

Київський національний університет імені Тараса Шевченка
Факультет комп'ютерних наук та кібернетики
Кафедра дослідження операцій

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА
за спеціальністю 113 «Прикладна математика»
на тему:

**Чисельне прогнозування показників на основі актуальних
відомостей**

Пептюк

студента 4 курсу
Пептюка Ярослава Андрійовича

Аее

Науковий керівник:
Асистент, доктор філософії
Тимошенко Андрій Анатолійович

Робота заслухана на засіданні кафедри дослідження операцій та
рекомендована до захисту в ЕК, протокол № 9 від 23.05 2023 р.

Завідувач кафедри ДО

Икср

проф. Іксанов О.М.

Київ – 2023

Зміст

ВСТУП	4
Розділ 1. Теорія чисельних методів у прогнозуванні показників	6
1.1 Розвиток чисельних методів у прогнозуванні економічних показників ..	6
1.2 Опис понять та термінів пов'язаних з прогнозуванням та чисельним аналізом.....	7
1.3 Огляд основних чисельних методів, що використовуються у прогнозуванні.....	8
1.3.1 Методи часового ряду	8
1.3.2 Методи машинного навчання.....	9
1.3.3 Методи інтерполяції та екстраполяції.....	10
1.4 Обґрунтування вибору конкретного методу прогнозування	11
Розділ 2. Збір та підготовка даних	13
2.1 Визначення даних для прогнозування	13
2.2 Підготовка даних до аналізу	14
2.3 Вибір методу для прогнозування.....	17
Розділ 3. Розробка моделі прогнозування	19
3.1 Алгоритм прогнозування даних методом ARIMA.....	19
3.2 Алгоритм прогнозування даних методом кубічної сплайн-інтерполяції	21
3.3 Обґрунтування конкретного програмного забезпечення.....	24
Розділ 4. Реалізація моделі прогнозування	30
4.1 Реалізація моделі ARIMA	30
4.2 Реалізація методу кубічної сплайн-інтреполяції.....	49
4.3 Аналіз та порівняння результатів.....	59
ВИСНОВКИ	64
Список використаних джерел.....	66

Додаток А. Код програмної реалізації для розділу 4	68
---	-----------

ВСТУП

Наукові дослідження, пов'язані з чисельним прогнозуванням показників на основі актуальних відомостей, є дуже популярним напрямом досліджень у сучасному науковому та виробничому середовищі. Чисельні методи використовуються для розв'язання складних задач прогнозування, де традиційні аналітичні підходи можуть бути обмежені або недостатньо ефективні.

Чисельні методи включають в себе широкий спектр алгоритмів та процедур, які базуються на математичних методах та побудовах різних статистичних моделей. Вони дозволяють наближено розв'язувати складні математичні задачі, з великою кількістю вхідних даних та складних взаємозв'язків між ними. Варіації чисельних методів охоплюють велику кількість підходів, включаючи методи інтерполяції, сплайнів, чисельного диференціювання та інтегрування, методи найменших квадратів, оптимізаційні методи, методи регресії та багато інших. Кожен з цих методів має свої особливості, переваги та недоліки, що дозволяють вирішувати конкретні завдання прогнозування [10]. Більш того, багато методів можуть використовуватися разом з іншими, а в результаті такої співпраці отримувати більш якісні результати.

Застосування чисельних методів у прогнозуванні показників дозволяє отримати значимі результати у різних сферах, включаючи економіку, фінанси, маркетинг, соціальні науки та інші. Ці методи дозволяють аналізувати інформацію з великого обсягу даних, виявляти тенденції, робити прогнози та приймати рішення на основі математичних розрахунків.

Для досягнення мети прогнозування показників на основі актуальних відомостей використовуються різні методи дослідження, такі як статистичний аналіз даних, машинне навчання, штучні нейронні мережі та інші [19].

Одержані результати досліджень з чисельного прогнозування показників на основі актуальних відомостей можуть мати практичний застосування. Наприклад, бути використані для прийняття рішень в управлінні, плануванні, стратегічному аналізі та інших сферах діяльності. Результати досліджень можуть також знайти застосування в інших дослідницьких роботах та статтях, сприяючи розвитку відповідної галузі науки та виробництва.

Отже, у даному дослідженні буде розглянуто проблему чисельного прогнозування показників на основі актуальних відомостей, використовуючи різні чисельні методи. Використання цих методів дозволить отримати цінні результати та висунути нові наукові положення щодо прогнозування в різних галузях. Ми зможемо порівняти отримані результати та зробити висновки, щодо застосованих методів прогнозування. Робота буде поділена на розділи, в кожному з яких будуть розглянуті певні аспекти проблеми та застосовані відповідні методи дослідження.

Розділ 1. Теорія чисельних методів у прогнозуванні показників

1.1 Розвиток чисельних методів у прогнозуванні економічних показників

У минулому столітті з'явилась потреба у більш точних та науково обґрунтованих методах прогнозування економічних показників. Це призвело до розвитку чисельних методів, які базуються на математичних моделях та статистичному аналізі. Перші згадки про використання чисельних методів у економічних дослідженнях можна віднести до середини ХХ століття [9].

Один з ранніх прикладів застосування чисельних методів у економічних прогнозах пов'язаний з регресійним аналізом. Відомий економіст Джон Мейнард Кейнс у своїх працях зазначав про важливість використання регресійних моделей для прогнозування економічних показників. Цей метод дозволяв оцінити залежність між різними змінними та передбачити їхнє майбутнє значення на основі наявних даних. Сама ж економетрика як наука була винайдена у 1930-му році [9].

Інший важливий розвиток в галузі чисельних методів пов'язаний зі статистичним аналізом часових рядів. Часові ряди є одним з ключових джерел інформації для прогнозування економічних показників, оскільки вони відображають залежності та тренди в часі. Використання статистичних методів, таких як ARIMA (Autoregressive Integrated Moving Average), дало змогу зробити більш точні прогнози на основі історичних даних. Деякі згадки про чисельні методи у прогнозуванні економічних показників включають використання інтерполяції та екстраполяції для заповнення пропусків в даних та прогнозування майбутніх значень, а також використання методів оптимізації для знаходження оптимальних параметрів моделей.

Історія розвитку чисельних методів у прогнозуванні економічних показників демонструє, що вони почали використовуватись як інструменти для покращення точності та наукової обґрунтованості прогнозів. Ці перші кроки стали основою для подальшого розвитку чисельних методів та їхнього використання в сучасному прогнозуванні в економічній та інших сферах суспільства.

У сучасному світі економічне середовище потребує все більш математично

точних та складних моделей для прогнозів та аналізу показників. З'являється все більше перспективних напрямків розвитку чисельних методів у цій сфері. Наприклад, штучний інтелект та машинне навчання, які дозволяють автоматизувати процес виявлення складних залежностей в економічних даних. Вони дозволяють прогнозувати майбутні значення з великою точністю. Нестационарні часові ряди та адаптивні моделі допомагають працювати з нестационарними економічними процесами. Методи що використовуються для них, як приклад, моделі з параметром, змінюваним у часі. Також, часто для прогнозування економічних значень, можна зустріти використання додаткових даних з інших галузей. Роблячи комбінації фінансових моделей, макроекономічних моделей, моделей поведінки споживача можна знайти кореляції між аспектами економічного розвитку.

1.2 Опис понять та термінів пов'язаних з прогнозуванням та чисельним аналізом

Наведемо список основних понять та термінів, які будуть використовуватися впродовж дослідження [16].

- Прогнозування – процес визначення майбутніх значень на основі актуальних відомостей. Включає в себе аналіз тенденцій, виявлення залежностей між даними та використання різних методів для отримання результату (прогнозу).
- Чисельний аналіз – методологія, що використовує числові методи для аналізу даних. Дозволяє знаходити характеристики даних такі як тренди, сезонність та багато інших.
- Вхідні дані – набір даних, що є основним для здійснення подальшого аналізу та прогнозу. Прикладом таких даних є числові ряди і числові значення.
- Вихідні дані – прогнозовані значення, отримані в результаті застосування чисельних методів до набору вхідних даних.
- Тренд – загальна тенденція зміни та розвитку показників з часом. Тренд може бути як лінійний (зростаючим чи спадаючим) так і нелінійний.
- Сезонність – періодичні коливання показників, що повторюються з певним інтервалом. Може бути річною, місячною, тижневою або іншою.

Далі перейдемо до розгляду конкретних методів прогнозування та аналізу вхідних даних.

1.3 Огляд основних чисельних методів, що використовуються у прогнозуванні

Розглянемо основні методи, що застосовуються для прогнозування та аналізу показників на основі актуальних відомостей. Кожен з методів може бути застосований для прогнозування даних, але слід уважно підходити до питання вибору конкретного методу відштовхуючись від вхідних даних.

1.3.1 Методи часового ряду

Часові ряди – ряд точок даних, що впорядковані у хронологічному порядку. Найчастіше представлення такого ряду є послідовністю. В прогнозуванні даних методи часового ряду є одними з найпоширеніших та одними з найбільш ефективних методів.

Сам метод прогнозування ґрунтується на припущенні залежності майбутніх значень показників від попередніх, тих що зустрічалися раніше. Зазвичай прогнозування включає в себе екстраполявання, тобто перенесення знайдених закономірностей на майбутнє. Метод екстраполяції ми розглянемо в наступних підрозділах окремо.

Прикладами методів часових рядів є просте експоненційне згладжування (SES), авторегресійна модель з ковзним середнім (ARMA), модель авторегресії з інтегрованим ковзним середнім (ARIMA) та інші [2].

Просте експоненційне згладжування (SES) – метод прогнозування часових рядів, полягає в тому, що у кожній наступній ітерації враховуються всі попередні значення ряду:

$$Y(t) = \alpha X(t) + (1 - \alpha)Y(t - 1)$$

Тут $Y(t)$ – прогнозоване значення на час t ,
 $X(t)$ – спостережуване значення на час t ,
 $Y(t - 1)$ – відповідно попереднє прогнозоване значення,
а α – коефіцієнт згладжування ($0 < \alpha < 1$).

Модель авторегресії з ковзним середнім (ARIMA) – модель, що включає в себе авторегресійну складову (AR), складову з інтегровним рядом (I) та ковзне середнє (MA) [16]. Модель ARIMA є більш гнучкою ніж ARMA та може забезпечувати більш точні прогнози для складних часових рядів:

ARIMA(p, d, q):

$$Y(t) = c + \varphi_1 Y(t - 1) + \dots + \varphi_p Y(t - p) + \theta_1 e(t - 1) + \dots + \theta_q e(t - q)$$

Тут c – константа,
 $\varphi_1, \dots, \varphi_p$ – коефіцієнти авторегресії,
 $\theta_1, \dots, \theta_q$ – коефіцієнти ковзного середнього,
а $e(t - 1), \dots, e(t - q)$ – залишкові члени.

1.3.2 Методи машинного навчання

Методи машинного навчання є одними з найсучасніших методів прогнозу даних, адже прикладом такого методу є використання нейронних мереж. Крім цього, методи машинного навчання включають в себе регресійні моделі.

Регресійні моделі використовуються для прогнозування значень показників на основі інших залежних змінних [4]. Вони побудовані на статистичних методах, що використовують як лінійні так і нелінійні залежності між змінними. Для досягнення оптимальних прогнозів, регресійні моделі вимагають вибору відповідного алгоритму, оцінку пріоритетності змінних та налаштування окремих параметрів моделі.

Нейронні мережі – потужний інструмент машинного навчання для

прогнозування й аналізу. Їх основою є моделювання нейронних структур, які складаються з трьох типів шарів: вхідних, прихованих та вихідних. Нейронні мережі здатні автоматично виявляти найскладніші залежності між вхідними та вихідними наборами даних. Нейронні мережі також вимагають налаштування та оптимізації параметрів моделі для більш точного моделювання.

1.3.3 Методи інтерполяції та екстраполяції

Методи інтерполяції та екстраполяції чудово підходять для прогнозування та аналізу показників на основі актуальних відомостей. Вони дозволяють встановлювати залежності між змінними та отримувати прогнози для відсутніх значень в початковому наборі, та майбутніх значень з певною точністю.

Метод інтерполяції базується на припущенні, що між вхідними точками існує гладка функція, що описує залежність між показниками. Прогнозовані значення в цьому методі отримуються якраз із цієї гладкої функції [18].

Одним з найпоширеніших методів інтерполяції є лінійна інтерполяція. Цей метод полягає в побудові прямих лінійних сегментів між парами сусідніх точок. Для прогнозування значення будь-якої проміжної точки між двома відомими, лінійна інтерполяція застосовує рівняння прямої:

$$y = y_1 + (x - x_1)(y_2 - y_1)(x_2 - x_1)$$

де x – наша проміжна точка, для якої треба спрогнозувати значення y , а (x_1, y_1) та (x_2, y_2) – дві відомі точки.

Наступний, більш точний метод інтерполяції – сплайн-інтерполяція. Він використовує поліноми нижчого порядку для побудови гладкої кривої, що проходить через усі відомі точки [3]. Найчастіше, для побудови використовують кубічні сплайни, які складаються з поліномів третього порядку та добре описує локальні залежності між точками, забезпечуючи гладку зміну значень між ними. Вигляд кубічного сплайну:

$$s(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

де x_i та x_{i+1} – відомі точки, a_i, b_i, c_i, d_i – коефіцієнти кубічного сплайна.

Звичайно, виникають умови, виконання яких є обов'язковим для застосування методу, та про них згадаємо пізніше, під час обговорення побудови моделі для прогнозування.

Перейдемо до методу екстраполяції. Метод полягає в продовженні тенденції який ми могли спостерігати в наборі вхідних даних на майбутні значення. Екстраполяція може бути лінійною, поліноміальною, експоненціальною, логарифмічною чи ще більш складною моделлю. Для кожного типу екстраполяції використовують відповідні математичні формули. Проте, слід враховувати, що екстраполяція є більш ризикованим методом в прогнозуванні, ніж, наприклад, модель ARIMA. Екстраполяція не має рекурсії і сильно залежить від правильності припущень, що ми робимо для початкового набору даних

1.4 Обґрунтування вибору конкретного методу прогнозування

Проведемо порівняння різних методів прогнозування і обґрунтуємо вибір конкретного методу для прогнозів. Існує багато факторів, які можуть впливати на обрання методу прогнозування [12]. Наприклад: точність прогнозування, розмір та якість початкової вибірки даних, обчислювальні вимоги, доступність програмного забезпечення та інші.

Почнемо з методів часового ряду, зокрема просте експоненційне згладжування (SES) та модель авторегресії (ARIMA).

Просте експоненційне згладжування – простіший метод з низькими обчислювальними вимогами. Підходить для прогнозування даних без сезонності чи зі стійким трендом. Проте, метод не враховує складні закономірності та залежності між значеннями.

Моделі авторегресії – складніша та більш універсальна модель. При обчисленнях вона враховує тренд, сезонність, автокореляцію та залишкові члени. ARIMA покращує результати для складних даних з трендом та сезонністю. Однак, модель

вимагає більшого набору вхідних даних та більше часу для обчислень.

Отже, для більшої точності обчислень при наявності досить великої вибірки даних та достатньої кількості часу краще застосовувати метод ARIMA

Методи машинного навчання є більш гнучкими і можуть виявляти складні залежності в даних. Проте, враховуючи обмеження вхідних даних, розмір вибірки та обчислювальні обмеження, використання машинного навчання для прогнозування може бути складним. Вони часто вимагають великої обчислювальної потужності та великої кількості даних для навчання. Через це, детально розглядати методи машинного навчання надалі я не збираюся.

Метод інтерполяції – класичний підхід до прогнозування на основі актуальних відомостей. Оскільки метод базується на побудові функції, що проходить через всі точки, робиться припущення про неперервну залежність між даними.

Метод сплайн-інтерполяції – побудова сплайнів – гладких кривих, що проходять через всі точки. Використовуючи сплайни, ми враховуємо локальні особливості в даних, тобто стрибки, вигини та інші неоднаковості.

Розділ 2. Збір та підготовка даних

2.1 Визначення даних для прогнозування

Процес збору та обробки даних для подальшого прогнозування є важливим початковим кроком в роботі, адже з «погано» підготовленими даними робити аналіз може бути неможливо.

Я вирішив робити прогнози для зовнішньоекономічних показників України та Німеччини за єдиний період часу, щоб пізніше була змога порівняти результати. Такі дані, будуть корисні в подальшому для експертів в відповідних сферах.

Розглянемо щомісячний імпорт та експорт з квітня 2022 року по березень 2023 року [1]. В таблиці 2.1 наведено дані для України [20], а в таблиці 2.2 для Німеччини [6].

Таблиця 2.1 – Дані експорту та імпорту для України

Україна	(млн.дол. США)	
	Експорт	Імпорт
Період	сезонно скориговані дані	сезонно скориговані дані
Квітень 2022	2685,1	2823,5
Травень 2022	2849,5	3857,9
Червень 2022	3361,2	4849,9
Липень 2022	3157,1	4421,1
Серпень 2022	3189,2	4446,0
Вересень 2022	4062,7	4571,2
Жовтень 2022	3775,0	4838,9
Листопад 2022	3541,7	4921,6
Грудень 2022	3361,7	5353,6
Січень 2023	3273,2	5561,0
Лютий 2023	3508,0	5624,1
Березень 2023	3544,9	5404,2

Таблиця 2.2 – Дані експорту та імпорту для Німеччини

Німеччина	(млн.дол. США)	
Період	Експорт	Імпорт
	сезонно скориговані дані	сезонно скориговані дані
Квітень 2022	134541,9	131985,8
Травень 2022	143688,5	138522,6
Червень 2022	144407,5	135981,1
Липень 2022	130674,7	125715,8
Серпень 2022	130879,4	130434,0
Вересень 2022	141157,5	133335,0
Жовтень 2022	133361,4	128480,9
Листопад 2022	147079,9	135043,7
Грудень 2022	131935,7	120538,3
Січень 2023	136077,3	125329,6
Лютий 2023	141889,4	123700,9
Березень 2023	155982,0	131892,2

2.2 Підготовка даних до аналізу

Спочатку треба перевірити якість даних. Під якістю маються на увазі помилки та пропущені значення. В загальному алгоритм обробки даних має такий вигляд:

1. Ідентифікація помилкових та пропущених даних. Виявлення пропущених значень, порівняння даних з іншими наявними чи з даними відомих джерел, пошук аномалій та відхилень через статистичні методи. Взагалі, існує багато методів, що дозволяють шукати аномалії – метод Z-оцінки, метод Грабса та інші [11]. Наприклад, аномалії можна знайти поширеним та простим методом міжквартильного діапазону:

- Обчислити міжквартильний діапазон (IQR)

$$IQR = Q3 - Q1$$

Де $Q1$ – перший квантиль (25-й перцентиль), а $Q3$ – третій квантиль (75-й перцентиль).

- Встановити верхню та нижню межі: $Q3 + 1.5 \cdot IQR$, $Q1 - 1.5 \cdot IQR$.
- Відкинути спостереження (викиди), що знаходяться поза межами.

В нашому випадку пропущених та помилкових даних немає.

2. Нормалізація даних. Іншими словами – преведення даних до одного масштабу.

Одним з найпоширеніших методів нормалізації є мінімаксна нормалізація [8] (масштабування):

Кожному значенню x ставимо у відповідність нормалізоване значення

$x_{norm} = (x - x_{min}) / (x_{max} - x_{min})$, де x_{max} та x_{min} – мінімальне та максимальне значення у виборці відповідно. Очевидно, що всі нормалізовані значення належать відрізьку $[0, 1]$.

3. Візуалізація даних. Найпростіший спосіб переконатися в правильності даних – відобразити їх на графіку. Для нашого випадку графіки мають такий вигляд – рисунок 2.1 – дані України, рисунок 2.2 – дані Німеччини:

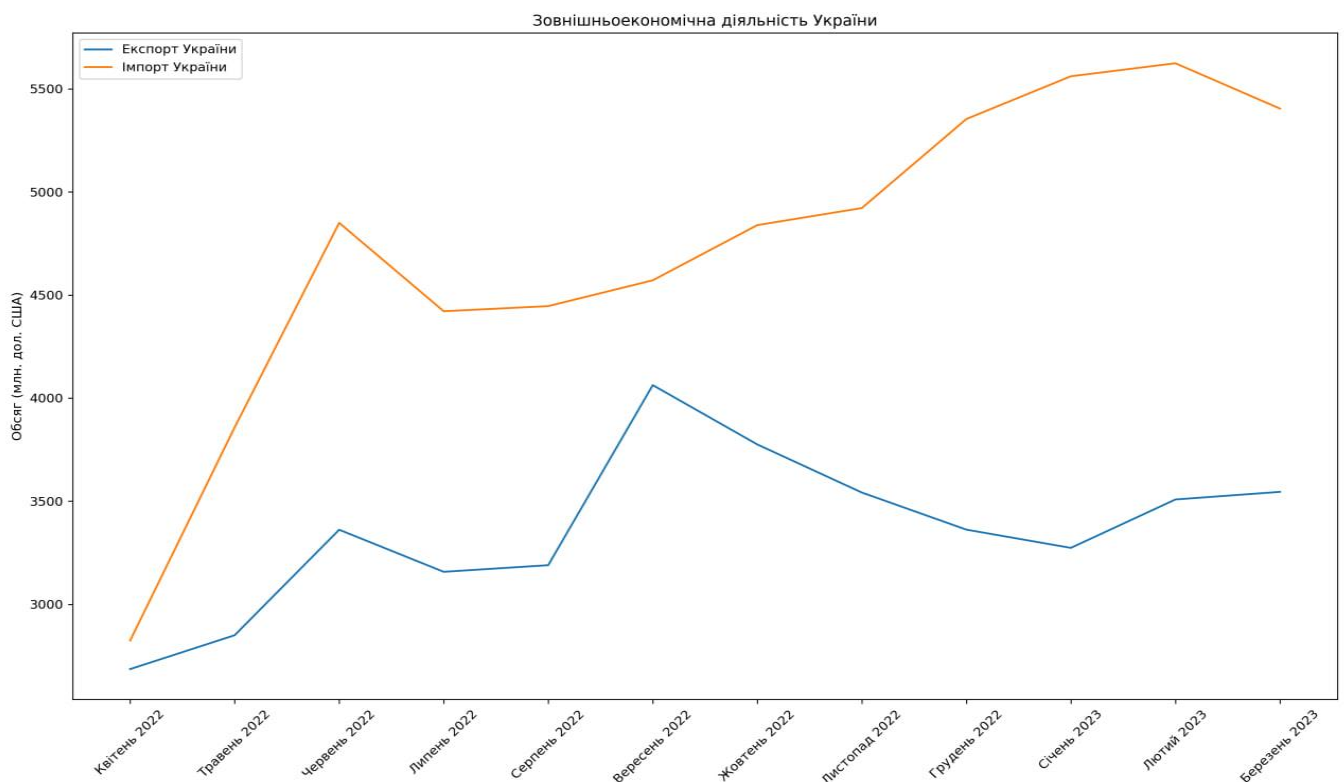


Рис 2.1.. Імпорт та Експорт України

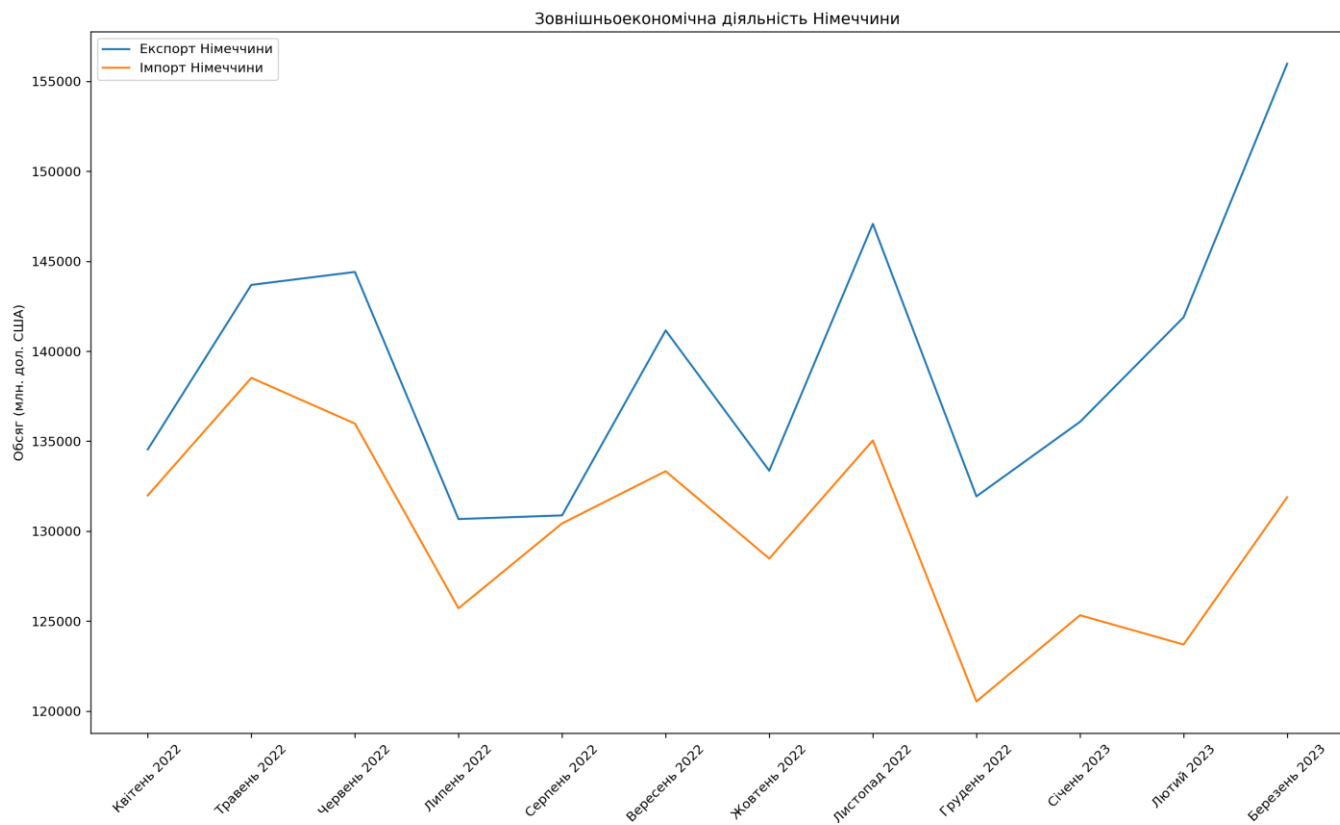


Рис 2.2. Імпорт та Експорт Німеччини

2.3 Вибір методу для прогнозування

Ми маємо зовнішньоекономічні показники, що представлені статистичним набором даних. Оскільки всі дані задаються для однакових інтервалів часу, буде зручно використовувати вищезгадані методи, що дозволять нам побудувати функції, що наближають графік зміни цих показників у часі та передбачать їх зміни в майбутні періоди часу.

В рамках нашої роботи, прогнозування будемо робити методом ARIMA та методом сплайн-інтерполяції. В попередньому розділі ми вже наводили огляд цих двох методів, та давайте трохи детальніше покажемо як цими методами можна проводити прогнозування даних.

Кубічна сплайн-інтерполяція – один з найпоширеніш методів інтерполяції, що застосовується в економетриці. Базується на побудові гладких кривих, що максимально наближають графік змін наших показників, за умовою того, що нам вже відомий певний набір значень. Інтервальні дані, які ми маємо зараз чудово підходять для інтерполювання. Крива сплайну розбивається на сегменти і на кожному з них ми апроксимуємо дані будуючи кубічний поліном. Таким чином будуть плавно враховані зміни в даних, їх тенденції.

Перевагами кубічної сплайн-інтерполяції є здатність адаптуватися до складних змін в наборі даних для аналізу, враховувати нелінійні тенденції та робити плавні прогнози. Крім того, побудова моделі не викликає великих труднощів, і вимоги щодо даних є не надто складно виконуваними.

ARIMA – один з найпоширеніш методів для аналізу та прогнозування часових рядів. Фактично, ця модель є комплексною, що складається з більш простих моделей, іншими словами поскладніше своїх попередників поступово (AR, MA, ARMA) [16].

- AR – модель авторегресії, що використовує зв'язок між наявними спостереженнями та спостереженнями із затримкою;
- I – інтегрованість, тобто застосування методу різниць для зведення вхідного часового ряду до стаціонарного;
- MA – модель рухомого середнього, що використовує залежність між наявними спостереженнями та залишковими членами.

Метод включає в себе три основні параметри:

- p – кількість лагів авторегресії,
- d – порядок різницевого перетворення,
- q – кількість лагів рухомого середнього.

Модель ARIMA дозволяє узагальнити автокореляційну структуру нашого часового ряду, включаючи як авторегресійну (AR), так і рухому середню (MA) компоненту. ARIMA моделює досить складні динамічні залежності в даних та прогнозує майбутні значення для наших часових рядів.

Перевагами моделі ARIMA є здатність моделювання лінійних та нелінійних залежностей в рядах, врахування сезонностей та змінюваної дисперсії.ч’

Обрані методи сплайн-інтерполяції та ARIMA дозволять зробити досить точні та релевантні прогнози майбутніх значень експорту та імпорту. На результатах такого дослідження зможемо виявити слабкі та сильні сторони кожного з методів.

Розділ 3. Розробка моделі прогнозування

3.1 Алгоритм прогнозування даних методом ARIMA

Побудуємо алгоритм прогнозування методом ARIMA.

Для початку слід знайти параметри моделі q , p , та d :

q – кількість лагів рухомого середнього, тобто порядок рухомого середнього.

Цей параметр шукається з аналізу ACF – автокореляційної функції. Зауважимо, що ACF – це графік загальної кореляції між різними функціями затримок. Отже, розглянемо графік автокореляційної функції та отримуємо з нього параметр q .

p – кількість лагів авторегресії, тобто кількість спостережень з затримкою, які включені в нашу модель. Шукається з аналізу графіка PACF – частково автокореляційної функції.

d – порядок різницевого перетворення, тобто скільки разів до спостереження застосовують метод різниць. Наш часовий ряд вхідних даних має бути стаціонарним для застосування методу ARIMA. Для того щоб перевірити стаціонарність ряду можна виконувати різні тести, наприклад, тест Дікі-Фуллера [10]. Якщо ряд виявляється нестаціонарним виконуємо диференціювання, тобто метод різниць. Тут параметр d і стане порядком диференціювання, тобто кількістю разів виконання такої операції.

Тест Дікі-Фуллера є одним з найпопулярніших тестів для перевірки часового ряду на стаціонарність. Тест має наступний вигляд:

$$Y(t) = (\rho - 1)Y(t - 1) + \varepsilon(t)$$

ε – похибка, $Y(t)$ – значення ряду. Треба виконати перевірку чи істотно вираз $(\rho - 1)$ відрізняється від нуля. Нульова гіпотеза буде мати такий вигляд:

$$H_0: \rho - 1 = 0$$

Якщо ж нульову гіпотезу рівності нулю виразу $(\rho - 1)$ відхиляємо, то ми отримали, що наш часовий ряд є стаціонарним.

У випадку коли ряд не є стаціонарним, ми будемо виконувати диференціювання порядку d [16]. Можна застосувати різницеву формулу

$$Y'(t) = Y(t) - Y(t - 1)$$

Тоді різницю $Y'(t)$ будемо називати першою різницею. Якщо потрібно робити різниці перетворення, то маємо знаходити ACF та PACF для нового ряду, адже значення цих функцій можуть змінитися. З графіків ACF та PACF ми не завжди можемо однозначно отримати значення параметрів кількості лагів авторегресії p та кількості лагів рухомого середнього q . Тому інколи доцільно обирати декілька значень для кожного з цих параметрів та проводити обчислення для набору моделей, що відповідають різним значенням цих параметрів.

Ми повинні провести розрахунок коефіцієнтів $\varphi_1, \dots, \varphi_p$ авторегресії, коефіцієнтів $\theta_1, \dots, \theta_q$ рухомого середнього, константи c та $e(t-1), \dots, e(t-q)$ – залишкових членів. Після чого отримаємо набір моделей, серед яких необхідно буде вибрати найкращу. Для цього виконаємо оцінку за критеріями AIC (Akaike Information Criterion) та BIC (Schwarz's Bayesian Information Criterion) за наступним алгоритмом:

1. Обчислюємо суму квадратів похибок моделі, яка відображає наскільки добре модель наближує наші спостереження. Ця сума позначається RSS (Residual Sum of Squares), вона є ключовою компонентою для методу найменших квадратів.
2. Визначаємо кількість параметрів моделі, позначимо через k .
3. Обчислюємо AIC за формулою:

$$AIC = 2k - 2 \ln(RSS)$$

Де RSS обчислюється як:

$$RSS = \sum (Y_i - \bar{Y})^2$$

тут Y_i – спостережуване значення часового ряду,
а \bar{Y} – середнє значення спостережень.

4. Обчислюємо BIC за формулою:

$$BIC = k \cdot \ln(n) - 2 \ln(RSS)$$

n – кількість спостережень у початковій вибірці

5. Порівнюємо значення AIC та BIC для нашого набору моделей. Модель, що має найменше значення AIC або BIC вважається найкращою з урахуваннями точності підгонки та складністю моделі. Цю модель ми й

обираємо як оптимальну для подальшого прогнозування.

Тепер опишемо процес розрахунку коефіцієнтів $\varphi_1, \dots, \varphi_p$ авторегресії, коефіцієнтів $\theta_1, \dots, \theta_q$ рухомого середнього та константи c .

Наведемо формули для обчислення відповідних значень:

1. Формули для оцінки коефіцієнтів φ_i авторегресії AR(p):

$$\varphi_i = \frac{\sum(Y(t)Y(t-i))}{\sum(Y(t-i))^2}$$

2. Формули для оцінки коефіцієнтів θ_i рухомого середнього MA(q):

$$\theta_i = \frac{\sum(e(t)e(t-i))}{\sum(e(t-i))^2}$$

3. Формули для обчислення константи c :

$$c = \frac{\sum Y - \sum \varphi_i \cdot Y_i - \sum \theta_i \cdot e_i}{n}$$

де Y_i та e_i – значення i -х спостережуваних значень та залишкових членів відповідно, а n – загальна кількість спостережень.

Обчислювати коефіцієнти не завжди зручно такими формулами, інколи застосовують метод найменших квадратів, метод Гаусса-Ньютона або інші методи. В програмних реалізаціях методу найчастіше використовують метод найменших квадратів.

Далі ми отримуємо $Y(t)$:

$$Y(t) = c + \varphi_1 Y(t-1) + \dots + \varphi_p Y(t-p) + \theta_1 e(t-1) + \dots + \theta_q e(t-q)$$

Усі майбутні значення знаходимо рекурсивно.

3.2 Алгоритм прогнозування даних методом кубічної сплайн-інтерполяції

Розглянемо загальний алгоритм побудови та прогнозування методом сплайн-інтерполяції

1. Розбиття на сегменти:

- Всі точки включені в сегменти
- Кожен сегмент складається з двох сусідніх точок

2. Будуємо кубічні функції:

- Для кожного сегменту визначаємо кубічну функцію
3. Розв'язання системи рівнянь:
- Для кожного сегмента будуємо рівняння за граничними умовами
 - Зазвичай граничні умови першого чи другого роду
 - З отриманої системи рівнянь методами розв'язання систем рівнянь (наприклад, метод Гаусса чи LU-розклад) отримуємо коефіцієнти
4. Побудова сплайну:
- Об'єднуємо всі кубічні функції
 - Утворено гладку, плавну криву апроксимації вхідних точок
5. Інтерполювання та прогнозування
- З отриманої функції можемо знайти проміжні значення між вхідними точками
 - Продовживши функцію на майбутній проміжок, отримаємо прогнозовані значення

Розглянемо детальніше кожен з кроків побудови кубічного сплайну. Маємо початковий набір даних $a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$.

Припускаємо, що існує неперервна на $[a, b]$ функція $Y(t)$, позначимо $Y(t_i) = Y_i$. Тоді сплайном цієї функції $Y(t)$ та вузлів $(t_i), i = 0, \dots, n$, називатимемо функцію $s(t)$, що задовільняє такі умови [18]:

- (1) На кожному сегменті $[t_{i-1}, t_i], i = 1, \dots, n$, функція $s(t)$ є поліномом третього степеня;
- (2) Неперервними на $[a, b]$ є сама функція $s(t)$, її перша та друга похідні;
- (3) Функція $s(t)$ співпадає з $Y(t)$ в вузлах $(t_i), i = 0, \dots, n$.

На кожному з сегментів $[t_{i-1}, t_i], i = 1 \dots n$ шукаємо функцію $s(t)$ у вигляді її складової $s_i(t)$ так:

$$s_i(t) = a_i + b_i(t - t_i) + \frac{c_i}{2}(t - t_i)^2 + \frac{d_i}{6}(t - t_i)^3$$

$$t_{i-1} \leq t \leq t_i, i = 1, \dots, n$$

Коефіцієнти a_i, b_i, c_i та d_i задовільняють умови (1), (2) та (3). Далі можемо побудувати систему рівнянь з якої отримаємо наші коефіцієнти для всіх сегментів.

$$s'_i(t) = b_i + c_i(t - t_i) + \frac{d_i}{2}(t - t_i)^2$$

$$s_i''(t) = c_i + d_i(t - t_i)$$

$$s_i'''(t) = d_i$$

Можемо підставити в наші рівності $t = t_i$. Звідси отримаємо:

$$a_i = s_i(t_i), b_i = s_i'(t_i), c_i = s_i''(t_i), d_i = s_i'''(t_i)$$

З умови (3) отримаємо:

$$a_i = Y(t_i)$$

Тобто

$$a_i = Y_i$$

З умови (2), а саме неперервність $s(t)$, отримаємо:

$$s_i(t_i) = s_{i+1}(t_i)$$

З попередніх результатів матимемо:

$$a_i = a_{i+1} + b_{i+1}(t - t_{i+1}) + \frac{c_{i+1}}{2}(t - t_{i+1})^2 + \frac{d_{i+1}}{6}(t - t_{i+1})^3$$

Позначимо $h_i = t_i - t_{i-1}$ і перепишемо вираз:

$$b_i h_i - \frac{c_i}{2} h_i^2 + \frac{d_i}{6} h_i^3 = Y_i - Y_{i-1}, i = 1, \dots, n$$

Аналогічно з умовами неперервності першої похідної:

$$s_i'(t_i) = s_{i+1}'(t_i)$$

$$c_i h_i - \frac{d_i}{2} h_i^2 = b_i - b_{i-1}, i = 2, \dots, n$$

І другої похідної:

$$d_i h_i = c_i - c_{i-1}, i = 2, \dots, n$$

Поєднавши усі вищеописані результати отримаємо $3n - 2$ рівняння для $3n$ невідомих (b_i, c_i та $d_i - n$ штук на кожний коефіцієнт). Два рівняння беремо з граничних умов нашого сплайну. Наприклад, нехай $Y''(a) = Y''(b) = 0$, тоді $s''(a) = s''(b) = 0$. Звідси $s_1''(t_0) = s_n''(t_n) = 0$, тобто $c_1 - d_1 h_1 = c_n = 0$.

Така система буде мати єдиний розв'язок, який можна отримати будь-яким з методів розв'язання систем. Отримавши всі коефіцієнти ми отримаємо поліном третього степеня для кожного сегменту. Склавши всі поліноми отримаємо шуканий кубічний сплайн.

Пошук коефіцієнтів також можна робити методом найменших квадратів [15] через функцію відхилень

$$E = \sum (Y_i - S(t_i))^2$$

Щодо прогнозування, метод кубічної сплайн-інтерполяції не дає змогу робити довгострокові прогнози. Проте просто побудувавши наш комплексний поліном зможете побачити найближчі зміни, які можуть здаватися цілком реалістичними. Тобто для прогнозування ми просто можемо підставити майбутній час $t = t_{n+1}$ в наш сплайн, але не варто відходити надто далеко вправо (на прямій) від точки $t = t_n = b$. В цьому ми зможемо переконатися в наступному розділі де відбудеться аналіз результатів прогнозування.

В цілому, з нашої побудови сплайна може виникнути ще одна ідея для прогнозування.

Спочатку треба визначити відрізок на якому збираємося робити прогнозування. Нехай таким відрізком буде $[t_s, t_e]$, де t_s – відповідно початкова точка (start), а t_e – кінцева (end). Якщо ми маємо деякий відрізок $[t_i, t_{i+1}]$ який містить точку t_s , то взявши кубічний поліном, що ми мали для цього відрізка зможемо знайти значення для всіх точок відріку $[t_s, t_{i+1}]$. Після цього ми можемо скористатися всіма вже знайденими кубічними поліномами для наших відрізків і додати новий відрізок який буде містити t_e – останній вузол. Тоді б ми мали змогу знайти коефіцієнти і для нього або отримати значення для $Y(t_e)$ більш близьке до істини.

3.3 Обґрунтування конкретного програмного забезпечення

Опишемо засоби програмного забезпечення, яке я використовував для реалізації алгоритмів для моделі ARIMA та кубічної сплайн-інтерполяції.

В цьому дослідженні я користувався мовою програмування Python з списком математичних бібліотек, що реалізують покроково алгоритми описані в попередніх розділах. Розглянемо ці бібліотеки більш детально [14]:

- NumPy. Бібліотека NumPy надає широкий спектр функцій для ефективно обробки числових даних, в нашому випадку часових рядів. Вона забезпечує потужні масиви даних та математичний апарат для операцій над ними. Ми використовували цю бібліотеку для обробки та аналізу даних перед

застосуванням алгоритмів прогнозування.

- **pandas.** Корисна бібліотека для маніпулювання та аналізу табличних даних. З нею можемо легко змінювати структуру даних, виконувати фільтрації та групування даних, також вміє робити обчислення. Бібліотека **pandas** часто є необхідною в роботі з іншими бібліотеками, здебільшого ми нею користувалися для підготовки даних перед застосуванням алгоритмів наших предиктивних моделей.
- **Scikit-learn.** Популярна бібліотека для підготовки та аналізу даних. Дозволяє оцінювати якість моделей, що може бути застосоване для перевірки результатів наших прогнозів. Бібліотека проінтегрована з іншими, наприклад, **NumPy** чи **pandas**, що полегшує обробку та візуалізацію даних.
- **statsmodels.** Основна бібліотека для прогнозування даних. Містить в собі велику кількість статистичних моделей, наприклад, **ARIMA** чи **SARIMA** моделі. Ми користувалися саме цією бібліотекою для прогнозу моделлю **ARIMA** та проведення подальшого аналізу результатів.
- **SciPy.** Потужна математична бібліотека, що містить усі необхідні функції для обрахунків, включаючи необхідні для інтерполяційних, оптимізаційних та інших статистичних методів. Обчислення та прогнозування методом кубічної сплайн-інтерполяції робилися саме через цю бібліотеку.
- **matplotlib.** Бібліотекою **matplotlib** ми користувалися для побудови та зображення загальних графіків з відображенням наших апроксимацій та наборами значень. Відображення деяких графіків включене в інші бібліотеки, наприклад графіки **ACF** та **PACF** робилися через бібліотеку **statsmodels**.

Обрані бібліотеки мають певні переваги, що робить їх зручними для реалізації наших алгоритмів прогнозування [17].

По-перше, всі бібліотеки мають широкий функціонал. Кожна з бібліотек надає ряд можливостей, необхідних для реалізації моделей. **NumPy** та **pandas** дозволяють ефективно адаптувати дані перед аналізом, а сам аналіз зручно виконувати через **statsmodels** та **SciPy**.

Наступною перевагою є документація та підтримка зазначених бібліотек.

Вони усі мають добре задокументовані API та активну спільноту користувачів. Це полегшує ознайомлення з їх використанням та можливість вдосконалення власного коду.

Також слід зазначити й саму популярність мови програмування Python. Вона є найбільш популярною з усіх мов сьогодення, і найбільш розвиненою для обробки даних чисельними методами. Python має чудові вбудовані інструменти для написання коду з ціллю виконувати математичні обчислення.

Зрештою, ключовим для точної обробки наших даних є ефективність. Бібліотеки NumPy та pandas базуються на мові C, що забезпечує їх швидкодію. Крім того, усі бібліотеки є оптимізованими та перевіреними на досвіді.

Наведемо приклади коду з штучними даними, які продемонструють використання обраних нами бібліотек для реалізації моделей. На рисунку 3.1 наведено приклад коду для обробки даних бібліотекою pandas.

```
import pandas as pd

# Завантаження даних з CSV файлу
data = pd.read_csv('data.csv')

# Перегляд перших рядків даних
print(data.head())

# Обробка пропущених значень
data = data.dropna() # Видалення рядків з пропущеними значеннями
data = data.fillna(0) # Заповнення пропущених значень нулями

# Масштабування числових змінних
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
data['числова змінна'] = scaler.fit_transform(data['числова змінна'].values.reshape(-1, 1))

# Розділення на вхідні та вихідні змінні
X = data.drop('вихідна змінна', axis=1)
y = data[['вихідна змінна']]
```

Рисунок 3.1. Обробка даних через pandas

Рисунок 3.2 показує використання функції CubicSpline з бібліотеки SciPy, що знаходить значення поліномів на сегментах. На рисунку 3.3 викликана функція бібліотеки statsmodels, яка дозволяє виконати побудови ARIMA моделі відповідно

до нашого алгоритму.

```
import numpy as np
from scipy.interpolate import CubicSpline

# Вхідні дані
x = np.array([1, 2, 3, 4, 5])
y = np.array([4, 2, 1, 3, 6])

# Побудова кубічного сплайна
cs = CubicSpline(x, y)

# Прогнозування на новому відрізку
x_new = np.array([6, 7, 8, 9, 10])
forecast = cs(x_new)

# Виведення прогнозованих значень
print(forecast)
```

Рисунок 3.2. Побудова кубічного сплайну через SciPy

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

# Завантаження даних
data = pd.read_csv('data.csv')

# Модель ARIMA
model = sm.tsa.ARIMA(data, order=(1, 0, 0))
results = model.fit()
predictions_arima = results.predict(start=..., end=...)
```

Рисунок 3.3. Побудова моделі ARIMA через statsmodels

Ці приклади ілюструють лише деякі можливості використання зазначених бібліотек у реалізації наших моделей. Враховуючи широкий функціонал іа можливості цих бібліотек ми мали можливість ефективно провести аналіз для нашого набору зовнішньоекономічних показників України та Німеччини.

Взагалі, як зазначалося у [1.3 Огляд основних чисельних методів, що використовуються у прогнозуванні] в сучасному світі існує багато засобів для прогнозування даних. Доречним буде провести порівняння обраної мови Python з іншими засобами.

Другим найбільш поширеним методом для аналізу даних є мова R [16]. Як Python так і R мають свої переваги та недоліки. Нижче наведено порівняння двох засобів на основі наших досліджень:

- Масштабність. Python має перевагу, коли йдеться про обробку великих обсягів даних. Крім значущої ефективності зазначених бібліотек, існують інші, наприклад, multiprocessing, що дозволяють використовувати паралельну обробку, що дає змогу значно пришвидшити обчислення.
- Візуалізація. R має величезну кількість пакетів для візуалізації даних, що і робить його другим за популярністю засобом аналізу та прогнозування даних. Однак, Python містить інструменти, що дозволяють досягати високої якості графіків та робити адаптивну, персоналізовану візуалізацію результатів.
- Моделі та алгоритми. Як R, так і Python мають багатий вибір бібліотек для реалізації різноманітних алгоритмів в побудові та аналізі моделей. Вони надають доступ до використання машинного навчання, статистичних методів та багатьох інших інструментів для аналізу даних. Для Python такими бібліотеками будуть: scikit-learn, statsmodels, SciPy. Для R: forecast, timeSeries та prophet. Проте, в Python зручніше створювати власні бібліотеки та алгоритми ніж в R.
- Спільнота користувачів. Python та R мають досить широку та, головне, активну спільноту розробників. Перемагає в їх кількості Python, як найпопулярніша мова програмування в світі на сьогодні. Це дозволяє швидше знаходити рішення, мати більше документації та інших матеріалів, до того ж, оновлення бібліотек та самої мови частіше відбуваються в Python.

Окрім мови R, існує багато інших засобів для аналізу даних. Ось деякі з них:

- MATLAB. MATLAB є популярним інструментом для математичних розрахунків та має значний набір функцій для прогнозування даних. Цей інструмент надає багато статистичних методів та алгоритмів, які дозволяють здійснювати прогнозування. Однак, MATLAB вимагає від

користувачів комерційної ліцензії, що не дає змоги звичайному користувачу використовувати всі можливості цього засобу. Також MATLAB менш часто оновлюється порівняно з Python. Python має велику спільноту користувачів, що сприяє швидкому розробленню нових бібліотек та інструментів.

- Excel. Excel є широко використовуваним інструментом для аналізу даних, включаючи прогнозування. Він має вбудовані функції для прогнозування, наприклад, експоненційне згладжування, та може бути зручним для простих моделей прогнозування. Однак, Excel має свої обмеження, найчастіше у часі, для складнішого аналізу даних та більших моделей. Python дозволяє поєднувати прогнозування з іншими інструментами та бібліотеками для розширеного аналізу даних та візуалізації і обмеження в часі не є проблемою, оскільки можна використовувати багатопоточність.
- Інші мови програмування. Існують також інші засоби, які можуть бути використані для прогнозування даних, такі як Julia, SAS, Scala та інші. Кожна з цих мов має свої переваги та недоліки, і вибір залежить від користувацьких вподобань та вимог проекту. Наприклад, Julia є мовою програмування, спрямованою на високу продуктивність та швидкість обчислень, що може бути корисним для великих обсягів даних або складних моделей прогнозування. SAS є комерційним інструментом, який надає широкий набір функцій для статистичного аналізу та прогнозування. Scala є мовою програмування, яка працює на Java Virtual Machine (JVM) і надає зручну інтеграцію з багатьма бібліотеками для прогнозування та обробки даних. Проте усі ці засоби прогнрають Python майже по всіх пунктах.

Обчислення на папері, тобто вручну, звичайно має обмеження та недоліки, особливо при роботі з великими обсягами даних і складними моделями прогнозування. Можемо навести декілька аргументів, що підтверджують недоречність цього підходу:

- Об'єм даних. У сучасному світі обсяги даних зростають експоненційно,

особливо в контексті експорту та імпорту. Ручні обчислення неспроможні ефективно обробляти великі обсяги даних, і це може призвести до помилок, втрати точності та затримок у роботі.

- **Складність моделей.** Сучасні моделі прогнозування можуть бути дуже складними, з великою кількістю параметрів та залежностей між даними. Ручні обчислення ускладнюють розробку та оцінку таких моделей, а також підвищують ризик помилок при введенні даних та виконанні обчислень.
- **Час та продуктивність.** Ручні обчислення вимагають багато часу та зусиль. Вони можуть бути підзадачою, особливо при необхідності швидкого аналізу даних та прогнозування в реальному часі. Використання програмних засобів, таких як Python, дозволяє автоматизувати процес обчислень, зменшити час і зусилля та підвищити продуктивність.

Розділ 4. Реалізація моделі прогнозування

Перейдемо до реалізації наших моделей ARIMA та методу кубічної сплайн-інтерполяції на мові програмування Python. Ми побачимо як чисельні розв'язки та функції у явному вигляді, так і графічне представлення наших прогнозів та апроксимацій.

4.1 Реалізація моделі ARIMA

Оскільки ми маємо дані за останні 12 місяців експорту та імпорту України і Німеччини в нас вже з'являється чотири задачі для аналізу, тобто чотири набори вхідних даних. Слід також враховувати можливість розгляду декількох значень параметрів моделі p , d , та q , що призводить до виникнення багатьох розгалужень для кожної задачі. Для розуміння реалізації алгоритму побудови моделі, я наведу детально процес побудови для даних експорту України [14]. Для решти показників будемо записувати процес отримання результатів стисло. Програмну реалізацію, тобто сам код на Python можна побачити у [Додаток А. Код програмної реалізації для розділу

4]. У таблиці 4.1 наведено дані для експорту України.

Таблиця 4.1. Експорт України

Україна	(млн.дол. США)
Період	Експорт
	сезонно скориговані дані
Квітень 2022	2685,1
Травень 2022	2849,5
Червень 2022	3361,2
Липень 2022	3157,1
Серпень 2022	3189,2
Вересень 2022	4062,7
Жовтень 2022	3775,0
Листопад 2022	3541,7
Грудень 2022	3361,7
Січень 2023	3273,2
Лютий 2023	3508,0
Березень 2023	3544,9

Далі можемо навести графік, що покаже динаміку зміни наших вхідних даних. Цей графік зображений на рисунку 4.1.

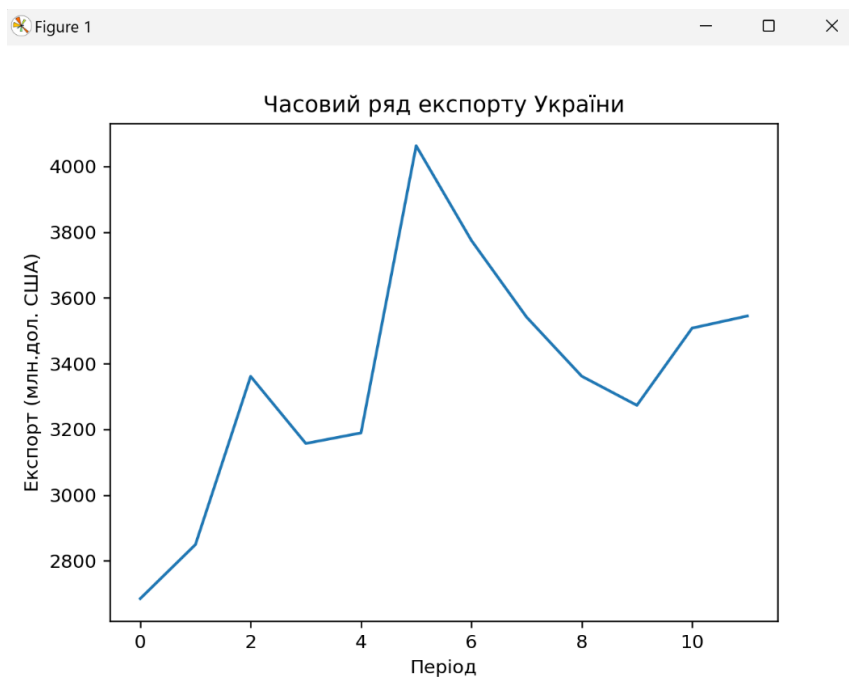


Рис. 4.1.. Графік експорту України за 12 місяців

Наступним кроком переглянемо графік автокореляційної та частково автокореляційної функції – рисунок 4.2.

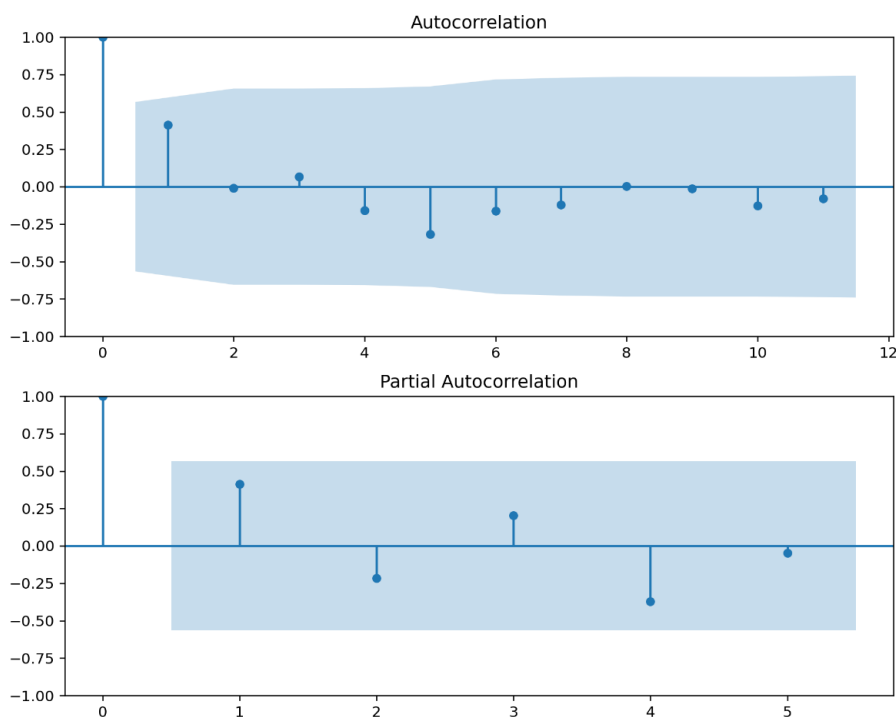


Рис. 4.2.. ACF та PACF для експорту України

З графіків можемо визначити значення p та q . Якщо вважати суттєвими значення більше 0.2 чи 0.3, то слід розглядати випадки $p \in \{0, 1, 4\}$, та $q \in \{0, 1, 5\}$.

Далі слід виконати тест Дікі-Фуллера для нашого початкового набору значень і дізнатися чи є наш ряд стаціонарним. Машинно пораховане p -значення для тесту рівне $4.32952 \cdot 10^{-17}$. Пригадуючи наш алгоритм та нульову гіпотезу теста:

$$H_0: \rho - 1 = 0$$

Оскільки наше p -значення близьке до нуля, то нульову гіпотезу відхиляємо, а отже, наш початковий часовий ряд є стаціонарним. Звідси отримуємо, що $d = 0$ і робити різниці перетворення не потрібно.

Для всіх випадків $p \in \{0, 1, 4\}$, та $q \in \{0, 1, 5\}$ ми можемо виписати побудову моделі, усі коефіцієнти, константи та інше. Проте, через велику кількість таких випадків обмежимося графіком апроксимації та порахованими значеннями АІС та ВІС. Після цього зможемо визначити модель що є найкращою, тобто має найменший АІС або ВІС. Таку модель вже розглянемо більш детально і подивимося на значення, зпрогнозовані нею.

До того ж, розглядати окремо випадок $p = 0, q = 0$ буде зайвим, адже в такому випадку в моделі не буде ні складових авторегресії (AR), ні рухомого середнього (MA), а це означало б, що ми просто не враховуємо минулі значення. Графік в такому випадку був би простою прямою, тому перейдемо одразу до решти ситуацій.

- Модель ARIMA з параметрами ($p = 0, q = 1, d = 0$).

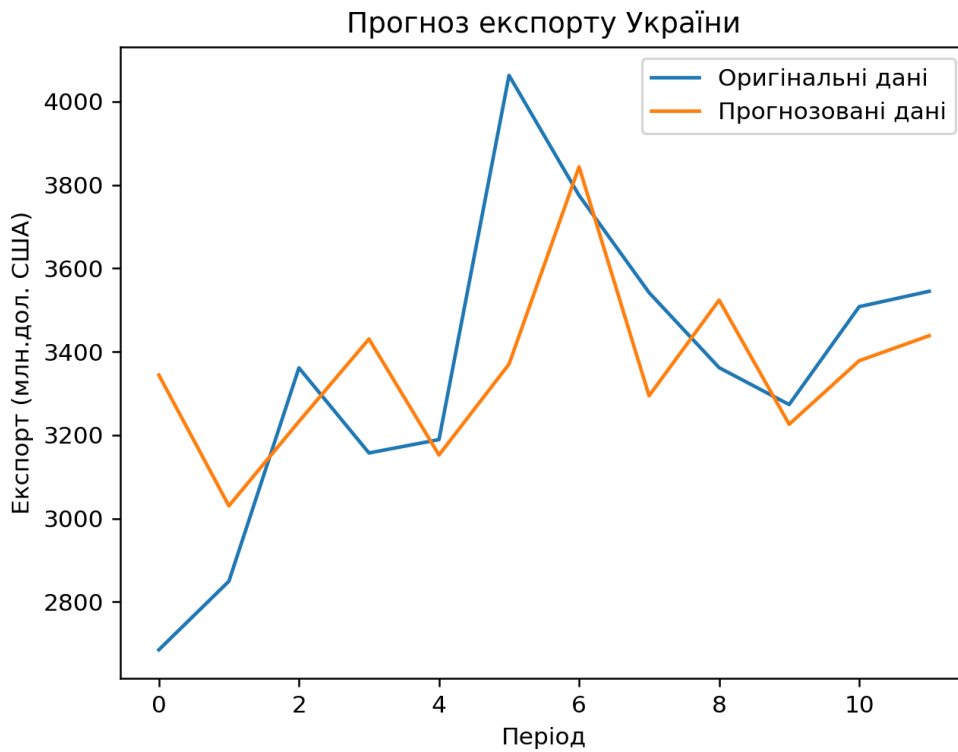


Рис. 4.3. Графік апроксимації для ARIMA(0, 1, 0)

В цьому та наступному випадку відсутня авторегресія (AR). Нагадаємо формули за якими обчислюємо AIC та BIC:

$$AIC = 2k - 2 \ln(RSS)$$

$$BIC = k \cdot \ln(n) - 2 \ln(RSS)$$

Для нашого випадку:

$$AIC = 176.732$$

$$BIC = 178.187$$

- Модель ARIMA з параметрами ($p = 0, q = 5, d = 0$).

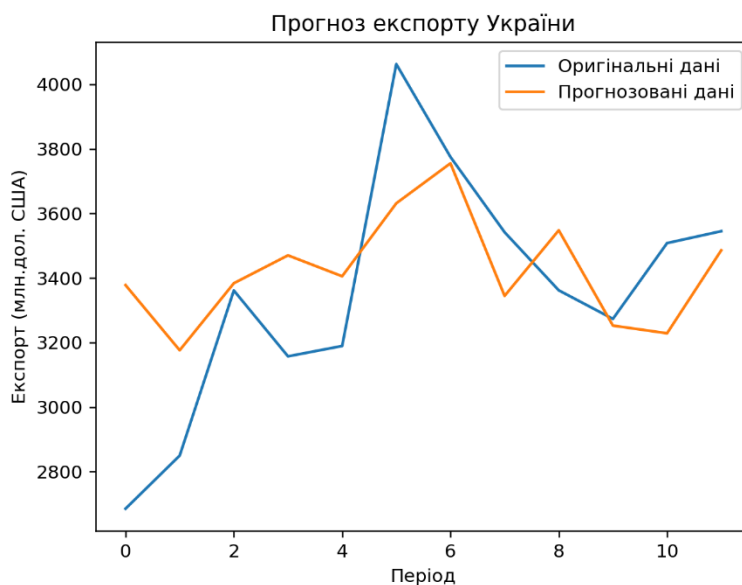


Рис. 4.4. Графік апроксимації для ARIMA(0, 5, 0)

Для нашого випадку:

$$AIC = 183.358$$

$$BIC = 186.752$$

- Модель ARIMA з параметрами ($p = 1, q = 0, d = 0$).

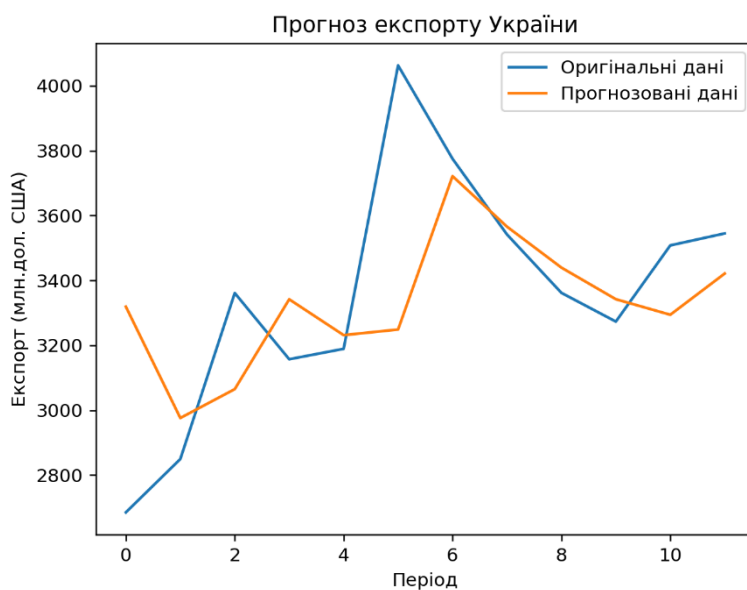


Рис. 4.5. Графік апроксимації для ARIMA(1, 0, 0)

Для нашого випадку:

$$AIC = 178.165$$

$$BIC = 179.620$$

- Модель ARIMA з параметрами ($p = 1, q = 1, d = 0$).

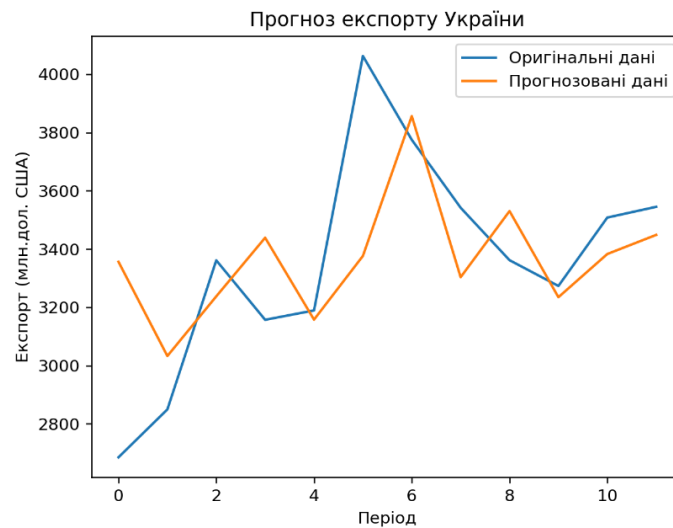


Рис. 4.6. Графік апроксимації для ARIMA(1, 1, 0)

Для нашого випадку:

$$AIC = 178.739$$

$$BIC = 180.678$$

- Модель ARIMA з параметрами ($p = 1, q = 5, d = 0$).

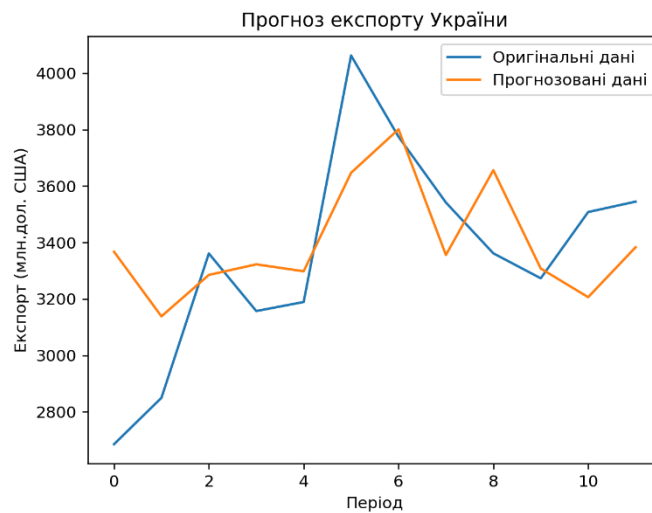


Рис. 4.7. Графік апроксимації для ARIMA(1, 5, 0)

Для нашого випадку:

$$AIC = 184.591$$

$$BIC = 188.471$$

- Модель ARIMA з параметрами ($p = 4, q = 0, d = 0$).

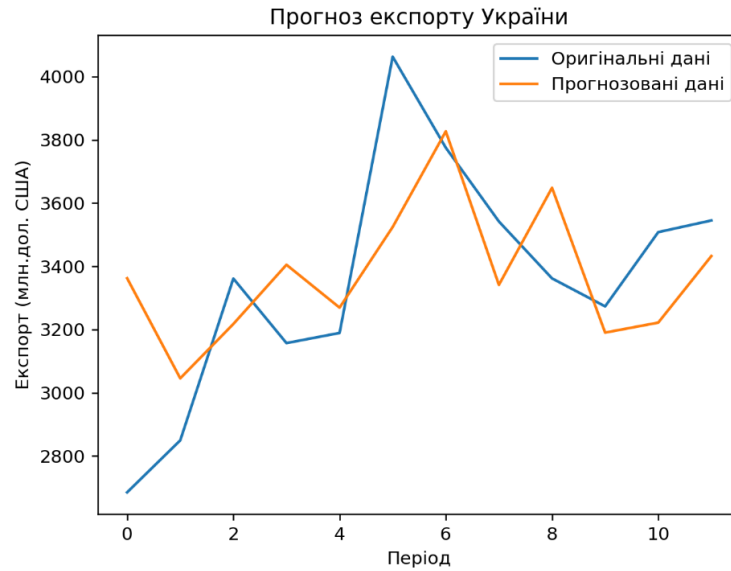


Рис. 4.8. Графік апроксимації для ARIMA(4, 0, 0)

Для нашого випадку:

$$AIC = 181.904$$

$$BIC = 184.814$$

- Модель ARIMA з параметрами ($p = 4, q = 1, d = 0$).

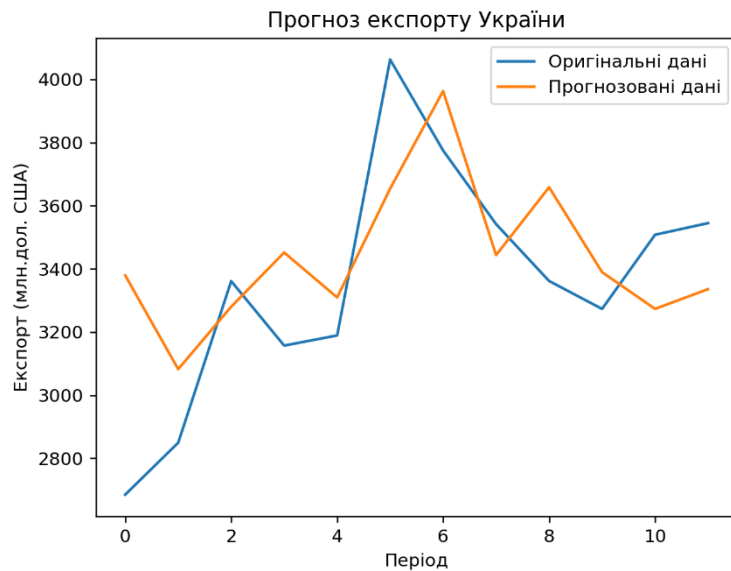


Рис. 4.9. Графік апроксимації для ARIMA(4, 1, 0)

Для нашого випадку:

$$AIC = 182.875$$

$$BIC = 186.269$$

- Модель ARIMA з параметрами ($p = 4, q = 5, d = 0$).

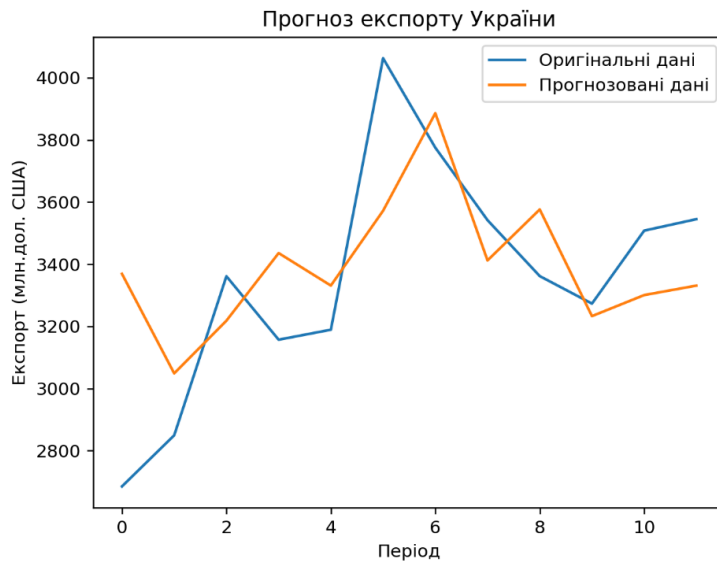


Рис. 4.10. Графік апроксимації для ARIMA(4, 5, 0)

Для нашого випадку:

$$AIC = 190.783$$

$$BIC = 196.117$$

Звідси маємо, що найкращою буде модель ARIMA($p = 0, q = 1, d = 0$), тобто модель без авторегресії (AR).

Для цієї моделі коефіцієнти авторегресії φ_i відсутні, єдиний коефіцієнт рухомго середнього $\theta_1 = 0.72914484$, константа $c = 3343.8821041600654$.

Маємо

$$Y(t) = c + \theta_1 \cdot e(t - 1) = 3343.8821041600654 + 0.72914484 \cdot e(t - 1)$$

де для залишкових членів $e(t)$ нашої моделі будемо мати значення, що наведені в таблиці 4.2.

Таблиця 4.2. Залишкові члени для $ARIMA(p=0, q=1, d=0)$

Місяць, t	e(t)
Квітень 2022, 1	-658.782104
Травень 2022, 2	-180.768105
Червень 2022, 3	128.589711
Липень 2022, 4	-273.370745
Серпень 2022, 5	36.855527
Вересень 2022, 6	692.491822
Жовтень 2022, 7	-68.403784
Листопад 2022, 8	247.411910
Грудень 2022, 9	-162.040093
Січень 2023, 10	47.280471
Лютий 2023, 11	129.672742
Березень 2023, 12	106.510182

Зробимо прогноз на три місяці уперед та покажемо майбутні зміни графічно, можемо бачити його на рисунку 4.11.



Рис. 4.11. Прогнозовані значення для експорту України

Отже, отримали для квітня 2023 значення 3421.524900, травень 2023 - 3343.882104 і червень 2023 - 3343.882004.

Таким чином ми визначили яка з моделей $ARIMA(p, d, q)$ краще всього підходить для прогнозування експорту України на три місяці вперед на базі попередніх 12-ти місяців.

Процес побудови моделей та вибору кращої з них для інших показників буде таким самим. Тому для імпорту України та імпорту експорту Німеччини одразу покажемо найкращу модель та отримані нею результати прогнозування.

Покажемо імпорт України за 12 місяців зображений на рисунку 4.12.

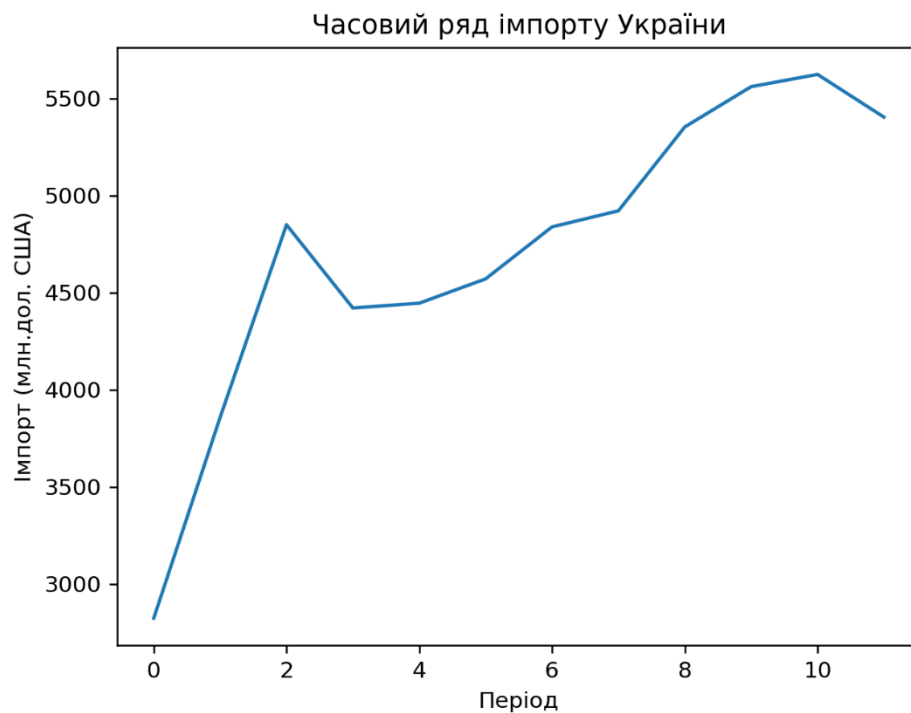


Рис. 4.12. Імпорт України

Для такого часового ряду будемо мати автокореляційну та частково автокореляційну функції у вигляді, щл зображений на рисунку 4.13.

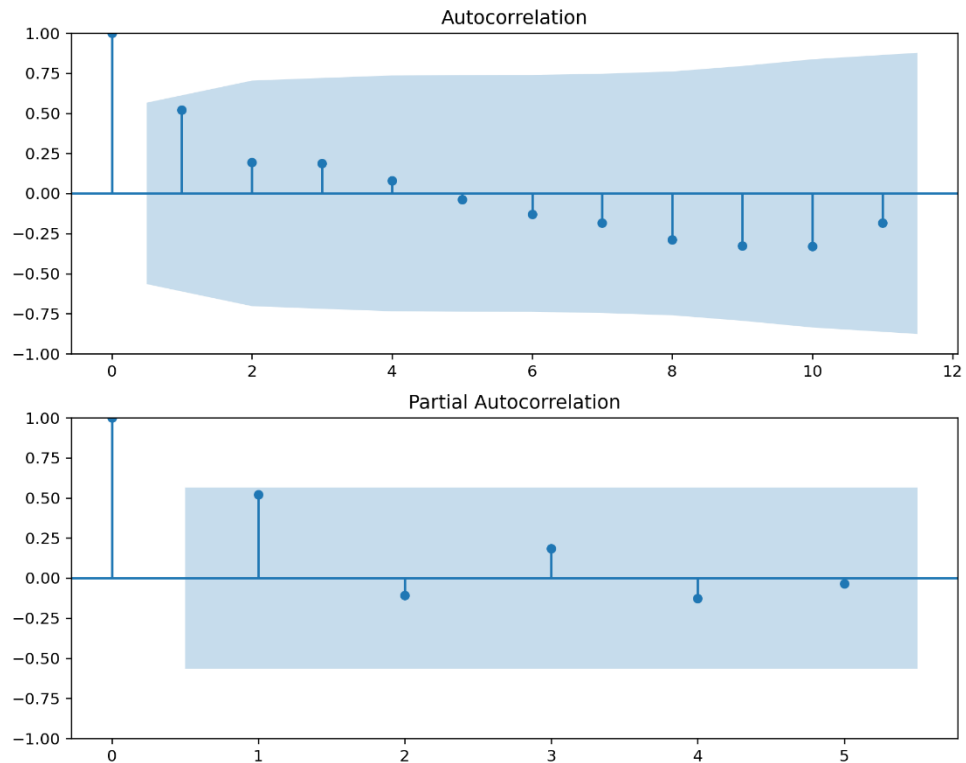


Рис. 4.13. ACF та PACF для імпорту України

Звідси можемо бачити такі можливі випадки $p \in \{0, 1\}$, $q \in \{0, 1, 9, 10\}$.

p -значення, що показав тест Дікі-Фуллера рівне 0.0098281494475299, що знову відхиляє нашу нульову гіпотезу. Отже, початковий часовий ряд є стаціонарним і можемо переходити до побудови моделей з $d = 0$. Наведена таблиця 4.3 з порахованими значеннями AIC та BIC для усіх моделей.

Таблиця 4.3. Моделі та їх значення AIC, BIC

Модель	AIC	BIC
ARIMA($p = 0, d = 0, q = 1$)	191.091	192.545
ARIMA(0, 0, 9)	200.582	205.916
ARIMA(0, 0, 10)	202.325	208.144
ARIMA(1, 0, 0)	189.812	191.267
ARIMA(1, 0, 1)	189.788	191.727
ARIMA(1, 0, 9)	203.850	209.669
ARIMA(1, 0, 10)	204.195	210.498

Звідси за найкращу модель обираємо ARIMA($p = 1, d = 0, q = 1$). Відповідні значення коефіцієнтів та інших параметрів:

$$\varphi_1 = 0.77481998$$

$$\theta_1 = 0.50857023$$

$$c = 4721.462221640915$$

Залишкові члени показані в таблиці 4.4.

Таблиця 4.4. Залишкові члени для $ARIMA(p=1, q=1, d=0)$

Місяць, t	e(t)
Квітень 2022, 1	-1897.962222
Травень 2022, 2	795.494125
Червень 2022, 3	462.677374
Липень 2022, 4	-625.144569
Серпень 2022, 5	271.724783
Вересень 2022, 6	-74.630768
Жовтень 2022, 7	271.791326
Листопад 2022, 8	-29.054334
Грудень 2022, 9	491.842479
Січень 2023, 10	99.611502
Лютий 2023, 11	201.487855
Березень 2023, 12	-119.114645

$$Y(t) = c + \varphi_1 Y(t - 1) + \theta_1 e(t - 1)$$

$$Y(t) = 4721.462221640915 + 0.77481998 \cdot Y(t - 1) + 0.50857023 \cdot e(t - 1)$$

Наближення, що ми отримали зображене на рисунку 4.14. Тут наведені апроксимації для наших вхідних даних, без майбутніх періодів.

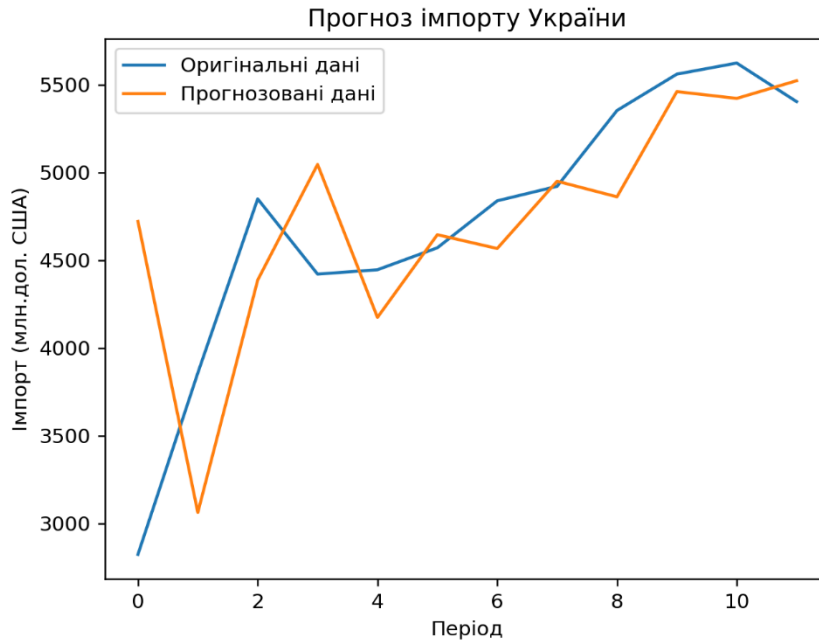


Рис. 4.14. Апроксимовані значення для імпорту України

Відповідний прогноз на три місяці зображений на рисунку 4.15.

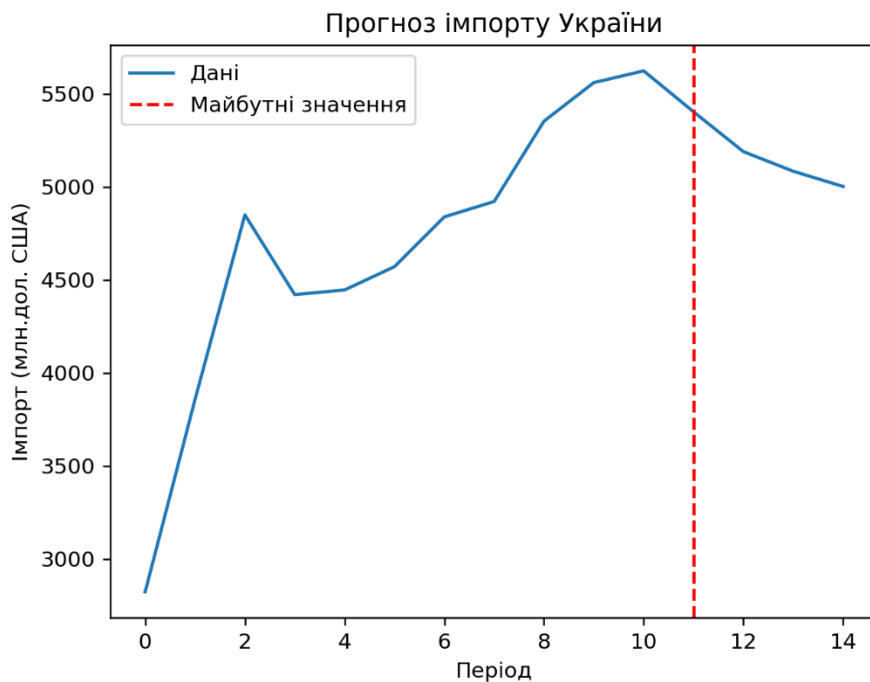


Рис. 4.15. Майбутні прогнозовані значення для імпорту України

Отже, отримали для квітня 2023 значення 5189.882943, травень 2023 - 5084.403955 і червень 2023 - 5002.676728.

Тепер виконаємо аналогічні дії для імпорту та експорту з Німеччини. В пункті [2.2 Підготовка даних до аналізу] зображені вхідні дані обох показників для Німеччини.

Ряд за тестом Дікі-Фуллера має p -значення рівне 0.12255629985018096, тобто

нульову гіпотезу відхилити не можемо. Після застосування методу різниць один раз маємо p -значення рівне 0.0010958338014145867, а значить новий ряд буде стаціонарним при $d = 1$.

Маємо наступні автокореляційну та частково автокореляційну функції для експорту на рисунку 4.16.

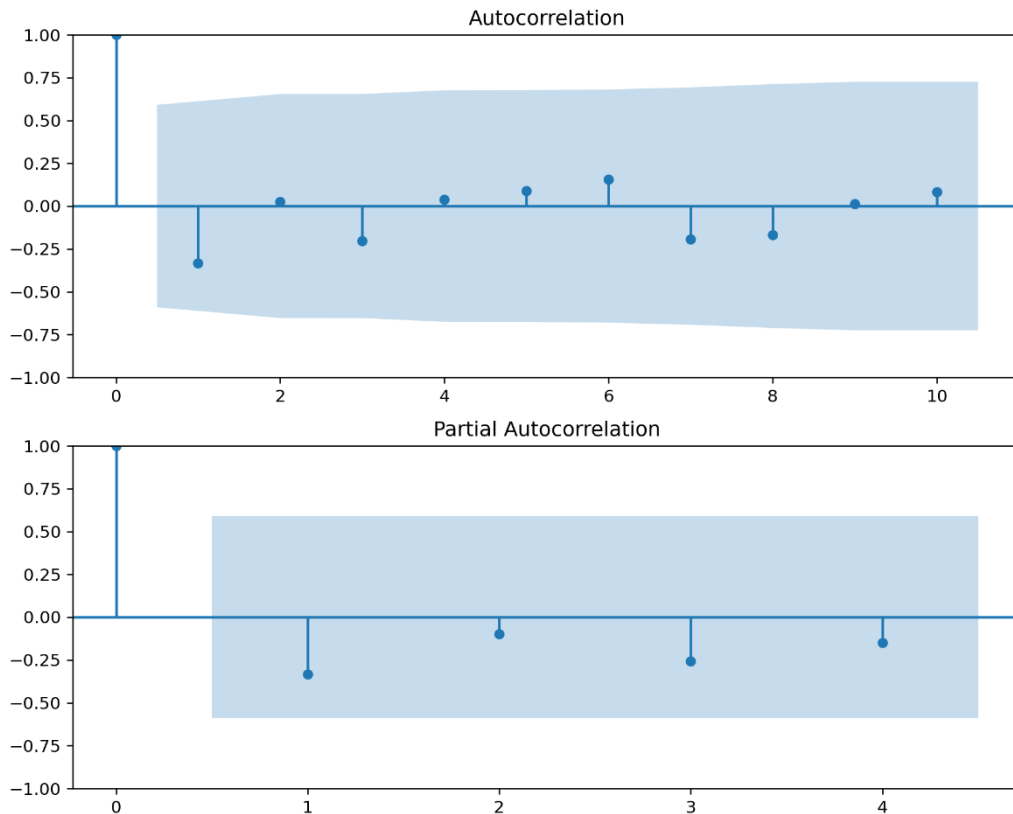


Рис. 4.16. ACF та PACF для експорту Німеччини

Звідси візьмемо $p \in \{0, 1, 3\}$, а $q \in \{0, 1\}$. Отримали набір моделей та значення AIC та BIC для них у таблиці 4.5.

Таблиця 4.5. Моделі та їх AIC, BIC

Модель	AIC	BIC
ARIMA($p = 0, d = 1, q = 1$)	237.740	238.536
ARIMA(1, 1, 0)	237.552	238.348
ARIMA(1, 1, 1)	239.752	240.946
ARIMA(3, 1, 0)	239.073	240.665
ARIMA(3, 1, 1)	240.932	242.922

Обираємо модель ARIMA($p = 1, d = 1, q = 0$). Для неї будемо мати такі параметри:

$$\varphi_1 = 0.0462906$$

Коефіцієнти для рухомого середнього (МА) θ_i – відсутні

$$c = 0.04629060379835924$$

Залишкові члени показані в таблиці 4.6.

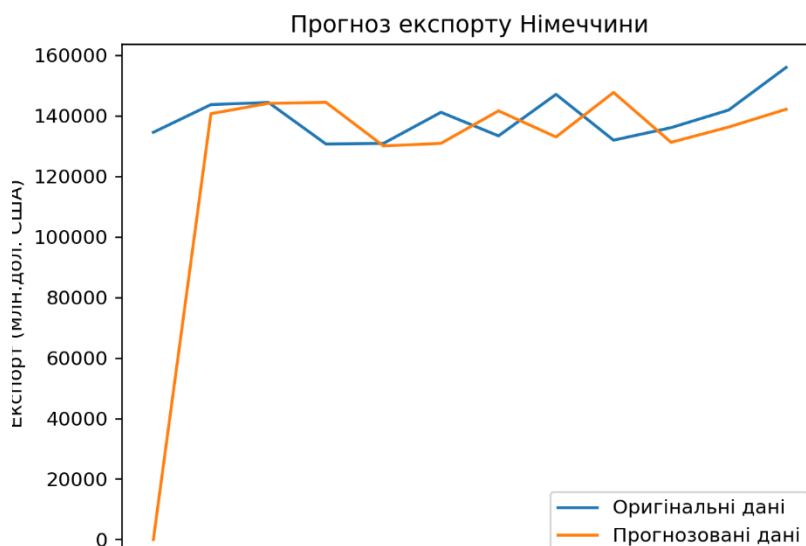
Таблиця 4.6. Залишкові члени для $ARIMA(p=1, q=0, d=1)$

Місяць, t	e(t)
Квітень 2022, 1	134541.900000
Травень 2022, 2	2988.141839
Червень 2022, 3	295.598363
Липень 2022, 4	-13766.082944
Серпень 2022, 5	840.399604
Вересень 2022, 6	10268.624313
Жовтень 2022, 7	-8271.879455
Листопад 2022, 8	14079.386176
Грудень 2022, 9	-15779.237648
Січень 2023, 10	4842.634162
Лютий 2023, 11	5620.382835
Березень 2023, 12	13823.554382

$$Y(t) = c + \varphi_1 Y(t - 1)$$

$$Y(t) = 0.04629060379835924 + 0.0462906 \cdot Y(t - 1)$$

На наступному рисунку 4.17 можемо бачити апроксимацію:



Та відповідні прогнозовані майбутні значення на рисунку 4.18.

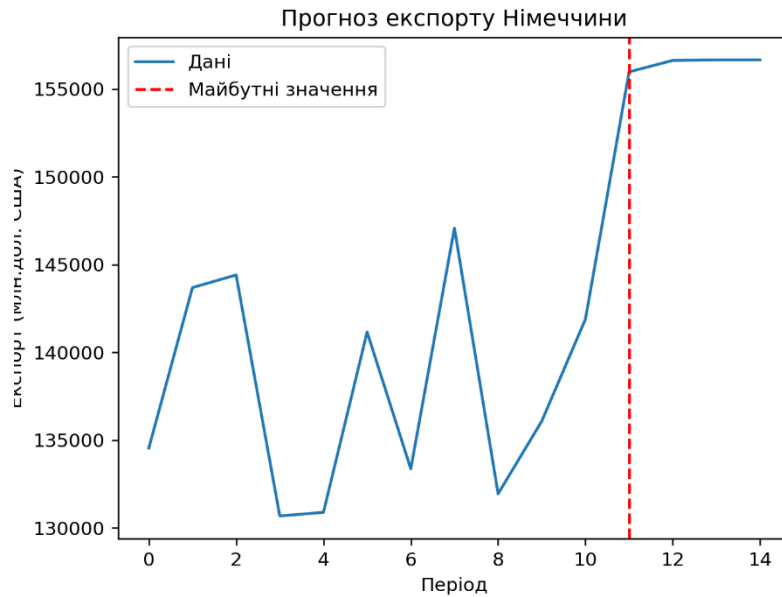


Рис. 4.18. Прогноз майбутніх значень експорту Німеччини

Отже, отримали для квітня 2023 значення 156634.354963, травень 2023 - 156664.552868 і червень 2023 - 156665.950747.

Перейдемо до останнього показника імпорт Німеччинни. Тест Дікі-Фуллера показав, що p -значення = 0.08836, тобто нульову гіпотезу відхилити не можемо. Виконуємо диференціювання та отримуємо нове p -значення = $2.0055 \cdot 10^{-5}$, тобто новий ряд буде стаціонарним і $d = 1$. Графіки ACF та PACF на рисунку 4.19.

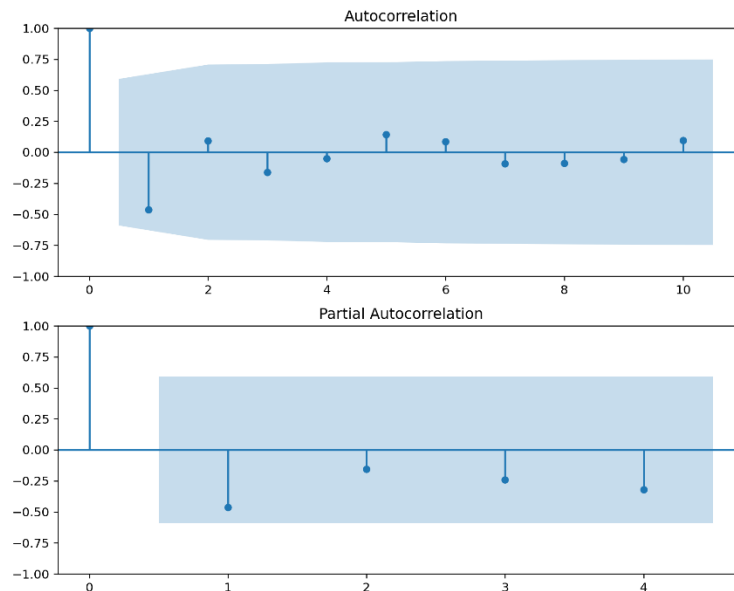


Рис. 4.19. ACF та PACF для імпорту Німеччини

Звідси беремо наші параметри $p \in \{0, 1, 4\}$, $q \in \{0, 1\}$. Для них маємо набір моделей $ARIMA(p, d, q)$ наведений в наступній таблиці 4.7.

Таблиця 4.5. Моделі та їх AIC, BIC

Модель	AIC	BIC
$ARIMA(p = 0, d = 1, q = 1)$	230.584	231.380
$ARIMA(1, 1, 0)$	230.305	231.101
$ARIMA(1, 1, 1)$	233.077	234.270
$ARIMA(4, 1, 0)$	233.088	235.077
$ARIMA(4, 1, 1)$	234.655	237.042

Обираємо модель $ARIMA(p = 1, d = 1, q = 0)$. Для неї будемо мати такі параметри:

$$\varphi_1 = 0.0342849$$

Коефіцієнти для рухомого середнього (МА) θ_i – відсутні

$$c = 0.03428489841263385$$

Залишкові члени в таблиці 4.8.

Таблиця 4.8. Залишкові члени для $ARIMA(p=1, q=0, d=1)$

Місяць, t	$e(t)$
Квітень 2022, 1	131985.800000
Травень 2022, 2	2128.525684
Червень 2022, 3	-2765.613524
Липень 2022, 4	-10178.164931
Серпень 2022, 5	5070.144768
Вересень 2022, 6	2739.236992
Жовтень 2022, 7	-4953.560490
Листопад 2022, 8	6729.222325
Грудень 2022, 9	-14730.404931
Січень 2023, 10	5288.616165
Лютий 2023, 11	-1792.969234
Березень 2023, 12	8247.139814

$$Y(t) = c + \varphi_1 Y(t - 1)$$

$$Y(t) = 0.03428489841263385 + 0.0342849 \cdot Y(t - 1)$$

На рисунку 4.20 можемо бачити апроксимацію:

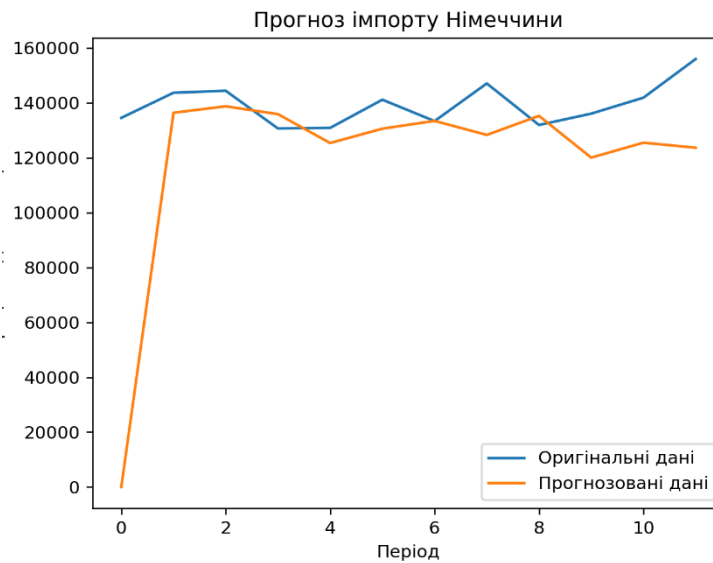


Рис. 4.20. Апроксимовані значення для імпорту Німеччини

Та відповідні прогнозовані майбутні значення на рис. 4.21:

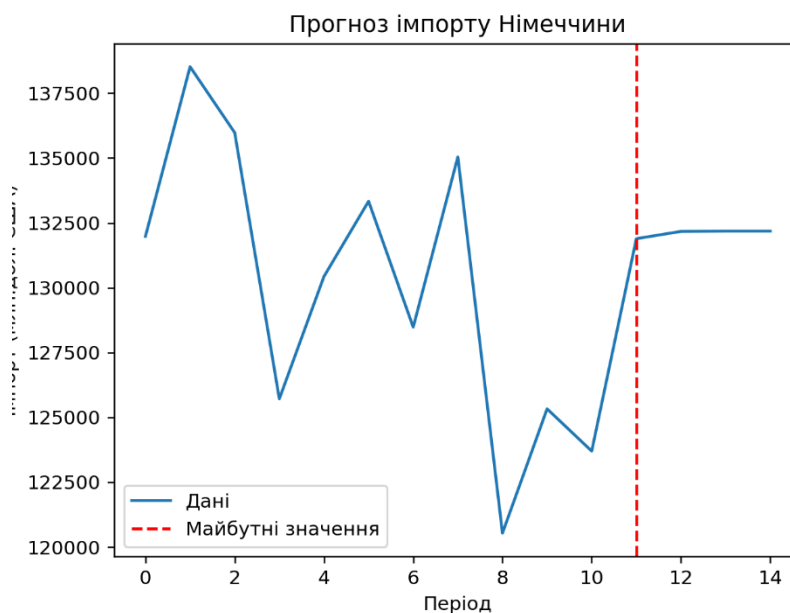


Рис. 4.21. Прогнозовані майбутні значення для імпорту Німеччини

Отже, отримали для квітня 2023 значення 132173.037888, травень 2023 - 132182.666387 і червень 2023 - 132182.996499.

Отже, прогноз для зовнішньоекономічних показників України та Німеччини методом ARIMA виконано.

4.2 Реалізація методу кубічної сплайн-інтреполяції

Реалізація методу кубічної сплайн-інтреполяції у Python відбувається простіше. Розглянемо отримані результати для наших показників.

Почнемо з експорту України. Місяці пронумеровані від 1 до 12. Будемо мати наступний вигляд кубічних поліномів для наших сегментів у таблиці 4.9.

Таблиця 4.9. Кубічні поліноми для експорту України

Проміжок	Кубічний поліном
1	$-259.921427 + -259.921427 \cdot (t - 1) + 236.507136 \cdot (t - 1)^2 + 265.892883 \cdot (t - 1)^3$
2	$-259.921427 + 236.507136 \cdot (t - 2) + 265.892883 \cdot (t - 2)^2 + -694.878667 \cdot (t - 2)^3$
3	$236.507136 + 265.892883 \cdot (t - 3) + -694.878667 \cdot (t - 3)^2 + 511.021786 \cdot (t - 3)^3$
4	$265.892883 + 694.878667 \cdot (t - 4) + 511.021786 \cdot (t - 4)^2 + -133.608479 \cdot (t - 4)^3$
5	$694.878667 + 511.021786 \cdot (t - 5) + -133.608479 \cdot (t - 5)^2 + 22.312128 \cdot (t - 5)^3$
6	$511.021786 + -133.608479 \cdot (t - 6) + 22.312128 \cdot (t - 6)^2 + 82.559966 \cdot (t - 6)^3$
7	$-133.608479 + 22.312128 \cdot (t - 7) + 82.559966 \cdot (t - 7)^2 + -120.751993 \cdot (t - 7)^3$
8	$22.312128 + 82.559966 \cdot (t - 8) + -120.751993 \cdot (t - 8)^2 + -120.751993 \cdot (t - 8)^3$
9	$-12027.256298 + -120.751993 \cdot (t - 9) + -120.751993 \cdot (t - 9)^2 + 953.414282 \cdot (t - 9)^3$
10	$-120.751993 + -120.751993 \cdot (t - 10) + 953.414282 \cdot (t - 10)^2 + 173.65 \cdot (t - 10)^3$
11	$-120.751993 + 953.414282 \cdot (t - 11) + 173.65 \cdot (t - 11)^2 + -606.114282 \cdot (t - 11)^3$

Можемо зобразити (рис. 4.22) як ці полімноми наближають наші вхідні дані:

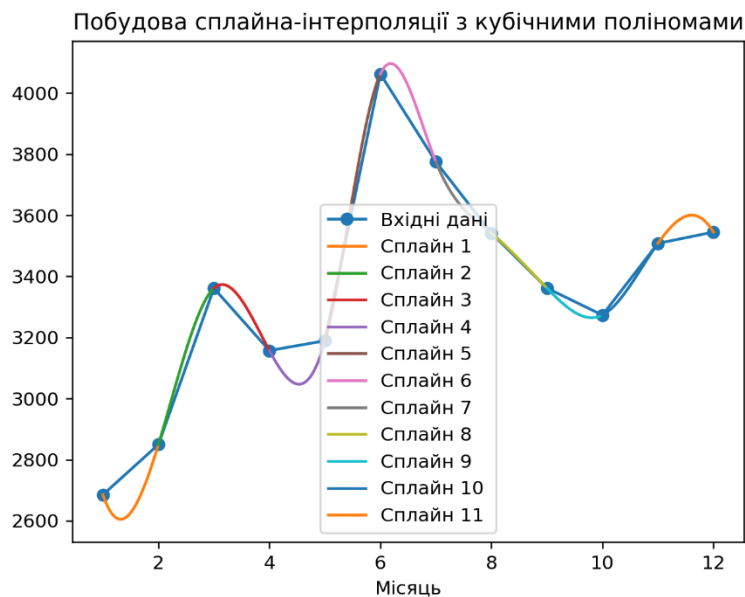


Рис. 4.22. Апроксимація сплайнами на сегментах експорту України

Прогнози що ми отримали для квітня 2023 - 183824.20755703552, травень 2023 - 313417.63022814196 і червень 2023 - 554593.1755703548. Зразу видно, що значення на другий та третій місяць в майбутньому не виглядають реалістично.

Маємо графік для прогнозу на один місяць рис. 4.23.

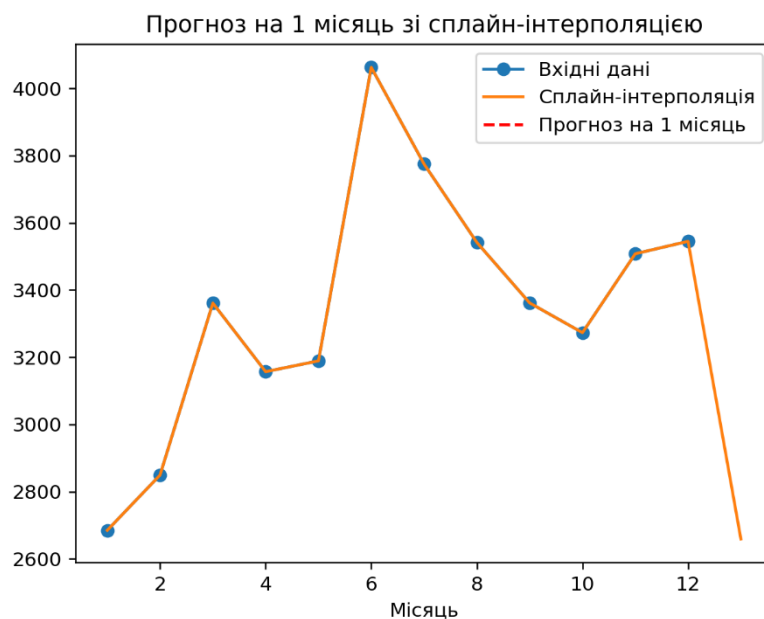


Рис. 4.23. Прогнозовані сплайном майбутні значення для експорту України на 1 місяць

Та графік на три наступних місяці на рисунку 4.24.

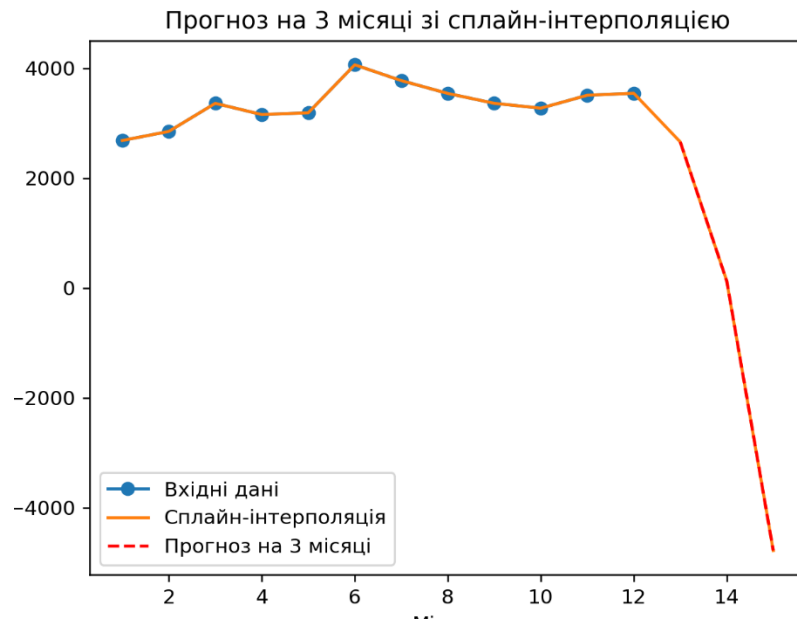


Рис. 4.24. Прогнозовані сплайном майбутні значення для експорту України на 3 місяці

Розглянемо тепер імпорт України. Поліноми на відповідні проміжки мають наступний вигляд таблиці 4.10.

Таблиця 4.10. Кубічні поліноми для імпорту України

Проміжок	Кубічний поліном
1	$-403.525734 + -403.525734 \cdot (t - 1) + 639.228672 \cdot (t - 1)^2 + -278.888953 \cdot (t - 1)^3$
2	$-403.525734 + 639.228672 \cdot (t - 2) + -278.888953 \cdot (t - 2)^2 + 122.927140 \cdot (t - 2)^3$
3	$639.228672 + -278.888953 \cdot (t - 3) + 122.927140 \cdot (t - 3)^2 + -170.619608 \cdot (t - 3)^3$
4	$-278.888953 + 122.927140 \cdot (t - 4) + -170.619608 \cdot (t - 4)^2 + 232.051291 \cdot (t - 4)^3$
5	$122.927140 + -170.619608 \cdot (t - 5) + 232.051291 \cdot (t - 5)^2 + -223.285557 \cdot (t - 5)^3$
6	$-170.619608 + 232.051291 \cdot (t - 6) + -223.285557 \cdot (t - 6)^2 + 87.190936 \cdot (t - 6)^3$

7	$232.051291 + -223.285557 \cdot (t - 7) + 87.190936 \cdot (t - 7)^2 + -45.178187 \cdot (t - 7)^3$
8	$-223.285557 + 87.190936 \cdot (t - 8) + -45.178187 \cdot (t - 8)^2 + -45.178187 \cdot (t - 8)^3$
9	$87.190936 + -45.178187 \cdot (t - 9) + -45.178187 \cdot (t - 9)^2 + 1189.377203 \cdot (t - 9)^3$
10	$-45.178187 + -45.178187 \cdot (t - 10) + 1189.377203 \cdot (t - 10)^2 + -21.2 \cdot (t - 10)^3$
11	$-45.178187 + 1189.377203 \cdot (t - 11) + -21.2 \cdot (t - 11)^2 + -1231.777203 \cdot (t - 11)^3$

Можемо зобразити (рис. 4.25) як ці поліноми наближають наші вхідні дані:

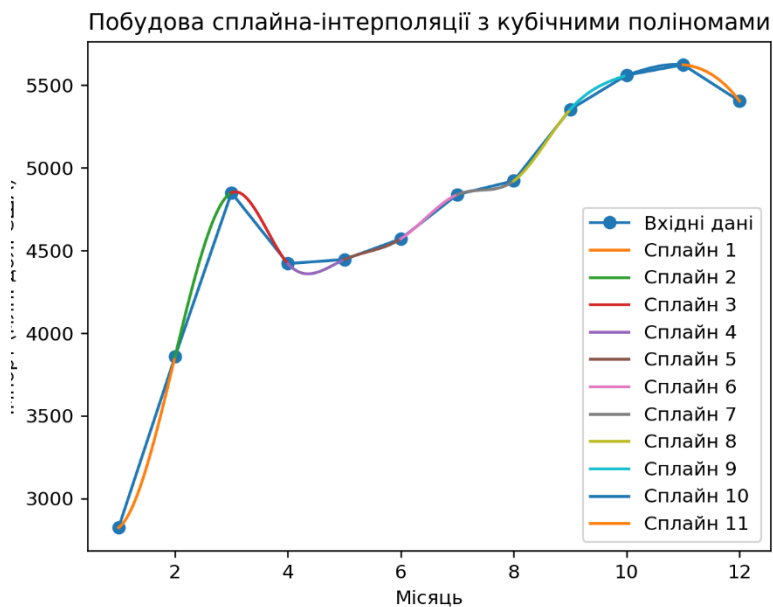


Рис. 4.25. Апроксимація сплайнами на сегментах імпорту України

Прогнози що ми отримали для квітня 2023 - 183824.20755703552, травень 2023 - 313417.63022814196 і червень 2023 - 554593.1755703548. Зразу видно, що значення на другий та третій місяць в майбутньому не виглядають реалістично.

Маємо графік для прогнозу на один місяць рис. 4.26.

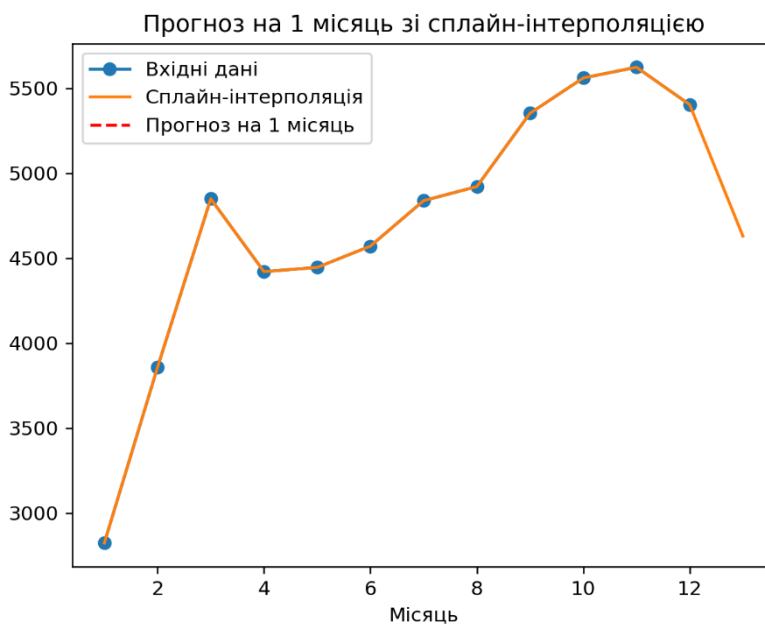


Рис. 4.26. Прогнозовані сплайном майбутні значення для імпорту України на 1 місяць

Та графік на три наступних місяці рис. 4.27.

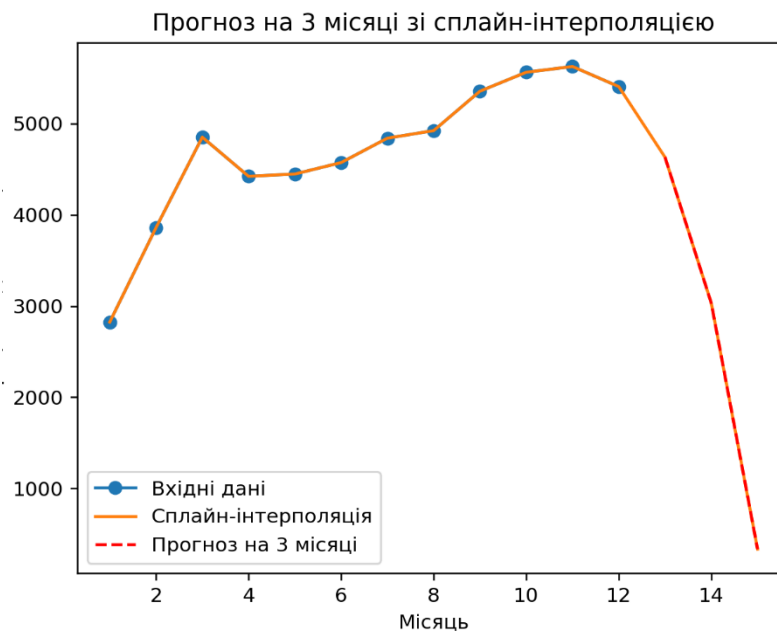


Рис. 4.27. Прогнозовані сплайном майбутні значення для експорту України на 3 місяці

Далі експорт Німеччини. Поліноми на відповідні проміжки мають наступний вигляд табл. 4.11.

Таблиця 4.11. Кубічні поліноми для експорту Німеччини

Проміжок	Кубічний поліном
1	$-2762.805910 + -2762.805910 \cdot (t - 1) + 7789.829549 \cdot (t - 1)^2 + -7.212288 \cdot (t - 1)^3$
2	$-2762.805910 + 7789.829549 \cdot (t - 2) + -7.212288 \cdot (t - 2)^2 + -11625.080398 \cdot (t - 2)^3$
3	$7789.829549 + -7.212288 \cdot (t - 3) + -11625.080398 \cdot (t - 3)^2 + 18359.933878 \cdot (t - 3)^3$
4	$-7.212288 + -11625.080398 \cdot (t - 4) + 18359.933878 \cdot (t - 4)^2 + -22225.855114 \cdot (t - 4)^3$
5	$-11625.080398 + 18359.933878 \cdot (t - 5) + -22225.855114 \cdot (t - 5)^2 + 20166.186580 \cdot (t - 5)^3$
6	$18359.933878 + -22225.855114 \cdot (t - 6) + 20166.186580 \cdot (t - 6)^2 + -10290.391205 \cdot (t - 6)^3$
7	$-22225.855114 + 20166.186580 \cdot (t - 7) + 10290.391205 \cdot (t - 7)^2 + 3380.078241 \cdot (t - 7)^3$
8	$20166.186580 + 10290.391205 \cdot (t - 8) + 3380.078241 \cdot (t - 8)^2 + 3380.078241 \cdot (t - 8)^3$
9	$10290.391205 + 3380.078241 \cdot (t - 9) + 3380.078241 \cdot (t - 9)^2 + 4074.617730 \cdot (t - 9)^3$
10	$3380.078241 + 3380.078241 \cdot (t - 10) + 4074.617730 \cdot (t - 10)^2 + -4213.8 \cdot (t - 10)^3$
11	$3380.078241 + 4074.617730 \cdot (t - 11) + -4213.8 \cdot (t - 11)^2 + -12502.217730 \cdot (t - 11)^3$

Можемо зобразити (рис. 4.28) як ці полімноми наближають наші вхідні дані:

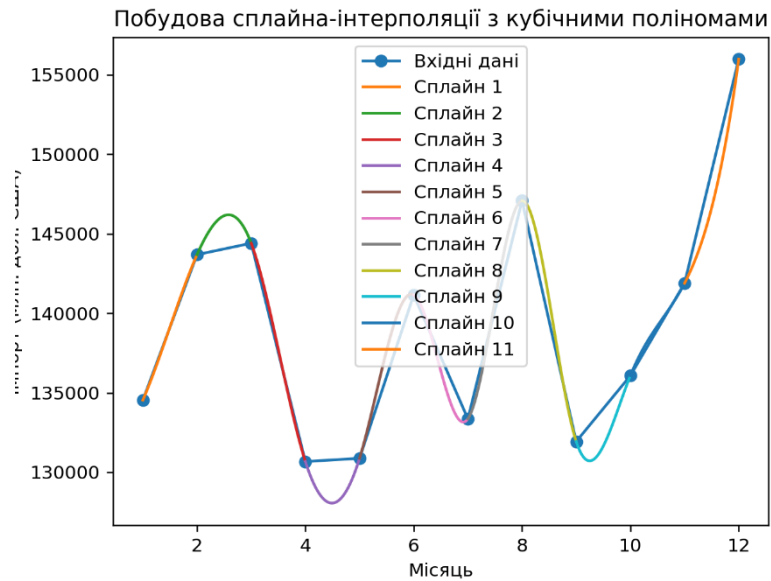


Рис. 4.28. Апроксимація сплайнами на сегментах експорту Німеччини

Прогнози що ми отримали для квітня 2023 - 4630.230876803548, травень 2023 - 3031.1235072141944 і червень 2023 - 335.8087680354888. Одразу видно, що значення на другий та третій місяць в майбутньому не виглядають реалістично.

Маємо графік для прогнозу на один місяць рис. 4.29.

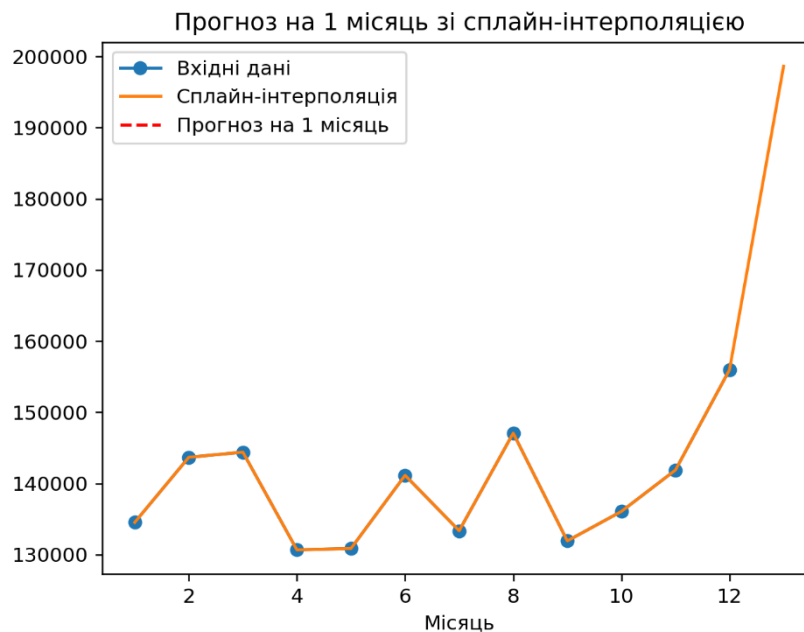


Рис. 4.29. Прогнозовані сплайном майбутні значення для експорту Німеччини на 1 місяць

Та графік на три наступних місяці рис. 4.30.

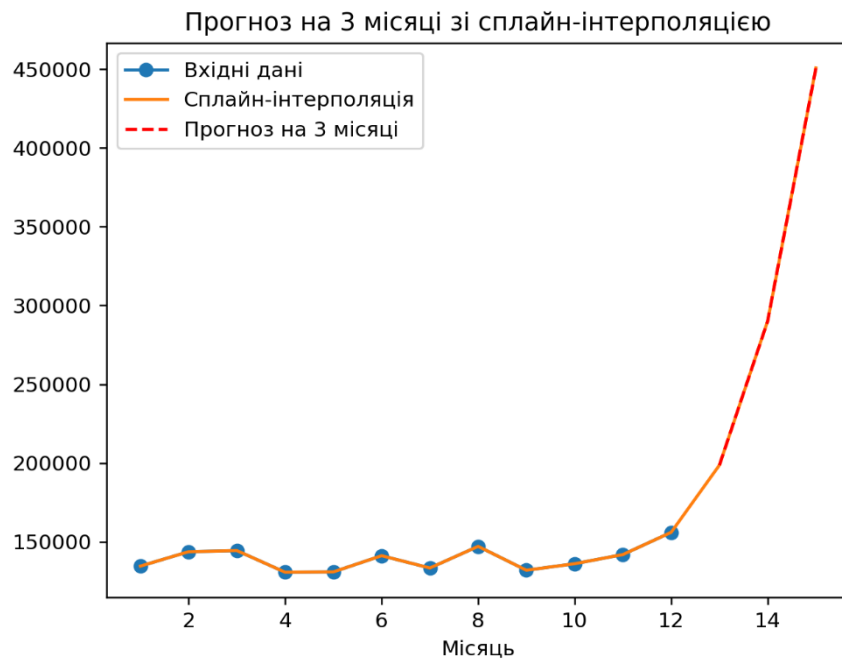


Рис. 4.30. Прогнозовані сплайном майбутні значення для експорту Німеччини на 3 місяці

Останній показник - імпорт Німеччини. Будемо мати наступний вигляд кубічних поліномів для наших сегментів у таблиці 4.12:

Таблиця 4.12. Кубічні поліноми для імпорту Німеччини

Проміжок	Кубічний поліном
1	$-1196.200061 + -1196.200061 \cdot (t - 1) + 7335.500306 \cdot (t - 1)^2 + -5438.501164 \cdot (t - 1)^3$
2	$-1196.200061 + 7335.500306 \cdot (t - 2) + -5438.501164 \cdot (t - 2)^2 + -2382.195651 \cdot (t - 2)^3$
3	$7335.500306 + -5438.501164 \cdot (t - 3) + -2382.195651 \cdot (t - 3)^2 + 9029.383769 \cdot (t - 3)^3$
4	$-5438.501164 + -2382.195651 \cdot (t - 4) + 9029.383769 \cdot (t - 4)^2 + -14563.339425 \cdot (t - 4)^3$
5	$-2382.195651 + 9029.383769 \cdot (t - 5) + -14563.339425 \cdot (t - 5)^2 + -16738.873931 \cdot (t - 5)^3$
6	$9029.383769 + -14563.339425 \cdot (t - 6) + -16738.873931 \cdot (t - 6)^2 + -12027.256298 \cdot (t - 6)^3$

7	$-14563.339425 + 16738.873931 \cdot (t - 7) + -12027.256298 \cdot (t - 7)^2 + 5653.451260 \cdot (t - 7)^3$
8	$16738.873931 + -12027.256298 \cdot (t - 8) + 5653.451260 \cdot (t - 8)^2 + 5653.451260 \cdot (t - 8)^3$
9	$-12027.256298 + 5653.451260 \cdot (t - 9) + 5653.451260 \cdot (t - 9)^2 + -950.549816 \cdot (t - 9)^3$
10	$5653.451260 + 5653.451260 \cdot (t - 10) + -950.549816 \cdot (t - 10)^2 + -4539.15 \cdot (t - 10)^3$
11	$5653.451260 + -950.549816 \cdot (t - 11) + -4539.15 \cdot (t - 11)^2 + -8127.750184 \cdot (t - 11)^3$

Можемо зобразити (рис. 4.31) як ці поліноми наближають наші вхідні дані:

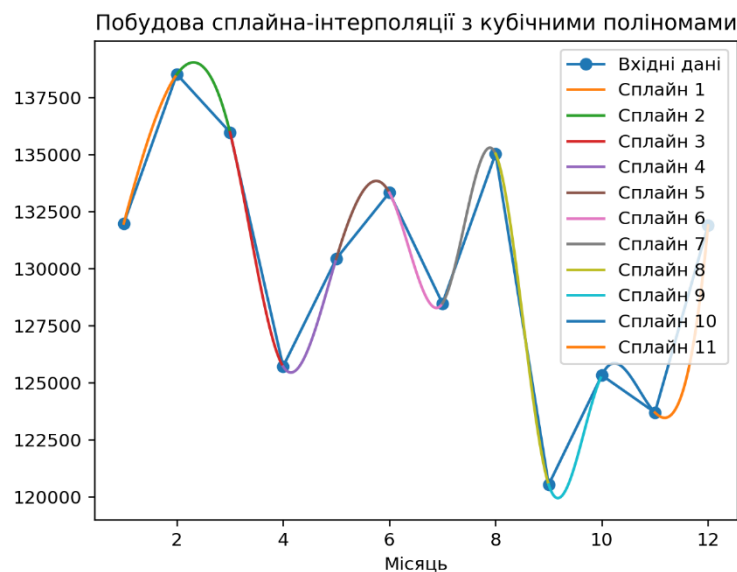


Рис. 4.31. Апроксимація сплайнами на сегментах імпорту Німеччини

Прогнози що ми отримали для квітня 2023 - 183824.20755703552, травень 2023 - 313417.63022814196 і червень 2023 - 554593.1755703548. Зразу видно, що значення на другий та третій місяць в майбутньому не виглядають реалістично.

Маємо графік для прогнозу на один місяць рис. 4.32.

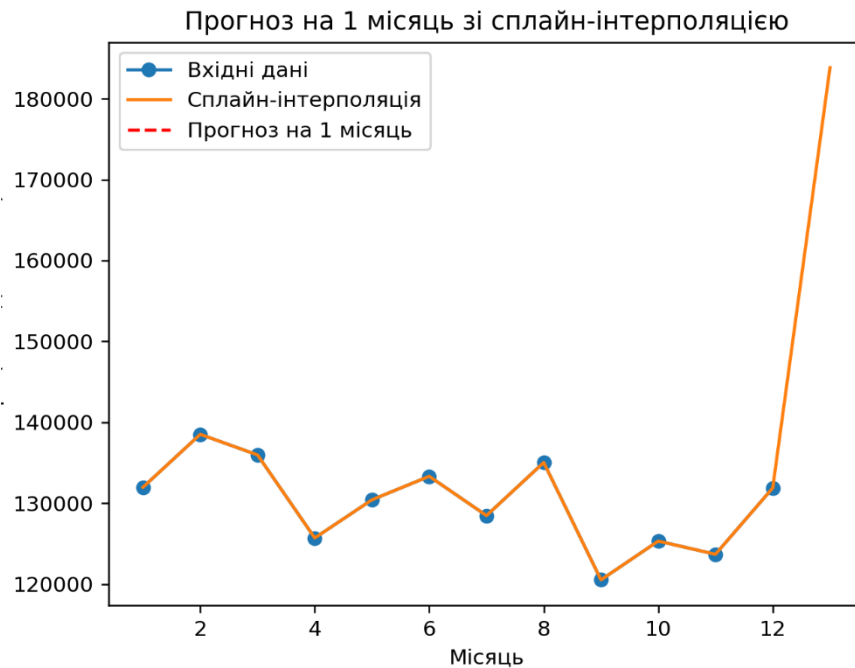


Рис. 4.32. Прогнозовані сплайном майбутні значення для імпорту Німеччини на 1 місяць

Та графік на три наступних місяці рис. 4.33.

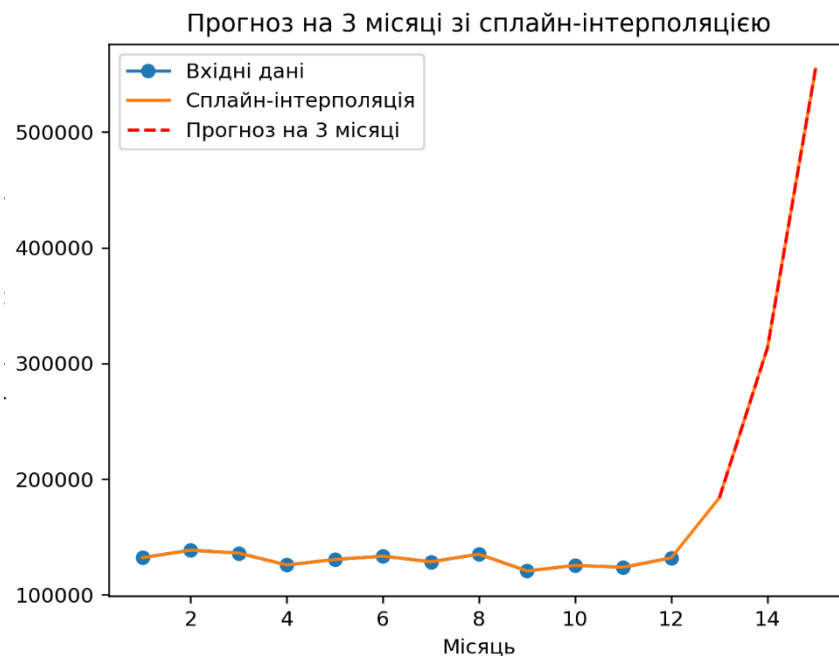


Рис. 4.33. Прогнозовані сплайном майбутні значення для імпорту Німеччини на 3 місяці

Таким чином, ми завершили реалізацію методу кубічної сплайн-інтерполяції для прогнозування зовнішньоекономічних показників України та Німеччини і зараз можемо переходити до аналізу наших результатів.

4.3 Аналіз та порівняння результатів

Визначимо, які саме значення ми збираємося рахувати, щоб оцінити наші моделі ARIMA. Можемо навести список таких метрик та навести короткі відомості, щодо алгоритму їх обчислення [19]:

1. MSE (Mean Squared Error). Метрика порівнює значення середньоквадратичної помилки та є однією з найбільш поширених в оцінці моделей ARIMA. Алгоритм обчислення наступний:
 - a. Рахуємо квадрати різниці - $(Y_i - \bar{Y}_i)^2$
 - b. Знаходимо середнє
2. MAE (Mean Absolute Error). Такою метрикою ми можемо порахувати абсолютну різницю між нашими початковими даними та їх апроксимованими значеннями. Навідміну від MSE, метрика MAE вимірює похибки без урахування їх напрямку, тобто не використовує квадратів помилок. Алгоритм обчислення наступний:
 - a. Рахуємо абсолютні різниці - $|Y_i - \bar{Y}_i|$
 - b. Знаходимо для них середнє
3. RMSE (Root Mean Square Error). Це метрика, що визначається як квадратний корінь з MSE [16]. Така операція дає змогу оцінити середню квадратичну похибку між зпрогнозованими та фактичними значеннями. Алгоритм обчислення наступний:
 - a. Корінь квадратних з MSE - \sqrt{MSE}
4. MAPE (Mean Absolute Percentage Error). Метрика, що обчислює середню відсоткову помилку в прогнозуванні. Дозволяє відносно оцінити точність прогнозу. Алгоритм обчислення наступний:
 - a. Рахуємо абсолютну відсоткову різницю - $\left| \frac{(Y_i - \bar{Y}_i)}{Y_i} \right| \cdot 100$
 - b. Знаходимо середнє

Далі наведемо результати цих метрик для наших вхідних даних.

Таблиця 4.13 показує значення метрик для експорту України.

Таблиця 4.13. Значення метрик для експорту України

Метрика	Значення
MSE	96782.0587528853
MAE	227.68143310144265
RMSE	311.0981497098389
MAPE	6.9439509140048425

Таблиця 4.14 показує значення метрик для імпорту України.

Таблиця 4.14. Значення метрик для імпорту України

Метрика	Значення
MSE	441723.62498226477
MAE	445.04466504776866
RMSE	664.6229193928425
MAPE	11.853045797616781

Таблиця 4.15 показує значення метрик для експорту Німеччини

Таблиця 4.15. Значення метрик для експорту Німеччини

Метрика	Значення
MSE	1597330257.685273
MAE	18759.8184768534
RMSE	39966.61428849425
MAPE	13.737728419711488

Таблиця 4.16 показує значення метрик для імпорту Німеччини

Таблиця 4.16. Значення метрик для імпорту Німеччини

Метрика	Значення
MSE	1496270000.734812
MAE	16384.116571540868
RMSE	38681.64940556196
MAPE	12.549105230178956

Тепер можемо сказати про якість наших прогнозів наступне [16]:

- Прогнози експорту України мають помірні значення метрик, що свідчить про задовільну точність моделей. RMSE для експорту складає 311.098, що вказує на середню помилку в наших прогнозах в розмірі близько 311 тис. доларів США. MAPE становить 6.94% і вказує це значення на відносну помилку в прогнозах.
- Прогнози імпорту України мають вищі значення метрик, що свідчить про меншу точність моделей. RMSE для імпорту України становить 664.623, тобто помилки в нашому прогнозі в середньому становлять близько 664.6 тис. доларів США, більш ніж в два рази перевищує помилку для експорту. MAPE для імпорту України складає 11.85%, що означає, що середня відсоткова помилка прогнозу складає близько 12%.
- Прогнози експорту та імпорту Німеччини мають високі значення метрик, що свідчить про невисоку точність моделей. Проте слід враховувати, що і суми імпорту та експорту в Німеччині значно перевищують аналогічні показники в Україні. RMSE для експорту Німеччини складає 39966.614 – помилки прогнозу на середньому становлять близько 39966.6 тис. євро. MAPE для експорту Німеччини показує, що середня відсоткова помилка прогнозу складає 13.74%. З цього показника вже можна сказати, що дійсно моделі для експорту України мали більш точні прогнози, при чому суттєво більш точними.
- Аналогічно, прогнози імпорту Німеччини мають високі значення метрик. RMSE для імпорту Німеччини становить 38681.649, маємо близько 38681.6 тис. євро помилки в прогнозах. MAPE для імпорту Німеччини складає 12.55%, ці дані не сильно відрізняються від MAPE в моделі для імпорту України, а тому й моделі приблизно одної якості.

Загалом, для експорту України спостерігається краща точність прогнозів порівняно з імпортом України, а також експортом і імпортом Німеччини. На графіках з підпункту 4.1 можемо бачити, що апроксимація дійсно не дуже точна, а

прогнозовані майбутні значення для українських показників виглядають більш реалістично порівняно з німецькими. Щодо майбутніх розрахунків – точність моделей може бути покращена шляхом оптимізації або використання більш складних моделей прогнозування.

Перейдемо до аналізу результатів дії методу кубічної сплайн-інтерполяції. Насправді, ми так само можемо шукати значення MSE, RMSE, MAE і додати сюди R-квадрат статистику. R-квадрат набуває значень від 0 до 1, чим ближче значення до одиниці, тим точніше апроксимує наш сплайн вхідні дані. Обчисливши ці метрики програмно для наших показників ми отримаємо інформацію про те, наскільки добре апроксимуються їх значення кубічними сплайнами. Результат для експорту України машинно оцінює перші три з зазначених метрик значеннями від 10^{-13} до 10^{-26} степені, а R-квадрат статистику взагалі одиницею. Для решти показників України та Німеччини результатом MSE, RMSE та MAE будуть нулі, а R-квадрат – одиниці. Тут слід враховувати машинну похибку, адже такі значення ідеалізовані. Можемо сказати напевне, що всі показники досить близькі до ідеальних значень, тобто метод кубічної сплайн-інтерполяції дуже точно апроксимує значення [3]. Проте, щодо прогнозів, тут ситуація й близько не ідеалізована. Сплайн інтерполяція не є найкращим методом для прогнозування майбутніх значень, особливо на віддалені періоди, які виходять за межі наявних даних. Сплайн інтерполяція працює значно ефективніше, коли мова йде про відтворення та апроксимацію наявних даних.

Ми спостерігаємо гарні результати інтерполяції для невеликого майбутнього періоду, в нашому випадку – до одного місяця. Такий результат може бути обумовлено тим, що сплайн добре підлаштовується під локальну поведінку даних та відтворює малі коливання. Однак, чим правіше ми віддаляємося від кінцевої-правої точки вхідних даних, тим гіршу точність прогнозів отримуємо, оскільки сплайн інтерполяція не має достатніх даних для адекватного апроксимування.

Якщо потрібно прогнозувати майбутні значення, рекомендується використовувати інші методи, такі як екстраполяція, регресійні моделі, часові ряди або прогнозування на основі машинного навчання. Ці методи зазвичай дають кращі результати для прогнозування на великі віддалені періоди, оскільки вони враховують тренди, сезонність та інші фактори, які впливають на майбутні значення показників.

Для покращення результатів прогнозування даних методом кубічної сплайн

інтерполяції треба більш ретельно підходити до питання обробки даних та виключати з них «ненормальність». Тобто прибирати складні тенденції та залежності, сезонності та інші завади на шляху прогнозу.

Порівнюючи методи ARIMA та кубічної сплайн інтерполяції можемо сказати таке. Обидва методи мають свої слабкі та сильні сторони. Моделі ARIMA не дають змогу відновлювати проміжні чи знаходити пропущені значення. Вони не мають точної апроксимації наших вхідних даних, але мають досить близькі до реальних прогнози майбутніх значень на досить тривалі періоди майбутнього. Метод кубічної сплайн-інтерполяції дозволяє відновлювати пропущені чи помилкові дані. На відміну від моделей ARIMA цей метод не має змоги дати точні чи навіть приблизні прогнози для досить далекого майбутнього. Проте, сплайни дозволяють детально наближувати функцію на відомих значенням, враховувати локальну поведінку вхідних значень та відтворювати найменші коливання, що дає змогу робити точні короткострокові прогнози, інколи, навіть, точніше ніж це робить метод ARIMA.

ВИСНОВКИ

Дослідження почалося з огляду чисельних методів для прогнозування даних. Ми переконалися в необхідності та ефективності такого аналізу даних, зваживши на сучасні об'єми інформації. Зокрема, цікавим було дослідити економічні показники для України та Німеччини.

Під час проведення дослідження було виконано прогнозування даних за допомогою методу ARIMA та кубічної сплайн-інтерполяції. Ми детально пояснили суть методів та алгоритм роботи з ними. Також, важливою частиною роботи була підготовка даних для подальшого аналізу. Нам вдалося досягти необхідних умов для наших вхідних даних, щоб потім успішно виконати апроксимацію та отримати точний прогноз.

Метод ARIMA виявився ефективним в прогнозуванні на середньострокові періоди. Ми отримали реалістичні прогнози на три місяці уперед, що є досить суттєвим результатом, адже вхідний набір даних містив інформацію про імпорт та експорт України та Німеччини за останні 12 місяців.

Метод кубічної сплайн-інтерполяції показав дуже точні результати для апроксимації та відтворення локальної поведінки даних в наш річний проміжок часу. Завдяки високій точності методу на наявних даних та відображенню коливань на всіх сегментах, ми змогли отримати досить чіткі короткострокові прогнози показників. В прогнозуванні на довші терміни використовувати метод недоречно, отримані результати виявлялися надто нереалістичними, а інколи й недосяжними в реальному житті.

Таким чином, для короткострокових прогнозів варто користуватися методом кубічної сплайн-інтерполяції, а для середньострокових та довгострокових – методом ARIMA. Також, якщо є потреба відновити пропущені чи помилкові дані – сплайни з цим чудово впораються. Більше того, оскільки метод ARIMA потребує суттєвої обробки даних перед прогнозуванням, то у випадку помилкових вхідних даних, ми можемо скористатися методом сплайн-інтерполяції у ролі допоміжного методу обробки та підготовки даних. Отже, наші методи можуть як виконувати ролі методів для прогнозу даних на відповідні терміни, так і доповнювати один одного.

Отримані результати – це прогнозовані значення сучасних

зовнішньоекономічних показників, тому вони можуть бути корисними для експертів в цих галузях. За допомогою прогнозів в економіці можна застерегтися від помилок та бути готовим майбутніх змін

В подальшому розвитку дослідження можна було б використати й інші методи прогнозування, такі як регресійні моделі, машинне навчання або нейронні мережі. Всі ці методи потребують більшої обчислювальної потужності, але й здатні враховувати складніші тенденції та залежності в даних. Таким чином можна буде отримати більш точні прогнози навіть на віддалені періоди.

Список використаних джерел

1. **Bank World** [Онлайновий] // The World Bank. - World Bank Group, 2023 р.. - <https://www.worldbank.org/>.
2. **Box G. E., Jenkins, G. M., & Reinsel, G. C.** Time series analysis: forecasting and control [Книга]. - [місце видання невідоме] : John Wiley & Sons, 2015.
3. **De Boor C.** A Practical Guide to Splines (Revised Edition) [Книга]. - New York : Springer, 2001.
4. **Draper N. R., & Smith, H.** Applied Regression Analysis (3rd Edition) [Книга]. - New York : Wiley, 1998.
5. **Enders W.** Applied Econometric Time Series (4th Edition) [Книга]. - Hoboken, NJ : Wiley, 2014.
6. **Eurostat** Eurostat [Онлайновий] // data-revision. - 2023 р.. - <https://ec.europa.eu/eurostat>.
7. **Granger C. W. J.** Some recent developments in a concept of causality [Книга]. - [місце видання невідоме] : Journal of econometrics, 1988.
8. **Hastie T., Tibshirani, R., & Friedman, J.** The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition) [Книга]. - New York : Springer, 2009.
9. **Hendry D. F.** Econometrics: Alchemy or science? [Книга]. - [місце видання невідоме] : Economica, 1993.
10. **Hyndman R. J., & Athanasopoulos, G.** Forecasting: Principles and Practice (2nd Edition) [Книга]. - 2018.
11. **Izenman A. J.** Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning [Книга]. - [місце видання невідоме] : Springer, 2013.
12. **James G., Witten, D., Hastie, T., & Tibshirani, R.** An Introduction to Statistical Learning [Книга]. - New York : Springer, 2013.
13. **Makridakis S., Wheelwright, S. C., & Hyndman, R. J.** Forecasting: Methods and Applications [Книга]. - [місце видання невідоме] : John Wiley & Sons, 1998.
14. **McKinney W.** Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd Edition) [Книга]. - [місце видання невідоме] : O'Reilly Media,

2018.

15. **Press W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P.** Numerical Recipes: The Art of Scientific Computing (3rd Edition) [Книга]. - Cambridge : Cambridge University Press, 2007.
16. **Robert H. Shumway David S. Stoffer** Time Series Analysis and Its Application With R Examples (3rd Edition) [Книга]. - New York : Springer, 2011.
17. **VanderPlas J.** Python Data Science Handbook: Essential Tools for Working with Data [Книга]. - [місце видання невідоме] : O'Reilly Media, 2016.
18. **І. П. Гаврилюк В. Л. Макаров** Методи обчислень [Книга]. - Київ : Вища школа, 1995. - Т. 369.
19. **МакКінлі Джеймс С., & Козлов, Александр А.** Методи прогнозування: теорія та практика [Книга]. - [місце видання невідоме] : «Юрайт», 2015.
20. **України Державна служба статистики Ukrstat** [Онлайновий] // Ukrstat. - Державна служба статистики України, 2023 р.. - <https://ukrstat.gov.ua/>.
21. **Хайзенбіль Артур Г., Меджор, Джон Т., & Орел, Сергій І.** Бізнес-прогнозування: методи, моделі та аналіз даних [Книга]. - Київ : Видавничий дім "Києво-Могилянська академія", 2019.

Додаток А. Код програмної реалізації для розділу 4

Наведемо код для побудови моделі ARIMA та кубічних сплайнів. На рисунку А.1 можемо бачити задання даних.

```
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Дані України
export_ukraine = [2685.1, 2849.5, 3361.2, 3157.1, 3189.2, 4062.7, 3775.0, 3541.7, 3361.7, 3273.2, 3508.0, 3544.9]
import_ukraine = [2823.5, 3857.9, 4849.9, 4421.1, 4446.0, 4571.2, 4838.9, 4921.6, 5353.6, 5561.0, 5624.1, 5404.2]

# Дані Німеччини
export_germany = [134541.9, 143688.5, 144407.5, 130674.7, 130879.4, 141157.5, 133361.4, 147079.9, 131935.7, 136077.3, 141889.4, 155982.0]
germany_imports = [131985.8, 138522.6, 135981.1, 125715.8, 130434.0, 133335.0, 128480.9, 135043.7, 120538.3, 125329.6, 123700.9, 131892.2]

# Перетворення даних у формат pandas.Series
export_ukraine_series = pd.Series(export_ukraine)
import_ukraine_series = pd.Series(import_ukraine)
export_germany_series = pd.Series(export_germany)
germany_imports_series = pd.Series(germany_imports)

# Графік часового ряду експорту України
germany_imports_series.plot()
plt.xlabel('Період')
plt.ylabel('Імпорт (млн.дол. США)')
plt.title('Часовий ряд імпорту Німеччини')
plt.show()
```

Рис. А.1. Задання даних

На рисунку А. 2 виконуємо тест Дікі-Фуллера та за необхідності використовуємо метод різниць, доки ряд не стане стаціонарним. Також на цьому рисунку показано побудова графіків ACF та PACF.

```

from statsmodels.tsa.stattools import adfuller

# Виконання тесту Дікі-Фуллера
result = adfuller(germany_imports_series)

# Отримання значення р-значення з результатів тесту
p_value = result[1]

# Виведення результату тесту
print(p_value)
if p_value < 0.05:
    print("Ряд стаціонарний")
else:
    print("Ряд нестационарний")

# Виконання різницевого перетворення у випадку нестационарності
diff_data = germany_imports_series.diff().dropna()
# Отримання значення р-значення з результатів тесту
result = adfuller(diff_data)
p_value = result[1]

# Виведення результату тесту
print(p_value)
if p_value < 0.05:
    print("Ряд стаціонарний")
else:
    print("Ряд нестационарний")

# Аналіз ACF та PACF
fig, ax = plt.subplots(2, 1, figsize=(10, 8))
plot_acf(diff_data, ax=ax[0])
plot_pacf(diff_data, ax=ax[1], lags=4)
plt.show()

```

Рис. А. 2. Тест Дікі-Фуллера, метод різниць та побудова ACF та PACF

Далі на рисунку А. 3, маємо задання можливих значень параметрів p , q , d та цикл, що будує усі ці моделі оцінюючи їх коефіцієнти та роблячи апроксимацію.

```

# Визначення параметрів моделі ARIMA
p_values = [0, 1, 4] # Кількість лагів авторегресії
d = 1 # Різницеве перетворення порядку d
q_values = [0, 1] # Кількість лагів ковзного середнього

# побудова моделей до можливих значень лагів
results = []

for p in p_values:
    for q in q_values:
        if p == 0 and q == 0:
            continue
        print("p= " + p.__str__() + ", q= " + q.__str__() + ", d= " + d.__str__())
        # Побудова та навчання моделі ARIMA
        model = sm.tsa.ARIMA(germany_imports_series, order=(p, d, q))
        model_fit = model.fit()

        # Отримання оцінок коефіцієнтів
        c = model_fit.params[0]
        ar_coefs = model_fit.arparams
        ma_coefs = model_fit.maparams

        # Виведення оцінок коефіцієнтів в консоль
        print("Оцінка константи (c):", c)
        print("Оцінки коефіцієнтів AR:", ar_coefs)
        print("Оцінки коефіцієнтів MA:", ma_coefs)

        # Обчислення залишкових членів
        residuals = model_fit.resid

        # Виведення залишкових членів в консоль
        print("Залишкові члени:", residuals)

        # Прогнозування значень на існуючих даних
        predicted_values = model_fit.predict()

```

Рис. А. 3. Задання можливих параметрів та побудова відповідних моделей

На рисунку А. 4 можемо бачити обчислення АІС та ВІС для всіх моделей та отримання найкращої моделі. Для найкращої моделі далі виконуємо прогноз, отримуючи прогнозований набір значень.

```

# Графік прогнозних значень та оригінальних даних
plt.plot(export_germany_series, label='Оригінальні дані')
plt.plot(predicted_values, label='Прогнозовані дані')
plt.xlabel('Період')
plt.ylabel('Імпорт (млн.дол. США)')
plt.title('Прогноз імпорту Німеччини')
plt.legend()
plt.show()

# Отримання AIC та BIC значень
aic = model_fit.aic
bic = model_fit.bic

# Збереження результатів
results.append((p, q, aic, bic))

for p, q, aic, bic in results:
    print(f"Модель (p={p}, q={q}): AIC={aic}, BIC={bic}")

best_model = min(results, key=lambda x: x[2])
best_p, best_q, best_aic, best_bic = best_model

print(f"\nНайкраща модель: (p={best_p}, q={best_q}), AIC={best_aic}, BIC={best_bic}")

best_to_forecast = sm.tsa.ARIMA(germany_imports_series, order=(best_p, d, best_q))
model_fit = best_to_forecast.fit()
# Прогнозування майбутніх значень
forecasted_values = model_fit.forecast(steps=3)

# Створення об'єкту Series без задання індексів
forecast_series = pd.Series(forecasted_values, name='Forecast')

```

Рис. А. 4. Визначення найкращої моделі та прогнозування на 3 місяці

Тепер, можемо перейти до реалізації кубічної сплайн-інтерполяції. На рисунку А. 5 можемо бачити побудову функції сплайна методом CubicSpline з бібліотеки SciPy, виведення отриманих кубічних поліномів в консоль в явному вигляді та задання графіків для виведення прогнозованих значень.

```

# Побудова кубічного сплайну для початкових значень та майбутніх прогнозів
cs = CubicSpline(x, y)

# Відрізки для побудови кубічного полінома
segments = len(x) - 1

# Виведення кубічних поліномів для кожного відрізка
for i in range(segments):
    segment_x = np.linspace(x[i], x[i + 1], num=100)
    segment_y = cs(segment_x)
    plt.plot(segment_x, segment_y, label=f'Сплайн {i+1}')
    polynomial = f'{cs.c.item(i):.6f} + {cs.c.item(i+1):.6f}*(x - {x[i]}) + {cs.c.item(i+2):.6f}*(x - {x[i]})**2 + {cs.c.item(i+3):.6f}*(x - {x[i]})**3'
    print(f'Сплайн {i+1}: {polynomial}')

# Місяці для побудови графіків
months = np.arange(1, 16)

# Графік з прогнозом на 1 місяць
plt.figure(2)
plt.plot(x[:12], y[:12], 'o-', label='Вхідні дані')
plt.plot(months[:13], cs(months[:13]), label='Сплайн-інтерполяція')
plt.plot(x[-1] + 1, cs(x[-1] + 1, extrapolate=True), 'r--', label='Прогноз на 1 місяць')

# Графік з прогнозом на 3 місяці
plt.figure(3)
plt.plot(x, y, 'o-', label='Вхідні дані')
plt.plot(months, cs(months), label='Сплайн-інтерполяція')
plt.plot(months[-3:], cs(months[-3:], extrapolate=True), 'r--', label='Прогноз на 3 місяці')

```

Рис. А. 6. Побудова сплайну та виведення результатів