

**Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”
Фізико-технічний інститут**

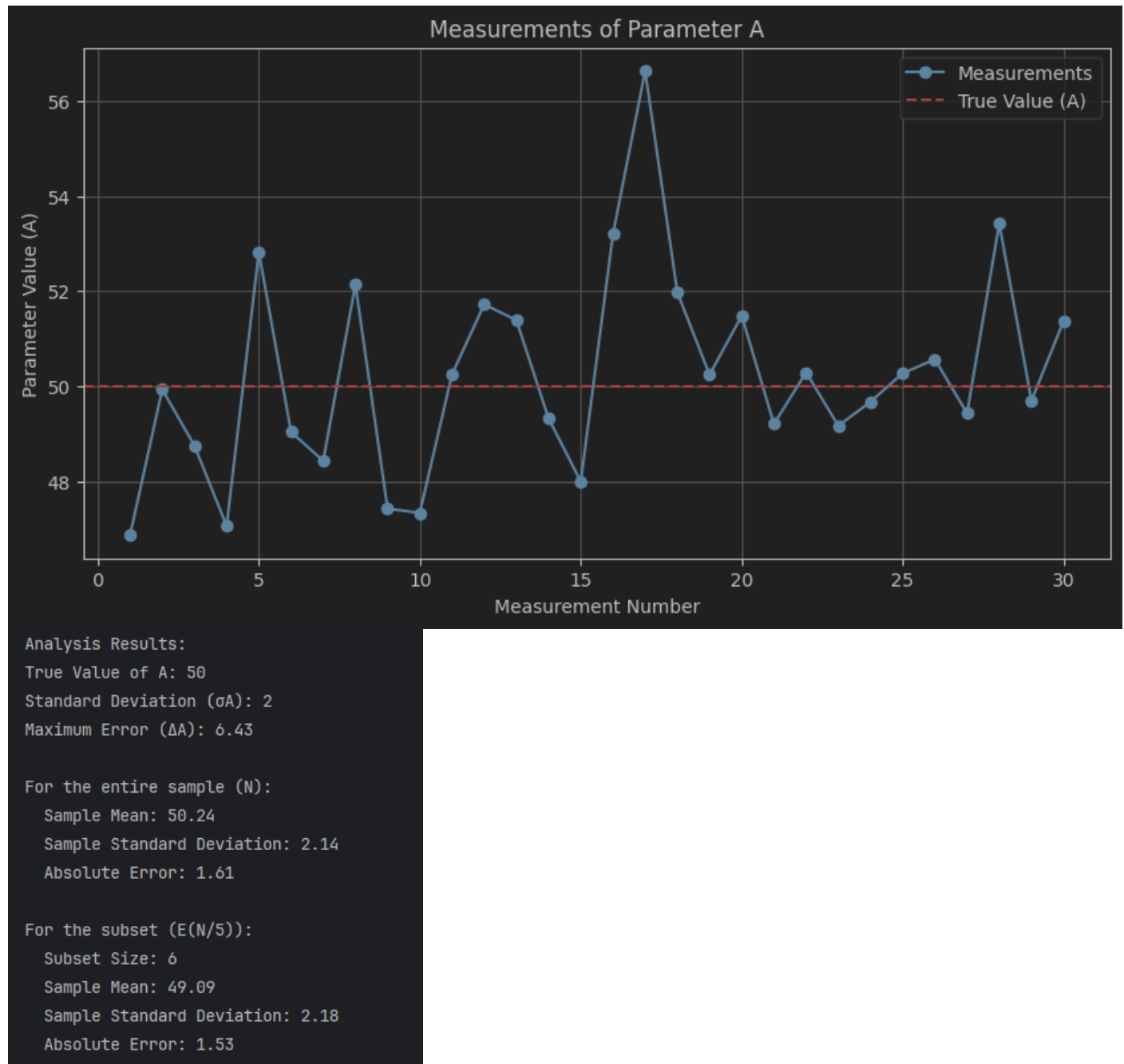
Методи обробки та розпізнавання даних
Лабораторна робота №1

Виконав:
студент ФІ-31мн
Шевченко Ю.
Перевірив:
Юзефович В. В.

Лабораторна робота №1. Формалізація, згладжування та прогнозування даних

Хід роботи:

1.1



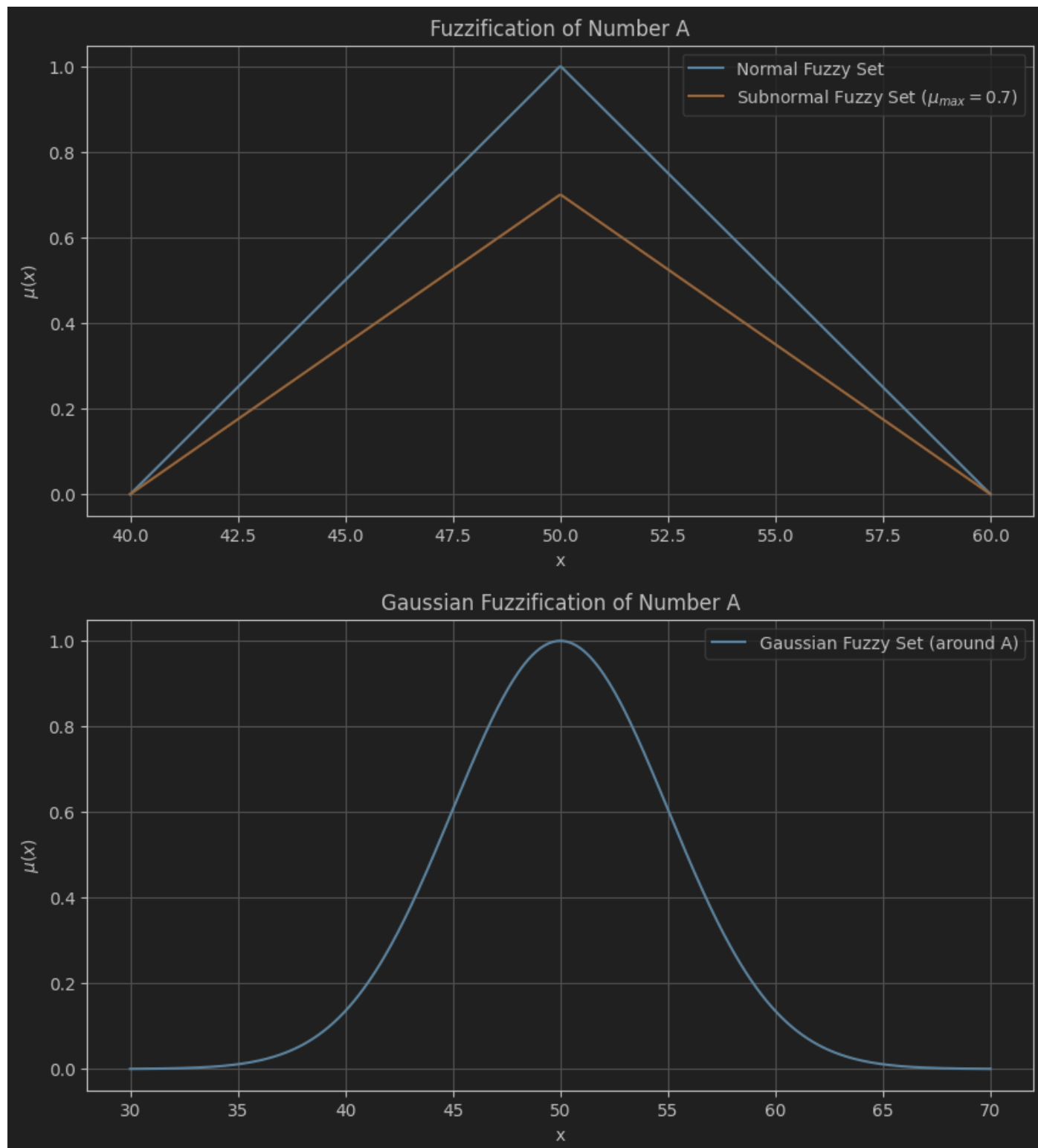
У ході виконання завдання було згенеровано множину даних із заданими параметрами: середнім значенням $A \approx 50$ та стандартним відхиленням $\sigma_A = 2$, що розподілені за нормальним законом. Максимальну похибку вимірювань визначено за правилом «трьох сигм» як $\Delta A = 6.43$.

Результати оцінки параметрів для всієї вибірки ($N = 30$) свідчать, що вибіркове середнє значення A дорівнює 50.24, що є близьким до заданого, а вибіркова стандартна похибка складає 2.14, що підтверджує адекватність моделі. Абсолютна похибка для всієї вибірки дорівнює 1.61, що також узгоджується із допустимими межами.

Аналіз підвибірki, яка складає $1/5$ від загальної вибірки ($E(N/5) = 6$), показав дещо знижене середнє значення (49.09) і підвищену вибіркoву стандартну похибку (2.18). Абсолютна похибка в цьому випадку склала 1.53. Ці результати демонструють вплив меншого обсягу вибірки на точність оцінки.

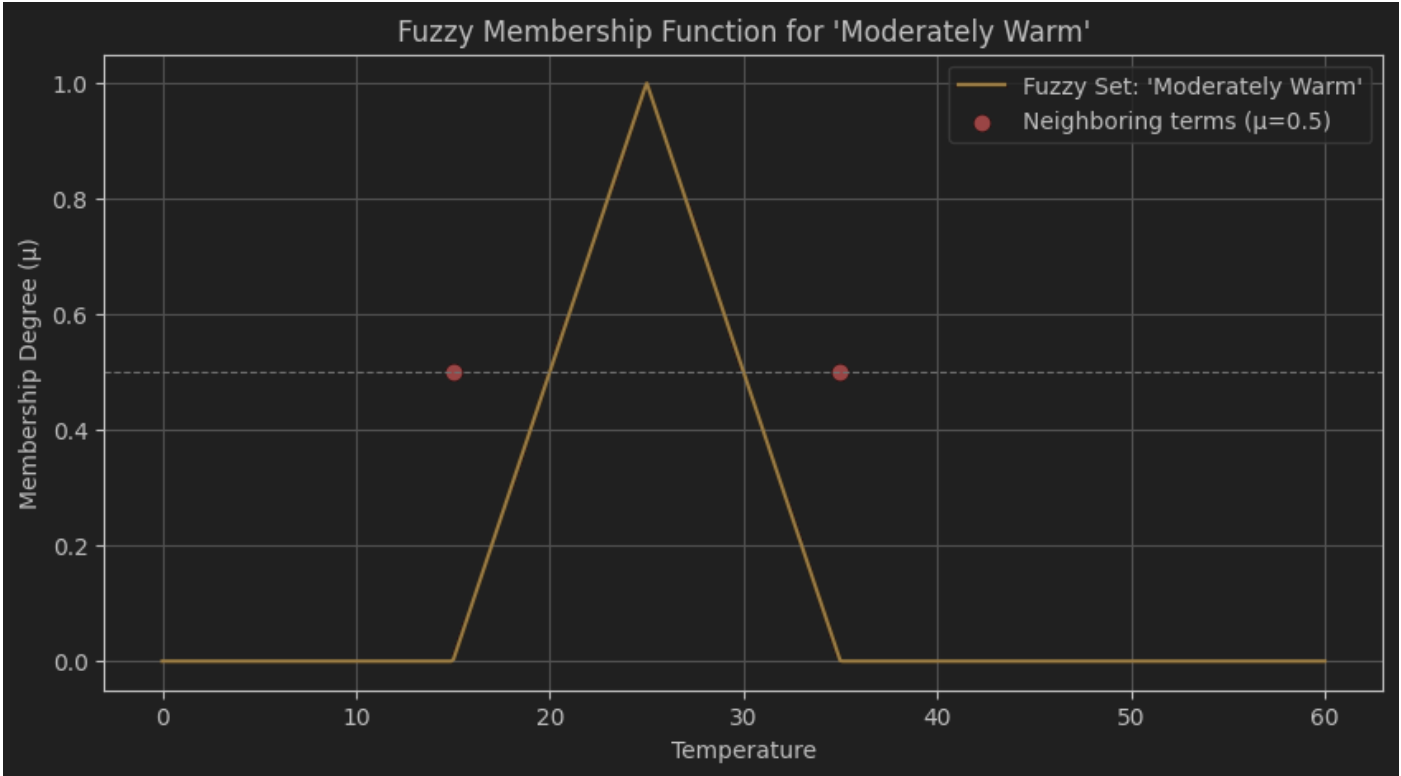
Таким чином, результати для всієї вибірки краще відображають реальні характеристики ряду, ніж аналіз малої підвибірki, що підтверджує важливість використання достатнього обсягу даних для точного оцінювання параметрів. Всі отримані показники узгоджуються із заданими значеннями параметрів A , σ_A та ΔA .

1.2.1

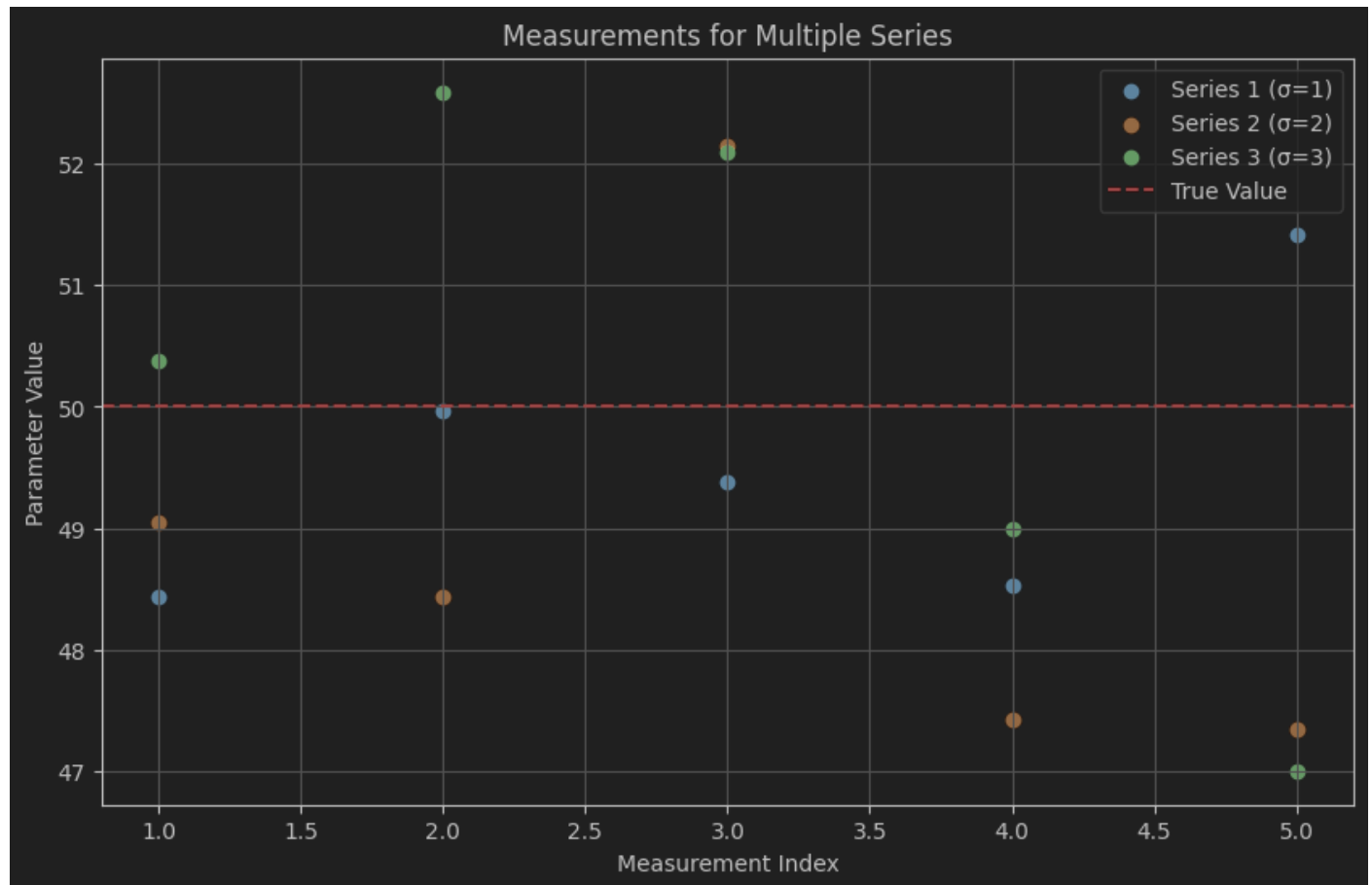




1.2.2



1.2.3



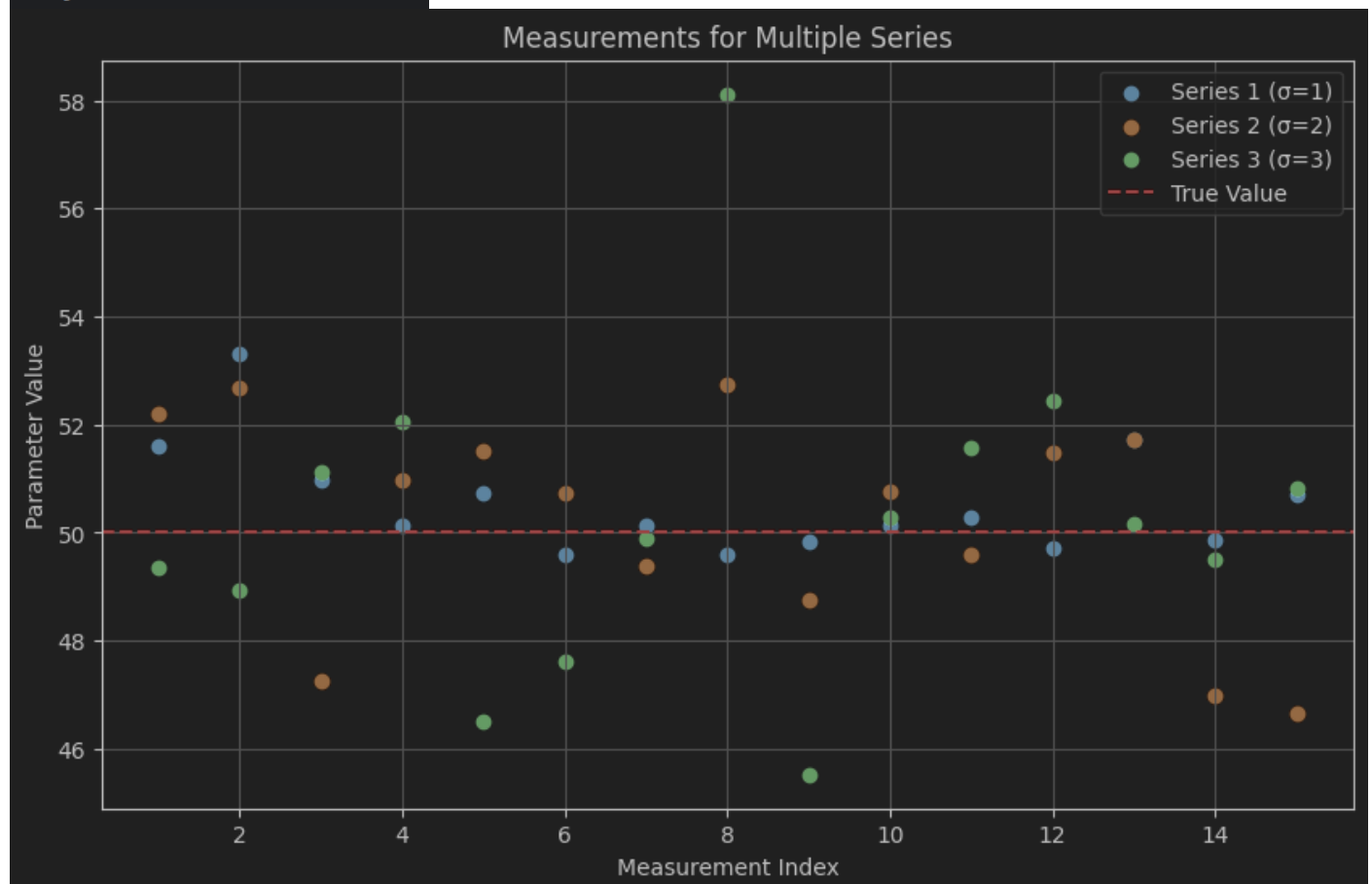
Results for $N = 5$

Unweighted Mean: 49.55

Unweighted Standard Deviation: 1.82

Weighted Mean: 49.48

Weighted Standard Deviation: 0.38



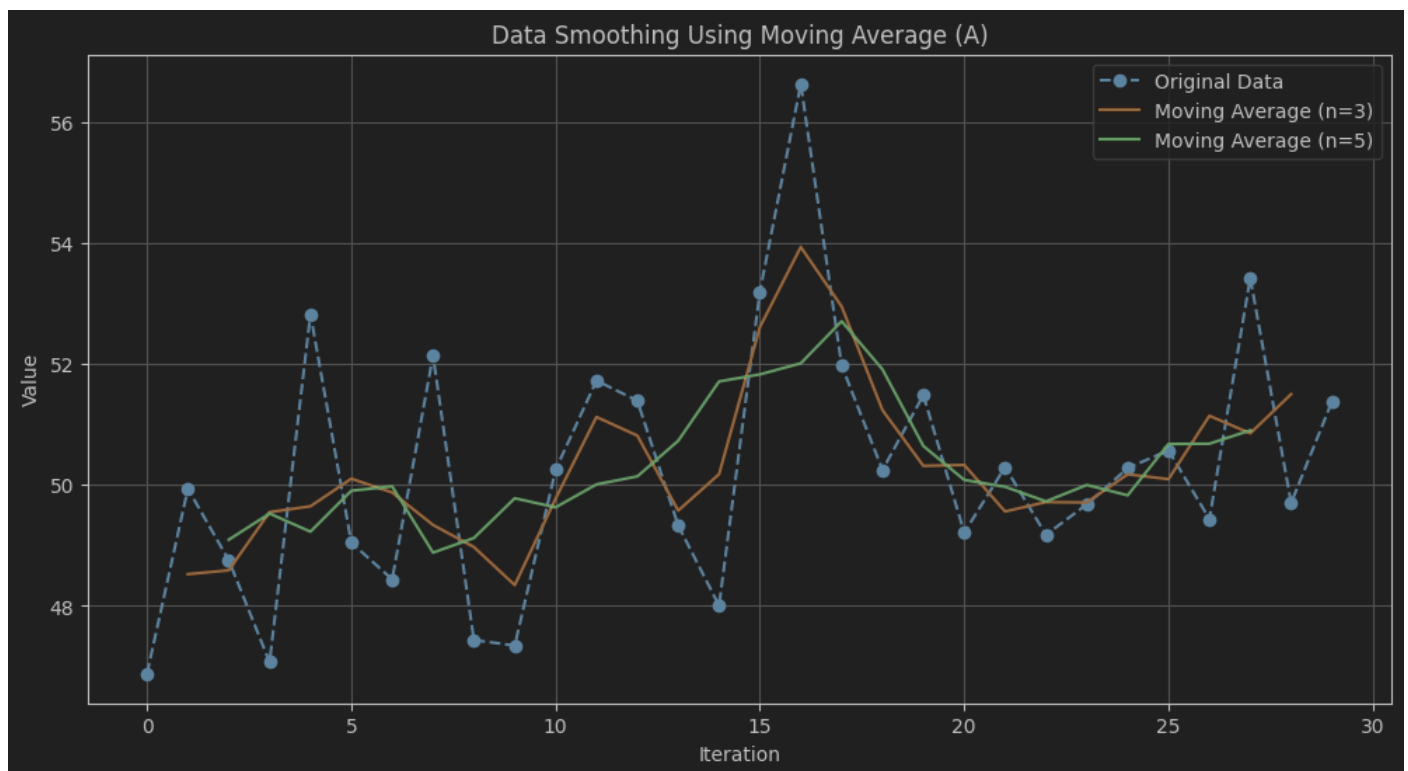
```
Results for N = 15
Unweighted Mean: 50.35
Unweighted Standard Deviation: 2.09
Weighted Mean: 50.47
Weighted Standard Deviation: 0.22
```

У результаті виконання завдання були згенеровані три серії вимірювань параметра A із різними значеннями середньоквадратичного відхилення (σ_A) для аналізу впливу точності вимірювань. Для вибірки з $N=5$ було отримано невиважені значення математичного сподівання 49.55 і стандартного відхилення 1.82. Водночас виважені значення дали більш точні результати: математичне сподівання 49.48 і стандартне відхилення 0.38. Це підтверджує, що врахування точності окремих серій вимірювань суттєво знижує похибки.

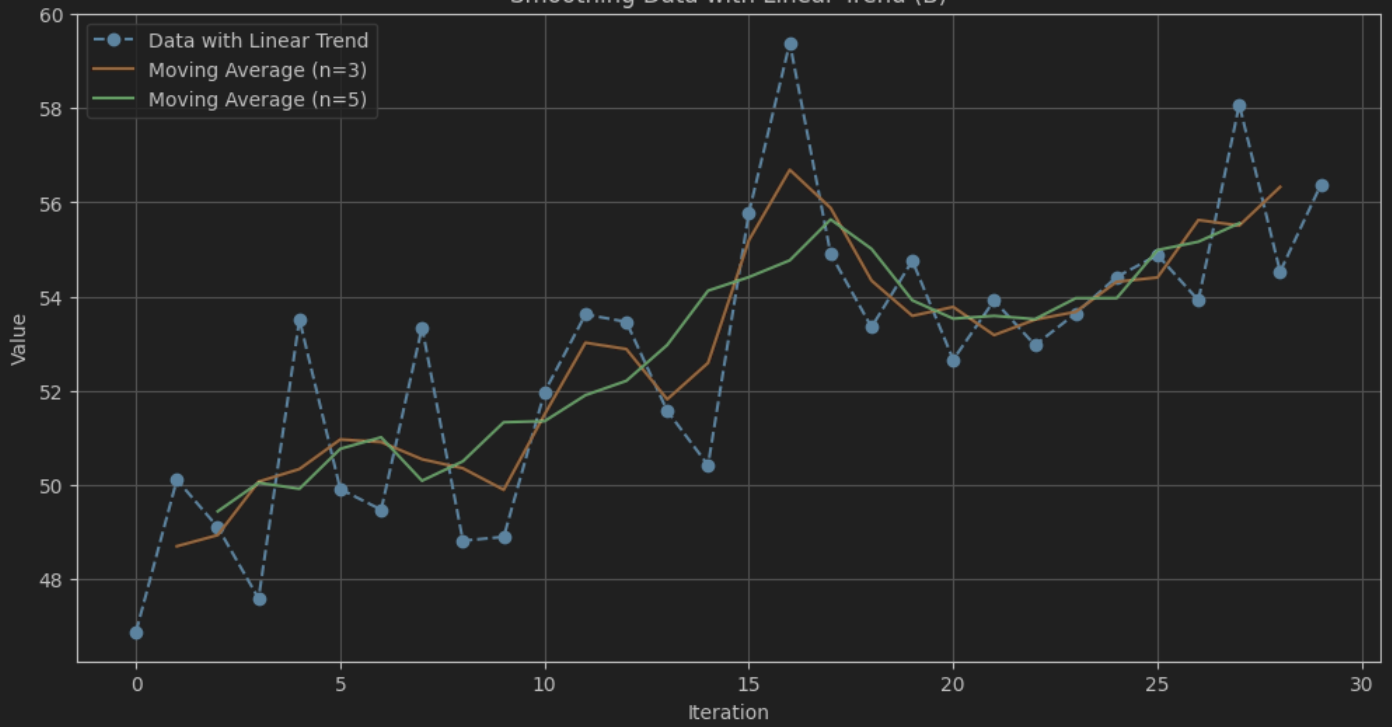
При збільшенні вибірки до $N=15$ невиважені результати також дещо відрізнялися: математичне сподівання становило 50.35, а стандартне відхилення зросло до 2.09. Водночас виважені результати знову демонстрували менші похибки: математичне сподівання 50.47, а стандартне відхилення всього 0.22. Це свідчить про підвищення точності оцінки при збільшенні обсягу даних.

Отримані результати підтверджують, що врахування різної точності вимірювань значно знижує стандартне відхилення, роблячи оцінки більш достовірними. Крім того, збільшення вибірки підвищує надійність результатів як для виважених, так і для невиважених даних.

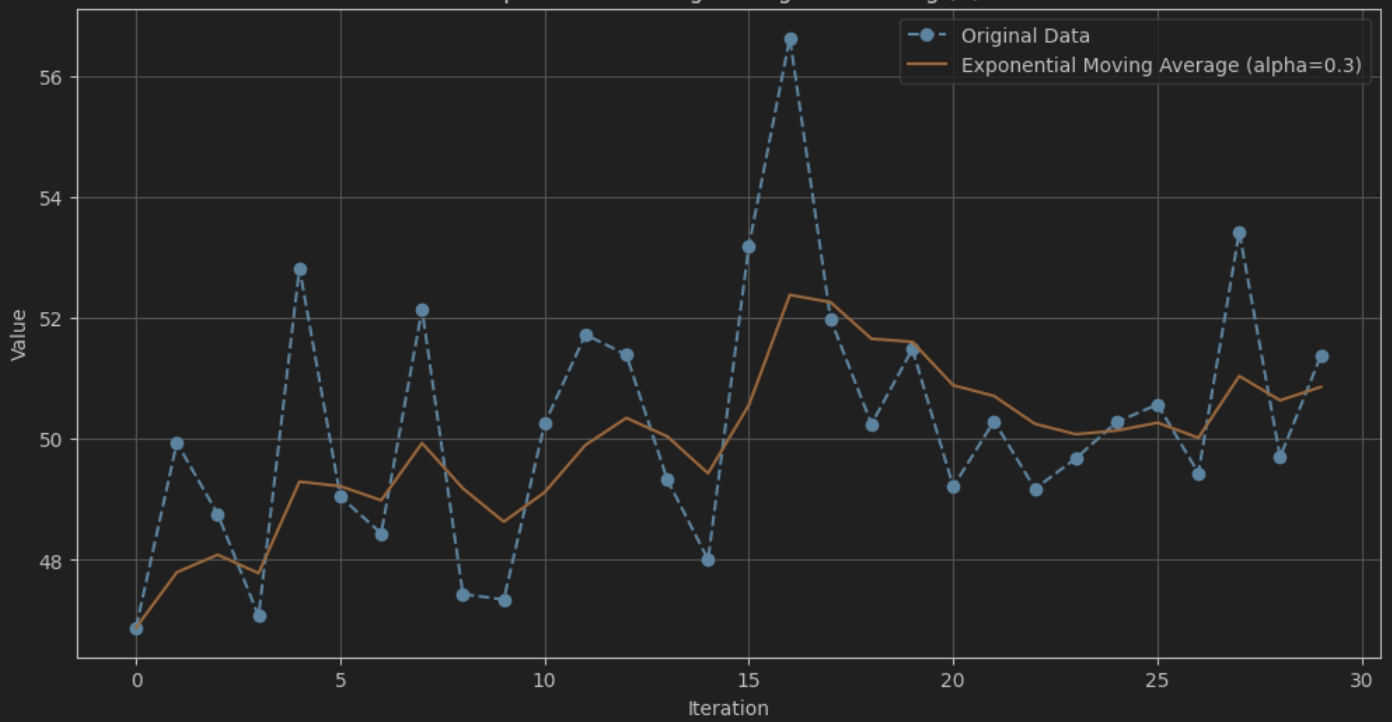
2.1.1

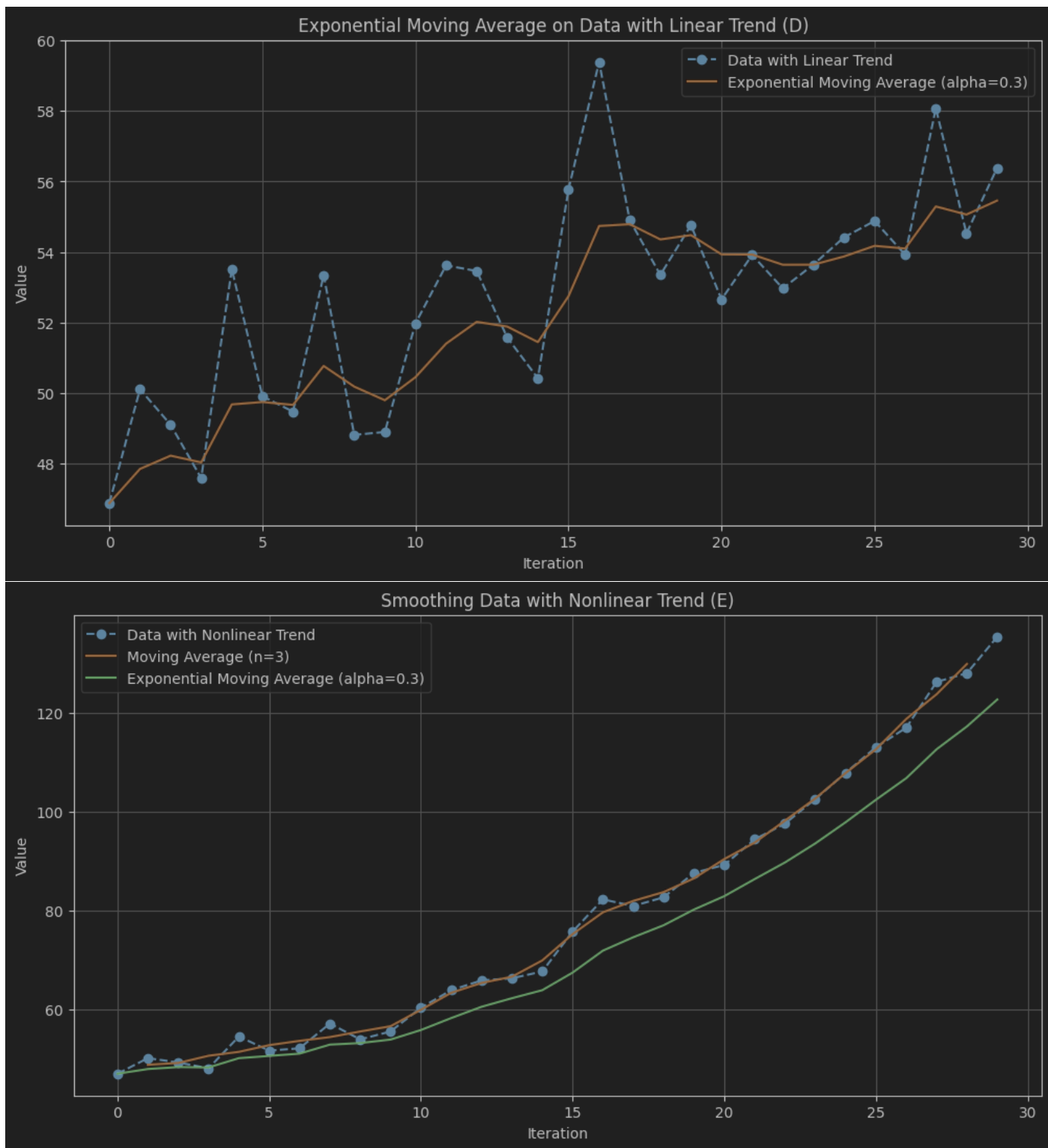


Smoothing Data with Linear Trend (B)



Exponential Moving Average Smoothing (C)



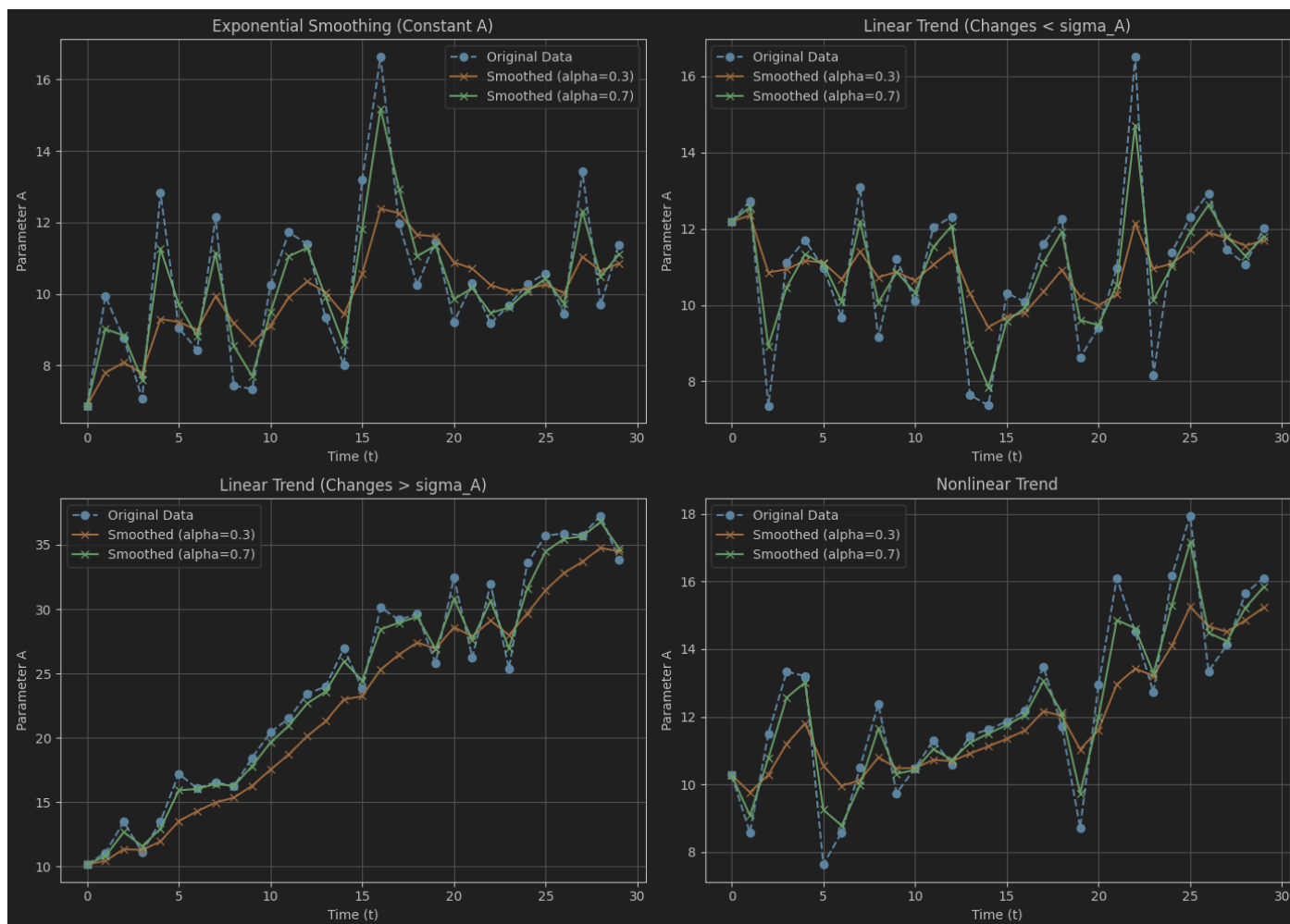


У завданні А метод ковзного середнього добре згладжує коливання в даних, при цьому $n=3$ зберігає більше варіативності, а $n=5$ робить результат більш рівним, але втрачаються дрібні деталі. У завданні Б, де був лінійний тренд, метод ковзного середнього не порушує тренд, але з великою кількістю точок втрачаються важливі коливання.

У завданні В експоненціально зважене ковзне середнє чутливіше до останніх змін і дозволяє краще відображати актуальні зміни, зберігаючи більше деталей при високому значенні α . У завданні Г, де є лінійний тренд, EWM краще зберігає тренд, порівняно з ковзним середнім, ефективно згладжуючи шум.

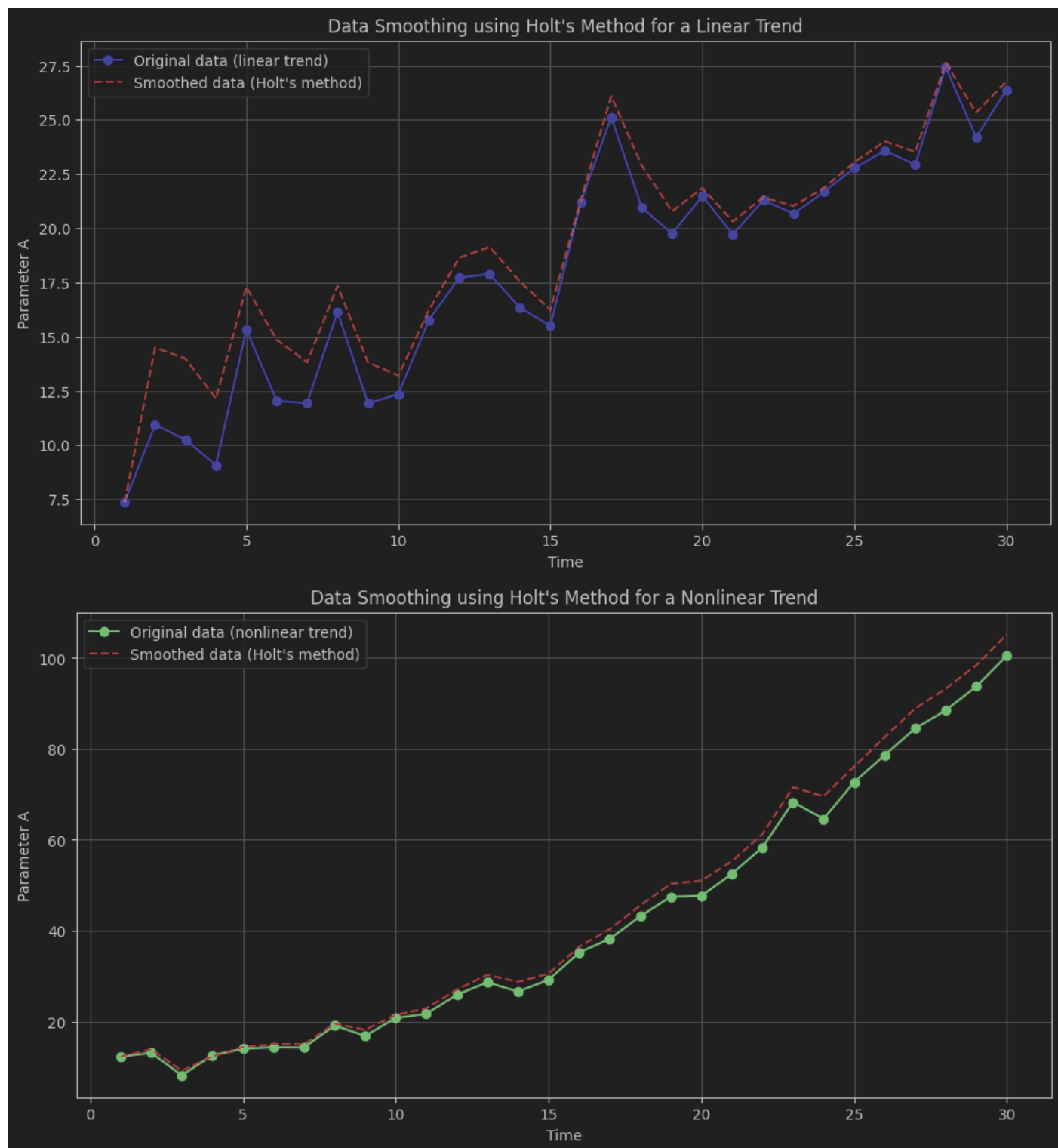
Загалом, ковзне середнє підходить для згладжування простих даних, а EWM краще справляється з трендами та чутливими змінами.

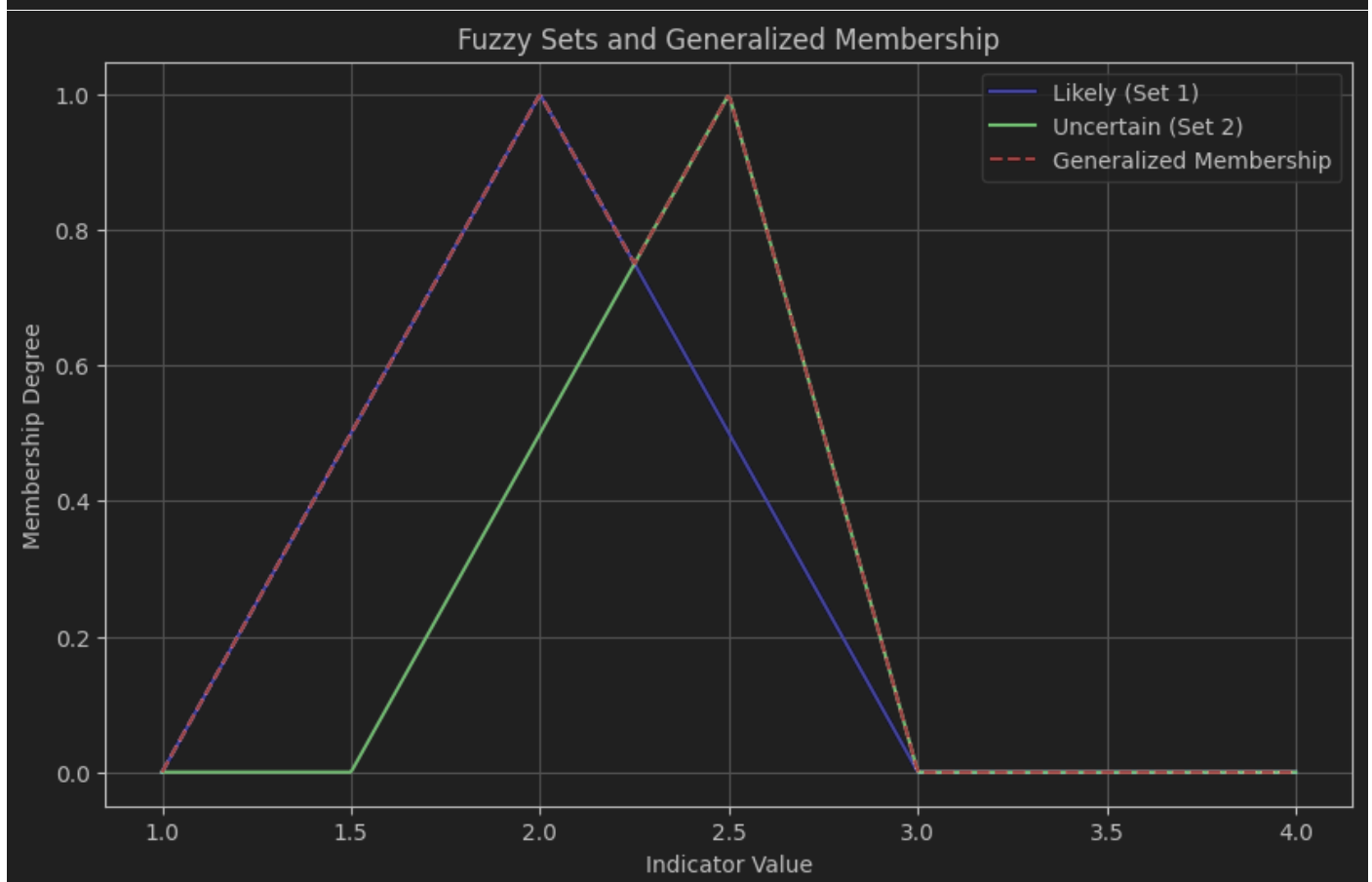
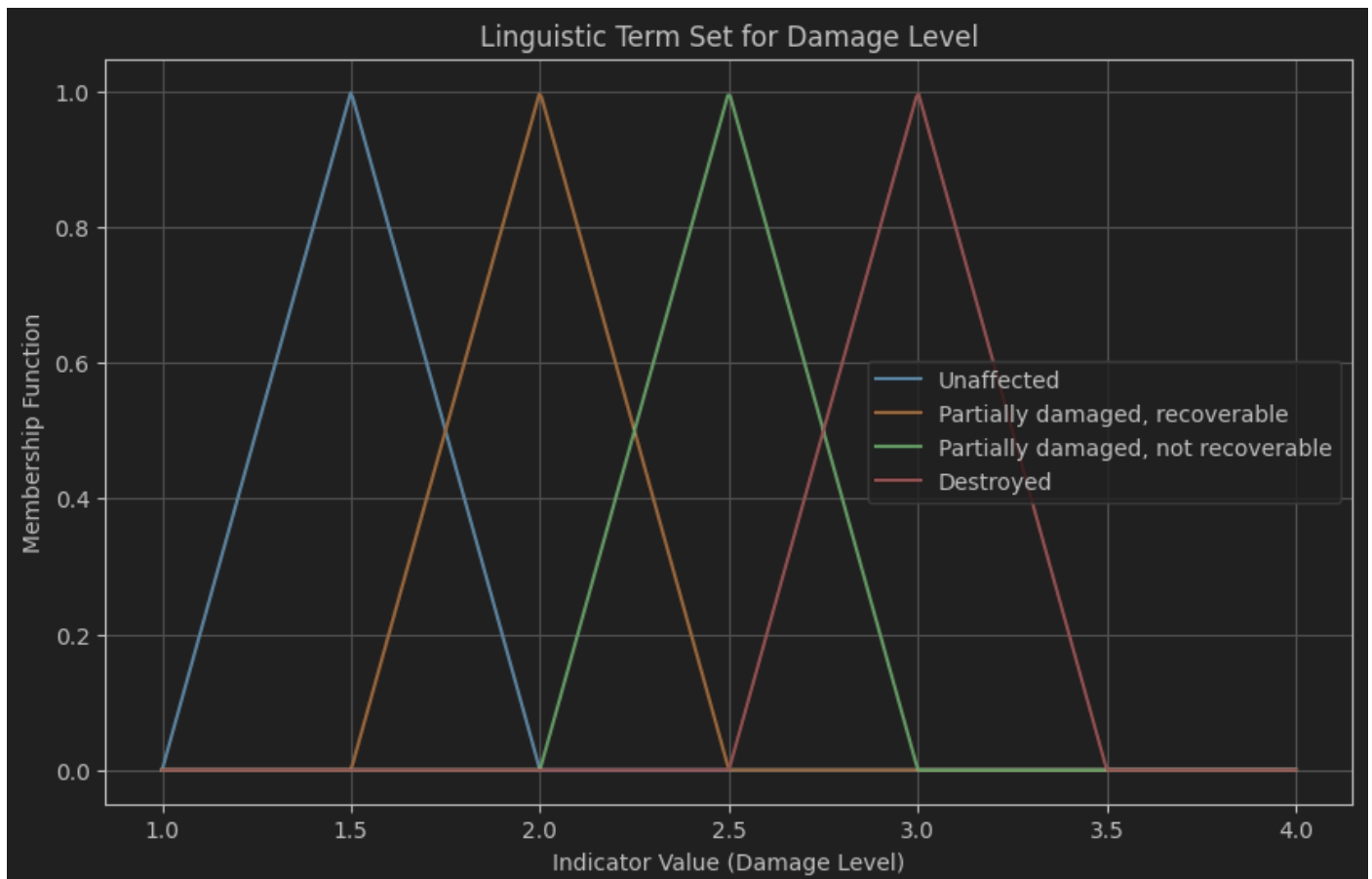
2.1.2



Результати виконання коду показують, як експоненціальне згладжування працює з різними типами даних. У частині А, де застосовувалось експоненціальне згладжування до константного параметра А, метод ефективно згладжує шум. Для $\alpha=0.3$ згладжування відбувається м'яко, а для $\alpha=0.7$ – швидше, що робить його чутливішим до змін. У частині Б, де застосовувався лінійний тренд з малими змінами, метод згладжування зберігав тренд, причому для $\alpha=0.7$ результат був ближчим до реального тренду, ніж для $\alpha=0.3$. У частині В, де лінійний тренд мав великі коливання, для $\alpha=0.7$ згладжування краще зберігало тренд, а для $\alpha=0.3$ флуктуації були більше згладжені. Частина Д показала, як експоненціальне згладжування працює з нелінійним трендом: метод ефективно згладжує коливання і зберігає тенденцію, при цьому для $\alpha=0.7$ результати були точнішими. Загалом, експоненціальне згладжування добре справляється з усуненням шуму та збереженням тенденцій, а параметр α дозволяє налаштувати чутливість до останніх змін.

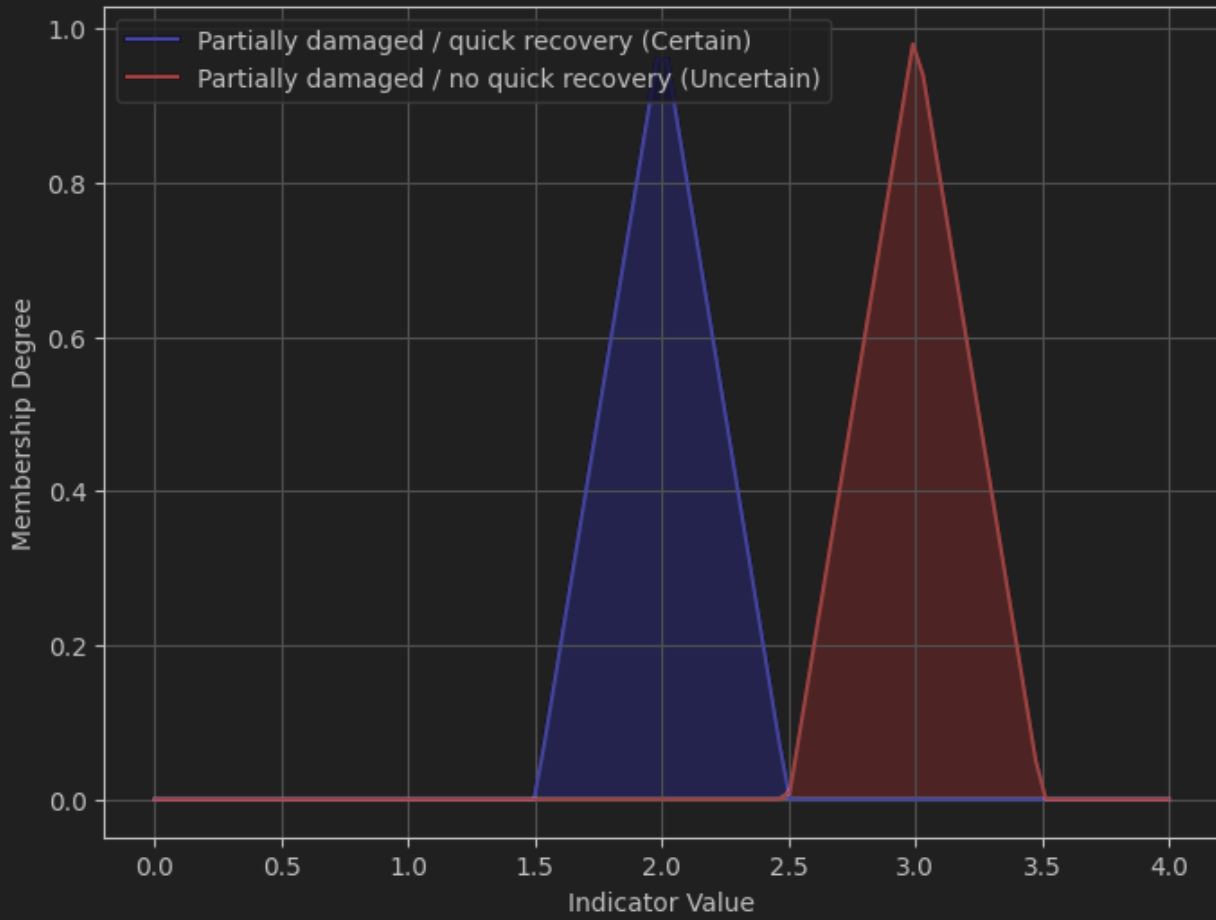
2.1.3



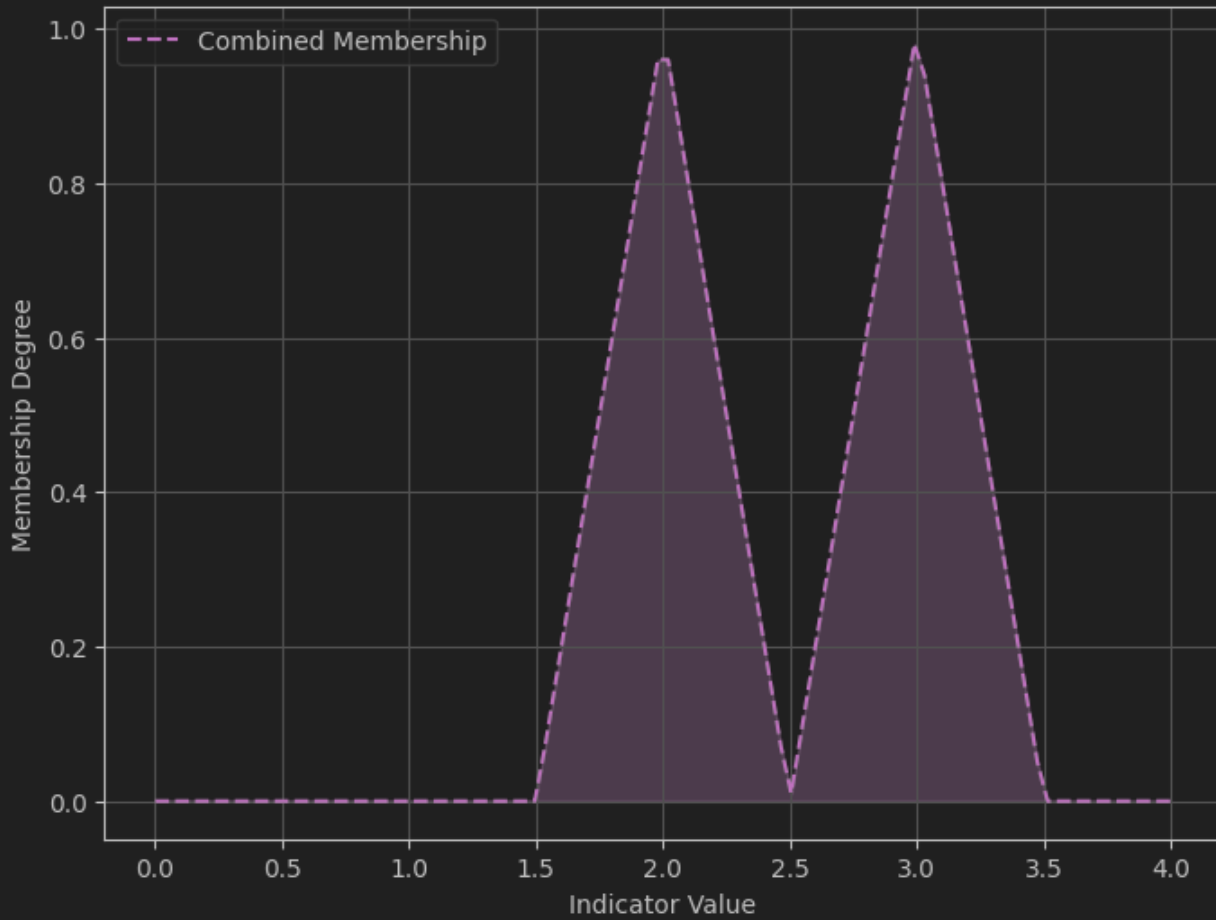


Confidence for Message 1: 2.99399
 Confidence for Message 2: 2.99098
 Generalized Confidence: 2.99399

Triangular Membership Functions for Two Messages



Combined Membership Function



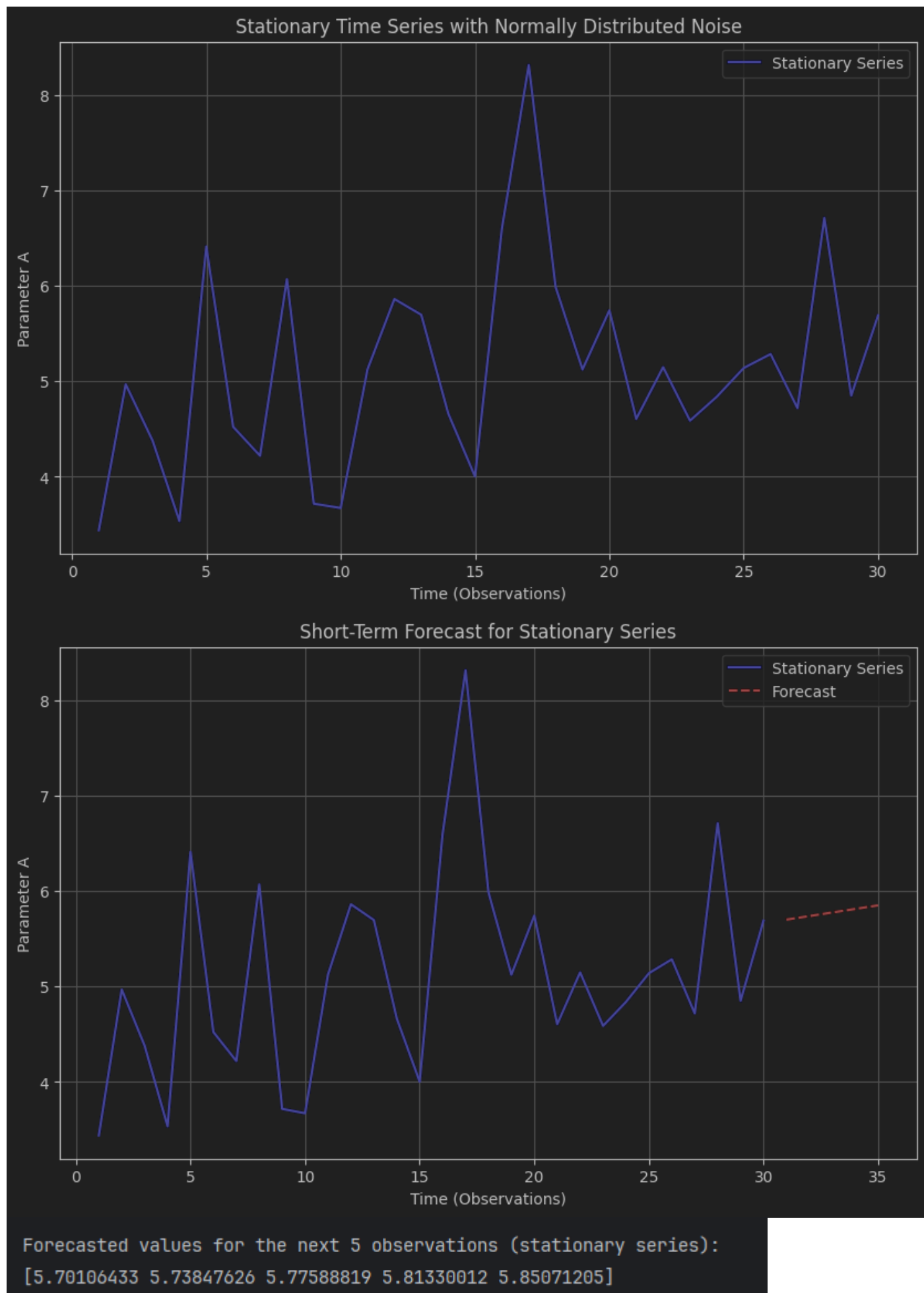
Final Standardized Certainty Level: 2.242

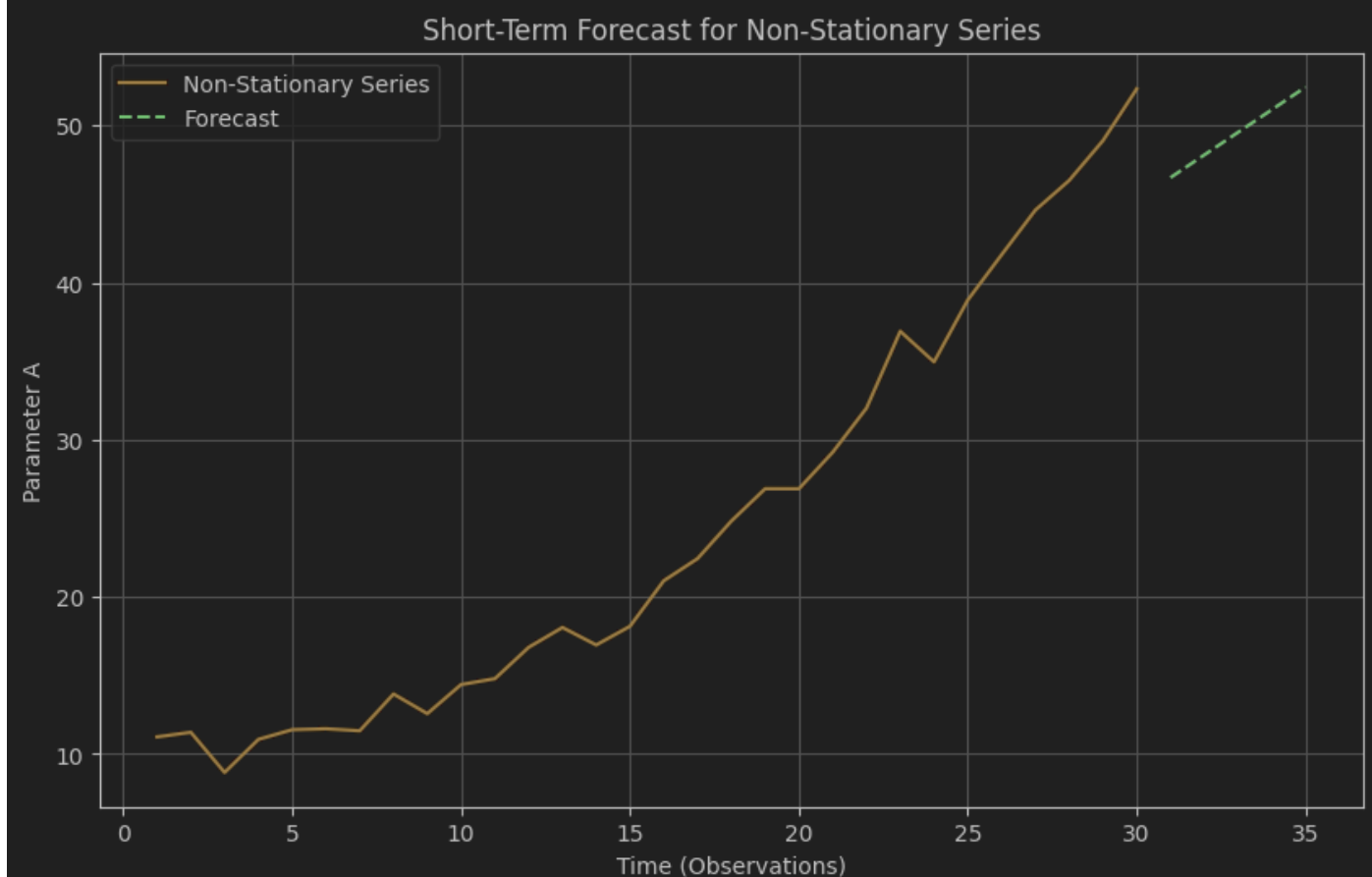
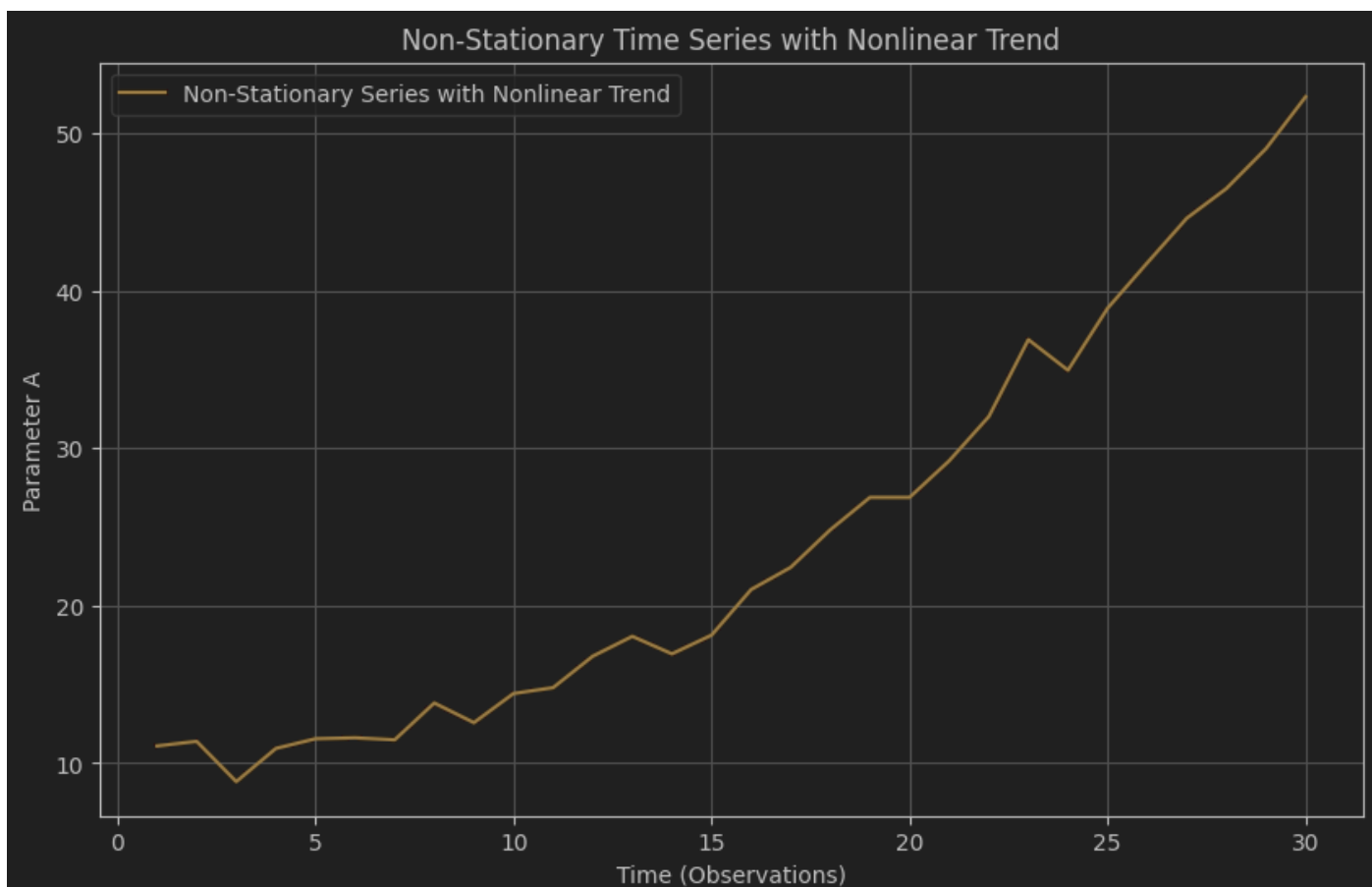
У цьому завданні створено терм-множину з чотирьох лінгвістичних значень, що описують ступінь ураження об'єкта: «Неуражений», «Частково уражений/підлягає швидкому відновленню», «Частково уражений/не підлягає швидкому відновленню», «Знищений». Для кожного з цих значень побудована трикутна функція належності, яка визначає відповідність кожного значення показника певному класу. Графіки для цих функцій дозволяють візуалізувати, як вони перекриваються між собою, відображаючи ступінь належності кожного значення до класу.

Для частини 2.2.1 були сформовані два повідомлення з однаковим значенням показника, але з різною достовірністю: перше — «ймовірно», друге — «сумнівно». Для кожного повідомлення побудовані трикутні функції належності, які відображають ступінь належності до відповідного класу. Після їх об'єднання за допомогою операції нечіткого об'єднання отримано комбіновану достовірність, що дорівнює 2.99399. Це значення показує, що комбіноване повідомлення має високу достовірність, оскільки воно ближче до значення «достовірно». Оскільки значення достовірності для обох повідомлень були схожими, комбіноване значення відображає високий рівень впевненості.

Для частини 2.2.2 були сформовані два повідомлення з різними значеннями показника та різною достовірністю: перше — «достовірно», друге — «сумнівно». Комбіноване значення достовірності, після нормалізації до стандартного рівня $K=3$, становить 2.242. Це показує середній рівень достовірності для комбінованого значення, що свідчить про компроміс між різними рівнями достовірності для кожного з повідомлень.

3.1





Forecasted values for the next 5 observations (non-stationary series):
 [46.68535181 48.1257923 49.56623278 51.00667327 52.44711375]

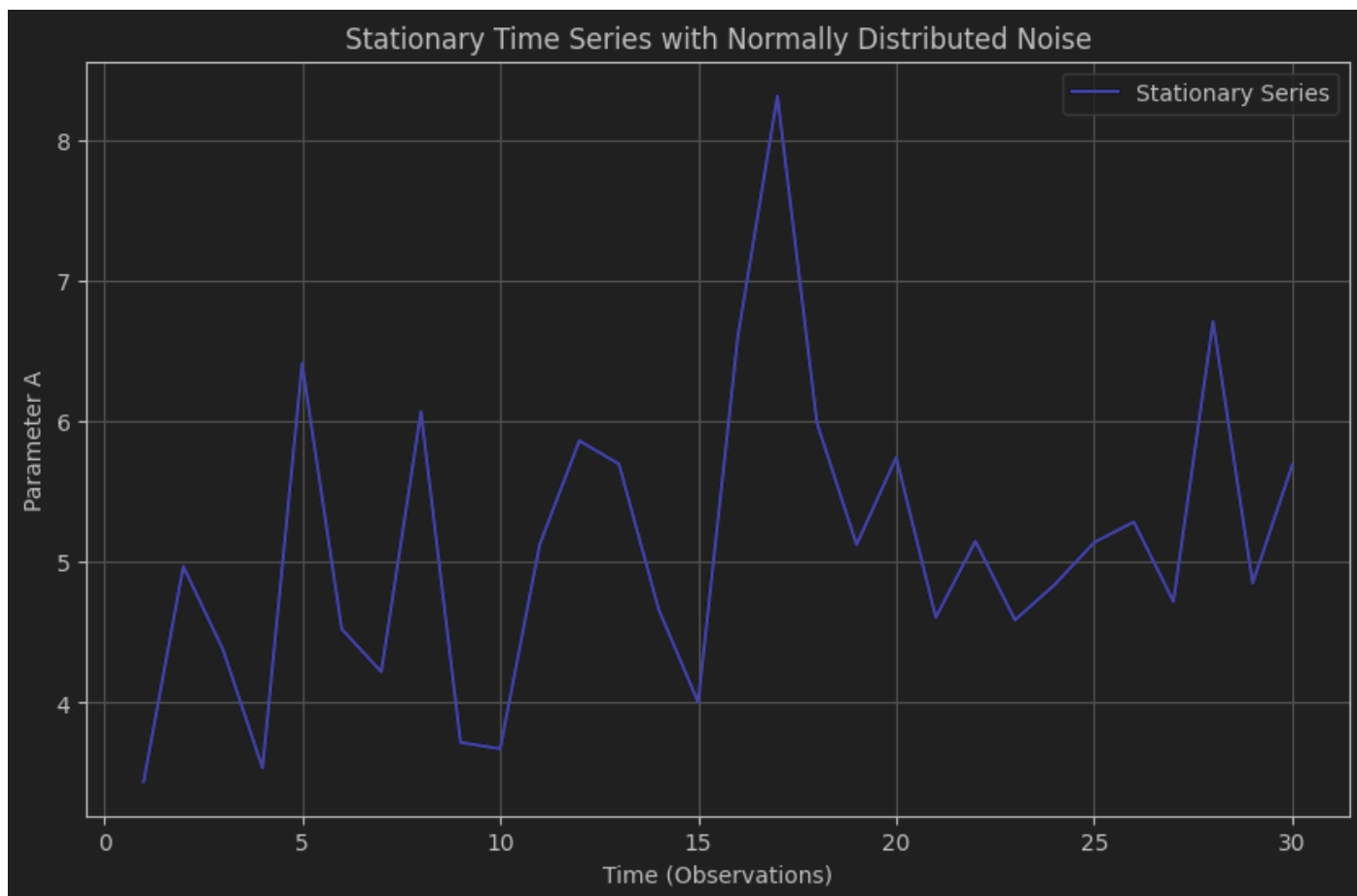
Для частини 3.1.1, де була згенерована стаціонарна серія з N вимірювань, ми отримали прогноз на наступні 5 спостережень за допомогою методу найменших квадратів.

Прогнозовані значення показують стабільний тренд, що характерно для стаціонарних рядів, де зміни в даних є мінімальними і варіації знаходяться в межах похибки вимірювань. Результати прогнозу: 5.701, 5.738, 5.775, 5.813 та 5.850, що свідчить про відсутність значних змін у даних за цей період.

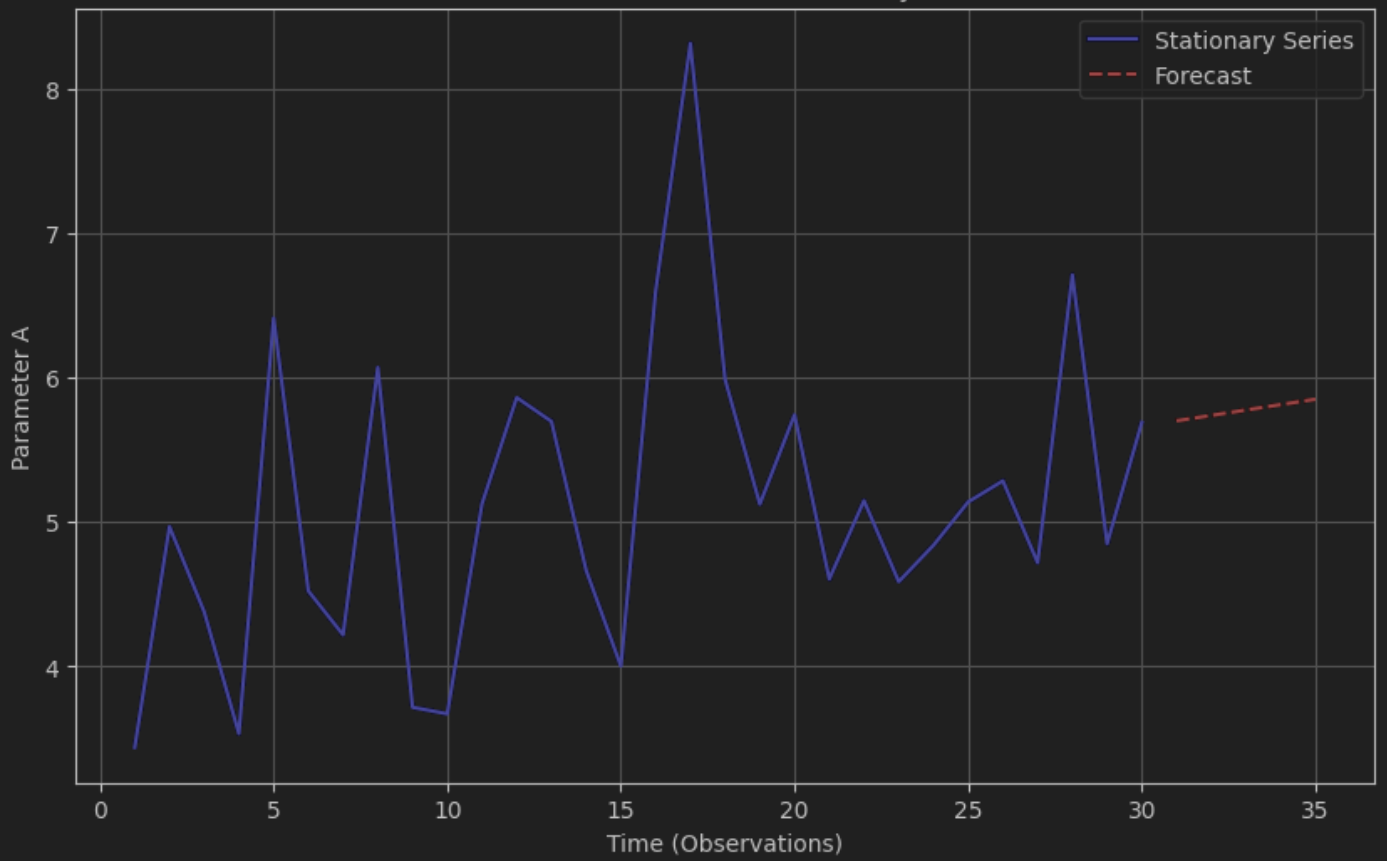
Для частини 3.1.2, де до ряду був доданий нелінійний тренд, прогноз на наступні 5 спостережень показав значне збільшення значень: 46.685, 48.126, 49.566, 51.007 та 52.447. Це підтверджує, що доданий тренд значно впливає на динаміку даних, роблячи їх нестационарними. Прогнозування за допомогою лінійної авторегресійної моделі виявилось менш точним через нелінійний характер змін у даних.

Візуально порівнюючи ці два результати, можна побачити, що для стаціонарного ряду прогнози зберігають постійний тренд, тоді як для нестационарного ряду значення прогнозу суттєво зростають. Це демонструє важливість врахування тренду при прогнозуванні нестабільних серій.

3.2

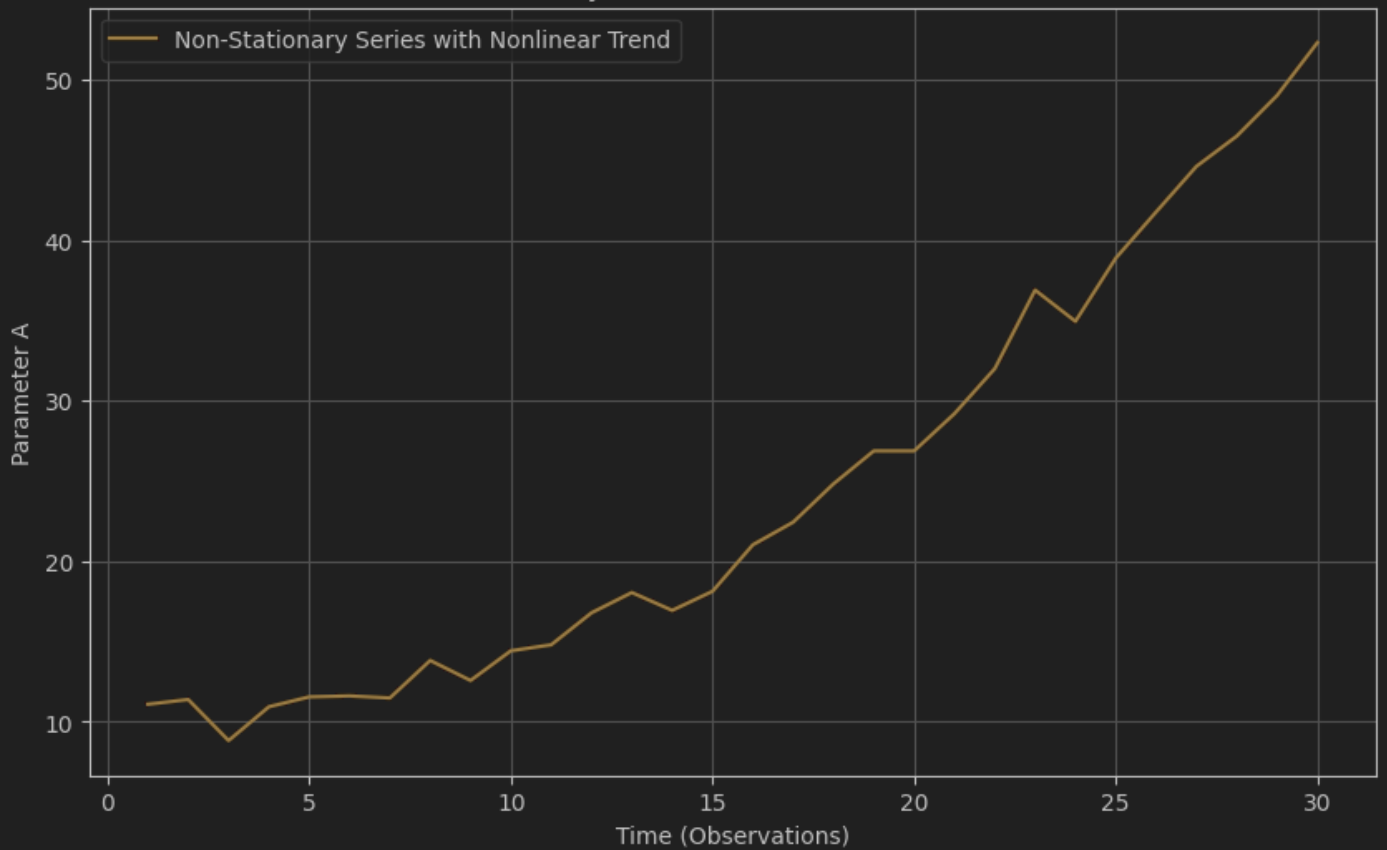


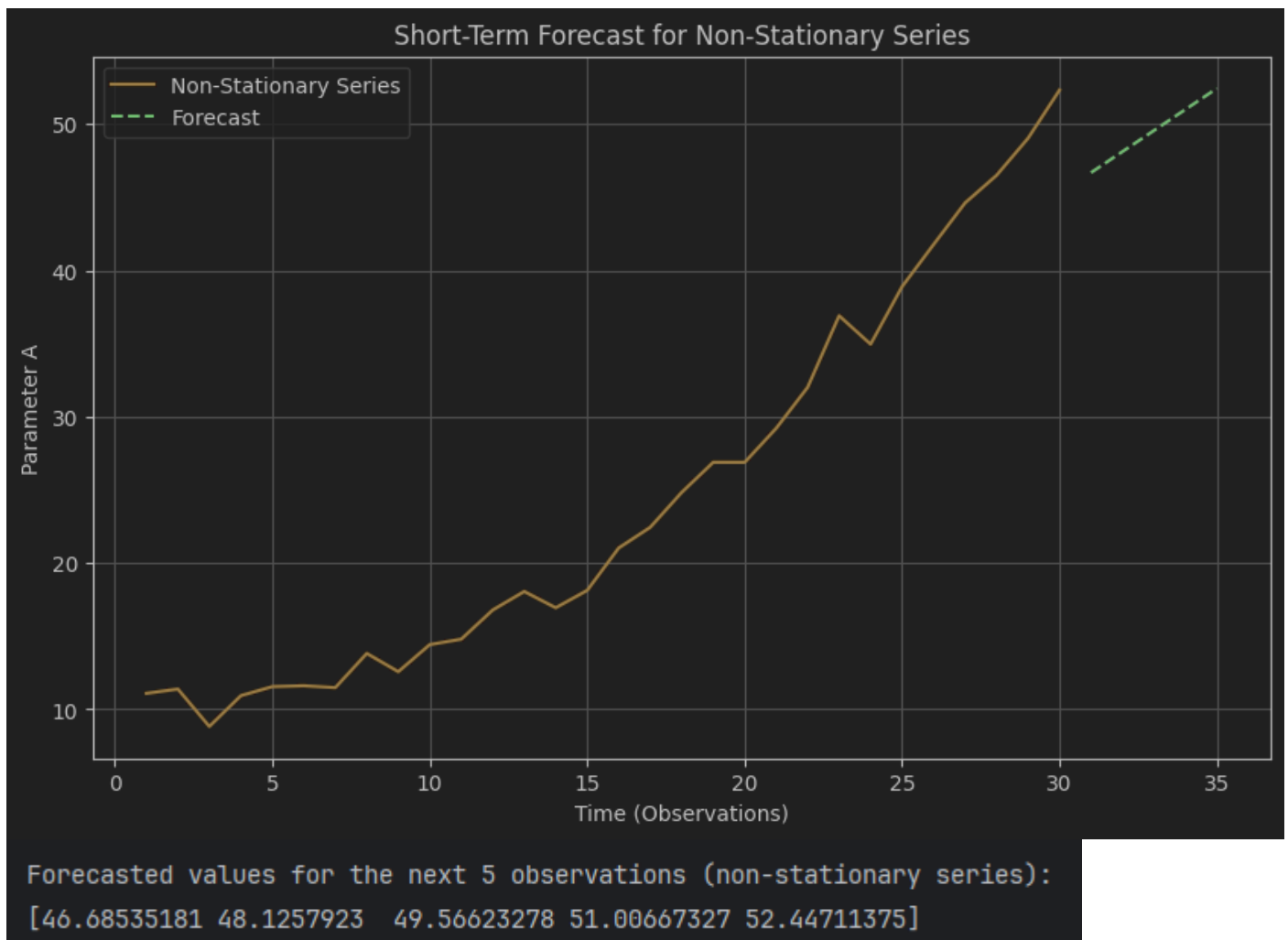
Short-Term Forecast for Stationary Series



Forecasted values for the next 5 observations (stationary series):
[5.70106433 5.73847626 5.77588819 5.81330012 5.85071205]

Non-Stationary Time Series with Nonlinear Trend

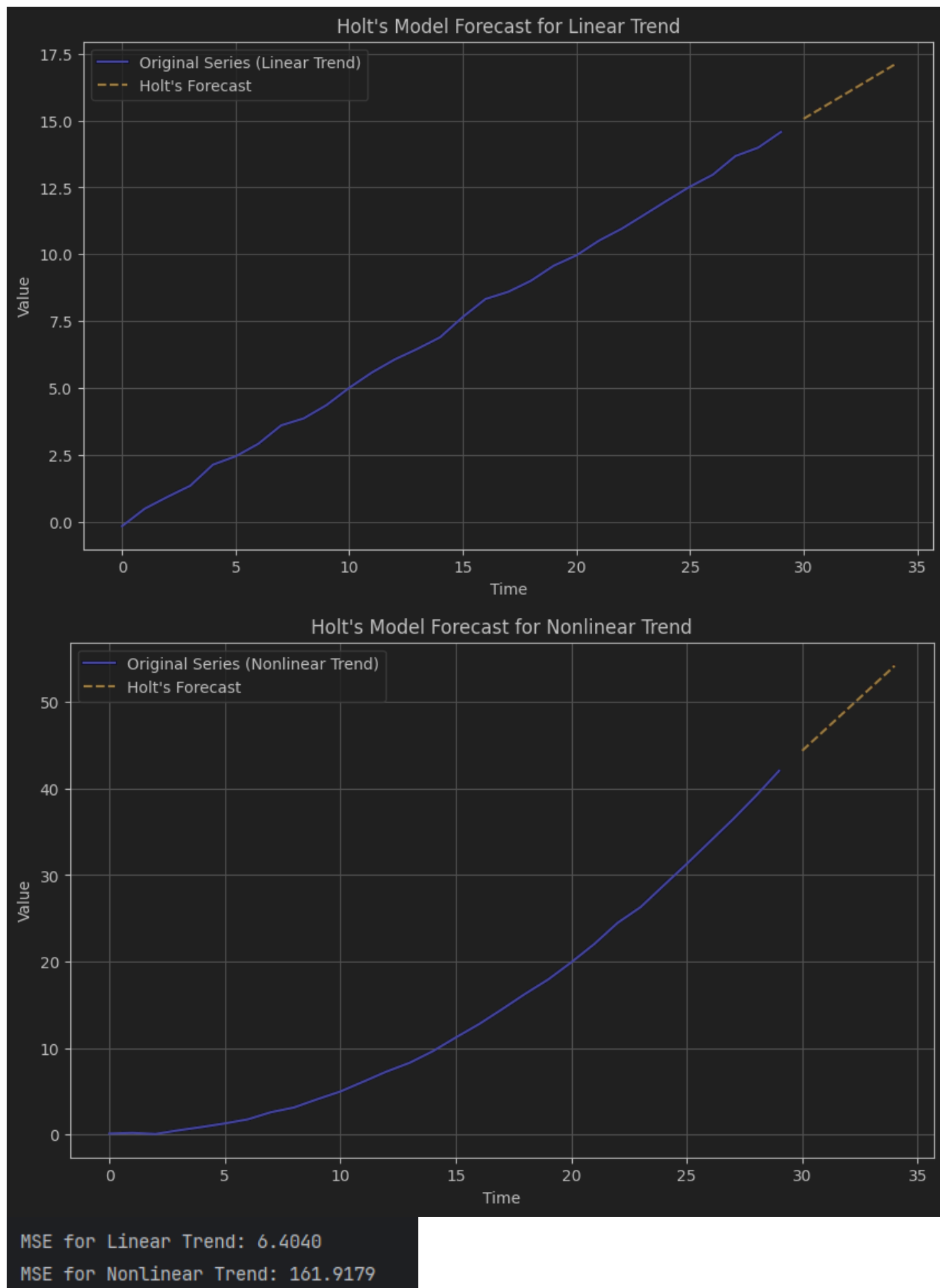




Для частини 3.2.1, де було здійснено прогноз для стаціонарного ряду за допомогою авторегресійної моделі Брауна, ми отримали прогнозовані значення для наступних 5 відліків, які відповідають таким самим трендам, як і в попередньому завданні. Це свідчить про те, що модель Брауна добре справляється з прогнозуванням стаціонарних рядів, де зміни в даних є мінімальними та стабільними. Прогнозовані значення: 5.701, 5.738, 5.776, 5.813 і 5.851 не виявляють істотних змін.

Для частини 3.2.2, де був використаний нестаціонарний ряд із доданим нелінійним трендом, результат прогнозування показав таке ж збільшення значень, як і в попередньому завданні: 46.685, 48.126, 49.566, 51.007 і 52.447. Модель Брауна намагається передбачити тренд, але через нелінійність даних точність прогнозу є меншою, порівняно з прогнозами для стаціонарного ряду.

Візуальне порівняння показує, що модель Брауна більш точно прогнозує стаціонарні ряди, де зміни незначні, тоді як для нестаціонарного ряду точність прогнозу знижується через доданий тренд. Це підтверджує важливість вибору відповідної моделі для типу ряду.



У завданні 3.3.1, де використано авторегресійну модель Хольта для прогнозування ряду з лінійним трендом, результати показали, що модель добре справляється з прогнозуванням даних із малими випадковими коливаннями. Можна спостерігати значення середньоквадратичної помилки (MSE) на рівні 6.404, що свідчить про адекватність прогнозу при наявності лінійного тренду, де різниця між початковими та кінцевими значеннями значна, а випадкова складова незначна.

У завданні 3.3.2, де дані містять нелінійний тренд, модель Хольта дала значно вищу середньоквадратичну помилку — 161.918. Це підтверджує, що модель не є оптимальною для прогнозування рядів з нелінійними трендами, де зміни є більш складними і непередбачуваними. Візуальне порівняння результатів прогнозування для лінійного та нелінійного рядів показує, що для лінійного тренду модель працює краще, а для нелінійного — прогноз має більшу похибку.

Висновки:

У ході виконання лабораторної роботи були досліджені різні моделі прогнозування, зокрема метод найменших квадратів, модель Брауна та модель Хольта. Для стаціонарних рядів, де зміни невеликі, метод найменших квадратів та модель Брауна продемонстрували хороші результати з низькою похибкою при короткостроковому прогнозуванні. У випадку рядів з нелінійними трендами моделі не змогли адекватно передбачити зміни, що призвело до значно більшої похибки, що показує необхідність застосування спеціалізованих методів для таких даних. Модель Хольта продемонструвала кращі результати для лінійних трендів, однак для даних з нелінійним трендом вона дала значну помилку, що підтверджує її обмеження для складніших залежностей. Загалом, різні моделі вимагають належного підбору залежно від типу даних та характеру їх змін.