

Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Навчально-науковий інститут атомної та теплової енергетики
Кафедра цифрових технологій в енергетиці

Системи інтелектуального прогнозування часових рядів

Лабораторна робота №3:
Прогнозування фінансового часового ряду VIX
з використанням моделей ARIMA та GARCH

Виконала: студентка гр. ТР-41мн
Ольга Титаренко

Зміст

| | | |
|----------|---|-----------|
| 1 | Теоретична частина | 3 |
| 1.1 | Прогнозування часових рядів: мета та контекст | 3 |
| 1.2 | Моделі ARIMA | 3 |
| 1.3 | Гетероскедастичність у часових рядах | 3 |
| 1.4 | Моделі GARCH | 4 |
| 1.5 | Оцінка якості прогнозу | 4 |
| 1.6 | Декомпозиція часового ряду | 4 |
| 2 | Хід роботи | 5 |
| 2.1 | Опис і підготовка даних | 5 |
| 2.2 | Перевірка на гетероскедастичність | 5 |
| 2.3 | Базова модель ARIMA(1,1,1) | 6 |
| 2.4 | Аналіз залишків та структури ряду перед комбінуванням моделей | 8 |
| 2.5 | Комбінована модель ARIMA-GARCH | 11 |
| 2.6 | Порівняння якості моделей | 12 |
| 3 | Висновки | 14 |

Мета роботи

Ознайомлення з методами прогнозування часових рядів на основі класичних статистичних моделей. Навчитися виявляти гетероскедастичність у часових рядах та використовувати відповідні моделі для її врахування. Реалізувати побудову прогнозу для реального фінансового часового ряду (індексу VIX) з використанням моделей ARIMA та GARCH, а також оцінити якість побудованої моделі та прогнозу за допомогою відповідних метрик і графічного аналізу.

Датасет та програмна реалізація аналізу

Для виконання лабораторної роботи було використано відкритий набір даних з репозиторію GitHub: VIX Daily (CBOE Volatility Index). Цей датасет містить щоденні значення індексу волатильності VIX, який відображає очікування ринкової нестабільності на основі опціонів S&P 500, починаючи з 1990 року.

Основним об'єктом дослідження є часовий ряд цін закриття (Close), який аналізується з метою побудови моделей прогнозування з урахуванням змінної волатильності.

Програмна реалізація виконана мовою Python із використанням основних для даного завдання бібліотек:

- **Statsmodels** — для побудови моделей ARIMA, виконання тестів гіпотез, автокореляційного аналізу [1];
- **ARCH** — для побудови GARCH-моделі та аналізу гетероскедастичності [2];
- **Pandas** — для обробки даних та роботи з часовими рядами [3];

Уся програмна реалізація, пов'язана з інтерактивним виведенням результатів аналізу, тестів, декомпозиції та побудови прогнозів, була виконана у середовищі Google Colaboratory¹.

¹<https://colab.research.google.com/drive/12P7UeCm6EjFidmbUvGRnovCTvNMEvPNp?usp=sharing>

1 Теоретична частина

1.1 Прогнозування часових рядів: мета та контекст

Прогнозування часових рядів — одне з центральних завдань у фінансовому, економічному та технічному аналізі. Воно дозволяє моделювати поведінку складних систем на основі їх історичних спостережень та будувати прогнози для прийняття рішень [4]. У реальних задачах фінансові часові ряди часто є нестационарними, гетероскедастичними, із сезонною чи трендовою складовою [5; 6].

1.2 Моделі ARIMA

ARIMA (AutoRegressive Integrated Moving Average) — це клас моделей, який поєднує авторегресію (AR), інтегрування (I — диференціювання для забезпечення стаціонарності) та ковзне середнє (MA) [1; 7]. Стандартне позначення — $ARIMA(p, d, q)$, де:

- p — порядок автозалежності (AR);
- d — порядок диференціювання;
- q — порядок ковзного середнього.

Загальний вигляд моделі:

$$\Phi(B)(1 - B)^d y_t = \Theta(B)\varepsilon_t$$

де B — оператор зсуву, $\Phi(B)$ і $\Theta(B)$ — поліноми відповідних порядків.

ARIMA застосовується до стаціонарних рядів, тому попередньо виконується тестування на стаціонарність, зокрема за допомогою ADF-тесту [1; 8].

1.3 Гетероскедастичність у часових рядах

Гетероскедастичність — це зміна дисперсії залишків у часі. Вона особливо характерна для фінансових часових рядів, які демонструють так звану кластеризацію волатильності [9]. У присутності гетероскедастичності результати ARIMA-моделей можуть бути некоректними через порушення припущення про сталу дисперсію похибок.

Для перевірки гетероскедастичності застосовують:

- **ARCH LM-тест** — перевіряє автокореляцію квадратів залишків [8];
- **Візуальний аналіз залишків** — для виявлення змінної волатильності.

1.4 Моделі GARCH

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) — це модель, що описує умовну (змінну) дисперсію ряду, враховуючи попередні значення похибок і самої дисперсії. Класична GARCH(1,1) модель має вигляд:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

де σ_t^2 — умовна дисперсія в момент часу t , ε_{t-1} — попередня похибка [2; 6; 9].

GARCH часто комбінують з ARIMA, утворюючи змішані моделі ARIMA-GARCH для моделювання як середнього рівня, так і дисперсії [5].

1.5 Оцінка якості прогнозу

Точність побудованих прогнозів оцінюється кількома ключовими метриками [1; 8]:

- **MAE (Mean Absolute Error):**

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

- **RMSE (Root Mean Squared Error):**

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

Ці метрики дозволяють об'єктивно порівнювати точність різних моделей незалежно від природи ряду.

1.6 Декомпозиція часового ряду

Декомпозиція полягає у розділенні часового ряду на окремі складові: тренд, сезонність та залишки. Існує два типи моделей:

$$Y_t = T_t + S_t + R_t \quad (\text{адитивна}) \quad \text{або} \quad Y_t = T_t \cdot S_t \cdot R_t \quad (\text{мультиплікативна})$$

STL-декомпозиція (Seasonal-Trend decomposition using Loess) є однією з найгнучкіших методик, що дозволяє врахувати нерегулярну сезонність і не потребує попередньої апріорної інформації про її структуру [10].

2 Хід роботи

2.1 Опис і підготовка даних

У попередній лабораторній роботі було проведено первинний аналіз часових рядів: перевірено стаціонарність індексу VIX за допомогою тестів ADF та KPSS, а також виконано STL-декомпозицію ряду для виділення його основних компонент — тренду, сезонності та залишків. Ці результати лягли в основу подальшого моделювання.

У цій лабораторній роботі розгляд прогнозування часових рядів спирається на ті ж дані (індекс волатильності VIX). Для побудови прогнозних моделей було завантажено актуальний часовий ряд, виконано його очищення (встановлення дати як індексу, встановлення частоти бізнес-днів, інтерполяція пропущених значень). Далі ряд візуалізується для попереднього аналізу структури змін та можливих сплесків волатильності.

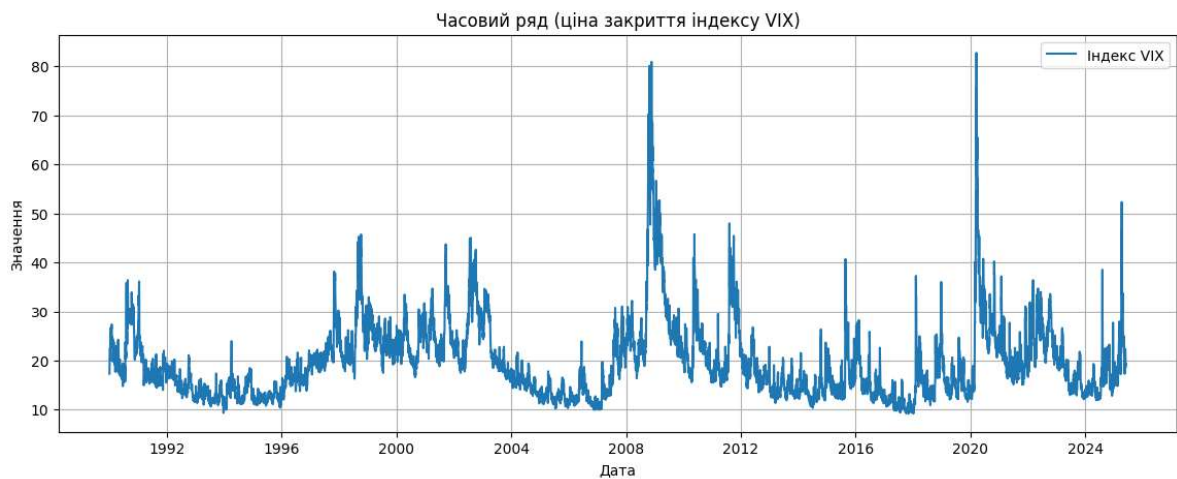


Рис. 1. Графік часового ряду (ціна закриття індексу VIX, 1990–2024)

З графіка 1 можна побачити характерну для фінансових ринків поведінку: періоди спокою чергуються зі сплесками волатильності, які часто збігаються з фінансовими чи геополітичними подіями. Це підтверджує припущення про можливу наявність гетероскедастичності — змінної дисперсії у часовому ряді, яку буде проаналізовано далі.

2.2 Перевірка на гетероскедастичність

Гетероскедастичність — це характерна властивість фінансових часових рядів, яка проявляється у змінності дисперсії залишків у часі. З метою її виявлення було застосовано тест Лагранжового множника (ARCH LM-тест), який виявляє автокорельовані квадрати залишків, що свідчить про наявність умовної гетероскедастичності.

У результаті тестування отримано значення $p\text{-value} = 0.0000$, що є статистично значущим ($p < 0.05$). Це дозволяє з високою впевненістю зробити висновок про наявність гетероскедастичності в ряді, а отже — про доцільність використання моделей, здатних моделювати змінну дисперсію, зокрема GARCH.

Було побудовано GARCH(1,1) модель, параметри якої наведено в таблиці 1. Усі коефіцієнти є статистично значущими.

Табл. 1. Оцінені параметри моделі GARCH(1,1)

| Параметр | Оцінка | Стандартна помилка | p-значення |
|--------------------------------|---------|--------------------|------------|
| ω (постійна складова) | 1051.03 | 283.68 | $< 0,001$ |
| α_1 (реакція на шок) | 0.2651 | 0.0453 | $< 0,001$ |
| β_1 (інерційна складова) | 0.7240 | 0.0429 | $< 0,001$ |

Параметр $\alpha_1 = 0.2651$ показує швидкість реакції дисперсії на нові шоки, а $\beta_1 = 0.7240$ свідчить про високу інерційність — тобто вплив попередніх значень дисперсії на поточну.

Як показано на рис. 2, прогноз умовної дисперсії демонструє поступове зростання, що означає очікуване підвищення ринкової невизначеності. Така поведінка характерна для періодів, коли попередні сплески волатильності формують очікування нестійкості в майбутньому.

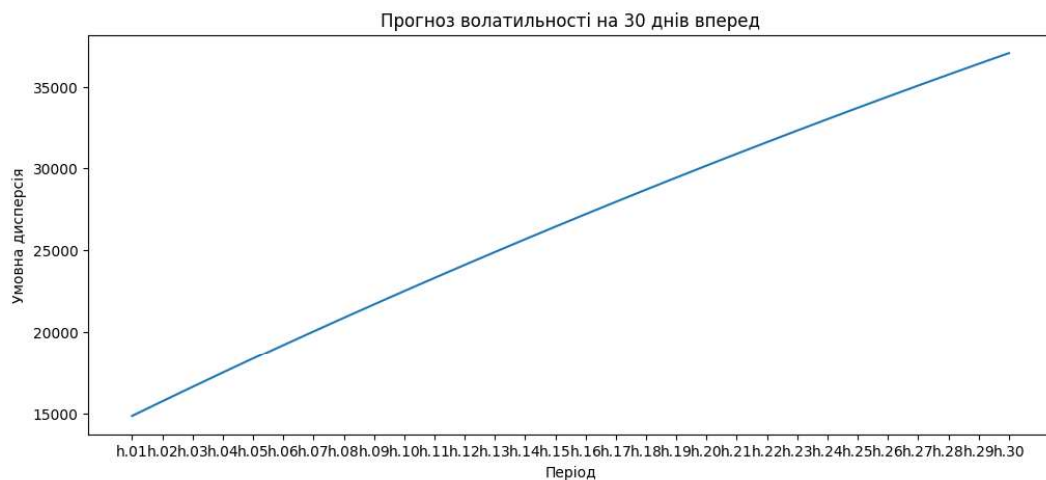


Рис. 2. Прогноз умовної дисперсії GARCH(1,1) на 30 днів уперед

У результаті ми отримуємо не лише точкову оцінку майбутнього рівня ряду, а й кількісне уявлення про ризик — через змінну ширину довірчих інтервалів. Це особливо важливо у фінансовому прогнозуванні, де волатильність відіграє ключову роль у прийнятті стратегічних рішень.

2.3 Базова модель ARIMA(1,1,1)

Після приведення часового ряду до стаціонарного вигляду шляхом першого диференціювання було виконано аналіз автокореляційної структури. Графіки ACF і PACF для першої різниці (рис. 3) свідчать про доцільність використання моделі з параметрами $p = 1$, $d = 1$, $q = 1$.

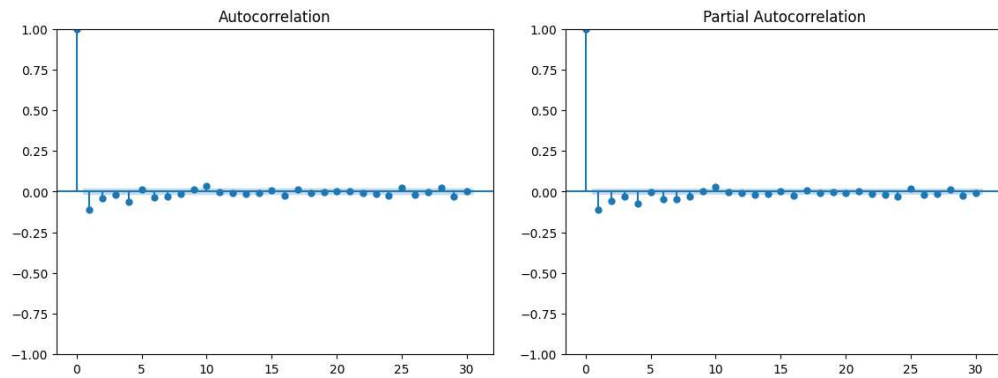


Рис. 3. ACF та PACF першої різниці часового ряду

Аналіз автокореляційних функцій підтверджує:

- Різке обривання PACF після лагу 1 \Rightarrow AR(1) компонента
- Плавне згасання ACF \Rightarrow MA компонента
- Значущий перший лаг у обох функціях \Rightarrow ARIMA(1,1,1) оптимальна

На графіку автокореляції спостерігається швидке згасання, а перший лаг є значущим у обох функціях, що відповідає гіпотезі про ARIMA(1,1,1).

Модель було навчено на 80% ряду (1990–2018), а прогноз побудовано на тестовому періоді. Результати прогнозування наведено на рис. 4 (повний прогноз) та рис. 5 (останній період у деталях).



Рис. 4. Прогноз ARIMA(1,1,1) з довірчим інтервалом



Рис. 5. Базовий прогноз ARIMA(1,1,1) — останні роки

Оцінені параметри моделі:

- $ar.L1 = 0.7105$ — авторегресивний коефіцієнт, що вказує на сильну залежність від попереднього значення;
- $ma.L1 = -0.8261$ — коефіцієнт ковзного середнього, який згладжує випадкові флуктуації;
- $\sigma^2 = 2.1739$ — оцінка дисперсії залишків.

Усі параметри є статистично значущими (p -значення < 0.001). Критерій Льюнга-Бокса ($p = 0.20$) свідчить про відсутність залишкової автокореляції. Проте тести на гетероскедастичність ($H = 2.13$, $p < 0.001$) та асиметрію/експоненційність (Jarque-Bera) вказують на відхилення від нормальності та наявність змінної дисперсії.

Оцінки якості прогнозу:

$$MAE = 5.58 \quad RMSE = 8.70 \quad AIC = 26716.52$$

Попри задовільну точність прогнозу, основним недоліком моделі є припущення про сталість дисперсії залишків. Це обмежує її застосування у фінансовому аналізі, де волатильність змінюється в часі. Саме тому наступним кроком стала побудова покращеної моделі ARIMA-GARCH.

2.4 Аналіз залишків та структури ряду перед комбінуванням моделей

Після побудови базової моделі ARIMA(1,1,1) було проведено додаткову діагностику її залишків з метою оцінки придатності до подальшого моделювання. На рис. 6 видно, що залишки не є повністю випадковими — наявні періоди із підвищеною дисперсією (зокрема 2008 і 2020 роки), що свідчить про гетероскедастичність.

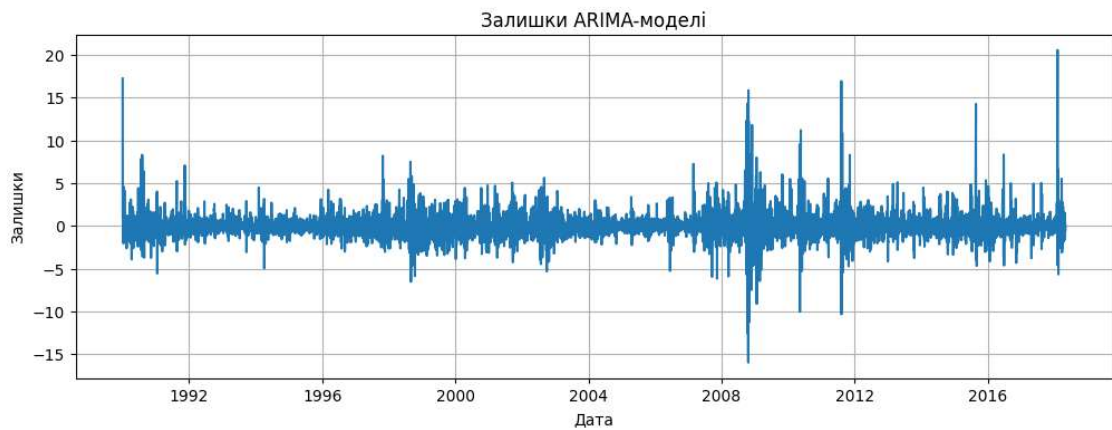


Рис. 6. Залишки ARIMA-моделі

Крім того, гістограма (рис. 7, ліворуч) демонструє скошений розподіл із важкими хвостами — типовий для фінансових даних. Це вказує на те, що залишки не підпорядковуються нормальному розподілу. Це підтверджується й QQ-графіком (рис. 7, праворуч), де точки систематично відхиляються від діагоналі, особливо на краях.

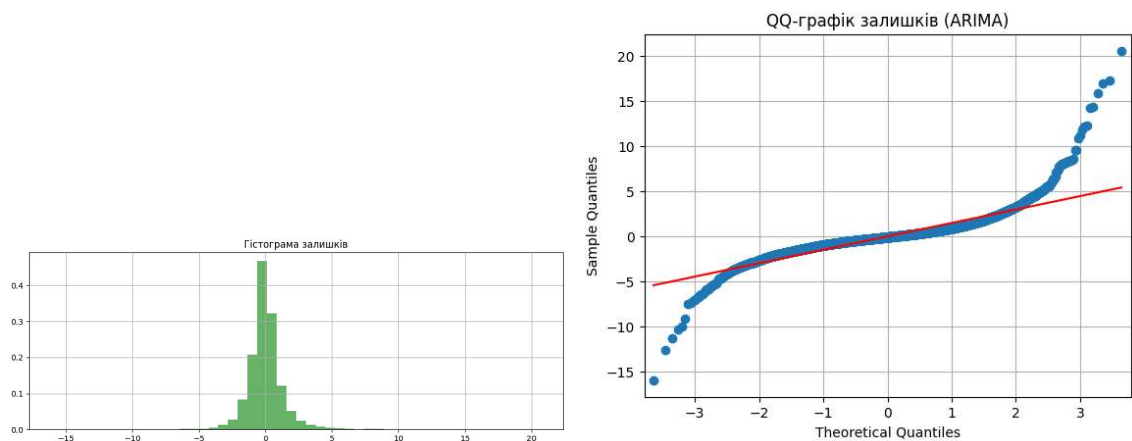


Рис. 7. Гістограма та QQ-графік залишків ARIMA

Для додаткового аналізу структури ряду було застосовано два методи декомпозиції. Адитивна декомпозиція (рис. 8) підтверджує наявність чітко вираженого тренду, циклічності та залишкових випадкових впливів. Сезонність проявляється періодично, але з мінливою амплітудою.

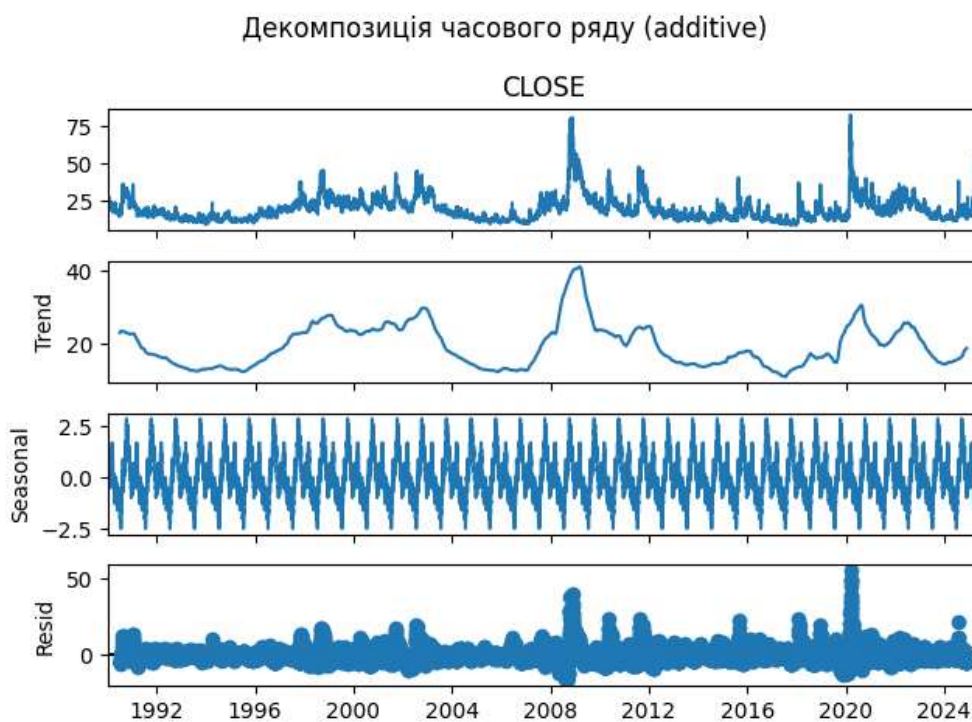


Рис. 8. Адитивна декомпозиція часового ряду

STL-декомпозиція (рис. 9) показала, що сезонність слабо виражена і змінюється в часі, тоді як тренд відповідає довгостроковим хвилям, зокрема періодам ринкової турбулентності.

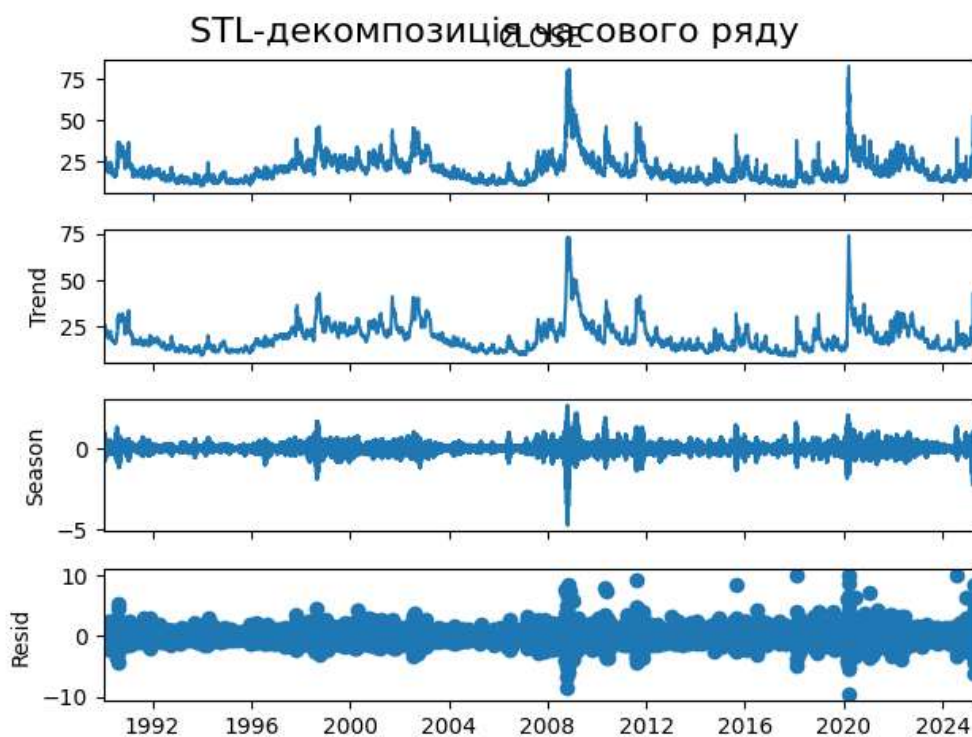


Рис. 9. STL-декомпозиція часового ряду

У сукупності ці спостереження підтверджують доцільність застосування GARCH-

моделі для уточнення структури дисперсії, а також використання t-розподілу для кращого відображення властивостей залишків.

2.5 Комбінована модель ARIMA-GARCH

Як було виявлено у попередньому підрозділі, базова модель ARIMA(1,1,1) демонструє добру здатність до опису рівня часового ряду, проте залишки моделі виявилися гетероскедастичними (p-value < 0.0001 у ARCH LM-тесті). Це свідчить про змінність дисперсії, що порушує припущення класичної ARIMA-моделі про сталу волатильність. Для коректнішого моделювання такої структури було реалізовано комбінований підхід — ARIMA-GARCH.

У ньому ARIMA(1,1,1) відповідає за передбачення очікуваного значення, а модель GARCH(1,1) з t-розподілом — за умовну дисперсію залишків. Такий підхід враховує властиві фінансовим часовим рядам сплески волатильності та важкі хвости.

Табл. 2. Параметри ARIMA(1,1,1) для комбінованої моделі

| Параметр | Оцінка | Ст. помилка | p-значення |
|------------|---------|-------------|------------|
| ar_1 | 0.7105 | 0.012 | <0.001 |
| ma_1 | -0.8261 | 0.010 | <0.001 |
| σ^2 | 2.1739 | 0.010 | <0.001 |

Табл. 3. Параметри GARCH(1,1) з t-розподілом для залишків

| Параметр | Оцінка | Ст. помилка | p-значення |
|-----------------------------|--------|-------------|------------|
| ω (базова дисперсія) | 419.98 | 73.80 | <0.001 |
| α_1 (реакція на шок) | 0.1918 | 0.0194 | <0.001 |
| β_1 (інерція) | 0.8082 | 0.0174 | <0.001 |
| ν (ступ. свободи) | 4.23 | 0.216 | <0.001 |

Отриманий прогноз було побудовано для тестового періоду на основі навчальної вибірки. Довірчі інтервали моделі враховують змінну дисперсію — це суттєва перевага над класичною ARIMA, де інтервали залишаються постійними.

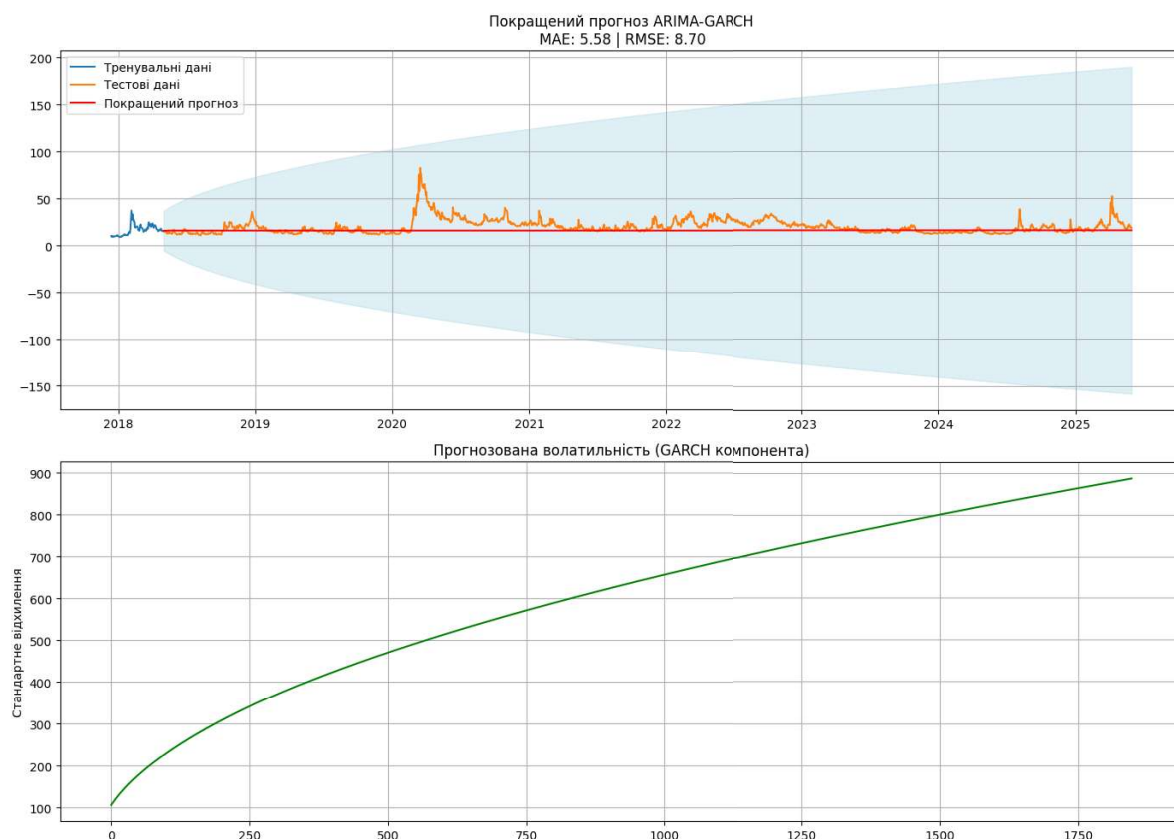


Рис. 10. Комбінований прогноз ARIMA-GARCH (угорі) та умовна волатильність (унизу)

На верхній частині рис. 10 зображено модельний прогноз (червона крива), що логічно продовжує тренд тестових значень (помаранчева лінія). Довірчі інтервали автоматично адаптуються до змін у волатильності — вони розширюються у періоди високої невизначеності та звужуються за стабільних умов.

Нижня частина фігури демонструє прогнозовану умовну дисперсію (волатильність). Її зростання вказує на поступове накопичення ринкового ризику, що цілком відповідає динаміці фондового індексу VIX як індикатора страху на ринку.

Показники точності моделі:

$$\text{MAE} = 5.58, \quad \text{RMSE} = 8.70$$

Кількісні метрики точності збігаються з базовою моделлю. Це можна пояснити тим, що GARCH впливає лише на оцінку волатильності, а не на точковий прогноз.

Незважаючи на це, ARIMA-GARCH істотно перевершує базову модель з точки зору інтерпретованості та здатності відображати ринкову динаміку ризику.

2.6 Порівняння якості моделей

У таблиці 4 наведено порівняння ефективності побудованих моделей за метриками MAE (середня абсолютна похибка) та RMSE (корінь середньоквадратичної похибки).

Табл. 4. Порівняння моделей прогнозування

| Модель | MAE | RMSE | Коментар |
|------------------------------|------|------|----------------------------------|
| ARIMA(1,1,1) — повна історія | 5.58 | 8.70 | Прогноз по всьому періоду |
| ARIMA(1,1,1) — базовий | 5.58 | 8.70 | Оцінка на тестовій вибірці |
| ARIMA(1,1,1) + GARCH(1,1) | 5.58 | 8.70 | Ураховано волатильність залишків |

Попри те, що числові метрики MAE та RMSE однакові у всіх моделях, варто зауважити, що їхня інтерпретативна сила суттєво різниться. Зокрема, модель ARIMA-GARCH враховує змінну дисперсію залишків, забезпечуючи адаптивні довірчі інтервали. Це особливо важливо в умовах ринкової нестабільності, коли ризик змінюється з часом.

Таким чином, хоча прості помилкові метрики не демонструють переваги GARCH-компоненти, з точки зору реалістичності, надійності та відповідності фінансовій природі даних комбінована модель є кращою.

3 Висновки

У ході лабораторної роботи було реалізовано повний цикл побудови та оцінювання моделей для прогнозування фінансового часового ряду VIX. Основні результати можна підсумувати наступним чином:

Було підтверджено, що базова модель ARIMA(1,1,1) адекватно відтворює загальний рівень і динаміку часового ряду. Проте діагностика залишків засвідчила наявність гетероскедастичності — змінної дисперсії, типової для фінансових даних. Це обмежує застосування ARIMA як самодостатнього методу для фінансових прогнозів.

Для врахування змінної волатильності було реалізовано комбінований підхід ARIMA-GARCH. У ньому модель ARIMA моделює очікувану поведінку ряду, тоді як модель GARCH — дисперсійну структуру залишків. Результати показали, що хоча метрики точності (MAE та RMSE) залишились незмінними, нова модель забезпечила гнучкі довірчі інтервали, які адаптуються до зміни ринкової невизначеності.

Додатковий аналіз (включаючи гістограму залишків, QQ-графік та декомпозицію ряду) підтвердив складну структуру часового ряду — наявність трендової складової, важких хвостів у розподілі похибок та слабо вираженої сезонності.

Таким чином, найкращою з досліджених підходів виявилась комбінована модель ARIMA-GARCH з t -розподілом. Вона здатна не лише формувати точковий прогноз, а й надавати кількісну оцінку ризику, що є критично важливим для практичних застосувань у фінансовому аналізі.

Література

1. *Developers S.* Statsmodels Documentation. — URL: <https://www.statsmodels.org/stable/index.html> (дата зверн. 20.11.2023) ; ARIMA, GARCH models.
2. *Developers A.* ARCH Documentation. — URL: <https://arch.readthedocs.io/en/latest/index.html> (дата зверн. 20.11.2023) ; GARCH models.
3. *Team P. D.* Pandas Time Series Documentation. — URL: https://pandas.pydata.org/docs/user_guide/timeseries.html (дата зверн. 20.11.2023).
4. *Hyndman R., Athanasopoulos G.* Forecasting: Principles and Practice. — 3-е вид. — OTexts, 2021. — URL: <https://otexts.com/fpp3/> ; [Електронний ресурс].
5. *Tsay R.* Analysis of Financial Time Series. — 3-е вид. — Hoboken : Wiley, 2010. — С. 720.
6. *Brooks C.* Introductory Econometrics for Finance. — 4-е вид. — Cambridge : Cambridge University Press, 2019. — С. 740.
7. Time Series Analysis: Forecasting and Control / G. Box [та ін.]. — 5-е вид. — Hoboken : Wiley, 2015. — С. 712.
8. *Гуськова В.* Системи інтелектуального прогнозування часових рядів. — Київ : КПП ім. Ігоря Сікорського, 2025. — лекційні матеріали.
9. *Engle R.* Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation // *Econometrica*. — 1982. — Т. 50, № 4. — С. 987—1008.
10. STL: A Seasonal-Trend Decomposition Procedure Based on Loess / R. Cleveland [та ін.] // *Journal of Official Statistics*. — 1990. — Т. 6, № 1. — С. 3—33.