $x(k)$ – часовий ряд статистичних даних, на основі яких будується модель; $u(k)$ і $\eta(k)$ – це стохастичні незалежні процеси білого шуму. У формулюванні Тейла припускається, що ψ_0 та ψ_1 дорівнюють нулю. Процес AR(1) з інновацією $y(k)$ в часовому ряду, що визначається рівнянням (12), пояснює існування можливої авторегресії в процесі $x(k)$. Традиційні труднощі оцінювання параметрів моделі стохастичної волатильності можна вирішити завдяки розвитку комп'ютерних технологій, які фактично дали можливість реалізації методу Монте Карло для марковських ланцюгів (МКМЛ). Таким чином, на сьогодні практичне застосування МКМЛ можливе так само, як і моделювання рівнянь АРУГ.

Порогова стохастична модель волатильності

Нижче пропонується новий підхід до фіксації асиметрії збурень часового ряду. Оскільки було встановлено, що дисперсія має тенденцію зростати під впливом поганих (невтішних) новин, то цілком імовірно, що динаміка процесу в рівнянні (14) визначається величиною доходів на попередньому відрізку часу. В цьому відношенні порогові моделі з достатньою адекватністю описують асиметрію механізму генерації дисперсії. Існує гіпотеза, що величина доходів залежить від знаку збурень. Цей вид асиметрії доходів також можна взяти до уваги, узагальнюючи рівняння (12) до кусочно-лінійної структури. Отже, природніше буде розглянути порогові нелінійні структури, ніж авторегресійні процеси, представлені рівняннями (12) – (14).

Визначено набір випадкових змінних $s(t)$ Бернуллі:

$$s(k) = \begin{cases} 0, & \text{якщо } x(k-1) < 0, \\ 1, & \text{якщо } x(k-1) \geq 0. \end{cases} \quad (15)$$

Порогова модель стохастичної волатильності має такий вигляд:

$$x(k) = \psi_{0s(k)} + \psi_{1s(k)}x(k-1) + y(k), \quad (16)$$

$$y(k) = \sqrt{h(k)} u(k), \quad u(k) \sim N(0,1), \quad (17)$$

$$\log h(k+1) = \alpha_{s_k+1} + \phi_{s_k+1} \log h(k) + \eta(k), \quad \eta(k) \sim N(0, \sigma^2). \quad (18)$$

Як і в початковому формулюванні, $u(k)$ та $\eta(k)$ є стохастично незалежними. В момент часу $k-1$, коли має місце несподіване пониження ціни внаслідок надходження невтішних новин, $x(k-1) < 0$ та $s(k) = 0$. Навпаки, якщо є хороші новини в момент часу $k-1$, то $x(k-1) > 0$ і $s(k) = 1$. Тому значення $s(k)$ визначається знаком величини $x(k-1)$. В пороговій моделі стохастичної волатильності значення параметрів ψ_0, ψ_1, α та ϕ переключаються між цими двома режимами, які відповідають підвищенню та падінню цін активів.

У симетричному випадку обидва набори параметрів ідентичні між собою. Зокрема, якщо $\phi_0 = \phi_1$, то із того факту, що $\alpha_0 \geq \alpha_1$, випливає, що дисперсія буде вищою, коли минулий дохід від'ємний, ніж коли він додатний. В узагальненій моделі ϕ_0 може відрізнитись від ϕ_1 . Дійсно, коефіцієнт, ϕ_s , характеризує величину впливу минулого доходу на поточну дисперсію. Якщо ϕ_0 більше, ніж ϕ_1 , то дисперсія в минулі періоди буде мати більший вплив на поточну дисперсію після падіння цін, ніж після підвищення цін. Цей вид асиметрії не розглядався в літературі з аналізу стохастичної волатильності.

Приклади застосування СППР

Приклад 1. Безпосереднє моделювання часового ряду

Наведемо приклад застосування створеної СППР для ідентифікації та прогнозування ряду спостережень ціни акцій компанії "Мотор-Січ" на фондовій біржі, динаміка якого наведена на рис 3.

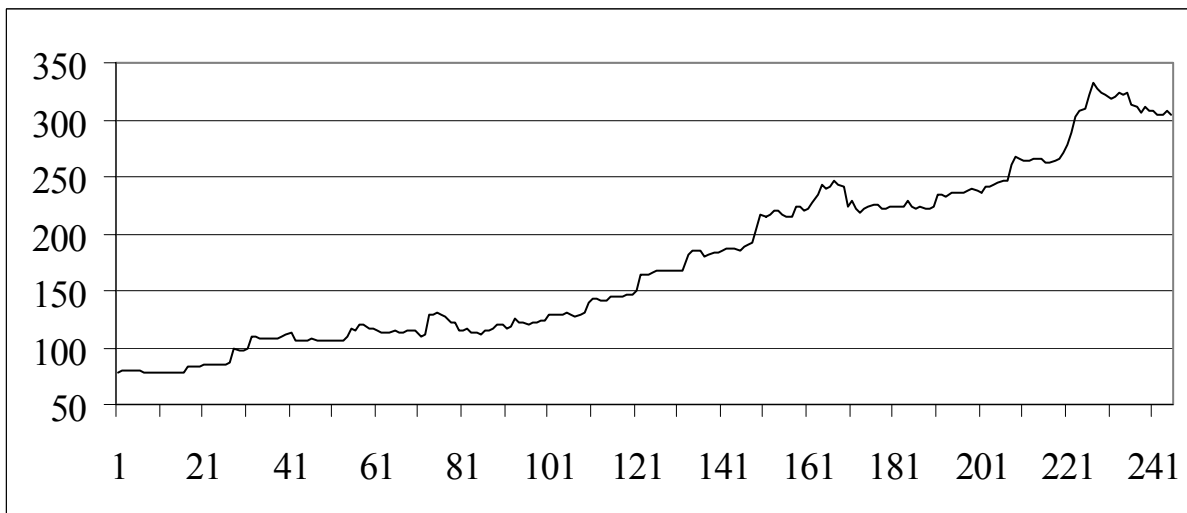


Рис. 3. Динаміка спостережень ціни акцій компанії "Мотор-Січ" за 2006-2007 роки

Користувачу надається можливість самостійно обрати модель для прогнозування процесу на основі графічного зображення результатів пробного прогнозування та порівняльної таблиці якості наведених моделей та отриманих на їх основі прогнозів.

Авторегресійні моделі для прогнозування процесів будуються на основі аналізу часткового коефіцієнта кореляції. Так, авторегресійна модель наведеного процесу має такий вигляд: $y_{n+1} = 0,7246 + 1,1260y_n - 0,0893y_{n-1} - 0,367y_{n-5} + \varepsilon$

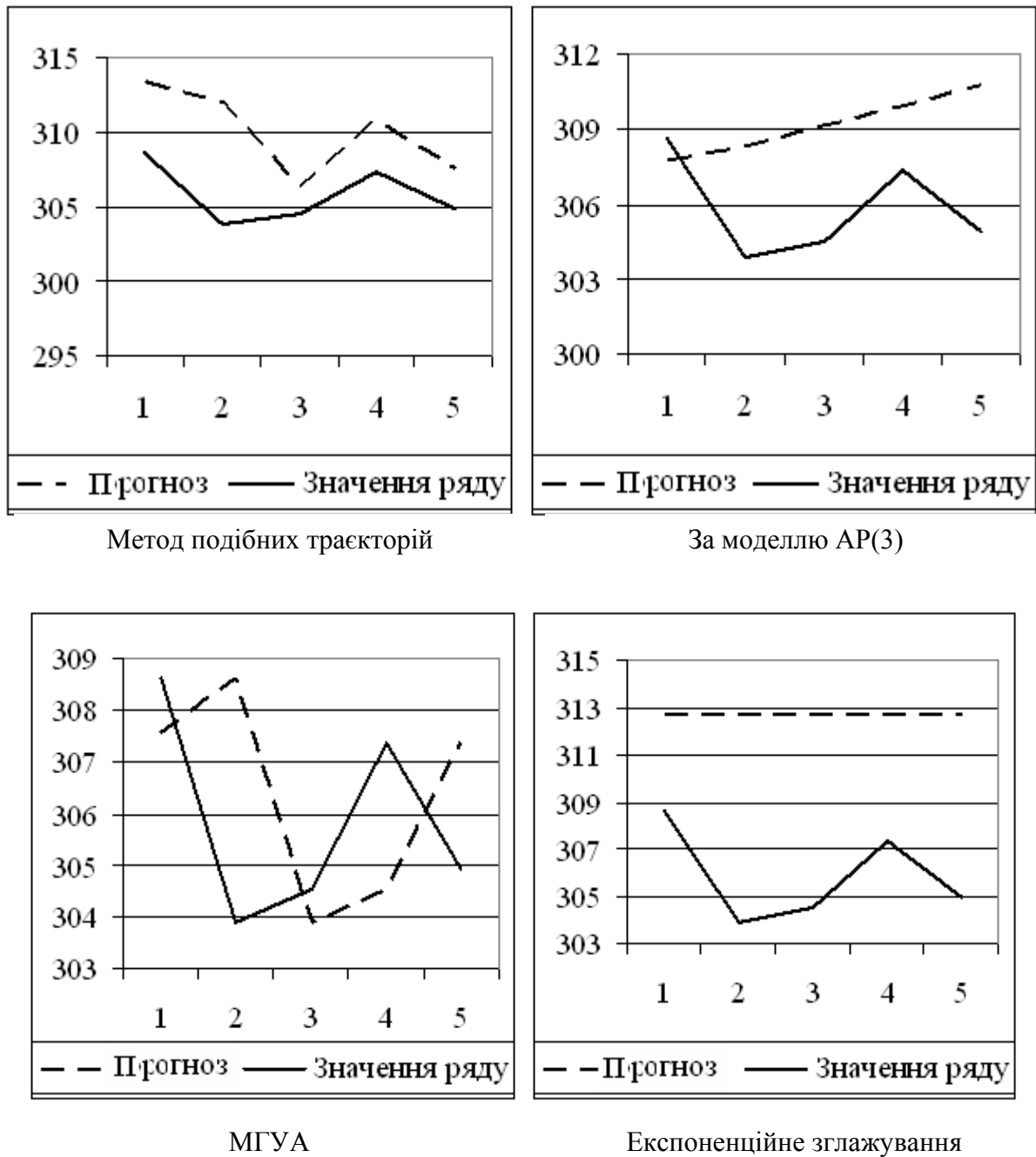


Рис. 4. Динаміка прогнозів ціни акцій компанії "Мотор-Січ" за 2006-2007 роки, отриманих різними методами

Таблиця 1

Типи моделей	Якість моделей			Якість прогнозу		
	R^2	$\sum \epsilon^2(k)$	DW	СКП	САПП	к-т Тейла
МГУА	0,9924	4187,52	1,6345	37,4634	0,761	0,011
Метод подібних траєкторій	–	–	–	110,459	–	–
АР(3)	0,9969	3921,39	1,9936	82,8984	1,2022	0,0597
Експоненційне згладжування	0,9496	12377,63	0,2200	247,5735	2,2274	0,01914

Приклад 2. Моделювання та прогнозування часового ряду як суми тренду та коливальної складової

Для роботи з рядами, що містять детермінований тренд, необхідно визначитись:

- чи можливі моделювання та прогнозування ряду безпосередньо за допомогою вихідної моделі ряду;
- чи потрібно попередньо описати та вилучити тренд і моделювати процес як дві окремі складові (тренд та коливальна складова).

Для прикладу розглянемо ряд спостережень ціни акцій компанії “Азовсталь” на фондовій біржі, динаміка якого наведена на рис 5.

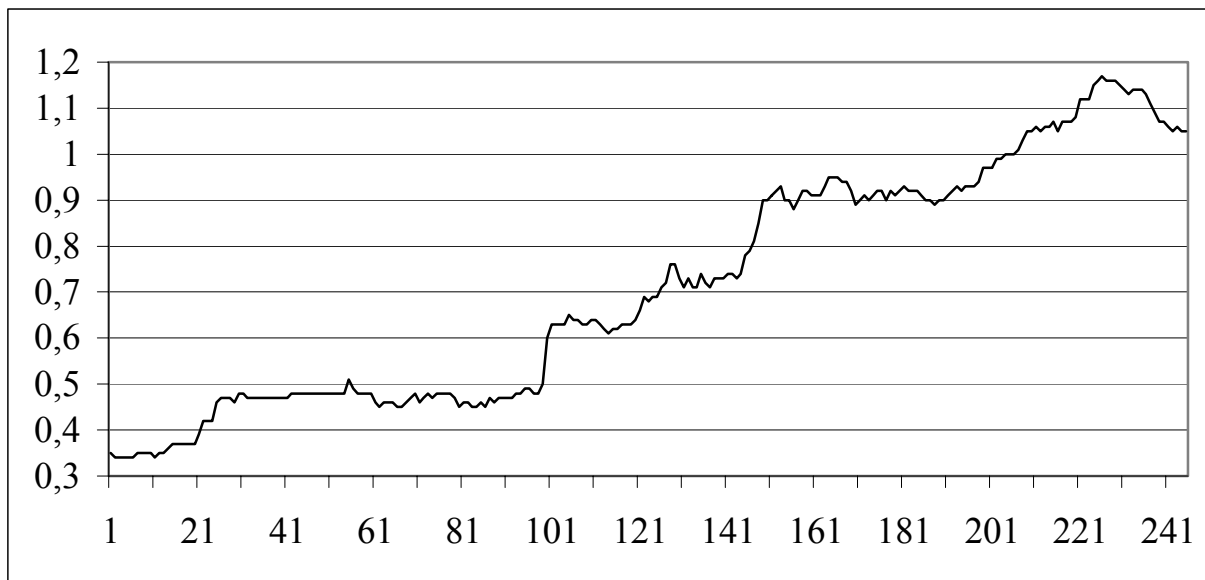


Рис. 5. Динаміка спостережень ціни акцій компанії “Азовсталь” за 2006-2007 роки

Авторегресійна модель наведеного процесу має такий вигляд: $y_{n+1} = 0,0043 + 1,1602y_n - 0,1627y_{n-1} + \epsilon$. Тренд даного процесу описується наступним поліномом шостого ступеня:

$$y_n = 0,2277 + 0,0206n - 0,0006n^2 + 9,29E - 06n^3 - 6,14E - 08n^4 + 1,94E - 10n^5 - 2,37E - 13n^6 .$$

Модель коливальної складової така :

$$x_{n+1} = -0,0004 + 1,1179x_n - 0,1973x_{n-1} + \epsilon .$$

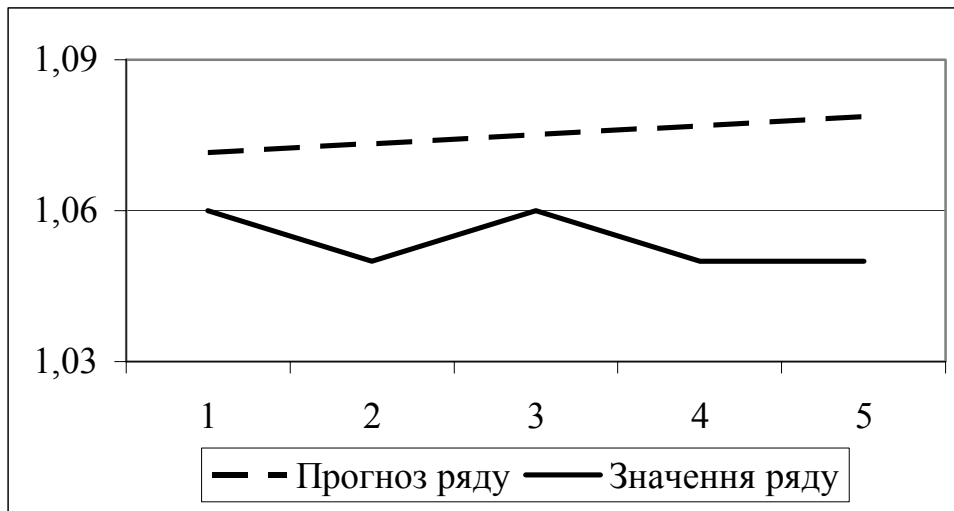


Рис. 6. Прогноз безпосередньо за моделлю AR(2)



Рис. 7. Прогноз тренду

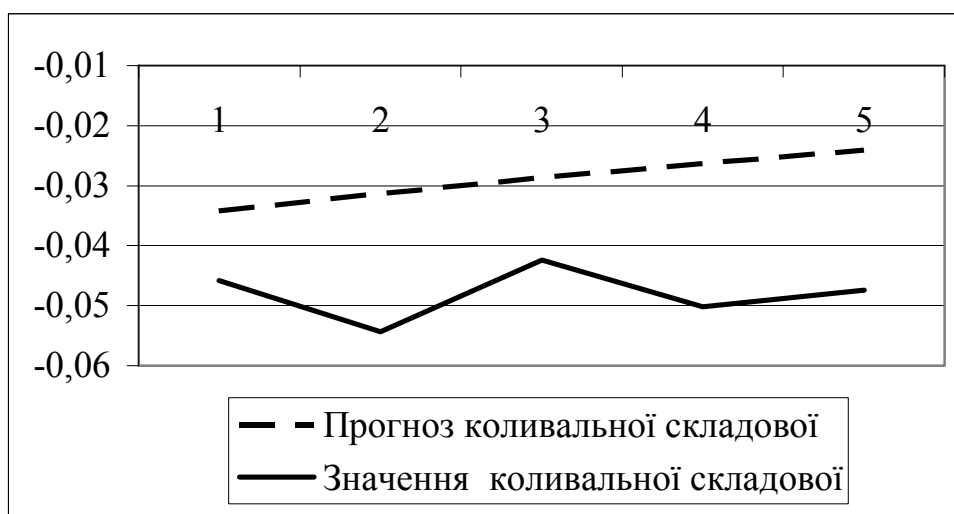


Рис. 8. Прогноз коливальної складової

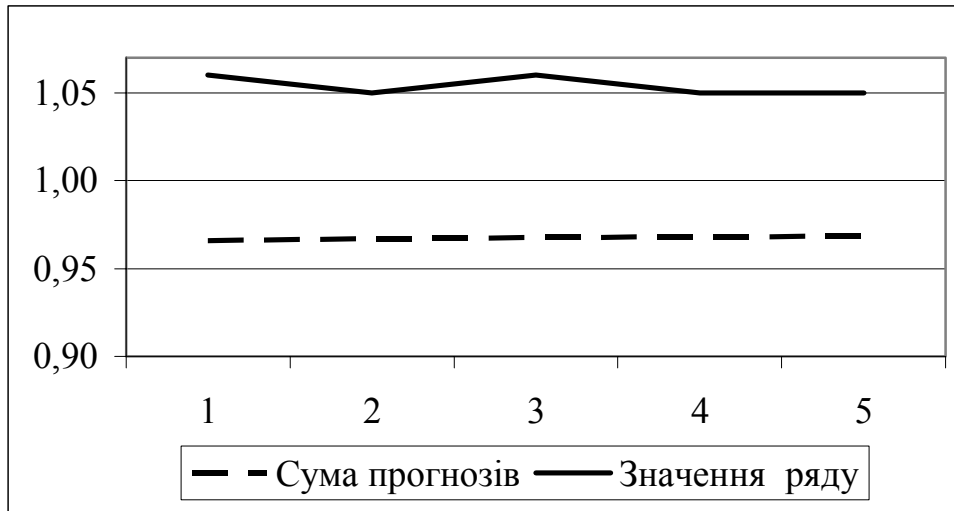


Рис. 9. Сумарний прогноз ряду

Таблиця 2

Типи моделей	Якість моделей			Якість прогнозу		
	R^2	$\sum e^2(k)$	DW	СКП	САПП	к-т Тейла
АР(2)	0,9969	0,0482	1,9948	0,0025	2,0073	0,0569
Тренд	0,9750	0,3979	0,1267	0,0816	11,5906	0,0267
Коливальна складова	0,8801	0,0458	2,0134	0,0020	39,3742	0,8347
Сума прогнозів (тренд+коливальна складова)				0,0376	8,2133	

Приклад 3. Моделювання дисперсії

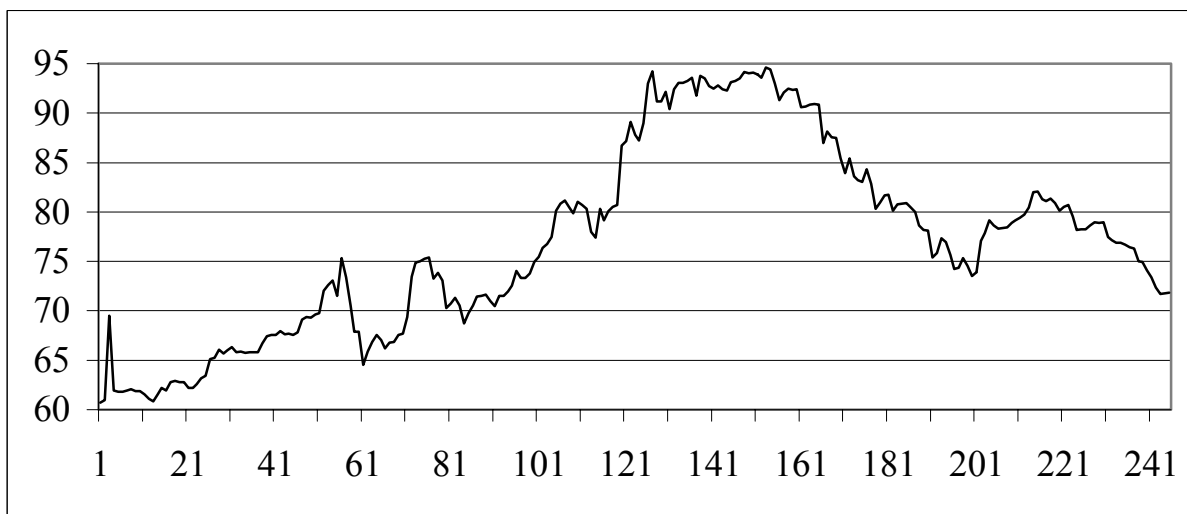


Рис. 10. Динаміка спостережень ціни акцій компанії “Укрнафта” за 2006-2007 роки

Після вилучення тренду як поліному:

$$y_n = 54,6321 + 1,2944n - 0,0485n^2 + 0,0008n^3 - 5,58E - 06n^4 + 1,83E - 08n^5 - 2,24E - 11n^6 .$$

Коливальна складова описана за допомогою моделі авторегресії, має такий вигляд: $e_{n+1} = -0,0304 - 0,0473e_n - 0,001e_{n-1} + \varepsilon$. Графіки квадратів залишків даної моделі та умовної дисперсії наведена на рис. 10, 11.

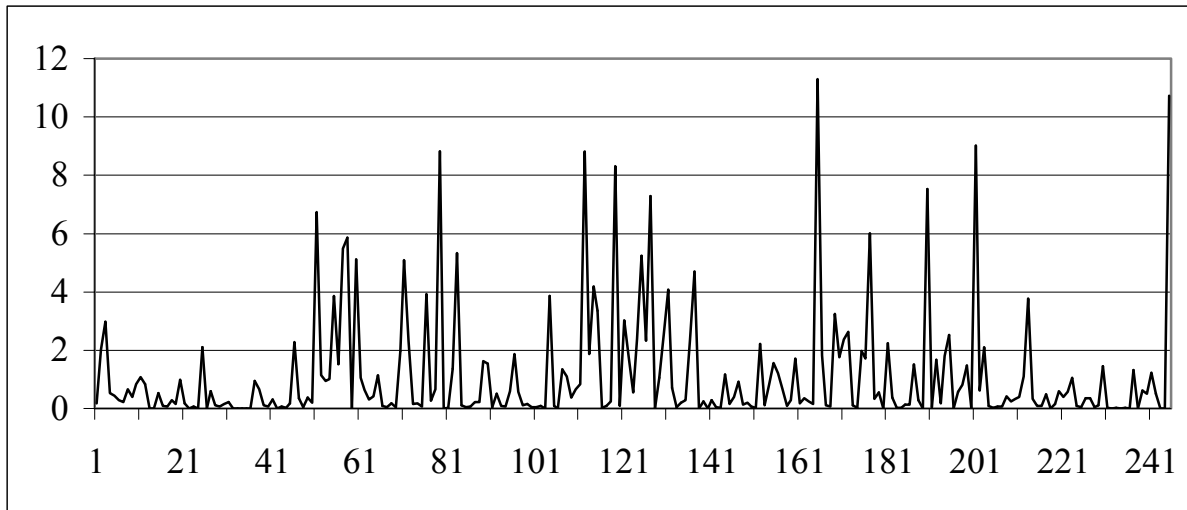


Рис. 10. Графік квадратів залишків

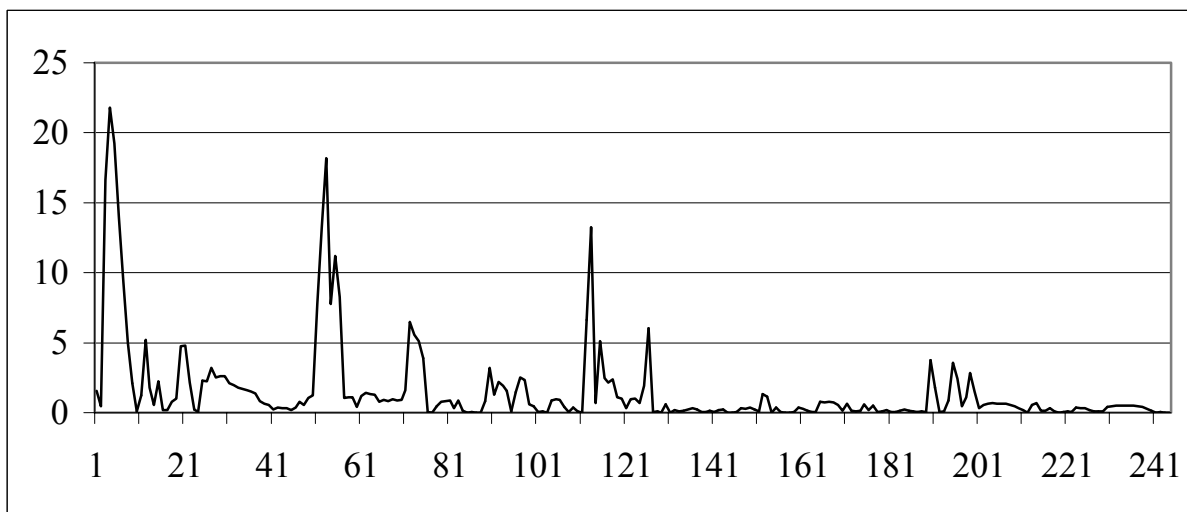


Рис. 11. Графік умовної дисперсії

Для прогнозування поведінки ряду залишків побудуємо такі моделі:

- АРУГ:
$$h_{n-1} = 0,0144 - 0,0013\varepsilon_n^2 + 0,0055\varepsilon_{n-1}^2 + 0,0075\varepsilon_{n-2}^2 + 0,005\varepsilon_{n-3}^2 + 0,0035\varepsilon_{n-4}^2 + 0,0017\varepsilon_{n-5}^2 + \varepsilon_1;$$

- УАРУГ:

$$h_{n-1} = 0,0269 - 0,0097\varepsilon_n^2 + 0,0032\varepsilon_{n-1}^2 - 0,0049\varepsilon_{n-2}^2 - 0,0007\varepsilon_{n-3}^2 + 0,0036\varepsilon_{n-4}^2 + 0,0014\varepsilon_{n-5}^2 + 0,1368h_n + 0,04h_{n-1} + 0,215h_{n-2} + 0,1095h_{n-3} + \varepsilon_1;$$

- ЕУАРУГ:

$$\log h_{n-1} = -3,7235 + 0,0042 \frac{|\varepsilon_n|}{\sqrt{h_n}} + 0,0018 \frac{|\varepsilon_{n-1}|}{\sqrt{h_{n-1}}} + 0,0032 \frac{|\varepsilon_{n-2}|}{\sqrt{h_{n-2}}} + 0,0031 \frac{|\varepsilon_{n-2}|}{\sqrt{h_{n-2}}} + 0,000793 \frac{\varepsilon_n}{\sqrt{h_n}} - 0,0029 \frac{\varepsilon_{n-1}}{\sqrt{h_{n-1}}} + 0,0002 \frac{\varepsilon_{n-2}}{\sqrt{h_{n-2}}} + 0,0025 \frac{\varepsilon_{n-3}}{\sqrt{h_{n-3}}} + 0,018 \log h_n + 0,0914 \log h_{n-1} + 0,1736h_{n-2} + 0,0633h_{n-3} + \varepsilon_1.$$

Для порівняння ефективності прогнозів, отриманих різними методами, наведемо графіки прогнозів рис. 12 та критерії якості як моделей, так і прогнозів табл. 3.

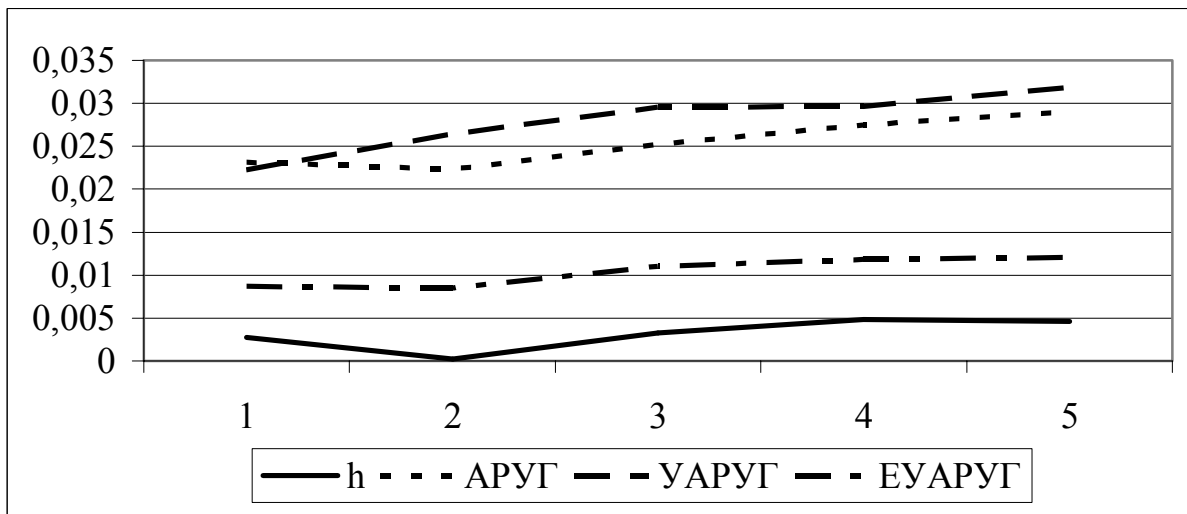


Рис. 12. Прогнози поведінки дисперсії, отримані за АРУГ, НАРУГ і ЕУАРУГ моделями, в порівнянні з точними значеннями

Таблиця 3

Типи моделей	Якість моделей			Якість прогнозу		
	R ²	∑e ² (k)	DW	СКП	САПП	к-т Тейла
АР(2)	0,0319	3,9702	1,8735	0,0005	2216,1531	0,7122
Тренд	0,0656	3,8317	1,9946	0,0006	2580,9828	0,7440
Коливальна складова	0,0526	835,752	1,9946	0,0006	2580,9828	0,7440
Сума прогнозів (тренд+коливальна складова)				5,382E-05	801,1343	0,9615

Після вилучення тренду, за допомогою різниць першого порядку, коливальна складова, описана моделлю авторегресії, має такий вигляд:

$$e_{n+1} = -0,0054 + 0,0004e_n + 0,0073e_{n-1} - 0,0633e_{n-5} + 0,0517e_{n-6} + 0,0823e_{n-8} + 0,0652e_{n-9} + \varepsilon.$$

Графіки квадратів залишків даної моделі та умовної дисперсії наведена на рис. 13, 14.

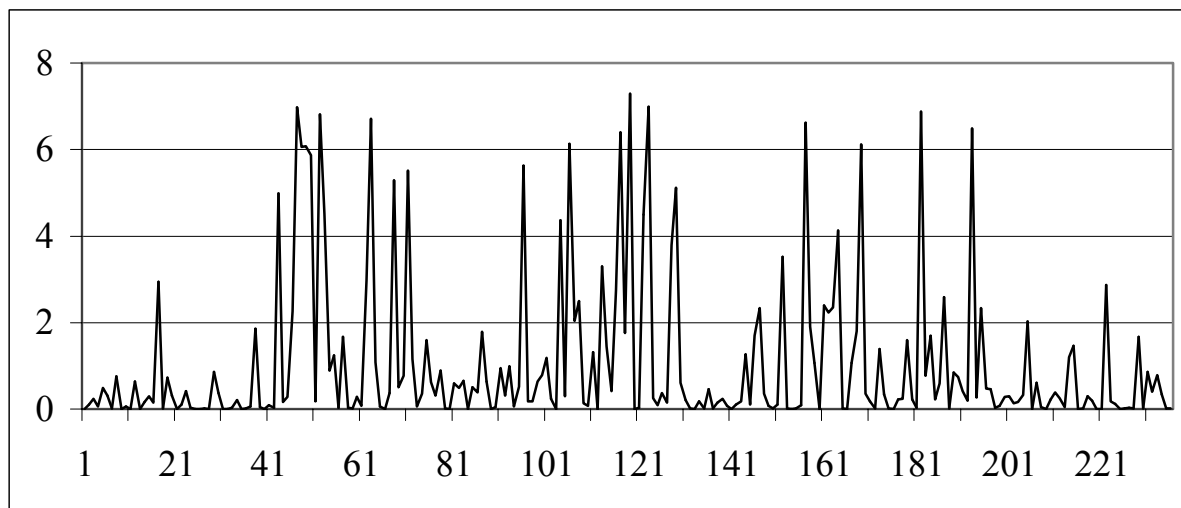


Рис. 13. Графік квадратів залишків

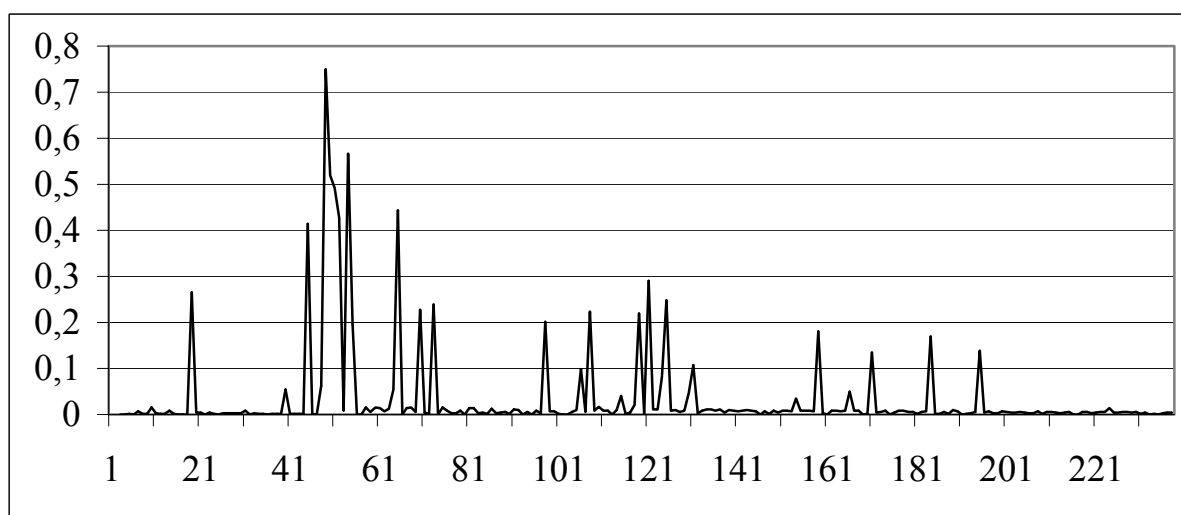


Рис. 14. Графік умовної дисперсії

Для прогнозування поведінки ряду залишків побудуємо такі моделі:

- АРУГ:

$$h_{n+1} = 0,0098 + 0,0089\varepsilon_n^2 + 0,0072\varepsilon_{n-1}^2 + 0,0057\varepsilon_{n-2}^2 + 0,0022\varepsilon_{n-3}^2 + 0,0059\varepsilon_{n-4}^2 + 0,0033\varepsilon_{n-5}^2 - 0,0065\varepsilon_{n-6}^2 - 0,0057\varepsilon_{n-7}^2 + \varepsilon_1;$$

- УАРУГ:

$$h_{n+1} = 0,0242 + 0,0002\varepsilon_n^2 - 0,0036\varepsilon_{n-1}^2 - 0,0007\varepsilon_{n-2}^2 - 0,0059\varepsilon_{n-3}^2 - 0,001\varepsilon_{n-4}^2 + 0,0031\varepsilon_{n-5}^2 - 0,0062\varepsilon_{n-6}^2 + 0,0003\varepsilon_{n-7}^2 + 0,1814h_n + 0,2243h_{n-1} + 0,1197h_{n-2} + 0,1769h_{n-3} + 0,1475h_{n-4} - 0,0006h_{n-6} - 0,1219h_{n-7} + \varepsilon_1;$$

- ЕУАРУГ:

$$\begin{aligned} \log h_{n+1} = & -3,4354 + 0,0001 \frac{|\varepsilon_n|}{\sqrt{h_n}} + 0,002 \frac{|\varepsilon_{n-1}|}{\sqrt{h_{n-1}}} - 0,0007 \frac{|\varepsilon_{n-5}|}{\sqrt{h_{n-5}}} + 0,0002 \frac{|\varepsilon_{n-6}|}{\sqrt{h_{n-6}}} + \\ & + 0,0017 \frac{|\varepsilon_{n-8}|}{\sqrt{h_{n-8}}} - 0,0011 \frac{|\varepsilon_{n-9}|}{\sqrt{h_{n-9}}} + 0,002 \frac{\varepsilon_n}{\sqrt{h_n}} - 0,0006 \frac{\varepsilon_{n-1}}{\sqrt{h_{n-1}}} - \\ & - 0,0019 \frac{\varepsilon_{n-5}}{\sqrt{h_{n-5}}} + 0,0012 \frac{\varepsilon_{n-6}}{\sqrt{h_{n-6}}} - 0,0026 \frac{\varepsilon_{n-8}}{\sqrt{h_{n-8}}} + 0,0009 \frac{\varepsilon_{n-9}}{\sqrt{h_{n-9}}} + \\ & + 0,1095 \log h_n + 0,0964 \log h_{n-1} + 0,0572 h_{n-2} + 0,0123 h_{n-3} + \\ & + 0,0572 h_{n-4} - 0,0113 h_{n-6} + 0,0377 h_{n-7} + \varepsilon_1. \end{aligned}$$

Для прогнозування умовної дисперсії також застосуємо МГВА. Порівняння ефективності прогнозів, отриманих різними методами, наведено на рис. 15, а критерії якості моделей і прогнозів наведено в табл. 4.

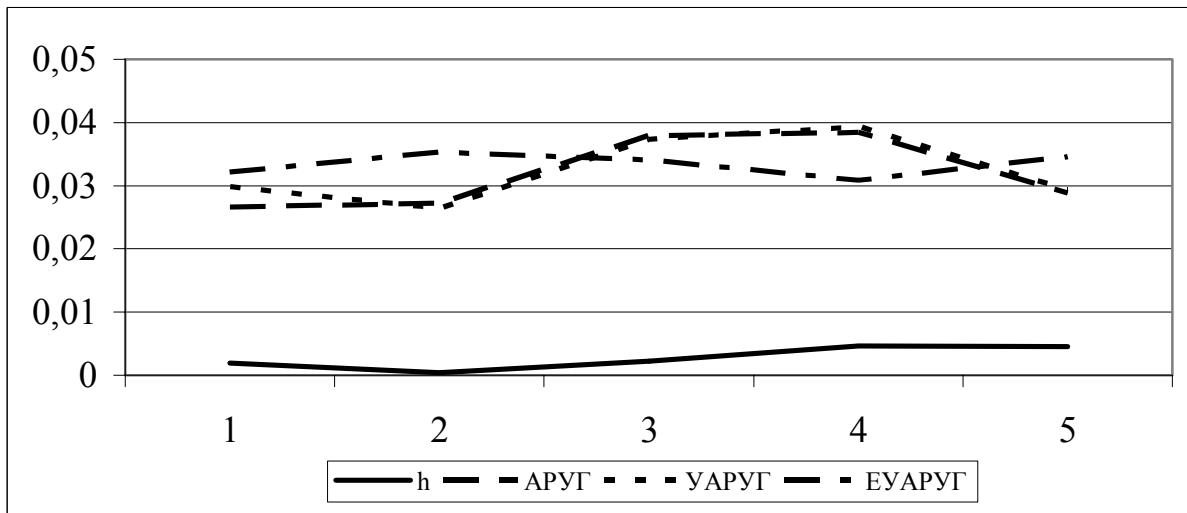


Рис. 15. Прогнози поведінки дисперсії, отримані за АРУГ, УАРУГ і ЕУАРУГ моделями, в порівнянні з точними значеннями

Таблиця 4

Типи моделей	Якість моделей			Якість прогнозу		
	R ²	∑e ² (k)	DW	СКП	САПП	к-т Гейла
АРУГ	0,0319	3,9702	1,8735	0,0005	2216,1531	0,7122
УАРУГ	0,0656	3,8317	1,9946	0,0006	2580,9828	0,7440
ЕУАРУГ	0,0526	835,752	1,9946	5,382E-05	801,1343	0,9615

Отже, виходячи з результатів наведених таблиць, використання запропонованої СППР дає можливість отримати високоякісний прогноз ціни акцій компанії Укрнафта, про що свідчать статистичні показники якості прогнозу.

Висновки

Таким чином, в роботі запропоновано комп'ютерну систему обробки даних, побудови математичних моделей та прогнозування розвитку нестационарних процесів. Вона дозволяє користувачам з різним рівнем підготовки отримувати надійні та обґрунтовані моделі розвитку нестационарних процесів та короткострокові прогнози їх

поведінки. Система має гнучку архітектуру, що передбачає розширення числа функцій, введення нових алгоритмічних модулів для оцінювання моделей та прогнозування. СППР дозволяє ідентифікувати та прогнозувати нестационарні процеси з детермінованими та стохастичними трендами, а також гетероскедастичні процеси без використання експертних оцінок.

Побудовано дерево прийняття рішень при аналізі часових рядів, використання якого дозволяє дослідити властивості даних, визначити наявність нестационарностей та коінтегрованості процесів. Застосування дерева прийняття рішень дає можливість структурувати проблему прийняття рішень при моделюванні та прогнозуванні нестационарних процесів.

У подальших дослідженнях планується розширення СППР новими тестами на наявність нелінійностей, моделями гетероскедастичних процесів, а також автоматизація вибору величини запізнення і побудови функцій прогнозування.

ЛІТЕРАТУРА

1. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах // Открытые системы. – 1997. – № 4. – С. 41-44.
2. Tony Murphy, Achieving Business Value from Technology. A Practical Guide for Today's Investment. Gartner Group, John Wiley & sons, 2002.
3. Олексюк О.С. Системи підтримки прийняття фінансових рішень на мікрорівні. – К.: Наукова думка, 1998. – 507 с.
4. Бідюк П.І., Половцев О.В. Аналіз та математичне моделювання економічних процесів перехідного періоду. – К.: ПЛАБ-75, 1999. – 209 с.
5. Скрипкин К.Г. Экономическая эффективность информационных систем. – М.: ДМК пресс, 2002.
6. Зайченко Ю.П. Основы проектирования интеллектуальных систем. – К.: Слово. – 2004. – 352 с.
7. Сахаров А.А. Концепции построения и реализации информационных систем, ориентированных на анализ данных // Системы управления базами данных. – 1996. – № 4. – С. 55-70.
8. Бідюк П. І., Савенков О.І., Баклан І.В. Часові ряди: моделювання та прогнозування. – К.: ЕКМО, 2004. – 144 с.
9. Бідюк П.І., Щербань В.Ю., Демківський Є.О. Проектування адаптивного інтерфейсу користувача систем підтримки прийняття рішень // Вісник КНУТД. – 2004. – № 6. – С. 37-43.
10. Бідюк П.І., Щербань В.Ю., Демківський Є.О. Проектування природно-мовного інтерфейсу користувача систем підтримки прийняття рішень // Вісник КНУТД. – 2005. – № 2. – С. 38-43.
11. Бідюк П.І., Щербань Ю.Ю., Щербань В.Ю., Демківський Є.О. Системи підтримки прийняття рішень – проектування та реалізація. – К.: КНУТД, 2004. – 112 с.
12. Демківський Є.О. Методи моделювання детермінованих і стохастичних трендів часових рядів. // Праці II міжнародної школи-семінару “Теорія прийняття рішень”. – Ужгород, 2004. – С. 31.
13. Лук'янченко І., Краснікова Л. Економетрика: Практикум з використанням комп'ютера. – К.: Знання, 1998. – 217 с.