**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**
**«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Кафедра комп’ютерної математики і аналізу даних

# ЗВІТ

про виконання Лабораторної роботи №5 за темою
«Ідентифікація і дослідження моделі АР(2) часового ряду»
з дисципліни «Аналіз даних і часових рядів»

Група КН-122а

| | |
|---|---|
| Виконавець | Жарський Н.Д. |
| Викладач | Гардер С.Є. |

Харків 2025

**Мета роботи**

Ознайомитися з методами автокореляційного аналізу (АКФ, PACF), оцінити та валідувати AR(2)-модель методом Юла–Уоккера, змоделювати процес та виконати короткостроковий прогноз

**Теоретичні відомості**

- Автокореляційна функція (АКФ) показує кореляцію між значеннями ряду на відстані (лазі) $\tau$
- Часткова автокореляційна функція (PACF) оцінює прямий вплив попередніх значень із врахуванням проміжних лагів
- Метод Юла–Уоккера використовує рівняння на основі автокореляцій для оцінки параметрів AR(p)
- AutoReg (statsmodels) спрощує оцінку AR-моделі та побудову прогнозу

**Опис даних**

- Початковий ряд із 50 спостережень (колонка X1 із ts.csv), розширено до 100 точок AR(2)–генерацією з шумом
- Для моделювання використано перші 97 спостережень, останні 3 — для перевірки прогнозу

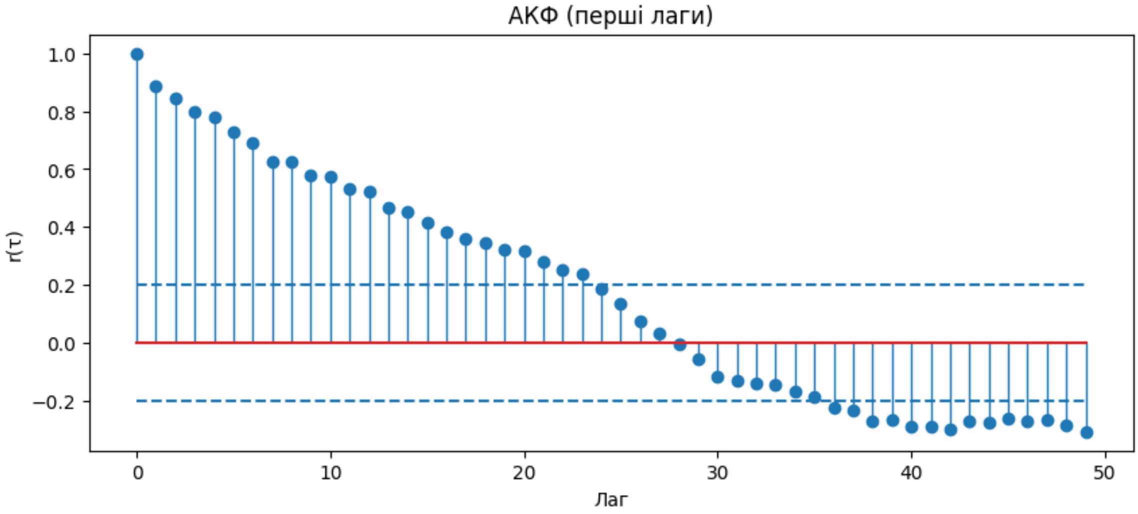
**Завдання роботи**

1. Обчислити та побудувати АКФ ряду (лаги 0…49), визначити кількість значущих лагів
2. Обчислити та побудувати PACF (лаги 0…19), визначити кількість значущих лагів
3. Оцінити параметри AR(2) методом Юла–Уоккера, виконати t-тест для $\hat{\rho}_1$, $\hat{\rho}_2$
4. Змоделювати AR(2)-процес за оціненими параметрами, накласти на оригінал
5. Зробити прогноз на 1–3 кроки вперед, порівняти з фактичними значеннями

**Результати**

1. **Автокореляція (АКФ)**

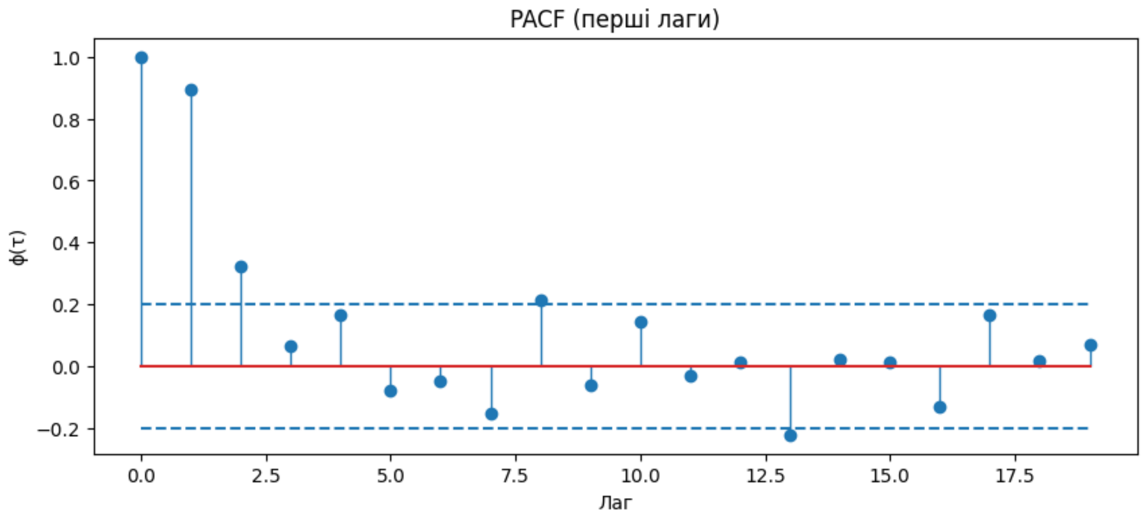
- **Графік АКФ (лаги 0–49):**




- **Кількість значущих лагів:** 38 (|r(τ)| > 1.96/√97 ≈ 0.199)
- **Поводження:** від лагу 1 (≈0.88) спостерігається повільне, майже лінійне зниження r(τ), яке не досягає нуля до лагу 28, а потім переходить у від'ємну зону. Далі автокореляція "хвилюється" навколо нуля, затухаючи до ≈–0.3 на кінцях
- **Довірчі межі** ±1.96/√97 ≈ ±0.199 (пунктир)
- Таке повільне згасання АКФ вказує на сильну й тривалу пам'ять у процесі — класична властивість слабозгасаючих AR(2) з великими коефіцієнтами. Автокореляція зберігається довго, що може ускладнювати короткострокове прогнозування, адже вплив попередніх значень залишається відчутним навіть через багато періодів

2. **Часткова автокореляція (PACF)**

- **Графік PACF (лаги 0–19):**



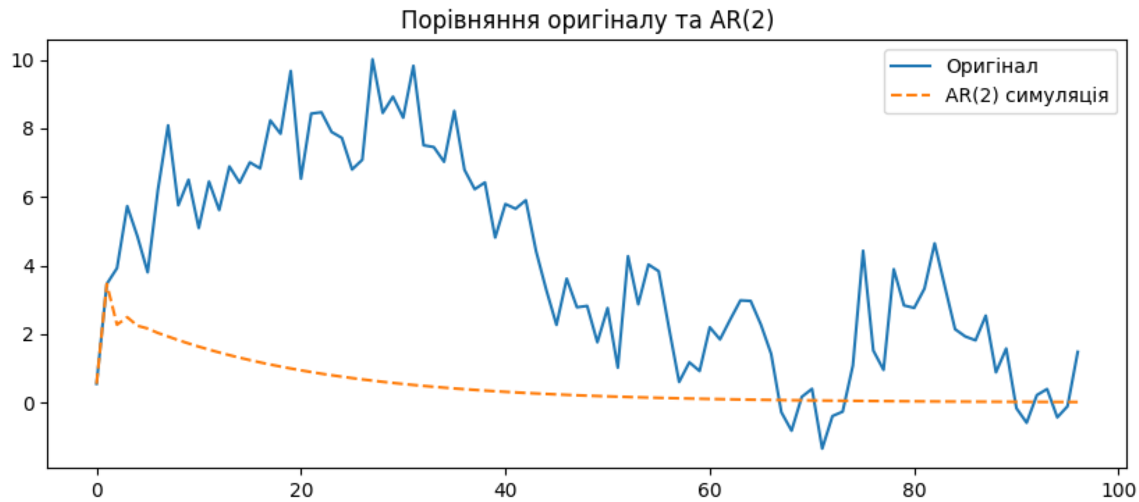

- **Кількість значущих лагів:** 5
- **Поводження:** на лагу 1 $\hat{\phi}_1 \approx 0.89$ — найбільший стрибок, на лагу 2 $\hat{\phi}_2 \approx 0.32$. Далі прямі кореляції зменшуються, але утримуються трохи вище довірчого порогу на лагах 8 (≈0.21), 10 (≈0.14) та 17 (≈0.17)
- **Довірчі межі** ті ж ±0.199
- Яскравий спад від $\hat{\phi}_1$ до $\hat{\phi}_2$ та подальше затухання підтверджує, що AR(2) є базовою моделлю, але слабкі "відлуння" на вищих лагах вказують на присутність додаткової динаміки — можливо, варто розглянути AR(p) із p>2

### 3. Оцінка AR(2)

- $\hat{\rho}_1 = 0.6054$ ($t = 5.331$, $p < 0.001$)
- $\hat{\rho}_2 = 0.3226$ ($t = 2.841$, $p = 0.005$)
- $\sigma^2$ шуму = 1.25075
- Ці значення добре узгоджуються з графіками: великий $\hat{\rho}_1$ забезпечує високий початковий рівень АКФ, а невеликий $\hat{\rho}_2$ — сповільнює спад кривої

### 4. Симуляція AR(2) vs оригінал

- **Графік порівняння:**




- **Спостереження:** змодельований ряд затихає експоненційно, оригінал має довгі коливання
- **Поводження симуляції:** експоненціальне згасання до нуля вже на 15–20 відлік, далі шум коливається близько нуля
- **Порівняння:** оригінал має складнішу структуру — тривалий період підйому (до ≈10), різкі коливання та зниження
- синя лінія — оригінальні 97 точок; помаранчева пунктирна — згенеровані AR(2) значення
- AR(2) передає базове затухання кореляції, але не здатен відобразити нерівномірні цикли та піки реального ряду. Це підкреслює обмеження простої AR(2)-моделі для рядів з вираженою довготривалою динамікою

**5. Прогноз 1–3 кроки вперед**

| Крок | Прогноз | Фактичне | Помилка |
|---|---|---|---|
| 1 | 1.1391 | 0.2407 | –0.8984 |
| 2 | 1.4089 | 0.6042 | –0.8048 |
| 3 | 1.4763 | 2.1786 | +0.7023 |

Модель систематично недооцінює перші два кроки (висока негативна помилка) та завищує третій — це результат стабільного затухання AR(2), яке не враховує можливі сплески чи відскоки в даних

**Висновки**

1. **Потужна автокореляція:** розширений ряд демонструє дуже повільне згасання АКФ (до 38 значущих лагів), що є наслідком великих $\hat{\rho}$-параметрів AR(2)
2. **Структура PACF:** два основні лага (1 і 2) визначають поведінку процесу, але невеликі "відлуння" на високих лагах свідчать про складніший цикл
3. **Адекватність AR(2):** хоча обидві оцінки $\hat{\rho}$ є статистично значущими, сама модель занадто спрощена для відтворення реального ряду
4. **Прогноз:** AR(2) не забезпечує високої точності короткострокових прогнозів для цього ряду через відсутність у моделі трендів і циклів

**Контрольні питання**

**1. Що відображає автокореляційна функція (АКФ) та як інтерпретувати довготривалу пам'ять у АКФ?** АКФ показує кореляцію між значеннями ряду, віддаленими на $\tau$ кроків. Високе значення $r(\tau)$ свідчить, що спостереження на відстані $\tau$ сильно залежить від попередніх. Довготривала пам'ять проявляється в тому, що $r(\tau)$ уповільнено спадає і залишається значущим на великих лагах, що вказує на тривалу авто-кореляцію процесу

**2. Чому PACF важливіша за АКФ при виборі порядку AR-моделі?** PACF показує тільки прямий вплив значення на лагу $\tau$, відсікаючи проміжні ефекти між лагами. Отже, якщо PACF різко обривається після лагу p ($|\hat{\phi}(\tau)|$ стає незначущим для $\tau>p$), це ідеальний індикатор порядку AR(p). АКФ у цьому випадку може спадати поступово й не дає чіткого "обриву"

**3. Як працює метод Юла–Уоккера і чому він дає оцінки $\hat{\rho}$ без зсуву?** Метод Юла–Уоккера будує систему лінійних рівнянь на основі автокореляційних коефіцієнтів $r(\tau)$ (рівняння Янґа–Якобі) і вирішує її для параметрів AR(p). Оскільки під час побудови модель вважає, що процес вже в стаціонарному стані, оцінки виходять несмещені (без зсуву) під час великих вибірок

**4. Яку роль відіграє t-тест у валідації параметрів AR-моделі?** Після оцінки параметрів $\hat{\rho}_1,\ldots,\hat{\rho}_p$ необхідно перевірити, чи відрізняються вони статистично від нуля. Для кожного $\hat{\rho}$ обчислюють t-статистику ($\hat{\rho}$/SE), де SE — стандартна помилка. Якщо p-значення $< \alpha$ (зазвичай 0.05), то відповідний параметр є значущим і вносить реальний внесок у модель

**5. Чому проста AR(2)-модель може бути недостатньою для складних часових рядів?** AR(2) описує лише дві прямі лінійні залежності (від двох попередніх кроків) і випадковий шум. У разі наявності складних циклів, сезонних коливань або довготривалої пам'яті на більших лагах знадобиться AR(p) із більшим p або більш складні моделі (ARIMA/SARIMA, експоненційне згладжування), які враховують тренди, сезонність чи інші компоненти