Міністерство освіти і науки України

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ПРОГНОЗУВАННЯ ФІНАНСОВИХ ПРОЦЕСІВ

Практикум

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня магістра
за освітніми програмами «Системний аналіз фінансового ринку» та «Системний аналіз і управління»
спеціальності 124 Системний аналіз

Укладачі: Н.В. Кузнєцова, П.І. Бідюк

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2023

УДК 004.89:33.021](076)
П89

Укладачі: *Кузнєцова Наталія Володимирівна*, доктор технічних наук, доцент
Бідюк Петро Іванович, доктор технічних наук, професор

Рецензент *Положаєнко С. А.*, доктор технічних наук, професор,
Одеський національний політехнічний університет

Відповідальний редактор *Данилов В.Я.*, доктор технічних наук, професор

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 3 від 07.12.2023 р.)
за поданням Вченої ради навчально-наукового інституту прикладного системного аналізу
(протокол № 9 від 30.10.2023 р.)

П89 **Прогнозування фінансових процесів** [Електронний ресурс]: **практикум** : навч. посіб. для здобувачів ступеня магістра за освіт. програмами «Системний аналіз фінансового ринку» і «Системний аналіз і управління» спеціальності 124 Системний аналіз / КПІ ім. Ігоря Сікорського ; уклад.: Н.В. Кузнєцова, П.І. Бідюк. – Електрон. текст. дані (1 файл: 3,91 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2023. – 106 с.

Посібник містить задачі та рекомендаційні вказівки для студентів стосовно виконання практичних завдань з навчальної дисципліни «Сучасні методи прогнозування». Наведені ілюстративні приклади розв'язання практичних задач прогнозування для сучасних фінансових процесів з різних економічних галузей. Показано практично всі аспекти аналізу і прогнозування фінансових процесів із застосуванням не тільки класичних підходів і методів, а й сучасного інструментарію (Facebook Prophet, Python, R). Навчальний посібник розроблений і призначений для здобувачів ступеня магістра за освітніми програмами «Системний аналіз фінансового ринку» та «Системний аналіз і управління» спеціальності 124 Системний аналіз, проте буде також корисним для студентів, аспірантів та викладачів, які спеціалізуються в галузі розв'язання задач статистичного аналізу даних, математичного моделювання і прогнозування динаміки фінансово-економічних та процесів іншої природи, представлених статистичними даними.

УДК 004.89:33.021](076)

Реєстр. № НП 23/24-155. Обсяг 3,25 авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Берестейський, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

ЗМІСТ

ВСТУП

Навчальний посібник призначений для допомоги у виконанні практичних і самостійних робіт з дисципліни «Сучасні методи прогнозування» орієнтований на студентів, які навчаються за спеціальністю 124 Системний аналіз і здобувають ступінь магістра за освітніми програмами «Системний аналіз фінансового ринку» та «Системний аналіз і управління». Посібник-практикум містить короткі теоретичні інформаційні матеріали, а також практичні застосунки, що пропонуються студентам у вибірковій дисципліні «Сучасні методи прогнозування», яка є логічним продовженням дисциплін «Аналіз часових рядів» та «Аналіз фінансово-економічних даних». Основним завданням вибіркової дисципліни, окрім формування загальних компетентностей, таких як здатність до абстрактного мислення, аналізу та синтезу (ЗК 1), здатність до пошуку, оброблення та аналізу інформації з різних джерел (ЗК4) і застосування їх у практичних ситуаціях, є формування фахових компетентностей. Серед таких: здатність інтегрувати знання та здійснювати системні дослідження, застосовувати методи математичного та інформаційного моделювання складних систем та процесів різної природи (ФК1), здатність моделювати, прогнозувати та проєктувати складні системи і процеси на основі методів та інструментальних засобів системного аналізу (ФК5), здатність застосовувати теорію і методи Data Science для здійснення інтелектуального аналізу даних з метою виявлення нових властивостей та генерації нових знань про складні системи (ФК6). Студенти після вивчення дисципліни отримують і додаткові фахові компетенції, які орієнтовані саме на застосуванні їх знань, вмінь і навичок за фахом у практичному середовищі на робочих місцях. Зокрема, це здатність формувати та перевіряти сформовані гіпотези щодо розподілу фінансово-економічних даних, розробляти власні прогнозні моделі

та оцінювати ймовірність появи ризиків та обсяги можливих втрат, застосовувати різні стратегії мінімізації та нейтралізації ризиків, здатність застосовувати методи індуктивного моделювання та математичний апарат нечіткої логіки, нейронних мереж, теорії ігор та обчислювального інтелекту в задачах системного аналізу фінансового ринку, а також здатність розробляти алгоритми прогнозування умовних дисперсій гетероскедастичних процесів в складних системах різної природи. Програмними результатами вивчення дисципліни є: вміння і навички будувати та досліджувати моделі складних систем і процесів, застосовуючи методи системного аналізу, математичного, комп'ютерного та інформаційного моделювання (ПР2), розробляти та застосовувати методи, алгоритми та інструменти прогнозування розвитку складних систем і процесів різної природи (ПР4), застосовувати методи машинного навчання та інтелектуального аналізу даних, математичний апарат нечіткої логіки, теорії ігор та розподіленого штучного інтелекту для розв'язання складних задач системного аналізу (ПР6), розробляти та застосовувати моделі, методи та алгоритми прийняття рішень в умовах конфлікту, нечіткої інформації, невизначеності та ризиків (ПР9). Додатковими програмними результатами навчання, визначеними за освітньою програмою і отриманими після вивчення дисципліни, є: знання методів системного підходу до аналізу ризиків, VAR-технології, IRB-підходу до аналізу та оцінювання ризиків; уміння застосовувати сучасні засоби інтелектуального аналізу даних при розв'язанні реальних фінансово-економічних задач, уміння створювати математичні моделі складних систем та проектувати алгоритми підтримки прийняття рішень в умовах проектування систем штучного інтелекту за допомогою методів індуктивного моделювання, нечіткої логіки, нейронних мереж, теорії ігор, генетичних методів оптимізації, еволюційного

моделювання, знання методів прогнозування умовних дисперсій гетероскедастичних процесів.

Метою викладання навчальної дисципліни «Сучасні методи прогнозування» є формування у студентів поглибленої системи теоретичних знань і практичних навичок з моделювання і прогнозування сучасних процесів різної природи. Для цього навчальна дисципліна формує у студентів уміння збирати та коректно обробляти статистичні дані; будувати математичні і статистичні моделі різних видів; формувати та перевіряти сформовані гіпотези щодо розподілу даних; оцінювати параметри математичних моделей; застосовувати сучасні методи прогнозування та інформаційні технології ІАД до прогнозування реальних подій і процесів; розуміти, моделювати та розв’язувати практичні задачі прогнозування фінансово-економічного характеру.

У посібнику наведені основні етапи вирішення практичних задач та варіанти і можливості для студентів отримання реальних даних та постановки власних задач. Для практичного завдання студентам пропонується обрати дані, що представлені часовими рядами, зокрема фінансовими, надаються джерела та агрегатори таких даних. Також запропонована коротка покрокова методика побудови моделей на основі сучасних методів прогнозування, студентам пропонується порівняти якість побудованих моделей на основі сучасних методів прогнозування (наприклад, з використанням Facebook Prophet) і порівняти якість самого прогнозування з використанням більш складних класичних моделей та моделей для прогнозування волатильності. Наведено основні бібліотеки і особливості реалізації і дослідження трендів на фінансових ринках для оцінювання і моделювання фінансових процесів.

1. Регресійні моделі та різницеві рівняння для прогнозування процесів різної природи

Нагадаємо основні типи регресійних моделей, їх математичне представлення і особливості побудови для подальшого прикладного застосування до фінансових процесів.

Авторегресія: рівняння авторегресії описує пам'ять процесу, тобто воно відображає вплив значень попередніх станів на його поточний стан [1]:

$$y(k) = a_0 + a_1 y(k-1) + ... + a_p y(k-p) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + \varepsilon(k),$$

де a_i, $i = 1,..., p$ – коефіцієнти моделі, які оцінюють на основі значень часового ряду; p – порядок авторегресії, який визначається числом затриманих в часі значень ряду, що використовуються у правій частині рівняння для описання динаміки змінної в момент k; $k = 1, 2,....$ дискретний час; $\varepsilon(k)$ – випадкова величина, поява якої в моделі зумовлена такими причинами:

- вплив випадкових збурень на процес, що моделюється;
- похибки рівняння, зумовлені неточно вибраною структурою (можливо, що не враховано деякі регресори, введено непотрібні незалежні змінні або робиться спроба моделювати складний нелінійний процес за допомогою лінійного рівняння);
- методичні й обчислювальні похибки, які з'являються при обчисленні оцінок параметрів (коефіцієнтів) рівняння.

Парна регресія – вона включає у правій частині незалежну змінну (регресор):

$$y(k) = a_0 + a_1 x(k) + \varepsilon(k),$$

де $x(k)$ – *регресор* (*незалежна* або *екзогенна* змінна, або *предиктор*), тобто $x(k)$ має чотири назви. Залежну змінну $y(k)$ називають ще *головною* або *ендогенною* змінною.

Множинна регресія – множинна регресія відображає вплив декількох незалежних змінних на залежну:

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + ... + a_p x_p(k) + \varepsilon(k),$$

де $x_1(k),..., x_p(k)$ – регресори рівняння. Таке рівняння може включати також авторегресійну частину [2].

Авторегресія + множинна регресія = змішана регресія.

Модель у вигляді авторегресії + множинна регресія отримала назву розширеної авторегресії (autoregressive extended (ARX) – у англомовній термінології):

$$y(k) = a_0 + \sum_{i=1}^{l} a_i y(k-i) + b_1 x_1(k) + b_2 x_2(k) + ... + b_p x_p(k) + \varepsilon(k).$$

При цьому досить мати один регресор у правій частині.

Авторегресія з ковзним середнім порядку (p,q) **(**АРКС**(** p,q **)):**

$$y(k) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j \varepsilon(k-j) + \varepsilon(k).$$

Розширена авторегресія з ковзним середнім – модель у вигляді авторегресії з ковзним середнім та одним або більше регресорами (autoregressive moving average extended (ARMAX) – у англомовній термінології):

$$y(k) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j \varepsilon(k-j) + c_1 x(k) + \varepsilon(k).$$

Регресія, нелінійна відносно змінних **(***псевдолінійна регресія***)**:

$$y(k) = a_0 + a_1 x(k) + a_2 x^2(k) + ... + a_p x^p(k) + \varepsilon(k).$$

Тобто в даному випадку це поліноміальна регресія порядку p. Коефіцієнти псевдолінійної регресії оцінюють такими ж методами, що і лінійної, наприклад за методом найменших квадратів (МНК) або максимальної правдоподібності (ММП).

Регресія, нелінійна стосовно параметрів. Моделі, нелінійні стосовно параметрів, містять адитивні члени, які включають в себе добутки параметрів моделей або інші види зв'язку (крім адитивного) між параметрами:

$$y(k) = a_0 + a_1 e^{bx(k)} + \varepsilon(k).$$

Для оцінювання параметрів таких моделей необхідно застосовувати методи нелінійного оцінювання – нелінійний метод найменших квадратів, метод максимальної правдоподібності, Монте- Карло та інші.

Моделі гетероскедастичних процесів [3-6] – моделі процесів, дисперсія яких змінюється в часі. Рівняння для умовної дисперсії (авторегресія першого порядку):

$$h(k) = \beta_0 + \varepsilon^2(k-1) + \varepsilon_1(k),$$

де $h(k)-$ умовна дисперсія процесу в момент k; $\varepsilon^2(k)-$ квадрат залишків; $\varepsilon_1(k)-$ похибка цієї моделі в момент k. Докладно моделі гетероскедастичних процесів будуть розглянуті нижче.

Авторегресійна умовно гетероскедастична модель порядку p (АРУГ(p)):

$$h(k) = \beta_0 + \sum_{i=1}^{p} \beta_i \varepsilon^2(k-i) + \varepsilon_1(k).$$

Узагальнена авторегресійна умовно гетероскедастична модель (УАРУГ(p,q)):

$$h(k) = \beta_0 + \sum_{i=1}^{p} \beta_i \, \varepsilon^2(k-i) + \sum_{i=1}^{q} \alpha_i h(k-i) + \varepsilon_1(k),$$

де $\alpha, \beta, \gamma \geq 0$ (для того щоб уникнути появи від'ємних значень умовних дисперсій).

Експоненціальна модель УАРУГ (умовна дисперсія як асиметрична функція ε, тобто моделювання впливу попередніх значень $\varepsilon(k-i)$ на волатильність) [7]:

$$\log(h(k)) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \frac{|\varepsilon(k-i)|}{h(k-i)} + \sum_{i=1}^{p} \gamma_i \frac{\varepsilon(k-i)}{h(k-i)} + \sum_{i=1}^{q} \beta_j \log(h(k-i)).$$

Модель УАРУГ-М (модифікована) – моделювання премії за ризик:

$$y(k) = \beta + \gamma h(k) + \varepsilon(k),$$

$$h(k) = a_0 + a \sum_{i=1}^{p} \varepsilon^2(k-i) + \sum_{i=1}^{q} h(k) + \varepsilon_2(k).$$

Модель для прогнозування волатильності за допомогою УАРУГ:

$$h(k+1) = \beta_0 + \beta_1 \varepsilon^2(k) + \gamma h(k),$$

$$h(k+j) = \beta_0 + (\beta_1 + \gamma_1) h(k+j-1).$$

Рівняння для умовної дисперсії та коваріації:

$$h(s(k)) = \alpha_0 + \alpha_1 \varepsilon^2(s(k-1)) + \alpha_2 h(s(k-1)),$$

$$h(f(k)) = \beta_0 + \beta_1 \varepsilon^2(f(k-1)) + \beta_2 h(f(k-1)),$$

$$\operatorname{cov}[s(k), f(k)] = \gamma_0 + \gamma_1 \varepsilon(s(k-1)) + \gamma_2 \operatorname{cov}[s(k-1), f(k-1)].$$

Коефіцієнт хеджування при використанні двофакторної моделі:

$$\hat{b}^*(k) = \frac{\operatorname{cov}[s(k), f(k)]}{\hat{h}^2(f)}.$$

Моделі коінтегрованих процесів, тобто нестаціонарних процесів, які можна об'єднати в межах однієї моделі, що буде мати стаціонарні статистичні характеристики.

Концепція коінтегрованості змінних передбачає існування довгострокового зв'язку між значеннями змінних. Тобто ми припускаємо існування спільної врівноваженої траєкторії руху цих змінних, від якої вони можуть відхилятися на коротких проміжках часу. Однак економічні механізми діють таким чином, що рівновага відновлюється і зберігається на довгих часових інтервалах шляхом корегування відповідних відхилень від врівноваженого стану.

У випадку коінтегрованості змінних $x(k)$ і $y(k)$ може бути побудована модель корегування похибки, яка поєднує динаміку змінних на коротких проміжках часу з довгостроковим врівноваженим зв'язком і має такий вигляд:

$$\Delta x(k) = a_{10} + \sum_{i=1}^{p} b_{1i}\,\Delta x(k-i) + \sum_{i=1}^{p} c_{1i}\,\Delta y(k-i) + \lambda_1 e(k-1) + \varepsilon_1(k),$$

$$\Delta y(k) = a_{20} + \sum_{i=1}^{p} b_{2i}\,\Delta y(k-i) + \sum_{i=1}^{p} c_{2i}\,\Delta x(k-i) + \lambda_2 e_x(k-1) + \varepsilon_2(k).$$

Коефіцієнти λ_1, λ_2 у наведених рівняннях називають швидкістю пристосування (корегування). Моделями такого типу можна скористатись для побудови систем керування фінансово-економічними процесами, оскільки вони поєднують у собі вхідні та вихідні змінні процесів, які аналізуються. А наявність коінтеграції – це також позитивний факт стосовно створення систем керування, тому що вона свідчить про існування суттєвого впливу одного процесу на інший.

Вище ми навели тільки деякі структури (типи) математичних моделей стаціонарних та нестаціонарних процесів, які використовуються для опису динаміки фінансово-економічних та процесів іншої природи.

Запитання для самоконтролю

1. Які регресійні моделі варто застосовувати для моделювання фінансових процесів?
2. Яка методика побудови і оцінювання параметрів регресійних моделей?
3. Як побудувати гетероскедастичні моделі?
4. Як оцінити волатильність фінансових процесів?
5. Як перевірити наявність тренду у ряді?
6. За допомогою яких відомих моделей можна описати детермінований тренд?
7. Що таке сезонні ефекти і в чому причина їх виникнення?
8. Що таке коротко- і середньострокове прогнозування?

2. Послідовність виконання моделювання і прогнозування процесів на фінансовому ринку регресійними моделями

В цьому розділі наведено покрокову інструкцію виконання практичних завдань моделювання фінансових процесів за допомогою різних видів регресійних моделей та використання сучасних моделей для прогнозування дисперсії (волатильності) процесів.

Метою запропонованих практичних робіт є використання сучасних методів прогнозування для аналізу і моделювання реальних фінансових процесів з високою волатильністю, які потребують урахування їх гетероскедастичності та сезонності.

Дані для роботи: вхідними даними для практичного завдання є ціни на акції провідних світових компаній, отримані з Нью-Йоркської, Австралійської, Японської фондової біржі (NYSE, NASDAQ, D-J, ASX), котирування акцій, співвідношення цін на валютні пари, криптовалюти, індекси світової економіки, тощо.

Сам етап збору та попередньої обробки даних є одним з найбільш складних і затратних у часі для аналітика в компанії. Найпоширеніше питання, яке виникає перед студентами під час виконання практичних, курсових та дисертаційних робіт: **а де ж брати дані?** У цьому розділі ми наведемо сучасні джерела даних, які можуть бути запропоновані і використані студентами.

1) *Змагання на платформі Kaggle* [8]. https://www.kaggle.com/datasets
Тут викладаються реальні дані компаній і здійснюється постановка задачі моделювання. Команди вирішують задачі та завантажують результати власного моделювання. Команда, яка знайшла вирішення задачі найкращим способом, отримує приз. Переваги: найсучасніші, найактуальніші дані, є можливість

порівняти результати власного моделювання з найкращими результатами команд-учасників змагань.

Недоліки: незрозумілі назви змінних, частково інформація є захешованою, тому інколи складно зрозуміти причинно-наслідкові зв'язки між змінними.

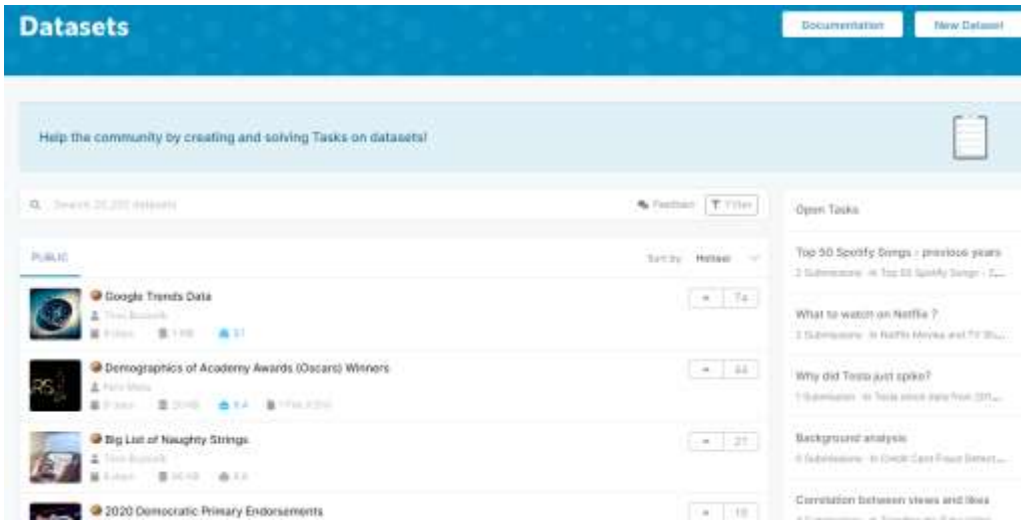


Рис. 1. Скріншот з видом платформи з даними Kaggle

2) *На фінансових порталах Yahoo Finance* [9], *Finam*

https://finance.yahoo.com/

Фінансовий портал для тих, хто працює з фінансовими інструментами на фондових ринках, і для тих, хто лише хоче почати займатись трейдингом. На сайті є можливості експорту цін на акції за вказаний період з погодинною, щоденною деталізацією. Наприклад, ціни на акції S&P, Nasdaq, Dow30, Russell 2000, Crude Oil, Gold, криптовалюти, ціни на державні облігації США, тощо.

https://finance.yahoo.com/quote/%5EGSPC?p=%5EGSPC

Будьте уважні з деталізацією періоду завантаження даних! Оскільки для погодинної деталізації при генеруванні файлу може виникнути помилка через

великий обсяг даних. Рекомендація: вивантажувати дані за невеликі періоди окремими файлами (листами в Excel).

3) Використання спеціальних статистичних пакетів для отримання даних з сайтів: *https://finance.yahoo.com/* Yahoo Finance; Google Finance; Federal Reserve Bank of St. Louis FRED®; Oanda (FOREX); з локальних БД MySQL, двійкових файлів R (.RData и .rda); текстових файлів CSV.

4) Власні дані: продажі, співвідношення цін, економічні дані, ВВП.

5) Отримати дані у викладача, скориставшись варіантами, наведеними у таблиці 15 розділу 8.

Задача практичного завдання полягає у розробці адекватної моделі прогнозування, яка відображає тенденцію зміни цін у наступні періоди і дозволяє аналітику/трейдеру приймати рішення щодо купівлі/продажу у наступні періоди. Прогнозування цін слід здійснювати на 1-5 кроків вперед.

Порядок виконання роботи:

1. Представити дані у вигляді гістограм, графіків, тощо. На основі такого попереднього аналізу є можливість оцінити ряд на наявність сезонних ефектів, викидів, пропусків, тощо і використати дану інформацію для подальшого моделювання.

2. Визначити часові періоди для побудови моделі (розбиття на навчальну та перевірочну вибірку) та вибрати кількість кроків, на скільки здійснюється прогнозування (для оцінювання якості прогнозу).

3. Використати тести на перевірку рядів на стаціонарність (Дікі-Фулера, KPSS, тощо). В разі наявності нестаціонарності ряду (в більшості варіантів задач) визначити, чи присутній тренд в цьому ряді.

3.1. Побудувати рівняння АРКС, яке найбільш адекватно описує відповідний набір даних (дивіться алгоритм підбора рівняння). Обрати кращий

підхід, що описує даний ряд. Кращий результат вивести у вигляді таблиці статистичних характеристик і результатів прогнозування.

4. Перевірити дані на наявність трендів. Побудувати моделі АРІКС та моделі з трендом (порядок тренду визначається експериментально!).

5. Побудовати моделі для оцінювання волатильності фінансових процесів з використанням гетероскедастичних моделей АРУГ і УАРУГ, моделей зі змінною волатильністю: МСВ, ЕУАРУГ, ЧІУАРУГ, тощо.

6. Навести **<u>дві таблиці</u>** результатів оцінювання якості моделей на основі критеріїв R^2, IKA, DW, BS, тощо:

а) моделей фінансового процесу для прогнозування абсолютного значення фінансової змінної з використанням моделей АР, АРКС, АРІКС; з трендом різного порядку.

б) моделей для оцінювання дисперсії фінансового ряду (з використанням різних гетероскедастичних моделей АРУГ, УАРУГ, МСВ, ЕУАРУГ, тощо).

7. Використати побудовані моделі для прогнозування значень ціни валюти(акцій) на 1-5 кроків вперед та значення дисперсії на 1 крок.
8. Порівняти результати прогнозування з реальними даними (для 5 значень). Оцінити точність прогнозів на основі статистичних критеріїв.

Основні кроки побудови моделі АРКС:

<u>1. Побудова АРКС(p,q) коли КС будується по залишкам АР(p) рівняння моделі.</u>

1.1. Визначення p – порядку авторегресійної складової.

1.1.1. Обчисліть АКФ та ЧАКФ для 12 лагів часового ряду

1.1.2. По отриманим значенням АКФ та ЧАКФ визначте p – порядок авторегресійної складової [3, 4].

1.2. Визначення $q-$порядку КС.

1.2.1. Обчисліть коефіцієнти моделі АР(р) : $y(k)=a_0+\sum_{i=1}^{p} y(k-i)$.

1.2.2.Обчисліть коефіцієнти АКФ та ЧАКФ для 12 лагів залишків моделі АР(р).

1.2.3. По отриманим значенням АКФ та ЧАКФ визначте $q-$порядок КС.

Підхід №1. Застосування стандартного КС. Побудуйте АРКС(р,q).

Отримані результати занесіть в таблицю 1. Буде отримано чотири рівняння (1) для випадку, коли просте КС з $N=5$; (2) просте КС з $N=10$; (3) експоненційне КС з $N=5$; (4) експоненційне КС з $N=10$.

Підхід №2. Застосування власного КС.

Побудуйте просте КС за залишками АР(р) з розміром вікна КС $N=5$ і $N=10$.

Побудуйте АРКС(р,q) та занесіть в таблицю 1 отримані результати.

Побудуйте експоненційне КС по залишкам АР(р). Побудуйте АРКС(р,q) та занесіть в таблицю 2 отримані результати.

2. Побудова АРКС(р,q), коли КС будується по вихідному сигналу y

2.1. Визначення $p-$ порядку авторегресійної складової.

2.1.1. Обчисліть АКФ та ЧАКФ для 12 лагів часового ряду

2.1.2. По отриманим значенням АКФ та ЧАКФ визначте $p-$ порядок авторегресійної складової.

2.2. Побудова КС по y.

Побудуйте просте КС по y з розміром вікна КС $N=5$ та $N=10$.

Побудуйте експоненційне КС по y з розміром вікна КС $N=5$ та $N=10$

2.3. Визначення $q-$порядку КС.

Для цього побудуйте ЧАКФ(КС) та визначте q.

2.4. Побудуйте АРКС(p,q).

Підхід №1. Застосування власних коефіцієнтів при КС.

Обчисліть коефіцієнти $b_1,...,b_q$ за формулою

$$b_i = \frac{(1-\alpha)^i}{\sum_{j=1}^{q}(1-\alpha)^j}, \text{ де } \alpha = \frac{2}{q+1}.$$

Після чого обчисліть коефіцієнти рівняння:

$$y1(k) = y(k) - mv(k) - \sum_{j=1}^{q} b_j \cdot mv(k-j) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i),$$

де $mv-$КС. Отримані результати занесіть в таблиці 2 та 3. Буде отримано чотири рівняння (1) для випадку, коли просте КС з $N=5$; (2) просте КС з $N=10$; (3) експоненційне КС з $N=5$; (4) експоненційне КС з $N=10$.

Підхід №2. Обчислення коефіцієнтів АРКС(p,q).

В даному випадку коефіцієнти $b_1,...,b_q$ вважаються невідомими і їх необхідно обчислити разом з коефіцієнтами $a_0, a_1,..., a_p$.

$$y(k) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i) + mv(k) + \sum_{j=1}^{q} b_j \cdot mv(k-j).$$

Отримані результати занесіть в таблиці 2 та 3. Буде отримано чотири рівняння (1) для випадку, коли просте КС з $N=5$; (2) просте КС з $N=10$; (3) експоненційне КС з $N=5$; (4) експоненційне КС з $N=10$.

Таблиця 1

Зразок порівняльної таблиці всіх моделей

	R - squared	Sum squared resid	Akaike	Durbin – Watson stat
АРКС(p,q), де КС побудоване по залишкам АР(p) рівняння				
АРКС(p)				

АРКС(p,q) із застосуванням КС системи Eviews				
АРКС(p,q) із застосуванням власного простого КС, при N=5.				
АРКС(p,q) із застосуванням власного простого КС, при N=10.				
АРКС(p,q) із застосуванням власного експоненційного КС, при N=5.				
АРКС(p,q) із застосуванням власного експоненційного КС, при N=10.				
Побудова АРКС(p,q), де КС побудоване по вихідному сигналу y				
Підхід №1. Застосування власних коефіцієнтів при КС.				
АРКС(p,q) із застосуванням власного простого КС по y, при N=5.				
АРКС(p,q) із застосуванням власного простого КС по y, при N=10.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=5.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=10.				
Підхід №2. Обчислення коефіцієнтів АРКС(p,q).				
АРКС(p,q) із застосуванням власного простого КС по y, при N=5.				
АРКС(p,q) із застосуванням власного простого КС по y, при N=10.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=5.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=10.				

Послідовність побудови моделей з трендом

Детермінований тренд описують вибраною функцією, наприклад, поліномом від часу, сплайном, експонентою, комбінацією тригонометричних функцій та інше. Часто використовують поліноми від часу вигляду:

$$y(k) = a_0 + a_1 \cdot k + a_2 \cdot k^2 + \ldots + a_m \cdot k^m + \varepsilon(k), \qquad (2.1)$$

де k – дискретний час, який зв'язаний з неперервним реальним часом t через період реєстрації (дискретизації) даних: $t = kT_s$; $\varepsilon(k)$ – випадкова змінна, оцінку якої можна знайти після оцінювання рівняння: $\hat{\varepsilon}(k) = e(k)$, де $e(k)$ – похибка моделі. Очевидно, що після оцінювання моделі послідовність значень $\{e(k)\}$ буде містити всі коливання, що накладаються на тренд.

Описуючи тренд рівнянням (2.1), ми фактично видаляємо його з процесу і повна модель процесу буде складатись щонайменше з двох рівнянь: рівняння (2.1) для тренду і рівняння АРКС(p,q), яке описує коливання, що накладаються на тренд.

Тренд може бути видалений з процесу (даних) за допомогою різниць. Так, перші різниці видаляють тренд першого порядку (лінійний тренд), другі різниці видаляють квадратичний тренд і т.д. Наприклад, нехай $y(k) = a_0 + a_1 \cdot k$. Перші різниці цього процесу

$$\Delta y(k) = y(k) - y(k-1) = a_0 + a_1 \cdot k - [a_0 + a_1 \cdot (k-1)] = a_1$$

приводять до видалення лінійного тренду. Очевидно, що після видалення тренду ми вже не зможемо його спрогнозувати.

Рівняння тренду другого порядку в даному випадку записується як

$$y(k) = a_0 + a_1 \cdot k + a_2 \cdot k^2 + \varepsilon(k),$$

де $k-$ дискретний час, який є послідовністю натуральних чисел від 1 до 128, а $\varepsilon(k)=resid$, тобто залишок, отриманий після оцінювання рівняння $y(k)=a_0+a_1\cdot k+a_2\cdot k^2$.

В процесі побудови математичних моделей зустрічаються статистичні дані, представлені досить великими числами, наприклад, десятки і сотні тисяч, мільйони і мільярди. Такі числа необхідно перетворювати у більш прийнятні для обчислювального процесу величини. Оскільки для алгоритмів оцінювання параметрів моделей характерним є накопичення похибок обчислень, а в деяких випадках необхідно обчислювати обернені матриці, то великі числа необхідно логарифмувати або нормувати у вибраному діапазоні значень.

Таблиця 2

Приклади запису трендів

$\log y(k)=25{,}4749+0{,}0157\cdot k$	тренд 1-го порядку
$\log y(k)=25{,}4749+0{,}0157\cdot k+0{,}666\cdot k^2$	тренд 2-го порядку
$\log y(k)=25{,}4749+0{,}0157\cdot k+0{,}666\cdot k^2+0{,}13\cdot k^3$	тренд 3-го порядку
$\log y(k)=d\log y(k)+\log y(k-1)$, де $d\log y(k)=0{,}0228-0{,}0151\cdot d\log y(k-1)-0{,}0073\cdot d\log y(k-3)+$ $+0{,}9817\cdot d\log y(k-4)-0{,}9389\cdot ma(k-4)$	АРІКС(4,1,4)
$\log y(k)=0{,}0838+1{,}0037\cdot\log y(k-1)$	АР(4)
$\log y(k)=22{,}8322+1{,}004\cdot\log y(k-1)+0{,}2542\cdot ma(k-1)-$ $-0{,}2756\cdot ma(k-3)+0{,}6569\cdot ma(k-4)$	АРКС(4,4)

Якщо процес містить **сезонний ефект**, то він враховується шляхом введення в праву частину рівняння залежної змінної із затримкою, що дорівнює періоду сезонного ефекту, або введення складової ковзного середнього з тією

ж затримкою. Для зменшення дисперсії процесу з сезонним ефектом застосовують так звані “сезонні” різниці, тобто різниці вигляду:

$$\Delta_4 y(k) = y(k) - y(k-4),$$

де “4” означає періодичність сезонного ефекту.

Порівняння моделей на основі статистичних критеріїв і обрання кращої:

Таблиця 3

Статистичні критерії моделей

	R2	Sum squared resid	IKA	DW	BS
Тренд 1-го порядку					
Тренд 2-го порядку					
Тренд k-го порядку					
АРІКС(p,k,q)					
АР(p)					

Статичне прогнозування на основі тренду першого порядку може бути представлено у вигляді таблиці 4.

Таблиця 4

Значення змінних АРІКС моделі при статичному однокроковому прогнозуванні

Час k	Прогнозне значення ряду ***“dlogy”*** dlogy_f	Значення ряду ***“logy”***	Прогнозне значення ряду “logy” ***logy(k)=dlogy(k)+logy(k-1)*** logy_f
1990/4	–	27.472889	–
1991/1	-0.01995	27.457827	27.45294
1991/2	0.034743	27.476812	27.49257
1991/3	0.009746	27.473663	27.48656
1991/4	0.050085	27.551847	27.52375

Тепер для відновлення **прогнозу реальних значень** (не логарифмованих) потрібно просто взяти експоненту від останньої колонки (e^{logy_f}) і порівняти отримані прогнозні значення з реальними значеннями ряду.

Послідовність побудови гетероскедастичних моделей.

1. Знайдіть середнє значення ряду, стандартне відхилення і побудуйте для введених даних модель АР(1).

2. Побудуйте автокореляційну функцію (АКФ) для отриманого після побудови моделі АР(1) ряду із залишків $\hat{\varepsilon}(k)$. Залишки мають ідентифікатор *resid,* а для побудови АКФ можна скористатися командами: *Quick —> Series Statistics —> Correlogram —> OK —> Level —> OK.*

3. Згенеруйте новий ряд із квадратів залишків: $\{\hat{\varepsilon}^2(k)\} = \{e^2(k)\}$, де $e(k)$ – залишки (похибки) моделі АР(1). Для цього можна скористатися командами: *Proc —> Generate Series —> resid_2=resid*resid.*

4. Обчисліть автокореляційну функцію для ряду із квадратів залишків.

5. За допомогою МНК обчисліть оцінки коефіцієнтів рівняння першого порядку для дисперсії залишків:

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1\hat{\varepsilon}^2(k-1) + v(k),$$

де $v(k)$ – похибки моделі. Таку модель називають авторегресійною умовно гетероскедастичною або скорочено АРУГ (в даному випадку АРУГ(1)).

6. Обчисліть коефіцієнти рівняння АРУГ(р):

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1\hat{\varepsilon}^2(k-1) + \alpha_2\hat{\varepsilon}^2(k-2) + \alpha_3\hat{\varepsilon}^2(k-3) + \alpha_4\hat{\varepsilon}^2(k-4).$$

7. Побудуйте ряд умовних дисперсій і побудуйте узагальнену авторегресійну умовно гетероскедастичну (УАРУГ) модель процесу у вигляді [4, 5,6]:

$$h(k) = \alpha_0 + \sum_{i=1}^{q}\alpha_i\hat{\varepsilon}^2(k-i) + \sum_{i=1}^{p}\beta_i h(k-i) + v(k),$$

де q – визначається за допомогою АКФ процесу $\{\varepsilon^2(k)\}$; p – визначається за допомогою АКФ для $\{h(k)\}$; $h(k)$ – умовна дисперсія процесу, яка розраховується послідовно для досліджуваного ряду за виразом:

$$h(k) = E_k[y^2(k)] = \frac{1}{k-1}\sum_{i=1}^{k}[y(i) - \bar{y}_k]^2, \ \ k = 2,3,...,N,$$

де N – довжина ряду; $\bar{y}_k = \frac{1}{k-1}\sum_{i=1}^{k} y(i), \ k = 2,...,N$ – умовне математичне сподівання.

9. Побудуйте модель гетероскедастичного процесу

$$h(k) = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon^2(k-i) + \sum_{i=1}^{p}\beta_i h(k-i),$$

де

$$h(k) = E_k[\varepsilon^2(k)] = \frac{1}{k-1}\sum_{i=1}^{k}[\varepsilon(i) - \bar{\varepsilon}_k]^2, \ \ k = 2,3,...,N,$$

$$\bar{\varepsilon}_k = \frac{1}{k-1}\sum_{i=1}^{k}\varepsilon(i), \ k = 2,...,N.$$

10. Наведіть графіки процесів: $y(k)$, $h(k)$, $\varepsilon(k)$, $\varepsilon^2(k)$ та автокореляційні функції, використані при побудові моделей.

11. Побудова моделі EGARCH:

$$y_t = \sigma_t\varepsilon_t, \quad t = 1,2,...,T$$

$$log\left(\sigma_t^2\right) = \alpha_0 + \left[1 - \sum_{j=1}^{q}\beta_j B^j\right]^{-1}\left[1 + \sum_{i=1}^{p}\alpha_i B^i\right]g\left(\varepsilon_{t-1}\right),$$

де всі змінні і параметри визначаються так само, як у моделі АРУГ; крім того, B^j та B^i – лінійні оператори зсуву, для яких: $x_t B^j = x_{t-j}$, та $g(\varepsilon_t) = \lambda_1\varepsilon_1 + \lambda_2\left(|\varepsilon_t| - E\left(|\varepsilon_t|\right)\right)$, де λ_1 і λ_2 додаткові параметри.

12. Побудова моделі FIGARCH(*p*,*q*)) (FIGARCH = Fractionally Integrated GARCH):

$$y_t = \sigma_t \varepsilon_t, \quad t = 1,2,\ldots,T$$

$$\sigma_t^2 = \alpha_0 \left[1 - \sum_{j=1}^{q} \beta_j B^j \right]^{-1} + \left\{ 1 - \left[1 - \sum_{j=1}^{q} \beta_j B^j \right]^{-1} \sum_{i=1}^{p} \alpha_i B^i (1-B)^d \right\} y_t^2,$$

де всі змінні та параметри визначені як у моделях АРУГ, за винятком того, що B^j та B^i – оператори, для яких: $x_t B^j = x_{t-j}$, а d – дробовий параметр різниці.

13. Порівняйте моделі, побудовані на кроках 6 – 12 у вигляді таблиці 5:

Таблиця 5

Таблиця статистичних характеристик моделей

	R2	Sum squared resid	IKA	DW	BS
ARCH(p)					
GARCH(p,q)					
EGARCH(p,q)					
FIGARCH(p,q)					

Запитання для самоконтролю

1. Опишіть основі етапі побудови авторегресійної моделі з ковзним середнім за вихідним сигналом y.
2. В чому різниця підходів 1 і 2 до побудови АРКС(p,q)?
3. Як побудувати модель АРКС по залишкам?
4. Як можна урахувати тренд?
5. Як будується модель АРУГ?
6. За якими статистичними критеріями обирається краща модель прогнозування?
7. Як оцінити якість прогнозу?

3. Аналіз та дослідження трендів фінансового ринку у середовищі R

В цьому розділі ми покажемо, як підключити і налаштувати програмне середовище R [10] для виконання дослідження фінансових рядів, а також побудови найпростіших моделей.

Нижче буде показано елементи коду, написаного на R, для автоматизованого видобування даних з фінансового сайту Finam.

```
library(rusquant)  # Бібліотека для завантаження котировок з сайту
ts1 <- getSymbols("EURUSD", src="Finam", from='2014-01-01', to ='2014-12-31', auto.assign=FALSE)
head(ts1)                          # Вивели 5 перших рядків
adf.test(ts1[,4], alternative="stationary")  # Тест ADF.
```

Для розширення можливостей R використовуються додаткові пакети-розширення. Щоб встановити пакет, треба в консолі R виконати команду:

```
install.packages("НазваПакету")
```

Другим параметром можна вказати, щоб автоматично були встановлені всі інші пакети, які потрібні для установки даного пакета, наприклад:

```
install.packages("xts", dependencies=TRUE)
install.packages("PerformanceAnalytics", dependencies=TRUE)
```

Будемо використовувати такі пакети:

- tseries для аналізу часових рядів;
- quantmod для скачування біржових даних;
- TTR для роботи з технічними індикаторами;

- quantstrat для тестування стратегій;
- PerformanceAnalytics для аналізу ефективності і ризиків;
- FinancialInstrument для зберігання відомостей про фінансові інструменти (тікер, ціна тика, місяці експірації, ...)

За допомогою пакета quantmod можна отримати котирування з різних джерел:

- **Yahoo Finance** (ціни відкриття, максимуми дня, мінімуми, ціни закриття і обсяги; котирування валют недоступні);
- **Google Finance** (ціни відкриття, максимуми дня, мінімуми, ціни закриття і обсяги);
- **Federal Reserve Bank of St. Louis FRED®;**
- **Oanda** (FOREX і метали, тільки ціни закриття);
- локальна база даних MySQL;
- виконавчі файли R (.RData і .rda);
- текстові файли CSV.

Відкрийте RStudio і встановіть пакет quantmod.

1. Формуємо файл вхідних даних з офіційного сайту за допомогою пакету quantmod

1.1. Завантажуємо бібліотеку в пам’ять.

> library(quantmod)

1.2. Тепер завантажуємо з сайту Yahoo Finance котирування акції Microsoft (тікер MSFT):

> getSymbols("MSFT", src = "yahoo")

2. Будуємо біржові графіки:

chartSeries(MSFT)

2.1. Для виділення певних значень кварталу використовуємо команду:

chartSeries(MSFT["YY-MM/YY-MM"])

За замовчуванням будуються графіки на чорному фоні, що не є зручним і презентативним для звітів та практичних робіт, тому рекомендовано одразу змінити налаштування на «білу тему» за допомогою команди:

chartSeries(MSFT["2014-12/2020-02"], theme="white")

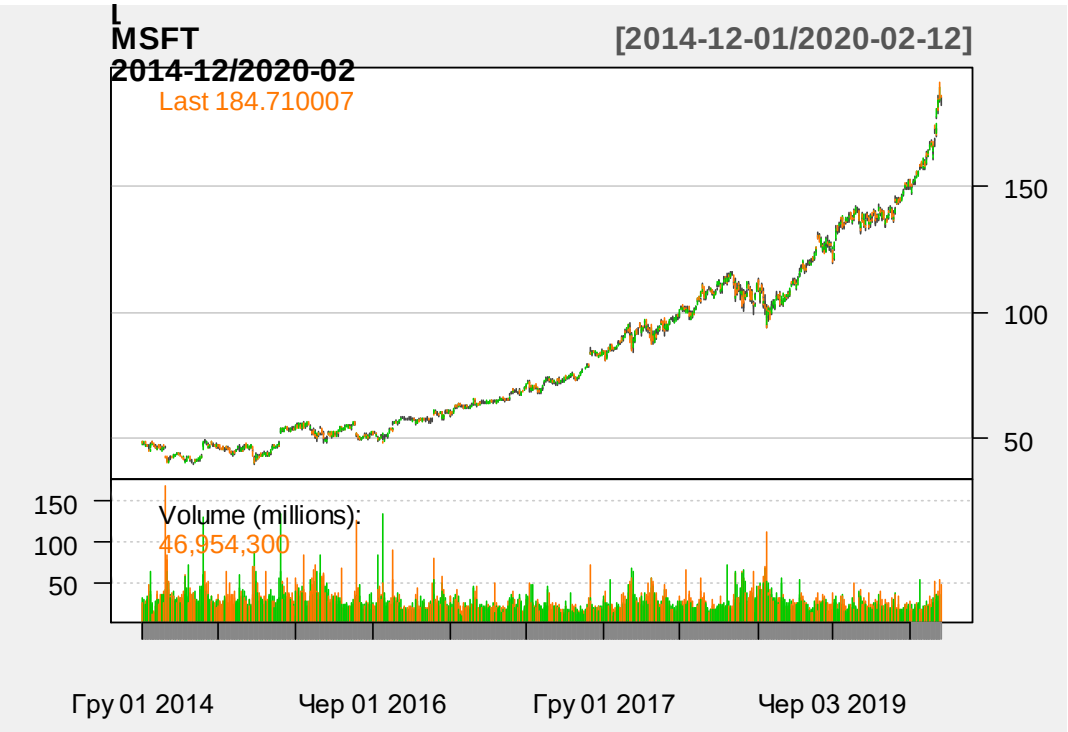


Рис. 2. Візуалізація даних на білому фоні

Для виведення даних (наприклад за січень 2019) у вигляді таблиці за місяць необхідно виконати наступну команду:

MSFT["2019-01"]

2.2. Задамо тижневий часовий період (week), оберемо для відображення тільки 2017 і 2020 роки, змінимо колір фону на білий, свічки «бики» зобразимо зеленим, а «ведмеді» - червоним кольором:

candleChart(to.weekly(MSFT["2018/2020"]), theme="white", up.col='green', dn.col='red')

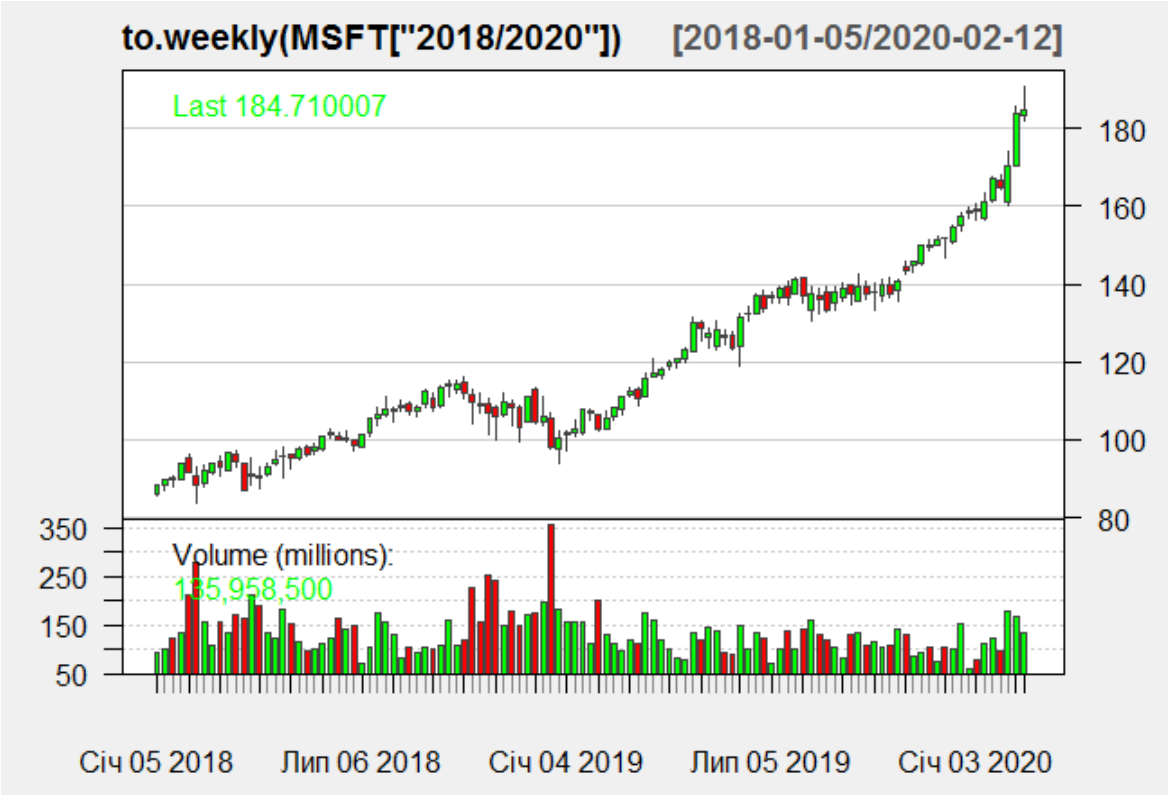


Рис. 3. Візуалізація даних за 2018-2020 рр.

2.3. У вигляді барелів побудуємо графік за 2019 рік:

barChart(MSFT["2019"],theme="white")

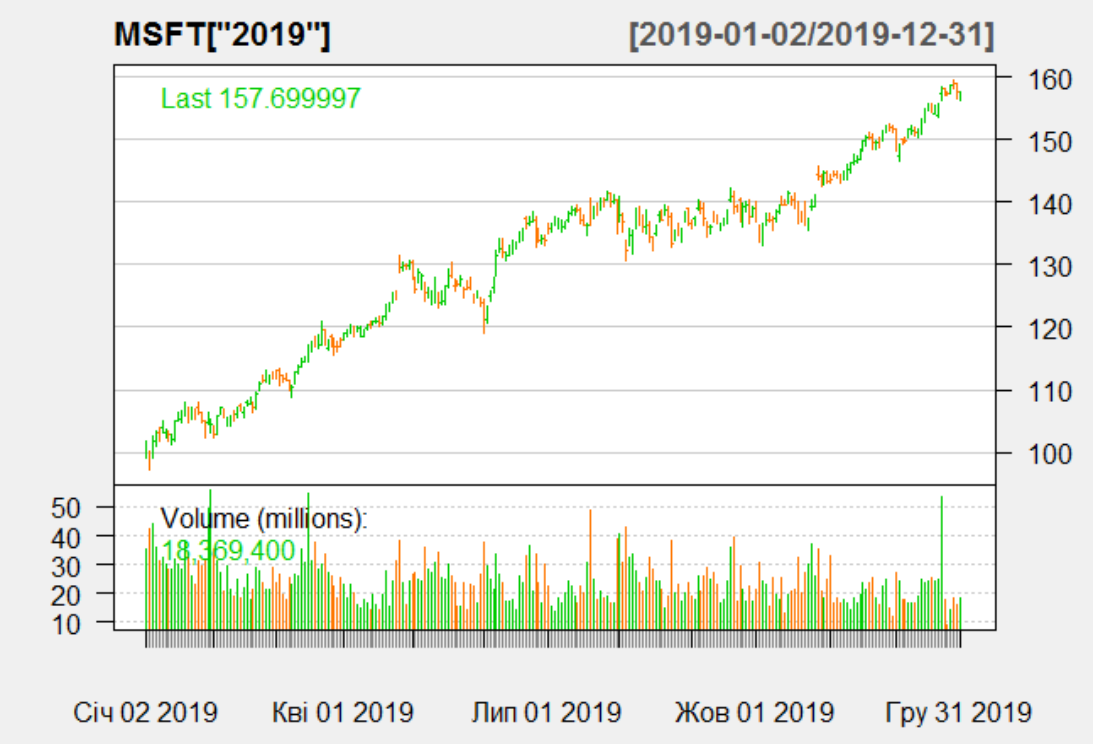


Рис. 4. Візуалізація даних за 2019 р.

Тепер спробуємо оцінити технічні індикатори ринку. Для цього скористаємось бібліотекою TTR:

library(TTR)

і додамо на графіки з рис. 2 обрані нами індикатори (MACD і смуги Болліндженра):

addMACD()

addBBands()

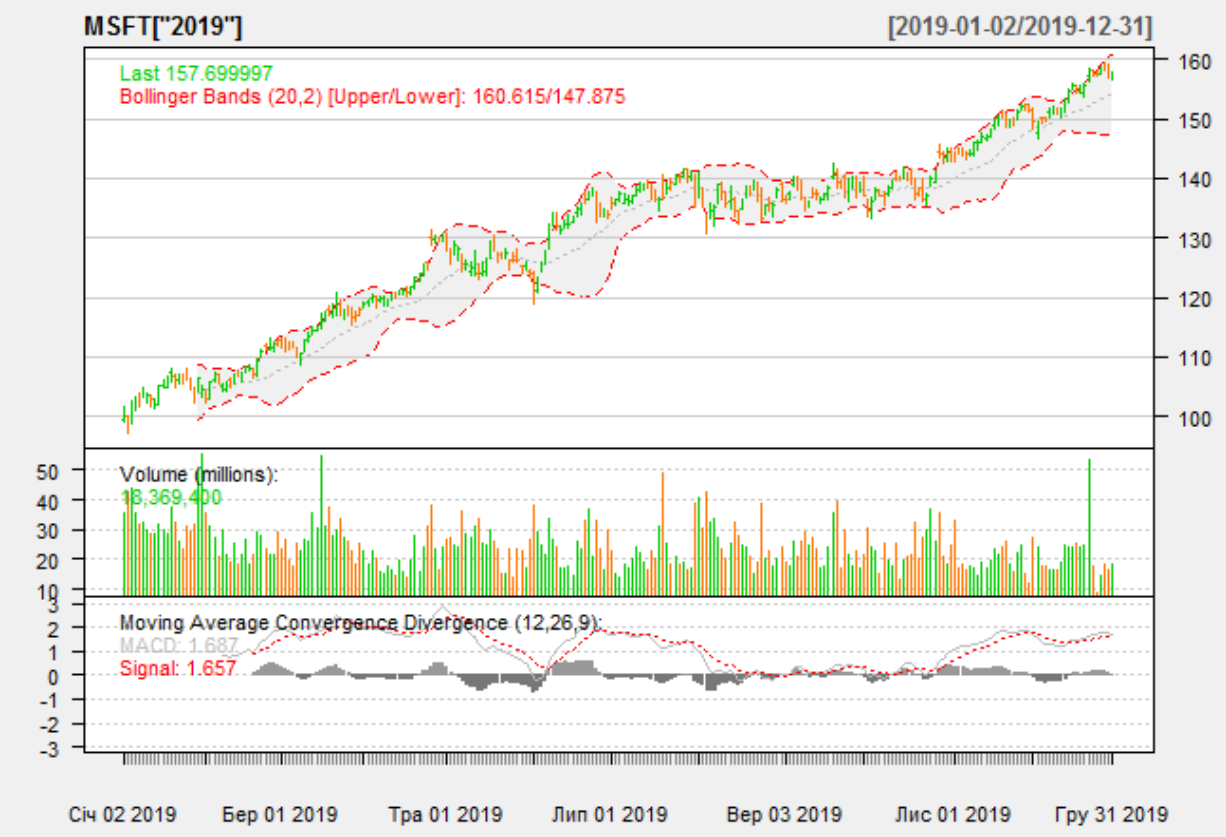


Рис. 5. Візуалізація даних зі смугами Болліндже́ра

3. Тепер перевіримо гіпотезу про стаціонарність наших рядів.

Розширений тест Дікі - Фуллера (Augmented Dickey-Fuller test, ADF) використовується для перевірки часового ряду на стаціонарність. За нульову гіпотезу розглядається наявність одиничного кореня (unit root, тобто хоча б один з коренів характеристичного полінома лежить на одиничному колі), тобто нестаціонарність ряду. Тест ADF є одностороннім: в якості альтернативної гіпотези за замовчуванням вважається гіпотеза про стаціонарність ряду (всі корені характеристичного полінома лежать поза одиничним колом); інший (рідкісний для економетричних рядів) варіант - наявність коренів всередині одиничного кола (вибухове зростання елементів ряду).

3.1. Дані MSFT є матрицею, де в першій колонці зберігаються ціни відкриття, в другій колонці – максимальна ціна протягом даного дня, в третій

колонці – мінімальна ціна за даний день, в четвертій колонці – ціна закриття, в п’ятій колонці – біржовий обсяг, а в шостій колонці – скоригована ціна закриття. Для того, щоб перевірити стаціонарність ряду MSFT.OPEN, введемо наступну команду:

adf.test(MSFT[,1], alternative="stationary")

В результаті отримаємо:

Augmented Dickey-Fuller Test

data: MSFT[, 1]

Dickey-Fuller = 3.6761, Lag order = 14, p-value = 0.99

alternative hypothesis: stationary

Оскільки p-value є значущим, то немає підстав відкидати гіпотезу про наявність одиничного кореня і ряд можна вважати нестаціонарним.

3.2. Перевіримо на стаціонарність ряд MSFT.Close (він є 4ю колонкою нашої матриці):

> adf.test(MSFT[,4], alternative="stationary")

Augmented Dickey-Fuller Test

data: MSFT[, 4]

Dickey-Fuller = 3.5733, Lag order = 14, p-value = 0.99

alternative hypothesis: stationary

Знову отримуємо, що ряд можна вважати нестаціонарним.

3.2. Аналогічно перевіримо стаціонарність обсягів торгів (MSFT.Volume):

> adf.test(MSFT[,5], alternative="stationary")

Augmented Dickey-Fuller Test

data: MSFT[, 5]

Dickey-Fuller = -10.328, Lag order = 14, p-value = 0.01

alternative hypothesis: stationary

Оскільки p-value є малим, то з рівнем значущості 0.01 ряд можна вважати стаціонарним.

Можна перевіряти стаціонарність певного підряду. Для цього треба вказати діапазон рядків ряду, які ми хочемо виділити в якості підряду.

adf.test(MSFT[78:93, 4], alternative = "stationary")

Отже, перші висновки щодо подальшого моделювання: для даних обсягів торгів можна застосувати моделі АР, АРКС, а от для цін відкриття/закриття потрібно перевірити ряд на нелінійність і будувати моделі з трендом та гетероскедастичними моделями.

4. Побудова ACF та PACF для стаціонарного ряду обсягів торгів.

acf(MSFT[,4])

> acf(MSFT[3000:3301,4],10)

> acf(MSFT[,5],10)

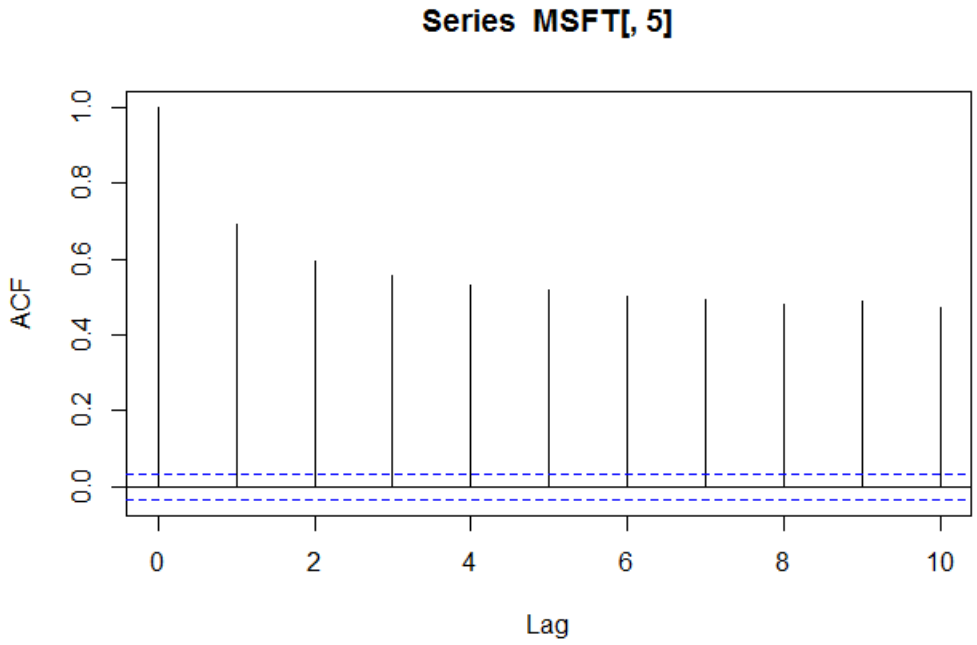


Рис. 6. Лаги для АКФ

pacf(MSFT[,5],lag.max = 10)

5. PACF для значення обсягів з максимальним лагом 10.

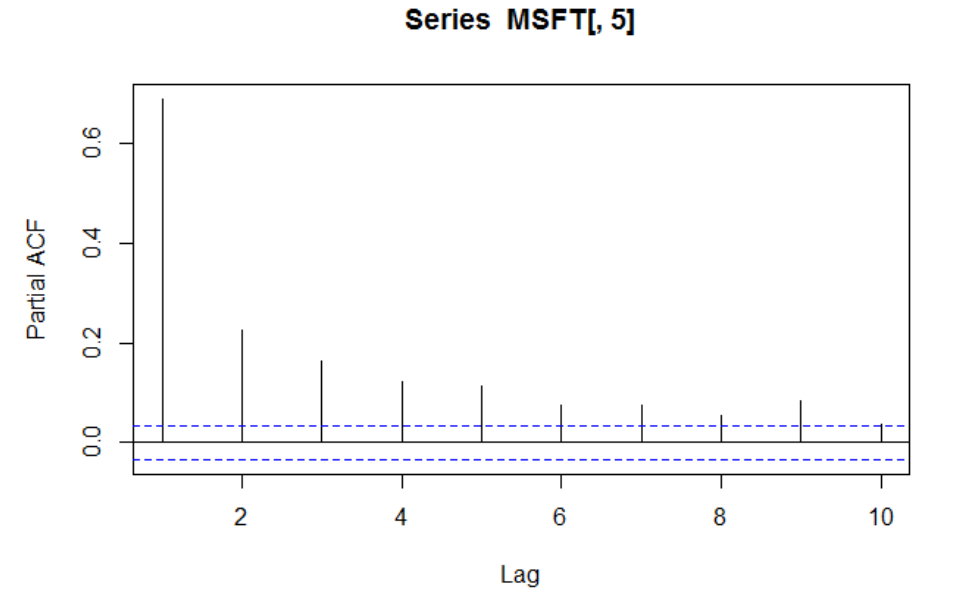


Рис. 7. Лаги для ЧАКФ

Отже, 2й порядок моделі для нас є значимим.

6. Побудуємо графік цін закриття:

PriceData<-ts(MSFT[,4], frequency = 5)
> plot(PriceData)

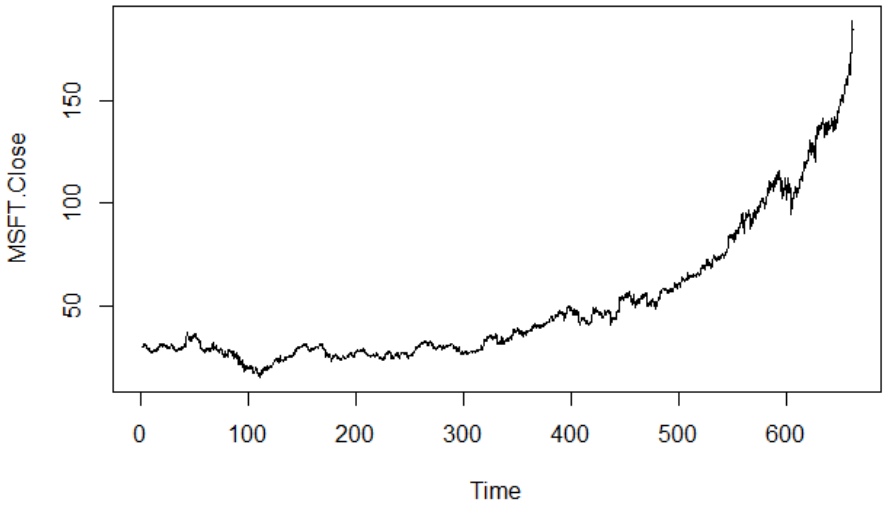


Рис. 8. Вид графіку цін закриття

Дійсно, навіть візуально ряд є нестаціонарним.

А ось графік для обсягів акцій:

PriceData1<-ts(MSFT[,5], frequency = 5)

> plot(PriceData1)

7. Встановлюємо пакети для GARCH

install.packages("rugarch")

8. Скачати дані S&P

```
SYM <- as.character(
read.csv('http://trading.chrisconlan.com/SPstocks_current.csv',
                    stringsAsFactors = FALSE, header = FALSE)[,1] )
```

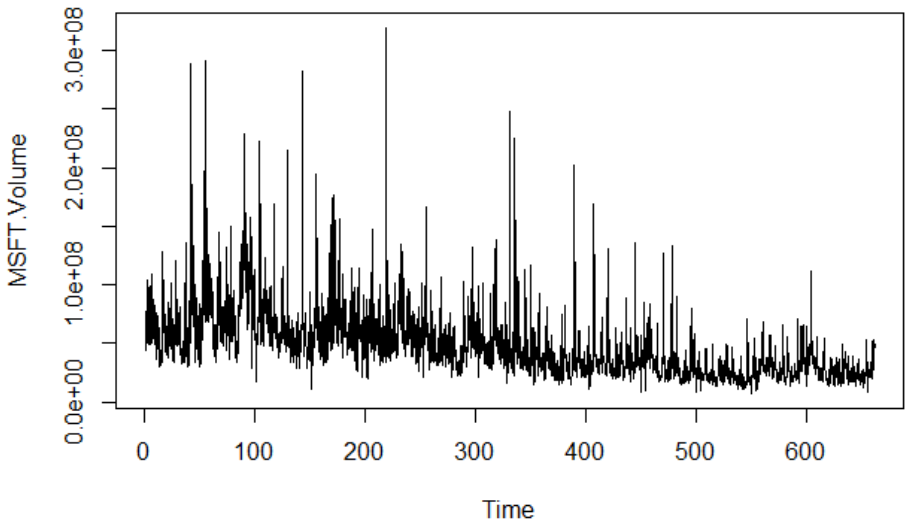


Рис. 9. Візуалізація щоденних обсягів продажів

Отже, ми показали можливості і послідовність побудови простих регресійних моделей для реальних фінансових даних і представлення (візуалізації) в зручному для користувачів вигляді.

Запитання для самоконтролю

1. Які пакети в R використовуються для видобування і аналізу даних з фінансових ринків?
2. Як оцінити технічні індикатори і побудувати смуги Болліндже́ра?
3. Як перевірити стаціонарність ряду?
4. Як побудувати ЧАКФ в R?

4. Методика дослідження і прогнозування фінансових процесів, реалізована у Python

Розглянемо особливості реалізації і застосування існуючих програмних застосунків та бібліотек при виконанні моделювання і прогнозування у програмному середовищі Python.

Поетапне виконання дослідження і прогнозування фінансових процесів.

1. Завантажуємо дані з ринку акцій

У кожному рядку зберігається дата (Date) в форматі рік-місяць-день; ціна відкриття (Open); максимальна (High) і мінімальна ціна (Low), досягнута за даний день; ціна закриття (Close); біржовий обсяг (Volume) і скоригована ціна закриття (Adjusted Close). Останній стовпець враховує поділ акцій і виплату дивідендів; для індексів і валют (а також ф'ючерсів на них) значення в останньому стовпці збігаються зі значеннями в стовпці Close.

2. Створюємо новий набір даних, з тією ж кількістю рядків, але нам потрібні тільки ціни закриття і обсяги.

3. Додаємо до нового набору даних 2 стовпці, які будуть зберігати затримані ціни закриття, зрушені на 1 день (стовпець Lag1) і на 2 дні (Lag2).

4. Створюємо ще один набір даних, тільки в кожному i-му рядку замість ціни закриття p_i будемо зберігати зміну цієї ціни порівняно з попереднім днем, виражену у відсотках по відношенню до попереднього дня, за формулою:

5. $$d_i = \frac{(p_i - p_{i-1})}{p_{i-1}} \times 100\%$$

```
ts = pd.DataFrame(index=ts2.index)
ts["Volume"] = ts2["Volume"]
ts["Today"] = ts2["Today"].pct_change()*100.0
```

6. Замінюємо нульові значення малими числами (наприклад, 0.0001), тому що нульові значення будуть заважати роботі деяких алгоритмів навчання (див. нижче).

```
for i,x in enumerate(ts["Today"]):
    if (abs(x) < 0.0001):
        ts["Today"][i] = 0.0001
print ts.head(3)
```

7. Додаємо до нового набору даних ті ж 2 стовпці для затриманих даних, але замість ціни закриття будемо зберігати процентну зміну (d). Крім того, нам буде потрібно ще один стовпець Direction (Напрямок): якщо ціна закриття виросла в порівнянні з попереднім днем, то в цей стовпець запишемо 1, а якщо ціна закриття зменшилася, то значення -1.

```
for i in xrange(0, count):
    ts["Lag%s" % str(i+1)] = ts2["Lag%s" % str(i+1)].pct_change()*100.0

ts["Direction"] = np.sign(ts["Today"])
ts = ts[count+1:]  # Пропустили перші count днів, оскільки для них деякі дані не визначені (NaN)
print ts.head(7)
```

8. Розділяємо наш набір даних на дві частини: навчальну і тестову. Першу будемо використовувати для навчання моделі, а другу - для тестування результатів.

```
start_test = datetime.datetime(2014,1,1) # Початок тестового набору даних - 1 січн 2014
```

9. Для навчання моделі будемо використовувати значення з стовпців Lag1 і Lag2:

```
cols = []
for i in xrange(0, count):
    cols.extend(["Lag%s" % str(i+1)])
print cols
```

10. Цільові значення – напрямки зміни ціни закриття, тобто значення зі шпальти Direction.

```
x = ts[cols] # Вхідні дані для навчання
y = ts["Direction"]     # Цільові значення

x_train = x[x.index < start_test] # Вхідні дані для навчання(березень)
print x_train[0:5] # Вивели для перевірки перші 5 рядків навчальної вибірки
x_test = x[x.index >= start_test] # Вхідні дані для тестування (квітень)
print x_test[0:5] # Вивели для перевірки перші 5 рядків навчальної вибірки

y_train = y[y.index < start_test] # Цільові значення для навчання (березень)
print y_train[0:5]
y_test = y[y.index >= start_test] # Цільові значення для навчання (квітень)
print y_test[0:5]

#from sklearn import preprocessing
#scaler = preprocessing.StandardScaler().fit(x_train)
#x_train = scaler.transform(x_train)
#x_test = scaler.transform(x_test)
#print x_train[0:5] # Вивели для перевірки перші 5 рядків навчальної
вибірки після масштабування

d = pd.DataFrame(index=y_test.index) # Набір даних для перевірки моделі
d["Actual"] = y_test                # Реальні зміни цін закриття
```

11. Використання логістичної регресії для передбачення зміни ціни:

```
from sklearn.linear_model import LogisticRegression
model1 = LogisticRegression()
model1.fit(x_train, y_train)     # Навчання (підбір параметрів моделі)
d['Predict_LR'] = model1.predict(x_test) # Тест
```

```
# Розраховуємо процент правильно передбачених направлень змін ціни:
d["Correct_LR"] = (1.0+d['Predict_LR']*d["Actual"])/2.0
print d
hit_rate1 = np.mean(d["Correct_LR"])
print "Процент вірних передбачень: %.1f%%" % (hit_rate1*100)
```

В результаті за допомогою логістичної регресії вдалося правильно спрогнозувати 56 випадків зі 100, правильно передбачивши напрям руху ціни за день. Всі виконані кроки – це пошук певних закономірностей на площині: ми намагалися спрогнозувати напрям зміни ціни по змінам за два попередні дні. Тепер можна спробувати збільшити розмірність завдання і врахувати зміну ціни на акції протягом дня. Для цього буде потрібна деталізація щогодини.

Запитання для самоконтролю

1. Які бібліотеки в Python використовуються для аналізу даних з фінансових ринків?
2. Яка деталізація цін використовується для побудови прогнозних моделей?
3. Щоденна деталізація цін дозволяє здійснювати коротко- чи середньострокове прогнозування?
4. Як спрогнозувати напрям зміни ціни на фінансовому ринку?
5. Які ще моделі варто застосовувати для прогнозування цін на акції, а які для прогнозування напряму зміни (зростання чи спадання)?

5. Дослідження і прогнозування фінансового ряду цін на акції компанії Tesla

Метою даного розділу є ілюстрація виконання процесу моделювання фінансового процесу і основних завдань фінансового аналітика на прикладі загальновідомої компанії Tesla.

Вхідними даними є ціни на акції компанії Tesla Inc. за період 2020 – 2023 роки, отримані з ресурсу Yahoo Finance. (https://finance.yahoo.com/quote/TSLA?p=TSLA). Датасет містить такі дані як: дата (Date), ціна акцій при відкритті на задану дату (Open), найвища ціна на акції компанії протягом цього дня (High), найнижча ціна на акції протягом дня (Low), скорегована ціна (Adj. Close), обсяг акцій, реалізованих протягом дня (Volume). Графіки візуалізації даних виглядають наступним чином (рис. 10):

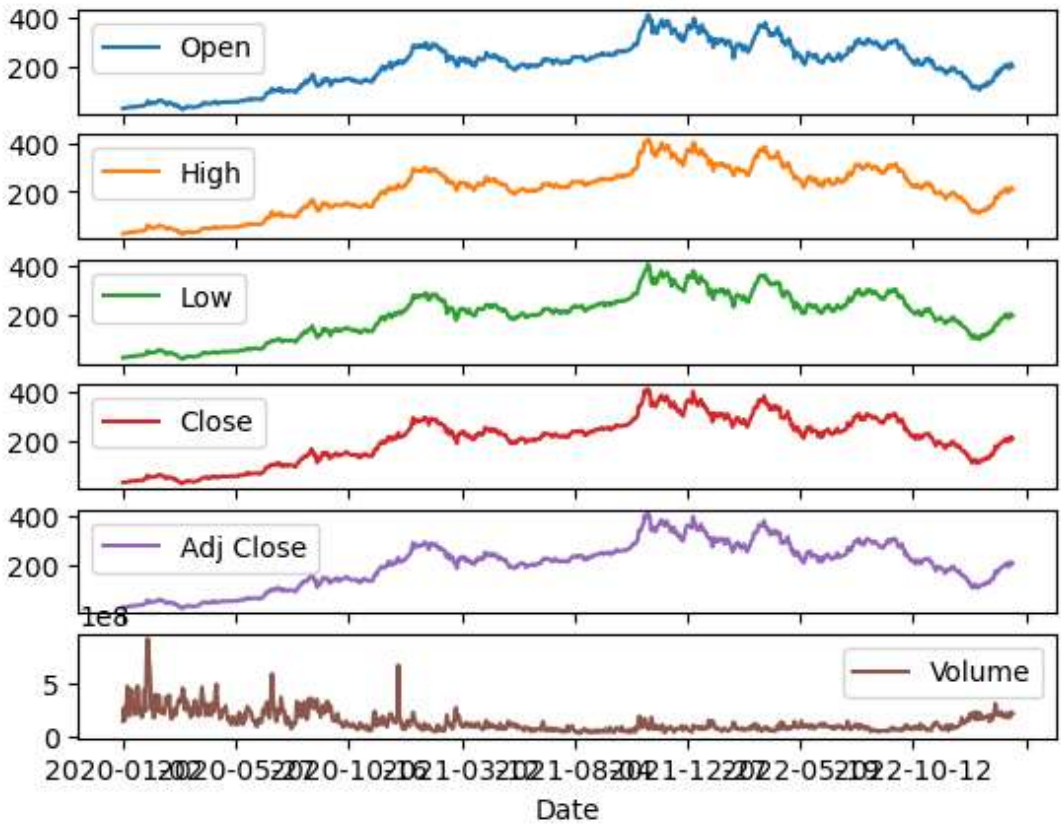


Рис. 10. Візуалізація часових рядів для акцій компанії Tesla

Проведене дослідження кореляції змінних підтверджує, що змінні між собою сильно корелюють, це показано у вигляді матриці кореляцій на рис. 11.

	Open	High	Low	Close	Volume	Adj Close
Open	1.000000	0.999623	0.999605	0.999233	0.407515	0.999233
High	0.999623	1.000000	0.999521	0.999691	0.416466	0.999691
Low	0.999605	0.999521	1.000000	0.999656	0.397615	0.999656
Close	0.999233	0.999691	0.999656	1.000000	0.406907	1.000000
Volume	0.407515	0.416466	0.397615	0.406907	1.000000	0.406907
Adj Close	0.999233	0.999691	0.999656	1.000000	0.406907	1.000000

Рис. 11. Матриця кореляцій

Для побудови моделей і прогнозування обираємо часовий ряд, в даному прикладі – це ціни на акції компанії Tesla на момент закриття. Візуалізацію графіку цін на акції представлено на рис. 12.

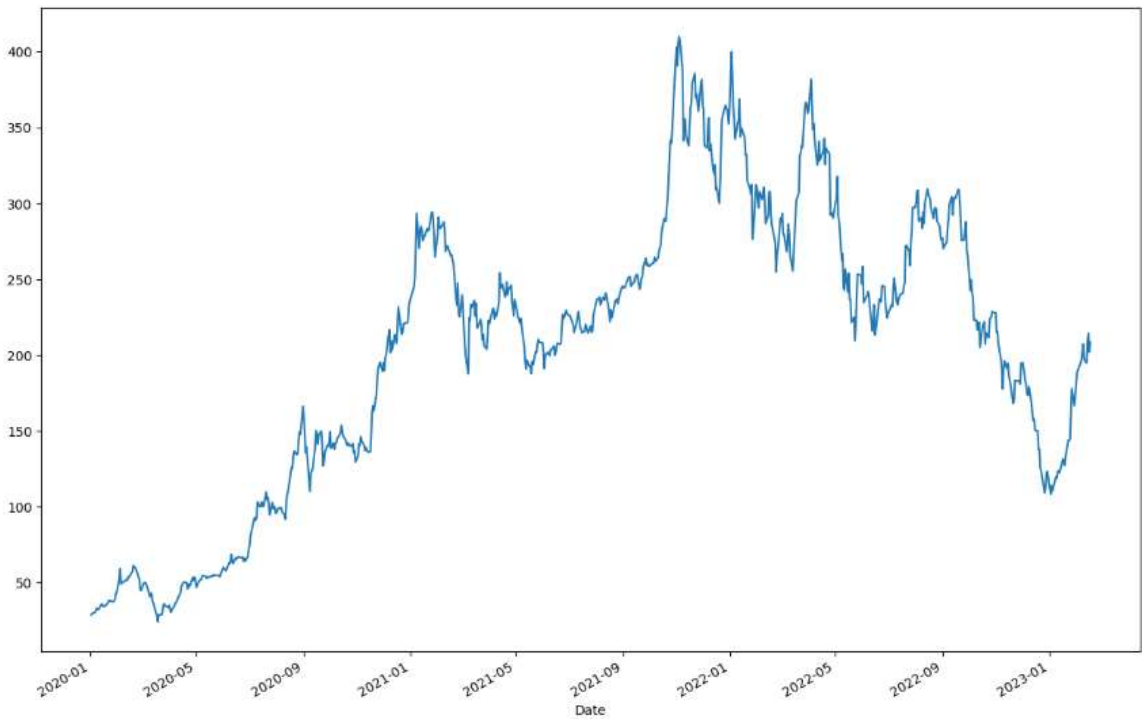


Рис. 12. Часовий ряд цін на акції компанії на момент закриття

Перевіримо ряд цін на стаціонарність, для цього використаємо тести Дікі-Фулера та KPSS. $ADFStatistics: 4.99404289799802$, $p-value=1.0$,

$KPSSStatistic=6.368771256451674$, $p-value=0.01$.

Обидва тести дали схожі результати, тобто ряд є нестаціонарним.

Покажемо способи розв'язання типових завдань фінансового аналітика, які зазвичай виконуються при більш глибокому дослідженні фінансового ряду, базуючись на попередньо отриманих результатах візуального і статистичного аналізу даних.

Завдання 1. Побудова регресійних моделей AP і АРКС

Для визначення порядку моделей обчислимо АКФ і ЧАКФ:

Date: 02/28/23 Time: 10:56
Sample: 1 789
Included observations: 789

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.993	0.993	781.50	0.000
		2	0.987	0.024	1554.1	0.000
		3	0.981	-0.009	2317.7	0.000
		4	0.974	-0.005	3072.4	0.000
		5	0.967	-0.035	3817.4	0.000
		6	0.961	0.003	4553.1	0.000
		7	0.954	0.003	5279.5	0.000
		8	0.947	-0.065	5995.5	0.000
		9	0.940	0.040	6702.1	0.000
		10	0.932	-0.067	7398.0	0.000
		11	0.925	0.028	8084.0	0.000
		12	0.918	0.008	8760.2	0.000
		13	0.910	-0.022	9426.5	0.000
		14	0.903	-0.006	10083.	0.000
		15	0.896	0.033	10730.	0.000
		16	0.889	0.003	11368.	0.000
		17	0.882	0.011	11997.	0.000
		18	0.876	0.012	12617.	0.000
		19	0.868	-0.064	13228.	0.000
		20	0.861	0.002	13830	0.000

Dependent Variable: CLOSE
Method: Least Squares
Date: 02/28/23 Time: 10:58

Sample (adjusted): 2 784

Included observations: 783 after adjustments

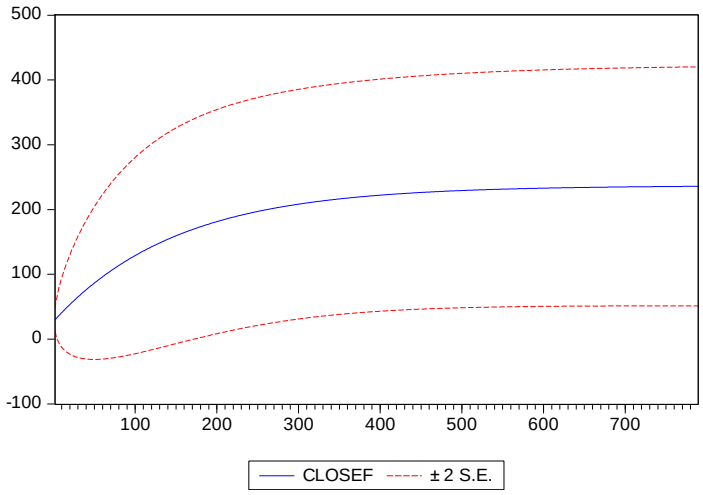
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	236.8891	50.91043	4.653056	0.0000
AR(1)	0.993385	0.003358	295.8066	0.0000

R-squared	0.991153	Mean dependent var	204.6287
Adjusted R-squared	0.991142	S.D. dependent var	94.73100
S.E. of regression	8.915753	Akaike info criterion	7.216068
Sum squared resid	62082.20	Schwarz criterion	7.227978
Log likelihood	-2823.090	Hannan-Quinn criter.	7.220648
F-statistic	87501.57	Durbin-Watson stat	2.068476
Prob(F-statistic)	0.000000		

Inverted AR Roots	.99

Рис. 13. Визначення порядку і характеристик моделі в Eviews



Forecast: CLOSEF	
Actual: CLOSE	
Forecast sample: 1 789	
Adjusted sample: 2 789	
Included observations: 788	
Root Mean Squared Error	61.78275
Mean Absolute Error	48.96496
Mean Abs. Percent Error	34.86613
Theil Inequality Coefficient	0.143966
Bias Proportion	0.013628
Variance Proportion	0.502184
Covariance Proportion	0.484188

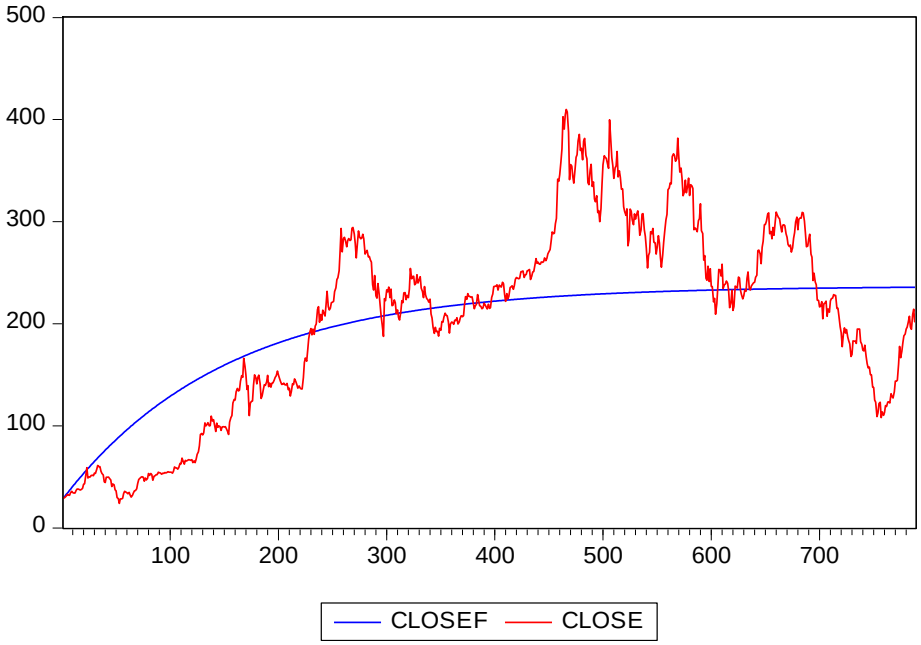


Рис. 14. Візуалізація побудованої моделі на графіку

Для побудови моделей обчислимо АКФ і ЧАКФ залишків моделі AP:

Date: 03/14/23 Time: 13:07
Sample: 1 789
Included observations: 788

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|*******	.|*******	1	0.989	0.989	773.26	0.000
.|*******	.| |	2	0.978	0.028	1531.1	0.000
.|*******	.| |	3	0.968	-0.007	2273.6	0.000
.|*******	.| |	4	0.957	-0.013	3000.7	0.000
.|*******	.| |	5	0.945	-0.065	3710.4	0.000
.|*******	.| |	6	0.933	0.006	4403.5	0.000
.|*******	.| |	7	0.922	0.012	5080.7	0.000
.|*******	*| |	8	0.908	-0.106	5739.0	0.000
.|******|	.| |	9	0.896	0.053	6380.4	0.000
.|******|	*| |	10	0.882	-0.099	7002.4	0.000
.|******|	.| |	11	0.869	0.038	7606.7	0.000
.|******|	.| |	12	0.856	0.017	8194.0	0.000
.|******|	.| |	13	0.842	-0.048	8763.3	0.000
.|******|	.| |	14	0.828	-0.017	9314.6	0.000
.|******|	.| |	15	0.815	0.044	9849.4	0.000
.|******|	.| |	16	0.803	0.006	10369.	0.000
.|******|	.| |	17	0.790	0.019	10873.	0.000
.|******|	.| |	18	0.778	-0.001	11363.	0.000
.|******|	*| |	19	0.764	-0.102	11835.	0.000
.|***** |	.| |	20	0.750	-0.007	12292.	0.000
.|***** |	.| |	21	0.737	0.016	12732.	0.000
.|***** |	.| |	22	0.723	-0.018	13157.	0.000
.|***** |	.| |	23	0.711	0.045	13568.	0.000

.|***** |		.|	|	24	0.699	0.033	13966.	0.000

Побудуємо АРКС(1,1) і АРКС(1,8)

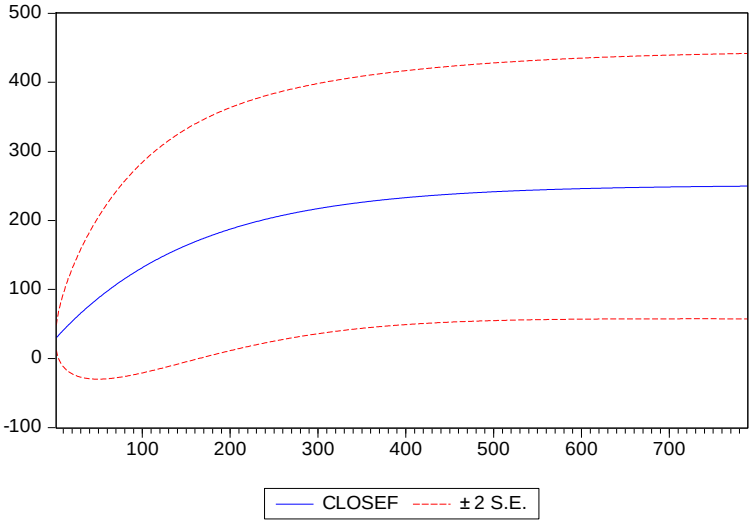
АРКС(1,1):

Dependent Variable: CLOSE
Method: Least Squares
Date: 03/14/23 Time: 17:07
Sample (adjusted): 2 700
Included observations: 699 after adjustments
Convergence achieved after 6 iterations
MA Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	251.2038	57.68965	4.354400	0.0000
AR(1)	0.993750	0.003353	296.3702	0.0000
MA(1)	-0.034569	0.038056	-0.908376	0.3640

R-squared	0.991639	Mean dependent var	208.4240
Adjusted R-squared	0.991615	S.D. dependent var	98.81973
S.E. of regression	9.048885	Akaike info criterion	7.247443
Sum squared resid	56990.09	Schwarz criterion	7.266969
Log likelihood	-2529.981	Hannan-Quinn criter.	7.254991
F-statistic	41273.97	Durbin-Watson stat	2.000004
Prob(F-statistic)	0.000000		

Inverted AR Roots	.99
Inverted MA Roots	.03



Forecast: CLOSEF
Actual: CLOSE
Forecast sample: 1 789
Adjusted sample: 2 789
Included observations: 788

Root Mean Squared Error	59.67563
Mean Absolute Error	47.53785
Mean Abs. Percent Error	35.56147
Theil Inequality Coefficient	0.135907
Bias Proportion	0.001112
Variance Proportion	0.441873
Covariance Proportion	0.557015

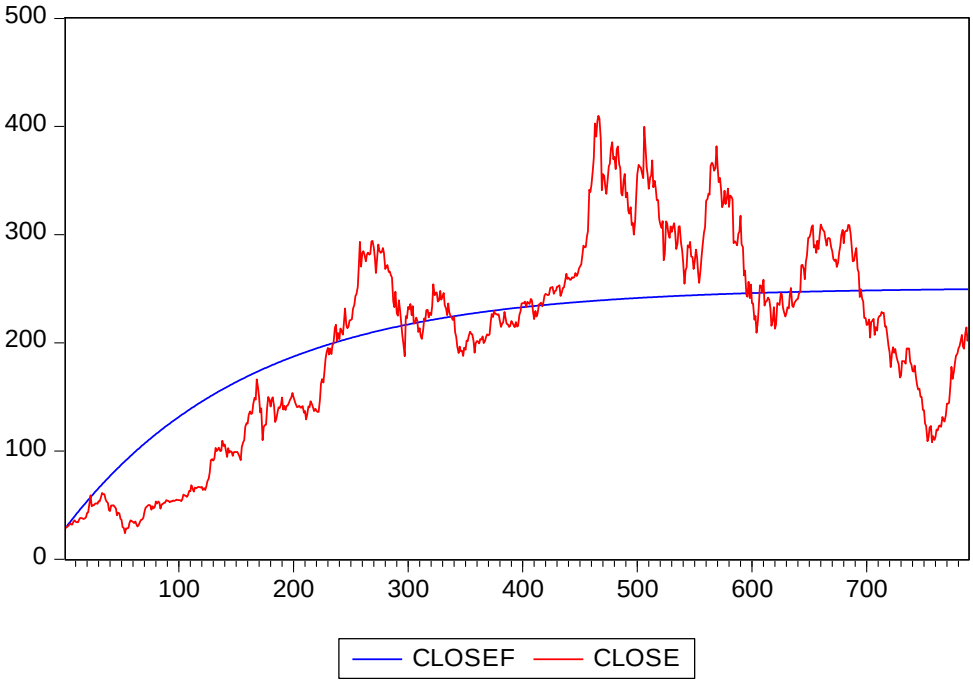


Рис. 15. Скріншоти побудови моделі АРКС (1,1)

АРКС(1,8)

Dependent Variable: CLOSE
Method: Least Squares
Date: 03/14/23 Time: 13:08
Sample (adjusted): 2 780
Included observations: 779 after adjustments
Convergence achieved after 11 iterations
MA Backcast: -6 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	234.6448	50.33618	4.661554	0.0000
AR(1)	0.992886	0.003700	268.3238	0.0000
MA(1)	-0.015849	0.036208	-0.437719	0.6617
MA(2)	-0.002254	0.035882	-0.062808	0.9499
MA(3)	0.005515	0.035895	0.153651	0.8779
MA(4)	0.066761	0.035903	1.859509	0.0633
MA(5)	-0.029031	0.035899	-0.808683	0.4189
MA(6)	-0.017462	0.035925	-0.486064	0.6271
MA(7)	0.124661	0.036039	3.459031	0.0006
MA(8)	-0.062184	0.036401	-1.708297	0.0880

R-squared	0.991362	Mean dependent var	204.6495
Adjusted R-squared	0.991261	S.D. dependent var	94.97327
S.E. of regression	8.878320	Akaike info criterion	7.217856
Sum squared resid	60616.09	Schwarz criterion	7.277650
Log likelihood	-2801.355	Hannan-Quinn criter.	7.240855
F-statistic	9806.423	Durbin-Watson stat	2.008891
Prob(F-statistic)	0.000000		

Inverted AR Roots	.99			
Inverted MA Roots	.55-.38i	.55+.38i	.50	.13+.72i
	.13-.72i	-.54+.61i	-.54-.61i	-.78

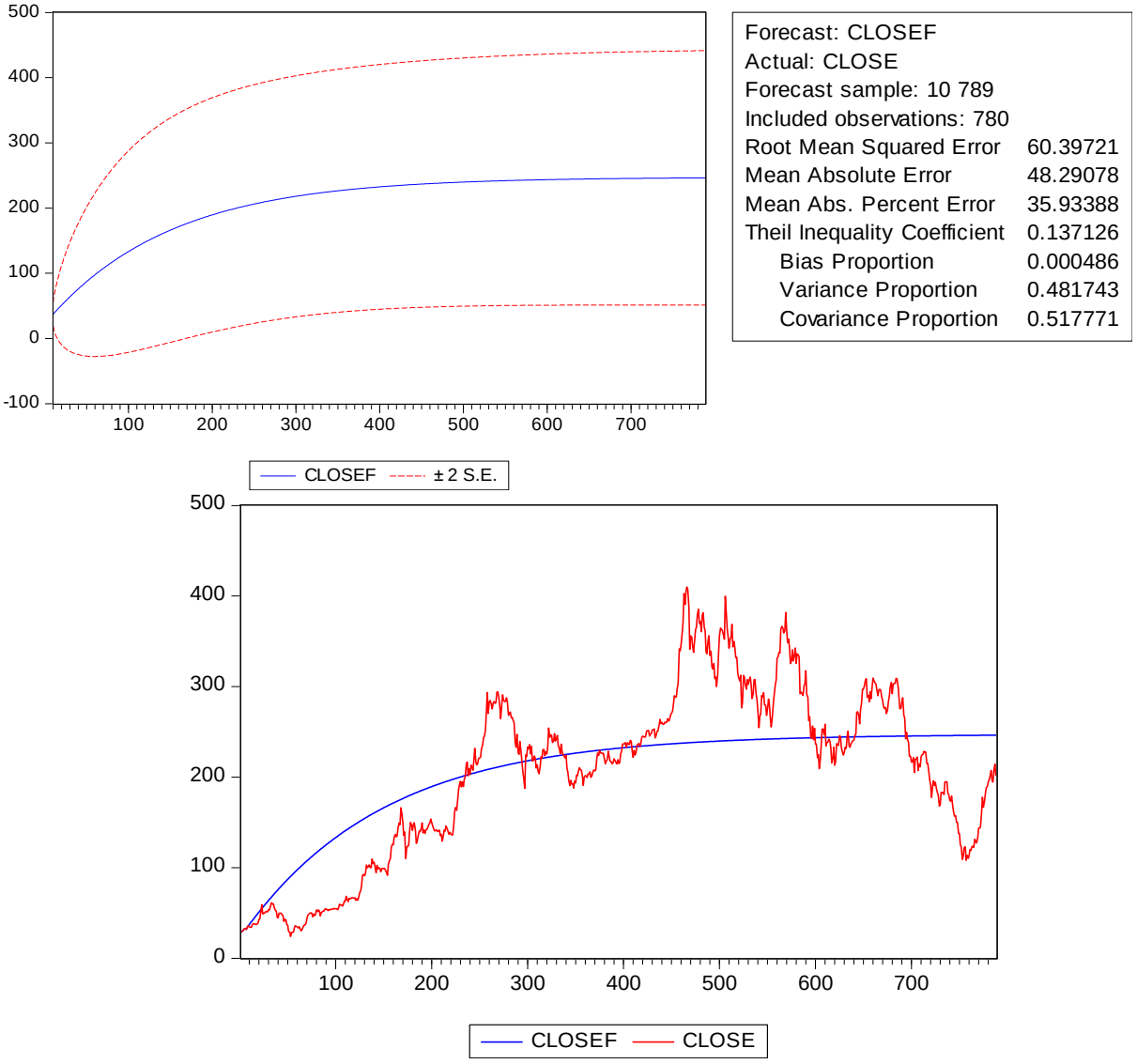


Рис. 16. Скріншоти етапів побудови моделі АРКС(1,8)

Оскільки часовий ряд не стаціонарний, визначимо порядок тренду шляхом видалення тренду з часового ряду і перевірки нового часового ряду на стаціонарність. Після видалення тренду першого порядку отримаємо за статистиками Дікі-Фуллера і KPSS, що ряд все ще залишився нестаціонарним.

```
ADF Statistic: -2.083275644398734
p-value: 0.2512925056636728
KPSS Statistic: 2.4209447857890125
p-value: 0.01
Time series is likely non-stationary
```

Після видалення тренду другого порядку, отримуємо за статистиками Дікі-Фуллера і KPSS, що ряд є стаціонарним.

```
ADF Statistic: -11.837219334937716
p-value: 7.739597960709718e-22
KPSS Statistic: 0.03899775516638495
p-value: 0.1
Time series is likely stationary
```

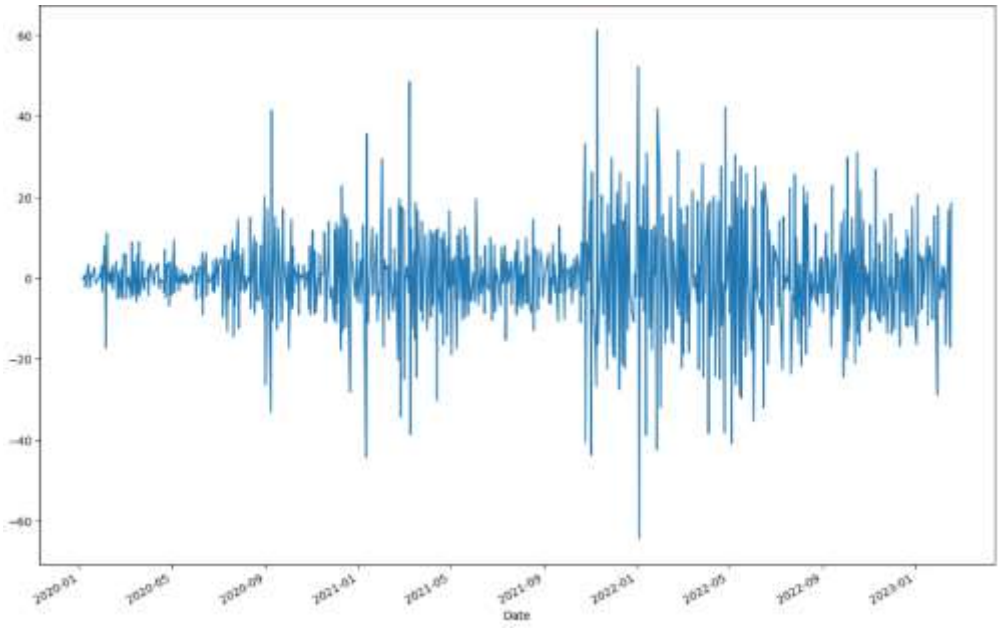


Рис. 17. Візуалізація ряду

Далі будуємо моделі авторегресії з інтегрованим трендом 2-го порядку.

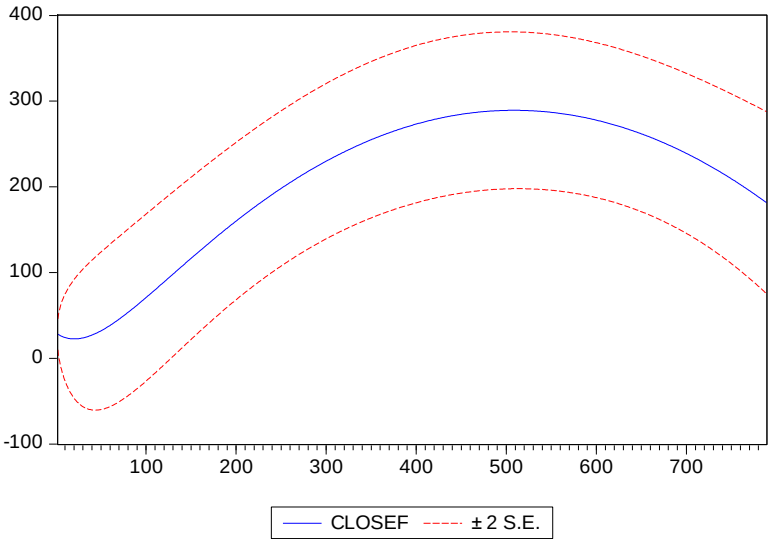
Dependent Variable: CLOSE
Method: Least Squares
Date: 03/14/23 Time: 13:20
Sample (adjusted): 2 789
Included observations: 788 after adjustments

Convergence achieved after 10 iterations
MA Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-63.35514	53.58248	-1.182385	0.2374
@TREND	1.390083	0.272258	5.105750	0.0000
@TREND^2	-0.001370	0.000300	-4.559823	0.0000
AR(1)	0.977786	0.007494	130.4799	0.0000
MA(1)	-0.028605	0.036552	-0.782595	0.4341

R-squared	0.991182	Mean dependent var	204.6355
Adjusted R-squared	0.991137	S.D. dependent var	94.43117
S.E. of regression	8.890031	Akaike info criterion	7.214063
Sum squared resid	61882.56	Schwarz criterion	7.243692
Log likelihood	-2837.341	Hannan-Quinn criter.	7.225453
F-statistic	22003.53	Durbin-Watson stat	2.000231
Prob(F-statistic)	0.000000		

Inverted AR Roots	.98
Inverted MA Roots	.03



Forecast: CLOSEF	
Actual: CLOSE	
Forecast sample: 1 789	
Adjusted sample: 2 789	
Included observations: 788	
Root Mean Squared Error	39.71002
Mean Absolute Error	31.68039
Mean Abs. Percent Error	17.65007
Theil Inequality Coefficient	0.088657
Bias Proportion	0.001397
Variance Proportion	0.068904
Covariance Proportion	0.929699

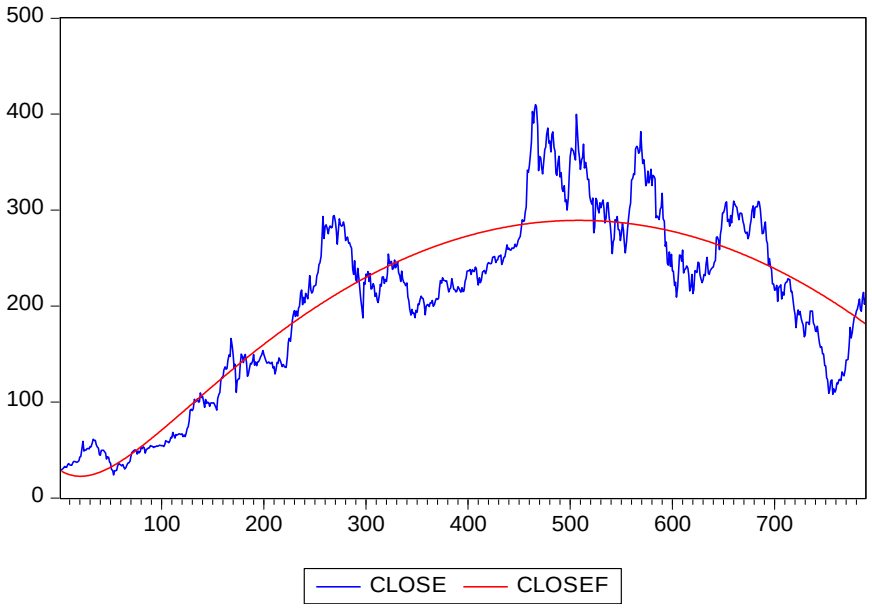


Рис. 18. Скріншоти етапів побудови авторегресії з інтегрованим трендом 2-го порядку ARIMA(1,2,1)

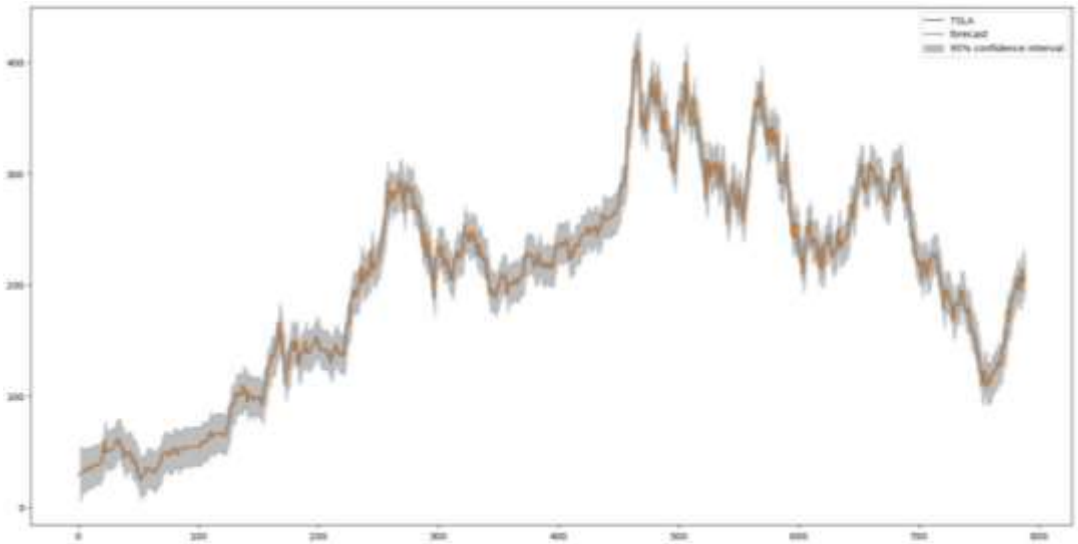


Рис. 19. Графік реальних даних і прогнозу моделі ARIMA (1,2,8)

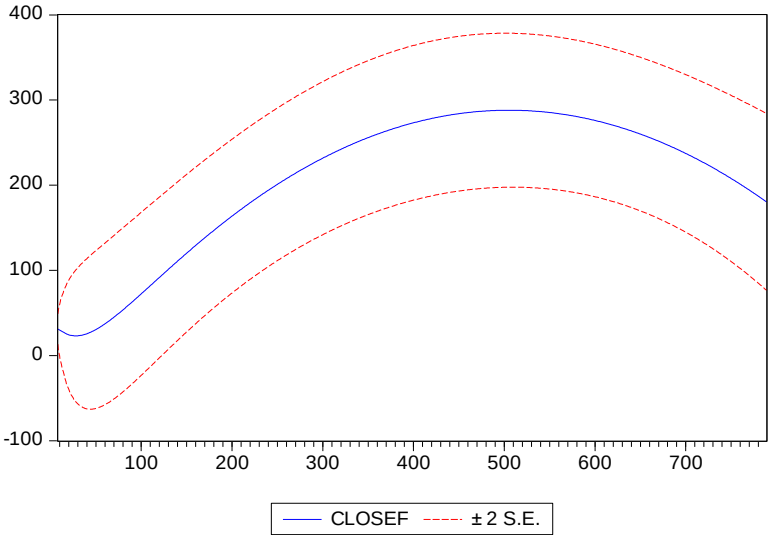
ARIMA(1,2,8)

Dependent Variable: CLOSE
Method: Least Squares
Date: 03/14/23 Time: 13:21
Sample (adjusted): 2 789
Included observations: 788 after adjustments
Convergence achieved after 12 iterations
MA Backcast: -5 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-51.85011	42.75854	-1.212626	0.2256
@TREND	1.348454	0.227576	5.925291	0.0000
@TREND^2	-0.001337	0.000259	-5.158741	0.0000
AR(1)	0.968369	0.010311	93.91605	0.0000
MA(1)	-0.005575	0.036691	-0.151945	0.8793
MA(2)	0.002575	0.036644	0.070262	0.9440
MA(3)	0.020464	0.036634	0.558611	0.5766
MA(4)	0.085200	0.036358	2.343340	0.0194
MA(5)	-0.026127	0.036579	-0.714259	0.4753
MA(6)	-0.004023	0.036517	-0.110158	0.9123
MA(7)	0.137240	0.036476	3.762429	0.0002

R-squared	0.991351	Mean dependent var	204.6355
Adjusted R-squared	0.991239	S.D. dependent var	94.43117
S.E. of regression	8.838649	Akaike info criterion	7.210006
Sum squared resid	60700.58	Schwarz criterion	7.275190
Log likelihood	-2829.742	Hannan-Quinn criter.	7.235065
F-statistic	8905.554	Durbin-Watson stat	2.013390
Prob(F-statistic)	0.000000		

Inverted AR Roots	.97			
Inverted MA Roots	.66+.35i	.66-.35i	.20+.72i	.20-.72i
	-.50+.59i	-.50-.59i	-.72	



Forecast: CLOSEF	
Actual: CLOSE	
Forecast sample: 8 789	
Included observations: 782	
Root Mean Squared Error	39.99773
Mean Absolute Error	32.11487
Mean Abs. Percent Error	17.79146
Theil Inequality Coefficient	0.088989
Bias Proportion	0.001946
Variance Proportion	0.084698
Covariance Proportion	0.913356

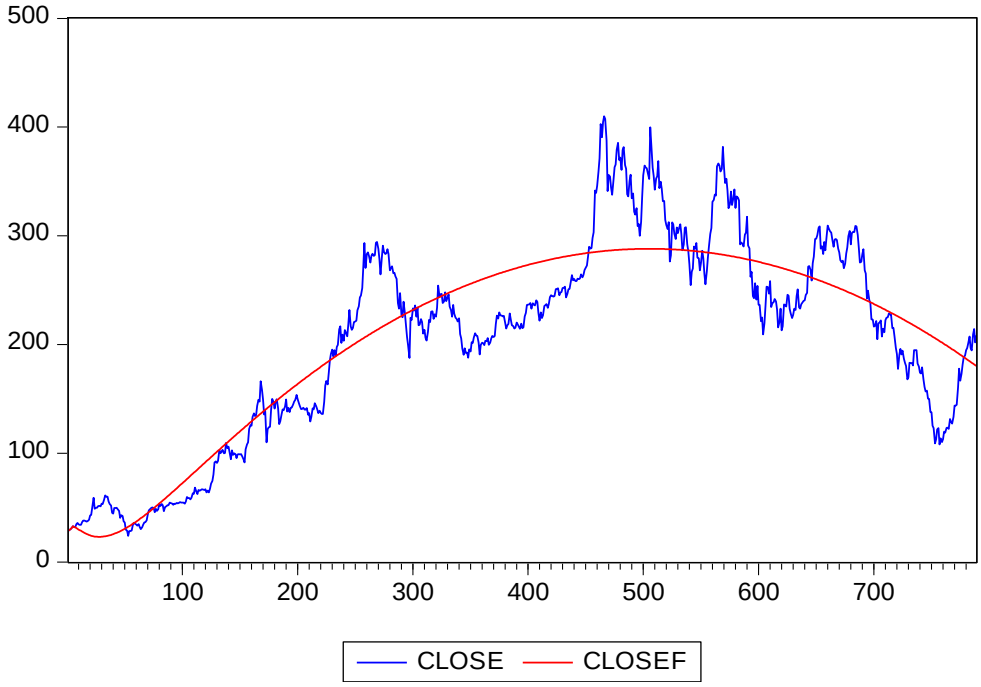


Рис. 20. Скріншоти етапів побудови авторегресії з інтегрованим трендом 2-го порядку ARIMA(1,2,8)

Порівняємо побудовані моделі за основними статистичними характеристиками і оберемо кращу модель (таблиці 6–7).

Таблиця 6

Таблиця статистичних характеристик моделей

Тип моделі	**R2**	**Sum squared resid**	**IKA**	**DW**
AR(1)	0,991	62082,2	7,246	2,069
ARMA(1,1)	0,991	60342,57	7,215	1,998
ARMA(1,8)	0,991	60616.09	7.217	2.009
ARIMA(1,2,1)	0,991	61882.56	7,214	2,0002
ARIMA(1,2,8)	0,991	60700.58	7,21	2,013

Таблиця 7

Таблиця характеристик якості прогнозу моделей

Тип моделі	**RMSE**	**MAE**	**MAPE**	**Theil's U**
AR(1)	0,991	62082,2	7,246	2,069
ARMA(1,1)	60,14	48,02	36,31	0,13
ARMA(1,8)	60,39	48,29	35,93	0,14
ARIMA(1,2,1)	39,71	31,68	17,65	0,089
ARIMA(1,2,8)	39,99	32,11	17,79	0,089

За статистичними характеристиками моделі відрізняються зовсім несуттєво, проте за характеристиками прогнозу видно, що моделі з трендом 2-го і 3-го порядку є набагато точнішими, що пояснюється нестаціонарністю часового ряду.

Наступним завданням є дослідження волатильності фінансового ряду. Для цього виконаємо побудову гетероскедастичних моделей [11].

Завдання 2. Побудова гетероскедастичних моделей

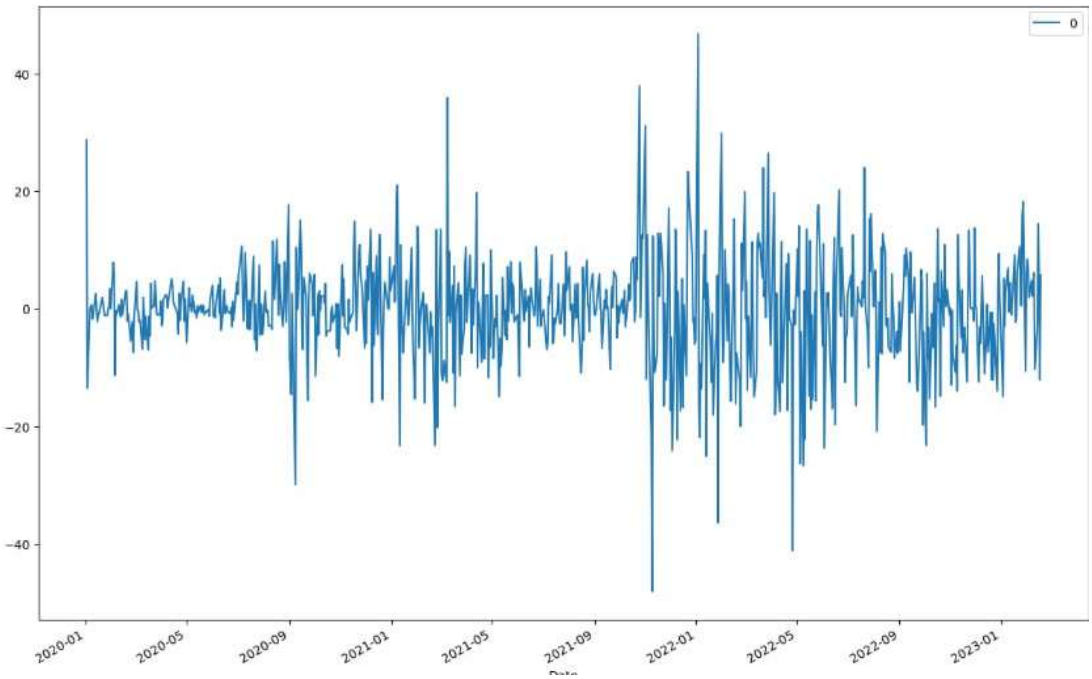


Рис. 21. Гістограма залишків моделі ARIMA(1,2,1)

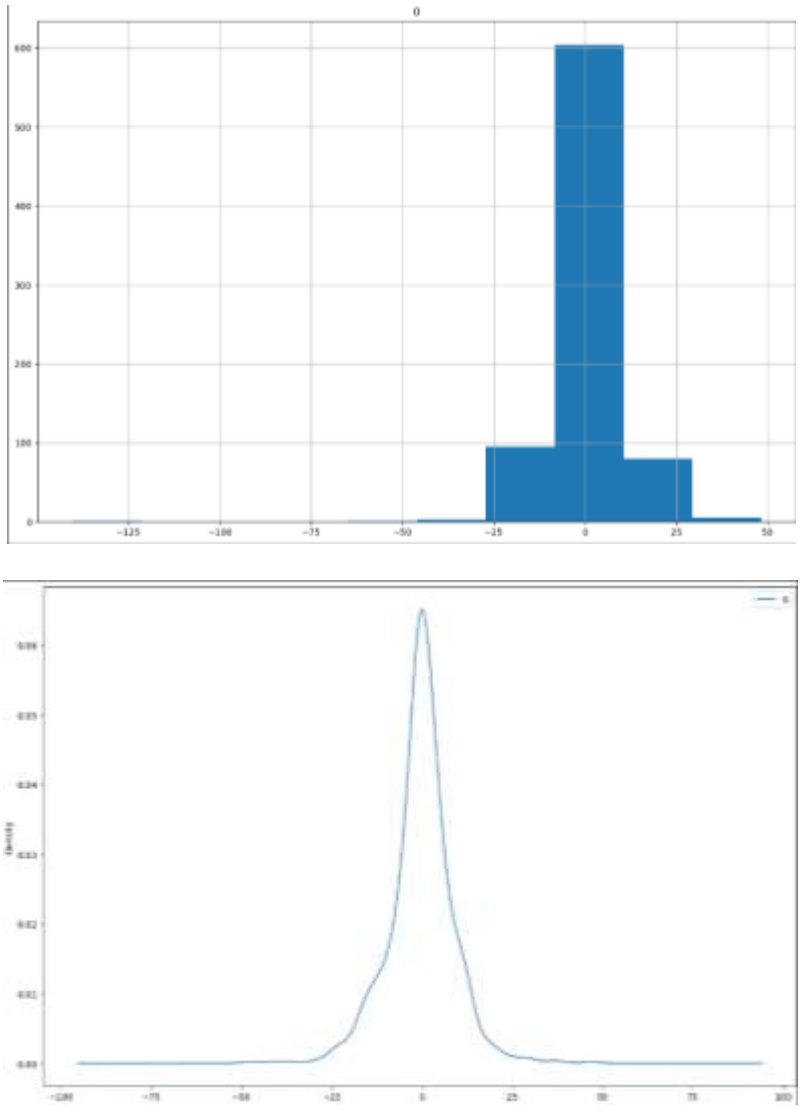


Рис. 22. Залишки моделі ARIMA(1,2,1)

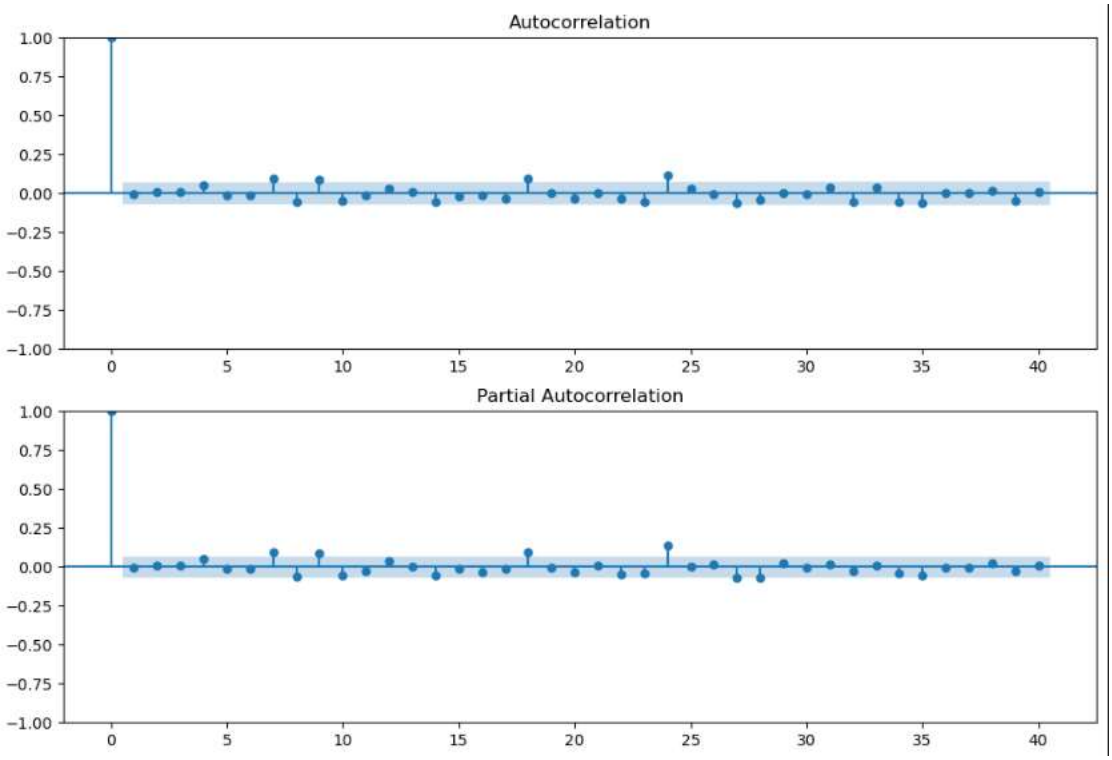


Рис. 23. АКФ і ЧАКФ для залишків

Побудуємо АКФ і ЧАКФ для залишків в квадраті

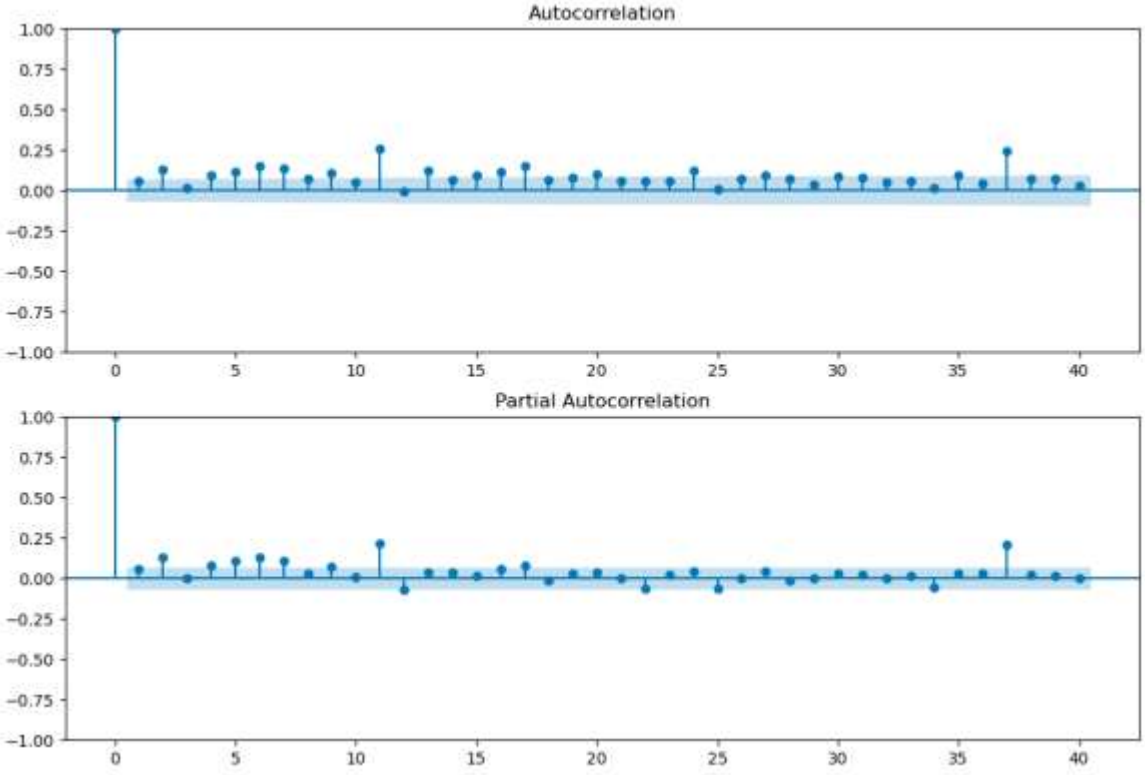


Рис. 24. АКФ і ЧАКФ для залишків в квадраті

Модель ARCH(1)

```
                     Constant Mean - ARCH Model Results
==============================================================================
Dep. Variable:                    TSLA   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                       ARCH   Log-Likelihood:               -2308.57
Distribution:                  Normal   AIC:                           4623.14
Method:            Maximum Likelihood   BIC:                           4637.15
                                        No. Observations:                  788
Date:                Tue, Mar 28 2023   Df Residuals:                      787
Time:                        16:28:42   Df Model:                            1
                                Mean Model
==============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
```

Модель ARCH(2)

```
                     Constant Mean - ARCH Model Results
==============================================================================
Dep. Variable:                    TSLA   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                       ARCH   Log-Likelihood:               -2287.40
Distribution:                  Normal   AIC:                           4582.79
Method:            Maximum Likelihood   BIC:                           4601.47
                                        No. Observations:                  788
Date:                Tue, Mar 28 2023   Df Residuals:                      787
Time:                        16:29:59   Df Model:                            1
...
```

Модель GARCH(1,1)

```
                    Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                    TSLA   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                      GARCH   Log-Likelihood:               -2264.83
Distribution:                  Normal   AIC:                           4537.66
Method:            Maximum Likelihood   BIC:                           4556.33
                                        No. Observations:                  788
Date:                Tue, Mar 14 2023   Df Residuals:                      787
Time:                        16:40:11   Df Model:                            1
                                Mean Model
...
beta[1]       0.8978  4.058e-02     22.122 1.952e-108   [  0.818,  0.977]
==============================================================================

Covariance estimator: robust
```

Модель EGARCH(1,1)

```
                     Constant Mean - EGARCH Model Results
==============================================================================
Dep. Variable:                    TSLA   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                      EGARCH   Log-Likelihood:               -2265.77
Distribution:                   Normal   AIC:                           4539.54
Method:            Maximum Likelihood   BIC:                           4558.22
                                        No. Observations:                  788
Date:                Tue, Mar 14 2023   Df Residuals:                      787
...
beta[1]        0.9612  2.911e-02     33.022 3.964e-239    [  0.904,  1.018]
==============================================================================

Covariance estimator: robust
```

Модель FIGARCH(1,1)

```
                    Constant Mean - FIGARCH Model Results
==============================================================================
Dep. Variable:                    TSLA   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                     FIGARCH   Log-Likelihood:               -2262.79
Distribution:                   Normal   AIC:                           4533.58
Method:            Maximum Likelihood   BIC:                           4552.25
                                        No. Observations:                  788
Date:                Tue, Mar 14 2023   Df Residuals:                      787
Time:                        16:41:06   Df Model:                            1
                              Mean Model
...
beta          0.3815      0.134      2.852  4.348e-03 [  0.119,  0.644]
==============================================================================

Covariance estimator: robust
```

Таблиця 8

Таблиця статистичних характеристик моделей

Тип моделі	**AIC**	**BIC**
ARCH(1)	4623.14	4637.15
ARCH(2)	4582.79	4601.47
GARCH(1,1)	4537.66	4556.33
EGARCH(1,1)	4539.54	4558.22
EGARCH(4,1)	4532.18	4564.86
FIGARCH(0,1)	4558.92	4535.57
FIGARCH(1,1)	4533.58	4552.25

Завдання 3. Прогнозування значень фінансового ряду на 5 кроків вперед

Метою аналізу і побудови моделей, що найкраще відповідають процесу на фінансовому ринку, є прогнозування значення фінансового ряду у наступні періоди. Слід зазначити, що фінансовий ринок зазнає багатьох впливів і збурень протягом навіть дня торгів, тому нам здається доречним здійснити прогнозування фінансового ряду на 5 кроків вперед (на тиждень) за допомогою побудованих вище моделей.

- AR(1)

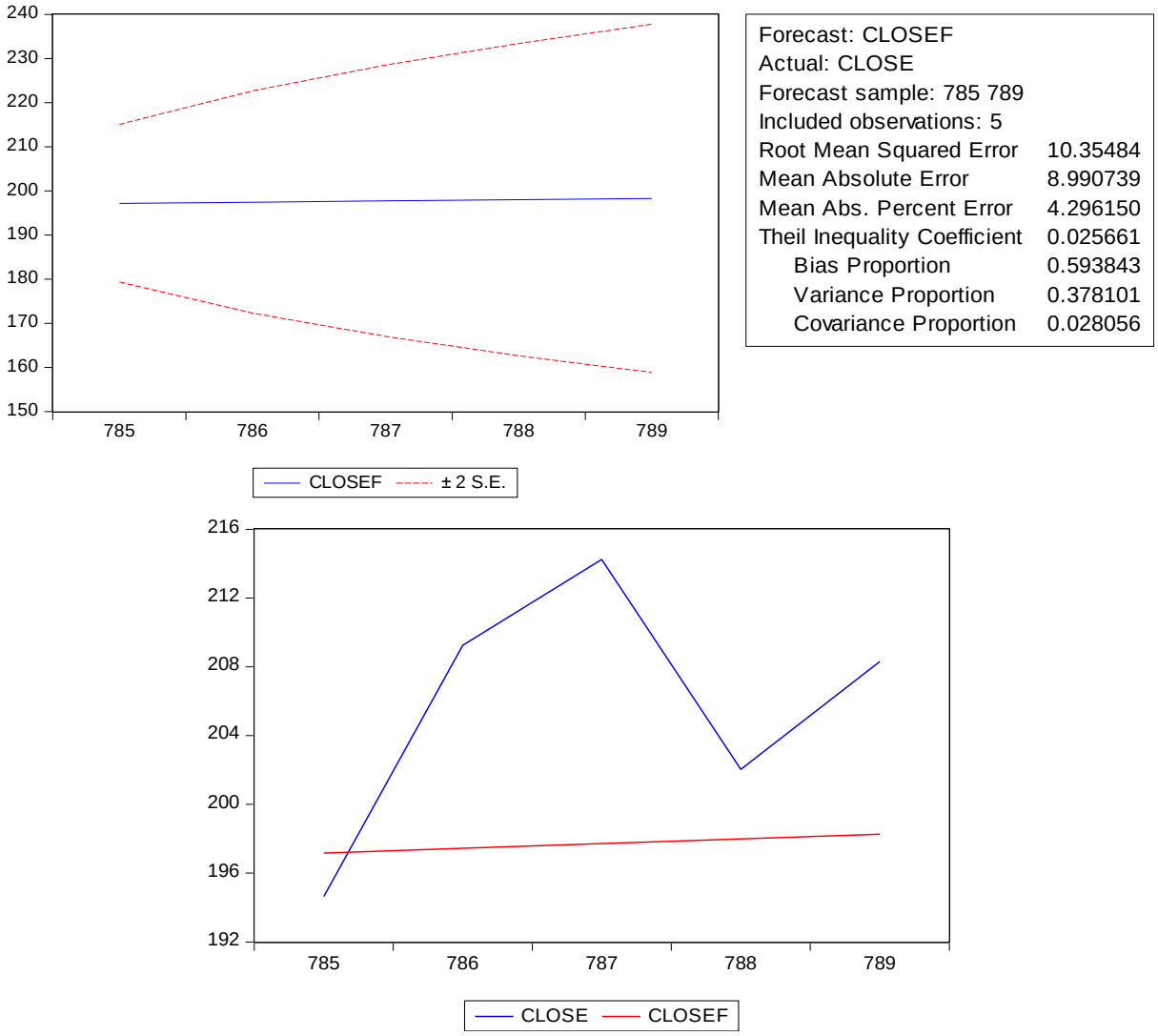
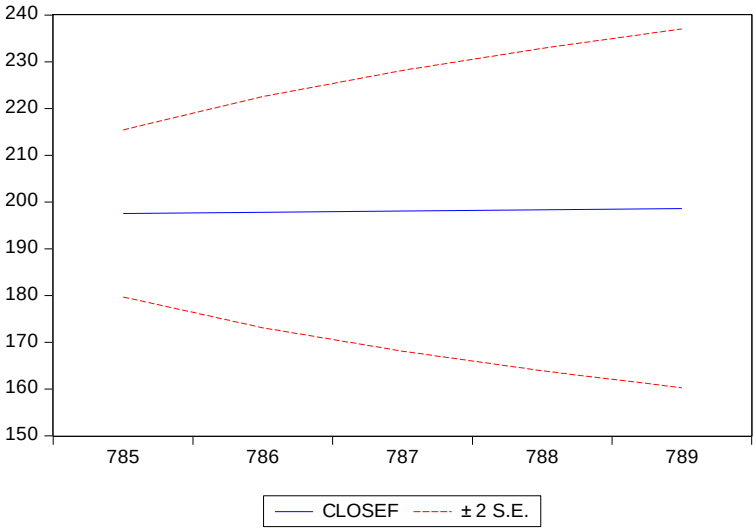


Рис. 25. Візуалізація ціни закриття і прогнозу моделі авторегресії AP(1)

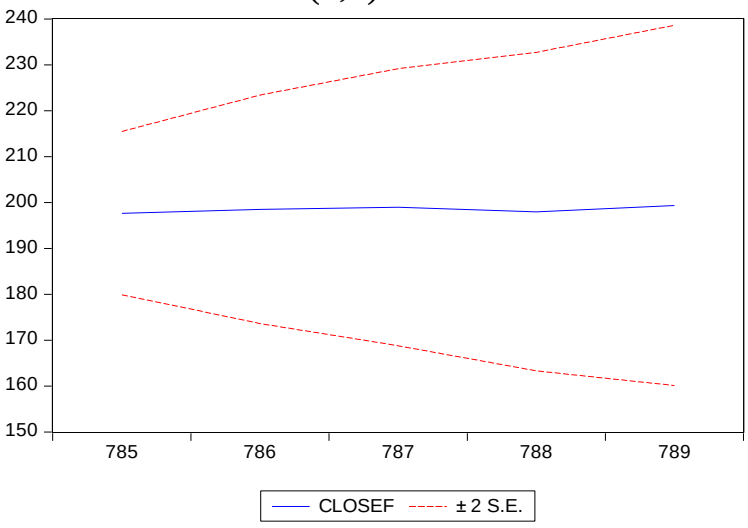
- ARMA(1,1)



Forecast: CLOSEF	
Actual: CLOSE	
Forecast sample: 785 789	
Included observations: 5	
Root Mean Squared Error	10.08256
Mean Absolute Error	8.779215
Mean Abs. Percent Error	4.197146
Theil Inequality Coefficient	0.024964
Bias Proportion	0.571283
Variance Proportion	0.399624
Covariance Proportion	0.029092

Рис. 26. Візуалізація ціни закриття і прогнозу моделі авторегресії АРКС(1,1)

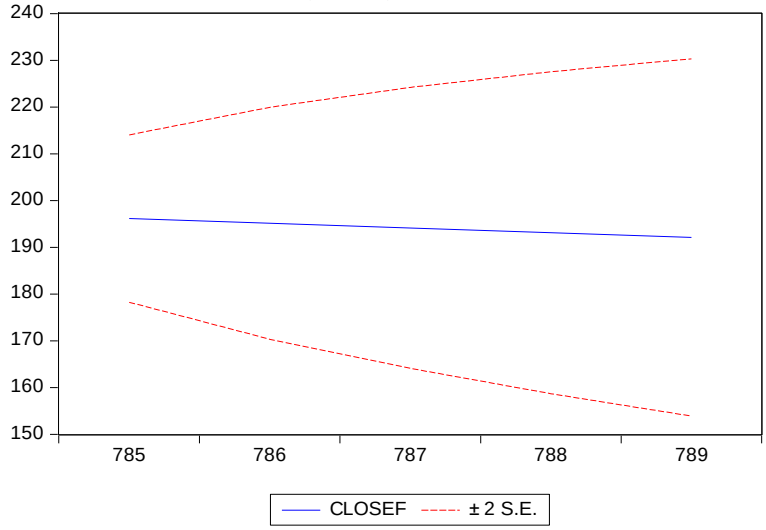
- ARMA(1,8)



Forecast: CLOSEF	
Actual: CLOSE	
Forecast sample: 785 789	
Included observations: 5	
Root Mean Squared Error	9.546937
Mean Absolute Error	8.420039
Mean Abs. Percent Error	4.029205
Theil Inequality Coefficient	0.023614
Bias Proportion	0.571256
Variance Proportion	0.413347
Covariance Proportion	0.015397

Рис. 27. Візуалізація ціни закриття і прогнозу моделі авторегресії АРКС(1,8)

- ARIMA(1,2,1)



Forecast: CLOSEF	
Actual: CLOSE	
Forecast sample: 785 789	
Included observations: 5	
Root Mean Squared Error	13.78982
Mean Absolute Error	12.18732
Mean Abs. Percent Error	5.827688
Theil Inequality Coefficient	0.034482
Bias Proportion	0.706390
Variance Proportion	0.148958
Covariance Proportion	0.144652

Рис. 28. Візуалізація ціни закриття і прогнозу моделі авторегресії АРКС(1,8)

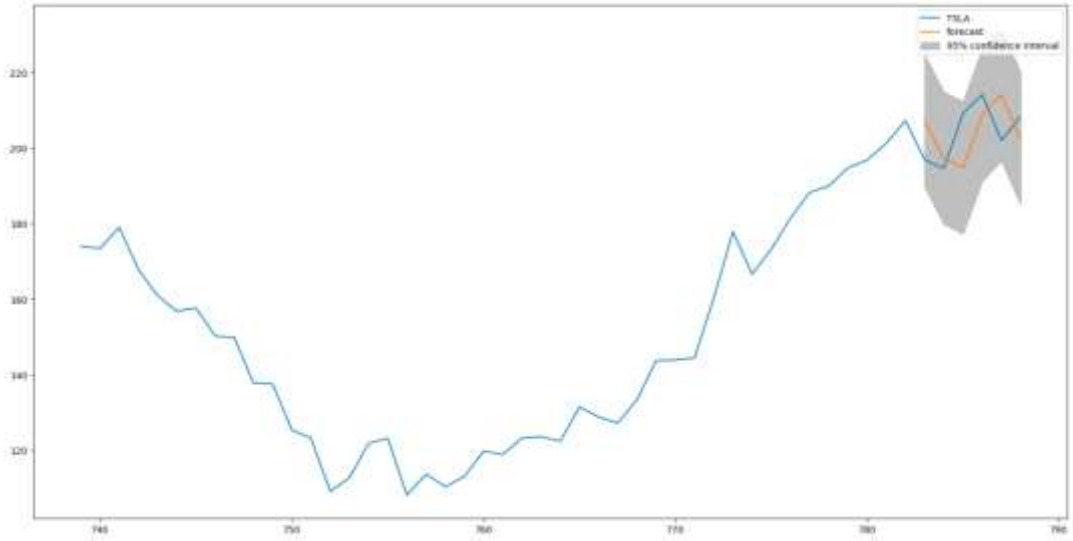


Рис. 29. Візуалізація ціни закриття і прогнозу моделі авторегресії АРІКС(1,2,1)

Таблиця 9

Тип моделі	**RMSE**	**MAE**	**MAPE**	**Theil's U**
AR(1)	10.35	8.99	4.3	0.026
ARMA(1,1)	10.08	8.78	4.2	0.024
ARMA(1,8)	9.55	8.42	4.03	0.024
ARIMA(1,2,1)	13.79	12.19	5.83	0.034

Таблиця 10

Прогнозування волатильності на крок вперед

Тип моделі	Прогноз	Абсолютна різниця
ARCH(1)	4.4335	1.30324277
ARCH(2)	4.8979	0.83886498
GARCH(1,1)	4.6164	1.12032047
EGARCH(4,1)	5.1061	0.630647
FIGARCH(1,1)	4.66249	1.07429179
FIGARCH(0,1)	4.66257	1.0742075

Завдання 4. Прогнозування діапазону значень фінансового ряду для інвестора

Оскільки волатильність фінансового ряду показує коливання значення фінансового ряду відносно математичного сподівання, тобто відповідає коливанню ціни на акції протягом дня відносно середньої денної ціни, то нам вважається необхідним надати інвестору прогноз у вигляді «червоних» ліній, що будуть відповідати прогнозу ціни (High, Low) протягом дня. Для цього комбінуємо моделі, які використовувались для прогнозування значення ціни і моделі для прогнозування волатильності (таблиці 11-12).

Таблиця 11

Результати прогнозування на 5 кроків вперед з урахуванням волатильності ARIMA(1,2,1) + GARCH(1,1)

Номер кроку	Прогнозований інтервал	Справжнє значення
1	[198.8423, 208.0751]	208.309998
2	[199.9178, 209.1506]	197.369995
3	[200.3096, 209.5424]	200.860001
4	[202.2282, 211.461]	202.070007
5	[200.1291, 209.3619]	196.880005

Таблиця 12

Результати прогнозування на 5 кроків вперед з урахуванням волатильності ARIMA(1,2,1) + FIGARCH(0,1)

Номер кроку	Прогнозований інтервал	Справжнє значення
1	[198.7961, 208.1213]	208.309998
2	[199.8716, 209.1968]	197.369995
3	[200.2634, 209.5886]	200.860001
4	[202.2021, 211.5273]	202.070007
5	[200.0829, 209.4081]	196.880005

В результаті моделювання реальних фінансових процесів на прикладі даних компанії Tesla з метою отримання точніших прогнозів було показано можливості застосування регресійних моделей (АР, АРКС, АРІКС) і гетероскедастичних моделей (АРУГ, УАРУГ), а також їх комбінації для формування прогнозу ціни на декілька вперед (до 5 кроків). Як видно з результатів було отримано прогнози моделей, а також інтервал значень прогнозу від найнижчої до найвищої ціни. Аналіз показав, що незважаючи на можливість формування прогнозу на декілька кроків вперед, доцільно здійснювати прогнозування лише на крок вперед. Це пов’язано як з нелінійністю і нестаціонарністю фінансових процесів, так і значними коливаннями і впливами, які трапляються на ринку.

Запитання для самоконтролю

1. Які тенденції в зміні ціни компанії Tesla Ви помітили у 2020-22р.?
2. З чим пов’язані ці зміни і яким чином їх можна врахувати?
3. Які моделі варто застосовувати для прогнозування цін на фінансовому ринку?
4. Які моделі дозволяють оцінювати і прогнозувати волатильність ряду? Як сформувати довірчий інтервал цін для інвестора?

6. Прогнозування фінансових індикаторів за допомогою Facebook Prophet

Створення зручного і сучасного інструменту Facebook Prophet для аналізу даних соціальних медіа, соціальних мереж, використання мобільного трафіку надало поштовх для застосування його у прогнозному моделюванні. Тому у даному розділі і нашому курсі ми демонструємо можливості, математичні особливості і програмні реалізації даного пакету у різних програмних середовищах.

Короткі теоретичні відомості

Facebook Prophet – це процедура прогнозування даних часових рядів на основі адитивної моделі, де нелінійні тенденції відповідають річній, щотижневій та щоденній сезонності, а також ефектам відпустки. Формально, модель складається з таких компонентів як тренд, сезонність та враховує святкові періоди і описується наступним рівнянням [12-16]:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t),$$

де g(t) – функція, що описує тренд, s(t) – функція, що описує сезонність та h(t) – функція, що враховує ефект святкових днів/періодів. Похибка або нев'язка моделі $\varepsilon(t)$ є випадковою компоненту і не підлягає поясненню тенденцією, сезонністю та святами і описує зміни, які не враховуються моделлю. Важливо зауважити, що конкретні функціональні форми та параметри кожної компоненти визначаються автоматично Facebook Prophet на підставі аналізу вхідних даних та їх характеристик. Також, $\varepsilon(t) \sim N(0,1)$.

Тренд. Для описання тренду в Prophet може застосовуватись модель насиченого зростання та кусково-лінійна функція.

Оскільки модуль Facebook використовують найчастіше для того, щоб спрогнозувати зростання веб-трафіку, а веб-трафік формують користувачі, то

тренд у цьому модулі моделюється по аналогії з популяцією – моделюється як раніше прийшла аудиторія та як очікується, що вона буде збільшуватись у подальшому. Наприклад, в обраному регіоні під пропускною здатністю для кількості користувачів Facebook розуміють частку населення регіону, що має доступ до Інтернету. Зростання цих користувачів можна змоделювати за допомогою логістичної форми:

$$g(t) = \frac{C}{1 + \exp(-k(t - m))},$$

де C – пропускна здатність, k – коефіцієнт росту, m – параметр зміщення.

Однак у цьому варіанті є декілька проблем:

– Пропускна здатність не може бути константою, оскільки доля населення, що має доступ до Інтернету з часом лише збільшується. Саме тому потрібно змінити константу на функцію від часу – C(t).
– Також, коефіцієнт росту не є постійною величиною, оскільки нові продукти та технології можуть пришвидшувати темп росту. Тому потрібно забезпечити можливість коефіцієнту приймати різні значення.

Окрім цього у моделі є можливість враховувати точки, де змінюється тренд. Нехай є множина S точок, де тренд змінюється, в момент часу s_j, j=1,…,S. Вектор $\delta \in \mathbb{R}^S$ – вектор змін темпів росту, тобто, δ_j – зміна темпу в момент часу s_j. Маємо, що темп росту у довільний час t складається з таких компонентів:

$$k + \sum_{j:t>s_j} \delta_j.$$

Додатково визначимо вектор $\boldsymbol{a}(t)$ компоненти якого:

$$a_j(t) = \begin{cases} 1, t \geq s_j \\ 0, \text{інакше} \end{cases}$$

Отримуємо, що темп росту в момент часу t – $k + \boldsymbol{a}^T(t)\delta$. Тепер ми маємо визначений темп росту у будь-який момент часу, отже, тепер потрібно, щоб параметр зміщення також був скорегований відповідно до кожного інтервалу – потрібно, щоб усі інтервали були з'єднані:

$$\gamma_j = \left(s_j - m - \sum\nolimits_{l<j} \gamma_l\right)\left(1 + \frac{k + \sum_{l<j}\delta_l}{k + \sum_{l\leq j}\delta_l}\right).$$

Остаточно отримуємо модель для тренду:

$$g(t) = \frac{C(t)}{1 + \exp(-(k + \boldsymbol{a}^T(t)\delta)(t - (m + \boldsymbol{a}^T(t)\gamma)))}.$$

У випадку, якщо потрібно спрогнозувати значення, якому не характерне стрімке зростання, часто достатньо лінійної моделі тренду:

$$g(t) = (k + \boldsymbol{a}^T(t)\delta)t + (m +)\boldsymbol{a}^T(t)\gamma.$$

Сезонність. У часових рядах, які пов'язані з людською поведінкою, часто присутня мультиперіодична сезонність. Наприклад, буденні дні – тижнева сезонність, відпустки та/або літні канікули – річна сезонність і т.д.

У моделі Prophet сезонність описується рядами Фур'є:

$$s(t) = \sum_{n=1}^{N}\left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right),$$

де P відповідає сезонному періоду, наприклад, у випадку тижневої сезонності P=7, а у випадку річної P=365.25.

Збільшення порядку N дозволяє ідентифікувати закономірності, які швидко змінюються, але водночас це може призвести до перенавчання моделі. Емпіричним шляхом було обрано найбільш оптимальні значення N: для річної сезонності N=10, а для тижневої – N=3. Інший спосіб обрати оптимальні значення N – це порівнювати критерій Акайке для набору значень N та залишити те значення, яке відповідає найменшому критерію.

Святкові дні та значимі події. Святкові дні часто буває складно враховувати, оскільки їх ефект інколи неможливо змоделювати. Наприклад, у США День подяки святкують четвертого четверга у листопаді місяці, чорні п'ятниці так само повторюються не циклічно – такі події складно запрограмувати та врахувати усі можливі зміни.

Модуль Prophet дозволяє створювати самостійно календар святкових днів та інших значимих подій або ж дає можливість користуватись уже створеним для кожної країни календарем. В залежності від цільової змінної, що прогнозується, можна обрати один з варіантів: якщо спеціаліст знає певні особливості, з якими можуть бути пов'язані різкі зміни поведінки часових рядів, то варто їх включити в модель, якщо ж поведінка часового ряду повністю відповідає календарним святам – обиратиметься стандартний підхід Prophet.

Нехай наслідки від кожного святкового дня – незалежні. Визначимо для кожного свята i множину дат D_i – множина днів, на які раніше припадало свято та коли в майбутньому буде це свято. Визначимо індикаторну функцію

$$Z(t) = [\mathbf{1}(t \in D_1), \dots, \mathbf{1}(t \in D_L)],$$

яка є індикатором святкового дня. Нехай кожному святу i відповідає параметр k_i, тоді ефект від цього свята – функція:

$$h(t) = Z(t)\boldsymbol{k}, \boldsymbol{k} \sim N(0,1).$$

Окрім ефекту святкових днів також може бути присутній ефект від днів до та після свята. У моделі визначається «вікно», яке відповідає святковим дням, та визначаються додаткові параметри для інтервалів, що потрапляють у відповідне «вікно».

Реалізація Facebook Prophet у програмному середовищі Python

Facebook Prophet реалізовано в середовищі Python і можна легко викликати за допомогою відповідної бібліотеки prophet [12].

Послідовність налаштування у Python:

1) Встановлення бібліотеки prophet відбувається стандартно за допомогою менеджера пакетів pip, оскільки prophet є на PyPI

python -m pip install prophet

Починаючи з версії 0.6, Python 2 більше не підтримується. Починаючи з версії 1.0, ім'я пакета на PyPI — «prophet»; на більш старших версіях до версії 1.0 це був пакет під назвою "fbprophet".

Починаючи з версії 1.1, мінімальна підтримувана версія Python становить 3.7. Після встановлення можна приступати до побудови моделі і прогнозування.

2) Anaconda. Prophet також можна встановити через conda-forge.

conda install -c conda-forge

3) Встановлення на Python - Версія розробника

Щоб отримати найновіші зміни коду під час їх об'єднання, можна клонувати цей репозиторій і створити з джерел вручну. Не гарантовано, що це буде стабільно.

git clone https://github.com/facebook/prophet.git

cd prophet/python

python -m pip install -e .

За замовчуванням Prophet використовуватиме фіксовану версію cmdstan (завантажуючи та встановлюючи її за потреби) для компіляції виконуваних файлів моделі. Якщо це небажано, і ви бажаєте використати власну наявну інсталяцію cmdstan, ви можете встановити для змінної середовища PROPHET_REPACKAGE_CMDSTAN значення False:

export PROPHET_REPACKAGE_CMDSTAN=False;
python -m pip install -e.

Для Linux:

Переконайтеся, що встановлено компілятори (gcc, g++, build-essential) і засоби розробки Python (python-dev, python3-dev). У системах Red Hat встановіть пакети gcc64 і gcc64-c++. При використанні віртуальної машини необхідно нонайменше 4 ГБ пам’яті для встановлення Prophet і щонайменше 2 ГБ для використання Prophet.

Для Windows:

Для використання cmdstanpy з Windows неообхідно встановити Unix-сумісний компілятор C, наприклад mingw-gcc. Якщо спочатку встановлюється cmdstanpy, його можна встановити за допомогою команди cmdstanpy.install_cxx_toolchain.

Основні кроки використання бібліотеки і можливостей Facebook Prophet, описані у попередньому підрозділі, будуть реалізовуватись наступним чином у Python.

1. Імпорт бібліотеки та підготовка даних:

import pandas as pd

from fbprophet import Prophet

Завантаження власних даних у DataFrame

data = pd.read_csv('your_time_series_data.csv')

Приведення даних до очікуваного формату (колонки 'ds' і 'y')

data.rename(columns={'timestamp_column': 'ds', 'value_column': 'y'}, inplace=True)

2. Побудова та навчання моделі:

Побудова математичної моделі на основі методу Prophet

model = Prophet()

Навчання моделі на завантаженому наборі даних

model.fit(data)

3. Створення фрейму для формування прогнозів

Створення фрейму з вказаними майбутніми датами для прогнозування

future = model.make_future_dataframe(periods=365)

для прикладу обрано середньострокове прогнозування для 1 року вперед

Формування безпосередньо прогнозу моделі на обраний інтервал прогнозування

forecast = model.predict(future)

4. Візуалізація результатів:

Візуалізація прогнозу разом з інтервалом невизначеності

fig = model.plot(forecast)

Візуалізація окремих компонентів моделі: тренду, сезонності та свят

fig_comp = model.plot_components(forecast)

model.plot_components(forecast)

Інсталяція і застосування Facebook Prophet у R

1) Різні варіанти встановлення Facebook Prophet в залежності від версії [12].

Встановлення R в CRAN

⚠ Версія CRAN Facebook Prophet є досить застарілою. Якщо все ж таки необхідно отримати саме CRAN-версію, то для встановлення пакету CRAN слід використовувати:

install.packages('prophet')

Для доступу до найновіших виправлень помилок і оновлених даних і для встановлення новішої версії рекомендується виконати наступне:

install.packages('remotes')

remotes::install_github('facebook/prophet@*release', subdir = 'R')

Для налаштування експериментального бекенду можна вибрати експериментальний альтернативний сервер Stan під назвою cmdstanr. Після встановлення prophet виконати наступні кроки, щоб використовувати cmdstanr замість rstan як серверну частину:

Рекомендується запустити в новому сеансі R або перезапустити поточний

install.packages(c("cmdstanr", "posterior"), repos = c("https://mc-stan.org/r-packages/", getOption("repos")))

Якщо раніше не встановлювався cmdstan, виконати:

cmdstanr::install_cmdstan()

Інакше можна вказати cmdstanr замість шляху cmdstan:

cmdstanr::set_cmdstan_path(path = <your existing cmdstan>)

Встановити змінну середовища R_STAN_BACKEND

Sys.setenv(R_STAN_BACKEND = "CMDSTANR")

У Windows для R потрібен компілятор, тому потрібно дотримуватись інструкцій rstan. Ключовим кроком є встановлення Rtools перед спробою встановлення пакета.

Якщо є бажаними спеціальні налаштування компілятора Stan, то слід встановлювати його з вихідного коду, а не з двійкового файлу CRAN.

Основні кроки застосування у R

1) Завантаження бібліотек та підготовка даних:

library(prophet)

data <- read.csv("your_time_series_data.csv")

Приведення даних до очікуваного формату (колонки ‘ds’ і ‘y’)

data <- data.frame(ds = as.Date(data$timestamp_column), y = data$value_column)

library(prophet)

2) Побудова та навчання моделі:

Побудова моделі Prophet

model <- prophet()

Навчання моделі на тренувальних даних

model <- fit.prophet(model, data)

3) Створення фрейму для прогнозів та безпосередньо прогнозування:

future <- make_future_dataframe(model, periods = 365) # аналогічно як і для Python обрано період прогнозування на 1 рік вперед

forecast <- predict(model, future)

5) Візуалізація результатів:

plot(model, forecast)

prophet_plot_components(model, forecast)

Це базовий опис налаштування і можливостей застосування бібліотеки prophet для прогнозування часових рядів. За допомогою додаткових параметрів та налаштувань можна підвищити точність та адаптувати модель до конкретного специфічного сценарію і налаштувати горизонт прогнозування та інтервал невизначеності.

Бібліотека prophet надає також можливість враховувати святкові ефекти та інші фактори, які можуть впливати на дані, забезпечуючи зручний інтерфейс для складних моделей прогнозування часових рядів.

Оскільки на даний момент бібліотека Facebook Prophet не має офіційної Java-версії, слід використовувати Python-пакет prophet або його R-версію в своєму Java-проекті. Якщо необхідно інтегрувати прогнозування на базі Facebook Prophet у Java-програму, можна використовувати взаємодію між мовами програмування за допомогою різних підходів, таких як використання системи підпроцесів, веб-сервісів або віддаленого виконання коду. Важливо враховувати, що ці підходи можуть вимагати додаткової конфігурації та обслуговування та можуть вплинути на продуктивність та стабільність вашої програми.

Запитання для самоконтролю

1. Які основні можливості модуля Facebook Prophet?
2. Як враховуються сезонні зміни?
3. Які особливості налаштування і реалізації Facebook Prophet у Python?
4. Як здійснити короткострокове прогнозування за допомогою Facebook Prophet? А довгострокове прогнозування?
5. Як враховуються святкові і вихідні дні в моделі?
6. Які оптимальні значення N для річної і тижневої сезонності?

7. Змішані лінійні моделі в задачах фінансового аналізу і прогнозування

Крім класичних достатньо відомих видів регресійних моделей, практичне застосування яких показано у нашому посібнику у розділах 2-5, актуальним стає використання інтегрованих і комбінованих підходів, які об’єднують переваги різних методів і дозволяють уникнути обчислювальних обмежень. Далі покажемо можливості застосування змішаних лінійних моделей і перспективність їх впровадження для фінансового ризик-менеджменту.

7.1. Короткі теоретичні відомості

Змішані лінійні моделі(ЗЛМ) (англ. Mixed Linear Models, скорочено MLMs) - це статистичні моделі, які комбінують лінійні моделі з фіксованими та випадковими ефектами [17-20]. Вони широко використовуються для аналізу даних, коли наявні незалежні об’єкти або вимірювання, які мають внутрішні варіації або ієрархічну структуру.

Основні характеристики змішаних лінійних моделей:

- **Фіксовані та випадкові ефекти:** Вони дозволяють моделювати як фіксовані, так і випадкові ефекти. Фіксовані ефекти - це загальні ефекти, які впливають на всі об’єкти дослідження, а випадкові ефекти - це ефекти, які враховують внутрішню варіацію або ієрархічну структуру даних.
- **Загальна структура моделі:** Змішані лінійні моделі дозволяють моделювати як залежність між змінними та відповідями (фіксовані ефекти), так і різні рівні варіацій у даних через випадкові ефекти.
- **Ієрархічні моделі:** MLMs особливо корисні для аналізу даних з ієрархічною структурою, де дані розподілені на різні рівні або групи, наприклад, учні в класах, споживачі в магазинах або пацієнти в лікарнях.

- **Пом’якшення впливу викидів:** Випадкові ефекти можуть допомогти зменшити вплив викидів або аномалій у даних на оцінки моделі, оскільки вони допомагають врахувати внутрішні залежності між спостереженнями.
- **Оцінка невизначеності:** Змішані лінійні моделі можуть надавати інформацію про невизначеність оцінок параметрів моделі, включаючи довірчі інтервали.
- **Продуктивність:** Ці моделі можуть бути обчислювально складними, особливо для складних ієрархічних структур даних.

Змішані лінійні моделі знаходять своє застосування в різних галузях, таких як соціологія, медицина, економіка, біологія та інші, де досліджуються дані зі складною структурою та внутрішньою залежністю.

Переваги змішаних лінійних моделей:

- **Моделювання випадкових та фіксованих ефектів:** ЗЛМ дозволяють враховувати як фіксовані, так і випадкові ефекти одночасно. Це робить їх потужним інструментом для аналізу даних зі складною структурою та внутрішньою залежністю.
- **Урахування ієрархічної структури даних:** ЗЛМ особливо корисні для аналізу даних з ієрархічною або груповою структурою, де дані розділені на різні рівні.
- **Управління невизначеністю:** ЗЛМ надають засоби для оцінки невизначеності оцінок параметрів моделі за допомогою довірчих інтервалів.
- **Пом’якшення впливу викидів:** Випадкові ефекти допомагають зменшити вплив викидів або аномалій на оцінки моделі.

- **Аналіз варіацій:** ЗЛМ дозволяють досліджувати рівень варіації між групами та в межах груп.

Недоліки змішаних лінійних моделей:

- **Складність обчислень:** ЗЛМ можуть бути обчислювальноскладними, особливо для складних моделей та великих наборів даних.
- **Вимоги до даних:** Для успішного застосування ЗЛМ може бути необхідним мати велику кількість спостережень, особливо якщо є багато випадкових ефектів або груп.
- **Суб'єктивність вибору моделі:** Вибір фіксованих та випадкових ефектів, а також структури моделі може вимагати певного рівня експертного знання та суб'єктивного вибору.
- **Залежність від припущень:** ЗЛМ, як і будь-які статистичні моделі, ґрунтуються на певних припущеннях про розподіли та структуру даних, і некоректне використання може призвести до недостовірних результатів.
- **Обмежена гнучкість:** ЗЛМ можуть бути обмежені в своїй гнучкості порівняно з більш складними моделями.

При використанні змішаних лінійних моделей важливо враховувати контекст дослідження, специфіку даних та цілі аналізу, а також здійснювати адекватну валідацію та перевірку припущень моделі.

7.2. Реалізація змішаних моделей у SAS

Змішані моделі в SAS реалізовані через процедуру PROC MIXED. Ця процедура дозволяє моделювати випадкові та фіксовані ефекти у даних з ієрархічною структурою або груповими впливами. Процедура PROC MIXED є потужним інструментом для виконання аналізу змішаних моделей в SAS [20].

Ось загальна структура використання PROC MIXED для реалізації змішаних моделей:

```
PROC MIXED DATA=data;
   CLASS categorical_variables;  /* Категоріальні змінні */
   MODEL response = fixed_effects / SOLUTION;
   RANDOM random_effects;
RUN;
```

Де:

- DATA=data: Вказує набір даних, на якому буде виконуватися аналіз.
- CLASS categorical_variables: Вказує категоріальні змінні, які використовуються як фактори або групи.
- MODEL response = fixed_effects / SOLUTION: Вказує модель, де response - це залежна змінна, fixed_effects - фіксовані ефекти або предиктори, які ви хочете включити в модель, а SOLUTION вказує вивід рішення, що містить результати аналізу.
- RANDOM random_effects: Вказує випадкові ефекти або групові впливи.

У реальному використанні ви повинні відповідно налаштувати та адаптувати параметри для вашого конкретного аналізу. Більш докладну інформацію про параметри та можливості процедури PROC MIXED ви можете знайти у документації SAS або в літературі зі статистичного аналізу.

7.3. Реалізація змішаних лінійних моделей у Python

В Python для реалізації змішаних лінійних моделей (ЗЛМ) ви можете використовувати бібліотеку statsmodels [19]. Вона надає зручний інтерфейс для

аналізу даних та побудови різноманітних статистичних моделей, включаючи ЗЛМ. Ось як це можна зробити:

Встановлення бібліотеки

Спочатку переконайтеся, що у вас встановлена бібліотека statsmodels. Ви можете встановити її за допомогою менеджера пакетів pip:

pip install statsmodels

Ось загальний підхід до реалізації ЗЛМ за допомогою бібліотеки statsmodels:

```
import pandas as pd
import statsmodels.api as sm
# Завантажте ваші дані у DataFrame
data = pd.read_csv('your_data.csv')
# Перетворіть категоріальні змінні в даммі-змінні (бінарні)
dummies = pd.get_dummies(data['categorical_variable'], drop_first=True)
data = pd.concat([data, dummies], axis=1)
# Визначте матрицю фіксованих ефектів та випадкових ефектів
X = data[['fixed_effect_1', 'fixed_effect_2', ...]]  # фіксовані ефекти
Z = data[['random_effect_1', 'random_effect_2', ...]]  # випадкові ефекти
# Визначте залежну змінну
y = data['dependent_variable']
# Побудуйте змішану модель
mixedlm_model = sm.MixedLM(y, X, Z)
# Оцініть модель
mixedlm_results = mixedlm_model.fit()
# Виведіть результати
print(mixedlm_results.summary())
```

У Python бібліотека statsmodels підтримує реалізацію змішаних лінійних моделей (ЗЛМ) через підмодуль statsmodels.regression.mixed_linear_model. Цей

підмодуль надає можливість моделювати різні типи ЗЛМ з різними структурами фіксованих та випадкових ефектів.

Ось деякі типи ЗЛМ, які підтримуються у бібліотеці statsmodels:

- ***Модель з випадковими ефектами***:

MixedLM: Загальний клас для моделювання змішаних лінійних моделей з випадковими ефектами. Вона дозволяє задати різні структури випадкових ефектів (наприклад, незалежні, структурні або авторегресійні).

- ***Модель з фіксованими та випадковими ефектами***:

MixedLM: Цей клас дозволяє включити як фіксовані, так і випадкові ефекти.

- ***Модель змішаних ефектів для бінарних даних***:

MixedLMBin: Модель змішаних ефектів для аналізу бінарних даних.

- ***Модель змішаних ефектів для пуассонівського розподілу***:

MixedLMCount: Модель змішаних ефектів для аналізу лічильних даних, таких як кількість випадків.

Це лише деякі приклади типів ЗЛМ, які можна реалізувати за допомогою statsmodels.regression.mixed_linear_model. Кожен тип моделі має свої властивості та параметри, які можуть бути налаштовані згідно з конкретними потребами аналізу.

7.4. Перспективи і напрями розвитку змішаних лінійних моделей

Змішані лінійні моделі (ЗЛМ) є потужним інструментом для аналізу даних зі складною структурою та внутрішньою залежністю. Розвиток ЗЛМ

продовжується, і існують кілька перспективних напрямів для їхнього вдосконалення та розширення. Ось деякі з можливих напрямів розвитку:

- **Розширення варіаційних компонентів:** Важливо розробити більш гнучкі моделі для моделювання різних типів варіації в даних. Це може включати розширення випадкових ефектів для більшої гнучкості в моделюванні.
- **Моделі для нелінійних та нелінійно-випадкових залежностей:** Розробка ЗЛМ для моделювання нелінійних залежностей між змінними та випадковими ефектами може розширити можливості аналізу.
- **Врахування більш складних ієрархічних структур:** Розробка методів для моделювання складніших ієрархічних структур даних дозволить точніше враховувати ієрархічну структуру в аналізі.
- **Адаптація до великих об'ємів даних:** Розвиток методів для аналізу ЗЛМ на великих наборах даних допоможе зробити їх ефективнішими та доступнішими для великих досліджень.
- **Інтеграція з іншими моделями:** Розробка методів для інтеграції ЗЛМ з іншими статистичними та машинними методами дозволить вирішувати складніші задачі аналізу даних.
- **Обробка даних різних типів:** Розробка ЗЛМ для аналізу даних різних типів, таких як бінарні, категоріальні, лічильні та інші, може розширити застосування моделей у різних галузях.
- **Адаптація до нових технологій:** Розвиток методів для аналізу даних, зібраних за допомогою нових технологій, таких як сенсори, датчики, Інтернет речей, дозволить вирішувати нові завдання аналізу даних.

Загалом, розвиток ЗЛМ полягає у розширенні їхнього функціоналу, зростанні їхньої гнучкості та ефективності, а також у забезпеченні засобів для адаптації до різних типів даних та завдань аналізу.

7.5. Практичні завдання застосування змішаних лінійних моделей для розв'язання фінансових задач

Змішані лінійні моделі можуть бути застосовані в фінансовому аналізі для різних завдань, включаючи аналіз часових рядів цін акцій, портфельний аналіз, аналіз волатильності та багато інших. Один з можливих прикладів застосування ЗЛМ в фінансах - це аналіз залежностей між доходами акцій різних компаній.

Розглянемо приклад аналізу залежностей між прибутковостями акцій двох різних компаній. Для спрощення припустимо, що ми маємо дані щодо прибутковості акцій компаній A та B протягом декількох років.

Завдання 1: Аналіз залежностей між прибутковістю акцій

Припустимо, що у нас є дані про щоквартальну прибутковість акцій компаній A та B протягом останніх 5 років. Ми хочемо дослідити, чи є статистично значуща залежність між прибутковістю акцій цих двох компаній.

```
 import pandas as pd
import statsmodels.api as sm
# Завантажте дані
data = pd.read_csv('stock_data.csv')
# Перетворення даних
```

```
quarterly_dummies = pd.get_dummies(data['quarter'], drop_first=True)
data = pd.concat([data, quarterly_dummies], axis=1)

# Визначення залежної та незалежних змінних
y = data['stock_A_return']
X = data[['stock_B_return', 'Q2', 'Q3', 'Q4']]  # Включаємо індикатори кварталів
# Побудова змішаної моделі
mixedlm_model = sm.MixedLM(y, X)
# Оцінка моделі
mixedlm_results = mixedlm_model.fit()
# Виведення результатів
print(mixedlm_results.summary())
```

У цьому прикладі ми завантажуємо дані про доходи акцій компаній A та B, створюємо даммні-змінні для кварталів, визначаємо залежні та незалежні змінні, будуємо змішану модель за допомогою statsmodels, оцінюємо модель та виводимо результати.

Змішані лінійні моделі також можуть бути застосовані для складніших завдань, таких як аналіз волатильності, аналіз портфелів, аналіз динаміки цін та багатьох інших фінансових аспектів.

Завдання 2. Застосування змішаних моделей для прогнозування цін на акції з врахуванням змішаних факторів

Змішані лінійні моделі можуть бути використані для прогнозування цін на акції, моделювання економічних процесів, як складові інших фінансових інструментів. Прогнозування цін на акції є складною задачею через їхню низьку

передбачуваність та залежність від впливу багатьох факторів. Змішані моделі можуть бути корисним інструментом для врахування складності та динаміки таких факторів.

Моделюйте ризик портфеля з урахуванням як фіксованих, так і випадкових факторів, таких як ринковий індекс і індивідуальні акції.

```
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf

data = pd.read_csv('portfolio_risk_data.csv')
model = smf.mixedlm("PortfolioRisk ~ MarketIndex + Stock1 + Stock2", data, groups=data['Portfolio'])
result = model.fit()
forecasted_risk = result.predict(data)
```

Завдання 3. Прогнозування цін на акції з врахуванням інших факторів

Припустимо, що ми маємо дані про ціни на акції певної компанії протягом деякого періоду часу, а також дані про макроекономічні показники (наприклад, ставку безризикової облігації, індекс споживчих цін тощо). Ми хочемо побудувати модель, яка прогнозує майбутні ціни на акції, враховуючи вплив цих макроекономічних показників.

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
# Завантажте дані
```

```
data = pd.read_csv('stock_price_data.csv')
# Перетворення даних
data['lagged_price'] = data['price'].shift(1)  # Додаткова змінна - попередня ціна
data['log_return'] = np.log(data['price']) - np.log(data['lagged_price'])  # Логарифмічний дохід
# Включення макроекономічних показників
macro_data = pd.read_csv('macroeconomic_data.csv')
data = pd.merge(data, macro_data, on='date', how='left')
# Визначення залежної та незалежних змінних
y = data['log_return']
X = data[['lagged_price', 'risk_free_rate', 'inflation_rate']]  # Макроекономічні показники
# Побудова змішаної моделі
mixedlm_model = sm.MixedLM(y, X)
# Оцінка моделі
mixedlm_results = mixedlm_model.fit()
# Виведення результатів
print(mixedlm_results.summary())
```

Завантажуємо дані цін на акції компанії та макроекономічні показники, створюємо додаткову змінну «попередня ціна» та логарифмічний дохід, включаємо макроекономічні показники в якості незалежних змінних, будуємо змішану модель та оцінюємо її.

Завдання 4. Прогнозування цін на акції з урахуванням змішаних часових та групових ефектів:

Для моделювання зміни цін на акції з урахуванням як часових, так і групових ефектів використовуйте змішані лінійні моделі з фіксованими та випадковими ефектами.

```
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf
data = pd.read_csv('stock_price_data.csv')
model = smf.mixedlm("StockPrice ~ Time + (1 | Stock)", data)
result = model.fit()
predicted_prices = result.predict(data)
```

Завдання 5. Прогнозування волатильності ринку з урахуванням їх динаміки за допомогою GARCH-змішаних моделей

Ціни на акції різних компаній можуть мати значну волатильність і можуть мати вплив на ринок в цілому. Тому для прогнозування волатильності ринку з урахуванням часової змінності є сенс використати GARCH-змішані моделі.

```
import pandas as pd
from arch import arch_model
import statsmodels.formula.api as smf
data = pd.read_csv('market_volatility_data.csv')
model = smf.mixedlm("Volatility ~ Time", data, groups=data['MarketIndex'])
result = model.fit()
# Перевірка кожного групового ефекту (ARCH)
arch_model_results = [arch_model(group_data['Volatility']).fit() for group,
group_data in data.groupby('MarketIndex')]
# Прогноз волатильності для кожної групи
```

forecasted_volatility = [model.predict(group_data) for model, group_data in zip(arch_model_results, data.groupby('MarketIndex'))]

Завдання 6. Застосування змішаних лінійних моделей для прогнозування прибутковості портфелю акцій

Змішані лінійні моделі можуть бути корисним інструментом для прогнозування прибутковості портфелю акцій. Прогнозування прибутковості портфелю акцій є важливим завданням для інвесторів та фінансових аналітиків, оскільки допомагає приймати обґрунтовані рішення щодо розподілу інвестицій та управління ризиком.

Припустимо, що у нас є дані про прибутковість окремих акцій та загальну ринкову прибутковість протягом деякого періоду часу. Ми хочемо побудувати модель для прогнозування прибутковості нашого портфелю акцій, враховуючи індивідуальні прибутковості акцій та ринкову прибутковість.

import pandas as pd

import statsmodels.api as sm

Завантажте дані

data = pd.read_csv(‘stock_returns_data.csv’)

Визначення залежної та незалежних змінних

y = data[‘portfolio_return’] # Прибутковість портфелю

X = data[[‘stock1_return’, ‘stock2_return’, ‘market_return’]] # Прибутковості акцій та ринкова прибутковість

Побудова змішаної моделі

mixedlm_model = sm.MixedLM(y, X)

Оцінка моделі

mixedlm_results = mixedlm_model.fit()

Виведення результатів

print(mixedlm_results.summary())

Завантажуємо дані прибутковості портфелю акцій, індивідуальні прибутковості акцій та ринкову прибутковість, визначаємо залежні та незалежні змінні, будуємо змішану модель за допомогою statsmodels, оцінюємо модель та виводимо результати.

Завдання 7. Вибір оптимального портфелю інвестора

Припустимо, що ми маємо дані щодо прибутковості різних акцій протягом деякого періоду часу, а також дані про ризик (наприклад, стандартне відхилення прибутковості) та інші характеристики акцій. Необхідно побудувати модель для вибору оптимального портфелю, який максимізує очікувану прибутковість при заданому ризику або мінімізує ризик при заданій прибутковості.

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import scipy.optimize as sco
# Завантажте дані
data = pd.read_csv('stock_data.csv')
# Визначення залежної та незалежних змінних
y = data['expected_return']  # Очікувана прибутковість
X = data[['risk', 'other_factors']]  # Ризик та інші фактори
# Побудова змішаної моделі для прибутковості
mixedlm_model = sm.MixedLM(y, X)
# Оцінка моделі для прибутковості
mixedlm_results = mixedlm_model.fit()
```

```
# Отримання оцінок коефіцієнтів для ризику та інших факторів
risk_coefficient = mixedlm_results.params['risk']
other_factors_coefficient = mixedlm_results.params['other_factors']
# Оптимізація вибору портфелю
def portfolio_objective(weights):
return-1*(weights[0] * risk_coefficient + weights[1] * other_factors_coefficient)
constraints = ({'type': 'eq', 'fun': lambda weights: np.sum(weights) - 1})
initial_guess = [0.5, 0.5]  # Початкові значення ваг акцій
optimal_portfolio=sco.minimize(portfolio_objective,initial_guess,method='SLS
QP', constraints=constraints)
# Виведення результатів вибору оптимального портфелю
print("Optimal Portfolio Weights:", optimal_portfolio.x)
print("Optimal Portfolio Expected Return:", -1 * optimal_portfolio.fun)
```

Послідовність вирішення завдання: завантажуємо дані щодо прибутковості та ризик акцій, визначаємо залежні та незалежні змінні, будуємо змішану модель для прибутковості за допомогою statsmodels, оцінюємо модель та отримуємо оцінки коефіцієнтів для ризику та інших факторів. Далі використовуємо оптимізацію портфелю для вибору таких ваг акцій, які максимізують очікувану прибутковість при заданому ризику або мінімізують ризик при заданій прибутковості.

Запитання для самоконтролю

1. Що таке змішані лінійні моделі?
2. Як за допомогою ЗЛМ можна прогнозувати ціни на акції з врахуванням інших факторів?
3. Які типи ЗЛМ реалізовані у бібліотеці statsmodels?

4. Як реалізувати змішані лінійні моделі в Python?

5. Які недоліки ЗЛМ обмежують їх використання?

6. Реалізуйте в Python задачу формування оптимального пакету інвестора на прикладі акцій певної галузі.

7. Спрогнозуйте прибутковість портфелю інвестора для обраних фінансових індикаторів.

8. Практичні завдання фінансового менеджера на фінансовому ринку

Даний розділ містить основні задачі, які необхідно вирішувати на фінансових ринках при оцінці прибутковості тих чи інших інструментів. Вам пропонується скористатись напрацьованими прикладами, обравши власні інструменти прибутковості і виконавши моделювання на основі запропонованих прикладів.

Приклад 1. Побудуємо математичну модель для 456 спостережень індексу місячного доходу для біржі S&P 500 [1]. Узагальнена авторегресійна умовно гетероскедастична модель процесу (УАРУГ(1, 1)) має вигляд:

$$y(k) = 0{,}583 + h(k)\varepsilon(k), \qquad (8.1)$$

$$h(k) = 2{,}975 + 0{,}079\, u^2(k-1) + 0{,}684 h(k-1), \qquad (8.2)$$

де $u(k) = h^{1/2}(k)\varepsilon(k)$. Застосування статистики Люнга-Бокса до нормованих залишків моделі та їх квадратів показало, що модель (8.1)-(8.2) має високу ступінь адекватності.

Стохастична модель волатильності для цього ж процесу:

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

$$\ln h(k) = \alpha_0 + \alpha_1 \ln h(k-1) + v(k), \qquad (8.3)$$

де $\{v(k)\} \sim N(0, \sigma_v^2)$. Для того щоб застосувати дискретизацію Гіббса (метод Монте Карло для марковських ланцюгів (МКМЛ)), скористаємось апріорними розподілами:

$$\mu \sim N(0;\ 7), \quad \alpha \sim N[\alpha_0,\ diag(0{,}08;\ 0{,}04)];$$

де $\alpha_0 = [0{,}3;\ 0{,}7]^T$ (початкові значення оцінок взяті з моделі (8.2) для умовної дисперсії $h(k)$). На основі вибірки даних обчислено наступні значені дисперсії випадкового процесу та середнього: $\sigma_v^2 = 0{,}5$ і $\mu = 0{,}73$. Значення $h(k)$ отримано за допомогою методу Монте Карло (дискретизація Гіббса) з обмеженням $0 \le h(k) \le 1{,}5s^2$, де s^2 – вибіркова дисперсія ряду $\log[y(k)]$.

Для того щоб оцінити параметри моделі (8.3), виконано 7000 ітерацій алгоритму дискретизації Гіббса, при цьому перші 500 ітерацій не враховано з огляду на перехідний процес, характерний для даної процедури оцінювання [1]. Апостеріорні середні і стандартні похибки оцінок чотирьох коефіцієнтів моделі (8.3) наведені в таблиці 13.

Таблиця 13

Результати оцінювання параметрів моделі

Параметри	μ	α_0	α_1	σ_v^2
Середнє	0,784	0,829	0,575	0,351
Станд. Похибка	0,196	0,207	0,077	0,064

Оскільки коефіцієнт $\hat{\alpha}_1 = 0{,}575$, це свідчить про існування високої кореляції між значеннями волатильності досліджуваного фінансового ряду. Виконано додаткові обчислювальні експерименти з оцінювання параметрів стохастичної моделі волатильності при різних значеннях кількості циклів методу МКМЛ, які свідчать про повторюваність та стійкість отриманих оцінок $h(k)$.

Приклад 2. В цьому прикладі розглядається двовимірна модель волатильності, побудована для двох змінних: логарифмів доходів по акціях фірми IBM за період часу з січня 1962 по грудень 1999 року і індексу S&P 500

за той же період часу [1]. Таким чином, $\mathbf{y}(k) = [ibm(k),\ sp(k)]^T$. Виберемо модель УАРУГ зі змінними коефіцієнтами кореляції та застосуємо розклад Холецького:

$$\mathbf{y}(k) = \beta_0 + \mathbf{u}(k), \tag{8.4}$$

$$g_{11}(k) = \alpha_{10} + \alpha_{11}\, g_{11}(k-1) + \alpha_{12}\, u_1^2(k-1), \tag{8.5}$$

$$g_{22}(k) = \alpha_{20} + \alpha_{21}\, u_1^2(k-1), \tag{8.6}$$

$$q_{21}(k) = \gamma_0, \tag{8.7}$$

Оцінки параметрів та їх стандартні похибки наведені в таблиці 14. Для визначення оцінок виконано 1500 ітерацій алгоритму оцінювання, при цьому перші п’ятсот ітерацій пропущено.

Таблиця 14

Результати оцінювання двовимірної моделі

Двовимірна модель УАРУГ(1,1) із змінними коефіцієнтами кореляції									
Параметр	β_{01}	β_{02}	α_{01}	α_{11}	α_{12}	α_{20}	α_{21}	γ_0	
Оцінка	1,07	0,84	3,78	0,81	0,14	11,42	0,06	0,41	
Станд. пох.	0,337	0,25	1,89	0,08	0,05	0,99	0,03	0,03	
Стохастична модель волатильності									
Параметр	β_{01}	β_{02}	α_{01}	α_{11}	σ_{1v}^2	α_{20}	σ_{2v}^2	γ_0	σ_η^2
Апост. сер.	0,89	0,91	0,52	0,82	0,07	1,69	0,43	0,37	0,09
Станд. пох.	0,37	0,25	0,26	0,06	0,03	0,13	0,05	0,03	0,03

З метою порівняння використано також рівняння (8.4) для основної змінної і стохастична модель волатильності такого вигляду:

$$\ln g_{11}(k) = \alpha_{10} + \alpha_{11} \ln g_{11}(k-1) + v_1(k), \quad Var[v_1(k)] = \sigma_{1v}^2, \tag{8.8}$$

$$\ln g_{22}(k) = \alpha_{20} + v_2(k), \quad Var[v_2(k)] = \sigma_{2v}^2, \tag{8.9}$$

$$q_{21}(k) = \gamma_0 + \eta(k), \qquad Var[\eta(k)] = \sigma_\eta^2. \qquad (8.10)$$

Апріорні розподіли:

$$\beta_{i0} \sim N(0{,}8;\ 4); \quad \alpha_1 \sim N[(0{,}4;\ 0{,}8)^T;\ diag(0{,}15;\ 0{,}05)]; \quad \alpha_{20} \sim N(5{,}5;\ 25);$$

$$\gamma_0 \sim N(0{,}5;\ 0{,}4); \quad \frac{10 \times 0{,}1}{\sigma_{1v}^2} \sim \chi_{10}^2; \quad \frac{5 \times 0{,}2}{\sigma_{2v}^2} \sim \chi_5^2; \quad \frac{5 \times 0{,}2}{\sigma_\eta^2} \sim \chi_5^2.$$

Дані апріорні розподіли є відносно неінформативними. Алгоритм дискретизації Гіббса було застосовано для реалізації 1500 ітерацій, при цьому перші 400 ітерацій в подальшому не використовувались для знаходження (усереднення) оцінок. Випадкові вибірки параметра $g_{ii}(k)$ отримано за допомогою решітчастої процедури Гіббса з використанням 500 точок в інтервалі $[0; 1{,}5s_i^2]$, де $s_i^2 -$ вибіркова дисперсія логарифму $y_i(k), i = 1{,}2$. Апостеріорні середні та стандартні похибки оцінок параметрів двовимірної стохастичної моделі також наведені в таблиці 14. Для перевірки збіжності алгоритму оцінювання з використанням дискретизації Гіббса сеанси оцінювання реалізовано 10 разів по 1500 ітерацій. Встановлено, що процедура оцінювання генерує стійкі повторювані оцінки.

Приклад 3. **Дисперсія – міра валютного ризику банку**. Оцінюванню валютного ризику приділяють у банківській діяльності значну увагу. Він представляє собою *можливість фінансових втрат внаслідок негативних змін (падіння) курсів іноземних валют* або внаслідок невідповідності встановлених банком курсів ринковій коньюнктурі (наприклад, масовий продаж валюти). Для визначення можливих фінансових втрат можна застосувати методику їх розрахунку, яка отримала назву VaR (Value-at-Risk) [1].

Значення VaR – це величина капіталу, яка з визначеною ймовірністю покриє можливі збитки на протязі деякого періоду часу.

Застосування VaR дає можливість стверджувати: "*Ми впевнені на* $X\,\%$, *що наші втрати не будуть перевищувати значення* Y *на протязі наступних* N *днів*". Тобто розраховане значення VaR можна інтерпретувати як таке, що вказує на можливі втрати із заданою ймовірністю, а для компенсації втрат необхідно мати відповідний резервний капітал. Для пояснення VaR можна скористатись означенням *квантиля*.

Квантиль розподілу визначається парою величин: *це деяке вибране значення випадкової змінної* x *та відповідна йому імовірність* c. Для квантиля x значення c визначається так:

$$c = F(x) = \int_{-\infty}^{x} f(u)\,du\,.$$

Тобто ймовірність того, що $x_i < x$ дорівнює c: $p(x_i < x) = c$. Ймовірність того, що $x_i > x$ визначається так: $p(x_i \geq x) = 1 - c$. Квантиль позначають через $Q(X, c)$. Очевидно, що **медіана** представляє собою 50%-й квантиль.

З точки зору теорії статистики VaR можна інтерпретувати як точку зрізу на розподілі втрат (прибутків) таку, що втрати не будуть мати місце з імовірністю, скажемо, $p = 95\%$. Якщо $f(u)$ – розподіл втрат (прибутків), то означення VaR зв'язане з виразом:

$$F(x) = \int_{-\infty}^{x} f(u)\,du = 1 - p\,,$$

де p – ймовірність того, що втрат не буде (ймовірність, зв'язана з правим хвостом розподілу). Тепер VaR можна визначити як відхилення між очікуваним значенням та квантилем:

$$VaR(c) = E[X] - Q(X, c),$$

де $c-$ ймовірність втрат, зв'язана з лівим хвостом розподілу; $Q(x, c)-$ відповідний квантиль розподілу.

Методика VaR для оцінювання валютного ризику

Дана методика використовується для оцінювання валютних ризиків, тобто ризиків, пов'язаних із використанням іноземних валют. Для оцінювання валютних позицій використовують курс НБУ.

Вихідні (початкові) дані, необхідні для розрахунку VaR:

- значення курсів НБ для валют за останні 60 робочих днів (за 60 днів можна зібрати необхідну за об'ємом статистику відносно курсів валют); наприклад, можна розглянути три найбільш валюти: USD, EUR і JPY;
- значення відкритих валютних позицій для вибраних валют, тобто щоденні об'єми операцій для кожної валюти.

Розглянемо виконання обчислювальних операцій методики у вигляді послідовності декількох кроків.

Крок 1. Першим кроком реалізації методики є збір статистичних даних стосовно курсів валют та відкритих валютних позицій.

Крок 2. На другому кроці розраховують темпи зміни курсів валют за виразом:

$$x^i(k) = \ln\left(\frac{Курс^{\text{i}}(k)}{Курс^{\text{i}}(k-1)}\right),\ k = 1,2,3,\ldots N,\ \ i = 1,2,\ldots,n,$$

де $Курс^{\text{i}}(k)-$ значення курсу $i-$ї валюти в момент k; $N-$ загальне число значень статистичних даних; $n-$ кількість валют. З обчислених значень динаміки курсів формують матрицю вимірів, $\mathbf{X}$, вимірністю $[N \times n]$.

Крок 3. Розрахунок коваріаційної та кореляційної матриць для $\mathbf{X}$. Елементи кореляційної матриці можна розрахувати в такій послідовності:

– Нормування елементів матриці $\mathbf{X}$:

$$x_{ij}^{н} = \frac{x_{ij} - \mu_i}{\sqrt{\text{var}[x_i]}} = \frac{x_{ij} - \mu_i}{\sigma_{x_i}}, \quad i = 1,...,n; \quad j = 1,...,N,$$

де μ_i – середнє значення динаміки зміни $i-$ї валюти, тобто, $i-$го стовпчика матриці $\mathbf{X}$; $\text{var}[x_i]$ – дисперсія динаміки курсу $i-$ї валюти; σ_{x_i} – відповідне стандартне відхилення. В результаті такого нормування отримаємо матрицю нормованих значень $\mathbf{X}_H$ динаміки курсів валют. Діагональні елементи коваріаційної матриці $\mathbf{C}$ представляють собою дисперсії $\sigma_i, i = 1,...,n$ відповідних динамік курсів валют.

– Обчислення кореляційної матриці $\mathbf{R}[n \times n]$:

$$\mathbf{R} = \mathbf{X}_H^T \mathbf{X}_H .$$

Крок 4. Прогнозування волатильності. Деякі методики обчислення прогнозу пропонують такий вираз: $\sigma_i^{прогн.} = a_i \sigma_i$, де $a_i = F(\mathbf{R}, \sigma_i)$. Пропонуються, наприклад, такі значення: $a = 0{,}5 \div 3{,}0$. При цьому чим ближче значення a до 1, тим нижчою є волатильність валюти.

Коректним підходом до прогнозування волатильності є застосування рівнянь гетероскедастичних процесів, які розглянуті вище. Кращі результати, як правило, можна отримати за допомогою рівнянь УАРУГ та ЕУАРУГ.

Крок 5. Розрахунок величини VaR. Значення VaR визначають за виразом:

$$VaR_i = \alpha \sigma_i^{прогн.} V_i \sqrt{N}, \quad i = 1,...,n,$$

де α – квантиль для довірчого інтервалу (для інтервалу 95% він складає 1,65, а для 99% дорівнює 2,33); $\sigma_i^{прогн.}$ – значення прогнозу волатильності, отримане

на попередньому кроці; $V_i -$ об'єм операцій для $i-$ї валютної позиції на визначений момент часу, який визначають за виразом:

$$V_i = Позиція_i \cdot Курс_i .$$

Обчислені значення VaR_i утворюють вектор:

$$VaR = [VaR_1 \ VaR_2 \ ... VaR_n]^T ,$$

який використовують для обчислення загального значення $VaR_{заг.}$:

$$VaR_{заг.} = \sqrt{VaR^T \cdot \mathbf{R} \cdot VaR} .$$

Значення $\sigma_i^{прогн.}$ можуть корегуватись у бік зростання в залежності від того, яка є додаткова інформація стосовно можливих змін курсів.

Виконані розрахунки свідчать, що можливі втрати (наступного дня) за валютними позиціями не перевищать значення VaR_i.

Варіанти практичних та індивідуальних завдань

У таблиці 15 наведено варіанти для виконання практичного завдання студентами.

Таблиця 15

Варіанти та дані для практичного завдання

Варіант	Приклад 1	Приклад 2	Приклад 3
1	S&P 500 за період з січня 2001 по грудень 2023 року	S&P 500 і ціни на акції Adobe за період з січня 2001 по грудень 2023 року	USD, EUR і NOK на період 90 днів
2	S&P 500 за період з січня 2000 по грудень 2022 року	S&P 500 і ціни на акції Apple за період з січня 2000 по грудень 2022 року	USD, EUR і NOK на період 60 днів
3	S&P 500 за період з січня 1999 по грудень 2021 року	S&P 500 і ціни на акції AT&T за період з січня 1999 по грудень 2021 року	USD, EUR і KDN на період 90 днів
4	S&P 500 за період з січня 1998 по грудень 2020 року	S&P 500 і ціни на акції Autodesk за період з	USD, EUR і KDN на період 60 днів

		січня 1998 по грудень 2020 року	
5	S&P 500 за період з січня 1997 по грудень 2019 року	S&P 500 і ціни на акції Bank of America за період з січня 1997 по грудень 2019 року	USD, EUR і PLN на період 90 днів
6	S&P 500 за період з січня 1996 по грудень 2018 року	S&P 500 і ціни на акції Avery Dennison за період з січня 1996 по грудень 2018 року	USD, EUR і PLN на період 60 днів
7	S&P 500 за період з січня 1995 по грудень 2017 року	S&P 500 і ціни на акції Baxter International за період з січня 1995 по грудень 2017 року	USD, EUR і AUD на період 90 днів
8	S&P 500 за період з січня 1994 по грудень 2016 року	S&P 500 і ціни на акції Boeing за період з січня 1994 по грудень 2016 року	USD, EUR і AUD на період 60 днів
9	S&P 500 за період з січня 1993 по грудень 2015 року	S&P 500 і ціни на акції Caterpillar Inc. за період з січня 1993 по грудень 2015 року	USD, EUR і CNY на період 90 днів
10	S&P 500 за період з січня 1992 по грудень 2014 року	S&P 500 і ціни на акції Chevron Corporation за період з січня 1992 по грудень 2014 року	USD, EUR і CNY на період 60 днів
11	S&P 500 за період з січня 1991 по грудень 2013 року	S&P 500 і ціни на акції Cisco за період з січня 1992 по грудень 2014 року	USD, EUR і JPY на період 90 днів
12	S&P 500 за період з січня 1999 по грудень 2012 року	S&P 500 і ціни на акції IBM за період з січня 1999 по грудень 2012 року	USD, EUR і JPY на період 60 днів
13	NASDAQ за період з січня 2001 по грудень 2023 року	S&P 500 і ціни на акції Coca-Cola Company за період з січня 2001 по грудень 2023 року	USD, EUR і SEK на період 60 днів
14	NASDAQ за період з січня 2000 по грудень 2022 року	S&P 500 і ціни на акції Colgate-Palmolive за період з січня 2000 по грудень 2022 року	USD, EUR і SEK на період 90 днів
15	NASDAQ за період з січня 1999 по грудень 2021 року	S&P 500 і ціни на акції CSX за період з січня 1999 по грудень 2021 року	USD, EUR і CHF на період 60 днів

16	NASDAQ за період з січня 1998 по грудень 2020 року	S&P 500 і ціни на акції Disney за період з січня 1998 по грудень 2020 року	USD, EUR і CHF на період 90 днів
17	NASDAQ за період з січня 1997 по грудень 2019 року	S&P 500 і ціни на акції Duke Energy за період з січня 1997 по грудень 2019 року	USD, EUR і GBF на період 60 днів
18	NASDAQ за період з січня 1996 по грудень 2018 року	S&P 500 і ціни на акції Edison International за період з січня 1996 по грудень 2018 року	USD, EUR і GBF на період 90 днів
19	NASDAQ за період з січня 1995 по грудень 2017 року	S&P 500 і ціни на акції ExxonMobil за період з січня 1995 по грудень 2017 року	USD, EUR і CZK на період 60 днів
20	NASDAQ за період з січня 1994 по грудень 2016 року	S&P 500 і ціни на акції Johnson & Johnson за період з січня 1994 по грудень 2016 року	USD, EUR і CZK на період 90 днів
21	NASDAQ за період з січня 1993 по грудень 2015 року	S&P 500 і ціни на акції FedEx за період з січня 1993 по грудень 2015 року	USD, EUR і TRY на період 60 днів
22	NASDAQ за період з січня 1992 по грудень 2014 року	S&P 500 і ціни на акції Goldman Sachs за період з січня 1992 по грудень 2014 року	USD, EUR і TRY на період 90 днів
23	NASDAQ за період з січня 1991 по грудень 2013 року	S&P 500 і ціни на акції HP Inc. за період з січня 1993 по грудень 2013 року	USD, EUR і NZD на період 60 днів
24	NASDAQ за період з січня 1990 по грудень 2012 року	S&P 500 і ціни на акції Intel за період з січня 1990 по грудень 2012 року	USD, EUR і NZD на період 90 днів

Індивідуальне завдання прогнозування цін на акції компанії регресійними моделями та Facebook Prophet

Таблиця 16

Варіанти та дані для індивідуального завдання

Варіант	Показник або індикатор для прогнозування	Моделі
1	Ціни на акції компанії Adobe на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
2	Ціни на акції компанії Apple на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
3	Ціни на акції компанії AT&T на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
4	Ціни на акції компанії Autodesk на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
5	Ціни на акції Bank of America на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
6	Ціни на акції Avery Dennison на 5 днів 2024 року	1) Регресійні Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
7	Ціни на акції компанії Baxter International на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
8	Ціни на акції компанії Boeing на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
9	Ціни на акції компанії Caterpillar Inc. на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
10	Ціни на акції компанії Chevron Corporation на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet

11	Ціни на акції компанії Cisco на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
12	Ціни на акції компанії IBM на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
13	Ціни на акції компанії Coca-Cola Company на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
14	Ціни на акції компанії Colgate-Palmolive на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
15	Ціни на акції компанії CSX на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
16	Ціни на акції компанії Disney на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
17	Ціни на акції компанії Duke Energy на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
18	Ціни на акції компанії Edison International на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
19	Ціни на акції компанії ExxonMobil на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
20	Ціни на акції компанії Johnson & Johnson на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
21	Ціни на акції компанії FedEx на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
22	Ціни на акції компанії Goldman Sachs на 5 днів 2024 року	1) Регресійні 2) Тренд

		3) Гетероскедастичні 4) Facebook Prophet
23	Ціни на акції компанії HP Inc. на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet
24	Ціни на акції компанії Intel на 5 днів 2024 року	1) Регресійні 2) Тренд 3) Гетероскедастичні 4) Facebook Prophet

Індивідуальне завдання для власної задачі прогнозування

Мета роботи: навчитися застосовувати сучасні методи прогнозування (методи ІАД, мультифакторні моделі прогнозування виживання, регресійні моделі зі змінною волатильністю (УАРУГ, АРІКС, ЕУАРУГ, ЧІУАРУГ, ЧІЕУАРУГ, МСВ тощо), відновлення пропущених значень ознак) для вирішення задач аналізу даних та оцінювання ризиків різної природи.

Порядок виконання роботи:

1. Постановка власної задачі моделювання.

2. Виконати пошук даних та формування файлу навчальних даних. Дані повинні задовольняти наступним вимогам:

- Обсяг даних – щонайменше 1000 записів;
- Наявність причинно-наслідкового зв’язку між компонентами даних.

3. Розбити навчальні дані на навчальну та перевірочну вибірку (крос-валідація). Для виконання завдання можна використовувати RStudio, Python, SAS EM/SAS EG, тощо).

4. За навчальною вибіркою побудувати структури моделей різними методами інтелектуального аналізу даних (логістична регресія, регресійні моделі, нейронні мережі, мережі Байєса, дерева рішень тощо) і обрати кращу з них.

4.1. За перевірочною вибіркою оцінити якість моделей.

4.2. Покращити якість побудованих моделей.

4.3. Обрати кращу модель

5. За власним вибором побудувати моделі інтелектуального аналізу даних, мультифакторні моделі прогнозування виживання (Кокса, Каплан-Майєра на основі статистичних критеріїв: LogRank, Wilcoxon, Likelihood Ratio; непропорційні) та відновлення пропущених даних (в задачах розпізнавання та кластеризації).

6. Порівняти побудовані моделі та зробити висновки щодо доцільності використання прикладних методів прогнозування та типів задач і прогнозних змінних, в яких це доцільно виконувати.

7. Оформити звіт.

8. Захист роботи (при наявності кодів програми, вибірки даних та протоколу).

Запитання для самоконтролю

1. Які моделі дозвляють оцінити волатильність фінансових процесів?
2. Чи спостерігаються певні тренди на обраному Вами наборі даних?
3. Які дані найкраще моделювати за допомогою моделі Facebook Prophet?
4. Як визначається квантиль розподілу?
5. Як застосувати регресійні моделі для оцінювання ризиків за методологією VaR?
6. Якими моделями доцільно прогнозувати курси валют?

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Бідюк П.І., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів: навч. посіб. Київ: НТУУ «КПІ», 2013. 600 с.
2. Бідюк П. І., Коршевнюк Л. О., Кузнєцова Н. В. Моделі і методи прикладної статистики: навч. посіб. з грифом МОН України. Київ: НТУУ «КПІ», 2014. 722 с.
3. Кузнєцова Н.В., Бідюк П.І. Моделі і методи коротко- і середньострокового прогнозування фінансових процесів: навч. посіб. Київ : КПІ ім. Ігоря Сікорського, 2023. 334 с.
4. Лук’яненко І. Г. Аналіз часових рядів. Частина друга : побудова VAR і VECM моделей з використаням пакета E.Views 6.0 : практичний посібник для роботи в комп'ютерному класі. К.: [НаУКМА], 2013. 174 с.
5. Bidyuk P. I., Konovalyuk M. M., Kuznetsova N.V., Pudlo I. V. Adaptive Short-Term Forecasting of Selected Financial Processes. Research bulletin of NTUU “KPI”. 2014. N1. P.35–41.
6. Chen B., Gel Y. R., Balakrishna. N., Bovas A. Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. 2011. 25p. URL: http://www.academia.edu/21503708/
7. Кузнєцова Н.В., Бідюк П.І. Теорія і практика аналізу фінансових ризиків: системний підхід: монографія. Київ: Видавництво Ліра-К, 2020. 400 с.
8. Kaggle [Електронний ресурс]. URL: https://www.kaggle.com/datasets

9. Yahoo Finance [Електронний ресурс]. URL: *https://finance.yahoo.com/*

10. Hodeghatta U. R., Nayak U. Business Analytics Using R – A Practical Approach. New York: Apress, 2017. 280 p.

11. N. Kuznietsova and P. Bidyuk, "Heteroskedasticity Models for Financial Processes Modelling and Forecasting," *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, Lviv, Ukraine, 2020, pp. 310-315, doi: 10.1109/DSMP47368.2020.9204313.

12. [Електронний ресурс] URL: https://github.com/gumdropsteve/intro_to_prophet

13. Kuznietsova, N., Bidyuk, P., Kuznietsova, M. Data Mining Methods, Models and Solutions for Big Data Cases in Telecommunication Industry// In: Lecture Notes on Data Engineering and Communications Technologies, 2022, 77, pp. 107–127. https://link.springer.com/chapter/10.1007/978-3-030-82014-5_8. DOI 10.1007/978-3-030-82014-5_8.

14. Robson, Winston A. “The Math of Prophet — Future Vision.” Medium, Future Vision, 17 June 2019, URL: https://medium.com/future-vision/the-math-of-prophet-46864fa9c55a

15. Robson, Winston A. “The Prophet on Walmart — Comprehensive Intro to FbProphet.” *Medium*, Future Vision, 9 July 2019, URL: https://medium.com/future-vision/intro-to-prophet-9d5b1cbd674e

16. Taylor SJ, Letham B. 2017. Forecasting at scale. Peer J Preprints 5:e3190v2. URL: https://doi.org/10.7287/peerj.preprints.3190v2

17. Gill J. Generalized linear models: a unified approach. USA: New Delli, 2001. 110 p.

18. Brown, Violet. (2021). An Introduction to Linear Mixed-Effects Modeling in R. Advances in Methods and Practices in Psychological Science. 4. 251524592096035. 10.1177/2515245920960351.

19. Brady et al. 2014: Brady T. West, Kathleen B. Welch, Andrzej T. Galecki, Linear Mixed Models: A Practical Guide Using Statistical Software (2nd Edition), 2014.

20. Christine R. Wells (2021) SAS for Mixed Models: Introduction and Basic Applications, The American Statistician, 75:2, 231, DOI:10.1080/00031305.2021.1907997.