

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ПРИКЛАДНОГО
СИСТЕМНОГО АНАЛІЗУ**

Кафедра математичних методів системного аналізу

До захисту допущено:

Завідувач кафедри

_____ Оксана ТИМОЩУК

«__» _____ 2024 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системний аналіз і управління»

спеціальності 124 «Системний аналіз»

на тему: «Моделі і прогнози демографічних процесів»

Виконала:

студентка IV курсу, групи КА-05

Заржецька Анастасія Олександрівна _____

Керівник:

Ст. викладач кафедри ММСА, PhD

Левенчук Людмила Борисівна _____

Консультант з економічного розділу:

Доцент, к. е. н.

Рощина Надія Василівна _____

Консультант з нормоконтролю:

К.ф.-м.н.

Статкевич Віталій Михайлович _____

Рецензент:

Доцент кафедри ШІ, PhD

Гуськова Віра Геннадіївна _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студентка _____

Київ – 2024 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Навчально-науковий інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Оксана ТИМОЦУК

« ____ » _____ 2024 р.

ЗАВДАННЯ

на дипломну роботу студентці

Заржецькій Анастасії Олександрівні

1. Тема роботи «Моделі і прогнози демографічних процесів», керівник роботи Левенчук Людмила Борисівна ст. викладач кафедри ММСА, PhD, затверджені наказом по університету від « ____ » _____ 2024 р. № _____
2. Термін подання студентом роботи 11.06.2024.
3. Вихідні дані до роботи: демографічні дані, взяті з дослідницької платформи Macrotrends.
4. Зміст роботи: особливості розвитку демографічних процесів та їх математичне моделювання, вибір та опис авторегресійних моделей та їх побудова, прогнозування демографічних процесів.

5. Перелік ілюстративного матеріалу: основні демографічні проблеми, авторегресійні моделі, критерії якості прогнозування для вибору кращої моделі, аналіз результатів.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	Доцент, к. е. н. Рощина Надія Василівна		

7. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Формулювання тематики (напрямку) дослідження.		Виконано
2	Аналіз актуальності задач стосовно тематики дослідження		Виконано
3	Аналіз відомих результатів стосовно тематики дослідження		Виконано
4	Формулювання задач дослідження		Виконано
5	Уточнення теми дипломної роботи		Виконано
6	Збір статичних даних, попередній аналіз даних		Виконано
7	Розробка програмного продукту для виконання обчислювальних експериментів		Виконано
8	Виконання обчислювальних експериментів, аналіз та оформлення результатів		Виконано
9	Оформлення пояснювальної записки у цілому		Виконано
10	Підготовка презентації для захисту		Виконано
11	Попередній захист дипломної роботи		Виконано

Студентка

Анастасія ЗАРЖЕЦЬКА

Керівник

Людмила ЛЕВЕНЧУК

РЕФЕРАТ

Дипломна робота: 113 с., 20 табл., 38 рис., 3 дод., 22 джерела.

ДЕМОГРАФІЧНІ ПРОЦЕСИ, АВТОРЕГРЕСІЙНІ МОДЕЛІ,
ПРОГНОЗУВАННЯ, ЧАСОВІ РЯДИ, PYTHON

Тема: Моделі і прогнози демографічних процесів

Дана робота зосереджена на проблемі моделювання та прогнозування демографічних процесів з використанням авторегресійних моделей, а також розробці програмного продукту для аналізу даних і побудови моделей.

Об'єкт: дослідження демографічних процесів в Японії та Данії.

Предмет дослідження: вибір та аналіз структур математичних моделей, які адекватно описують демографічні процеси для побудови якісного прогнозу.

Мета роботи: оцінити поточну демографічну ситуацію в Японії та Данії шляхом моделювання демографічних показників та надати оцінку прогнозу на наступні 4 роки.

Методи дослідження: методи аналізу часових рядів, зокрема використання авторегресійних моделей та їх модифікацій.

В роботі представлені результати моделювання та аналізу демографічних процесів з використанням рівнянь авторегресії. Проведено прогнозування та оцінено побудовані моделі для різних демографічних процесів. Розроблено математичні моделі та програмний продукт для їхньої реалізації, що дозволяє виконувати прогнозування та аналіз демографічних тенденцій.

ABSTRACT

Thesis: 113 p., 20 tables, 38 fig., 3 app., 22 references.

DEMOGRAPHIC PROCESSES, AUTOREGRESSIVE MODELS,
FORECASTING, TIME SERIES, PYTHON

Topic: Models and forecasts of demographic processes

This work focuses on the problem of modeling and forecasting demographic processes using autoregressive models, as well as the development of a software product for data analysis and model construction.

Object: study of demographic processes in Japan and Denmark.

Subject of research: selection and analysis of the structures of mathematical models that adequately describe demographic processes for the construction of quality forecasts.

Purpose: to assess the current demographic situation in Japan and Denmark by modeling demographic indicators and provide an assessment of the forecast for the next 4 years.

Methods of research: time series analysis methods, specifically the use of autoregressive models and their modifications.

The study presents the results of modeling and analyzing demographic processes using autoregressive equations. Forecasting and evaluation of the constructed models for various demographic processes have been conducted. Mathematical models and a software product have been developed for their implementation, enabling forecasting and analysis of demographic trends.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ	9
ВСТУП	10
РОЗДІЛ 1 ОСОБЛИВОСТІ РОЗВИТКУ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ ТА ЇХ МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ	11
1.1 Особливості демографічних процесів та актуальність дослідження. .	11
1.1.1 Демографічна ситуація в світі	12
1.1.2 Демографічні процеси в Україні	15
1.2 Підходи до математичного моделювання та прогнозування демографічних процесів.	16
1.3 Комп’ютерні системи для побудови моделей демографічних процесів 20	
1.4 Висновки до розділу 1	22
Постановка задачі.....	23
РОЗДІЛ 2 ВИБІР І ДОКЛАДНИЙ ОПИС МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ	24
2.1 Структура математичної моделі	24
2.1.1 Узагальнений алгоритм побудови моделі	24
2.1.2 Аналіз на наявність нелінійностей	25
2.1.3 Аналіз на стаціонарність	26
2.1.4 Гомо- та гетероскедастичність	27
2.1.5 Аналіз на коінтегрованість.....	29
2.2 Математичні моделі	30
2.2.1 AR(p) – модель авторегресії.....	31

2.2.2	ARMA(p,q) – модель авторегресії ковзного середнього	31
2.2.3	ARIMA – модель авторегресії з інтегрованим ковзним середнім 32	
2.3	Критерії адекватності моделі	33
2.4	Критерії вибору кращого прогнозу	36
2.5	Висновки до розділу 2.....	37
РОЗДІЛ 3 ПОБУДОВА МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ		39
3.1.	Вхідні дані	39
3.2.	Вибір інструментальної платформи	39
3.3.	Побудова математичних моделей демографічних процесів Японії. 40	
3.3.1	Кількість населення	41
3.3.2	Народжуваність	45
3.3.3	Смертність	51
3.4.	Побудова математичних моделей демографічних процесів Данії... 55	
3.4.1	Кількість населення	55
3.4.2	Народжуваність	59
3.4.3	Смертність	64
3.5.	Висновки до розділу 3.....	68
РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ		70
4.1	Постановка задачі проектування	71
4.2	Обґрунтування функцій програмного продукту	71
4.3	Обґрунтування системи параметрів програмного продукту	74
4.4	Аналіз експертного оцінювання параметрів.....	77

4.5	Аналіз рівня якості варіантів реалізації функцій	81	
4.6	Економічний аналіз варіантів розробки ПП	83	
4.7	Вибір кращого варіанту ПП техніко-економічного рівня	89	
4.8	Висновки до розділу 4.....	90	
ВИСНОВКИ ЗА РОБОТОЮ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ			
ДОСЛІДЖЕНЬ.....			91
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ			93
ДОДАТОК А (ілюстративний матеріал доповіді)			95
ДОДАТОК Б (статистичні дані)			107
ДОДАТОК В (код програмного продукту)			111

ПЕРЕЛІК СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ

МНК – метод найменших квадратів

АР – авторегресія

МА – ковзне середнє

АРМА – авторегресія з ковзним середнім

АРІМА – авторегресія з інтегрованим ковзним середнім

АКФ – автокореляційна функція

ЧАКФ – частково автокореляційна функція

DW – критерій Дарбіна-Уотсона

R^2 – коефіцієнт детермінації

AIC – інформаційний критерій Акайке

SSE – сума квадратів похибок

ВСТУП

В сучасному світі демографічні процеси відіграють ключову роль у формуванні та розвитку суспільства. Вони визначають динаміку народжуваності, смертності, міграції та інших важливих аспектів людської популяції, впливаючи на соціально-економічний розвиток країн та регіонів.

Однак, розуміння цих процесів та їхніх майбутніх тенденцій вимагає використання математичних моделей та прогностичних методів. В даній роботі було застосовано моделі та методи прогнозування у дослідженні демографічних процесів, для розуміння та аналізу взаємозв'язків між різними демографічними показниками, а також для здійснення прогнозів на їх майбутню динаміку.

У розділі 1 розглянуто загальні характеристики демографічних процесів та їх актуальність для сучасного суспільства, а також різні підходи до математичного моделювання та прогнозування демографічних процесів.

У розділі 2 розглянуто структуру математичних моделей, що використовуються для опису демографічних процесів. Представлено авторегресійні моделі та їх модифікації. Наведено критерії адекватності моделей та методи їх вибору для забезпечення якісного прогнозування.

Розділ 3 присвячено побудові математичних моделей демографічних процесів в Японії та Данії. Розглянуто вхідні дані, які використовуються для моделювання та обґрунтовано вибір інструментальної платформи для проведення досліджень. Представлено побудову конкретних моделей для різних демографічних процесів та здійснено їх аналіз.

У розділі 4 виконано функціонально-вартісний аналіз програмного продукту.

РОЗДІЛ 1 ОСОБЛИВОСТІ РОЗВИТКУ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ ТА ЇХ МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

1.1 Особливості демографічних процесів та актуальність дослідження.

У теперішньому світі постає досить велика кількість глобальних проблем, які чинять неабиякий вплив на життя кожної людини та загалом на розвиток людства. Однією з найважливіших серед них є демографічні процеси, що стосуються народонаселення [1].

Демографічні процеси — зміни чисельності, структури та розміщення населення, що відбуваються протягом певного часу. Ці процеси можуть бути наслідком різноманітних факторів, включаючи народжуваність, смертність, міграцію, як внутрішню, так і зовнішню та інші демографічні явища. А, власне, демографія – це наука, що вивчає закономірності у відтворенні населення в суспільно-історичній обумовленості даного процесу [2].

Демографічні процеси є важливим аспектом розвитку будь-якої країни, оскільки вони впливають на її економічну, соціальну та політичну ситуацію. Сьогодні особливо гостро постає питання нерівномірного зростання населення, яке у багатьох частинах світу не може бути цілком забезпечене усім необхідним для життя. В той час в інших регіонах Землі відбувається старіння націй. Такий дисбаланс має прямий, досить негативний вплив на майбутнє розвитку і життя людства. Неминучі виклики урбанізації також в свою чергу загрожують існуванню людства. Станом на сьогодні понад 60% мешканців планети проживають у містах, що призводить до перенаселення міст, збільшення промисловості та рівня забруднення, виснаження ресурсів і, як наслідок, погіршення здоров'я та психоемоційного стану людей. Задля протидії та запобігання виникнення кризових ситуацій і усунення загроз формується «демографічна політика».

Немає однозначного визначення цьому терміну, але ми можемо трактувати його як певні дії держави спрямовані на зміну параметрів народонаселення для отримання бажаного результату. Тобто, для економічно розвинених країн – підвищення рівня народжуваності, а для країн, що розвиваються – на зниження рівня народжуваності. Проблема демографії розглядається вже дуже давно. Багато науковців вважає, що за 30-35 років кількість людей працездатного віку суттєво зменшиться, тоді як значно зростатиме частка населення, що перебуває у віці старше 65-70 років.

1.1.1 Демографічна ситуація в світі

У ХХ ст. у світі відбулися найбільші зміни кількості населення. Цьому посприяли такі чинники: висока народжуваність, збільшення тривалості життя, покращення охорони здоров'я і, звичайно, зменшення смертності.

Населення Землі стрімко зростає, хоча ще чотири століття тому, зміна кількості населення мала значно повільніший характер.

Так можемо прослідкувати (рис. 1.1) збільшення кількості населення починаючи з 1960 року, коли кількість людей на планеті була 3,03 мільярда. Станом на 2022 рік це число зросло до 7,95 мільярдів осіб – це більше ніж у 2,5 рази, що свідчить про значно більше використання земних ресурсів. Разом з цим отримуємо нестачу їжі та дефіцит прісної води, від якого страждає четверта частина населення світу.

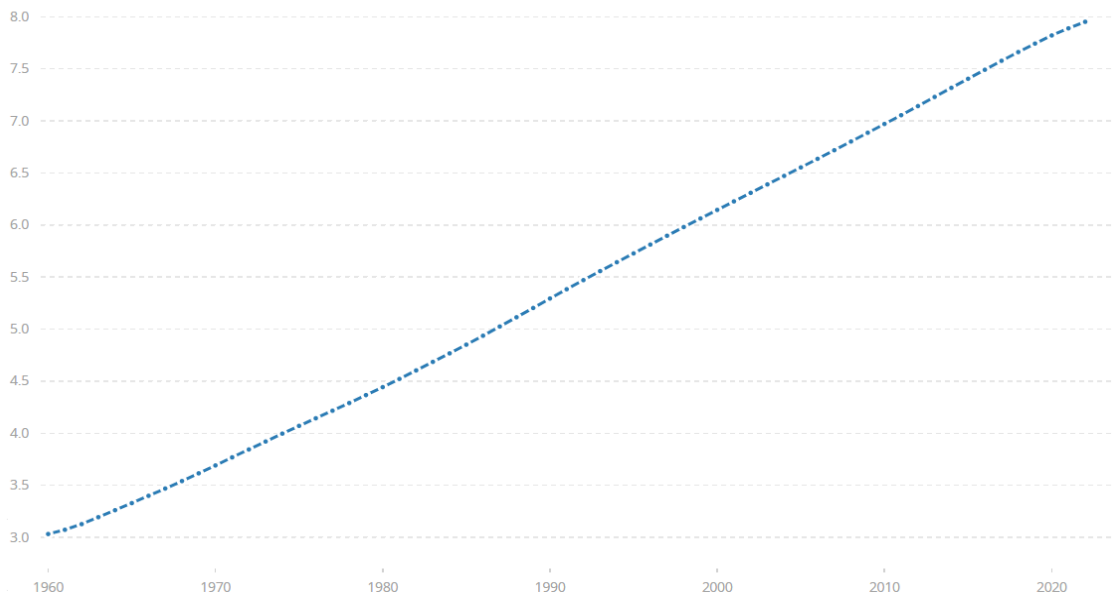


Рисунок 1.1 – Населення світу

Всього 20% населення світу проживає у розвинених країнах. Найшвидше зростає населення в країнах, що розвиваються, Азії, Африки, Латинської Америки. Причинами цьому є релігія, заборона абортів, ранній вік вступу в шлюби, традиції мати великі сім'ї, низький рівень освіченості та ін. Разом з цим постає велика проблема з безробіттям. Число людей, які не можуть знайти собі заробіток сягає сотні мільйонів у країнах, що розвиваються. Також для цих країн характерна відносно низька смертність, тут переважна частка – молоде, працездатне населення, більшість якого – чоловіки.

Протилежна ситуація в розвинених країнах, де відносно невеликі показники народжуваності, смертності та природного приросту. Таким країнам властиве «старіння» населення, адже частка людей віком від 65 років, переважає молодше населення. До них можна віднести країни Європи, США, Японія, Нова Зеландія, Австралія, Канада та інші. Порівняльні статистичні дані у відсотках станом на 2023 рік у декількох країнах і у світі загалом можна побачити нижче (рис. 1.2), де блакитним виділено відсоток людей до 15 років, а чорним – люди від 65.

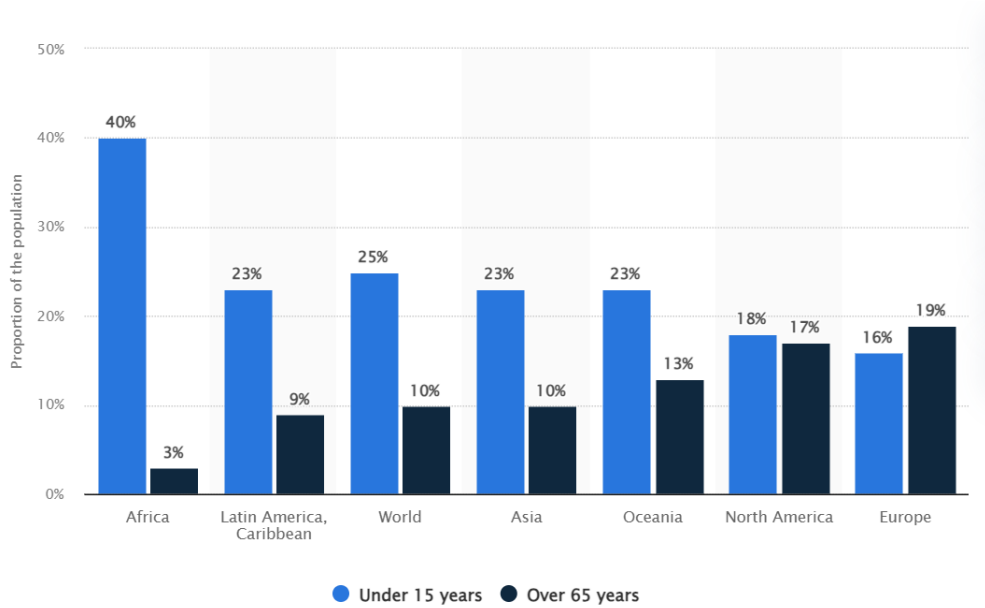


Рисунок 1.2 – Населення світу за віком і регіонами у 2023 році

Можемо помітити, що значна частка приросту населення припадає саме на Африку. Тут проживає наймолодше населення у світі. На континенті 40% населення молодше 15 років і близько трьох відсотків старше 65 років, що свідчить про низьку тривалість життя в деяких країнах. З іншого боку, в Європі більша частка населення старше 65 років, ніж населення віком до 15 років.

Розглянемо динаміку «старіння» населення на прикладі Японії (рис. 1.3).

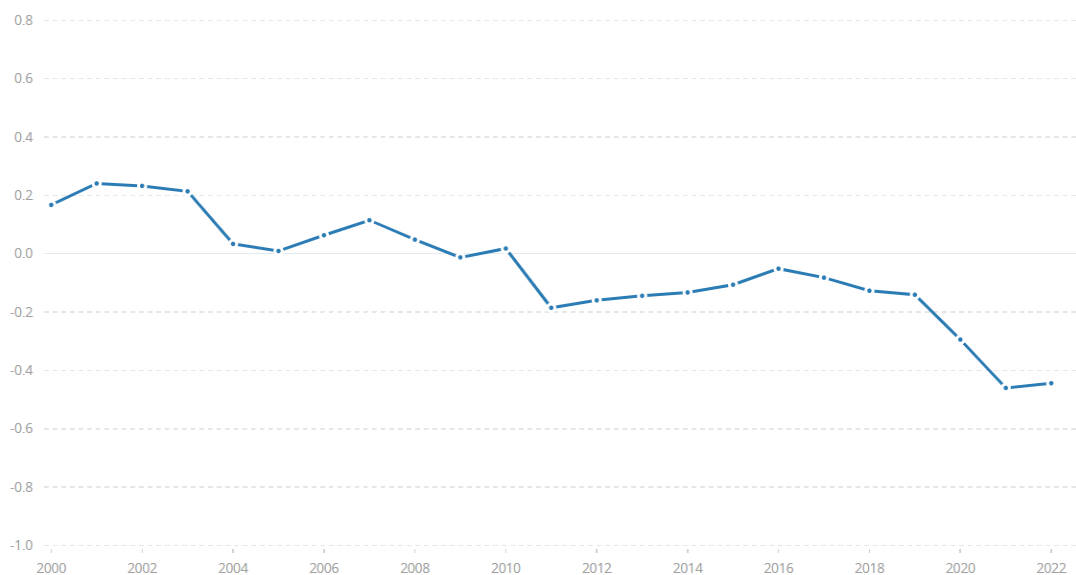


Рисунок 1.3 – Приріст населення у відсотках (Японія)

Населення Японії найстарше у світі. У 2022 році природний приріст становив $-0,4\%$, а люди у віці 65 років і старше становили приблизно 29% населення. Згідно з прогнозом, до 2070 року частка людей від 65 років становитиме близько 38% населення Японії. Ця країна стискається з низьким рівнем народжуваності, що призводить до зменшення населення [3].

1.1.2 Демографічні процеси в Україні

На превеликий жаль, у зв'язку з війною дуже важко сказати щось про кількість населення в Україні сьогодні, за даними Інституту Демографії станом на початок 2023 року це число варіювалося від 28 млн до 34 млн осіб. З початку повномасштабного вторгнення Україну покинуло близько 8 млн біженців. Кожного дня збільшується число загиблих та зниклих безвісті, що дуже трагічно впливає на демографію та економіку нашої держави.

Розглянемо як змінювалася кількість населення починаючи з 1960 року (рис. 1.4)

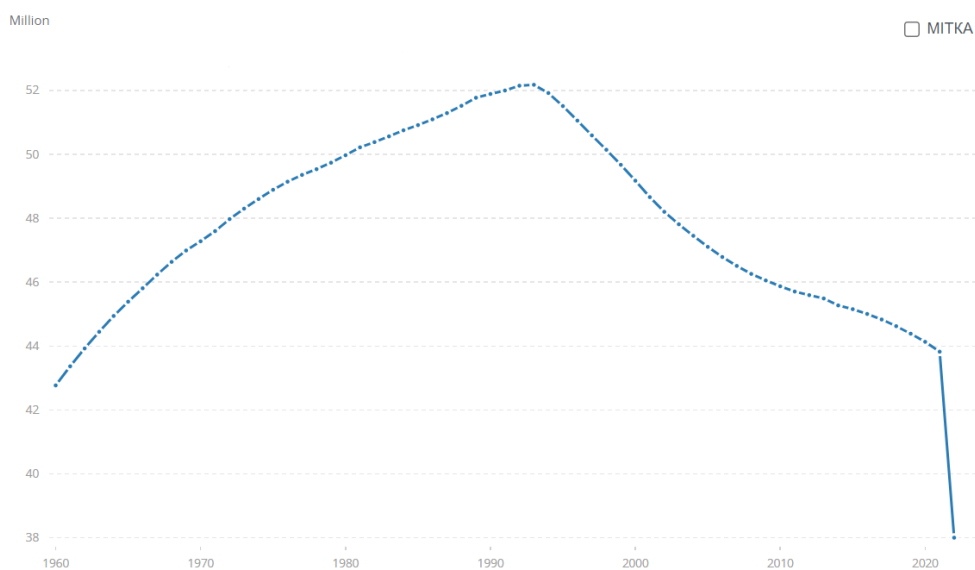


Рисунок 1.4 – Динаміка чисельності населення України

Спостерігаємо ріст населення до 1993 року, коли його кількість сягала 52 179 200 осіб, а потім стрімкий спад. Це пов'язано з низькою народжуваністю, високим рівнем смертності, як наслідок низький показник природного приросту, а також зі «старінням» нації, та трудовою міграцією. В більшості це молоді кваліфіковані працівники. Також переважна кількість населення – жінки.

До початку повномасштабної війни у 2021 році в Україні проживало близько 48,8 млн осіб, зараз це число значно менше, адже війна дуже загострила демографічну ситуацію в країні.

1.2 Підходи до математичного моделювання та прогнозування демографічних процесів.

Для всебічного дослідження демографічних процесів країни чи регіону, виявлення основних характеристик і закономірностей та вивчення взаємозв'язків процесів зміни чисельності населення з тенденціями соціально-економічного розвитку використовують методи прогнозування.

Демографічний прогноз - це науково-обґрунтоване судження на рахунок майбутньої демографічної ситуації в країні, регіоні чи світі. За допомогою методів демографічного прогнозування можна визначити наступні показники: чисельність населення, природний приріст, вікова та статева структури населення, народжуваність, смертність, внутрішня та зовнішня міграції. Прогноз майбутньої динаміки кількості населення для вирішення соціально-економічних проблем та усунення загроз виникнення небажаних наслідків демографічних процесів за допомогою побудови сценаріїв – це основна мета демографічного прогнозування. Адже це прямо впливає на майбутнє виробництво, а отже споживання продуктів харчування, ресурсів, на охорону здоров'я, будівництво, забезпеченням робочих місць, належною освітою та

пенсіями, розвиток підприємств, інфраструктури та на інше соціально-економічне планування. Прогнозування буває глобальним (якщо розглядати весь світ в цілому), регіональне, національне та територіальне. Статистичне моделювання і прогнозування ґрунтується на наступних типах експериментальних даних: часові ряди і часові перерізи.

Часовий ряд – це множина рівновіддалених в часі спостережень, які описують поведінку процесу на певному часовому інтервалі.

Часовий переріз – це множина спостережень змінних досліджуваних процесів на вибраний конкретний момент часу [4].

Тому надалі будемо працювати з моделями часових рядів.

На даний момент відомо близько 200 методів прогнозування з використанням часових рядів. Розглянемо одні з найвідоміших і найпоширеніших з них для прогнозування демографічних процесів:

- екстраполяційний;
- регресійні моделі;
- метод пересування вікових груп;
- метод марківських ланцюгів;
- аналітичний.

Екстраполяція — це техніка прогнозування, яка поширює минулі тенденції або моделі на майбутнє, щоб передбачити майбутнє значення змінної. Методи екстраполяції широко використовуються в демографії для прогнозування кількості населення, розподілу за віком, народжуваності, смертності та інших демографічних показників на основі історичних даних.

Даний метод заснований на безпосередньому використанні лінійних та експоненціальних функцій. Для аналізу демографічних тенденцій важливо мати дані про середньорічні зміни чисельності населення та середні темпи зростання або приросту населення. При наявності цих показників, чисельність населення можна розрахувати на будь-яку наступну кількість років, припускаючи, що вони залишаться незмінними протягом усього прогнозного

періоду. Прогнозований процес при екстраполяції виражається як функція часу, яка акумулює вплив інших факторів, що визначають його інтенсивність та напрямок. Як правило, всі прогнози засновані на екстраполяції тенденцій в тій чи іншій формі. Наприклад, метод експертних оцінок передбачає ретельне вивчення перебігу певного процесу експертами з різних галузей науки та екстраполяцію тренду [5].

Методи екстраполяції базуються на припущенні, що виявлені в минулому тенденції зберігатимуться і в майбутньому. Завдяки своїй простоті вони є найпоширенішим методом [5]. Вони дають уявлення динаміки демографічних явищ, якщо розвиток продовжується за встановленою траєкторією. Наприклад, розуміючи, що тривале зниження народжуваності спричинить старіння населення, яке триватиме і надалі, можна розрахувати ступінь цього старіння в майбутньому. Користуючись цим, можна визначити загальне зменшення мобільності населення, оскільки молодь характеризується найвищою міграційною мобільністю. Недолік екстраполяційних моделей полягає в тому, що вони не враховують особливості розвитку певних груп населення, таких як вікові групи або когорти. Саме з цієї причини даний метод не застосовується при їх прогнозуванні. Екстраполяційні прогнози точніше можна назвати методами оцінки тенденцій.

Регресійні моделі – це статистичний метод, який використовується для аналізу зв'язку між залежною змінною та однією чи декількома незалежними змінними. У демографії регресійні моделі часто використовуються для розуміння того, як демографічні фактори, такі як вік, стать, освіта та дохід, впливають на такі результати, як народжуваність, смертність і моделі міграції. У даному методі незалежною змінною є визначена матеріальна характеристика, що виступає у ролі фактору, а не час, як у простій екстраполяції [11]. Цей метод зручніше використовувати для регіонального прогнозування.

Метод пересування вікових груп або когортно-компонентний метод— це демографічний метод, який коригує дані про населення з урахуванням змін

у віковому розподілі з часом. Це передбачає перерозподіл чисельності населення від однієї вікової групи до іншої на основі спостережуваних тенденцій і припущень щодо демографічних процесів, таких як народжуваність, смертність і міграція. Такий метод можна вважати одним з найпоширеніших і найбільше використовуваних.

Метод марківських ланцюгів - це підхід імовірнісного моделювання, який ґрунтується на ймовірностях переходу одиниць генеральної сукупності з однієї групи в іншу, а також вибуття зі спостереження (смерть, міграція та інші). У демографії моделі ланцюгів Маркова можна використовувати для прогнозування динаміки населення шляхом моделювання переходів між демографічними станами (віковими групами, сімейним станом та іншими) з часом. Передбачається, що отримані коефіцієнти залишаються постійними (незмінними) протягом певного інтервалу часу. При цьому формується така структура, що залежить не від початкових умов, а лише від матриці переходу. Інколи в літературі цей процес зустрічається під назвою закон ергодичності (властивість системи “забувати” минулу структуру) [5].

Аналітичний метод відноситься до широкої категорії демографічних методів, які використовують математичні та статистичні моделі для аналізу динаміки населення та прогнозування. Цей метод передбачає використання різних аналітичних інструментів для розуміння змін у населенні та їхніх причин. Він заснований на тому, що на основі минулих демографічних тенденцій вибираються функції, які мають емпіричний характер та найкраще їх представляють. Даний метод в основному застосовується для прогнозування на короткотривалих проміжках часу.

Кожен з цих методів є важливим інструментом для дослідження та керування демографічними процесами. Кожен з них має свої переваги та обмеження, тому їхнє використання варіюється залежно від конкретного контексту дослідження. Використання різних методів аналізу демографічних процесів дає змогу отримати більш повне та точне уявлення динаміки змін у населенні та їхніх наслідків.

1.3 Комп'ютерні системи для побудови моделей демографічних процесів

Сьогодні різноманітні демографічні зміни та тенденції чинять вагомий вплив на соціально-економічний розвиток країн. Вивчення та прогнозування демографічних процесів є критично важливим для науковців, політиків та практиків.

Використання комп'ютерних систем і програмного забезпечення є одним із найважливіших елементів демографічних досліджень, моделювання та прогнозування демографічних процесів.

Ці інструменти дають змогу ефективно виявляти складні зв'язки, виконувати аналіз великих обсягів даних та робити прогнози, важливі для прийняття рішень у таких різноманітних сферах, як економіка, охорона здоров'я, освіта та соціальна політика. На сьогодні нараховується декілька тисяч ПЗ для обробки і аналізу даних на ринку статистичних пакетів. Розглянемо найпоширеніші з них:

– SAS (Statistical Analysis System) – потужна інтегрована система для аналізу даних, статистичного моделювання, візуалізації даних і бізнес-аналітики, що використовується в наукових дослідженнях, бізнесі, фінансах, медицині, демографії та багатьох інших сферах для обробки та аналізу великих обсягів даних. Має велику кількість функцій і процедур, високу ефективність, надає підтримку та навчання своїм користувачам. Серед недоліків можемо виділити дорожню вартість ліцензій (тому ця система доступна не всім) та залежність від одного постачальника програмного забезпечення, що може вплинути на гнучкість та відкритість системи. Незважаючи на наявність навчання, SAS є досить складною, початковим користувачам буде непросто розібратися.

– STATISTICA – пакет статистичного аналізу загального призначення, який реалізує основні математичні методи аналізу даних. Має широкий вибір статистичних методів і інструментів для аналізу даних, простий у використанні та має гнучкий інтерфейс, в STATISTICA наявна підтримка інтеграції з іншими програмами., можливість конвертувати дані в таблиці чи бази даних, порівняно з SAS, є значно легшою в освоєні. Серед недоліків – обмежена підтримка та розвиток.

– EViews (Econometric Views) – статистичний пакет для аналізу економетричних даних часових рядів, аналізу та моделювання даних, а також створення регресійних моделей. За допомогою цього пакету можна реалізовувати фінансовий аналіз, моделювання економічних процесів, макроекономічне прогнозування, аналізувати наукову інформацію та ін. Пакет EViews надає широку функціональність для аналізу даних часових рядів. Завдяки спрощеній і логічній ідеології побудови інтерфейсу, пакет EViews дуже простий у вивченні та містить розширену довідку, яка фактично є керівництвом до економетричних методів [12].

Завдяки модульній структурі системи Statistica можна користуватися різними модулями для вирішення різних класів завдань, таких як аналіз часових рядів, множинна регресія, нелінійне оцінювання та інші. Система дає можливість імпортувати дані і працювати з ними.

На відміну від Statistica, EViews не є модульною системою, але має об'єктну структуру робочого вікна, що дозволяє працювати з різними типами інформації. У програмі EViews є командний рядок для статистичного аналізу даних, а також можливість створення файлів з послідовністю команд для подальшого використання. В даній системі наявний великий набір функцій для аналізу, візуалізації та моделювання часових рядів та інших типів даних.

Система SAS є потужним комплексом з різними програмними продуктами. Вона має високо розвинену функціональність та набір статистичних алгоритмів, що робить її ефективним інструментом для

економетричного аналізу, моделювання і прогнозування. Система також має можливості програмування та роботи з базами даних, що робить її універсальним інструментом для різних видів аналізу даних.

Хоча кожна з цих систем має свої переваги, такі як гнучкість у Statistica, об'єктна структура у EViews та висока функціональність у SAS, вони також мають свої недоліки. Statistica є менш ефективною для роботи з великими обсягами даних, EViews вимагає деякої експертизи в програмуванні для повного використання його функцій, а SAS є недешевою системою, тому може бути недоступною для деяких користувачів.

Задля створення власних алгоритмів, адаптації їх до конкретних потреб і забезпечення оптимальної продуктивності можна використовувати мови програмування, наприклад, Python, R, C++, C# та інші. Перші дві є найбільш поширеними для аналізу та використовуються в роботі з часовими рядами. Python має широкий вибір бібліотек для аналізу даних, таких як Pandas, NumPy, SciPy. R надає багато вбудованих методів для статистичного аналізу, включаючи аналіз часових рядів та має широкий спектр пакетів для аналізу.

1.4 Висновки до розділу 1

Розглядаючи світові демографічні тенденції, можна зазначити, що популяція на планеті швидко зростає, але такий ріст спостерігається переважно в країнах третього світу, таких як Азійський та Африканський регіони та Південна Америка. У той же час, в розвинених країнах, як наслідок, низької народжуваності та «старіння» нації природній приріст є незначним, а в деяких випадках навіть від'ємним.

Дослідження різних пакетів для статистичної обробки даних та моделювання, а також інструментів для розробки власних програмних продуктів, дає змогу зробити вибір середовища, яке буде найбільш

ефективним для моделювання демографічних процесів та оцінювання прогнозів.

Враховуючи теперішню ситуацію в Україні, вимушену міграцію та аналізуючи статистику демографічного стану, можна прийти до невтішного висновку. Демографічні процеси демонструють специфічні особливості, такі як зменшення тривалості життя, низький рівень народжуваності та високий рівень смертності. Ці фактори сприяють депопуляції, скорочення природного приросту та старінню населення, що потребує розробки спеціальних стратегій для пом'якшення негативних наслідків.

Сьогодні демографічні проблеми є дуже актуальною та важливою темою для людства. Адже, вони досить тісно пов'язані з економічними, соціальними, екологічними, політичним та іншими процесами.

Щоб якомога ефективніше вивчити та проаналізувати цю проблему, можна використовувати аналіз часових рядів, новітні технології та методи математичного моделювання та прогнозування.

Постановка задачі

1. Зробити огляд існуючих моделей демографічних процесів.
2. Зібрати та обробити статистичні дані.
3. Розробити математичні моделі для прогнозування динаміки населення в майбутньому.
4. Здійснити оцінку якості результатів демографічних прогнозів.
5. Візуалізувати прогнозні значення та порівняти їх зі спостережуваними даними та зробити висновки про наявних тенденцій та прогнозування майбутніх демографічних процесів.

РОЗДІЛ 2 ВИБІР І ДОКЛАДНИЙ ОПИС МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ

2.1 Структура математичної моделі

Для того щоб приступити до застосування математичних методів прогнозування та побудови їх моделей, потрібно чітко сформулювати структуру моделі. Це критично важливий етап, оскільки некоректно побудована модель може призвести до неточних результатів або навіть вказати на неправильні залежності між вихідними даними та результатами прогнозування. Аналіз на інтегрованість допоможе виявити стійкість змін у часовому ряді. Для визначення структури моделі також важливо виявити можливість гетероскедастичності та коінтегрованості змінних. Гетероскедастичність може свідчити про нерівномірність дисперсії у часовому ряді, тоді як коінтегрованість вказує на наявність стійкого статистичного зв'язку між змінними. Коли процес проаналізовано, можна побудувати приблизну структуру моделі для майбутнього прогнозу та аналізу даних.

2.1.1 Узагальнений алгоритм побудови моделі

Загальний алгоритм побудови моделі такий:

- попереднє аналізування процесу, для якого будуємо модель;
- виконання обробки експериментальних даних;
- перевірка часових рядів на наявність нестационарності і нелінійності;

- вибір структур моделей-кандидатів, за допомогою обчислення та аналізування кореляційної матриці для часових рядів, визначення екзогенних змінних та врахування їх в модель;
- обрахунок автокореляційної (АКФ) та часткову автокореляційну функції (ЧАКФ) залежної змінної для визначення порядку авторегресійної частини моделі;
- вибір методів та оцінка параметрів математичних моделей;
- вибір кращої з оцінених моделей-кандидатів за допомогою критеріїв відбору;
- перевірка моделі на адекватність [13].

2.1.2 Аналіз на наявність нелінійностей

Аналіз нелінійності є важливою складовою процесу моделювання, оскільки багато реальних явищ не підкоряються простим лінійним залежностям. Для визначення нелінійності та їх типу користуються візуальним аналізом. Існують різні статистичні тести, які можуть допомогти виявити нелінійність у даних: тест на нормальність, тест на адекватність моделі та тест на гетероскедастичність. Розглянемо один з тестів на наявність нелінійності. Цей тест зазвичай використовується у випадку можливості набору декількох груп спостережень для того ж процесу:

$$\hat{F} = \frac{\frac{1}{m-2} \sum_{i=1}^m n_i (\bar{y}_i - \hat{y})^2}{\frac{1}{n-m} \sum_{i=1}^m n_i (\bar{y}_i - \hat{y})^2 \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2} \quad (2.1)$$

де \bar{y}_i – середнє значення для i -ї реалізації даних;
 \hat{y}_i – середнє для лінійної апроксимації даних;

- m – кількість груп даних;
- n_i – кількість вимірів в i -й групі;
- n – загальна кількість вимірів.

Фактично, наведена статистика представляє собою таке відношення:

$$\hat{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{Відхилення значень } y(k) \text{ від групових середніх}} \quad (2.3)$$

Якщо статистика з степенями свободи $\nu_1 = m - 2$ та $\nu_2 = n - m$ досягає або перевищує рівень значущості, то гіпотеза щодо лінійності має бути відхилена. Серед недоліків цього підходу можна виділити необхідність застосування декількох (не менше трьох) груп даних для одного і того ж процесу, що можуть бути отримані при виконанні повторних експериментів. Очевидно, що це не завжди можливо [6].

2.1.3 Аналіз на стаціонарність

До класу нестаціонарних процесів належать процеси з одиничними коренями. Стаціонарність можна розділи на слабку (по коваріації, другого порядку та в широкому сенсі) і сильну. Процеси стаціонарні по коваріації вимагають незалежне від часу математичне сподівання, постійну дисперсію та незмінну в часі автоковаріацію. На відміну від слабкої стаціонарності, сильна не потребує постійне математичне сподівання та дисперсію.

Для стохастичного процесу $AR(1)$ маємо такі умови стаціонарності:

- модуль коефіцієнта (корінь характеристичного рівняння) $|a_1| < 1$;
- $A=0$ (однорідний розв'язок) або забезпечується збіжність $a_1^k \rightarrow 0$ для процесів, які почалися давно в минулому.

Для $AP(p)$ умовою стаціонарності є рівність нулю його однорідного розв'язку.

Також, стаціонарним вважається будь-який процес ковзного середнього, що має скінченний порядок.

Для перевірки стаціонарності моделі часового ряду можна використовувати тест Дікі-Фуллера, KPSS-тест та інші статистичні тести, що допомагають визначити, чи є модель стаціонарною.

2.1.4 Гомо- та гетероскедастичність

Гомоскедастичність – постійна дисперсія залишків для кожного спостереження. Навпаки, гетероскедастичність – це явище зміни дисперсії залишків для кожного спостереження. Наслідком гетероскедастичності залишків може бути порушення ефективності оцінки параметрів моделі МНК (метод найменших квадратів).

Гетероскедастичні процеси відносяться до класу нестаціонарних процесів, до якого входять:

- процеси, у яких змінна дисперсія;
- процеси, які мають детерміновані і стохастичні тренди;
- процеси, які одночасно характеризуються змінною в часі дисперсією та змінним математичним сподіванням;
- процеси із змінною коваріацією (дисперсійно-коваріаційна нестаціонарність) [6].

Виявити наявність гетероскедастичності можна за допомогою таких тестів:

- Уайта;
- Бройша-Пагана (Годфрі);

- Голдфельда-Квандта;
- тест рангової кореляції Спірмена;
- тест Парка.

Розглянемо детальніше один з них.

Тест Уайта — статистичний тест, який встановлює наявність гомоскедастичності. З метою виявлення гетероскедастичності потрібно побудувати допоміжну модель регресії для квадратів залишків, що були згенеровані за допомогою застосування методу найменших квадратів до часових рядів [7].

Алгоритм роботи:

1. Оцінка коефіцієнтів основної регресії.
2. Оцінка регресії квадратів залишків, виявлення їх квадратів і констант.
3. Обрахунок R^2 – коефіцієнту множинної детермінації.
4. Обчислення тестової статистики за формулою NR^2 . При виконанні гіпотези про існування гетероскедастичності тестова статистика має розподіл $\chi^2(q)$, де q – кількість регресорів.
5. Порівняння отриманого значення тестової статистики з критичним. Якщо воно перевищує критичне значення, то гіпотеза про гомоскедастичність відкидається.

Зазвичай тест заперечує існування гетероскедастичності в моделі. Припускається, що залишки мають незмінну дисперсією. В такому разі, допоміжна регресія не чинить значний вплив на основне рівняння. Можна застосувати метод LM-статистики (тест Лагранжа) аби зробити перевірку гіпотези про відсутність гетероскедастичності. Основною метою в застосуванні цього методу є перевірка обмежень статистичної моделі. Можна вважати, що модель містить гетероскедастичність, якщо значення статистики перевищує критичний поріг [7].

2.1.5 Аналіз на коінтегрованість

Коінтеграція - статистичне вираження концепції довгострокових зв'язків між нестационарними (інтегрованими) змінними.

Для коінтегрованості часових рядів існує дві необхідні передумови: часові ряди повинні бути нестационарними та мати однаковий порядок інтеграції. Якщо ж друга умова не виконується і змінні мають різні порядки інтегрування, то вони взагалі не можуть бути коінтегровані.

Але це й не означає, що змінні з однаковим порядком інтегрування обов'язково коінтегровані.

Перевірити наявність коінтегрованості можна за допомогою різноманітних тестів, наприклад, двоетапного тесту Інгла-Гренджера, тесту Йогансена, непараметричного тесту Філіпса-Перрона та інших [8].

Перевірка за тестом Інгла-Гренджера передбачає послідовність таких етапів:

I. Оцінка коефіцієнтів рівняння довгострокової рівноваги. Якщо залишки рівняння довгострокової рівноваги $(y_{1t} = \gamma_0 + \gamma_1 y_{2t} + \varepsilon_t)$ є стаціонарним рядом, то коінтегрованість існує. Перевірити залишки на стаціонарність можна за допомогою тесту Дікі-Фуллера.

II. Процес виконання тесту Інгла-Грейнджера. На основі базової моделі тестуються наступні гіпотези:

$H_0: b = 0$; часові ряди не коінтегрують: залишки $\hat{\varepsilon}_t$ є нестационарними;

$H_1: b < 0$; часові ряди коінтегрують: залишки $\hat{\varepsilon}_t$ є стаціонарними;

Базовим рівнянням є регресійне рівняння:

$$\Delta \hat{\varepsilon}_t = b \hat{\varepsilon}_{t-1} + \sum_{i=1}^k c_i \Delta \hat{\varepsilon}_{t-1} + u_t \quad (2.4)$$

$$\Delta \hat{\varepsilon}_t = \hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}$$

де $\Delta \hat{\varepsilon}_t$ – випадкова величина;

b, c_i – коефіцієнти регресії;

$\hat{\varepsilon}_t$ – залишки рівняння довгострокової рівноваги.

Перша гіпотеза відкидається, якщо коефіцієнт $b < 0$ і розрахована t -статистика за абсолютною величиною більша за абсолютну величину критичного значення τ -статистики МакКіннона для тестування на коінтеграцію із заданим рівнем значущості. При цьому, вважається, що часові ряди коінтегрують. Для залишків рівняння довгострокової рівноваги, нормалізованого за другою змінною ($y_{2t} = \gamma_0 + \gamma_1 y_{1t} + \varepsilon_t$) можемо провести такий же тест. Якщо для них теж нульова гіпотеза відкидається, то можемо вважати, що часові ряди коінтегрують в двох випадках [9].

2.2 Математичні моделі

Розглянемо математичні моделі регресії, які використовуються для аналізу демографічних процесів. Регресійні моделі дозволяють визначити взаємозв'язок між залежною змінною (наприклад, рівнем народжуваності чи смертності) та незалежними змінними (наприклад, економічними показниками, рівнем освіти, віковою структурою населення тощо). Основними типами регресійних моделей, що використовуються в демографічних дослідженнях, є авторегресійні моделі (AR), моделі авторегресії ковзного середнього (ARMA) та моделі авторегресії з інтегрованим ковзним середнім (ARIMA).

2.2.1 AR(p) – модель авторегресії

Авторегресійна модель застосовується для прогнозування майбутніх значень часового ряду на основі попередніх значень цього ряду. Наприклад, у моделі авторегресії AR(p) порядку p поточне значення залежить від p попередніх значень часового ряду.

Рівняння авторегресії можемо записати так:

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t, \quad (2.5)$$

де y_t – значення часового ряду в момент часу t

ε_t – імпульс білого шуму

a_p – коефіцієнти моделі

c – константа

Умова стаціонарності: всі корені характеристичного рівняння мають перебувати у межах одиничного кола.

Авторегресійна модель є найпростішою, проте може не завжди точно відображати складні процеси. Дана модель є основою для складніших моделей.

2.2.2 ARMA(p,q) – модель авторегресії ковзного середнього

Модель авторегресії є узагальненням авторегресійної моделі (значення часового ряду залежить від його попередніх значень) та ковзного середнього (значення часового ряду залежить від попередніх значень шуму моделі) та також використовується для прогнозування стаціонарних часових рядів.

Модель авторегресії ковзного середнього ARMA(p,q), де цілі числа – p,q, що задають порядок моделі, можна описати таким рівнянням:

$$y_t = c + \varepsilon_t + \sum_{i=1}^p a_i y_{t-i} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad (2.6)$$

де a_i – коефіцієнти авторегресії

b_j – коефіцієнти ковзного середнього

c – константа

y_t – значення часового ряду в момент часу t

ε_t – імпульс білого шуму

Модель ARMA(p,q) має складнішу структуру, ніж моделі авторегресії (AR) або ковзного середнього (MA), які мають подібну поведінку. Однією з переваг цієї моделі є менша кількість параметрів.

Для побудови моделі ARMA потрібно визначити порядок моделі (числа p і q), за допомогою характеристик часового ряду, таких як автокореляційна функція (АКФ) та часткова автокореляційна функція (ЧАКФ) та обчислити коефіцієнти. Визначити коефіцієнти AR можна за допомогою методів найменших квадратів та максимальної правдоподібності.

2.2.3 ARIMA – модель авторегресії з інтегрованим ковзним середнім

Модель ARIMA є розширенням попередньої моделі ARMA, але застосовується для нестационарних часових рядів. Ці ряди можна перетворити на стаціонарні шляхом визначення різниць деякого порядку від вихідного часового ряду. Модель авторегресії з інтегрованим ковзним середнім записується таким чином:

$$\Delta y_t = c + \varepsilon_t + \sum_{i=1}^p a_i \Delta y_{t-i} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad (2.7)$$

де a_i – коефіцієнти авторегресії

b_j – коефіцієнти ковзного середнього

c – константа

y_t – значення часового ряду в момент часу t

ε_t – імпульс білого шуму

Δ – оператор різниці часового ряду

Спочатку оцінюємо наявність стаціонарності ряду. За допомогою тестів виявляємо наявність поодиноких коренів і порядок інтегрованості часового ряду. Якщо порядок інтегрованості більше від нуля, то ряд перетворюємо визначенням різниці порядків та після цього можемо побудувати модель авторегресії з ковзним середнім, адже передбачається, що наш процес являється стаціонарним [10].

2.3 Критерії адекватності моделі

Одну з ключових ролей в моделюванні та прогнозуванні часових рядів відіграють критерії адекватності моделі, оскільки вони допомагають забезпечити надійність та точність результатів аналізу та прогнозування. Вони дозволяють оцінити ефективність процесу побудови та визначити наскільки добре модель підходить для опису та прогнозування даних обраних процесів. За допомогою критеріїв адекватності моделей можна порівнювати різні моделі між собою і вибрати ту, яка найкраще підходить для аналізу та прогнозування часового ряду. В сучасному світі існує великий вибір таких

критеріїв, які надають оцінку адекватності тим чи іншим моделям. Попри це застосування деяких критеріїв може призвести до неоднозначних результатів вибору найкращої моделі. Тож, аби цього уникнути рекомендується оцінювати адекватність за кількома критеріями одночасно. Розглянемо найвідоміші з них:

– Статистика Стьюдента (t-критерій) – критерій значимості коефіцієнтів простої лінійної регресії, який записується так:

$$t = \frac{\hat{a} - a^0}{SE_{\hat{a}}} \quad (2.8)$$

де \hat{a} - оцінка коефіцієнта моделі

a^0 – нуль-гіпотеза щодо цієї оцінки

$SE_{\hat{a}}$ – стандартна похибка оцінки

Цілком логічно, що чим менше значення похибки, тим кращою є оцінка коефіцієнтів. Для перевірки значимості спочатку ми висуваємо нуль-гіпотезу, обчислюємо значення статистики Стьюдента для кожного коефіцієнта та знаходимо її критичне значення. Потім перевіряємо нуль-гіпотезу за нерівністю: $-t_{кр} < t < t_{кр}$.

– Коефіцієнт детермінації R^2

$$R^2 = \frac{var(\hat{y})}{var(y)} \quad (2.9)$$

де $var(\hat{y})$ – дисперсія залежної змінної

$var(y)$ – дисперсія вимірів залежної змінної

Виходячи з рівняння коефіцієнта детермінації, можемо зробити висновок, що для найкращої адекватної моделі він має прямувати до 1. В

найкращому випадку $R^2 = 1$. Також ми можемо вважати його як рівень інформативності моделі.

- Інформаційний критерій Акайке (AIC)

$$AIC = N \ln \left[\sum_{k=1}^N e^2(k) \right] + 2p, \quad (2.10)$$

AIC -критерій для вибору найкращої з кількох статистичних моделей. Тому за допомогою цього критерію потрібно декілька моделей, він не дає змогу оцінити кожну модель окремо на адекватність. Чим менше значення має критерій Акайке, тим кращою є модель. Він дозволяє врахувати довжину досліджуваної вибірки, кількість параметрів та суму квадратів похибок.

- Статистика Дарбіна-Уотсона

За допомогою цієї статистики маємо змогу визначити ступінь корельованості похибок моделі.

$$D = 2 - 2p \quad (2.11)$$

де p – коефіцієнт кореляції між сусідніми значеннями випадкової змінної. При відсутності кореляції між похибками статистика Дарбіна-Уотсона буде рівна 2 (найбільш прийнятне значення).

- Статистика Фішера (f-статистика)

За допомогою цієї статистики визначається ступінь адекватності моделей в цілому. Вона дає можливість порівняти дисперсії варіаційних рядів.

Фактичне значення статистики Фішера порівнюється з табличними значеннями.

Якщо фактичне значення критерію F більше табличного ($F > F_{\text{табл}}$), то модель можна вважати адекватною. Для багатофакторної регресії статистику Фішера можемо записати так:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{(N - p - 1)}{p} \quad (2.12)$$

Дана статистика, як ми можемо бачити, залежить від коефіцієнту детермінації. Тому при $R^2 \rightarrow 1$ статистика $F \rightarrow \infty$ [10].

2.4 Критерії вибору кращого прогнозу

Вибір кращого прогнозу також як і вибір кращої моделі є важливим етапом в прогнозуванні. Тому для оцінки якості прогнозів потрібно використовувати статистичні критерії. Так як і для перевірки адекватності моделі, для оцінки якості прогнозу доцільно користуватися декількома критеріями. Нижче наведено деякі з них [14]:

- СКП – середньоквадратична похибка:

$$\text{СКП} = \sqrt{\frac{1}{S} \sum_{i=1}^S (y(k + s) - \hat{y}(k + s, k))^2} \quad (2.13)$$

- СП – середня похибка прогнозу:

$$\text{СП} = \frac{1}{S} \sum_{i=1}^S y(k + s) - \hat{y}(k + s, k) \quad (2.14)$$

- СПП – середня похибка у відсотках:

$$\text{СПП} = \frac{1}{S} \sum_{i=1}^S \frac{y(k+s) - \hat{y}(k+s, k)}{y(k+s)} \cdot 100\% \quad (2.15)$$

- АСПП – середня абсолютна похибка у відсотках:

$$\text{АСПП} = \frac{1}{S} \sum_{i=1}^S \frac{|y(k+s) - \hat{y}(k+s, k)|}{|y(k+s)|} \cdot 100\% \quad (2.16)$$

- МАП – максимальна абсолютна похибка:

$$\text{МАП} = \max \{|y(k+1) - \hat{y}(k+1, k)|, \dots, |y(k+s) - \hat{y}(k+s, k)|\} \quad (2.17)$$

- МіАП – мінімальна абсолютна похибка:

$$\text{МіАП} = \min \{|y(k+1) - \hat{y}(k+1, k)|, \dots, |y(k+s) - \hat{y}(k+s, k)|\} \quad (2.18)$$

2.5 Висновки до розділу 2

Цей розділ присвячений математичному моделюванню та прогнозуванню демографічних процесів. У ньому було розглянуто різні аспекти побудови моделей та наведено узагальнений алгоритм побудови, який надає системний підхід до розв'язання завдань прогнозування. Одним з головних етапів при побудові моделі є визначення її структури, включаючи перевірку стаціонарності, лінійності, виявлення гетероскедастичності, а також аналіз коінтегрованості. Для цього було наведено декілька тестів. Для

виявлення стаціонарності можна використовувати тест Дікі-Фуллера, KPSS-тест та інші статистичні тести. Для виявлення нелійності використовують тест на нормальність, тест на адекватність моделі та тест на гетероскедастичність. Виявити наявність гетероскедастичності можна за допомогою тесту Уайта, Бройша-Пагана (Годфрі), Голдфельда-Квандта, тесту рангової кореляції Спірмената та тесту Парка. Здійснити перевірку на наявність коінтегрованості можна за допомогою двоетапного тесту Інгла-Гренджера, тесту Йогансена, непараметричного тесту Філіпса-Перрона та інших.

Аналіз наявності нелійностей дозволяє врахувати складні закономірності демографічних процесів.

Було розглянуто математичні моделі регресії, які використовуються для аналізу демографічних процесів, а саме, AR, ARMA та ARIMA.

Також було продемонстровано критерії адекватності моделі та якості прогнозу. За їх допомогою можна оцінити достовірність та точність отриманих результатів. Математичні моделі регресії, описані в цьому розділі, дозволяють враховувати вплив різних факторів на демографічні процеси, що сприяє більш точному та обґрунтованому прогнозуванню.

РОЗДІЛ 3 ПОБУДОВА МАТЕМАТИЧНИХ МОДЕЛЕЙ ДЕМОГРАФІЧНИХ ПРОЦЕСІВ

3.1. Вхідні дані

У даному розділі було розглянуто демографічні процеси Японії та Данії, як об'єкти дослідження для побудови математичних моделей. Адже обидві країни стикаються з важливими демографічними викликами, які мають серйозний вплив на їх соціально-економічний розвиток. В Японії демографічні процеси, зокрема зниження народжуваності, зростання середнього віку населення та зменшення приросту населення, є одними з найяскравіших у світі. Тому вивчення демографічних тенденцій у Японії є важливим та актуальним завданням.

Данія, хоч і не має таких виражених демографічних проблем, як Японія, також стикається з демографічними викликами. В Данії спостерігається зростання середнього віку населення та стабільне, але не дуже високе зростання народжуваності. Ці тенденції можуть мати значний вплив на соціально-економічну політику країни. Тому вивчення демографічних процесів у Данії також є актуальним і важливим завданням.

Основним джерелом даних було взято веб-ресурс Macrotrends. Це веб-сайт, який надає доступ до великого обсягу фінансових і економічних даних та аналізів. Він зосереджується на довгострокових тенденціях і макроекономічних показниках, що охоплюють широкий спектр тем, від фінансових ринків до демографічних і економічних індикаторів.

3.2. Вибір інструментальної платформи

Для побудови математичних моделей демографічних процесів було обрано Python як основну мову програмування. Python має величезну кількість

бібліотек для аналізу даних, моделювання, візуалізації та машинного навчання. Також було використане середовище Google Colab для виконання коду. Використання Google Colab надає низку переваг, таких як доступ до потужних обчислювальних ресурсів, можливість легко ділитися кодом і результатами з іншими дослідниками, а також інтеграція з Google Drive для зберігання даних. Код програми було реалізовано за допомогою таких бібліотек, як Pandas, NumPy, Matplotlib, Statsmodels, Scikit-learn, що значно спростило процес обробки даних та побудови моделей.

- Pandas – бібліотека для обробки та аналізу даних, що дозволяє зручно працювати з таблицями даних.
- NumPy – бібліотека призначена для виконання математичних операцій та роботи з масивами, за допомогою якої можна обчислювати статистичні показники та нормувати дані
- Statsmodels – бібліотека для статистичного моделювання, включаючи регресійний аналіз та моделі часових рядів, що дозволяє створювати регресійні моделі для аналізу демографічних даних
- Matplotlib – бібліотека для візуалізації даних, побудова графіків для аналізу та презентації результатів.
- Scikit-learn – бібліотека, що включає інструменти для підготовки даних, оцінки моделей допомогою критеріїв, таких як сума середньоквадратичних похибок (SSE) та коефіцієнт детермінації (R^2).

3.3. Побудова математичних моделей демографічних процесів Японії

Побудова математичних моделей для аналізу демографічних процесів є важливою задачею для розуміння та прогнозування змін у населенні країни. У випадку Японії, яка стикається з проблемами старіння населення і

зменшенням чисельності, такі моделі можуть бути особливо корисними для політичного планування та управління ресурсами.

3.3.1 Кількість населення

Кількість населення - це один з ключових демографічних показників, що вказує на загальну кількість людей, які проживають у країні чи регіоні в певний момент часу. Цей показник є важливим для визначення рівня розвитку та соціально-економічного статусу даної території.

Аналіз та прогнозування зміни кількості населення є важливими завданнями для урядів, дослідницьких установ, міжнародних та громадських організацій, оскільки вони дозволяють розробляти ефективні стратегії розвитку та управління соціально-економічними процесами.

Розглянемо динаміку зміни чисельності Японії в період з 1950-2023 роки (рис. 3.1)

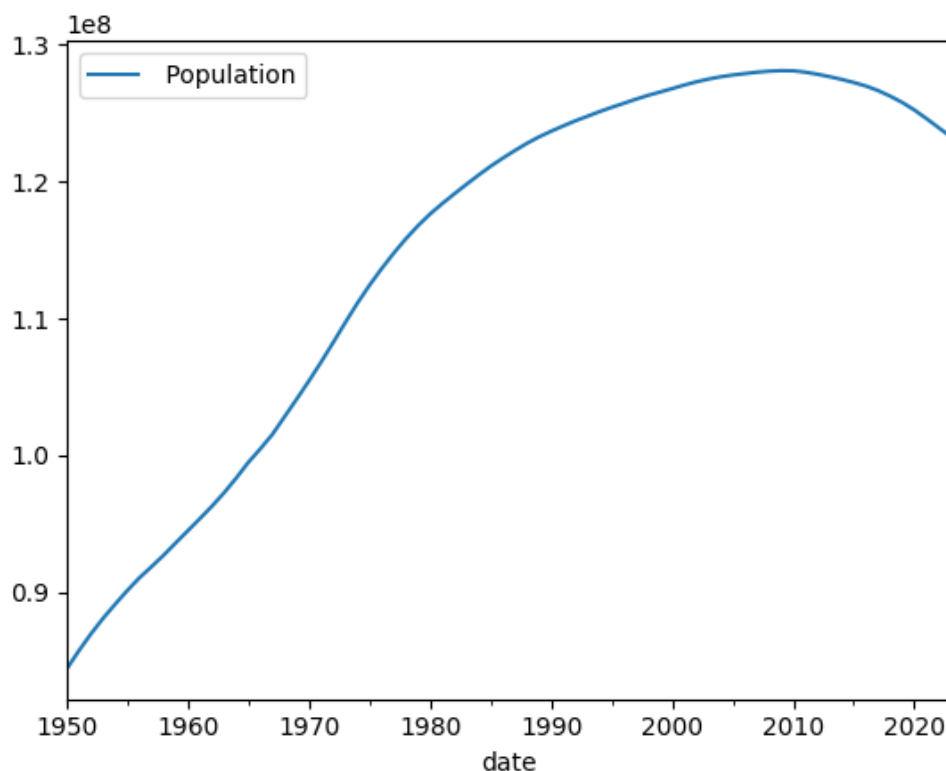


Рисунок 3.1 – Зміна кількості населення Японії в період 1950-2023

Будуємо автокореляційну (АКФ) та частково автокореляційну (ЧАКФ) функції з використанням 19 лагів для визначення порядку регресійних моделей: (рис. 3.2 та рис. 3.3).

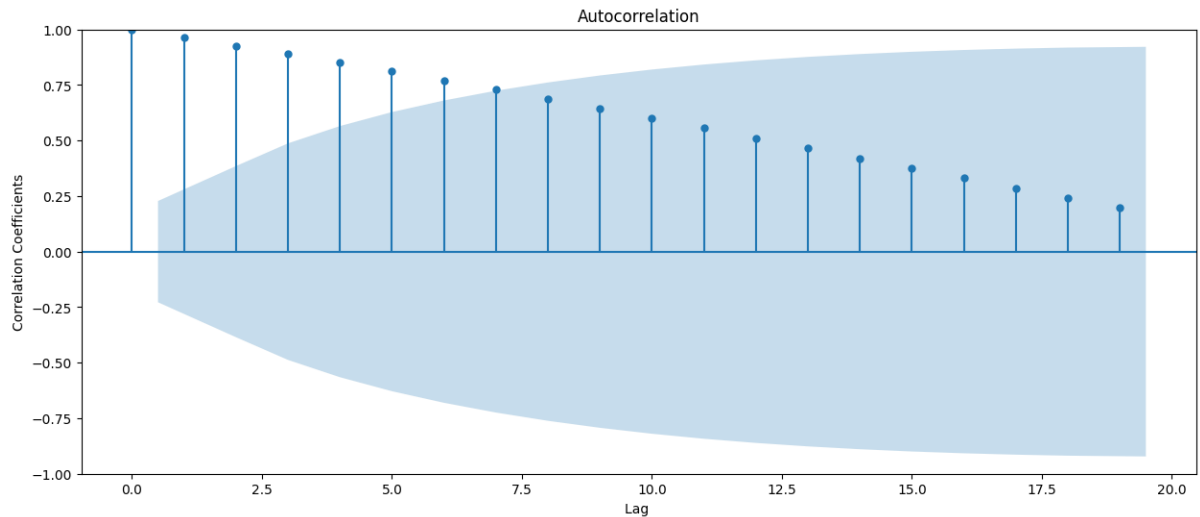


Рисунок 3.2 – АКФ

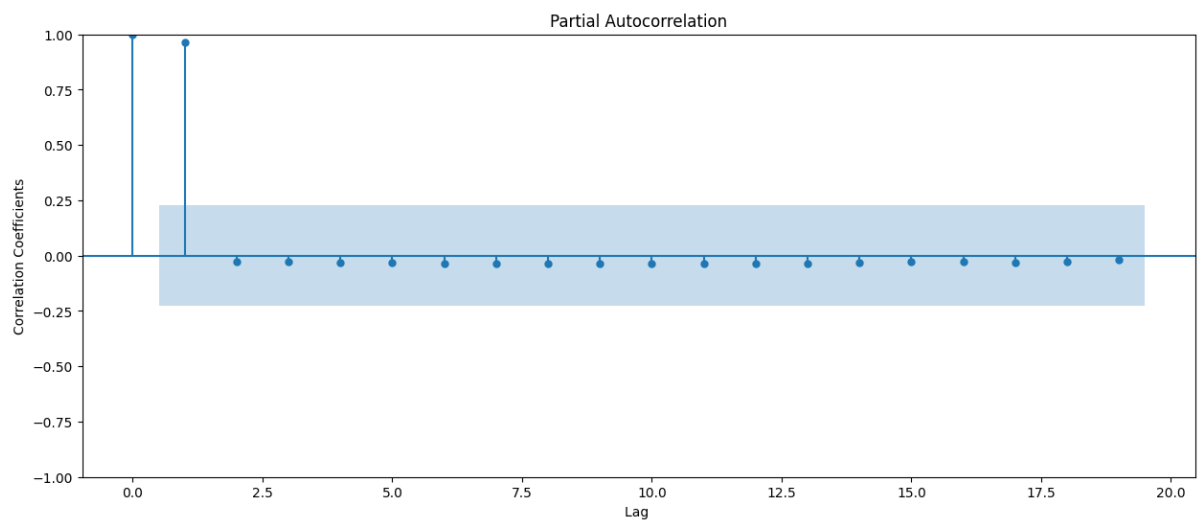


Рисунок 3.3 – ЧАКФ

Аналізуючи автокореляційну та частково автокореляційну функції, було вирішено побудувати авторегресійну модель другого порядку.

$$\text{Рівняння AR(2): } y(k) = c + a_1y(k-1) + a_2y(k-2) = 9.47e^7 + 1.9964y(k-1) - 0.9969y(k-2)$$

Тепер на основі цієї авторегресійної моделі побудуємо модель ARMA. Для цього нам потрібно визначити порядок моделі ковзного середнього (MA). За допомогою розробленої програми було побудовано 15 моделей вигляду: ARMA(2,1), ARMA(2,2), ARMA(2,3) і так далі. Кожна з цих моделей оцінювалася за трьома критеріями адекватності, а саме, SSE (сума квадратів похибок), R^2 (критерій детермінації) та DW (Статистика Дарбіна-Уотсона). Потім за цими критеріями обирався кращий порядок ковзного середнього і краща модель ARMA в цілому.

У нашому випадку це ARMA(2,1), рівнянням якої є:

$$y(k) = 9.47e^7 + 1.9595y(k-1) - 0.9907y(k-2) - 0,7501\varepsilon(k-1)$$

Порівняємо отримані моделі між собою за допомогою тих критеріїв, на основі яких обирали порядок ковзного середнього (табл. 3.1):

Таблиця 3.1 – Порівняння моделей кількості населення

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R^2	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	$10e^{12}$	0,7036	2,4353	$5,192e^{10}$	0,7272	0,024
ARMA(2,1)	$5e^{12}$	0,8566	1,826	$5,19e^{10}$	0,7224	0,024

Найкращі значення критеріїв:

- SSE – менше значення свідчить про кращу модель;
- R^2 – значення критерію детермінації прямує до 1 для кращої моделі;
- DW – значення статистики Дарбіна-Уотсона для кращої моделі повинно дорівнювати 2;

- СКП (середня квадратична похибка) – менше значення свідчить про кращу модель;
- САПП (середня абсолютна похибка в процентах) – менше значення свідчить про кращу модель;
- Коефіцієнт Тейла (вимірює точність прогнозів) – менше значення свідчить про кращу модель.

При оцінці моделей за допомогою критерію суми квадратів похибок отримали дуже великі числа, це пов'язано з масштабом даних нашого дослідження. Адже ми розглядаємо кількість населення починаючи з 1950 року, коли чисельність становила 84 353 049, а вже у 2009 році в Японії налічувалась найбільша кількість осіб – 128 117 042. Як ми можемо помітити за цей час кількість населення зросла в 1,5 рази.

Отже, порівнюючи ці моделі, кращою моделлю є ARMA(2,1), тому на основі цієї моделі побудуємо прогноз на майбутні 4 роки (рис. 3.4):

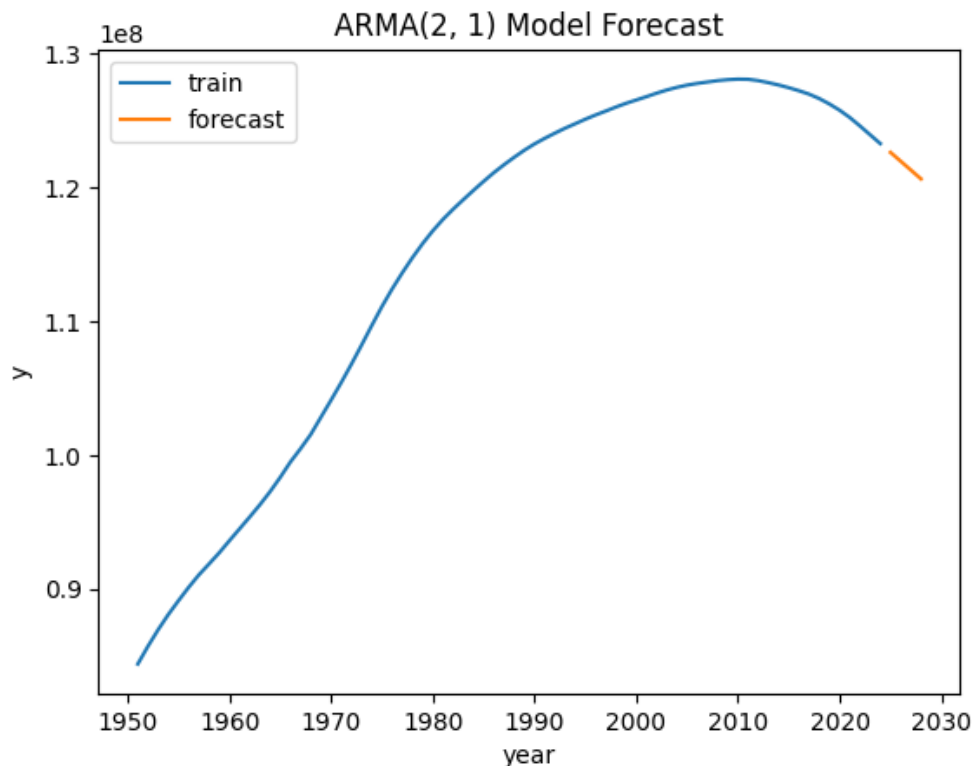


Рисунок 3.4 – Прогноз кількості населення на 4 роки

Чисельність населення Японії зменшується з кожним роком, починаючи з 2009. За побудованим прогнозом це буде й продовжуватися надалі як зазначено в табл. 3.2.

Таблиця 3.2 – Прогнозовані значення чисельності населення Японії

Рік	Чисельність населення
2024	1226368
2025	1229767
2026	1213150
2027	1206524

3.3.2 Народжуваність

Як і кількість населення, народжуваність є важливим параметром для дослідження демографії країн, оскільки вона визначає приріст населення та впливає на багато аспектів соціально-економічного життя країни. На рис. 3.5 можна побачити графік зміни коефіцієнту народжуваності з 1950 року.

Коефіцієнтом народжуваності вважаємо демографічний показник відношення кількості народжених за рік на 1000 жителів країни.

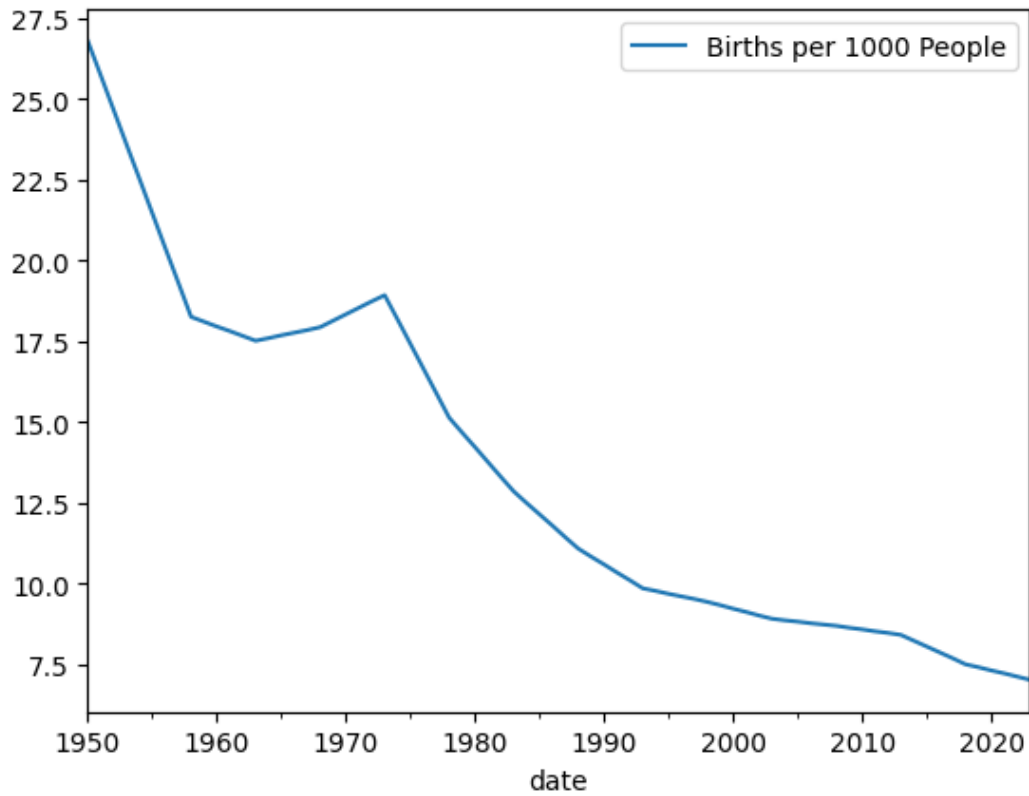


Рисунок 3.5 – Народжуваність Японії

Будуємо АКФ та ЧАКФ для коефіцієнту народжуваності (рис. 3.6 та рис 3.7):

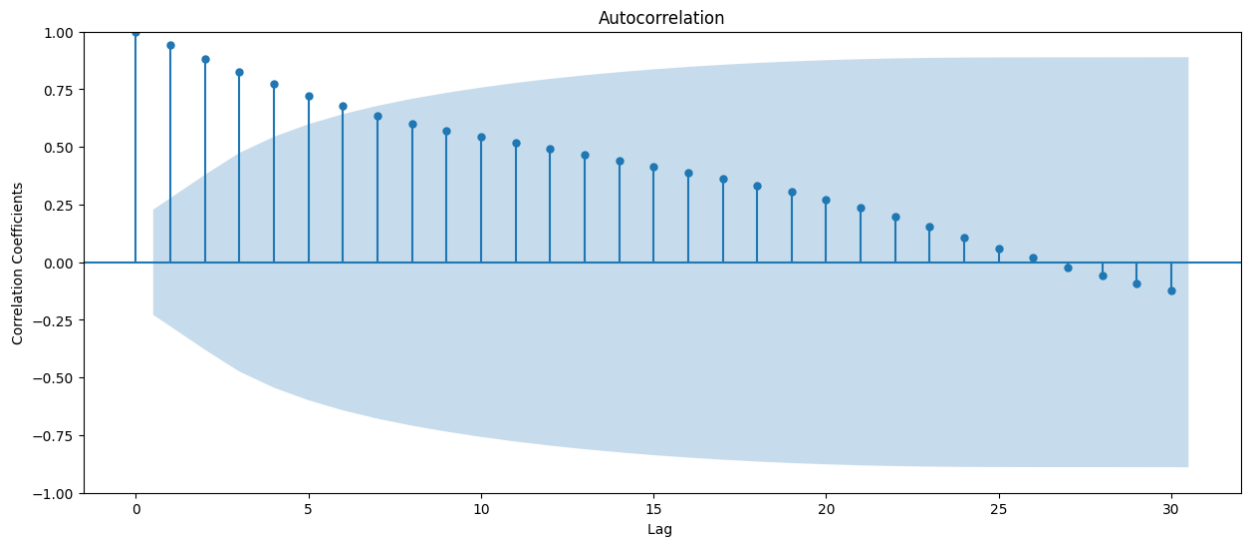


Рисунок 3.6 – АКФ для народжуваності

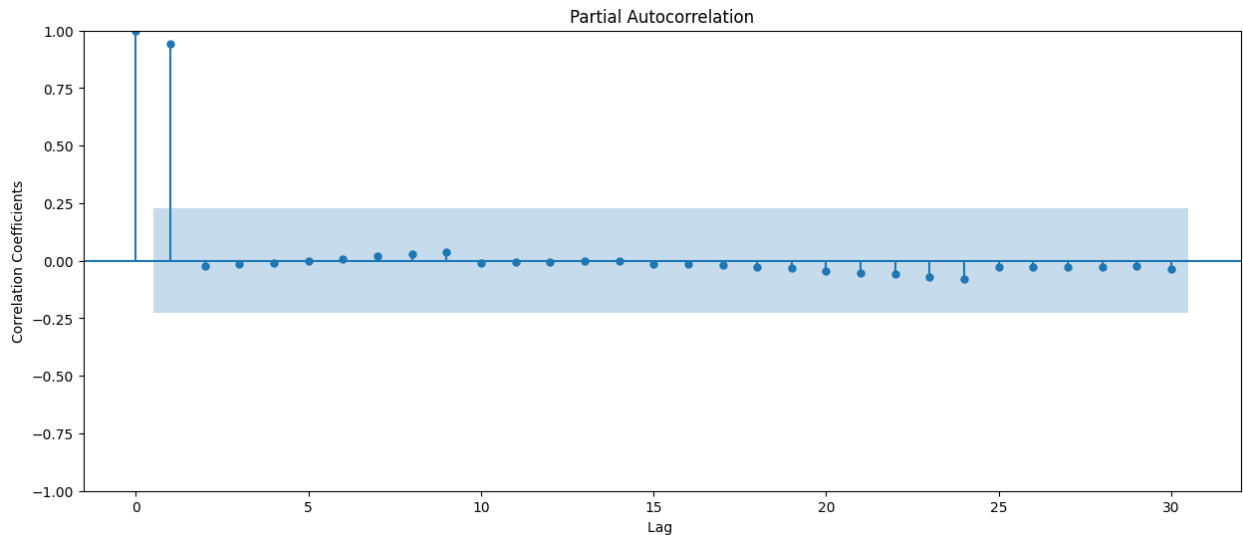


Рисунок 3.7 – ЧАКФ для народжуваності

Аналізуючи отримані функції було прийнято рішення будувати моделі авторегресії порядків 2 та 9:

$$AR(2): y(k) = 22,8578 + 1,9345y(k - 1) - 0,9532y(k - 2)$$

$$AR(9): y(k) = 22,793 + 1,8294y(k - 1) - 0,8216y(k - 2) - 0,0364y(k - 3) + 0,024y(k - 4) + 0,2099y(k - 5) - 0,3940y(k - 6) + 0,1725y(k - 7) - 0,007y(k - 8) - 0,0005y(k - 9)$$

Задля побудови моделі ARMA аналізуємо функції АКФ та ЧАКФ для залишків цих моделей (для AR(2) – рис. 3.8). та за допомогою програмного продукту підбираємо кращий порядок ковзного середнього (МА) на основі критеріїв, якими користувалися раніше:

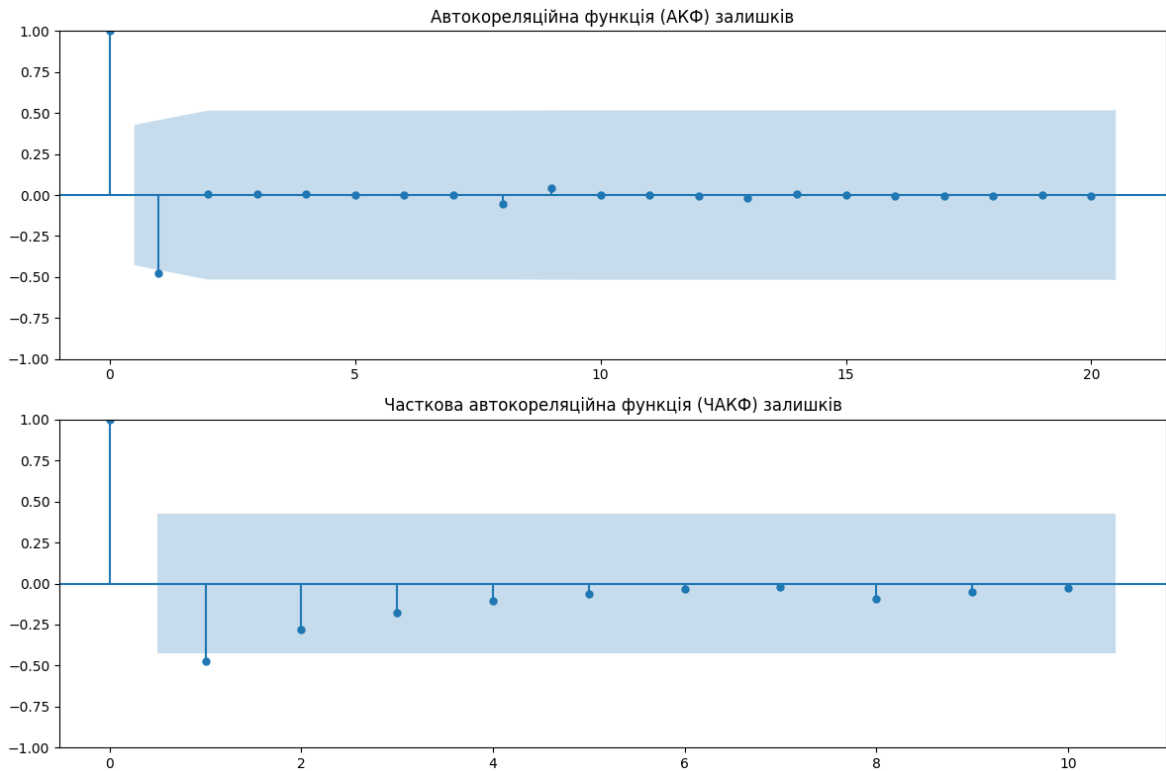


Рисунок 3.8 – АКФ та ЧАКФ для залишків AR(2)

Вирішено будувати модель авторегресії з ковзним середнім.

$$\text{ARMA}(2,8): y(k) = 22,3708 + 0,2872y(k-1) + 0,6183y(k-2) + 1,7280\varepsilon(k-1) + 1,7226\varepsilon(k-2) + 1,7706\varepsilon(k-3) + 1,7622\varepsilon(k-4) + 1,7491\varepsilon(k-5) + 1,7543\varepsilon(k-6) + 1,7054\varepsilon(k-7) + 0,9567\varepsilon(k-8)$$

Аналогічним чином підбираємо порядок моделі ковзного середнього та будуємо АКФ та ЧАКФ для AR(9) (рис. 3.9).

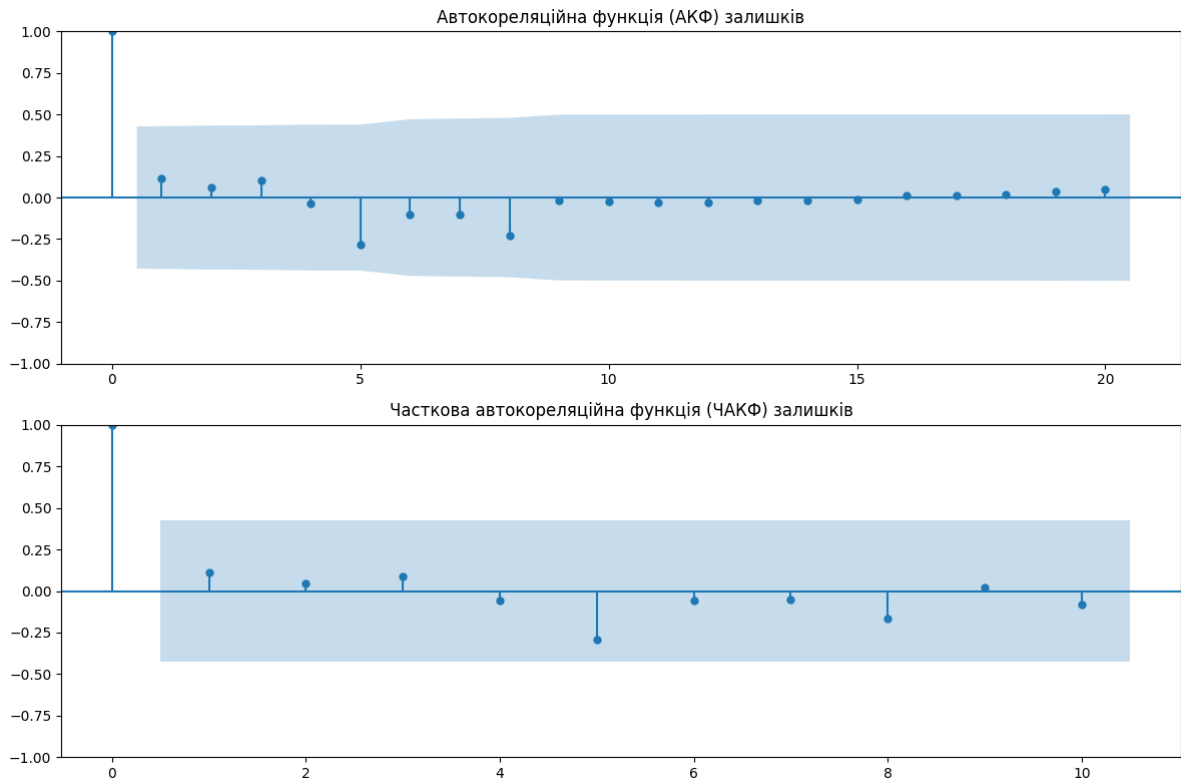


Рисунок 3.9 – АКФ та ЧАКФ для залишків AR(9)

Кращим порядком моделі ковзного середнього виявився MA(1).

$$\text{ARMA}(9,1): y(k) = 22,8803 + 0,864y(k-1) + 0,9444y(k-2) - 0,8169y(k-3) - 0,0191y(k-4) + 0,2501y(k-5) - 0,2065y(k-6) - 0,2181y(k-7) + 0,1629y(k-8) - 0,0005y(k-9) + 0,9605\varepsilon(k-1)$$

Для порівняння побудованих моделей та вибору кращої з них, зводимо таблицю з оцінками критеріїв адекватності (табл. 3.3):

Таблиця 3.3 – Порівняння моделей для народжуваності

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R^2	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	5,2285	0,3917	1,5856	0.8141	1.4523	0.0449
ARMA(2,8)	3,2291	0,6434	1,5652	1.0096	1.7371	0.05
AR(9)	2,8085	0,6732	1,5845	0.8031	1.3194	0.0447
ARMA(9,1)	0,4386	0,9489	1,5122	0.8353	1.3569	0.0456

Аналіз отриманих результатів показав, що кращою моделлю для народжуваності виявилася ARMA(9,1). Тому на основі цієї моделі будемо прогнозувати народжуваності на 4 роки (рис. 3.10):

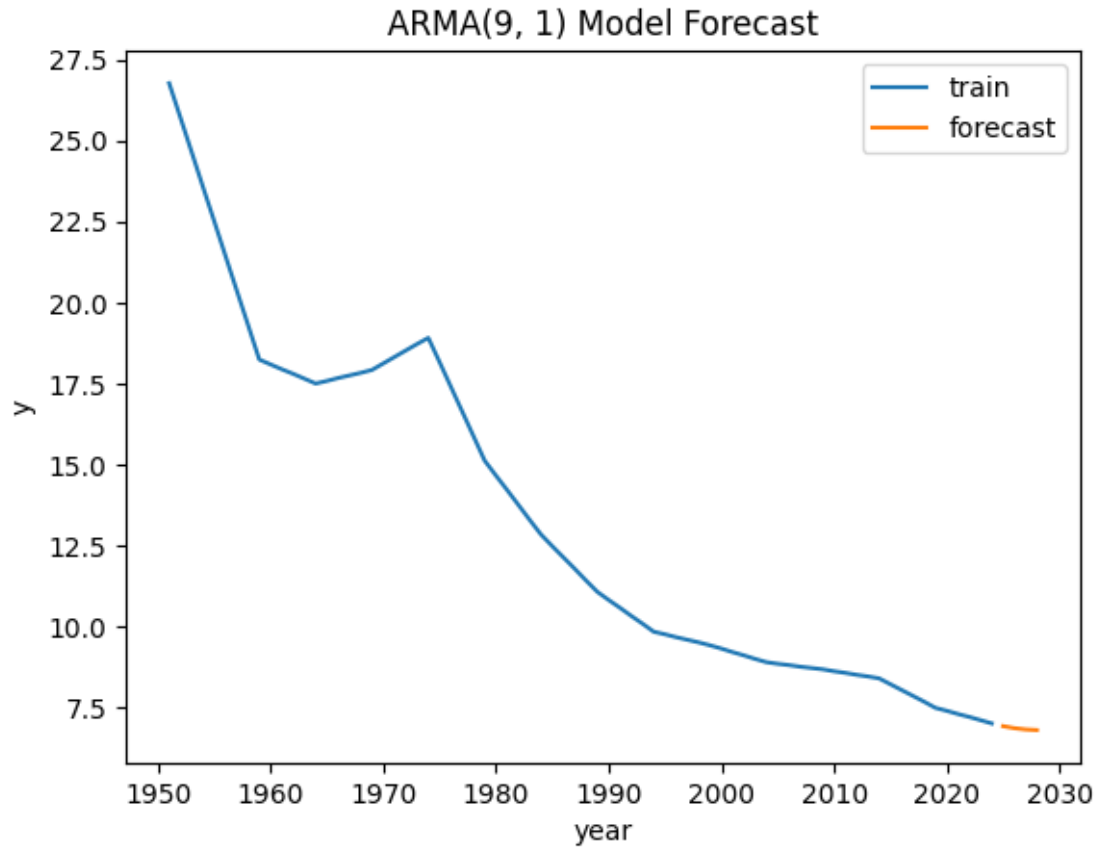


Рисунок 3.10 – Прогноз народжуваності на 4 роки

Прогнозовані значення показника народжуваності на наступні роки (табл. 3.4):

Таблиця 3.4 – Прогнозовані значення показнику народжуваності Японії

Рік	Показник народжуваності
2024	6.926020
2025	6.863779
2026	6.823790
2027	6.804682

Отже, як ми можемо бачити на графіку візуалізації змін показника народжуваності, з 1974 року показник народжуваності почав стрімко знижуватися. За нашим прогнозом він буде знижуватися й наступні 4 роки, але повільніше.

Якщо ця тенденція збережеться, це може мати серйозні наслідки для демографічної структури країни. Тому існує необхідність у впровадженні політичних заходів, спрямованих на стимулювання народжуваності. Заходи, спрямовані на поліпшення умов для сімейного життя, можуть допомогти зупинити або принаймні сповільнити цю тенденцію.

3.3.3 Смертність

Візуалізуємо часовий ряд для показника смертності (рис. 3.11):

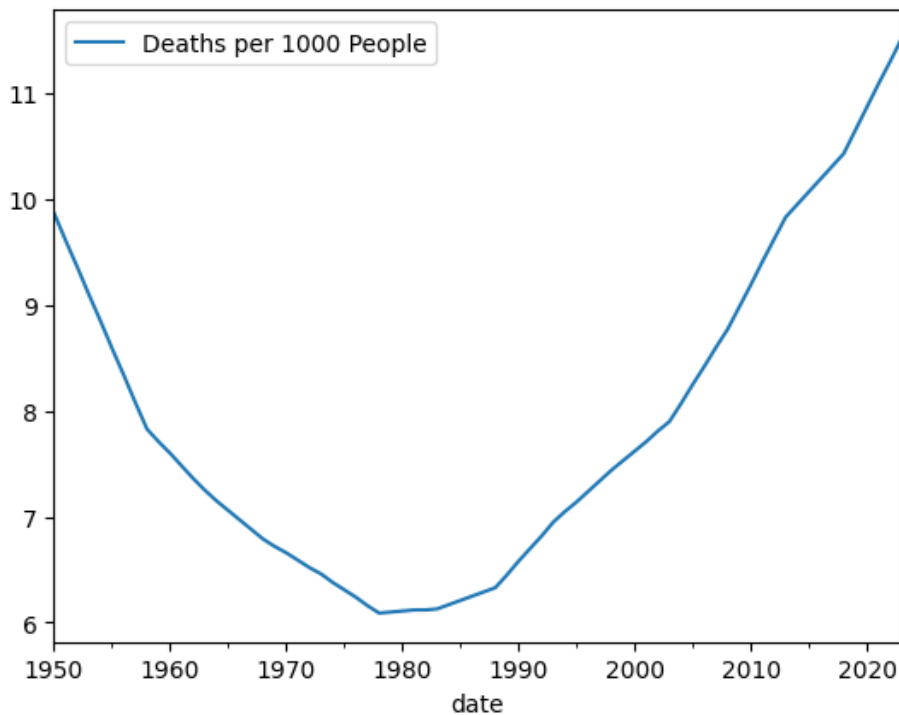


Рисунок 3.11 – Смертність Японії

Будуємо АКФ (рис 3.12) та ЧАКФ (рис 3.13):

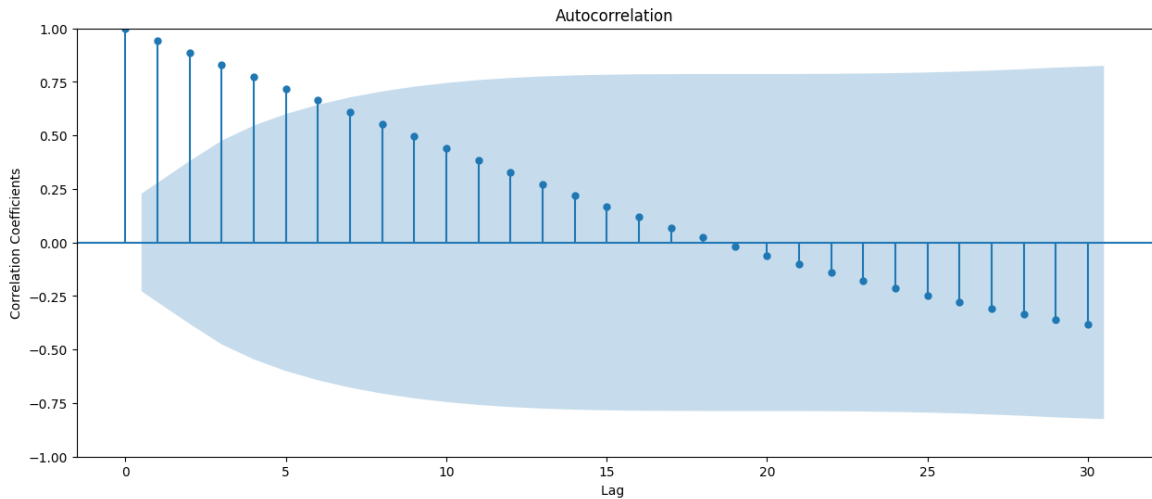


Рисунок 3.12 – АКФ для показника смертності

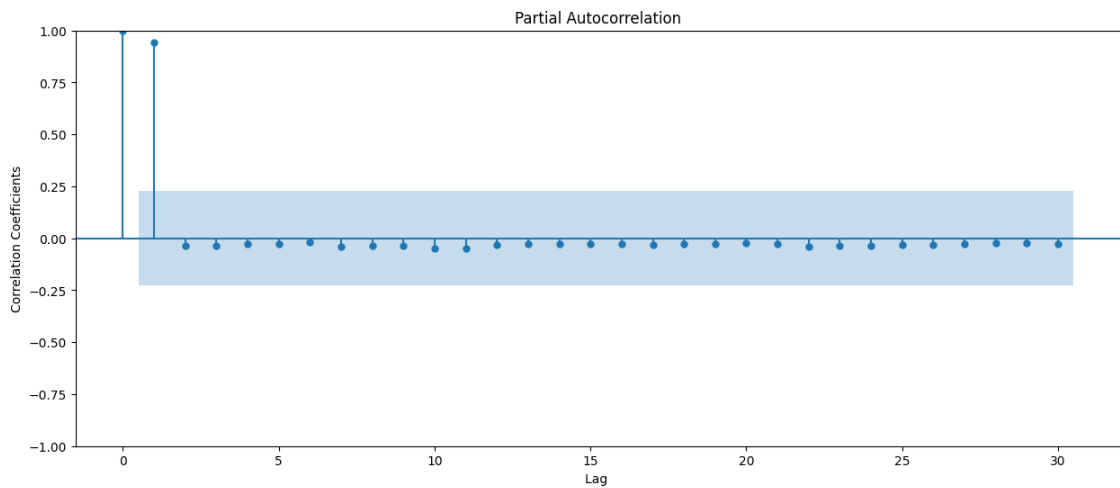


Рисунок 3.13 – ЧАКФ для показника смертності

На графіку часткової автокореляційної функції (ЧАКФ) видно, що значення значно відрізняються від нуля для перших двох лагів, а потім швидко падають до значень, що близькі до нуля. Це вказує на те, що найкращий порядок моделі авторегресії (AR) може бути 2. Також побудуємо AR(9), адже, в порівнянні з іншими порядками авторегресії, ця модель має найкращі значення критеріїв адекватності.

$$\text{Рівняння AR(2): } y(k) = 9,2012 + 1,9784y(k - 1) - 0,9848y(k - 2)$$

$$\begin{aligned} \text{Рівняння AR(9): } y(k) = & 8,7071 + 1,3852y(k - 1) + 0,3892y(k - 2) - \\ & 1,2681y(k - 3) + 0,8337y(k - 4) - 0,3191y(k - 5) - 0,3241y(k - 6) + \\ & 0,6529y(k - 7) - 0,5808y(k - 8) + 0,2162y(k - 9) \end{aligned}$$

На основі кожної з цих моделей будуюмо моделі ARMA, визначаючи порядок ковзного середнього по тому ж принципу. Будуюмо 15 моделей вигляду ARMA (9,1), ARMA (9,2), ARMA (9,3) ARMA (9,15) та 15 моделей вигляду ARMA (2,1), ARMA (2,2), ARMA (2,3) ARMA (2,15).

При оцінці найкращими виявилися моделі ARMA (9,1) та ARMA (2,1), рівняннями яких є:

$$\begin{aligned} \text{ARMA}(9,1): y(k) = & 9,0942 + 1,1902y(k-1) + 0,134y(k-2) - \\ & 0,3134y(k-3) + 0,4222y(k-4) - 0,4296y(k-5) - 0,1977y(k-6) - \\ & 0,0956y(k-7) + 0,6275y(k-8) - 0,3623y(k-9) + 0,9595\varepsilon(k-1) \end{aligned}$$

$$\begin{aligned} \text{ARMA}(2,1): y(k) = & 8,9855 + 1,9751y(k-1) - 0,9833y(k-2) + \\ & 0,0744\varepsilon(k-1) \end{aligned}$$

Краща модель обиралася шляхом порівняння критеріїв адекватності моделей-кандидатів (табл. 3.5)

Таблиця 3.5 – Порівняння моделей для смертності

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R ²	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	0,5299	0,9739	1,8141	0,0252	0,6309	0.02
AR(9)	1,4624	0,9281	1,417	0,0696	1,0731	0.0332
ARMA(2,1)	0,8643	0,9574	1,6274	0,0412	0,7485	0.0256
ARMA(9,1)	0,6889	0,9661	1,7081	0,0328	0,7674	0.0228

Хоча всі моделі показали непоганий результат, все таки кращою моделлю є AR(2), тому будуюмо прогноз, беручи її за основу (рис 3.14):

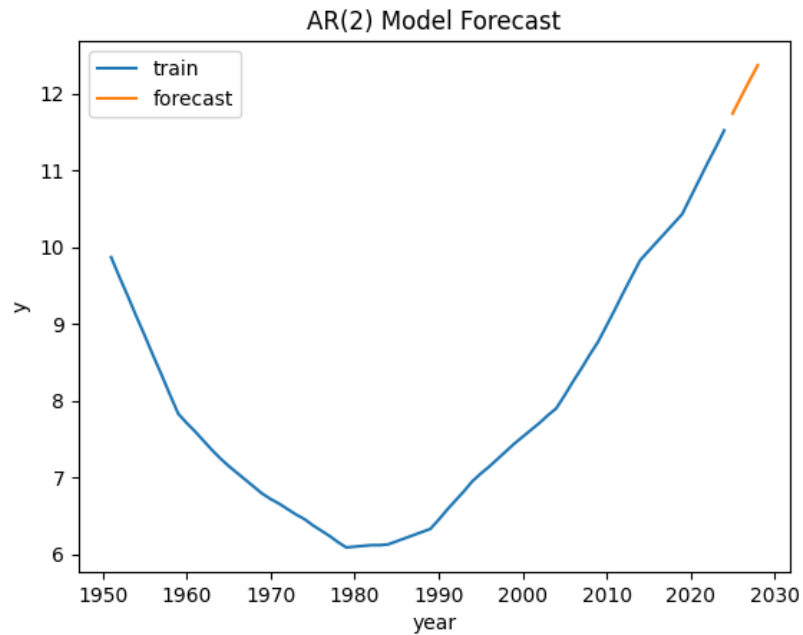


Рисунок 3.14 – Прогноз показника смертності на 4 роки

Результати прогнозування на наступні роки внесені до табл. 3.6.

Таблиця 3.6 – Прогнозовані значення показника смертності Японії

Рік	Показник смертності
2024	11.740077
2025	11.954634
2026	12.165080
2027	12.369402

Тенденція росту показника смертності починається приблизно з 1978 року. За нашим побудованим прогнозом це буде відбуватися й надалі. Збільшення смертності може бути наслідком старіння населення, що вимагає більш ефективних систем медичної допомоги та соціальної підтримки для літніх людей.

3.4. Побудова математичних моделей демографічних процесів Данії

На відміну від Японії, демографічна ситуація в Данії характеризується стабільним зростанням населення. Проте, для неї властивий відносно низький природний приріст. Високий рівень життя і стабільна економіка продовжують приваблювати іммігрантів, що підтримує демографічний баланс. У 2023 році до країни прибуло 79,561 іммігрантів. Данія стикається з низкою демографічних викликів, які можуть вплинути на її майбутній розвиток. Одним з таких викликів, як і для більшості розвинених країн, є старіння населення.

3.4.1 Кількість населення

Зобразимо динаміку змін в чисельності населення Данії з 1950 року (рис. 3.15)

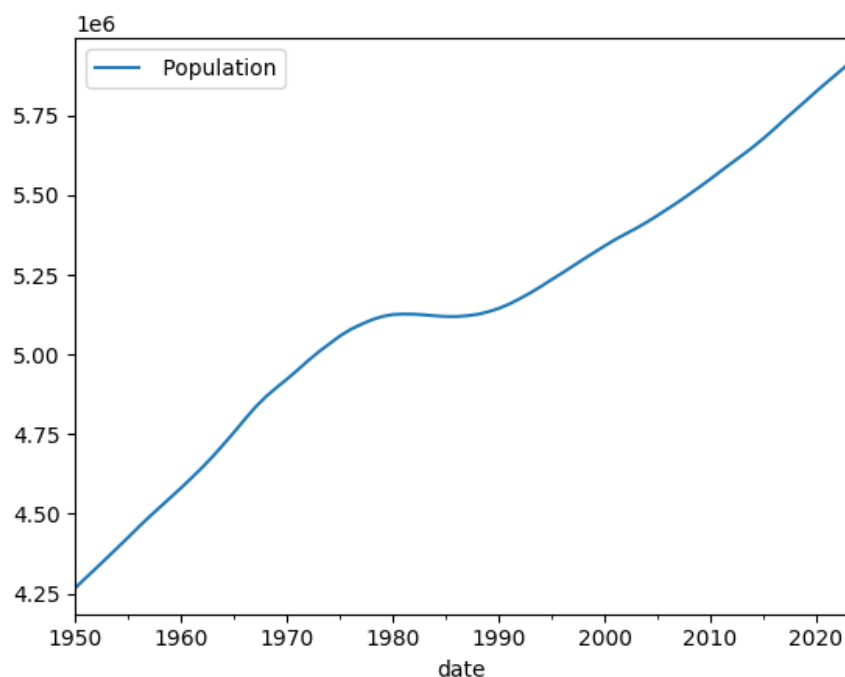


Рисунок 3.15 – Зміна кількості населення Данії в період 1950-2023рр.

Будуємо автокореляційну (АКФ) та часткову автокореляційну (ЧАКФ) функції з використанням 30 лагів для визначення порядку регресійних моделей (рис. 3.16 та рис. 3.17):

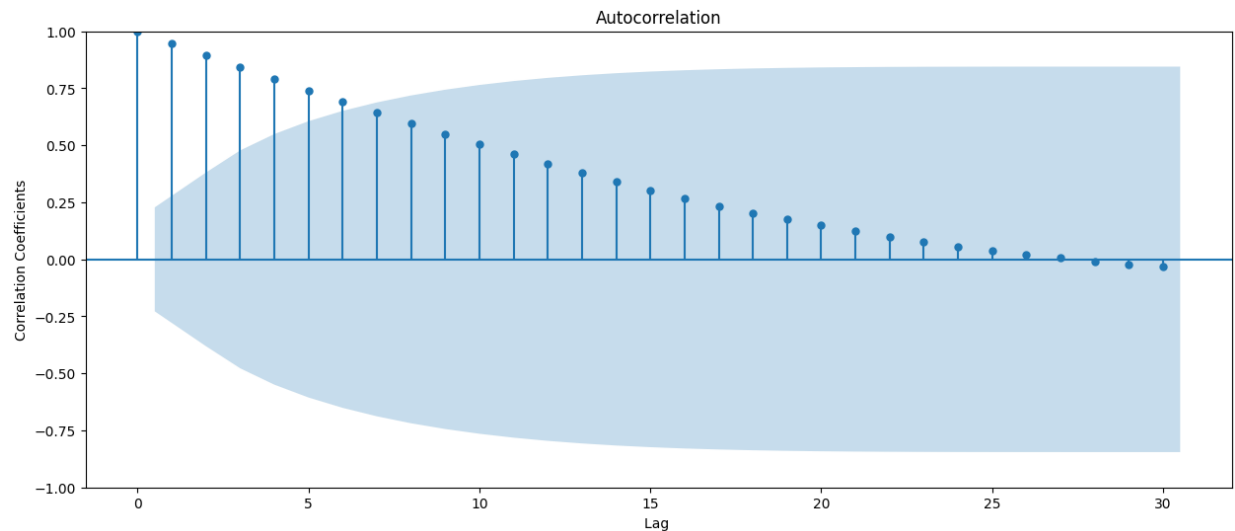


Рисунок 3.16 – АКФ для кількості населення Данії

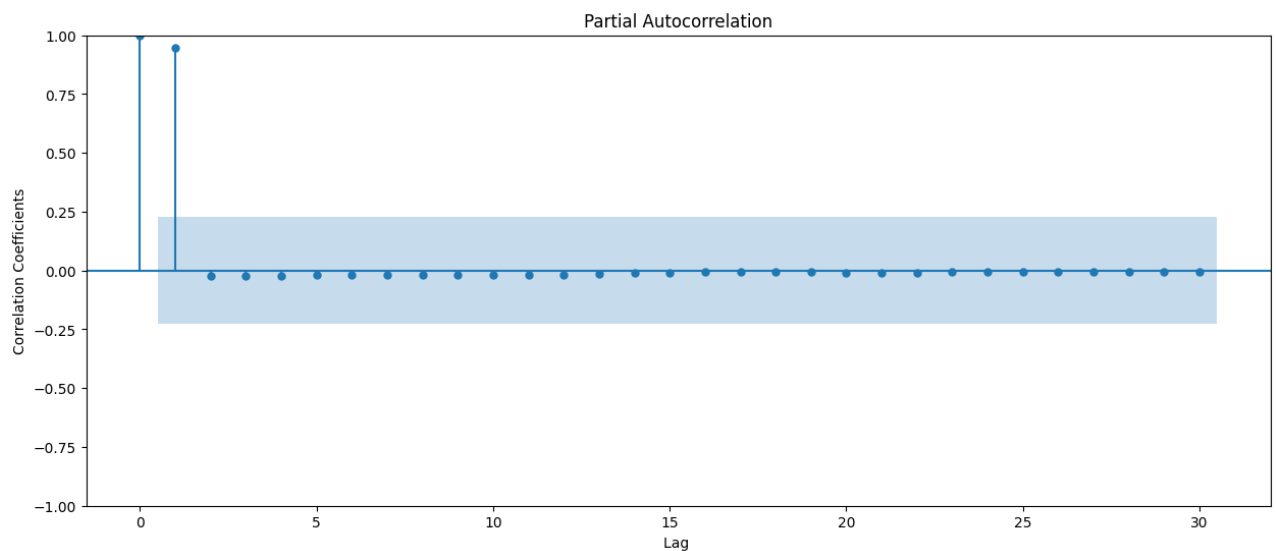


Рисунок 3.17 – ЧАКФ для кількості населення Данії

На графіку ЧАКФ спостерігаються значні автокореляції на перших лагах (0, 1, 2), а далі вони зменшуються і стають несуттєвими, тому можна обрати другий порядок для авторегресії – AR(2)

Рівняння AR(2): $y(k) = c + a_1y(k - 1) + a_2y(k - 2) = 4.591e^6 + 1.9931y(k - 1) - 0.9973y(k - 2)$

За допомогою розробленої програми було побудовано 15 моделей вигляду: ARMA(2,1), ARMA(2,2), ... ARMA(2,15). Цей підхід дозволяє визначити найкращий порядок для моделі ковзного середнього шляхом перебору та оцінки якості кожної моделі за допомогою різних критеріїв адекватності.

У нашому випадку це ARMA(2,1), рівнянням якої є:

$$y(k) = 4.591e^6 + 1.992y(k - 1) - 0.9956y(k - 2) + 0,996\varepsilon(k - 1)$$

Краща модель обиралася шляхом порівняння критеріїв адекватності моделей-кандидатів (табл. 3.7)

Таблиця 3.7 – Порівняння моделей для чисельності населення

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R^2	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	$10e^{11}$	0,73	2,45	$5e^9$	0,4247	0,0155
ARMA(2,1)	$5e^{11}$	0,88	2,3	$5e^9$	0,4164	0,0154

Для побудови прогнозу слід взяти другу модель (ARMA(2,1)). Результати критеріїв вказують на адекватність моделі, що добре пояснює дані без значних помилок або відхилень. За тестом Льюнга-Бокса можна стверджувати, немає суттєвої автокореляції залишків моделі. Тест Жарка-Бера вказує на нормальність розподілу залишків, адже немає підстав відхиляти нульову гіпотезу. Також можемо стверджувати про гомоскедастичність моделі (рис. 3.18).

Ljung-Box (L1) (Q):	0.24	Jarque-Bera (JB):	0.63
Prob(Q):	0.63	Prob(JB):	0.73
Heteroskedasticity (H):	2.68	Skew:	-0.19
Prob(H) (two-sided):	0.22	Kurtosis:	2.24

Рисунок 3.18 – Значення тестів для ARMA(2,1)

Візуалізуємо прогнозовані дані для чисельності населення (рис. 3.19).

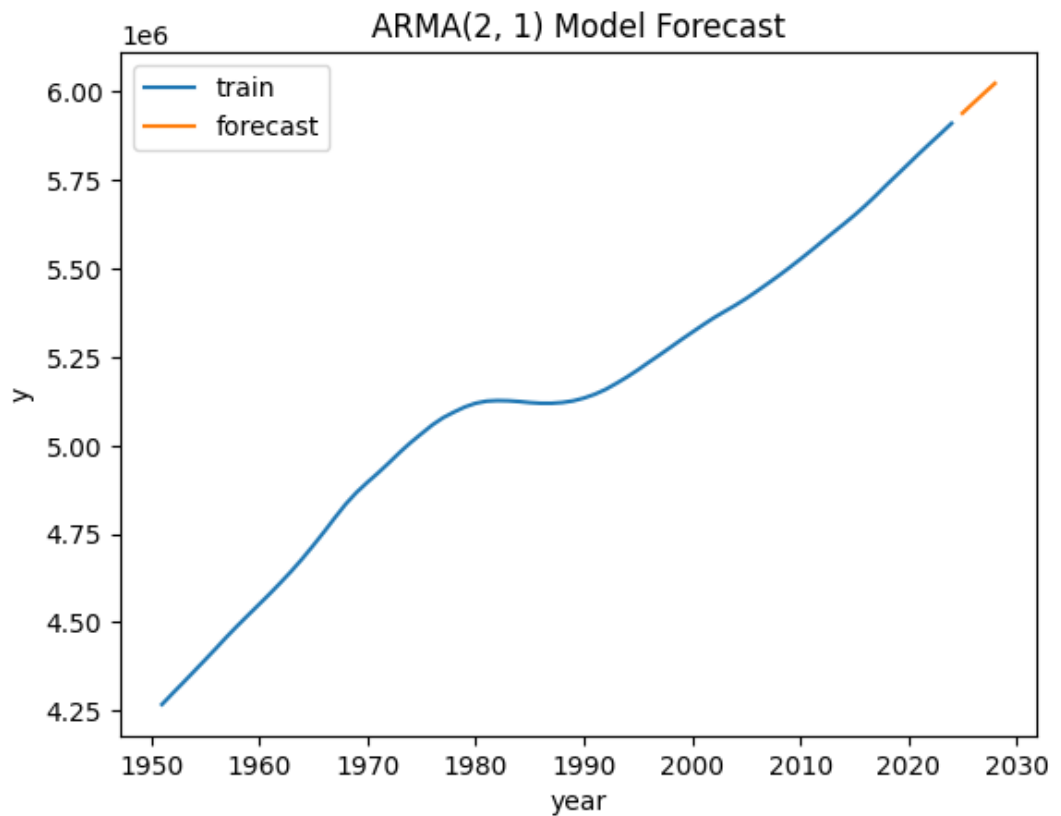


Рисунок 3.19 - Прогноз зміни кількості населення на 4 роки

Також зведемо отримані результати до таблиці 3.8.

Таблиця 3.8 – Прогнозовані значення чисельності населення Данії

Рік	Чисельність населення
2024	5939683
2025	5968092
2026	5996142
2027	6023834

Дивлячись на динаміку зміни чисельності населення, можемо спостерігати постійне збільшення цього показника. За нашим прогнозом ця тенденція збережеться й на наступні 4 роки і у 2027 році це число буде більше 6 мільйонів.

3.4.2 Народжуваність

Візуалізуємо часовий ряд для показника народжуваності Данії (рис. 3.20):

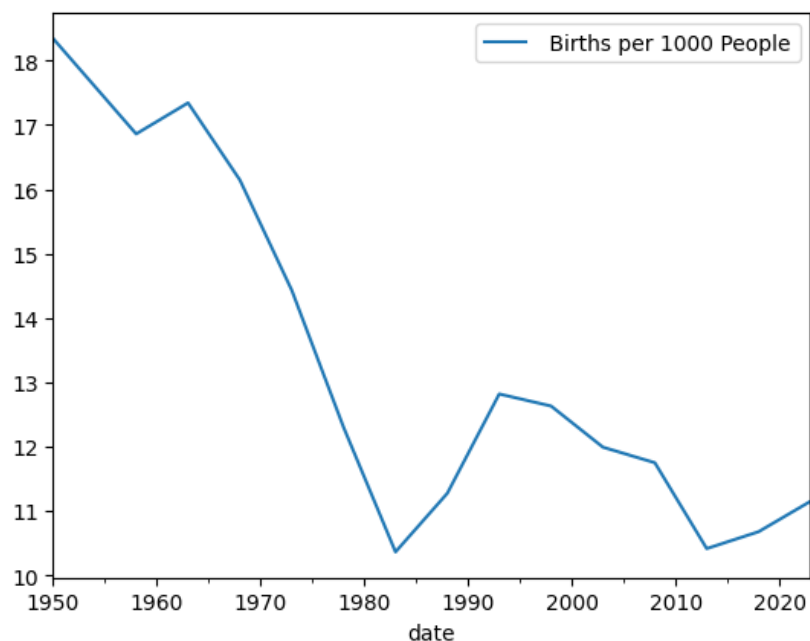


Рисунок 3.20 – Показник народжуваності Данії

Будуємо АКФ (рис 3.21) та ЧАКФ (рис 3.22):

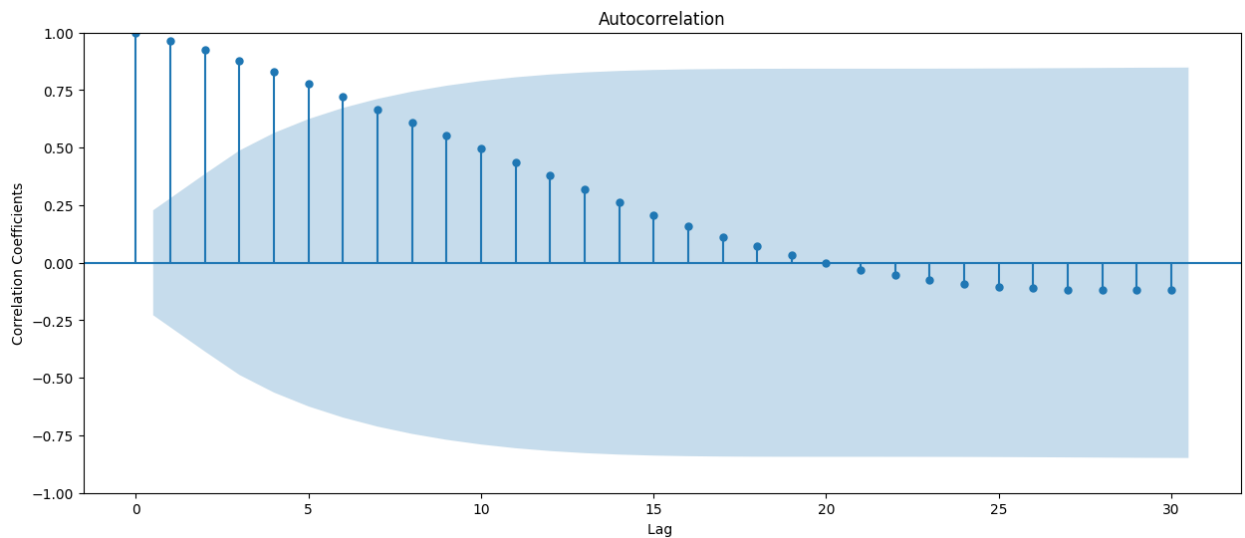


Рисунок 3.21 – АКФ для ряду народжуваності

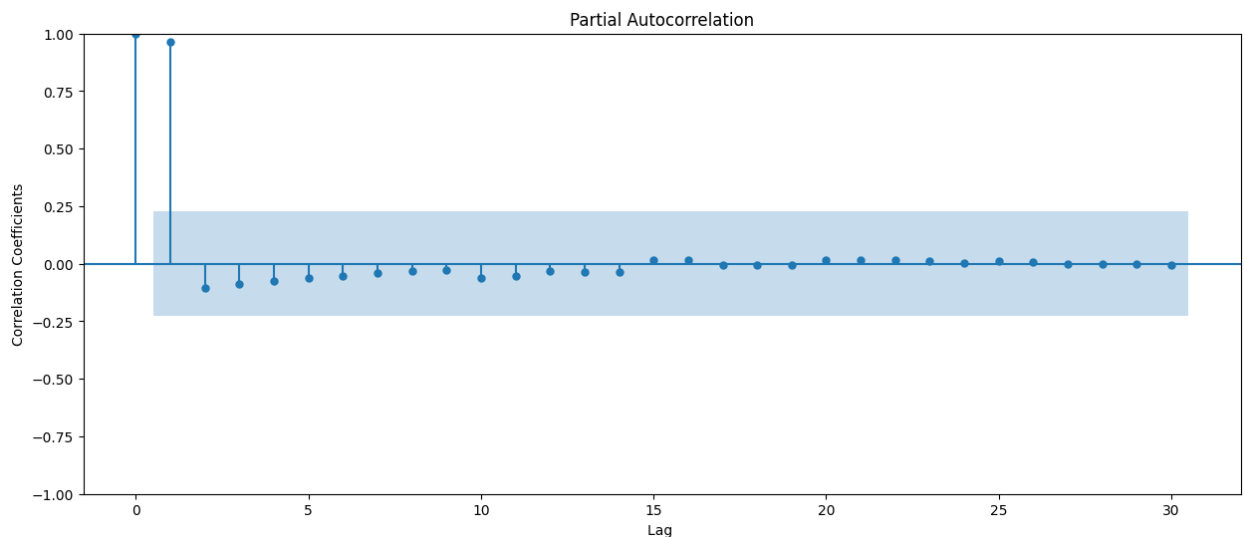


Рисунок 3.22 – ЧАКФ для ряду народжуваності

На графіку часткової автокореляційної функції (ЧАКФ) видно, що значення значно відрізняються від нуля для перших двох лагів, а потім швидко падають до значень, що близькі до нуля. Це вказує на те, що найкращий порядок моделі авторегресії (AR) може бути 2. Також побудуємо AR(4) та AR(6), адже, в порівнянні з іншими порядками авторегресії, ці моделі мають найкращі критерії адекватності.

$$\text{AR}(2): y(k) = 16.7868 + 1.8673y(k - 1) - 0.8869y(k - 2)$$

$$\text{AR}(4): y(k) = 16.8102 + 1.9284y(k - 1) - 0.9605y(k - 2) - 0.0525y(k - 3) + 0.067y(k - 4)$$

$$\text{AR}(6): y(k) = 16.7854 + 1.9108y(k - 1) - 0.9576y(k - 2) + 0.0693y(k - 3) - 0.0597y(k - 4) - 0.0579y(k - 5) + 0.0797y(k - 6)$$

Зробимо аналіз-порівняння отриманих моделей за допомогою таблиці 3.9.

Таблиця 3.9 – Порівняння моделей для показника народжуваності

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R^2	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	2,6476	0,72	2,36	0,1261	0,7367	0,0208
AR(4)	2,5733	0,83	2,23	0,12254	0,7242	0,0205
AR(6)	2,6459	0,98	1,94	0,126	0,721	0,0207

Всі обрані моделі авторегресії мають подібні оцінки якостей прогнозів. Кращою моделю, за результатами критеріїв адекватності, є AR(6). Зобразимо графіки АКФ та ЧАКФ залишків для цієї моделі (рис. 3.23).

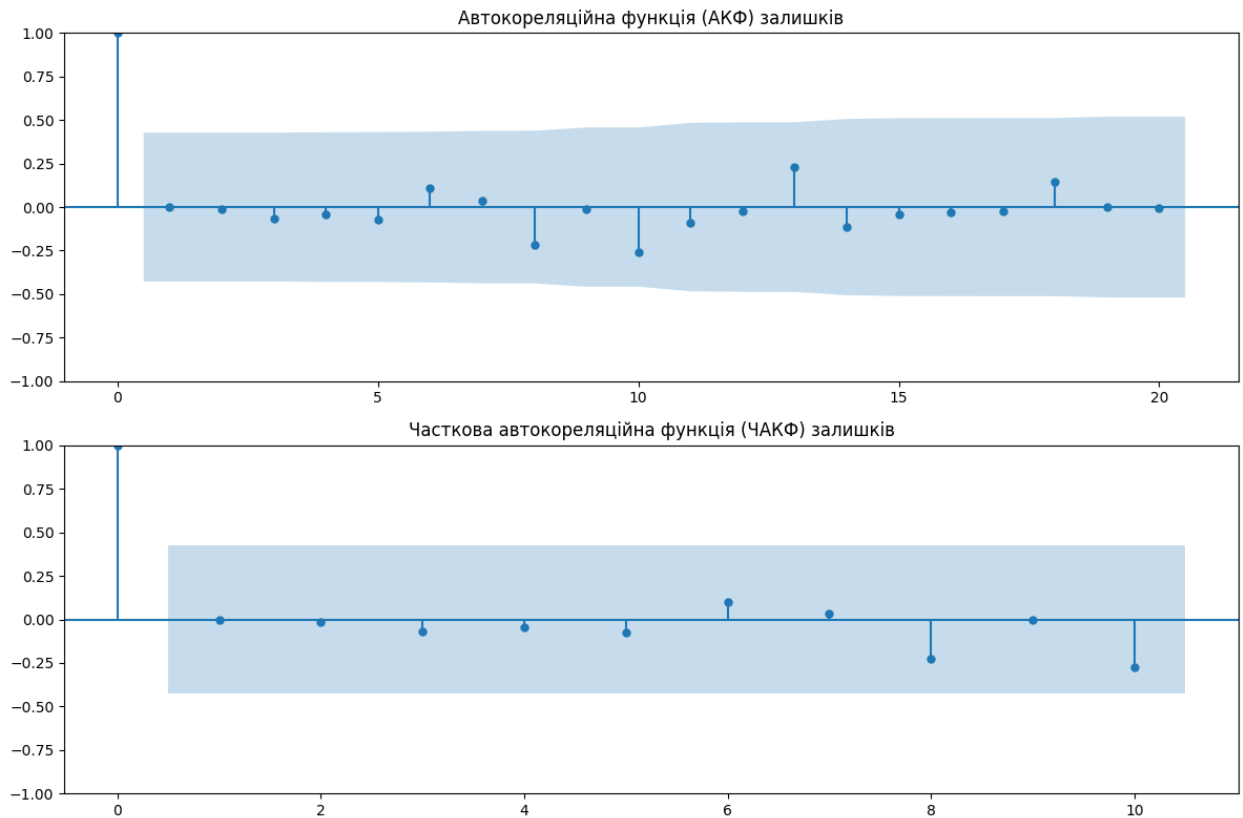


Рисунок 3.23 – АКФ та ЧАКФ для залишків AR(6)

На графіках всі значення знаходяться в межах області довіри (синя зона). Це свідчить про відсутність автокореляції в залишках. На основі цієї моделі можна побудувати модель ARMA. Проте, при виборі порядку ковзного середнього, спираючись на критерії адекватності, виявилось, що модель AR(6) має значно кращі характеристики. Тому будемо прогнозувати за цією моделлю (рис. 3.24).

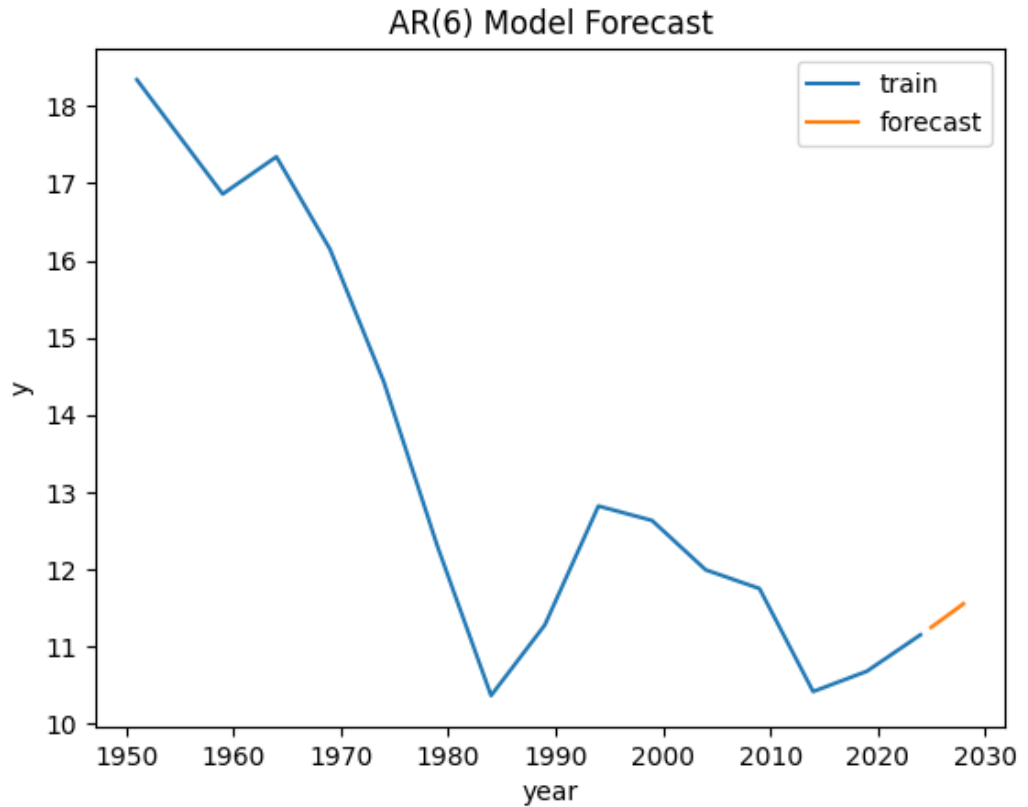


Рисунок 3.24 – Прогноз показника народжуваності на 4 роки

Отримали такі результати на наступні 4 роки у таблиці 3.10.

Таблиця 3.10 – Прогнозовані значення показника народжуваності Данії

Рік	Показник народжуваності
2024	11.246514
2025	11.347160
2026	11.448862
2027	11.550992

Отже, як ми можемо бачити на графіку візуалізації змін показника народжуваності, з 1964 року показник народжуваності почав стрімко

знижуватися, потім до 1993 року він зростає, до 2014 року знижувався, а потім почав плавно зростати. За нашим прогнозом він буде зростати й наступні 4 роки, але цей значення цього показника значно менші, ніж в 50х роках.

Хоч і у випадку Данії показник народжуваності зростає, але цей процес відбувається повільно. Його можна покращити шляхом впровадження політичних заходів, спрямованих на стимулювання народжуваності.

3.4.3 Смертність

Візуалізуємо часовий ряд для показника смертності Данії (рис. 3.25):

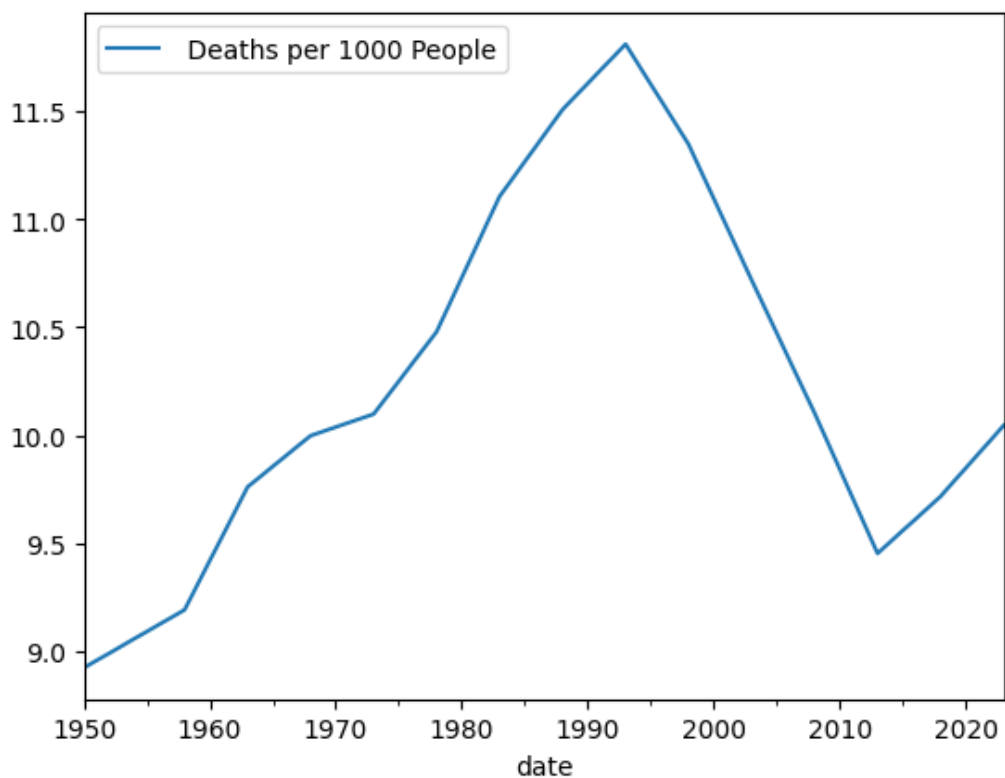


Рисунок 3.25 – Смертність Данії

Будуємо АКФ (рис 3.26) та ЧАКФ (рис 3.27):

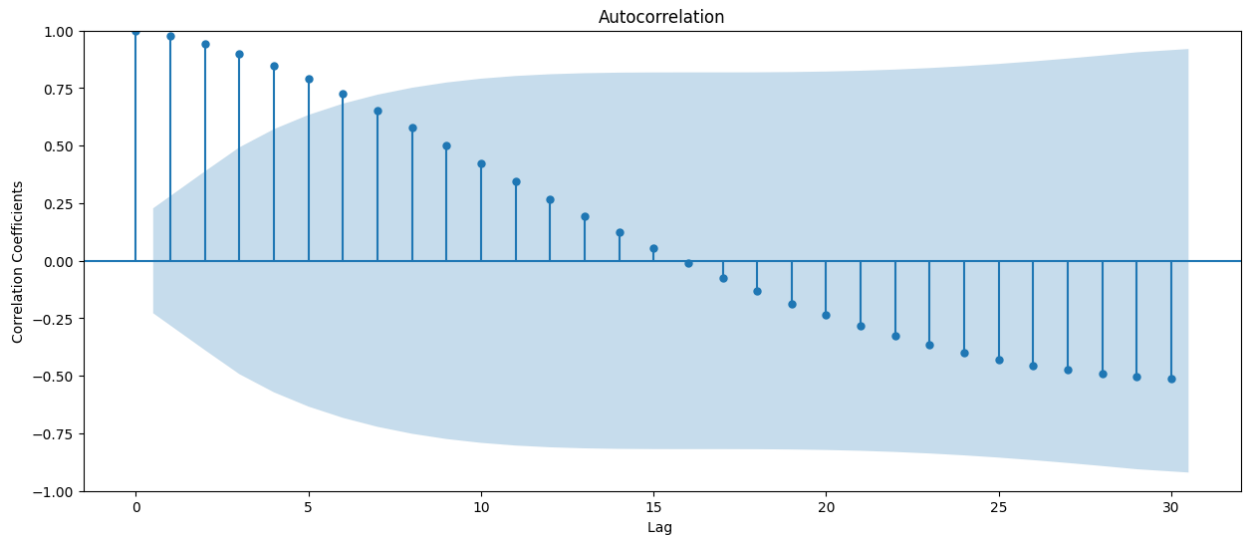


Рисунок 3.26 – АКФ для показника смертності

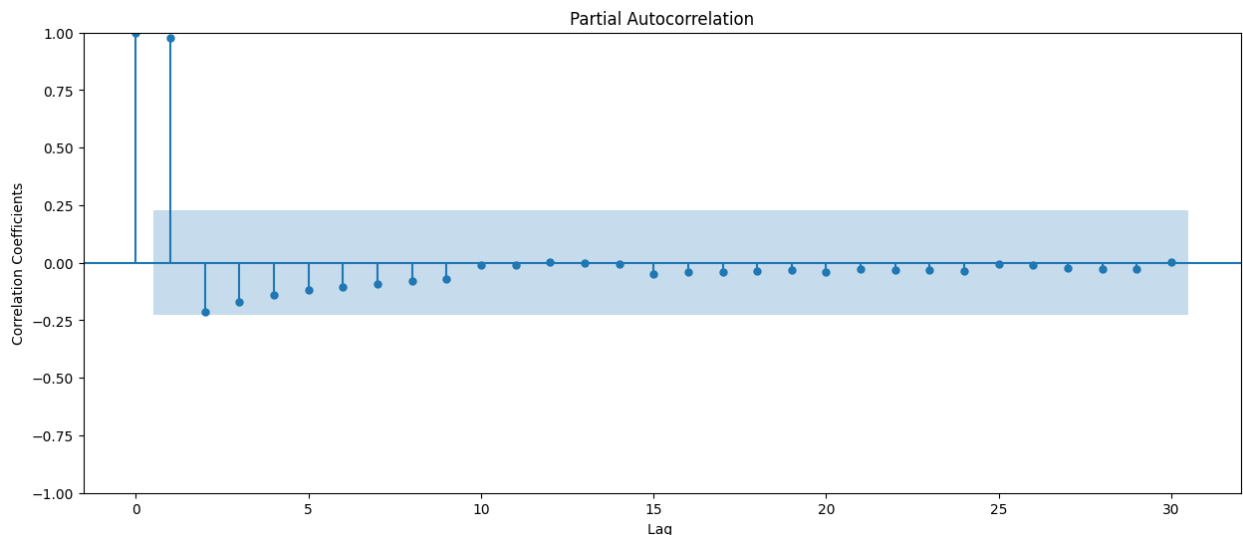


Рисунок 3.27 – ЧАКФ для показника смертності

Найкращий порядок авторегресійної моделі, зважаючи на графік ЧАКФ, буде 2 (AR(2)), оскільки тільки перші два лаги показують значущу автокореляцію.

Рівняння AR(2): $y(k) = c + a_1y(k - 1) + a_2y(k - 2) = 9.441 + 1.9079y(k - 1) - 0.9252y(k - 2)$

Також побудуємо модель AR(8), адже, в порівнянні з моделями інших порядків, вона має найкращі значення критеріїв адекватності.

$$AR(8): y(k) = 9.4537 + 1.0148y(k - 1) + 0,6626y(k - 2) - 0,9413y(k - 3) + 0.3796y(k - 4) - 0.5299y(k - 5) + 0.1898y(k - 6) - 0,7745y(k - 7) - 0,6705y(k - 8)$$

Порівняємо отримані моделі між собою за допомогою тих критеріїв адекватності та якості прогнозу (табл. 3.11):

Таблиця 3.11 – Порівняння моделей для показника смертності

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R^2	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	0,1116	0,69	2,38	0,0139	0,5011	0,0124
AR(8)	0,0267	0,83	1,93	0,0133	0,4246	0,0122

Кращі значення характеристик має модель AR(8). Візуалізуємо графіки АКФ та ЧАКФ лишків для цієї моделі (рис. 3.28)

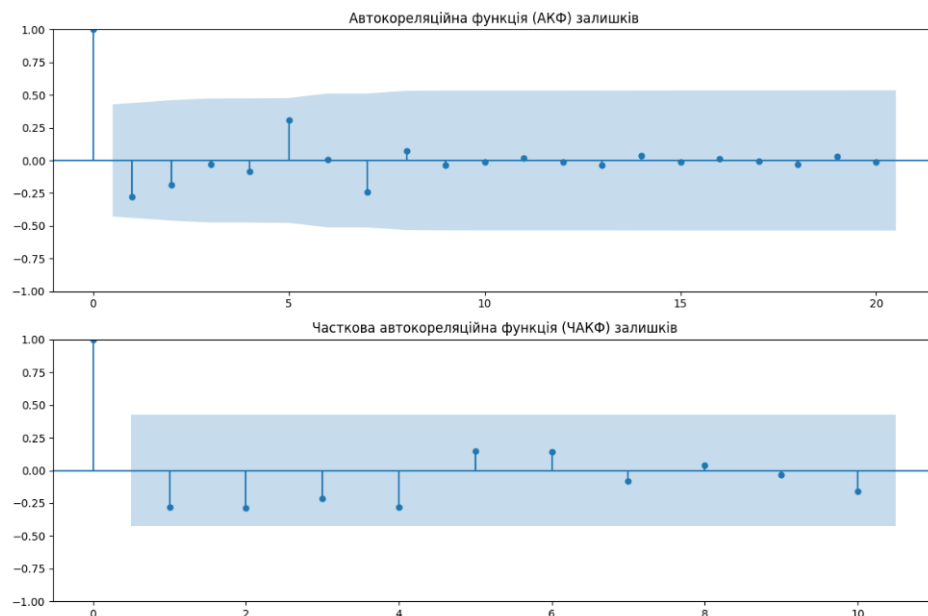


Рисунок 3.28 – АКФ та ЧАКФ залишків AR(8)

На графіках всі значення знаходяться в межах області довіри (синя зона). Це свідчить про відсутність автокореляції в залишках, що є гарним

показником. Можемо вважати, що залишки є білим шумом. Тому на основі саме цієї моделі будемо прогнозувати на 4 роки (рис. 3.29).

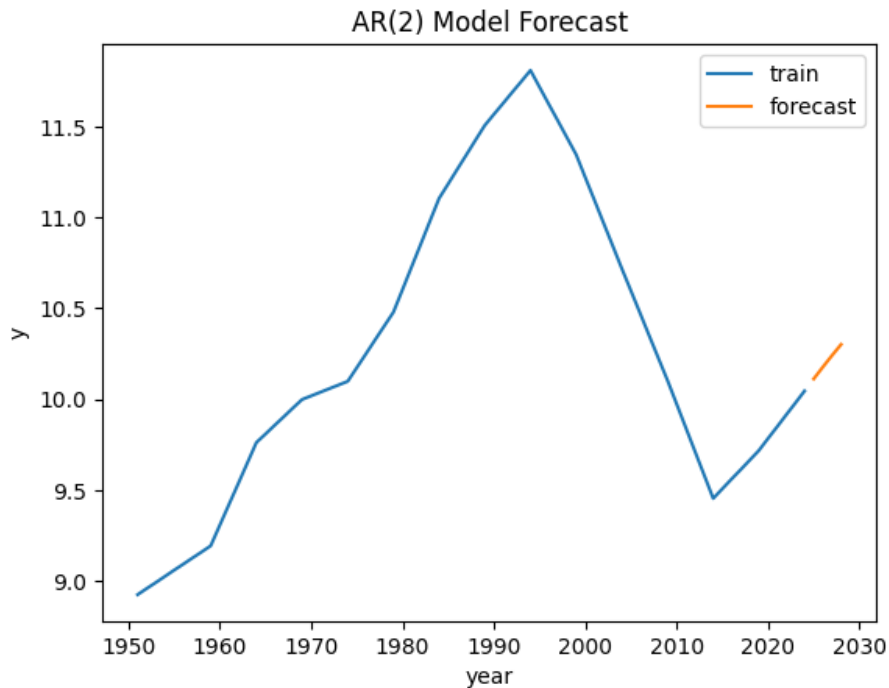


Рисунок 3.29 – Прогноз показника смертності Данії

Отримали такі результати на наступні роки у таблиці 3.12.

Таблиця 3.12 – Прогнозовані значення показника смертності Данії

Рік	Показник смертності
2024	10.111406
2025	10.176636
2026	10.240073
2027	10.301111

Тенденція росту показника смертності починається приблизно з 2012 року. За побудованим прогнозом це буде відбуватися й надалі. Збільшення смертності може бути наслідком старіння населення, що вимагає більш

ефективних систем медичної допомоги та соціальної підтримки для літніх людей.

3.5. Висновки до розділу 3

В даному розділі було здійснено побудову математичних моделей для аналізу демографічних процесів на прикладі Японії та Данії. Було проаналізовано дані про населення, народжуваність та смертність, використовуючи дані з веб-ресурсу Macrotrends. За допомогою розробки власного програмного продукту, написаного на мові програмування Python та програмного середовища Google Colab було побудовано моделі та прогнози.

Аналізуючи вхідні дані, було встановлено, що Японія стикається з серйозними демографічними викликами, такими як зниження народжуваності, зростання старшого віку населення та зменшення приросту населення.

Для побудови моделей були обрані статистичні методи, зокрема авторегресійні моделі, кожна з яких була протестована на адекватність та точність прогнозів. Проведений аналіз показав, що використані моделі здатні адекватно відображати демографічні процеси та надавати досить точні прогнози. Судячи з результатів отриманих прогнозів, кількість населення та показник народжуваності будуть й надалі зменшуватися, в той час як показник смертності буде зростати. Наслідком цього є зменшення природного приросту, що має значний вплив на демографію країни. Старіння нації призводить до зменшення кількості працездатного населення, що досить негативно відображається на економіці країни.

У випадку Данії, маємо протилежну ситуацію. Данія, хоча і не має таких виражених демографічних проблем, як Японія, але також стикається з певними викликами. Зростання середнього віку населення та стабільне, але не дуже високе зростання народжуваності можуть впливати на економічний розвиток та соціальну політику країни. Для побудови моделей демографічних

процесів у Данії були використані ті ж самі статистичні методи, що й для Японії. Аналіз показав, що, хоча демографічні зміни в Данії не настільки драматичні, все ж є тенденції до старіння населення та певної стабільності у народжуваності, що вимагає уваги до соціальних та економічних політик.

Для пом'якшення цих негативних наслідків та підтримки стабільного демографічного розвитку країни, потрібно аби були вжиті заходи з боку держави. Наприклад, стимулювання народжуваності, зокрема введення фінансових стимулів для сімей з дітьми, таких як податкові пільги, матеріальна допомога при народженні дитини, субсидії на дитячі садки та освіту, забезпечення гнучких умов праці для батьків, що дозволить їм краще поєднувати роботу з сімейними обов'язками. Щодо людей старшого віку, це може бути розробка програм для підвищення якості життя літніх людей, зокрема покращення медичних послуг, створення програм соціальної підтримки, пільги та підвищення пенсії. Для збільшення кількості населення та заповнення дефіциту робочої сили доцільно розробляти програми для залучення кваліфікованих іммігрантів.

Такі заходи допоможуть стабілізувати демографічну ситуацію, підтримати економічне зростання та забезпечити сталий розвиток суспільства у довгостроковій перспективі.

РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

В заданому розділі буде проведено оцінювання основних характеристик для майбутнього програмного продукту, що спеціалізується на дослідженні демографічного стану.

Дана реалізація буде сприяти проведенню усіх необхідних досліджень, що дасть змогу якісно дослідити питання не лише в Україні, проте у всьому світі.

Також в даному дослідженні показано різні варіанти реалізації для забезпечення найбільш коректної та оптимальної стратегії вибору, що має вплив на економічні фактори та сумісність з майбутнім програмним продуктом. Для цього застосовувався апарат функціонально-вартісного аналізу.

Функціонально-вартісний аналіз (ФВА) передбачає собою технологію, що дозволяє оцінити реальну вартість продукту або послуги незалежно від організаційної структури компанії. ФВА проводиться з метою виявлення резервів зниження витрат за рахунок ефективніших варіантів виробництва, кращого співвідношення між споживчою вартістю виробу та витратами на його виготовлення. Для проведення аналізу використовується економічна, технічна та конструкторська інформація.

Алгоритм функціонально-вартісного аналізу включає в себе визначення послідовності етапів розробки продукту, визначення повних витрат (річних) та кількості робочих часів, визначення джерел витрат та кінцевий розрахунок вартості програмного продукту.

4.1 Постановка задачі проектування

У роботі застосовується метод ФВА для проведення техніко-економічного аналізу розробки системи прогнозу стійкості фінансових показників. Оскільки рішення стосовно проектування та реалізації компонентів, що розробляється, впливають на всю систему, кожна окрема підсистема має її задовольняти. Тому фактичний аналіз представляє собою аналіз функцій програмного продукту, призначеного для збору, обробки та проведення аналізу даних по компанії.

Технічні вимоги до програмного продукту є наступними:

- функціонування на персональних комп'ютерах із стандартним набором компонентів;
- зручність та зрозумілість для користувача;
- швидкість обробки даних та доступ до інформації в реальному часі;
- можливість зручного масштабування та обслуговування;
- мінімальні витрати на впровадження програмного продукту.

4.2 Обґрунтування функцій програмного продукту

Головна функція F_0 – розробка можливого програмного продукту, яка дозволяє аналізувати різні характеристики, що безпосередньо впливають на стійкість підприємства. Беручи за основу цю функцію, можна виділити наступні:

F_1 – вибір самої програми.

F_2 – якісний аналіз даних.

F_3 – графічні показники.

Кожна з цих функцій має декілька варіантів реалізації:

Функція F_1 :

а) Eviews.

б) Python.

Функція F_2 .

а) Застосування вбудованих функцій.

б) Використання бібліотек.

Функція F_3 :

а) Використання шаблонних графіків.

б) Використання бібліотек.

Варіанти реалізації основних функцій наведені у морфологічній карті системи (рис. 4.1).

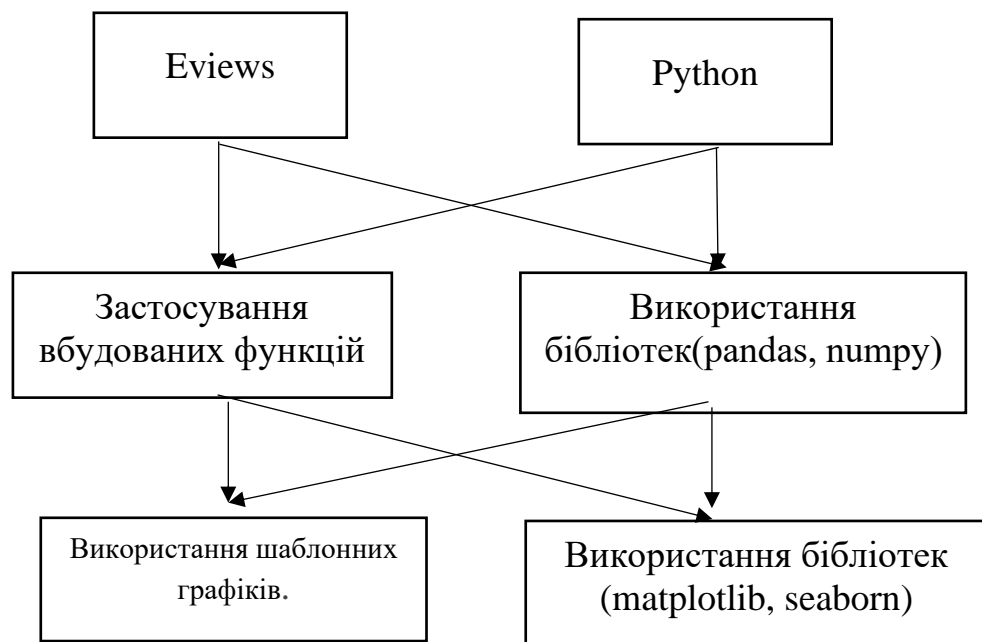


Рисунок 4.1 – Морфологічна карта

Морфологічна карта відображає множину всіх можливих варіантів основних функцій. Позитивно-негативна матриця показана в таблиці 4.1.

Таблиця 4.1 - Позитивно-негативна матриця

Функції	Варіанти реалізації	Переваги	Недоліки
F_1	<i>A</i>	Загальнодоступна програма, доступність багатьох бібліотек	Необхідність повної реалізації алгоритму
	<i>B</i>	Доступна в реалізації програма для різних обчислень	На написання коду необхідно мати базові навички та вміння
F_2	<i>A</i>	Доступність та легкість при написанні	Іноді не відповідає задачі яку треба розв'язати
	<i>B</i>	Ідеально описують усі необхідні характеристики	Достатньо затратно реалізовувати свої алгоритми для подальшої реалізації
F_3	<i>A</i>	Загально прийнята реалізація	Іноді не відповідає очікуваним значенням
	<i>B</i>	При виконанні власних досліджень краще може передавати висновки	Необхідно достатньо багато часу для написання програми для побудови та знаходження всього необхідного в задачі.

На основі аналізу позитивно-негативної матриці робимо висновок, що при розробці програмного продукту деякі варіанти реалізації функцій варто відкинути, тому, що вони не відповідають поставленим перед програмним продуктом задачам. Ці варіанти відзначені у морфологічній карті.

Функція F_1 :

Перевагу даємо загальнодоступності. Для спрощення роботи по написанню коду варіант Б має бути відкинтий.

Функція F_2 :

Програма допускає обрання обох варіантів. Можливо використати варіанти А чи Б.

Функція F_3 :

Реалізація першого варіанту є сприйнятливою для програми. Це варіант А.

Таким чином, будемо розглядати такі варіанти реалізації ПП:

$$F_{1a} - F_{2a} - F_{3a}$$

$$F_{1a} - F_{2б} - F_{3a}$$

Для оцінювання якості розглянутих функцій обрана система параметрів, описана нижче.

4.3 Обґрунтування системи параметрів програмного продукту

На основі даних, розглянутих вище, визначаються основні параметри вибору, які будуть використані для розрахунку коефіцієнта технічного рівня. Для того, щоб охарактеризувати програмний продукт, будемо використовувати наступні параметри:

- X_1 – швидкодія мови програмування;
- X_2 – об’єм пам’яті для обчислень та збереження даних;
- X_3 – час навчання даних;
- X_4 – потенційний об’єм програмного коду.

Гірші, середні і кращі значення параметрів вибираються на основі вимог замовника й умов, що характеризують експлуатацію програмного продукту, як показано у таблиці 4.2.

Таблиця 4.2 – Основні параметри програмного продукту

Назва Параметра	Умовні позначе ння	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія мови програмування	X1	оп/мс	60	80	110
Об'єм пам'яті	X2	Мб	60	50	30
Час попередньої обробки даних	X3	мс	80	70	60
Потенційний об'єм програмного коду	X4	кількість рядків коду	35	25	20

За даними таблиці 4.2 будуються графічні характеристики параметрів – рис. 4.2 – рис. 4.5.

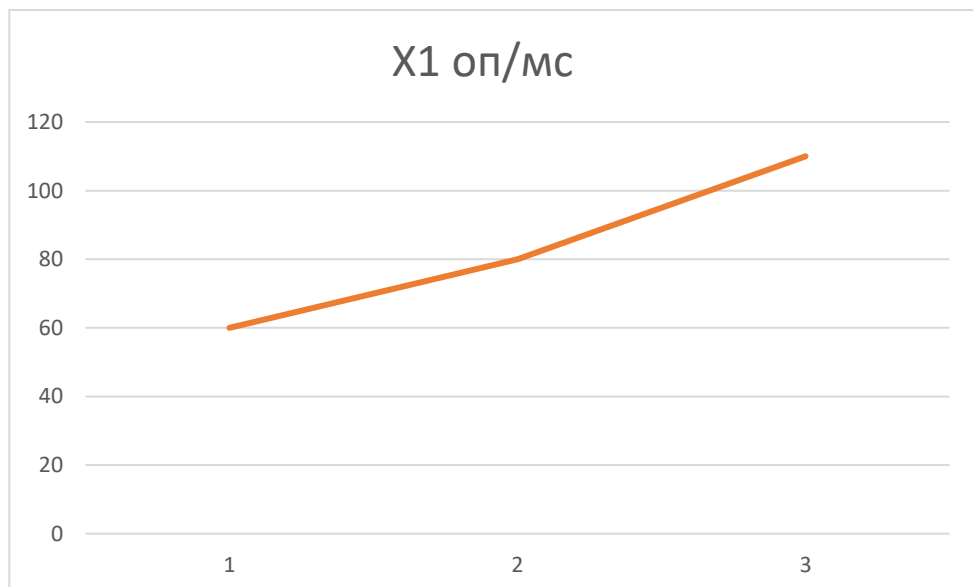


Рисунок 4.2 – X1, швидкодія мови програмування

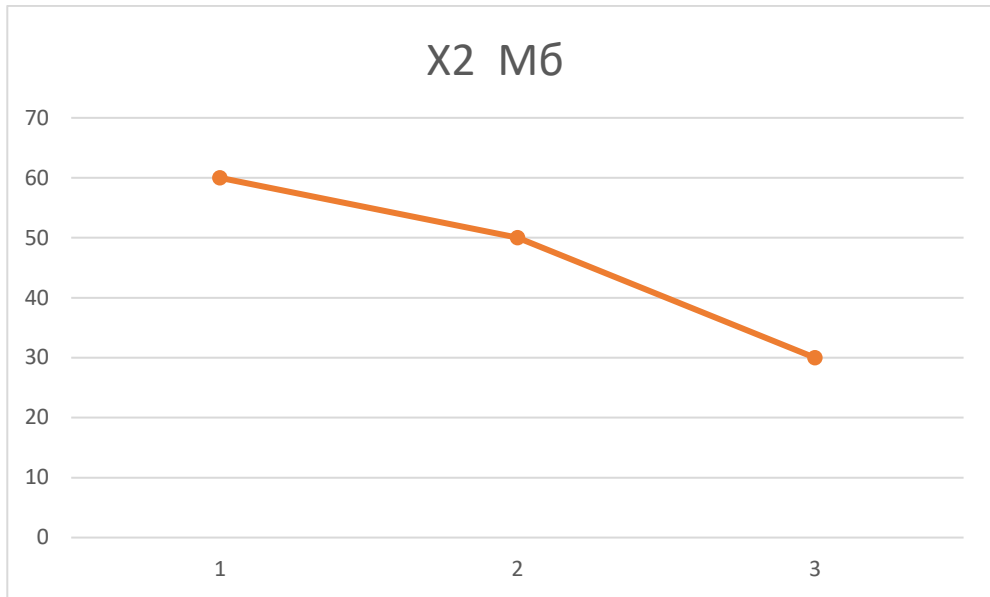


Рисунок 4.3 – X2, об'єм пам'яті

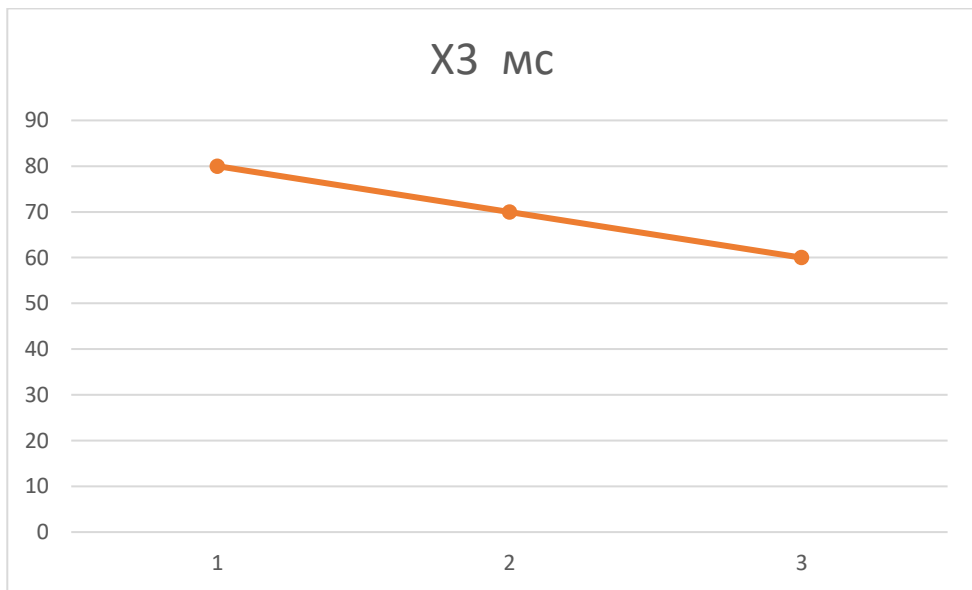


Рисунок 4.4 – X3, час попередньої обробки даних

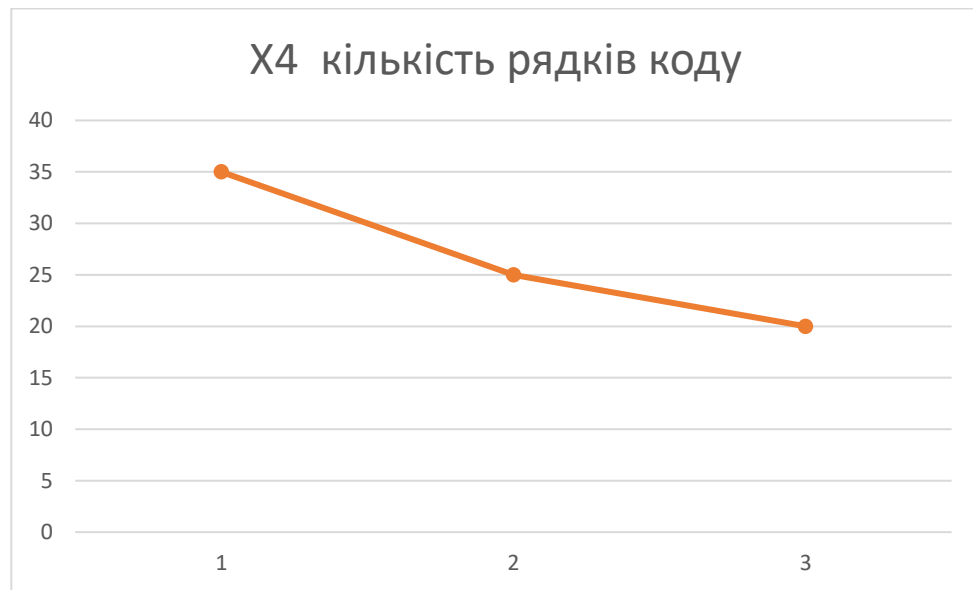


Рисунок 4.5 – X4, потенційний об'єм програмного коду

4.4 Аналіз експертного оцінювання параметрів

Після детального обговорення й аналізу кожний експерт оцінює ступінь важливості кожного параметру для конкретно поставленої цілі – розробка програмного продукту, який дає найбільш точні результати при знаходженні параметрів моделей адаптивного прогнозування і обчислення прогнозних значень.

Значимість кожного параметра визначається методом попарного порівняння. Оцінку проводить експертна комісія із 7 людей. Визначення коефіцієнтів значимості передбачає:

- визначення рівня значимості параметра шляхом присвоєння різних рангів;
- перевірку придатності експертних оцінок для подальшого використання;
- визначення оцінки попарного пріоритету параметрів;
- обробку результатів та визначення коефіцієнту значимості.

Результати експертного ранжування наведені у таблиці 4.3.

Таблиця 4.3 – Результати ранжування параметрів

Позначення параметра	Назва параметра	Одиниці виміру	Ранг параметра за оцінкою експерта							Сума рангів R_i	Відхилення Δ_i	Δ_i^2
			1	2	3	4	5	6	7			
X1	Швидкодія мови програмування	Оп/мс	2	2	1	2	2	2	1	12	-5,5	30,25
X2	Об'єм пам'яті	Мб	3	3	4	3	4	3	3	23	5,5	30,25
X3	Час попередньої обробки даних	мс	4	4	3	4	3	4	4	26	8,5	72,25
X4	Потенційний об'єм програмного коду	Кількість рядків коду	1	1	2	1	1	1	2	9	-8,5	72,25
	Разом		10	10	10	10	10	10	10	70	0	205

Для перевірки степені достовірності експертних оцінок, визначимо наступні параметри:

а) сума рангів кожного з параметрів і загальна сума рангів:

$$R_i = \sum_{j=1}^N r_{ij} R_{ij} = \frac{Nn(n+1)}{2} = 70, \quad (4.1)$$

де N – число експертів,

n – кількість параметрів;

б) середня сума рангів:

$$T = \frac{1}{n} R_{ij} = 17,5 \quad (4.2)$$

в) відхилення суми рангів кожного параметра від середньої суми рангів:

$$\Delta_i = R_i - T. \quad (4.3)$$

Сума відхилень по всіх параметрам повинна дорівнювати 0;

г) загальна сума квадратів відхилення:

$$S = \sum_{i=1}^N \Delta_i^2 = 205. \quad (4.4)$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{12S}{N^2(n^3 - n)} = \frac{12 \cdot 205}{7^2(4^3 - 4)} = 0,84 > W_k = 0,67. \quad (4.5)$$

Ранжування можна вважати достовірним, тому що знайдений коефіцієнт узгодженості перевищує нормативний, котрий дорівнює 0,67.

Скориставшись результатами ранжирування, проведемо попарне порівняння всіх параметрів і результати занесемо у таблицю 4.4.

Таблиця 4.4 – Попарне порівняння параметрів.

Параметри	Експерти							Кінцева оцінка	Числове значення
	1	2	3	4	5	6	7		
X1 і X2	<	<	<	<	<	<	<	<	0,5
X1 і X3	<	<	<	<	<	<	>	<	0,5
X1 і X4	>	>	<	>	>	>	<	>	1,5
X2 і X3	<	<	>	<	>	<	<	<	0,5
X2 і X4	>	>	>	>	>	>	>	>	1,5
X3 і X4	>	>	>	>	>	>	>	>	1,5

Числове значення, що визначає ступінь переваги i -го параметра над j -тим, a_{ij} визначається по формулі:

$$a_{ij} = \begin{cases} 1.5 \text{ при } X_i > X_j \\ 1.0 \text{ при } X_i = X_j \\ 0.5 \text{ при } X_i < X_j \end{cases} \quad (4.6)$$

З отриманих числових оцінок переваги складемо матрицю $A = \|a_{ij}\|$.

Для кожного параметра зробимо розрахунок вагомості K_{bi} за наступними формулами:

$$K_{bi} = \frac{b_i}{\sum_{i=1}^n b_i} \quad (4.7)$$

$$b_i = \sum_{i=1}^N a_{ij} \quad (4.8)$$

Відносні оцінки розраховуються декілька разів доти, поки наступні значення не будуть незначно відрізнятися від попередніх (менше 2%). На другому і наступних кроках відносні оцінки розраховуються за наступними формулами:

$$K_{Bi} = \frac{b'_i}{\sum_{i=1}^n b'_i}, \quad (4.9)$$

$$b'_i = \sum_{j=1}^N a_{ij} b_j \quad (4.10)$$

Як видно з таблиці 4.5, різниця значень коефіцієнтів вагомості не перевищує 2%, тому більшої кількості ітерацій не потрібно.

Таблиця 4.5 – Розрахунок вагомості параметрів

Параметри x_i	Параметри x_j				Перша ітер.		Друга ітер.		Третя ітер	
	X1	X2	X3	X4	b_i	K_{Bi}	b_i^1	K_{Bi}^1	b_i^2	K_{Bi}^2
X1	1	0,5	0,5	1,5	3,5	0,23	11,75	0,22	41,125	0,22
X2	1,5	1	0,5	0,5	3,5	0,23	12,75	0,24	44,625	0,24
X3	1,5	1,5	1	1,5	5,5	0,37	19,75	0,37	69,625	0,37
X4	0,5	0,5	0,5	1	2,5	0,17	8,75	0,17	30,875	0,17
Всього:					15	1	53	1	186,25	1

4.5 Аналіз рівня якості варіантів реалізації функцій

Визначаємо рівень якості кожного варіанту виконання основних функцій окремо.

Абсолютні значення параметрів X2 (Об'єм пам'яті), X3 (час попередньої обробки даних) та X4 (потенційний об'єм програмного коду) відповідають технічним вимогам умов функціонування даного ПП.

Абсолютне значення параметра $X1$ (швидкість роботи мови програмування) обрано не найгіршим.

Коефіцієнт технічного рівня для кожного варіанта реалізації ПП розраховується так (таблиця 4.6):

$$K_K(j) = \sum_{i=1}^n K_{\delta i,j} B_{i,j}, \quad (4.11)$$

де n – кількість параметрів;

$K_{\delta i}$ – коефіцієнт вагомості i -го параметра;

B_i – оцінка i -го параметра в балах.

Таблиця 4.6 – Розрахунок показників рівня якості варіантів реалізації основних функцій ПП

Основні функції	Варіант реалізації функції	Параметри	Абсолютне значення параметра	Бальна оцінка параметра	Коефіцієнт вагомості параметра	Коефіцієнт рівня якості
F1	A	X1	100	25	0,22	5,5
F3	A	X2	85	29	0,24	6,96
	B	X3	26	19	0,37	7,03
F4	A	X4	23	23	0,17	3,91

За даними з таблиці 4.6 за формулою:

$$K_K = K_{\text{ТУ}}[F_{1k}] + K_{\text{ТУ}}[F_{2k}] + \dots + K_{\text{ТУ}}[F_{zk}], \quad (4.12)$$

визначаємо рівень якості кожного з варіантів:

$$K_{K1} = 5,5 + 6,96 + 3,91 = 16,37 ;$$

$$K_{K2} = 5,5 + 7,03 + 3,91 = 16,44 .$$

Як видно з розрахунків, кращим є 2 варіант, для якого коефіцієнт технічного рівня має найбільше значення.

4.6 Економічний аналіз варіантів розробки ПП

Для визначення вартості розробки ПП спочатку проведемо розрахунок трудомісткості.

Всі варіанти включають в себе два окремих завдання:

1. Розробка проекту програмного продукту;
2. Розробка програмної оболонки;

Завдання 1 за ступенем новизни відноситься до групи А, завдання 2 – до групи Б. За складністю алгоритми, які використовуються в завданні 1 належать до групи 1; а в завданні 2 – до групи 3.

Для реалізації завдання 1 використовується довідкова інформація, а завдання 2 використовує інформацію у вигляді даних.

Проведемо розрахунок норм часу на розробку та програмування для кожного з завдань.

Загальна трудомісткість обчислюється як:

$$T_0 = T_P \cdot K_{\Pi} \cdot K_{СК} \cdot K_M \cdot K_{СТ} \cdot K_{СТ.М}, \quad (4.13)$$

де T_P – трудомісткість розробки ПП;

K_{Π} – поправочний коефіцієнт;

$K_{СК}$ – коефіцієнт на складність вхідної інформації;

K_M – коефіцієнт рівня мови програмування;

$K_{СТ}$ – коефіцієнт використання стандартних модулів і прикладних програм;

$K_{СТ.М}$ – коефіцієнт стандартного математичного забезпечення

Для першого завдання, виходячи із норм часу для завдань розрахункового характеру степеню новизни А та групи складності алгоритму 1, трудомісткість дорівнює: $T_P = 37$ людино-днів. Поправочний коефіцієнт, який враховує вид нормативно-довідкової інформації для першого завдання: $K_{П} = 1.8$. Поправочний коефіцієнт, який враховує складність контролю вхідної та вихідної інформації для всіх семи завдань рівний 1: $K_{СК} = 1$. Оскільки при розробці першого завдання використовуються стандартні модулі, врахуємо це за допомогою коефіцієнта $K_{СТ} = 0.9$. Тоді загальна трудомісткість програмування першого завдання дорівнює:

$$T_1 = 37 \cdot 1.8 \cdot 0.9 = 59,94 \text{ людино-днів.}$$

Проведемо аналогічні розрахунки для подальших завдань.

Для другого завдання (використовується алгоритм третьої групи складності, степінь новизни Б), тобто $T_P = 29$ людино-днів, $K_{П} = 0.9$, $K_{СК} = 1$, $K_{СТ} = 0.8$:

$$T_2 = 29 \cdot 0.9 \cdot 0.8 = 20.88 \text{ людино-днів.}$$

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їх трудомісткість:

$$T_I = (59,94 + 20.88 + 4.8 + 20.88) \cdot 8 = 852 \text{ людино-годин.}$$

$$T_{II} = (59,94 + 20,88 + 6,91 + 20,88) \cdot 8 = 868,88 \text{ людино-годин.}$$

Найбільш високу трудомісткість має варіант II.

В розробці беруть участь два програмісти з окладом 27000 грн., один аналітик в області даних з окладом 29500. Визначимо середню зарплату за годину за формулою:

$$C_{ч} = \frac{M}{T_m \cdot t} \text{ грн.}, \quad (4.14)$$

де M – місячний оклад працівників;

T_m – кількість робочих днів тиждень;

t – кількість робочих годин в день.

$$C_{ч} = \frac{27000 + 27000 + 29500}{3 \cdot 21 \cdot 8} = 165,67 \text{ грн.} \quad (4.15)$$

Тоді, розрахуємо заробітну плату за формулою:

$$C_{зп} = C_{ч} \cdot T_i \cdot K_d, \quad (4.16)$$

де $C_{ч}$ – величина погодинної оплати праці програміста;

T_i – трудомісткість відповідного завдання;

K_d – норматив, який враховує додаткову заробітну плату.

Зарплата розробників за варіантами становить:

$$I. C_{зп} = 165,67 \cdot 852 \cdot 1,2 = 169381 \text{ грн.}$$

$$II. C_{зп} = 165,67 \cdot 868,88 \cdot 1,2 = 172736,82 \text{ грн.}$$

Відрахування на єдиний соціальний внесок становить 22%:

$$I. C_{\text{ВІД}} = C_{\text{ЗП}} \cdot 0.22 = 169381 \cdot 0.22 = 37263,82 \text{ грн.}$$

$$II. C_{\text{ВІД}} = C_{\text{ЗП}} \cdot 0.22 = 172736,82 \cdot 0.22 = 38002,1 \text{ грн.}$$

Тепер визначимо витрати на оплату однієї машино-години. (C_M)

Так як одна ЕОМ обслуговує одного програміста з окладом 27000 грн., з коефіцієнтом зайнятості 0,2 то для однієї машини отримаємо:

$$C_T = 12 \cdot M \cdot K_3 = 12 \cdot 27000 \cdot 0,2 = 64800 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$C_{\text{ЗП}} = C_T \cdot (1 + K_3) = 64800 \cdot (1 + 0.2) = 77760 \text{ грн.}$$

Відрахування на соціальний внесок:

$$C_{\text{ВІД}} = C_{\text{ЗП}} \cdot 0.22 = 77760 \cdot 0,22 = 17107,2 \text{ грн.}$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 10000 грн.

$$C_A = K_{\text{ТМ}} \cdot K_A \cdot C_{\text{ПР}} = 1.4 \cdot 0.25 \cdot 10000 = 3500 \text{ грн.,}$$

де $K_{\text{ТМ}}$ – коефіцієнт, який враховує витрати на транспортування та монтаж приладу у користувача;

K_A – річна норма амортизації;

$C_{\text{ПР}}$ – договірна ціна приладу.

Витрати на ремонт та профілактику розраховуємо як:

$$C_P = K_{TM} \cdot C_{ПР} \cdot K_P = 1.4 \cdot 10000 \cdot 0.08 = 1120 \text{ грн.},$$

де K_P – відсоток витрат на поточні ремонти.

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$\begin{aligned} T_{EF} &= (D_K - D_B - D_C - D_P) \cdot t_3 \cdot K_B = (365 - 104 - 12 - 16) \cdot 8 \cdot 0.35 = \\ &= 627,2 \text{ години,} \end{aligned}$$

де D_K – календарна кількість днів у році;

D_B, D_C – відповідно кількість вихідних та святкових днів;

D_P – кількість днів планових ремонтів устаткування;

t – кількість робочих годин в день;

K_B – коефіцієнт використання приладу у часі протягом зміни.

Витрати на оплату електроенергії розраховуємо за формулою:

$$C_{EL} = T_{EF} \cdot N_C \cdot K_3 \cdot C_{EH} = 627.2 \cdot 0,25 \cdot 0,3 \cdot 5,21 = 245.08 \text{ грн.},$$

де N_C – середньо-споживча потужність приладу;

K_3 – коефіцієнтом зайнятості приладу;

C_{EH} – тариф за 1 КВт-годин електроенергії.

Накладні витрати розраховуємо за формулою:

$$C_H = C_{ПР} \cdot 0.67 = 10000 \cdot 0,67 = 6700 \text{ грн.}$$

Тоді, річні експлуатаційні витрати будуть:

$$C_{\text{ЕКС}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_{\text{А}} + C_{\text{Р}} + C_{\text{ЕЛ}} + C_{\text{Н}}, \quad (4.17)$$

$$C_{\text{ЕКС}} = 77760 + 17107,2 + 3500 + 1120 + 245,08 + 6700 = 106432,28 \text{ грн.}$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$C_{\text{М-Г}} = C_{\text{ЕКС}} / T_{\text{ЕФ}} = 106432,28 / 627,2 = 169,69 \text{ грн/год.}$$

Оскільки в даному випадку всі роботи, які пов'язані з розробкою програмного продукту ведуться на ЕОМ, витрати на оплату машинного часу, в залежності від обраного варіанта реалізації, складає:

$$C_{\text{М}} = C_{\text{М-Г}} \cdot T, \quad (4.18)$$

$$\text{I. } C_{\text{М}} = 169,69 \cdot 852 = 144527,88 \text{ грн.}$$

$$\text{II. } C_{\text{М}} = 169,69 \cdot 868,88 = 147440,25 \text{ грн.}$$

Накладні витрати складають 67% від заробітної плати:

$$C_{\text{Н}} = C_{\text{ЗП}} \cdot 0,67, \quad (4.19)$$

$$\text{I. } C_{\text{Н}} = 169381 \cdot 0,67 = 113485,27 \text{ грн.}$$

$$\text{II. } C_{\text{Н}} = 172736,82 \cdot 0,67 = 115733,67 \text{ грн.}$$

Отже, вартість розробки ПП за варіантами становить:

$$C_{\text{ПП}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_{\text{М}} + C_{\text{Н}}, \quad (4.20)$$

$$I. C_{III} = 169381 + 37263,82 + 144527,88 + 113485,27 = 464657,97 \text{ грн.}$$

$$II. C_{III} = 172736,82 + 38002,1 + 147440,25 + 115733,67 = 473912,84 \text{ грн.}$$

4.7 Вибір кращого варіанту III техніко-економічного рівня

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$K_{TEPj} = K_{kj} / C_{\Phi j}, \quad (4.21)$$

$$K_{TEP1} = 16,37 / 464657,97 = 3,52 \cdot 10^{-5},$$

$$K_{TEP2} = 16,44 / 473912,84 = 3,47 \cdot 10^{-5}.$$

Як бачимо, розраховані коефіцієнти майже однакові, але найбільш ефективним є перший варіант реалізації програми з коефіцієнтом техніко-економічного рівня $K_{TEP1} = 3,52 \cdot 10^{-5}$.

Після виконання функціонально-вартісного аналізу програмного комплексу що розроблюється, можна зробити висновок, що з альтернатив, що залишились після першого відбору двох варіантів виконання програмного комплексу оптимальним є перший варіант реалізації програмного продукту. У нього виявився найкращий показник техніко-економічного рівня якості $K_{TEP} = 3,52 \cdot 10^{-5}$.

Цей варіант реалізації програмного продукту має такі параметри:

- вибір програмного продукту – Eviews;
- реалізація важливої постановки з допомогою вбудованих функцій;
- використання стандартного інтерфейсу для побудови значень.

Даний варіант виконання програмного комплексу дає користувачу зручний інтерфейс, швидку реалізацію програми та доступний функціонал для роботи.

4.8 Висновки до розділу 4

В даному розділі було проведено повний функціонально-вартісний аналіз програмного продукту. Також було знайдено оцінку основних функцій програмного продукту.

В результаті виконання функціонально-вартісного аналізу програмного комплексу що розроблюється, було визначено та проведено оцінку основних функцій програмного продукту, а також знайдено параметри, які його характеризують.

ВИСНОВКИ ЗА РОБОТОЮ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Дипломна робота присвячена дослідженню демографічних процесів Японії та Данії та моделюванню демографічних показників.

Демографія є дуже важливою наукою для сучасного світу. Адже, всі демографічні процеси мають значний вплив на інші економічні, соціальні, екологічні та політичні процеси. Для ефективного управління демографічними змінами необхідні точні прогнози та обґрунтовані рекомендації. Використання регресійних моделей у прогнозуванні демографічних тенденцій дозволяє урядам та політикам розробляти більш точні та ефективні стратегії.

Аналізуючи глобальні демографічні тенденції, можна відзначити, що населення планети швидко зростає, але цей ріст спостерігається переважно в країнах, що розвиваються, таких як країни Азії, Африки та Південної Америки. Водночас у розвинених країнах через низьку народжуваність і "старіння" населення природний приріст є незначним, а в деяких випадках навіть негативним.

У цій роботі було проведено аналіз та дослідження тенденції поведінки демографічних показників для Японії та Данії, зокрема народжуваності, смертності і чисельності населення та для кожного з них побудовано прогноз на 4 роки. Обидві країни стикаються зі значними викликами через старіння населення. У Японії цей процес є особливо гострим, оскільки частка літніх людей у структурі населення стрімко зростає, що спричиняє зменшення працездатного населення та підвищує навантаження на соціальні та медичні системи.

Було описано структури математичних моделей, зокрема AR, ARMA, ARIMA та наведено узагальнений алгоритм їх побудови. В ході виконання роботи за допомогою власного програмного продукту, написаного на мові Python, були побудовані авторегресійні моделі різних порядків та прогнози на

4 роки. Кожна з них була оцінена на адекватність та точність прогнозів за допомогою критеріїв адекватності моделі (критерію детермінації, статистики Дарбіна-Уотсона, суми квадратів похибок) та якості прогнозів (середня квадратична похибка, середня абсолютна похибка у відсотках та коефіцієнт Тейла). Проведений аналіз показав, що використані моделі здатні адекватно відображати демографічні процеси та надавати досить точні прогнози.

Дослідження надає важливу інформацію для політиків, економістів, соціологів та інших фахівців, що займаються питаннями демографії та соціально-економічного розвитку.

Подальші дослідження повинні зосереджуватися на покращенні та адаптації регресійних моделей для більш точного прогнозування демографічних змін. Важливо враховувати нові фактори, такі як зміни в політичному, економічному та соціальному середовищі. Дослідження впливу війни на демографічні процеси в Україні, зокрема міграційні потоки, зміни в народжуваності та смертності. Вивчення цих змін допоможе краще розуміти поточну ситуацію та прогнозувати подальші тенденції. Адже війна має значний вплив на ринок праці, зокрема на доступність робочих місць, безробіття та міграцію робочої сили. Дослідження цих аспектів може допомогти у розробці політики підтримки економічної стабільності та відновлення після війни.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Григор'єва С. В. Основи демографії : конспект лекцій. Харків : НТУ „ХПІ”, 2010. 77 с.
2. Гудзеляк І. Основи демографії : навч. посіб. Львів : Вид. центр ЛНУ ім. І. Франка, 2013. 81 с.
3. Japan: population by gender 2003-2022 | Statista. Statista. URL: <https://www.statista.com/statistics/612246/japan-population-breakdown-total-gender/> (дата звернення: 21.04.2024).
4. Бідюк П. І., Романенко В. Д., Тимошук О. Л. Аналіз часових рядів : навч. посіб. Київ : Політехніка, 2010. 316 с.
5. Комплексний демографічний прогноз України на період до 2050 р. : монографія / Н. С. Власенко та ін. Київ : Укр. центр реформ, 2006. 122 с.
6. Бідюк П. І., Гожий О. П. Ймовірно-статистичні методи моделювання і прогнозування : монографія. Миколаїв : Чорном. держ. ун-т ім. Петра Могили, 2014. 440 с.
7. Бідюк П. І., Коршевнік Л. О. Проектування комп'ютерних інформаційних систем підтримки прийняття рішень : навч. посіб. Київ : ННК "ІПСА" НТУУ "КПІ", 2010. 340 с.
8. Лук'яненко І. Г., Жук В. Н. Аналіз часових рядів побудова VAR і VECM моделей з використанням пакета e.views 6.0 : навч. посіб. Київ : Нац. ун-т "Києво-Морил. акад.", 2013. 174 с.
9. Лук'яненко І., Гордієнко Ю. Сучасні економетричні методи у фінансах : навч. посіб. Київ : Літера ЛТД, 2002. 349 с.
10. Перерва А. С., Ковалюк Т. В. Порівняльний аналіз методів прогнозування стаціонарних і нестаціонарних часових рядів. Науковий огляд. 2018. С. 4–6.
11. Шевчук П. Україна: мультирегіональна модель народжуваності. Регіональна економіка. 2000. №2. С. 128–133.

12. Денисюк В. О. Огляд засобів статистичної обробки і аналізу даних. *Polish journal of science*. 2020. № 27. С. 14–23.
13. Бідюк П. І., Терентьев О. М., Просьянкіна-Жарова Т. І. Прикладна статистика : навч. посіб. Вінниця : ПП «ТД»Едельвейс і К», 2013. 304 с.
14. Находов В. Ф., Бориченко О. В., Іванько Д. О. Вибір необхідного складу критеріїв адекватності математичних моделей енергоспоживання в системах оперативного контролю енергоефективності. *Енергетика: економіка, технології, екологія*. 2013. № 3. С. 68–78.
15. Гладун О., Кулик Н., Рудницький О. Населення України. *Енциклопедія сучасної України*. Київ, 2020.
16. Шубенкова І. А., Петрова С. К., Бідюк П. І. Системний підхід до моделювання та прогнозування на основі регресійних моделей і фільтра Калмана. *Системні дослідження та інформаційні технології*. Київ, 2017. № 2. С. 52-62.
17. Юрченко М. Є. Прогнозування та аналіз часових рядів. *Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика» ЧНТУ*, 2018. 88 с.
18. Бідюк П. І., Баклан І. В., Литвиненко В. І. Моделювання і прогнозування гетероскедастичних процесів. *Системні дослідження та інформаційні технології*. Київ, 2004. С. 115 – 134.
19. Davis R. A, Brockwell P. J. *Introduction to Time Series and Forecasting: 2nd ed.* New York : Springer, 2016. 426 p.
20. Young P. *Forecasting and time series analysis. International Journal of Forecasting*. 1991. 248 p.
21. Stevens, L. J. *Time Series Modelling of Demographic Processes*. New York: *Journal of Population Research*, 2018. 212 p.
22. Бачо Р. Й. Екстраполяційне прогнозування демографічних показників Закарпатської області. *Науковий вісник УжНУ. Серія Економіка. Спецвипуск 33. Частина 1*. 2011. С. 11– 14.

ДОДАТОК А (ілюстративний матеріал доповіді)

Моделі і прогнози демографічних процесів

Виконала: студентка групи КА-05
Заржецька Анастасія Олександрівна
Науковий керівник: ст. викладач кафедри ММСА, PhD
Левенчук Людмила Борисівна

Об'єкт: дослідження демографічних процесів в Японії та Данії.

Предмет дослідження: вибір структури математичних моделей, що адекватно описують демографічні процеси та забезпечують якісну оцінку прогнозу.

Мета дослідження: оцінити поточну демографічну ситуацію в Японії та Данії за допомогою моделювання демографічних показників та оцінити прогноз на наступні 4 роки.

Постановка задачі

1. Зробити огляд існуючих моделей демографічних процесів
2. Зібрати та обробити статистичні дані
3. Розробити математичні моделі для прогнозування динаміки населення в майбутньому.
4. Здійснити оцінку якості результатів демографічних прогнозів.
5. Візуалізувати прогнозні значення та порівняти їх зі спостережуваними даними та зробити висновки щодо наявних тенденцій та прогнозування майбутніх демографічних процесів.

3

Основні демографічні проблеми

01

Старіння населення

Низький рівень народжуваності в розвинених країнах

02

Міграційні процеси

Виїзд кваліфікованих працівників до інших країн

03

Перенаселення землі

Активне зростання населення у країнах третього світу

04

Урбанізація

Масове переселення населення з сільських районів у міста.

05

Військові конфлікти

Збільшення смертності, зменшення народжуваності та міграційні кризи

4

Авторегресійні моделі

AR (авторегресійна модель): $y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t$,

ARMA (авторегресійна модель з ковзним середнім): $y_t = c + \varepsilon_t + \sum_{i=1}^p a_i y_{t-i} + \sum_{j=0}^q b_j \varepsilon_{t-j}$,

ARIMA (авторегресійна модель з інтегрованим ковзним середнім): $\Delta y_t = c + \varepsilon_t + \sum_{i=1}^p a_i \Delta y_{t-i} + \sum_{j=0}^q b_j \varepsilon_{t-j}$,

5

Критерії якості прогнозування для вибору кращої моделі

Критерії адекватності

моделі:

SSE (сума квадратів похибок)

R² (критерій детермінації)

DW (статистика Дарбіна-Уотсона)

Критерії якості прогнозів:

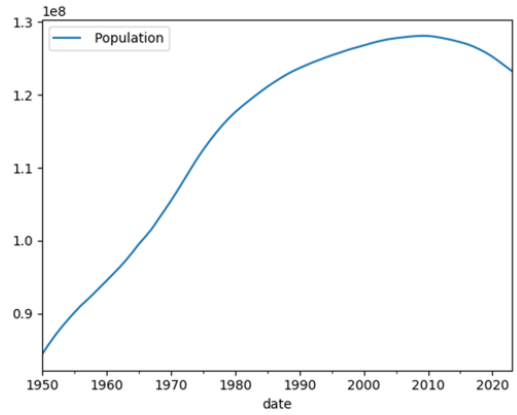
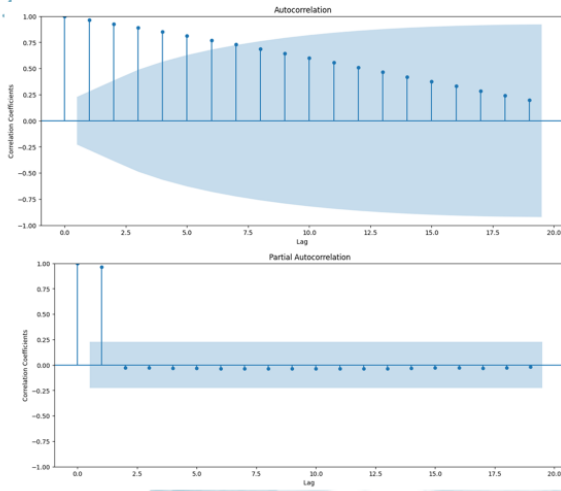
СКП (середня квадратична похибка)

САПП (середня абсолютна похибка в процентах)

Коефіцієнт Тейла (вимірює точність прогнозів)

6

Дослідження чисельності населення Японії



7

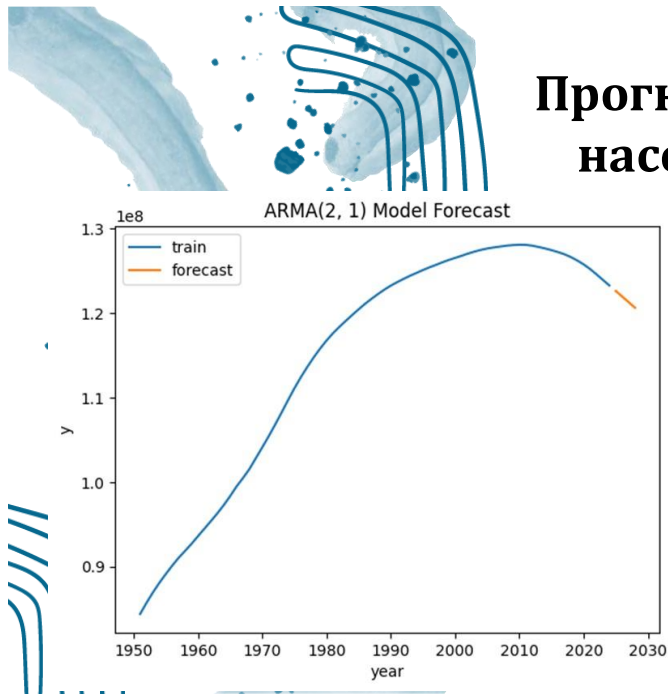
Кращою моделлю виявилася

ARMA(2,1): $y(k) = 9.47e^7 + 1.9595y(k - 1) - 0.9907(k - 2) - 0,7501\varepsilon(k - 1)$

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R ²	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	10e ¹²	0,7036	2,4353	5,192e ¹⁰	0,7272	0,024
ARMA(2,1)	5e ¹²	0,8566	1,826	5,19e ¹⁰	0,7224	0,024

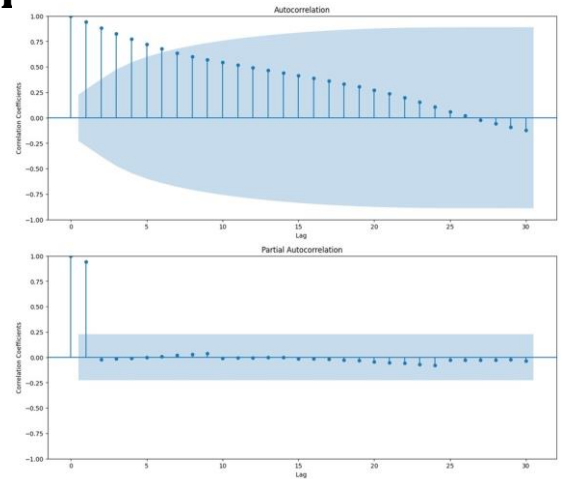
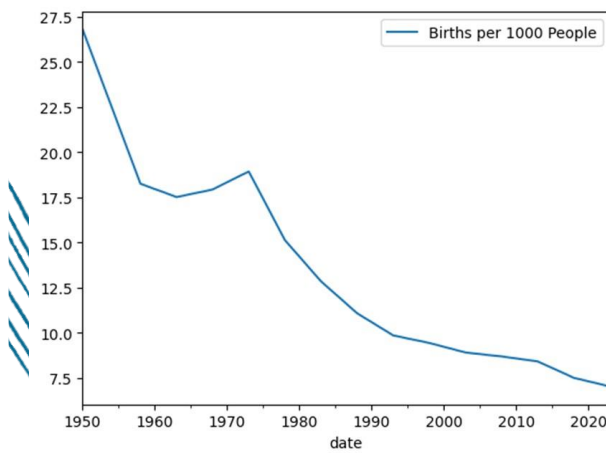
8

Прогноз чисельності населення Японії



Рік	Чисельність населення
2024	1226368
2025	1229767
2026	1213150
2027	1206524

Дослідження показника народжуваності Японії



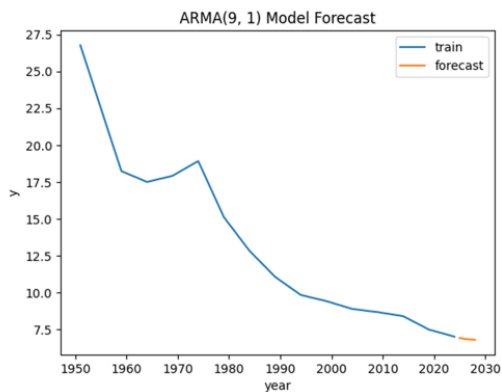
Моделі-кандидати для показника народжуваності Японії

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R ²	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	5,2285	0,3917	1,5856	0.8141	1.4523	0.0449
ARMA(2,8)	3,2291	0,6434	1,5652	1.0096	1.7371	0.05
AR(9)	2,8085	0,6732	1,5845	0.8031	1.3194	0.0447
ARMA(9,1)	0,4386	0,9489	1,5122	0.8353	1.3569	0.0456

11

Прогноз показника народжуваності Японії

Кращою моделлю виявилася ARMA(9,1) | ARMA(9,1): $y(k) = 22,8803 + 0,864y(k-1) + 0,9444y(k-2) - 0,8169y(k-3) - 0,0191y(k-4) + 0,2501y(k-5) - 0,2065y(k-6) - 0,2181y(k-7) + 0,1629y(k-8) - 0,0005y(k-9) + 0,9605\varepsilon(k-1)$



Рік	Показник народжуваності
2024	6.926020
2025	6.863779
2026	6.823790
2027	6.804682

12

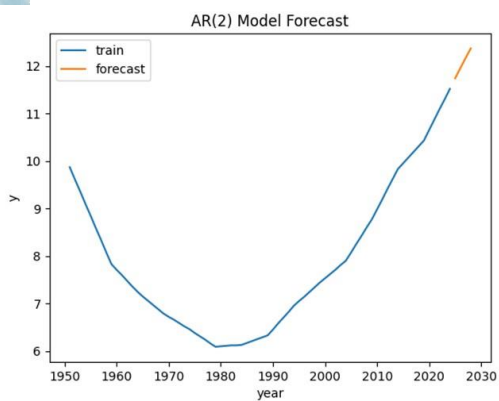
Моделі-кандидати для показника смертності Японії

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R ²	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	0,5299	0,9739	1,8141	0,0252	0,6309	0.02
AR(9)	1,4624	0,9281	1,417	0,0696	1,0731	0.0332
ARMA(2,1)	0,8643	0,9574	1,6274	0,0412	0,7485	0.0256
ARMA(9,1)	0,6889	0,9661	1,7081	0,0328	0,7674	0.0228

13

Прогноз показника смертності Японії

Кращою моделлю виявилася AR(2):



$$y(k) = 9,2012 + 1.9784y(k - 1) - 0.9848y(k - 2)$$

Рік	Показник смертності
2024	11.740077
2025	11.954634
2026	12.165080
2027	12.369402

14

Порівняння моделей-кандидатів для чисельності населення Данії

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R ²	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	10e ¹¹	0,73	2,45	5e ⁹	0,4247	0,0155
ARMA(2,1)	5e ¹¹	0,88	2,3	5e ⁹	0,4164	0,0154

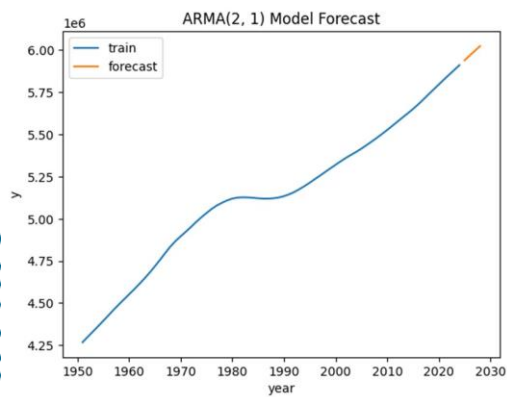
15

Прогноз чисельності населення Данії

Кращою моделлю виявилася

ARMA(2,1):

$$y(k) = 4.591e^6 + 1.992y(k-1) - 0.9956y(k-2) + 0.996\varepsilon(k-1)$$



Рік	Чисельність населення
2024	5939683
2025	5968092
2026	5996142
2027	6023834

16

Порівняння моделей-кандидатів для показника народжуваності Данії

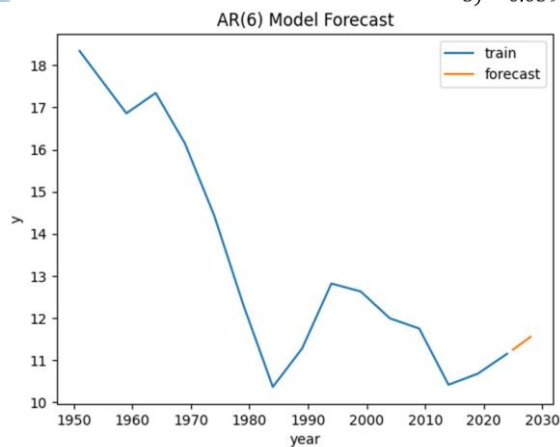
Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R ²	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	2.6476	0,72	2,36	0,1261	0,7367	0,0208
AR(4)	2,5733	0,83	2,23	0,12254	0,7242	0,0205
AR(6)	2,6459	0,98	1,94	0,126	0,721	0,0207

17

Прогноз показника народжуваності Данії

Краща модель AR(6)

$$AR(6): y(k) = 16.7854 + 1.9108y(k-1) - 0.9576y(k-2) + 0.0693y(k-3) - 0.0597y(k-4) - 0.0579y(k-5) + 0.0797y(k-6)$$



Рік	Показник народжуваності
2024	11.246514
2025	11.347160
2026	11.448862
2027	11.550992

18

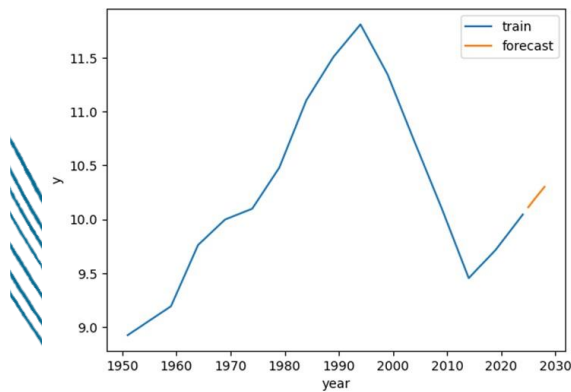
Порівняння моделей-кандидатів для показника народжуваності Данії

Тип моделі	Характеристики моделі			Характеристики прогнозу		
	SSE	R ²	DW	СКП	САПП	Коефіцієнт Тейла
AR(2)	0,1116	0,69	2,38	0,0139	0,5011	0,0124
AR(8)	0,0267	0,83	1,93	0,0133	0,4246	0,0122

19

Прогноз показника смертності Данії

Краща модель AR(8)



$$\text{AR}(8): y(k) = 9.4537 + 1.0148y(k-1) + 0.6626y(k-2) - 0.9413y(k-3) + 0.3796y(k-4) - 0.5299y(k-5) + 0.1898y(k-6) - 0.7745y(k-7) - 0.6705y(k-8)$$

Рік	Показник смертності
2024	10.111406
2025	10.176636
2026	10.240073
2027	10.301111

20

Аналіз результатів

Японія: кількість населення та показник народжуваності будуть й надалі зменшуватися, в той час як показник смертності буде зростати. Наслідком цього є зменшення природного приросту, що має значний вплив на демографію країни. Старіння нації призводить до зменшення кількості працездатного населення, що досить негативно відображається на економіці країни. Потрібно аби були вжиті заходи з боку держави.

Данія: Аналіз показав, що, хоча демографічні зміни в Данії не настільки драматичні, все ж є тенденції до старіння населення та певної стабільності у народжуваності, що вимагає уваги до соціальних та економічних політик.

21

Висновки

1. Здійснено огляд існуючих моделей демографічних процесів.
2. Досліджено демографічні процеси Японії та Данії.
3. Розроблено моделі для прогнозування.
4. Здійснено оцінку якості результатів прогнозів та виявлено кращі моделі.
5. Виконано аналіз отриманих прогнозів.

22

Перспективи подальших досліджень

Подальші дослідження повинні зосереджуватися на покращенні та адаптації регресійних моделей для більш точного прогнозування демографічних змін. Важливо враховувати нові фактори, такі як зміни в політичному, економічному та соціальному середовищі. Дослідження впливу війни на демографічні процеси в Україні, зокрема міграційні потоки, зміни в народжуваності та смертності.

Дякую за увагу!

ДОДАТОК Б (статистичні дані)

Основним джерелом даних Японії було взято веб-ресурс <https://www.macrotrends.net/global-metrics/countries/JPN/japan/population>, після чого їх було зведено до таблиці:

Таблиця Б.1 – Демографічні дані Японії

Рік	Популяція	Показник народжуваності	Показник смертності
1950	84353049	26,774	9,866
1951	85683474	25,708	9,612
1952	86921764	24,641	9,357
1953	88057140	23,574	9,102
1954	89093019	22,507	8,847
1955	90090281	21,44	8,592
1956	91021560	20,374	8,338
1957	91837665	19,307	8,083
1958	92671295	18,24	7,828
1959	93575985	18,092	7,711
1960	94464839	17,943	7,595
1961	95355458	17,795	7,478
1962	96267883	17,646	7,362
1963	97246712	17,498	7,245
1964	98319344	17,582	7,153
1965	99468746	17,666	7,062
1966	100466907	17,75	6,97
1967	101527852	17,834	6,879
1968	102819279	17,918	6,787
1969	104102077	18,117	6,721
1970	105416839	18,316	6,655
1971	106789355	18,516	6,588
1972	108220404	18,715	6,522
1973	109679470	18,914	6,456
1974	111100023	18,155	6,382
1975	112411640	17,396	6,308
1976	113624865	16,637	6,235
1977	114761516	15,878	6,161
1978	115809525	15,119	6,087
1979	116765509	14,662	6,096
1980	117624196	14,206	6,105
1981	118385723	13,749	6,115
1982	119098142	13,293	6,124
1983	119793752	12,836	6,133
1984	120470727	12,483	6,173

Продовження таблиці Б.1

1985	121119031	12,13	6,212
1986	121710305	11,777	6,252
1987	122268885	11,424	6,291
1988	122798302	11,071	6,331
1989	123270648	10,824	6,454
1990	123686321	10,578	6,577
1991	124073993	10,331	6,7
1992	124444693	10,085	6,823
1993	124779576	9,838	6,946
1994	125119269	9,755	7,045
1995	125433969	9,672	7,143
1996	125726223	9,589	7,242
1997	126027985	9,506	7,34
1998	126308390	9,423	7,439
1999	126555063	9,317	7,531
2000	126803861	9,212	7,622
2001	127065731	9,106	7,714
2002	127301750	9,001	7,805
2003	127502352	8,895	7,897
2004	127671288	8,851	8,072
2005	127798373	8,807	8,246
2006	127902167	8,763	8,421
2007	128006426	8,719	8,595
2008	128077633	8,675	8,77
2009	128117042	8,62	8,982
2010	128105431	8,565	9,193
2011	128007257	8,511	9,405
2012	127853688	8,456	9,616
2013	127678924	8,401	9,828
2014	127476735	8,219	9,948
2015	127250933	8,038	10,068
2016	126993857	7,856	10,187
2017	126662472	7,675	10,307
2018	126255866	7,493	10,427
2019	125791677	7,397	10,646
2020	125244761	7,301	10,865
2021	124612530	7,205	11,085
2022	123951692	7,109	11,304
2023	123294513	7,013	11,523

Для Данії було взято дані з того ж веб-сайту <https://www.macrotrends.net/global-metrics/countries/DNK/denmark/population> та занесено до таблиці Б.2.

Таблиця Б.2 – Демографічні дані Данії

Рік	Популяція	Показник народжуваності	Показник смертності
1950	4267694	18,343	8,923
1951	4299579	18,157	8,957
1952	4330757	17,972	8,990
1953	4362483	17,787	9,024
1954	4394320	17,602	9,058
1955	4426508	17,417	9,091
1956	4459559	17,231	9,125
1957	4491051	17,046	9,158
1958	4521329	16,861	9,192
1959	4551440	16,958	9,306
1960	4581727	17,054	9,420
1961	4613434	17,151	9,533
1962	4645994	17,247	9,647
1963	4680812	17,344	9,761
1964	4718016	17,104	9,808
1965	4756337	16,864	9,856
1966	4796403	16,624	9,903
1967	4834730	16,384	9,951
1968	4867734	16,144	9,998
1969	4896157	15,801	10,018
1970	4922963	15,458	10,038
1971	4951486	15,116	10,058
1972	4981078	14,773	10,078
1973	5008581	14,430	10,098
1974	5033573	14,003	10,174
1975	5058203	13,577	10,250
1976	5078773	13,150	10,327
1977	5094599	12,724	10,403
1978	5108930	12,297	10,479
1979	5119418	11,910	10,604
1980	5125392	11,523	10,730
1981	5126981	11,136	10,855
1982	5126398	10,749	10,981
1983	5124332	10,362	11,106
1984	5121314	10,545	11,186
1985	5119452	10,727	11,267
1986	5119369	10,910	11,347
1987	5121510	11,092	11,428
1988	5126107	11,275	11,508

Продовження таблиці Б.2.

1989	5133930	11,584	11,569
1990	5144623	11,892	11,629
1991	5158325	12,201	11,690
1992	5175469	12,509	11,750
1993	5193775	12,818	11,811
1994	5213800	12,780	11,718
1995	5235197	12,743	11,625
1996	5256007	12,705	11,533
1997	5277382	12,668	11,440
1998	5298992	12,630	11,347
1999	5319854	12,502	11,221
2000	5340655	12,374	11,096
2001	5360748	12,246	10,970
2002	5378697	12,118	10,845
2003	5396355	11,990	10,719
2004	5415586	11,942	10,596
2005	5436313	11,894	10,473
2006	5457864	11,845	10,349
2007	5479722	11,797	10,226
2008	5502751	11,749	10,103
2009	5526452	11,482	9,973
2010	5550849	11,215	9,843
2011	5576016	10,948	9,713
2012	5600959	10,681	9,583
2013	5625385	10,414	9,453
2014	5650653	10,466	9,506
2015	5677796	10,519	9,559
2016	5706857	10,571	9,611
2017	5737284	10,624	9,664
2018	5766686	10,676	9,717
2019	5795878	10,771	9,783
2020	5825641	10,865	9,848
2021	5854240	10,960	9,914
2022	5882261	11,054	9,979
2023	5910913	11,149	10,045

ДОДАТОК В (код програмного продукту)

```

import inspect
import time
import warnings
import matplotlib.pyplot as plt
import pandas as pd

from sklearn.metrics import mean_absolute_error
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima.model import ARIMA

warnings.filterwarnings("ignore")

df = pd.read_csv('data 1.csv')
df['date'] = pd.to_datetime(df['date'], format='%d.%m.%Y')
df.head()
print (df)
df.plot(x=df.columns[0], y=df.columns[1]) # показати населення
#df.plot(x=df.columns[0], y=df.columns[3]) # показати смертність
#df.plot(x=df.columns[0], y=df.columns[2]) # показати народжуваність

#y =df.drop(columns=[df.columns[2],df.columns[3]]) #якщо населення
y =df.drop(columns=[df.columns[1],df.columns[2]]) #якщо смертність
#y =df.drop(columns=[df.columns[1],df.columns[3]]) #якщо
народжуваність

print(y)
y.head()
fig, ax = plt.subplots(figsize= (15,6))
plot_acf(y.iloc[:,1], ax = ax)
plt.xlabel("Lag ")
plt.ylabel("Correlation Coefficients")
fig, ax = plt.subplots(figsize=(15,6))
plot_pacf(y.iloc[:,1], ax=ax)
plt.xlabel("Lag ")
plt.ylabel("Correlation Coefficients");
y_train = y.loc[:"2012-12-31"]
y_test = y.loc["2013-12-31":]

p_params = range(1,26,1)

from sklearn.preprocessing import MinMaxScaler

def sum_of_squared_errors(y_true, y_pred):
    return np.sum((y_true - y_pred) ** 2)

r2_sc_grid = dict()
sse_grid = dict()
dw_grid = dict()

best_model = None
best_sse = float('inf')

```

```

best_r2 = float('-inf')
best_dw = float('inf')
best_order = None

scaler = MinMaxScaler()

y_train_normalized = scaler.fit_transform(y_train.iloc[:,
1].values.reshape(-1, 1))

for p in p_params:
    sse_grid[p] = list()
    r2_sc_grid[p] = list()
    dw_grid[p] = list()

    order = (p, 0, 0)

    start_time = time.time()

    model = ARIMA(y_train_normalized, order=order,
enforce_stationarity=False).fit()
    elapsed_time = round(time.time() - start_time, 2)
    print(f"Trained ARIMA {order} in {elapsed_time} seconds.")

    y_pred = scaler.inverse_transform(y_pred_normalized.reshape(-1,
1))
    sse = mean_squared_error(y_train.iloc[:, 1], y_pred)
    r2 = r2_score(y_train.iloc[:, 1], y_pred)
    dw = durbin_watson(model.resid)
    sse_grid[p].append(sse)
    r2_sc_grid[p].append(r2)
    dw_grid[p].append(dw)

    if sse < best_sse:
        best_model = model
        best_sse = sse
        best_r2 = r2
        best_dw = dw
        best_order = order

print(f"Best Model Order: {best_order}")
print(f"Best SSE: {best_sse}")
print(f"Best R2: {best_r2}")
print(f"Best DW: {best_dw}")
print("SSE Grid:")
print(sse_grid)
print("R2 Grid:")
print(r2_sc_grid)
print("DW Grid:")
print(dw_grid)

model = ARIMA(y_train.iloc[:,1], order=(1, 0, 0))
result = model.fit()
result.summary()
y_pred_wfv = pd.Series()
y1 = y

history = y1.iloc[:,1].copy()

```

```
for i in range(0,4):
    model = ARIMA(history, order=(1,0,0)).fit()
    next_pred = model.forecast()
    y_pred_wfv = y_pred_wfv._append(next_pred)
    history = history._append(next_pred)

y_forecast =
pd.DataFrame({"date":["31.12.2024","31.12.2025","31.12.2026","31.12.20
27"], "Population":y_pred_wfv})
y_forecast['date'] = pd.to_datetime(y_forecast['date'],
format='%d.%m.%Y')
df.head()
print (y_forecast)

fig, ax = plt.subplots()
ax.plot(y1.iloc[:,0], y1.iloc[:,1], color="C0", label="train")
ax.plot(y_forecast.iloc[:,0],y_forecast.iloc[:,1], color="C1",
label="forecast")

ax.legend()
ax.set(xlabel="year", ylabel="y", title="AR(1) Model Forecast")
```