

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 №2:
Перевірка стаціонарності та
декомпозиція часового ряду**

Виконала: студентка гр. ТР-41мн
Ольга Титаренко

Зміст

1	Теоретична частина	3
1.1	Стаціонарність: суть та мета перевірки	3
1.2	Декомпозиція часового ряду: ідея та значення	3
2	Хід роботи	5
2.1	Візуалізація часового ряду	5
2.2	Перевірка на стаціонарність — тест ADF	5
2.3	Додаткова перевірка — тест KPSS	6
2.4	Декомпозиція часового ряду	6
3	Висновки	8

Мета роботи

Дослідити властивості часового ряду, перевірити його стаціонарність за допомогою статистичних тестів, а також виконати декомпозицію для виділення трендової, сезонної та залишкової складових.

Датасет та програмна реалізація аналізу

Для виконання лабораторної роботи було обрано датасет "US Stock Market 2020-2024" з Kaggle, який містить щоденні дані про акції та індекси фондового ринку США за період з 2020 по 2024 рік. Для аналізу буде використовуватися закриття торгів (Close) як основний часовий ряд.

Програмна реалізація виконана на мові Python з використанням бібліотек:

- **Pandas** для обробки та аналізу даних;
- **Statsmodels** для статистичного моделювання та тестування;
- **Matplotlib** та **Seaborn** для візуалізації результатів.

Уся програмна реалізація, пов'язана з попередньою обробкою даних, побудовою графіків, проведенням статистичних тестів та декомпозицією часового ряду, була виконана у середовищі Google Colaboratory¹.

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

1 Теоретична частина

Часовий ряд — це впорядкована хронологічно послідовність числових значень, яка виникає внаслідок вимірювання певного процесу. Задача аналізу часового ряду полягає у виявленні його структури, тенденцій, циклічних змін, сезонних ефектів і випадкових коливань.

1.1 Стационарність: суть та мета перевірки

Стационарність — це властивість часового ряду, яка означає, що його основні статистичні характеристики (середнє, дисперсія, автокореляція) залишаються незмінними в часі. Стационарні ряди простіші для аналізу, оскільки підлягають класичним методам математичного моделювання, зокрема ARIMA, і дозволяють будувати точні прогнози.

У нестационарному ряду середнє або варіація змінюються з часом. Такий ряд потрібно перетворити в стационарний, зокрема шляхом диференціювання.

Наступні тестування можуть бути використані для перевірки:

- *ADF-тест* (*Augmented Dickey–Fuller*) перевіряє нульову гіпотезу про наявність однічного кореня (тобто нестационарності). Якщо $p\text{-value} < 0.05$, ряд вважається стационарним.
- *KPSS-тест* — навпаки, перевіряє гіпотезу про стационарність. Якщо $p\text{-value} > 0.05$, гіпотеза не відхиляється і ряд вважається стационарним.

Перевірка на стационарність дозволяє:

- опінити придатність ряду до моделювання;
- вибрати правильну модель (наприклад, ARIMA vs SARIMA);
- уникнути хибних висновків через зміну структури ряду в часі.

1.2 Декомпозиція часового ряду: ідея та значення

Декомпозиція — це розкладання часового ряду на окремі складові:

- **Тренд (Trend)** — довготривала тенденція зростання або зниження, яка відображає глобальні зміни;
- **Сезонність (Seasonal)** — періодичні коливання з фіксованим інтервалом;
- **Залишкова складова (Residual)** — випадкові впливи, які не можна пояснити іншими компонентами.

Принцип роботи: Модель декомпозиції припускає, що часовий ряд можна подати у вигляді:

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

де T_t — тренд, S_t — сезонна складова, R_t — залишки.

Значення:

- Виявлення тренду допомагає зрозуміти довготривалі зміни.
- Сезонна складова дозволяє адаптувати моделі до періодичних закономірностей.
- Оцінка залишків дозволяє виявити аномалії, викиди або інші випадкові впливи.

Таким чином, декомпозиція є важливим етапом аналізу часового ряду та передує моделюванню й прогнозуванню.

2 Хід роботи

2.1 Візуалізація часового ряду

На графіку нижче показано динаміку індексу волатильності VIX за період з 1990 по 2024 роки.

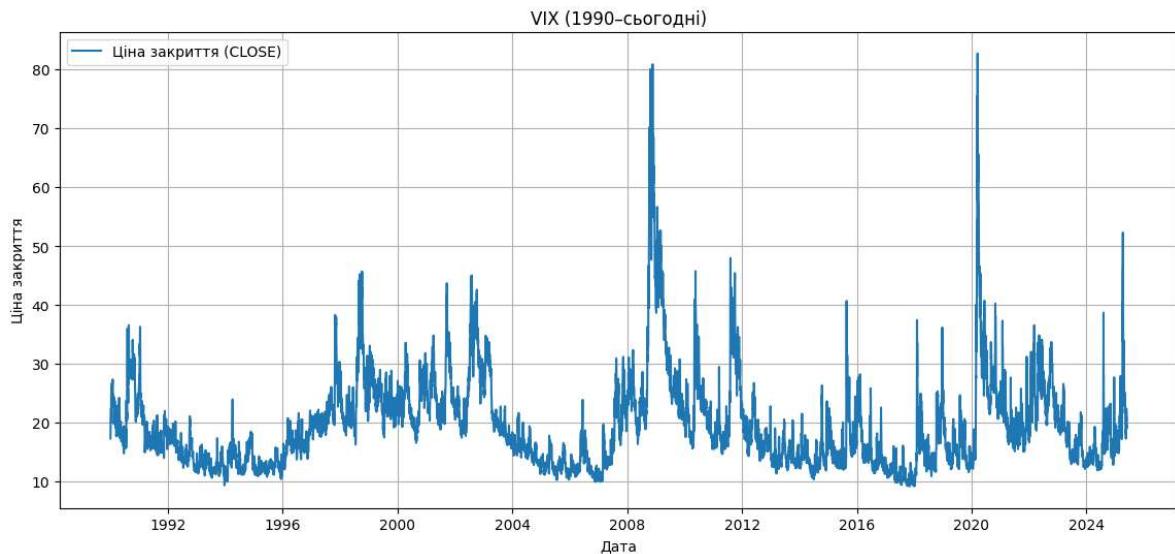


Рис. 1. Графік часового ряду (ціна закриття індексу VIX)

Отриманий графік характеризується періодами високої волатильності (кризи 2008 та 2020 років), коливанням навколо змінного середнього значення, та відсутністю явного лінійного тренду (тобто середнє коливається, але не зростає/спадає постійно).

Це вже може свідчити про стаціонарність, але потрібне формальне підтвердження.

2.2 Перевірка на стаціонарність — тест ADF

У результаті проведення тексту Дікі-Фуллера отримано наступні результати:

ADF статистика: -7.2031

p-value: 2.34e-10

Критичні значення:

1%: -3.431

5%: -2.862

10%: -2.567

Оскільки $p\text{-value} < 0.05$ та ADF статистика менша за всі критичні значення, відхиляємо нульову гіпотезу. Отже, ряд — **стаціонарний**.

2.3 Додаткова перевірка — тест KPSS

Задля того аби впевнитись у стаціонарності ряду, проведемо додатковий тест KPSS. Цей тест перевіряє нульову гіпотезу про стаціонарність. Результати тесту KPSS:

KPSS статистика: 0.3477

p-value: 0.0997

Критичні значення:

10%: 0.347

5%: 0.463

2.5%: 0.574

1%: 0.739

Як бачимо, p-value > 0.05, тому ми не відхиляємо нульову гіпотезу. Це ще раз підтверджує, що ряд є **стаціонарним**.

2.4 Декомпозиція часового ряду

Було виконано адитивну декомпозицію, що дозволила виділити тренд, сезонність і залишки.

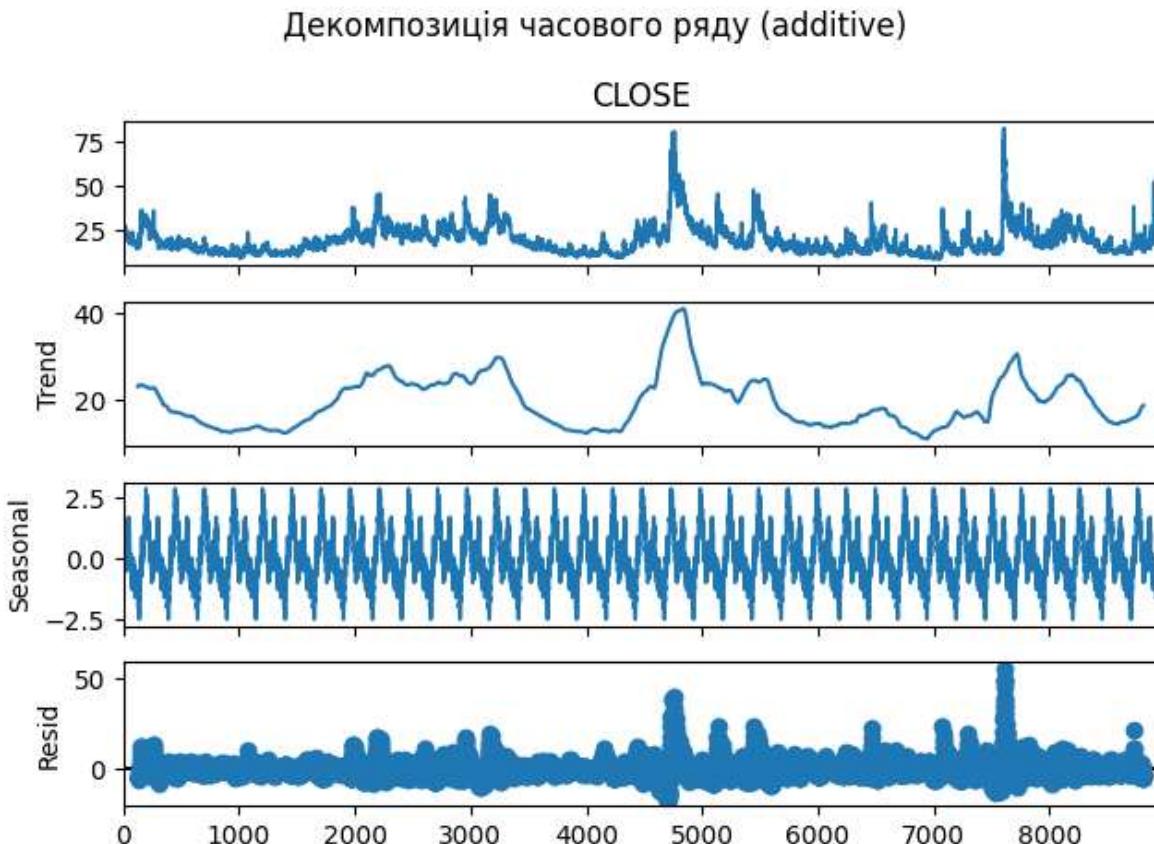


Рис. 2. Декомпозиція часового ряду (Trend, Seasonal, Residual)

На графіку декомпозиції часового ряду (рис. 2) спостерігаються наступні ключові елементи:

- **Трендова складова (Trend)** демонструє загальний напрямок зміни індексу волатильності VIX у довгостроковій перспективі. Видно періоди стабільного зниження і зростання, що відображає зміну економічної напруги в різні роки.
- **Сезонна складова (Seasonal)** показує регулярні коливання значень, які повторюються з певною періодичністю. Це можуть бути реакції ринку на щорічні події або квартальні звіти. Амплітуда сезонності є помірною, але стабільною протягом усього періоду.
- **Залишкова складова (Residual)** містить випадкові коливання, які не можуть бути пояснені трендом або сезонністю. Їхня амплітуда свідчить про наявність шуму в ряді, зумовленого непередбачуваними подіями (наприклад, фінансовими кризами або політичними інцидентами).

Виконана декомпозиція дозволяє зробити висновок, що часовий ряд містить усі основні компоненти: тренд, сезонність і залишки. Це свідчить про доцільність використання моделей, які враховують обидва аспекти — наприклад, SARIMA або STL-декомпозицію з подальшим моделюванням залишків.

Декомпозиція також дозволяє окремо досліджувати кожну складову і покращити якість майбутніх прогнозів за рахунок точного врахування періодичної структури та довгострокових тенденцій.

3 Висновки

У ході виконання лабораторної роботи було проведено комплексне дослідження часового ряду індексу волатильності VIX. Спочатку було побудовано графік часового ряду, що дозволило візуально оцінити його поведінку.

Для перевірки стаціонарності було застосовано два взаємодоповнюючих статистичних тести — тест Дікі-Фуллера (ADF) та тест Квіатковського–Філліпса–Шмідта–Шина (KPSS). Результати обох тестів виявилися узгодженими: ADF-тест дав р-значення значно менше за 0.05, що дозволяє відхилити нульову гіпотезу про наявність одиничного кореня, а KPSS-тест підтверджив відсутність суттєвої зміни дисперсії з часом. Таким чином, часовий ряд можна вважати стаціонарним, що дозволяє надалі використовувати класичні моделі типу ARMA чи ARIMA без додаткових перетворень.

Наступним етапом була декомпозиція часового ряду, яка дала змогу виокремити основні компоненти — тренд, сезонність та залишкову складову. Аналіз тренду показав наявність довгострокової зміни рівня волатильності в різні періоди часу. Сезонна складова виявилась регулярною і стабільною, що вказує на періодичну поведінку індексу, ймовірно пов’язану з ринковими циклами. Залишкова частина містила нерегулярні коливання, зумовлені випадковими зовнішніми подіями.

Отримані результати підтверджують ефективність декомпозиції як інструменту попереднього аналізу та обґрунтують вибір відповідних математичних моделей для подальшого прогнозування. Проведене дослідження є важливим етапом перед моделюванням часових рядів і дозволяє краще розуміти внутрішню структуру даних.

Список використаних джерел

1. Гуськова В.Г. Лекційні матеріали з дисципліни "Системи інтелектуального прогнозування часових рядів"
2. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. Time Series Analysis: Forecasting and Control. 5th ed. Hoboken, NJ: Wiley, 2015. 712 p. [Електронний ресурс]. – Режим доступу: http://repo.darmajaya.ac.id/4781/1/Time%20Series%20Analysis_%20Forecasting%20and%20Control%20%28%20PDFDrive%20%29.pdf
3. *Офіційна документація Python-бібліотеки statsmodels*: <https://www.statsmodels.org/stable/index.html>
4. *Документація Pandas (робота з часовими рядами)*: https://pandas.pydata.org/docs/user_guide/timeseries.html