НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

НН Інститут прикладного системного аналізу

Кафедра математичних методів системного аналізу

До захисту допущено:
Завідувач кафедри
_____ Оксана ТИМОЩУК
«___»_____2023 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системний аналіз і управління»

спеціальності 124 «Системний аналіз»

на тему: «Моделі і прогнози діяльності виробничого підприємства»

Виконав:
Студент IV курсу, групи Ка-95
Гулкевич Борис Юрійович _____

Керівник:
Проф., д.т.н., Бідюк П.І. _____

Консультант з економічного розділу:
Доц., к.е.н., Рощина Н.В. _____

Консультант з нормконтролю:
Доц., к.ф.-м.н.., Статкевич В.М. _____

Рецензент:
Проф., д.т.н., Сільвестров А.М. _____

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

Київ – 2023 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
НН Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)
Спеціальність – 124 «Системний аналіз»
Освітня програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Оксана ТИМОЩУК
«__» травня 2023 р.

ЗАВДАННЯ

на дипломну роботу студенту

Гулкевичу Борису Юрійовичу

1. Тема роботи «Моделі і прогнозування діяльності виробничого підприємства», керівник роботи Бідюк Петро Іванович, доктор технічних наук, професор, затверджені наказом по університету від «30» травня 2023 р. № **2065-с**

2. Термін подання студентом роботи 12.06.2023.

3. Вихідні дані до роботи звітність ПАТ "АвтоКрАЗ"

4. Зміст роботи особливості і методи прогнозування результатів діяльності підприємства, докладний огляд обраних моделей та методів, проведення прогнозування даних за допомогою програмного забезпечення, функціонально-вартісний аналіз програмного продукту

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): предмет, об'єкт та мета роботи, постановка задачі, морфологічний аналіз перший етап результат, результати прогнозування для фінансів ARIMA(1,2,0), прогнозування для обсягів виробництва ARIMA(1,0,1), висновки

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	Рощина Н. В., доцент кафедри ЕК		

7. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Інструктаж з техніки безпеки, затверджено тему дипломної роботи.	17.04.2023 –23.04.2023	Виконано
2	Ознайомлення з науковою літературою по темі.	23.04.2023– 30.04.2023	Виконано
3	Написання першого розділу дипломної роботи.	01.05.2023– 07.05.2023	Виконано
4	Написання другого розділу дипломної роботи, розпочато розроблення програмного продукту.	08.05.2023– 14.05.2023	Виконано
5	Доповнення першого та другого розділу дипломної роботи. Оформлення звіту з переддипломної практики.	15.05.2023– 21.05.2023	Виконано
6	Завершення реалізації програмного продукту та написання третього розділу дипломної роботи.	22.05.2023– 28.05.2023	Виконано
7	Написання четвертого розділу дипломної роботи, створення презентації та попередній захист дипломної роботи	29.05.2023– 04.06.2023	Виконано
8	Підготовка до захисту.	05.06.2023– 18.06.2023	Виконано
9	Захист дипломної роботи	19.06.2023 – 23.06.2023	Виконано

Студент Борис ГУЛКЕВИЧ

Керівник Петро БІДЮК

РЕФЕРАТ

Дипломна робота: 112 с., 17 рис., 21 табл., 2 додатки, 25 джерел.

АР, АРКС, АРІКС, АРУГ, ПРОГНОЗУВАННЯ, УАРУГ, ЧАСОВИЙ РЯД, МОРФОЛОГІЧНИЙ АНАЛІЗ.

Об'єкт дослідження: два пов’язані набори даних, де перший – експертні оцінки діяльності підприємства, а другий – вибірка даних, які взяті зі звітів про фінансові результати, а також фінансові коефіцієнти, розраховані на їх основі, що потребують подальшого аналізу із виявленням закономірностей та взаємозалежностей між елементами.

Предмет дослідження: метод морфологічного аналізу, методи та моделі прогнозування часових рядів: моделі авторегресії, авторегресії – ковзного середнього, авторегресії – інтегрованого ковзного середнього, авторегресії з умовною гетероскедастичністю та взаємодію з часовими рядами, підготовкою даних на їх основі.

Мета дослідження: розробка та реалізація програмного продукту для прогнозування результатів діяльності виробничого підприємства якісними та кількісними методами.

Результат дослідження: створення програмного продукту за допомогою мови програмування Python у програмному середовищі PyCharm. За допомогою розробленого програмного продукту було проведено морфологічний аналіз об’єкта, а потім побудовано та описано моделі для прогнозування дохідності компанії та обсягів її виробництва.

ABSTRACT

Bachelor thesis: 112 pages, 17 figures, 21 tables, 2 appendices, 25 sources.

AR, ARMA, ARIMA, ARCH, ANALYSIS, FORECASTING, GARCH, TIME SERIES, MORPHOLOGICAL ANALYSIS.

The object of the study: we have two related data sets. First one is expert evaluations of the enterprise's activity. Second dataset is a sample of data taken from reports of company’s financial results and calculated on their basis financial ratios. This elements required for further analysis of regularities and interdependencies between datasets elements.

Subject of research: method of morphological analysis, methods and models of time series forecasting: autoregression models, autoregression - moving average, autoregression - integrated moving average, autoregression with conditional heteroscedasticity and interaction with time series, data preparation based on them.

The purpose of the research is to develop and implement an information system for forecasting the results of the activity of a manufacturing enterprise by qualitative and several methods.

Research result: creation of a software product using the Python programming language in the PyCharm programming environment. We carry out a morphological analysis of the object with the help of the developed software product. Then we build and describe the models for forecasting the company's profitability and its production volumes.

ЗМІСТ

ВСТУП……...8

1 ОСОБЛИВОСТІ І МЕТОДИ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА……..9

1.1 Загальний огляд задач і методів прогнозування діяльності підприємств 9

1.2 Якісні методи для прогнозування 11

1.3 Кількісні методи для прогнозування 12

1.4 Огляд можливого програмного забезпечення..... 17

1.5 Підготовка даних..... 19

1.6 Висновки до розділу 20

1.7 Постановка завдання дослідження..... 21

2 ДОКЛАДНИЙ ОГЛЯД ОБРАНИХ МОДЕЛЕЙ ТА МЕТОДІВ……...23

2.1 Морфологічний аналіз..... 23

2.2 Огляд моделей авторегресії 32

2.2.1 Модель авторегресії..... 33

2.2.2 Модель з ковзним середнім 34

2.2.3 Модель авторегресії з ковзним середнім..... 34

2.2.4 Моделювання з умовною гетероскедастичністю 40

2.2.5 Узагальнена авторегресійна умовно гетероскедастична модель..... 43

2.3 Засоби для оцінки адекватності моделей 45

2.4 Метрики оцінювання якості прогнозів 49

2.5 Висновки до розділу..... 52

3 ПРОВЕДЕННЯ ПРОГНОЗУВАННЯ ДАНИХ ЗА ДОПОМОГОЮ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ……...53

3.1 Розгляд базових даних..... 53

3.2 Опис платформи для виконання аналізу 57

3.3 Робота з методом морфологічного аналізу 59

3.4 Робота з кількісними моделями..... 68

3.5 Висновки до розділу 74

4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ 75

4.1 Постановка задачі проектування 76
4.2 Обґрунтування функцій програмного продукту 77
4.3 Обґрунтування системи параметрів програмного продукту 80
4.4 Аналіз експертного оцінювання параметрів 82
4.5 Аналіз рівня якості варіантів реалізації функцій 86
4.6 Економічний аналіз варіантів розробки ПП 88
4.7 Вибір кращого варіанту ПП техніко-економічного рівня 95
4.8 Висновки до розділу 96
ВИСНОВКИ……..97
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ……99
ДОДАТОК А Лістинг програми……102
ДОДАТОК Б Презентація……104

ВСТУП

Людина завжди намагалася хоча б наближено передбачити майбутнє щоб скоригувати свої рішення у сторону найбільшої вигоди. Спочатку прийняття рішень засновувалося на власному досвіді та інтуїції, однак з часом виникла потреба у більш чітких інструментах, якими стали різні методи аналізу. Це також було зумовлено ускладненням систем у світі. Так, наприклад, те ж виробництво ставало складнішим і складнішими. Ускладнювався і світ навколо, додаючи більше і більше факторів, потребуючи йти на більші вкладення та залучення додаткових ресурсів. Внаслідок цього почалося ускладнення методів аналізу, випрацьовування чіткіших методик, що дозволяють зменшити невизначеність при прийнятті рішень. У даній роботі буде розглянуто якісні та кількісні методи при роботі з виробничими підприємствами, а також перевірка результатів одним типом методів за допомогою іншого.

У першому розділі було розглянуто що таке прогнозування, його види, методи обробки даних для цього процесу та інших можливостей роботи з цим.

У другому розділі було розглянуто морфологічний аналіз та вибрані методи кількісного аналізу.

У третьому розділі було проведено аналіз та прогнозування показників ПАТ “АвтоКрАЗ”, проведення аналізу обох видів прогнозування та підтвердження одного іншим.

1 ОСОБЛИВОСТІ І МЕТОДИ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА

1.1 Загальний огляд задач і методів прогнозування діяльності підприємств

Для виробничих підприємств аналіз і прогнозування мають дуже велику цінність, тому що їх процеси потребують постійної уваги. Так, власник підприємства бажає визначити ефективність його власності та знайти як покращити її. Для інвесторів аналіз допомагає визначити цінність вкладення у це підприємство, а для клієнтів та замовників це дає можливість зрозуміти, чи варто вести з компанією справу, чи закриється вона так і не виконавши замовлення. Відомий приклад – замовлення вантажівок на підприємстві ПАТ “АвтоКрАЗ”, яке з кожним роком зменшувало кількість виготовлених одиниць і більше починало зривати контракти.

Цілями аналізу діяльності та стану підприємства зазвичай є:

- визначення змін у діяльності підприємства;
- визначення факторів, які впливають на підприємство та його діяльність;
- передбачення того які зміни відбудуться у майбутньому і від яких факторів;
- зниження рівня невизначеностей при прийнятті рішень щодо компанії;
- визначення положення компанії у майбутньому;
- визначення результативності попередніх рішень в компанії;
- визначення напряму у якому приймати майбутні рішення.

Існує дуже багато методів для аналізу, тому буде зосереджено увагу лише на частині під назвою прогнозування.

Прогнозування – це визначення майбутнього стану підприємства на основі причинно-наслідкових зв’язків та закономірностей. Береться поведінка різних процесів раніше, з чого аналізується як вони залежні між собою та впливають один на одного. За допомогою отриманих результатів визначається, яка буде ситуація на підприємстві у майбутньому.

До згадки попередніх прикладів –можна зібрати дані про наявну діяльність “АвтоКрАЗ” і спробувати спрогнозувати, як підприємство почувало б себе у руках старих власників, які нині звинувачують у тому що держава вбиває ефективне виробництво. Тобто визначається наскільки є подібні звинувачення вірними та правильність уже прийнятого рішення. Це один приклад.

За часом прогнози можна поділити на: короткострокові, середньострокові та довгострокові.

- короткострокові - до 1 року – 1,5 роки, базується на якісній та кількісній оцінці змін попиту та пропозиції, об’ємів виробництва, конкурентоспроможності на ринку та індексів цін. При короткостроковому прогнозуванні враховують циклічні та нециклічні сезонні, часові, випадкові фактори;
- середньострокові – 4 – 6 років, характеризують планування вкладення капіталу та рух грошових коштів, виробниче планування, враховують циклічні та нециклічні фактори;
- довгострокові – більше 6 років, характеризують бізнес-планування номенклатури продукції, капіталовкладення, науково-дослідницьких робіт, місце розташування компанії.

Однак, необхідно розуміти, що і як прогнозується. Можливо шукати майбутні значення доходу, виготовлених одиниць, чи навіть зустрічалися моделі щодо кількості працівників. Це все вирішується поставленою задачею та наявними даними. На мою думку цьому випадку розглядаються одні з

основних показників – фінансово-економічний стан компанії. Їх можна назвати основним орієнтиром, який в більшості репрезентує загальний стан речей на підприємстві.

Цю думку можна знайти у багатьох сучасних авторів, як от Джордж Н. Рут III[1] чи Брайан Бірс[2]. Визначимо формальніше поняття для кращого розуміння.

Фінансовий стан підприємства – це показники підприємства, що характеризують фінансові ресурси, їх застосування, конкурентоспроможність та здатність підприємства виконувати власні зобов'язання перед інвесторами, замовниками чи партнерами. Прогнозування цього є складним процесом, що спирається в основному на елементарні моделі.

Також можна спробувати застосувати загальну експертну оцінку та на її основі вибудувати прогноз майбутнього підприємства. Як можна зрозуміти, подібні методи будуть сильно залежати від думки людей та можливих наданих альтернатив.

Обираючи класифікацію за рівнем формалізації, можна поділити методи прогнозу на кількісні та якісні.

1.2 Якісні методи для прогнозування

Якісні методи базуються на експертній оцінці, тобто на суб'єктивних оцінках та інтуїтивно-логічному мисленні. Це призводить до великої неточності подібних засобів. Більше того, технічно "дуже важко" загалом визначити точність подібних прогнозів. Тоді де ж це можна застосувати?

Якісні методи зазвичай застосовують у випадках з високою невизначеністю, коли у нас недостатньо чітких статистичних даних, які дозволяють нам застосувати кількісні методи. Або дані взагалі відсутні.

Однак, ці методи також можна застосовувати для аналізу та прогнозування і у випадку наявності даних. Це може знадобитися, наприклад, для додаткової допомоги у прийнятті рішень чи викритті недоліків, які не виявляться одним фінансовим чи просто кількісним прогнозуванням. У таких випадках якісні методи варто застосувати у тандемі з кількісними, а в кінці порівняти висновки, які можна отримати з обох результатів.

Це вже можна назвати, у певній мірі, передбаченням, але не до кінця, оскільки порівняння висновків проводиться більше якісно ніж кількісно. Дослідження цієї теми я спостерігав у Наталії Дмитрівни Панкратової.

Серед якісних методів прогнозування можна, для прикладу, виділити:

- Метод морфологічного аналізу - шляхом комбінування основних структурних елементів систем або ознак рішень дозволяє знайти найбільшу кількість можливих способів вирішення проблеми та потенційно найкращі з них.
- Метод експертних оцінок - це метод отримання та аналізу інформації про об'єкт прогнозування за допомогою опитування експертів та узагальнення даних
- Метод Дельфі - це метод отримання та аналізу інформації про об'єкт прогнозування за допомогою опитування експертів та узагальнення даних в систематичному порядку. Думки збираються та аналізуються у багатьох ітераціях з метою досягнення консенсусу.

1.3 Кількісні методи для прогнозування

Кількісні методи для прогнозування – методи які засновуються на математичних підходах та методах.

Роботи таких авторів, як Д.-Е. Бестенс, В. М. Ван ден Берга, К. Гренджера, Б. Джордана, Д. Дікі, К. Доугерті, Р. Інгла, Е. Сігела тощо присвячені аналізу та прогнозуванню явищ та процесів. У їхніх наукових роботах розглядаються моделі та методи прогнозування фінансових показників та економічних явищ з детальним описом їхніх характеристик.

Одним з видів кількісних методів є методи на основі часових рядів.

Часовий ряд - це множина спостережень (вимірів) за рівновіддалені проміжки часу. Спостереження, які входять до множини, характеризують поведінку процесу чи об'єкта на тому чи іншому визначеному часовому інтервалі.

Часові ряди ділять на такі компоненти: тренд, сезонна, періодична та випадкова компоненти. Розглянемо детальніше:

- тренд (трендова компонента) зміна процесу (чи явища) протягом певного проміжку часу у низхідному або висхідному напрямку;
- сезонна компонента відповідає за коливання навколо тренду, які повторюються через певні проміжки в часу.
- періодична компонета описує характеристику зміни процесу у низхідному або висхідному напрямі, яка повторюється через деякі проміжки часу чия довжина більша за рік.
- Нерегулярна (випадкова) компонета описує непостійні короткі коливання, що не мають опису закономірностями

При побудові моделі береться тільки сезонна та трендова компоненти.

Методи прогнозування на основі часових рядів засновуються на презумції, що той закон, за яким часовий ряд змінювався раніше, буде актуальним у майбутньому. Їхніми елементами є компоненти часового ряду, які виражаються коефіцієнтами, за якими потім рахують оцінку показника або явища. За тим як це шукають моделі поділяються на:

- адитивні – базуються на розрахунку суми всіх компонент щоб оцінити явище;

– мультиплікативні – базуються на розрахунку добутку всіх компонент щоб оцінити явище.

Також бувають негативні випадки коли при формування прогнозованого явища не вдається помітити стійкі закономірності. У таких випадках можна застосувати методи фільтрації чи згладжування. Основна ціль таких дій – замінити наявні значення іншими, прогнозними, які матимуть менше зайвої інформації та шуму.

Поширені різні типи фільтрів для різних задач. Так для усунення трендів, поділ на складові компоненти часових дядів, стиснення та згладжування даних і оцінювання спектрів часто застосовують цифрові фільтри. Їх можна описати як систему, що приймає на вхід дискретний набір даних x_t та перетворює його у дискретний набір даних y_t використовуючи лінійні співвідношення виду:

$$y_t = \sum_{i=-\infty}^{+\infty} \alpha_i x_{t-i}.$$

Існують такі види цифрових фільтрів:

1. Експоненційне згладжування – фільтр, на відфільтровані значення якого впливають всі попередні значення.

Записується таким рівнянням:

$$y_t = \theta x_t + (1-\theta) y_{t-1},$$

де x_t – знайдене значення у момент t;

y_t – відфільтроване значення у момент t;

θ – коефіцієнт фільтрації, $\theta \epsilon [0; 1]$;

Цей фільтр здатен легко усувати великі “стрибки” у даних, а також не потребує зберігати у пам’яті певну кількість попередніх значень.

2. Фільтр ковзного середнього – фільтр, на відфільтровані значення якого впливають лише останні N значень. Він засновується на розрахунку середнього на певному ковзному інтервалі:

$$y_t = \frac{1}{N}\sum_{i=0}^{N-1} x_{t+i}.$$

3. Медіанний фільтр – фільтр, що використовує відсортовані за зростанням останні N значень, серед яких як фільтроване обирається значення з середини відсортованого масиву.

Також варто згадати ще один популярний фільтр – фільтр Калмана. Його застосовують для знаходження оптимальних оцінок залишків та побудови короткострокових прогнозів, а також у пост-обробленні даних. Серед переваг цього алгоритму: простота реалізації, автоматичний підбір посилення у залежності від помилки прогнозу, можливість в реальному часі слідкувати за досліджуваною системою.

На першому етапі відбувається передбачення. В основі цього лежить модель, за якою потім будується прогноз значення використовуючи попереднє значення цього параметра:

$$x_k = F_k\widehat{x_{k-1}} + B_k u_k + w_k,$$

Дана формула називається рівнянням переходу стану. Тепер розберемо її:

x_k – прогноз поточного значення системи на k-му кроці;

F_k – матриця, яка визначає залежність поточного стану системи від попереднього;

$\widehat{x_{k-1}}$ – значення системи на $k-1$ кроці;

w_k – вектор гаусівського шуму;

B_k – матриця керуючого впливу на систему;

u_k – вектор керування.

Коваріаційну матрицю похибок оцінок значень $\widehat{x_k}$ у поточний момент часу P_k можна знайти за допомогою вірності:

$A_k = F_k\widehat{A_{k-1}}F_k^T + Q_k$;

$\widehat{A_{k-1}}$ – коваріаційна матриця помилок, яка змінена на попередньому кроці для уточнення;

Q_k – коваріаційна матриця гаусівського шуму.

На другому етапі дії фільтру Калмана уточнюється та відповідно скориговується значення системи. Для цього виконується певний набір кроків:

1) знаходження відхилення відомого стану системи z_k від передбаченого значення x_k:

$\overline{y_k} = z_k - H_k x_k$, H_k – матриця залежності фактичного стану системи від прогнозу,

2) визначення матриці оптимальних коефіцієнтів для посилення Калмана K_k :

$$K_k = A_k H_k^T (H_k A_k H_k^T + R_k)^{-1},$$

де R_k – коваріаційна матриця шуму системи;

3) корегування оцінки системи:

$\widehat{x_k} = \widehat{x_{k-1}} + K_k\overline{y_k}$, де $\widehat{x_{k-1}}$ та $\widehat{x_k}$ – скориговані значення на $k-1$ та k кроках.

Тоді коригування дисперсії стану системи буде відбуватися за такою формулою:

$$\widehat{A_k} = (1 - K_k H_k) A_k.$$

1.4 Огляд можливого програмного забезпечення

На сьогодні існує величезна кількість програмного забезпечення для реалізації прогнозування. Це стосується в основному кількісних методів, але так само існують програми для якісних методів.

До прикладу останніх можна віднести eDelphi, яка забезпечує роботу з методом Делфі з 1998 року. Одною з переваг цього ПЗ є його базування на принципах open source (відкритого коду). Базовий функціонал доступний на безкоштовній основі, але більше можливостей можна відкрити лише за додаткову оплату різних планів підтримки[3].

Для підтримки морфологічного аналізу можна застосовувати платформу MA/Carma (Computer-Aided Resource for Morphological Analysis). Вона служить платформою для розробки морфологічних моделей висновків і для створення лабораторій сценаріїв і стратегій. Серія програмного забезпечення MA була вперше розроблена в 1995 році, наразі йде її п'ята версія[4]. Розроблена на мові програмування C++ у Шведському морфологічному товаристві.

Тепер щодо програмного забезпечення для кількісних методів.

EViews — це статистичний пакет для операційної системи Microsoft Windows, який використовують для орієнтованого на часові ряди аналізу (в основному економетричного). Це програмне забезпечення розроблене

компанією Quantitative Micro Software, яка наразі є частиною IHS. Використання потребує придбання коштовної ліцензії. Версія Eviews 1.0 побачила великий світ у березні 1994 року та замінила MicroTSP.

Згадана вище MicroTSP є похідною від TSP ("Time Series Processor" – обробник часових рядів) - мови програмування для аналізу та прогнозування економетричних даних[5]. Розробка системи велася ще з 60-х років минулого століття, її компанія розробник була заснована у 1978 році, а у 1983 році проєкт розділився на дві гілки, одна з яких стала згаданою MicroTSP, а інша продовжила розроблятися як TSP. Хоча у багатьох джерелах остання версія датується 2009 роком, вдалося знайти її дані за оновлення у 2016 році[6].

Серед сучасних зразків програмного забезпечення для прогнозування можна виділити ForecastX. Розробники програми стверджують, що їхній продукт може надати прогноз буквально до хвилини та об'єднує в собі можливості Microsoft Excel разом з простим для використання інтерфейсом. Якщо вірити описам – ця система передбачає великий ступінь автоматизації та спрощення роботи між собою для аналітиків компанії.

Згаданий вище Microsoft Excel є табличним процесором, або ж програмою для роботи з електронними таблицями. Це програмне забезпечення глибоко використовується у величезній кількості галузей по всьому світу. І, серед можливого функціоналу є і варіант створення передбачень на основі часових рядів. Програма постачається у складі пакету Microsoft Office та потребує придбання ліцензії, хоча і існує безкоштовна онлайн-версія. Також варто зазначити про існування безкоштовних аналогів Excel, більшість з яких також здатні на прогнозування.

Одними з найпотужніших і ширших інструментів в області нині є мови програмування R та Python.

R — мова програмування та водночас програмне середовище для статистичних обчислень, аналізу та зображення даних. Передбачається величезна кількість різноманітних бібліотек, які дають можливість працювати з майже будь якою типовою задачею з прогнозування. Зручний

інтерфейс і досить простий синтаксис не вимагають від користувача дуже великих навичок у софтдеві.

Python – інтерпретована високорівнева мова програмування зі строгою динамічною типізацією. За час свого існування вона обросла величезною кількістю бібліотек, що дають дуже корисний та гнучкий інструментарій для роботи у різних сферах. Не оминуло це і аналіз даних та прогнозування, де мова вже встигла набути хорошої популярності. Одним з головних недоліків мови є її швидкість роботи, але навіть це можливо обійти за допомогою сучасних розширюючих бібліотек. На сьогодні це один з самих багатогранних інструментів для прогнозування.

1.5 Підготовка даних

Одним з найважливіших елементів прогнозування є вхідні статистичні дані. На жаль, у реальному житті не завжди стається так, що все повністю підходить для роботи чи здатні забезпечити прийнятний результат.

На практиці найпоширенішими є такі проблеми:

- шуми та викиди;
- дублікати даних;
- пропущені значення;
- аномальні значення.

Тому для роботи дані проходять певний процес підготовки.

В цілому процес підготовки можна виділити такі етапи:

1) нормалізацію даних для зменшення впливу шумів та викидів. Тут можна застосувати те саме нормування.
2) за потреби – дискретизацію даних, яка перетворює неперервні параметри у дискретні. Це буває потрібно для застосування деяких моделей (баєсівські мережі).

3) заповнення пропусків – заповнення пропущених даних за допомогою різних методів.

Один з часто використовуваних методів нормалізації даних полягає у застосуванні логарифмування. Для заміни пропущених значень можна керуватися різними методами.

Першим методом можна назвати просте ігнорування пропусків. Це можливо тоді коли їх небагато та допомагає уникнути спотворення даних. Також можна просто видалити частину з пропущеними даними.

До так само специфічних методів можна віднести і заміну пропусків нулями. Це є простим методом, але навіть так він вже спроможний дещо видозмінити та спотворити наявні дані.

Серед менш радикальних можна взяти заміну методом найближчих сусідів. Він ґрунтується на тому, що елемент належить до того самого класу що і більшість об’єктів які близькі до пропущеного за певними параметрами. Це потребує введення метрики на просторі, яку можна буде використати для вимірювання “відстані між сусідами”.

Також можна згадати метод заміни середнім або найбільш частим значенням. Спочатку рахується для якісних характеристик моду, а для кількісних – середнє. Потім цими значеннями заповнюються пропуски. Цей метод намагаються не використовувати через те що він може послабити міри асоціації та кореляції між елементами.

1.6 Висновки до розділу

Отже, прогнозування – один з важливих елементів при аналізі діяльності виробничого підприємства, що суттєво впливає на прийняття рішень та зменшує невизначеності у роботі. Можна застосовувати як і

кількісні методи (що спираються на статистичні дані), так і якісні методи (що спираються на думку експертів).

Одним з хороших узагальнюючих факторів за якими можна визначити успішність підприємства є фінансово-економічні показники, для аналізу та прогнозування яких на основі статистичних даних дуже добре підходять кількісні методи.

Оскільки у нас не завжди є достатньо прийнятна кількість даних, то можна застосувати і якісні методи аналізу. Хоча вони і є досить неточними, спираючись на думку експертів, їх можна сміливо застосувати у прогнозуванні разом з кількісними методами.

Серед кількісних методів для аналізу фінансово-економічних показників чудово підійдуть методи на базі часових рядів. Вони працюють зі спостереженнями за рівновіддалені проміжки часу. Однак, інколи визначити необхідні для часових рядів елементи не вдається. Тоді намагаються застосовувати різні методи фільтрації, одним з найпоширеніших та найпростіших в реалізації є фільтр Калмана.

Також перед використанням дані мають пройти й іншу підготовку, як от нормалізацію, заповнення пропусків і так далі. Це допоможе усунути різні спотворення, з якими зазвичай приходять реальні дані.

На сьогодні існує велика кількість інструментів для реалізації прогнозування, але одним з найуніверсальніших з ним є мова програмування Python та її широкий інструментарій.

1.7 Постановка завдання дослідження

Ціллю даної дипломної роботи є розгляд різних моделей та методів для прогнозування стану виробничого підприємства. Спочатку здійснюватиметься загальний прогноз стану за допомогою методу

морфологічного аналізу. Це дозволить нам, за допомогою експертної оцінки спрогнозувати можливі варіанти та найбільш реальний приблизний результат для підприємства. Застосування цього методу пояснюється невеликою кількістю даних, яку вдалося отримати, їхньою неточністю та загалом інтересом для порівняння результатів.

Далі, за допомогою кількісних методів буде розраховано фінансово-економічний стан підприємства. Як було розглянуто раніше у першій частині, це дозволить нам отримати досить добре репрезентуючий загальний прогноз по підприємству. Основою для роботи виступатимуть методи на основі часових рядів.

Для застосування цих методів використовуватимуться отримані статистичні дані. У будь якому випадку, дані буде проаналізовано та підготовлено до процесу прогнозування. Це допоможе досягнути більшої точності прогнозів.

2 ДОКЛАДНИЙ ОГЛЯД ОБРАНИХ МОДЕЛЕЙ ТА МЕТОДІВ

2.1 Морфологічний аналіз

Метод морфологічного аналізу було презентовано у 1969 році швейцарським астрофізиком Фріцем Цвікі (1898-1974 рр). Головною ідеєю методу є розв’язання певної проблеми проблеми за допомогою комбінування основних структурних елементів систем, або інших частин того чи іншого цілого (за завданням). За допомогою цього можна обрати найкращу для нашої задачі ситуацію.

З точки зору прогнозування, це можна взяти яе комбінування різних можливих станів певних частин нашої системи і за допомогою цього визначити, яка комбінація потенційного майбутнього нам більше підходить.

Дослідження за допомогою цього методу поділяється на два великі кроки. На першому проводиться морфологічна класифікація даної нам множини об’єктів. Для цього розглядається об’єкт зі сторони різних його аспектів, тобто класів. Також, на цьому етапі беруть не лише відомі, а й гіпотетичні системи, що можна віднести до нашого об’єкта.

На другому кроці оцінюються різні об’єкти кожного досліджуваного класу та відбираємо ті, які найбільше підходять для поставленого нам завдання.

Одною з важливих умов для об’єкта морфологічного дослідження є наявність великої кількості конфігурацій, або іншими словами – великої кількості можливих альтернатив. Причинами цього зазвичай є те що

досліджувана система розглядається у майбутньому та/або за задачею необхідно визначити, яка буде конфігурація системи та які рішення будуть прийняті відповідно до цього. Тобто, можна побачити, що за самою базою до розгляду задачі можна визначити, що метод морфологічного аналізу можна застосувати для прогнозування діяльності виробничого підприємства.

Загально, існує три види об’єктів опису морфологічної таблиці: об’єкт, дії та стану.

Першим можна взяти опис об’єкта, який власне описує певну систему чи об’єкт. Подібний випадок трапляється саме для синтезу нових речей чи покращення існуючих. Для подібних задач спочатку і замислювався метод морфологічного аналізу.

Другим іде опис дії, який створюється для опису певної взаємодії між об’єктами, або деякої події. У такому випадку відомий предмет дослідження, а невизначеність полягає у стані системи, тих системах чи предметах що беруть участь у події та характеристиках перебігу самої події. Подібні дослідження типово стосуються виведення на ринок нового продукту чи різні аварійні ситуації.

При описі стану ми маємо відому систему або об’єкт, а невизначеність стосується того що відбувається, або буде відбуватися на об’єкті. Основними параметрами при такому дослідженні виступатимуть характеристики та елементи системи. Прикладом подібного, власне, може бути прогноз стану підприємства. Тобто, ми у рамках дослідження методу морфологічного аналізу визначатимемо різні властивості стану підприємства.

Властивості стану підприємства, за якою ми проводитимемо класифікацію різновидів об’єкта, називають **характеристичним параметром** і позначається $F_i, i \in \overline{1, N}$, де N – кількість параметрів, а i – номер параметра.

У кожного з параметрів наявні **альтернативи**, які позначають взаємовиключні стани або значення, які можуть набувати відповідні

характеристичні параметри. Позначається це $a_j^{(i)} \in \overline{1, n_i}$, де n_i – кількість альтернатив у i-го параметра, j – номер альтернативи.

Заявлена вище **морфологічна таблиця** – множина характеристичних параметрів системи чи об'єкта, кожен з яких $F_i, i \in \overline{1, N}$ характеризується через множину альтернатив $a_j^{(i)} \in \overline{1, n_i}$.

Загальний вигляд морфологічної таблиці наведено у таблиці 2.1.

Таблиця 2.1 – Загальний вигляд морфологічної таблиці

F_1	F_2	F_3	…	F_N
$a_1^{(1)}$	$a_1^{(2)}$	$a_1^{(3)}$	…	$a_1^{(N)}$
$a_2^{(1)}$	$a_2^{(2)}$	$a_2^{(3)}$	…	$a_2^{(N)}$
…	…	…	…	…
$a_{n_1}^{(1)}$	$a_{n_2}^{(2)}$	$a_{n_3}^{(3)}$	…	$a_{n_N}^{(N)}$

Також є поняття **конфігурації морфологічної таблиці** – множину, де є по одній можливій альтернативі для кожного з параметрів морфологічної таблиці. Позначається через: $s = \left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}$.

З введенням позначень, ми можемо вказати загальний алгоритм для роботи морфологічного аналізу:

1) формулювання завдання та визначення характеристичних параметрів;
2) поділ параметрів на їх значення (визначення можливих альтернатив);
3) побудова морфологічної таблиці
4) зменшення кількості можливих конфігурацій за допомогою матриці взаємоузгодженості;
5) оцінювання і вибір з таблиці найкращої або найімовірнішої конфігурації.

Формалізація задачі:

За допомогою обробки експертного оцінювання ми отримуємо початкові наближення ймовірності $p_j^{(i)}$ для кожної з альтернатив $a_j^{(i)}$. Задача: знайти ймовірності $p_j'^{(i)}$ настання кожної з альтернатив $a_j^{(i)}$.

Процес експертної оцінки $\widetilde{p_j^{(i)}}$ відбувається за допомогою такої шкали рівнів з таблиці 2.2.

Таблиця 2.2 – Шкала рівнів

Рівень	**Якісна характеристика**	**Кількісна характеристика**
1	Практично неможливо	$[0, 0.1]$
2	Дуже мала ймовірність	$[0.1, 0.25]$
3	Мала ймовірність	$[0.25, 0.4]$
4	Середня ймовірність	$[0.4, 0.6]$
5	Велика ймовірність	$[0.6, 0..75]$
6	Дуже велика ймовірність	$[0.75, 0.9]$
7	Практично гарантовано	$[0.9, 1]$

Далі ми пронормуємо отримані оцінки альтернатив за формулою:

$$p_j^{(i)} = \frac{\widetilde{p_j^{(i)}}}{\sum_{j=1}^{n_i} \widetilde{p_j^{(i)}}}.$$

Одразу виникає питання про те, чому ми зразу не застосуємо отримані та оброблені оцінки. Головною проблемою цих результатів є те, що вони не враховують взаємозв'язки між параметрами. Тобто, отримані в результаті експертного оцінювання дані є наближеними, не репрезентують всієї ситуації та загалом не дають достатньо точної відповіді для заданої задачі.

Далі нам потрібно врахувати зв'язки між параметрами морфологічної таблиці. Для цього у методі морфологічного аналізу використовується числова матриця взаємозв'язків. У ній для кожної пари альтернатив $a_{j_1}^{(i_1)}$ та $a_{j_2}^{(i_2)}$ різних параметрів F_{i_1} та F_{i_2} присвоюється оцінка $c_{i_1 j_1 i_2 j_2} \in [-1,1]$. Оцінка відбувається за такою шкалою з таблиці 2.3.

Таблиця 2.3 – Шкала для оцінки взаємозв'язків

Оцінка	**Пояснення**
-1	Повністю несумісні альтернативи, конфігурація з якими неможлива
$(-1; 0)$	Частково несумісні альтернативи, наявність одної з них у конфігурації зменшує ймовірність наявності іншої
0	Альтернативи незалежні та не мають впливу одна на одну
$(0; 1)$	Пов'язані альтернативи, наявність одної з них у конфігурації збільшує ймовірність наявності іншої
1	Повністю пов'язані альтернативи, наявність одної з них у конфігурації означає і наявність іншої

Далі розглядається вирішення даної задачі, для чого записується рівняння ймовірностей. Розглядається загальний варіант з довільною кількістю параметрів.

$$p_1^{(1)} = \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} \dots \sum_{j_N=1}^{n_N} P\left(a_1^{(1)} \middle| \left\{a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right) P\left(\left\{a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right)$$

$$= \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} \dots \sum_{j_N=1}^{n_N} P\left(\left\{a_1^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right) =$$

$$= \sum_{j_2=1}^{n_2} \sum_{j_3=1}^{n_3} \dots \sum_{j_N=1}^{n_N} P\left(\left\{a_1^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\} \middle| a_{j_2}^{(2)}\right) p_{j_2}^{(2)}.$$

Для знаходження значень виразів $P\left(\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\} \middle| a_{j_k}^{(k)}\right)$ використовується такий спосіб:

$$P\left(\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\} \middle| a_{j_1}^{(1)}\right) =$$

$$= \frac{P'\left(\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\} \middle| a_{j_1}^{(1)}\right)}{\sum_{k_2=1}^{n_2} \sum_{k_3=1}^{n_3} \dots \sum_{k_N=1}^{n_N} P'\left(\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\} \middle| a_{j_1}^{(1)}\right)},$$

У якому знаходиться:

$$P'\left(\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\} \middle| a_{j_1}^{(1)}\right) = \prod_{m=2}^{N} p_{j_m}^{\prime(m)} \cdot \prod_{m=1}^{N-1} \prod_{l=m+1}^{N} \left(c_{mj_m,lj_l} + 1\right),$$

Тоді буде отримана у загальному виді система рівнянь виду:

$$\begin{cases} \mathrm{p}_1^{(1)} = \sum_{\mathrm{j}_2=1}^{\mathrm{n}_2}\sum_{\mathrm{j}_3=1}^{\mathrm{n}_3}\dots\sum_{\mathrm{j}_N=1}^{\mathrm{n}_N} \mathrm{P}\left(\left\{\mathrm{a}_1^{(1)}, \mathrm{a}_{\mathrm{j}_2}^{(2)}, \mathrm{a}_{\mathrm{j}_3}^{(3)}, \dots, \mathrm{a}_{\mathrm{j}_N}^{(\mathrm{N})}\right\}\middle|\mathrm{a}_{\mathrm{j}_2}^{(2)}\right)\mathrm{p}_{\mathrm{j}_2}^{(2)}\ ;\dots;\ \mathrm{p}_{\mathrm{n}_1}^{(1)} = \sum_{\mathrm{j}_2=1}^{\mathrm{n}_2}\sum_{\mathrm{j}_3=1}^{\mathrm{n}_3}\dots\sum_{\mathrm{j}_N=1}^{\mathrm{n}_N} \mathrm{P}\left(\left\{\mathrm{a}_{\mathrm{n}_1}^{(1)}, \mathrm{a}_{\mathrm{j}_2}^{(2)}, \mathrm{a}_{\mathrm{j}_3}^{(3)}, \dots, \mathrm{a}_{\mathrm{j}_N}^{(\mathrm{N})}\right\}\middle|\mathrm{a}_{\mathrm{j}_2}^{(2)}\right)\mathrm{p}_{\mathrm{j}_2}^{(2)}\ ; \\ \mathrm{p}_1^{(2)} = \sum_{\mathrm{j}_1=1}^{\mathrm{n}_1}\sum_{\mathrm{j}_3=1}^{\mathrm{n}_3}\dots\sum_{\mathrm{j}_N=1}^{\mathrm{n}_N} \mathrm{P}\left(\left\{\mathrm{a}_{\mathrm{j}_1}^{(1)}, \mathrm{a}_1^{(2)}, \mathrm{a}_{\mathrm{j}_3}^{(3)}, \dots, \mathrm{a}_{\mathrm{j}_N}^{(\mathrm{N})}\right\}\middle|\mathrm{a}_3^{(2)}\right)\mathrm{p}_{\mathrm{j}_3}^{(3)}\ ;\dots;\ \mathrm{p}_{\mathrm{n}_1}^{(1)} = \sum_{\mathrm{j}_2=1}^{\mathrm{n}_1}\sum_{\mathrm{j}_3=1}^{\mathrm{n}_3}\dots\sum_{\mathrm{j}_N=1}^{\mathrm{n}_N} \mathrm{P}\left(\left\{\mathrm{a}_{\mathrm{j}_1}^{(1)}, \mathrm{a}_{\mathrm{n}_2}^{(2)}, \mathrm{a}_{\mathrm{j}_3}^{(3)}, \dots, \mathrm{a}_{\mathrm{j}_N}^{(\mathrm{N})}\right\}\middle|\mathrm{a}_{\mathrm{j}_3}^{(3)}\right)\mathrm{p}_{\mathrm{j}_3}^{(3)}\ ; \\ \dots \\ \mathrm{p}_1^{(\mathrm{N})} = \sum_{\mathrm{j}_1=1}^{\mathrm{n}_1}\sum_{\mathrm{j}_2=1}^{\mathrm{n}_2}\dots\sum_{\mathrm{j}_{N-1}=1}^{\mathrm{n}_{N-1}} \mathrm{P}\left(\left\{\mathrm{a}_{\mathrm{j}_1}^{(1)}, \mathrm{a}_{\mathrm{j}_2}^{(2)}, \mathrm{a}_{\mathrm{j}_3}^{(3)}, \dots, \mathrm{a}_1^{(\mathrm{N}-1)}\right\}\middle|\mathrm{a}_{\mathrm{j}_1}^{(1)}\right)\mathrm{p}_{\mathrm{j}_1}^{(1)}\ ;\dots;\ \mathrm{p}_{\mathrm{n}_1}^{(1)} = \sum_{\mathrm{j}_1=1}^{\mathrm{n}_2}\sum_{\mathrm{j}_2=1}^{\mathrm{n}_3}\dots\sum_{\mathrm{j}_{N-1}=1}^{\mathrm{n}_{N-1}} \mathrm{P}\left(\left\{\mathrm{a}_{\mathrm{j}_1}^{(1)}, \mathrm{a}_{\mathrm{j}_2}^{(2)}, \mathrm{a}_{\mathrm{j}_3}^{(3)}, \dots, \mathrm{a}_{\mathrm{n}_N}^{(\mathrm{N}-1)}\right\}\middle|\mathrm{a}_{\mathrm{j}_1}^{(1)}\right)\mathrm{p}_{\mathrm{j}_1}^{(1)}; \\ \sum_{j=1}^{n_1} p_j^{(1)} = 1;\dots;\ \sum_{j=1}^{n_N} p_j^{(N)} = 1 \end{cases}$$

Розв'язком цієї системи буде:

$$\mathrm{p}_{\mathrm{j_k}}^{(i_k)} = \frac{\mathrm{p}_{\mathrm{j_k}}^{\prime(i_k)} \sum_{\mathrm{j}_1=1}^{\mathrm{n}_1}\dots\sum_{\mathrm{j}_{k-1}=1}^{\mathrm{n}_{k-1}}\sum_{\mathrm{j}_{k+1}=1}^{\mathrm{n}_{k+1}}\dots\sum_{\mathrm{j}_N=1}^{\mathrm{n}_N} C_{\mathrm{j}_1\mathrm{j}_2\dots\mathrm{j}_N} p_{j_1}^{\prime(i_1)}\dots p_{j_{k-1}}^{\prime(i_N)} p_{j_{k+1}}^{\prime(i_N)}\dots p_{j_N}^{\prime(i_N)}}{\sum_{\mathrm{j}_1=1}^{\mathrm{n}_1}\dots\sum_{\mathrm{j}_1=1}^{\mathrm{n}_1} C_{\mathrm{j}_1\mathrm{j}_2\dots\mathrm{j}_N} p_{j_1}^{\prime(i_1)}\dots p_{j_N}^{\prime(i_N)}},$$

Де: $C_{\mathrm{j}_1\mathrm{j}_2\dots\mathrm{j}_N} = \prod_{m=1}^{N-1}\prod_{l=m+1}^{N}\left(c_{mj_m,lj_l}+1\right)$,

Тоді отримується ймовірність конфігурації $s = \left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}$ у вигляді:

$$P(s) = \frac{C_{\mathrm{j}_1\mathrm{j}_2\dots\mathrm{j}_N} p_{j_1}^{\prime(i_1)}\dots p_{j_N}^{\prime(i_N)}}{\sum_{\mathrm{j}_1=1}^{\mathrm{n}_1}\dots\sum_{\mathrm{j}_1=1}^{\mathrm{n}_1} C_{\mathrm{j}_1\mathrm{j}_2\dots\mathrm{j}_N} p_{j_1}^{\prime(i_1)}\dots p_{j_N}^{\prime(i_N)}}.$$

Звідси ймовірність кожної альтернативи $\mathrm{a}_j^{(\mathrm{i})}$ знаходиться як сума ймовірностей всіх конфігурацій, які включають дану альтернативу. Результат можна записати у таблиці 2.4.

Таблиця 2.4 – Готова морфологічна таблиця першого етапу

F_1	F_2	F_3	…	F_N
$p_1^{(1)}$	$p_1^{(2)}$	$p_1^{(3)}$	…	$p_1^{(N)}$
$p_2^{(1)}$	$p_2^{(2)}$	$p_2^{(3)}$	…	$p_2^{(N)}$
…	…	…	…	…
$p_{n_1}^{(1)}$	$p_{n_2}^{(2)}$	$p_{n_3}^{(3)}$	…	$p_{n_N}^{(N)}$

Ці значення можна використовувати для визначення найбільш критичних станів параметрів досліджуваного об'єкта, сортування цих станів за ймовірністю виникнення, вибору найбільш ймовірних конфігурацій.

Також морфологічний аналіз буває двохетапним, коли на першому прогнозуються різні варіації сценаріїв, а на другому видаються найімовірніші стратегії прийняття рішень чи зміни стану. Це використовується у системах прийняття рішень.

Спочатку створюється таблиця з альтернативами до стратегії, аналогічно до того що було на першому етапі. Для зв’язку створеної та існуючої з попереднього етапу таблиці нам необхідно використати числову матрицю узгодженості. Якщо формально то це можна описати, що кожній парі альтернатив $a_{j1}^{(i1)}, a_{j2}^{(i2)}$ параметрів $F1_{i1}, F2_{i2}$ таблиць двох етапів морфологічного аналізу присвоюється оцінка $c_{i1j1i2j2} \in [-1,1]$. Значення для оцінювання вказують експерти за шкалою оцінки, яка вказана у таблиці 2.5.

Таблиця 2.5 – Шкала оцінки для другого етапу

Оцінка	**Пояснення**
-1	Альтернатива параметра другого етапу повністю не підходить при виборі відповідної альтернативи параметра першого етапу
$(-1,0)$	Вибір альтернативи параметра першого етапу зменшує ефективність альтернативи параметра другого етапу
0	Вибір параметра другого етапу не залежить від вибору параметра першого етапу
$(0,1)$	Вибір альтернативи параметра першого етапу збільшує ефективність альтернативи параметра другого етапу
1	Альтернатива параметра другого етапу повністю підходить при виборі відповідної альтернативи параметра першого

	етапу

Значення умовної результативності альтернативи $a_j^{(i)}, i \in [N+1, N+N']$ при конфігурації отриманої на першому етапі морфологічної таблиці $\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}$ буде:

$$R\left(a_j^{(i)} \middle| \left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right) = \frac{p_j'^{(i)} \cdot \prod_{m=1}^{N}\left(c_{mj_m,ij}+1\right)}{\sum_{k=1}^{n_i}\left(p_j'^{(i)} \cdot \prod_{m=1}^{N}\left(c_{mj_m,ij}+1\right)\right)},$$

За відсутності додаткової інформації значення буде виходити:

$$p_j'^{(i)} = \frac{1}{N}, i \in [N+1, N+N'].$$

Очікувана результативність деякої альтернативи $a_j^{(i)}$:

$$R_j^{(i)} = \sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\dots\sum_{j_N=1}^{n_N} R\left(a_j^{(i)} \middle| \left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right) P\left(\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right),$$

Тепер запишемо відстань до найефективнішого значення таблиці другого етапу – альтернативну оцінку:

$$W_j^{(i)} = \sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\dots\sum_{j_N=1}^{n_N}\left(R\left(a_{max_j}^{(i)} \middle| \left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right) - R\left(a_j^{(i)} \middle| \left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right)\right) \cdot P\left(\left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right) =$$

$$= R_{max}^{(i)} - R_j^{(i)},$$

Де $a_{max_j}^{(i)}$ – альтернатива при виборі якої $R\left(a_j^{(i)} \middle| \left\{a_{j_1}^{(1)}, a_{j_2}^{(2)}, a_{j_3}^{(3)}, \dots, a_{j_N}^{(N)}\right\}\right)$ буде з максимальним значенням

2.2 Огляд моделей авторегресії

Регресія – функціональна залежність математичного сподівання залежної змінної від однієї або декількох інших пояснювальних змінних.

Якщо простіше, то багато процесів (у тому числі економічних) можна представити з деяких кількісних характеристик. Встановлюється між цими параметрами певна залежність, та вона представляється у аналітичній формі. Це дозволяє не лише показати зв’язок досліджуваного параметра з іншими факторами за допомогою графіка, а й записати його за допомогою емпіричної формули та побудувати прогноз досліджуваної характеристики.

Моделі авторегресії – це моделі що “повертаються до себе”. Їх побудова засновується на взаємозалежності один від одного рівнів одного і того ж самого ряду, що є властивістю економічних процесів. Також, для такого випадку не є обов’язковою умова нормальності розподілу нашого ряду.

Авторегресійні моделі активно застосовуються для опису стаціонарних випадкових процесів. Для часових рядів, які описують ці процеси, характерний розвиток, що проходить у стабільних незмінних умовах без вираженої тенденції, тому незмінними від часу залишаються і ймовірнісні властивості рядів. Також сталість при зсуві часу захищають функціх розподілу.

2.2.1 Модель авторегресії

У даній “початковій” моделі поточне значення процесу $y = y_t, t \geq 1$ представляється у вигляді лінійної комбінації скінченної кількості попередніх значень процесу та білого шуму ε_t:

$$y_t = \alpha_t y_{t-1} + \alpha_t y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t,$$

де ε_t не корелює з лагами y_t.

Це називають авторегресією порядку p та позначається $AR(p)$.

За допомогою формули $\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p$ можна визначити так званий лаговий оператор L. Тоді модель авторегресії записується:

$$\varepsilon_t = y_t\left(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p\right) = \alpha(L) y_t,$$

Тепер записується умова стаціонарності. Для першого порядку вид авторегресії:

$$AR(1): y_t = \alpha_1 y_{t-1} + \varepsilon_t,$$

де ε_t – білий шум у якого $cov(\varepsilon_t, y_{t-1}) = 0, \alpha_1 = \alpha$.

Тоді з дисперсії двох частин: $\sigma_y^2 = \alpha\sigma_y^2 + \sigma_\varepsilon^2$, з чого умова стаціонарності буде $|\alpha| < 1$.

Для другого порядку $AR(2): y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$ умова стаціонарності буде $\alpha_1 \pm \alpha_2 < 1. -1 < \alpha_2 < 1.$

2.2.2 Модель з ковзним середнім

Модель ковзного середнього, вона ж КС та moving average(MA), описує лінійну залежність процесу y_t від q його попередніх значень ε:

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q},$$

Якщо взяти оператор ковзного середнього через , то модель буде записана як:

$$y_t = \beta(L)\varepsilon_t,$$

Де $\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_q L^q$.

Якщо на параметри β не накладається обмежень, то процес $KC(q)$ є стаціонарним. Математичне сподівання та дисперсія будуть відповідно:

$$E(y_t) = 0 \text{ та } \sigma_0^2 = \left(1 + \beta_1^2 + \cdots + \beta_q^2\right)\sigma_\varepsilon^2 = C\sigma_\varepsilon^2$$

2.2.3 Модель авторегресії з ковзним середнім

Дана модель є включенням і елементів авторегресії, і ковзне середнє. Це дозволяє за допомогою меншої кількості параметрів оцінити характеристики часового ряду. Процес назвали змішаним процесом авторегресії-ковзного середнього і позначається українською АРКС(p,q) та англійською ARMA(p,q):

$$y_t = \alpha_t y_{t-1} + \alpha_t y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t - \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q},$$

$$y_t\left(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p\right) = \varepsilon_t\left(1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_q L^q\right),$$

$$\beta(L)\varepsilon_t = \alpha(L) y_t,$$

де $\alpha(L)$ – оператор авторегресії, а $\beta(L)$ – оператор ковзного середнього.

З часом, з'явився ще один тип моделей – комбінована авторегресійна модель з проінтегрованим ковзним середнім. Вона позначається АРІКС(p, d, q) (або англійською $ARIMA(p, d, q)$) та застосовується якщо ряд зводиться до стаціонарного після того як взяти d послідовних різниць.

У окремий клас подібні нестаціонарні ряди виділили Бокс та Дженкінс. Вони визначили їх як ті, що допомогою взяття послідовних різниць можна привести до виду АРКС.

Підбір моделі АРІКС для певного ряду спостережень проходить за чотири етапи:

1) Вибір моделі що відповідає реальному процесу найбільше;
2) Використання методів регресії для отримання включених у модель оцінених параметрів;
3) Перевірка моделі на адекватність із застосуванням тестів на автокореляцію залишків, нормальний розподіл та якість специфікації моделі;
4) Побудова прогнозу за допомогою отриманої моделі.

Розглянемо питання оцінювання невідомих параметрів моделі АРКС(p, q). Для цього зазвичай використовують метод найменших квадратів (МНК), звичайний або нелінійний, де параметри розраховуються так, що сума квадратів залишків буде мінімальною. Також у оцінюванні можуть використовувати метод максимальної правдоподібності.

Розглядається робота з МНК. Спершу береться модель авторегресії p-го порядку:

$$AR(p): y_t = \alpha_t y_{t-1} + \alpha_t y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t,$$

де $cov(\varepsilon_t, y_{t-i}) = 0, i = 1, \dots, p.$

Як бачимо факторні зміни та залишки некорелюють, отже метод найменших квадратів дозволяє забезпечити конзистентні оцінки. Це дає підстави на твердження, що оцінювання цієї моделі буде подібним до оцінки лінійної регресії з лаговою змінною.

Для моделі з ковзним середнім КС(1): $y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q}$, де ми не спостерігаємо ε_{t-1}, а отже і не можемо застосувати метод регресії.

Візьмемо ε_{t-1} як функцію від спостережуваних y_t:

$$\varepsilon_{t-1} = \sum_{j=0}^{\infty} (-\beta)^J \ y_{t-j-1}.$$

Застосовуючи метод найменших квадратів, ми отримуємо:

$$S(\beta) \sum_{t=2}^{T} \left(y_t - \beta \sum_{j=0}^{\infty} (-\beta)^J \ y_{t-j-1} \right)^2,$$

$$\hat{S}(\beta) \sum_{t=2}^{T} \left(y_t - \beta \sum_{j=0}^{t-2} (-\beta)^J \ y_{t-j-1} \right)^2,$$

Якщо $T \to \infty, S(\beta) \neq \hat{S}(\beta)$, то мінімізуючи наближену суму квадратів $S(\beta)$ дає асимптотичну та конзистентну нормальну оцінку $\hat{\beta}$.

Перейдемо до прогнозування за допомогою комбінованої авторегресійної моделі з проінтегрованим ковзним середнім АРІКС. Для цієї моделі значення незалежної змінної прогнозується для певного моменту у майбутньому $(t+1)$, при цьому, можна вважати, що лаги критериальної змінної мають випадкове або фіксоване значення. Функцію залежності від інформаційної множини, яка містить попередні значення процесу y_t та лагові значення, буде вважатися прогнозним значенням. Обирати дану функцію ми будемо за допомогою мінімізації умовного математичного сподівання квадрату похибки прогнозу.

Береться модель з ковзним середнім та робиться припущення, що її коефіцієнти визначені, а значення y_t існують, $t \in [1, T]$:

$$\text{КС}(q):\ y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q},$$

Тоді прогнозом для моделі у моменту $T+1$ буде вважатися така рівність:

$$\widehat{y_{T+1}} = E[y_{T+1}|y_1, \dots, y_T] = E\left[\varepsilon_{T+1} + \beta_0 + \beta_1 \varepsilon_T + \cdots + \beta_q \varepsilon_{T-q}|y_1, \dots, y_T\right] =$$
$$= \beta_0 + \beta_1 E[\varepsilon_T|y_1, \dots, y_T] + \cdots + \beta_q\left[\varepsilon_{T-q+1}|y_1, \dots, y_T\right].$$

З цього отримується різниця між реальним та прогнозованим значенням:

$$E[\varepsilon_T|y_1, \dots, y_T] = y_T - \hat{y}_T$$

З цього слідує:

$$\hat{y}_{T+1} = \beta_0 + \beta_1(y_T - \hat{y}_T) + \cdots + \beta_q\left(y_{T-q+1} - \hat{y}_{T-q+1}\right).$$

Вводиться момент $h \geq q + 1$, тоді прогноз для нього буде $\hat{y}_{T+h} = \beta_0$.

Тоді, для похибки $y_{T+1} - \hat{y}_{T+1}$ дисперсія для одного кроку вперед буде виглядати:

$$var[y_{T+1} - \hat{y}_{T+1}|y_1, \dots, y_T] =$$
$$= E\left[\left(\varepsilon_{T+1} + \beta_0 + \beta_1\varepsilon_T + \cdots + \beta_q\varepsilon_{T-q}|y_1, \dots, y_T\right)|y_1, \dots, y_T\right]^2 = \sigma_\varepsilon^2.$$

Для прогнозу на два кроки дисперсія похибки буде:

$$var[y_{T+2} - \hat{y}_{T+2}|y_1, \dots, y_T] = (1 + \beta_1^2)\sigma_\varepsilon^2,$$

Для похибки прогнозу на h кроків при $h < q$:

$$var[y_{T+h} - \hat{y}_{T+h}|y_1, \dots, y_T] = (1 + \beta_1^2 + \cdots + \beta_h^2)\sigma_\varepsilon^2.$$

Якщо $h \geq q$, то дисперсія похибки прогнозу дорівнює дисперсії нашого процесу y_t.

Тепер візьмемо модель стаціонарного процесу авторегресії порядку p:

$$\mathrm{AR}(p)\colon y_t = \alpha_t y_{t-1} + \alpha_t y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t,$$

Тепер переглянемо запис прогнозу для такої моделі. Зазначимо, що математичне сподівання білого шуму ε є нульовим. Спочатку візьмемо для одного кроку вперед:

$$\hat{y}_{T+1} = E[y_{T+1}|y_1, \dots, y_T] =$$

$$= E\left[\alpha_0 + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1} + \varepsilon_T | y_1, \dots, y_T\right] =$$
$$= \alpha_0 + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1}.$$

Тепер записується цей вираз для другого кроку:

$$\hat{y}_{T+2} = E[y_{T+2} | y_1, \dots, y_T] =$$
$$= E\left[\alpha_0 + \alpha_1 y_{T+1} + \alpha_2 y_T + \alpha_3 y_{T-1} + \cdots + \alpha_p y_{T-p+2} + \varepsilon_T | y_1, \dots, y_T\right] =$$
$$\alpha_0 + \alpha_1 y_{T+1} + \alpha_2 y_T + \alpha_3 y_{T-1} + \cdots + \alpha_p y_{T-p+2}$$

З цього ми можемо побачити, що $E[y_T | y_1, \dots, y_T] = y_T,\ E[y_{T-1} | y_1, \dots, y_T] = y_{T-1}$ і це можна продовжити далі.

Для нашої моделі авторегресії $AR(p)$ ми отримуємо формулу для прогнозу на один крок вперед:

$$\hat{y}_{T+1} = \alpha_0 + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1}$$

$$y_{T+1} = \alpha_0 + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1} + \varepsilon_{T+1}$$

На жаль, не можна прогнозувати ε_T, тож аналогічно до дисперсії ковзного середнього, дисперсія похибки прогнозу рівна σ_ε^2.

Записується для прогнозу на два кроки вперед:

$$\hat{y}_{T+2} = \alpha_0 + \alpha_1 y_{T+1} + \alpha_2 y_T + \alpha_3 y_{T-1} + \cdots + \alpha_p y_{T-p+2} =$$
$$= \alpha_1\left(\alpha_0 + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1} + \varepsilon_{T+1}\right) + \alpha_2 y_T + \alpha_3 y_{T-1} +$$
$$+ \cdots + \alpha_p y_{T-p+2},$$

$$y_{T+2} = \alpha_0 + \alpha_1 y_{T+1} + \alpha_2 y_T + \alpha_3 y_{T-1} + \cdots + \alpha_p y_{T-p+2} + \varepsilon_{T+2} =$$
$$= \alpha_1\left(\alpha_0 + \alpha_1 y_T + \alpha_2 y_{T-1} + \cdots + \alpha_p y_{T-p+1} + \varepsilon_{T+1}\right) + \alpha_2 y_T + \alpha_3 y_{T-1} +$$

$$+ \cdots + \alpha_p y_{T-p+2} + \varepsilon_{T+2}.$$

Застосовуючи згадану раніше аналогію з дисперсією ковзного середнього, дисперсія похибки прогнозу на два кроки для нашої авторегресії буде:

$$\hat{\sigma} = (1 + \beta_1^2)\sigma_\varepsilon^2$$

Формула для прогнозу на деяку кількість кроків, d кроків, буде дорівнювати:

$$y_{t+d} = \alpha_0 + \sum_{i=0}^{p} \theta_i \hat{y}_{t+d-i},$$

Де θ_i – параметри.

2.2.4 Моделювання з умовною гетероскедастичністю

Авторегресійна умовно гетероскедастична модель (скорочено АРУГ), або AutoRegressive Conditional Heteroscedasticity (скорочено ARCH) полягає у тому, що у певний момент часу t дисперсія її вільного члену ε_t залежить від квадратів вільних членів попередніх моментів у часі. Ця модель вперше згадується у статті Роберта Енгла[7].

Далі необхідно записати саму модель. Якщо у якості $y = y_t$ взяти випадковий процес, що є гаусівською моделлю з $y_t = \sigma_t \varepsilon_t, \varepsilon = \varepsilon_t, t \geq 1$ (σ_t – волатильність) послідовністю випадкових величин з нормальним розподілом. В такому випадку АРУГ(p) буде умовною дисперсією у вигляді лінійної функції квадратів попередніх значень спостережуваного параметра:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2,$$

Де p – максимальний порядок моделі,

σ_t^2 – умовна дисперсія,

y_{t-i} – попередні значення спостережуваного параметра,

α_0, α_i – параметри моделі із заданою умовою $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, p$, але це може бути порушено при наявності великої кількості лагів.

Якщо ввести позначення $\theta_t = y_t^2 - \sigma_t^2$, то написана згори модель може бути записана:

$$y_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \theta_t.$$

Дана модель відповідає згаданій у попередньому підпункті моделі авторегресії $AR(p)$ для квадратів залишків, з чого слідує, що тільки у тому випадку, коли сума додатніх параметрів авторегресії менша за 1 (тобто $\sum_{i=1}^{p} \alpha_i < 1$), то досліджуваний процес є слабо стаціонарним.

Тепер розглянемо безумовну дисперсію. Спершу зазначимо, що вона не залежить від того чи іншого моменту часу t, тож для процесу має місце умовна гетероскедастичність:

$$E(\sigma^2) = \alpha_0 + \alpha_1 E(y_{t-1}^2) + \dots + \alpha_p E\left(y_{t-p}^2\right).$$

З цього можна визначити, що за умови $0 \leq \alpha_1 + \dots + \alpha_p < 1$ буде:

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}$$

Серед інших властивостей ARCH моделей можна відмітити те, що значення $y_t^2, y_{t-1}^2, \dots, y_{t-p}^2$ корельовні.

Оцінювання коефіцієнтів цієї моделі зазвичай проводиться за допомогою методу найменших квадратів. Якщо ж випадкові похибки мають нормальний розподіл, то можливо застосувати метод максимальної правдоподіюності з функцією:

$$\ln L = -\frac{1}{2}\sum_{t=1}^{T}\left(2\ln\sigma_t + \frac{\varepsilon_t^2}{2\sigma_t^2}\right)$$

Отже, тепер постає питання визначення наявності властивостей, які характерні для цих моделей. Роберт Енгл свого часу запропонував для цього використати тест множників Лагранджа (LM-тест)[7].

Покроковий процес тесту:

1) Для процесу y_t формуємо ряд залишків $e_t = y_t - \hat{y}_t$;
2) Проводиться оцінювання ось такої моделі регресії:

$$e_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + z_t$$

3) За початкову береться гіпотеза, що відсутні властивості АРУГ, тобто $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. Альтернативною гіпотезою ж береться те, що хоча б один з коефіцієнтів не дорівнює нулю, тобто $H_1: \alpha_i \neq 0$.
4) За допомогою статистичного тесту Фішера ми проводимо перевірку гіпотези, де з табличним значенням порівнюється фактичне

значення $\chi^2_{\text{факт}}$. Якщо розмір вибірки буде досить великим, то буде застосовуватися тест χ^2.

У випадку, якщо початкова гіпотеза приймається, то $\chi^2_{\text{факт}}$ розроділена за законом χ^2 з p ступенями свободи.

Одною з важливих “фіч” моделей АРУГ є можливість прогнозування умовної дисперсії. Розглянемо це. Для цього, запишемо альтернативну модель з середнім значенням:

$$\sigma_t^2 - \sigma^2 = \alpha_1(y_{t-1}^2 - \sigma^2) + \cdots + \alpha_p\left(y_{t-p}^2 - \sigma^2\right),$$

З цього випишемо формулу для прогнозу. Спочатку р прогнозом на один крок вперед:

$$\sigma_{t+1|t}^2 = \sigma^2 + \alpha_1(y_t^2 - \sigma^2) + \cdots + \alpha_p\left(y_{t+1-p|t}^2 - \sigma^2\right),$$

Тепер на k кроків уперед:

$$\sigma_{t+k|t}^2 = \sigma^2 + \alpha_1\left(\sigma_{t+k-1|t}^2 - \sigma^2\right) + \cdots + \alpha_p\left(\sigma_{t+k-p|t}^2 - \sigma^2\right),$$

Рахуючи те, що сума коефіцієнтів строго менше одиниці, то якщо k прямуватиме до нескінченності, то практичного сенсу прогноз не матиме.

2.2.5 Узагальнена авторегресійна умовно гетероскедастична модель

Generalized AutoRegressive Conditional Heteroscedasticity, скорочено GARCH (або українською УАРУГ) включає значення попередніх умовних

дисперсій, що дозволяє використовувати невеликі значення p та q. Запишемо модель УАРУГ(p,q):

$$\sigma_t^2 = \alpha_0 + \alpha(L)y_{t-1}^2 + \beta(L)\sigma_{t-1}^2,$$

Де $\alpha(L), \beta(L)$ – поліноми оператора зсуву.

На практиці дуже добре працює варіант моделі з $p = 1, q = 1$, що називатиметься УАРУГ(1,1). Записується формула до цього, яка матиме тільки три невідомі невід'ємні параметри:

$$\sigma_t^2 = \alpha_0 + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2,$$

Для кращого розуміння варто розглянути умовну дисперсію. За УАРУГ вона буде лінійною комбінацією p попередніх квадратів залишків і q лагів попередніх значень умовної дисперсії. Формула матиме вигляд:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-1}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-1}^2,$$

Де $\alpha_0, \alpha_i, \beta_i$ – параметри моделі, $i = 1, \dots, p, j = 1, \dots, q$, які мають властивості

$\alpha_0 > 0, \alpha_i \geq 0, \beta_i \geq 0.$

Для оцінки моделей GARCH використовується метод максимальної правдоподібності. Для цього рекурсивно розраховують спеціальний ряд $\{\sigma_t^2\}_{t=0}^{\infty}$. При цьому припускається, що значення залишків та дисперсій $\{\sigma_{-1}^2, \dots, \sigma_{-q}^2, y_{-1}^2, \dots, y_{-p}^2\}$ є довільними.

Як можна побачити, що є деякий випадковий процес моделі УАРУГ(p,q) з параметрами, для яких $\alpha_0 > 0, \alpha_i \geq 0, \beta_i \geq 0$. Тоді, заданий

процес буде стаціонарним тоді і тільки тоді, коли $\alpha + \beta < 1$, або, якщо розширити то для
$\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{p} \beta_i$. Тоді можна записати безумовну диспресію у вигляді:

$$\sigma^2 = \frac{\alpha_0}{1 - \left(\alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots + \beta_q\right)}.$$

Тоді прогноз для умовної дисперсії на d кроків буде мати вигляд:

$$\sigma^2_{t+d|t} = \sigma^2 + (\alpha + \beta)^d(\sigma_t^2 - \sigma^2).$$

2.3 Засоби для оцінки адекватності моделей

Оцінка адекватності моделей є важливим етапом у аналізі та прогнозуванні процесів, у тому числі показників виробничого підприємства. Оцінка дозволяє визначити, наскільки добре модель підходить для опису та прогнозування реальних економічних процесів.

Процес оцінювання адекватності моделі включає ряд критеріїв та тестів, які оцінюють, наскільки добре модель підходить до взятих даних і чи відповідає вона теоретичним припущенням.

Для вибору найкращої моделі можна взяти інформаційний критерій Акайке. Цей метод є досить інформативним, оскільки враховує кількість оцінюваних параметрів та довжину вибірки. Формула записується:

$$AIC = N \ln(SS_R) + 2n,$$

де $SS_R = \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} [\hat{y}_i - \bar{y}_i]^2$ – загальна сума квадратів регресії,
n – кількість параметрів статистичної моделі,

N – кількість спостережень.

Найкраща модель матиме найменше значення критерію Акайке.

Також у цього інформаційного критерію є модифікації, одна з яких – критерій Байєса-Шварца. Його особливість заключається у тому, що до збільшення критерію призводять збільшення кількості спостережуваних змінних та зміна залежних змінних. З цього слідує, щ значення критерію будуть ставати менші за зменшення розмірності моделі, тому не варто його застосовувати не малих вибірках. Формула задається як:

$$BSC = N\ln(SS_R) + n\ln(N),$$

де N – кількість спостережень,

n – це та кількість параметрів моделі, яка оцінюється нами з використанням статистичних даних і знаходиться за формулою: $n = p + q - 1$, де q – кількість параметрів КС, а p – кількість параметрів AR компоненти.

Серед методів, які можуть вказати на адекватність отриманої моделі, можна виділити статистику Дарбіна-Уотсона (Durbin-Watson statistics). Вона виражається формулою:

$$DW = \frac{\sum_{t=2}^{N}[\varepsilon_t - \varepsilon_{t-1}]^2}{\sum_{t=2}^{N}\varepsilon_t^2},$$

де ε_t – оцінені методом найменших квадратів залишки,

N – кількість спостережень.

Значення $DW \in [0,4]$, а якщо $DW = 2$, то автокореляція відсутня. Якщо значення менше 2, то присутня позитивна послідовна кореляція послідовних членів помилок. Якщо ж більше, то присутня негативна кореляція.

Також, для роботи з великою кількістю спостережень можна провести невелике перетворення:

$$\sum_{t=2}^{N}[\varepsilon_t - \varepsilon_{t-1}]^2 = \sum_{t=1}^{N}\varepsilon_t^2 - \varepsilon_1^2 - 2\sum_{t=1}^{N}\varepsilon_t\,\varepsilon_{t-1} + \sum_{t=2}^{N+1}\varepsilon_{t-1}^2 - \varepsilon_N^2,$$

З перетворення вище можна отримати:

$$DW = 2 - 2p\frac{\sum_{t=2}^{N}\varepsilon_t^2}{\sum_{t=1}^{N}\varepsilon_t^2} - \frac{\varepsilon_N^2+\varepsilon_1^2}{\sum_{t=1}^{N}\varepsilon_t^2} \approx 2 - 2p,$$

де p є вибірковою автокореляцією залишків.

Для перевірки значущості коефіцієнта в статистичному сенсі можна використати t-statistic, або статистику Стьюдента. Вона має вигляд:

$$t = \frac{\hat{a} - a^0}{SE_a},$$

де $\hat{a}$ – оцінка коефіцієнта моделі,

a^0 – початкова гіпотеза щодо нашої моделі,

SE_a – стандартна похибка оцінки.

Робота методу описується так: спочатку береться гіпотеза про несуттєву значимість коефіцієнта кореляції або параметра регресії, де $a^0 = 0$. Альтернативою до цього береться гіпотеза що згадані вагіанти не дорівнюють нулю, а саме $a^0 \neq 0$.

Надалі процес включає перевірку значення t, яке ми спостерігаємо. Для цього порівнюється зі значенням з відповідним значенням з таблиці розподілу Стьюдента. Це робиться за допомогою рівня значимості (позначається α) та ступенів свободи.

Оцінюється це за такою схемою:

– Якщо, $|t| > |t_{\text{табл}}|$, тобто спостережуване за модулем більше за взяте з таблиці, то коефіцієнт значимо відмінний від нуля з ймовірністю $1 - \alpha$.

– Якщо навпаки, тобто $|t| < |t_{\text{табл}}|$, то коефіцієнт незначимо відмінний від нуля, а отже гіпотеза приймається.

Також можна зазначити таку річ як коефіцієнт детермінації R^2, або R-Squared. Це статистичний показник у регресійній моделі, який визначає частку дисперсії залежної змінної, яку можна пояснити незалежною змінною. Іншими словами, показує, наскільки добре дані відповідають моделі регресії (відповідність). Формула має вигляд:

$$R^2 = \frac{var(\hat{y})}{var(y)} = 1 - \frac{SS_E}{SS_T} = \frac{SS_R}{SS_T},$$

де $var(\hat{y})$ – оцінена за допомогою моделі дисперсія залежної змінної;

$var(y)$ – дисперсія вимірів залежної змінної;

$SS_E = \sum_{i=1}^{N}[y_i - \hat{y}_i]^2$ – сума квадратів похибок моделі;

$SS_T = \sum_{i=1}^{N}[y_i - \bar{y}_i]^2$ - загальна сума квадратів;

$\bar{y}_i$ – середнє значення;

$SS_R = \sum_{i=1}^{N}[\hat{y}_i - \bar{y}_i]^2$ – загальна сума квадратів регресії.

Якщо $R^2 = 1$, то дисперсії значень змінних, які отримані з моделі та реальних змінних збігаються, тобто модель точно підібрана. Тим не менш, якщо кількість змінних рівна заданій кількості регресорів, то подібне значення коефіцієнта не має сенсу з економічної точки зору. Це виходить з одної з властивостей R^2, яка говорить що значення коефіцієнта зростає від додавання додаткового регресора.

Також, кажучи про властивості, варто згадати, що R^2 змінюється при будь-якому перетворенні залежної змінної.

Як можна зробити висновок із запису вище, додавання регресорів не зможе зменшити значення цього критерію, що можна віднести до недоліків. В такому випадку можна спробувати коригувати оцінки дисперсії на степені

свободи. Отримується скоригований коефіцієнт детермінації $R^2_{\text{скориг}}$, де передбачається певне покарання за включення у модель додаткових змінних, що допомагає зупинити зростання значення коефіцієнта. Обчислюється за формулою:

$$R^2_{\text{скориг}} = 1 - \frac{SS_R(n-1)}{(n-k)SS_T}$$

$R^2_{\text{скориг}}$ є строго меншим за R^2 в усіх випадках окрім ситуацій, коли модель складається тільки з одного константного члена.

F-тест, або статистика Фішера показує, як добре загальна дисперсія залежної змінної пояснюється моделлю, що розглядається. Формула для множинної регресії:

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

де p – кількість параметрів моделі,

N – кількість значень ряду.

За початкову гіпотезу береться те, що модель неадекватна, тобто: $H_0: a_0 = a_1 = \cdots = a_p = 0$. Альтернатива – хоча б одне a_i відмінне від 0. Тоді критичне значення $F_{\text{кр}}$ береться з таблиці розподілу Фішера і якщо отримане $F > F_{\text{кр}}$, то ми відхиляємо початкову гіпотезу про неадекватність моделі.

2.4 Метрики оцінювання якості прогнозів

Середня похибка - абсолютний показник зміщення прогнозу:

$$ME = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{y}_i,$$

де N – кількість спостережень,

y_i – реальні значення,

$\hat{y}_i$ – спрогнозовані значення.

Середня абсолютна похибка

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|,$$

Середній квадрат похибок

$$MSE = E((y_i - \hat{y}_i)^2) = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{N}$$

Згадані вище критерії не можуть самі повноцінно відображати характеристики якості прогнозу. для залежать від рівня залежної змінної. Показник зміщення прогнозу суттєво залежить від обсягу спостережень, який використовується для оцінки прогнозних можливостей економетричної моделі. Зі збільшенням обсягу спостережень зростає впевненість у наближенні середньої помилки до нуля, що свідчить про відсутність зміщення прогнозу.

Середня відсоткова похибка

$$MPE = \frac{1}{N}\sum_{i=1}^{N} \frac{y_i - \hat{y}_i}{y_i} \cdot 100\%,$$

Коефіцієнт невідповідності Тейла

$$K_T = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i{}^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i{}^2}}, K_T \in [0,1]$$

Чим ближчі до нуля MPE та коефіцієнт невідповідності Тейла, тим кращі прогнозні якості моделі.

Середня відсоткова абсолютна похибка:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|} \cdot 100\%$$

Для тлумачення цієї метрики задається таблиця 2.6.

Таблиця 2.6 – Тлумачення метрики MAPE

Значення $MAPE$	Висновки про прогноз
$< 10\%$	Висока якість
$10 - 20\%$	Непогана якість
$21 - 50\%$	Задовільна якість
$> 50\%$	Незадовільна якість

Сума квадратів похибок моделі

$$\sum_{i=1}^{N} e^2(k) = \sum_{i=1}^{N}\left(\hat{y}_i(k) - y_i(k)\right)^2,$$

За результатами цієї метрики обирається та модель, яка має найменшу суму квадратів похибок.

2.5 Висновки до розділу

У розділі були розглянуті та проаналізовані моделі та методи для можливого використання у роботі з прогнозуванням роботи підприємства. Було обрано один якісний метод, різні варіанти авторегресійних моделей. Також було розглянуто різноманітні методи та метрики для оцінки як самих моделей, так і отриманих прогнозів.

Для якісного прогнозу розглядається метод морфологічного аналізу, який хоч і створювався спершу для синтезу нових систем, але з часом отримав багато застосувань, серед яких і прогнозування. У цій ролі він може застосовуватися як і у випадках, коли нам невідомо нічого (це підприємство конкурентів), так і у випадку коли наявних статистичних даних мало і вони можуть погано репрезинтувати картину. Також, якісний метод може урахувати потенційні варіанти подій та станів, які можуть вплинути на загальний прогноз, але не можуть бути репрезентовані тільки через цифри. Недоліком морфологічного аналізу, як і всіх якісних методів, є повна залежність від думки експертів, тобто великий вплив людського фактору.

У якості кількісних методів розглядалися моделі авторегресії, які можуть застосовуватися для прогнозування різних економічно-фінансових характеристик. Одними з найбільш розповсюджених є моделі ARIMA, вони ж моделі Бокса-Дженкінса. Ці моделі є одними з найпопулярніших та мають велику простоту реалізації, що є дуже корисним у питаннях розробки того чи іншого програмного продукту. В залежності від нашої задачі ми можемо додатково застосовувати GARCH (та просто ARCH) моделі та інші. Великим недоліком всіх цих моделей є їхня лінійність та велика кількість параметрів, що веде до більших вимог до обчислювальних потужностей та часу.

Також було розглянуто різні методи та метрики для оцінювання отриманих моделей та прогнозів. В залежності від того, що ми хочемо оцінити, ми використовуємо різні способи та показники. Наприклад, ми можемо застосувати критерій Акайке для оцінки адекватності моделей.

При оцінюванні тих же отриманих прогнозів, ми зазвичай користуємося декількома метриками, оскільки не всі з них самостійно є достатньо репрезентативними. Та і бувають складні випадки, коли в залежності від розглядаваного параметра і задачі краще дивитися на різні метрики.

3 ПРОВЕДЕННЯ ПРОГНОЗУВАННЯ ДАНИХ ЗА ДОПОМОГОЮ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Розгляд базових даних

Для дослідження було обрано Приватне акціонерне товариство “АвтоКрАЗ”, яке також відоме як “Кременчуцький автомобільний завод”, що не до кінця є правдою, оскільки до складу розглядаваного підприємства входить не лише один завод, а декілька виробничих та складальних підприємств.

Основним видом діяльності цього підприємства є виробництво різних вантажних автомобільних транспортних засобів, які постачаються до Збройних Сил, Національної гвардії, ДСНС, комунальних служб та комунальних підприємств України та експортується до багатьох країн світу. Також, до основних видів діяльності підприємства належить встановлення різного спеціалізованого устаткування та виробництво військових транспортних засобів. Це може проходити як і на шасі власного виробництва, так і на ліцензованому або імпортованому з інших країн.

Частково питання розгляду цього прикладу пояснювалося у першій частині цієї роботи, де наводився приклад з конфліктом щодо націоналізації підприємства. Тож, “АвтоКрАЗ” є реальним підприємством, що попри наявність замовлень, в тому числі зі сторони держави, отримувало гірші і гірші фінансові показники (сукупний дохід рис. 3.1).

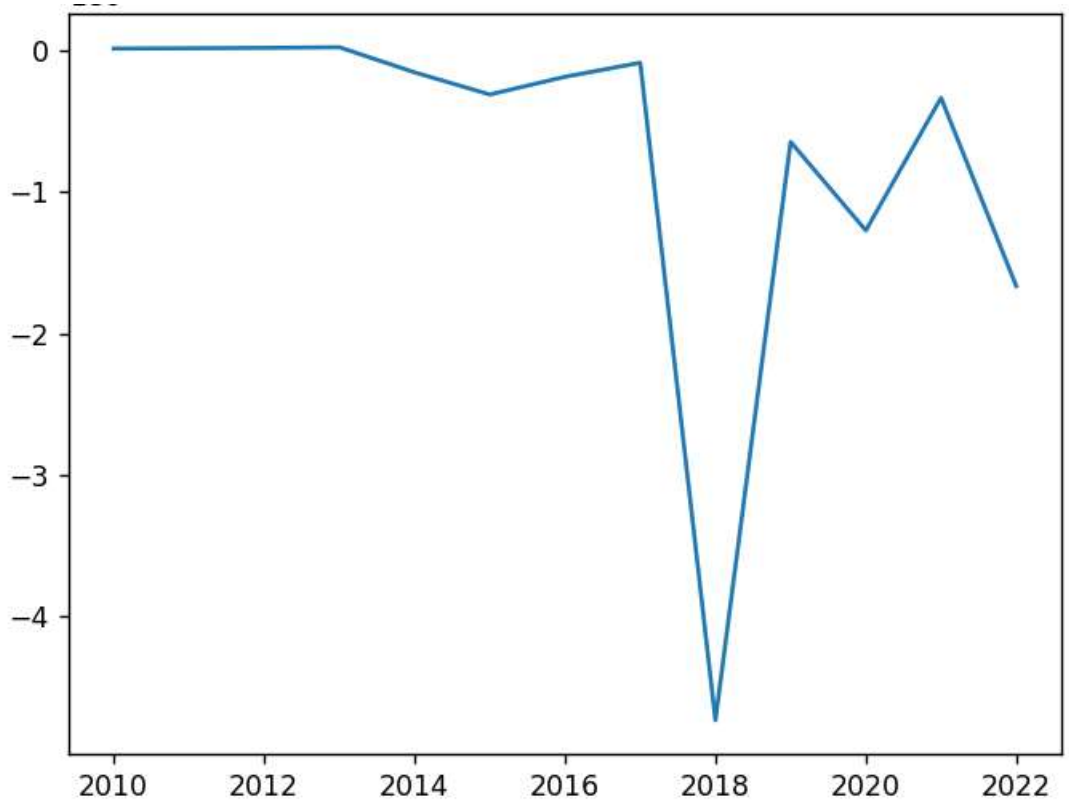


Рисунок 3.1 – Дохідність компанії

Також, варто зазначити, що ,на відміну від іноземних компаній, деякі вітчизняні підприємства (включно з тим яке розглядаємо) не одразу використовували міжнародні стандарти у звітності та у певні періоди не завжди дотримувалися правил оформлення звітності. З цього слідує, що отримані дані є не до кінця точними та не завжди містять повної необхідної інформації, або містять її з певною похибкою. Тож, це і буде цікавість у розгляді задачі.

Дані бралися з експертного оцінювання щодо роботи “АвтоКрАЗ” та відкритих даних про емінента, які містили Баланс та фінансову звітність підприємства.

Балансом підприємства називають звіт про фінансовий стан підприємства, який відображає на певну дату його активи, зобов'язання та власний капітал.

Звіт про фінансові результати (Звіт про сукупний дохід, “Форма 2”) – це звіт з інформацією про фінансові результати певного підприємства впродовж певного періоду часу, а саме його доходи та витрати. Фінальний результат розрізняється як різниця між доходами та витратами з відповідної категорії. З нього виділимо для себе такі пункти:

1. Чистий дохід від реалізації продукції (Revenue) – обсяг товарів та послуг який множиться на вартість цих товарів. Варто уточнити, що Net Revenue знаходиться врахуванням зі знижок, повернень та інших витяжок;
2. Чистий фінансовий результат (Net Income, прибуток, або якщо від’ємне то збиток), який включає до попереднього пункту усі й надходження та витрати та знаходиться у кінці звіту;
3. Валовий прибуток (Gross profit) – різниця між чистим доходом від реалізації продукції та собівартістю цієї самої продукції.
4. Операційний дохід є виручкою, з якої вираховують собівартість адміністративні та операційні витрати;
5. Прибутки від податкових відрахувань, благодійності та інше, теж записуються у різних рядках.
6. Прибуток на акцію (EPS) – чистий прибуток, який ділиться на середньозважену кількість випущених протягом звітного періоду акцій.

Ці показники активно використовуються у аналізі, прогнозуванні та розрахунку різних аналітичних фінансових коефіцієнтів. Останні є ще одним чудовим засобом для аналізу поточного стану компанії та прогнозування її стану через деякий час. Вони можуть вказати на проблеми прийняття рішень, де вони були прийняті не так і чи взагалі варто мати справу з даним підприємством. Для аналізу було обрано такі:

1. Gross margin – коефіцієнт маржинальності, описується як частина загального об’єму виручки, що залишається після врахування витрат щодо виробництва та реалізації товару. Даний коефіцієнт відображає ефективність прийняття рішень у компанії. Чим вищий даний результат – тим вище прибутковість компанії. Має формулу:

$$\text{Gross margin} = \frac{\text{Валовий прибуток}}{\text{Чистий дохід від реалізації продукції}},$$

2. ROA – характеризує ефективність використання активів при генеруванні кінцевих фінрезультатів. Цей коефіцієнт дозволяє оцінити рівень професійної придатності менеджменту підприємства та його рішення. Формула:

$$\text{ROA} = \frac{\text{Чистий фінрезультат}}{\text{Величина активів підприємства}},$$

Де *величина активів підприємства* береться із згаданого раніше Балансу (теж наявний у інформації про емінента), та береться як середнє на початку та в кінці розглядаваного періоду (це важливо, оскільки між ними може бути велика різниця, у нашому прикладі це є).

3. EBIT Margin – показник, який дозволяє оцінити рентабельність компанії, тобто розглянути її фінансові результати впускаючи вплив таких факторів, як розміри податків чи амортизаційна політика. Формула:

$$\text{EBIT Margin} = \frac{\text{Операційний дохід}}{\text{Чистий дохід від реалізації продукції}},$$

4. ATR – характеристика оборотності активів, яка визначає ефективність використання виробничих ресурсів. Формула:

$$ATR = \frac{\text{Чистий дохід від реалізації продукції}}{\text{Величина активів підприємства}}$$

3.2 Опис платформи для виконання аналізу

Для виконання аналізу було розроблено пропрієтарну програму, яка ділиться на дві частини. Перша проводить обрахункову частину методу морфологічного аналізу, а друга використовується для роботи з методами на основі часових рядів. Реалізація проводиться на основі мови програмування Python із застосуванням її різних бібліотек для перегляду, аналізу та прогнозування даних, а також загалом із обчислювальними роботами. Найважливішими є:

– NumPy – відкритий пакет засобів, що дає змогу оперувати з багатовимірними масивами і матрицями, а також забезпечує роботу з великою кількістю математичних функцій, які можуть застосовуватися як до стандартних типів Python, так і до доданих цим пакетом. Ще одною особливістю роботи даної бібліотеки є швидкодія. Так, Python відстає від “нижчих” мов програмування у стандартній швидкодії, тож пакет NumPy написаний був на швидшій базі, що дозволяє працювати з даними та обчисленнями суттєво швидше стандарту. Підтримується великою кількістю інших пакетів для мови програмування Python.

– pandas – бібліотека, яка дає широкий функціонал для аналізу та маніпуляцій з наборами або базами даних. Додається інструментарій

для роботи з одно та багатовимірними таблицями та часовими рядами.

- statsmodels – б модуль, який надає класи та функції для оцінки багатьох різних статистичних моделей, а також для проведення статистичних тестів і дослідження статистичних даних. Розширений список статистики результатів доступний для кожного оцінювача. Також результати перевіряються на існуючі статистичні пакети, щоб переконатися, що вони правильні.
- Matplotlib - це комплексна бібліотека для створення статичних, анімованих та інтерактивних візуалізацій на Python. Так само, сумісна з великою кількістю інших пакетів та бібліотек, що дає високу функціональність для широкого спектру задач, включно з аналізом і прогнозуванням даних.

Вибір цієї мови та бібліотек є логічним для подібного завдання. Як було вказано у вступі, то мова програмування Python є доступним багатоцільовим інструментом для аналізу даних, машинного навчання та прогнозування. Її синтаксис дозволяє забути про багато нюансів більш низькорівневих мов, а наведені вище бібліотеки містять широкий функціонал для роботи з даними, моделями для прогнозування та різними методами перевірки. Також тут присутній інструментарій для створення графічного інтерфейсу.

Створення спеціальних програм пояснюється особистою зручністю та полегшенням роботи. Так, обрана мова програмування та бібліотеки до неї є відкритими розробками, які в тій чи іншій мірі надаються на безоплатній основі, в той час як різне вже існуюче програмне забезпечення потребує придбання та має певні обмеження та використання.

Функціонал роботи частини щодо морфологічного аналізу:

1. Завантажуються дані щодо першої морфологічної таблиці;
2. Завантажується таблиця зв’язків;
3. Проводяться обчислення щодо першого етапу;

4. Виведення результатів для першого етапу: отриманої морфологічної таблиці та найімовірніших конфігурацій;
5. Завантаження даних другої морфологічної таблиці для другого етапу;
6. Обчислення для другого етапу;
7. Виведення результатів з найбільш ймовірними стратегіями та загальний результат прогнозів.

Тепер записується функціонал другої частини розробленої програми:

1. Завантаження даних часового ряду;
2. Побудова графіку;
3. Розклад ряду на три складові компоненти;
4. Перевірка ряду на стаціонарність за допомогою ADF, PP, KPSS тестів;
5. Аналіз автокореляційної та часткової автокореляційної функцій;
6. ЗА потреби – доведення до стаціонарності;
7. Побудова моделей ARMA та ARIMA та визначення підходящих для них параметрів, аналіз отриманих результатів;
8. Розгляд АКФ/ЧАКФ залишків та побудова моделей ARCH/GARCH, аналіз отриманого;
9. Побудова прогнозів та їх оцінка.

Для розробки використовується платформа JetBrains PyCharm, яка розроблена на основі рішення IntelliJ IDEA. Використовується версія Community Edition, яка розповсюджується на безкоштовній основі та дозволяє роботу для таких студентських проєктів як наявний.

3.3 Робота з методом морфологічного аналізу

Робота відкривається із запису початкових даних, які були отримані від експертної оцінки одного з досліджень щодо підприємства у попередні роки. Це в певній мірі допоможе оцінити отриманий результат з наявними даними.

На Таблицях 3.1 і 3.2 можна побачити морфологічну таблицю для першого етапу дослідження та її ймовірності:

Таблиця 3.1 – морфологічна таблиця

Характеристичні параметри					
Виготовлені машини на місяць	Локалізація	Фінансовий стан	Стабільна якість	Готовність йти назустріч	Забезпечення сервісу
1	2	3	4	5	6
1.1 До 20	2.1 Повна	3.1 Поліпшення	4.1 Так	5.1 Так	6.1 Нема
1.2 До 50	2.2 Частково імпорт	3.2 Ще більше погіршення	4.2 Ні	5.2 Ні(беріть як є)	6.2 Запчастини
1.3 До 100	2.3 Більшість імпорт	3.3 Збереження негативної тенденції			6.3 Ремонтні спеціалісти
1.4 Понад 100	2.4 Готові комплекти				6.4 Повний сервіс

Можна побачити, що кожній характеристиці та кожній альтернативи кожної характеристики присвоюється відповідний індекс. Це пов’язано зі спрощенням індентифікації отриманих результатів, бо повноцінні фрази дуже незручні у відображенні результатів програми у подальшому.

Таблиця 3.2 – морфологічна таблиця з ймовірностями

Характеристичні параметри					
Виготовлені машини на місяць	Локалізація	Фінансовий стан	Стабільна якість	Готовність йти назустріч	Забезпечення сервісу
1	2	3	4	5	6
1.1 = 0.2	2.1 = 0.1	3.1 = 0.3	4.1 = 0.4	5.1 = 0.4	6.1 = 0.3
1.2 = 0.3	2.2 = 0.2	3.2 = 0.2	4.2 = 0.6	5.2 = 0.6	6.2 = 0.3
1.3 = 0.3	2.3 = 0.3	3.3 = 0.5			6.3 = 0.3
1.4 = 0.2	2.4 = 0.4				6.4 = 0.1

Отримані дані завантажуються у програму. Далі ми завантажуємо матрицю взаємозвязків альтернатив параметрів. Вона представлена на таблиці 3.3. Експертами до неї надається таке пояснення:

1.1-3.3 та 1.2-3.3 прямий зв'язок що випливає з самого визначення фінстану. 1.4-4.1 та 1.3-4.2 При збільшенні серійності легше забезпечити конвеєрну якість та уніфікацію 2.1-3.2 або 3.3 Оскільки більшість старих моделей покладалися на імпорт в тій чи іншій мірі, для повної локалізації більше знадобиться робота з новими моделями 1.1-6.2 або 6.4 - Складно при малій серійності забезпечити запчастинами та нормальним ремонтом При утриманні стабільної якості досить ймовірно що виробник для цього прислухатиметься до вимог клієнта Також від готовності слухати замовника залежить обслуговування.

Таблиця 3.3 – Матриця взаємозв’язків першого етапу

	1				2				3			4		5	
	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3	4.1	4.2	5.1	5.2

2	2.1													
	2.2													
	2.3													
	2.4													
3	3.1					-0.2								
	3.2					0.6								
	3.3	-0.8	-0.2			0.4								
4	4.1			0.2	0.3									

Продовження **таблиці 3.3**

4	4.2														
5	5.1													0.1	
	5.2														
6	6.1											-0.1		-0.8	0.5
	6.2	-0.5												0.4	0.3
	6.3													0.6	-0.1
	6.4	-0.4										0.2		0.6	-0.8

Надалі ці завантажені таблиці використовуються у програмі для обчислення результатів першого етапу. Частини з отриманого можна побачити на рисунках 3.2 та 3.3. Це можливі комбінації, їх проміжні параметри та фінальні ймовірності.

	1	2	3	4	5	6	p0	C	p1	p_final
0	1.1	2.1	3.1	4.1	5.1	6.1	0.000288	0.15840	0.000046	0.000043
1	1.1	2.1	3.1	4.1	5.1	6.2	0.000288	0.61600	0.000177	0.000166
2	1.1	2.1	3.1	4.1	5.1	6.3	0.000288	1.40800	0.000406	0.000380
3	1.1	2.1	3.1	4.1	5.1	6.4	0.000096	1.01376	0.000097	0.000091
4	1.1	2.1	3.1	4.1	5.2	6.1	0.000432	1.08000	0.000467	0.000438

Рисунок 3.2 – Частина можливих комбінацій з проміжними параметрами

	1	2	3	4	5	6	p0	C	p1	p_final
763	1.4	2.4	3.3	4.2	5.1	6.4	0.000960	1.60000	0.001536	0.001441
764	1.4	2.4	3.3	4.2	5.2	6.1	0.004320	1.50000	0.006480	0.006079
765	1.4	2.4	3.3	4.2	5.2	6.2	0.004320	1.30000	0.005616	0.005269
766	1.4	2.4	3.3	4.2	5.2	6.3	0.004320	0.90000	0.003888	0.003648
767	1.4	2.4	3.3	4.2	5.2	6.4	0.001440	0.20000	0.000288	0.000270

Рисунок 3.3 – Частина можливих комбінацій з проміжними параметрами

Скориставшись Таблицею 3.1 можна розшифрувати отримані результати. Таким чином, отримується ймовірність кожного можливого варіанту розвитку подій. Власне кажучи, як найбільш ймовірний варіант ми можемо побачити до сотні виготовлених автомобілів, які будуть зібрані з готових комплектів, проблеми з якістю та готовністю йти назустріч. Питання фінансового становища має бути так само поганим. Решту результатів обчислень до цієї буде занесено до Exel файлу, який програма створить самостійно у директорії з виконуваним файлом.

Тепер, продовжуючи перший етап, дослідження рухається далі до отриманих значень ймовірностей для першої морфологічної таблиці. Вона знаходиться з попередньо розглянутого результату. Спочатку розглядається скріншот з програми на рисунку 3.4.

	1	2	3	4	5	6
0	0.103036	0.121584	0.316235	0.438103	0.412371	0.254441
1	0.295474	0.195204	0.228034	0.561897	0.587629	0.345506
2	0.355545	0.292805	0.455731	0.000000	0.000000	0.326559
3	0.245945	0.390407	0.000000	0.000000	0.000000	0.073494

Рисунок 3.4 – Фінальна морфологічна таблиця першого етапу

Тепер записуються знайдені оцінки у таблиці 3.4 для кращого розуміння.

Таблиця 3.4 – Фінальна морфологічна таблиця першого етапу

Характеристичні параметри					
Виготовлені машини на місяць	Локалізація	Фінансовий стан	Стабільна якість	Готовність йти назустріч	Забезпечення сервісу
1	2	3	4	5	6
1.1 = 0.103036	2.1 = 0.121584	3.1 = 0.316235	4.1 = 0.438103	5.1 = 0.412371	6.1 = 0.254441
1.2= 0.295474	2.2 = 0.195204	3.2 = 0.228034	4.2 = 0.561897	5.2= 0.587629	6.2 = 0.345506
1.3 = 0.355545	2.3 = 0.292805	3.3 = 0.455731			6.3 = 0.326559
1.4 = 0.245945	2.4 = 0.390407				6.4 = 0.073494

На цьому завершується перший етап морфологічного аналізу.

Другий етап морфологічного аналізу розглядає можливі кроки та з якою ймовірністю вони скоріше за все відбудуться, або підходять для

прийняття у якості рішення. Завантажимо потрібну морфологічну таблицю (таблиця 3.5).

Таблиця 3.5 – Початкова морфологічна таблиця другого етапу

7.1 Припинення підтримки	8.1 Ліквідація
7.2 Націоналізація	8.2 Продаж
7.3 Заключення нових контрактів	8.3 Покращення наявних можливостей виробництва
	8.4 Пошук нових контрактів та розширення під них
	8.5 Нічого не робити

Далі у програму завантажується оцінена матриця зв’язків альтернатив параметрів першого та другого етапів. Тут вона буде Таблиця 3.6. Пояснення до неї експерти надають таке:

Для держави важливими є об’єми виробництва підприємства, його можливість утримувати стабільну якість товару та його обслуговування. Для підприємства є важливими обсяги, а також діючий стан, з якого випливає як і чи взагалі діяти. Так, бездіяльність досить серйозно висока через особливу історію досліджуваного підприємства.

Програма видає нам результат, за яким спершу виводяться всі можливі комбінації разом з фінальним значенням ймовірності для другого етапу та всіма проміжними значеннями. Приклад початку на рисунку 3.5.

	1	2	3	4	5	...	pR81	pR82	pR83	pR84	pR85
0	1.1	2.1	3.1	4.1	5.1	...	0.000000	0.000009	0.000012	0.000008	0.000014
1	1.1	2.1	3.1	4.1	5.1	...	0.000000	0.000022	0.000057	0.000052	0.000036
2	1.1	2.1	3.1	4.1	5.1	...	0.000000	0.000054	0.000145	0.000133	0.000049
3	1.1	2.1	3.1	4.1	5.1	...	0.000000	0.000007	0.000038	0.000041	0.000006
4	1.1	2.1	3.1	4.1	5.2	...	0.000000	0.000046	0.000013	0.000022	0.000357

Рисунок 3.5 – Можливі комбінації другого етапу

Таблиця 3.6. - матриця зв’язків альтернатив параметрів першого та другого етапів

		7			8				
		7.1	7.2	7.3	8.1	8.2	8.3	8.4	8.5
1	1.1	0.8	0.5	-1	0.8	0.9	0.2	-0.6	0.7
	1.2	0.6	0.4	-0.1	0.2	0.7	0.7	0.1	0.7
	1.3	0.1	0.4	0.2	-0.3	0.1	0.8	0.6	0.8
	1.4	-0.4	0.1	0.6	-0.8	-0.1	0.6	0.8	0.7

Продовження **таблиці 3.6**

2	2.1	-0.2	0.2	0.4	-0.6	-0.1	0.5	0.6	-0.2
	2.2	-0.2	0.1	0.3	-0.2	0	0.7	0.3	0.2
	2.3	0	0.1	0	0	0.1	0.5	0	0.5
	2.4	0	0.1	0	0	0.2	0.5	0	0.6
3	3.1	0	0	0.1	0	0.1	0.1	0.2	0.5
	3.2	0	-0.1	0.1	0	-0.1	0.8	0.8	-0.8
	3.3	0	0	0.2	0	0	0.4	0.4	-0.3
4	4.1	-0.8	0.3	1	-1	-0.6	0	0.8	0.8

	4.2	0.9	0.3	-0.8	0.3	0.2	0.4	-0.8	0.4
5	5.1	-0.6	-0.8	0.6	-0.8	0	0.5	0.8	-0.7
	5.2	0.7	0.8	-0.7	0.1	0	-0.7	-0.1	0.4
6	6.1	0.8	0.7	-0.6	0.5	0.4	-0.5	-0.6	0.6
	6.2	-0.1	0.2	0.1	0.4	0.5	0	0.1	0.7
	6.3	-0.3	-0.8	0.2	0.3	0.6	0.1	0.2	0
	6.4	-0.4	-0.8	0.5	-0.2	0	0.4	0.8	-0.4

З цього розраховуються ймовірності для другої морфологічної таблиці, яка розміщена у таблиці 3.7. Для її розшифровки можна застосувати таблицю 3.5.

Таблиця 3.7 – Результати другого етапу морфологічного аналізу

Реакції держави	Реакції керівництва
7	8
7.1 = 0.327135	8.1 = 0.050224
7.2 = 0.369775	8.2 = 0.157151
7.3 = 0.303090	8.3 = 0.267476
	8.4 = 0.194699
	8.5 = 0.330450

Вважається, що якісні методи крім практики дуже складно перевірити на правильність. Хоча частину отриманих результатів ми зможемо оглянути за допомогою кількісних моделей, можливо зараз оцінити деякий результат з

цього аналізу. Експертні оцінки, за якими робився аналіз давалися літом 2022 року, майже за рік до здачі цієї роботи. За ними у нас вийшло, що найімовірнішою реакцією держави вийшла націоналізація, що дійсно сталося у листопаді 2022 року[17].

3.4 Робота з кількісними моделями

Розглянемо прогнозування фінасових показників підприємства, почнемо з загального доходу.

Спершу для роботи проглянемо наш ряд розкладеним на компоненти. Графік на рисунку 3.6.

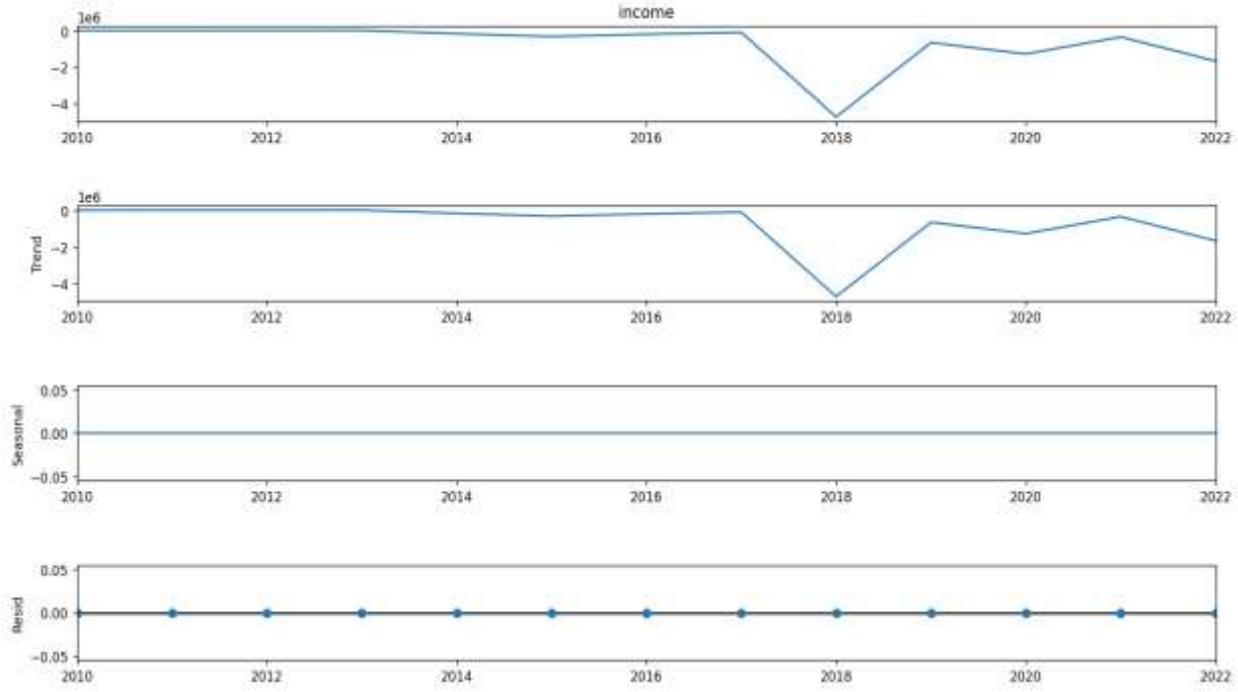


Рисунок 3.6 – Розклад ряду на компоненти

Далі необхідно перевірити завантажений ряд на стаціонарність. Спершу проглянемо результати тестів. Використовуватимуться тести KPSS та Дікі-Фуллера, результати яких відображені на рисунку 3.7.

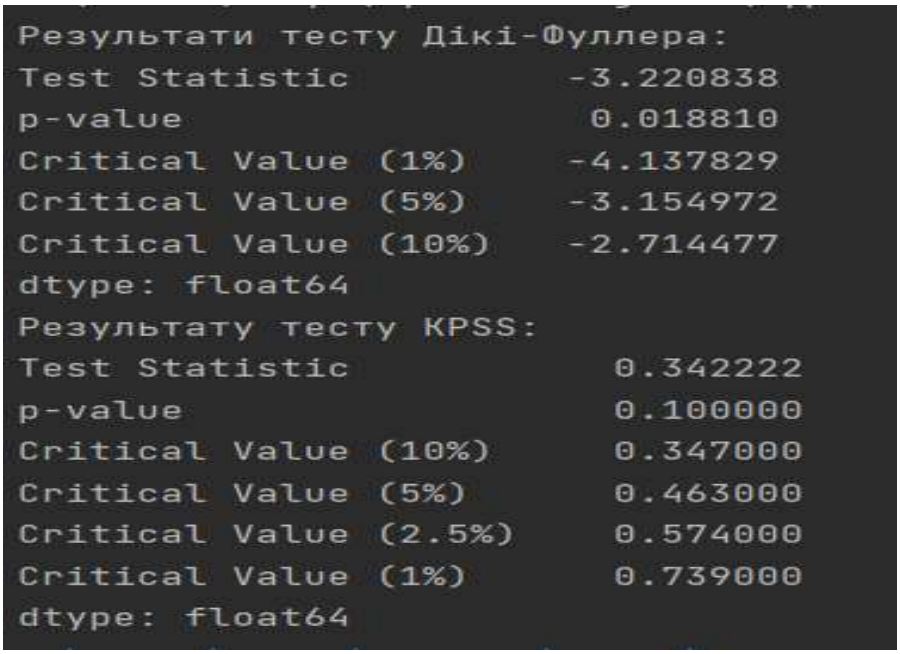


Рисунок 3.7 – Тести на стаціонарність

Результати дають нам право стверджувати що завантажені дані з самого початку є стаціонарними. Це може бути викликано помилкою невеликого датасета, але робота з малим масивом даних є частиною цього невеликого дослідження.

Програма має функціонал логарифмування та взяття різниць, що може дозволити покращити стаціонарність датасету. Варто зазначити, що у випадку від’ємності даних логарифмування може не працювати коректно. Зазвичай у фінансовому аналізі це не буде такою великою проблемою, але у нашому випадку це невелике виключення (як і вся діяльність цього підприємства).

Надалі розглядається автокореляційна та часткова автокореляційна функції. Результат побудови на рисунку 3.8.

Як бачимо, результати показують, що на рисунках нема суттєвих лагів, тож наш ряд є стаціонарним. Можемо продовжувати далі роботу. Найкращий результат із застосуванням ARIMA було отримано у ARIMA(1,2,0) з показниками AIC= 357.492 HQIc = 356.990 BIC=-358.288. Детальніше про модель у підсумку на рисунку 3.9.

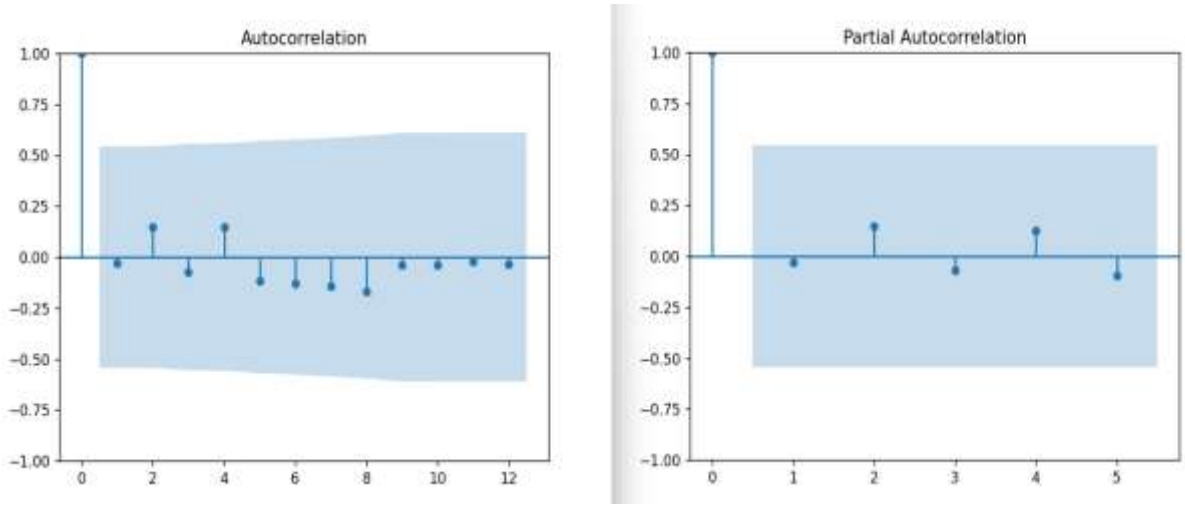


Рисунок 3.8 – АКФ та ЧАКФ моделі

Надані 4 графіки на основі декількох тестів дають підсумок моделі.

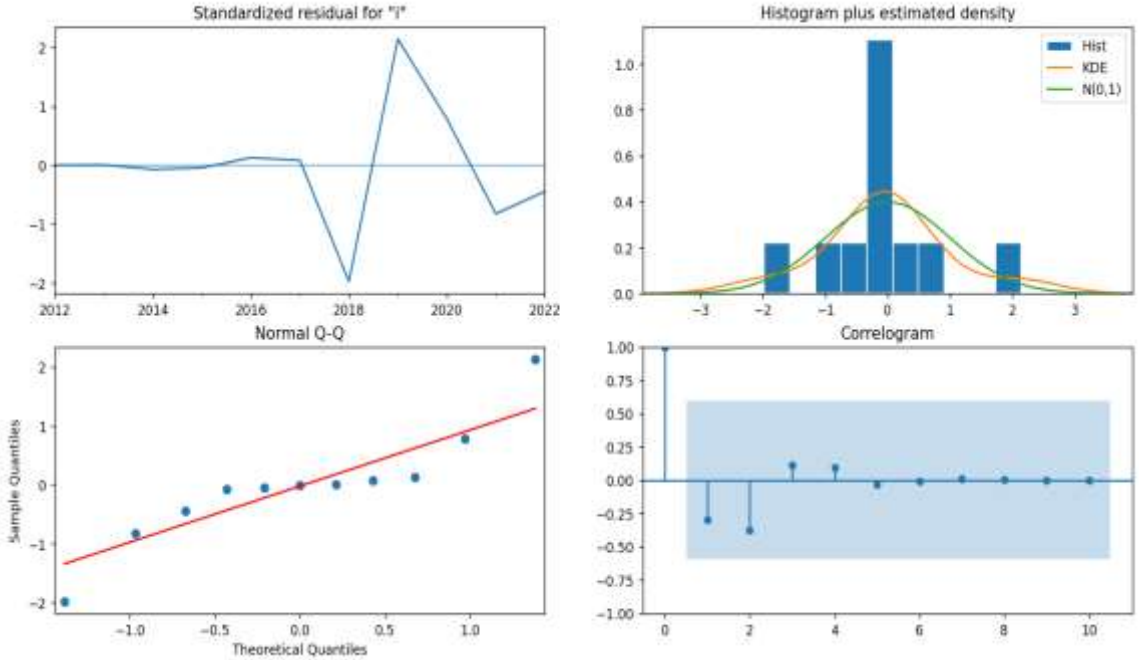


Рисунок 3.9 – Оцінки моделі

По першому можна побачити стандартизовані лишки для отриманої моделі. На останньому графік автокореляційної функції. Значення параметрів у даної моделі відповідно будуть: ar.L1== -0.5911, ma.L1 == -0.8728.

Далі ми будемо розглядати ARCH/GARCH моделі для залишків моделі, яка була отримана перед цим на таблиці 3.8.

Таблиця 3.8 – Результати різних ARCH та GARCH моделей

Модель	AIC	BIC
ARCH(1)	412.777	415.037
ARCH(2)	412.777	415.037
GARCH(1,1)	412.174	414.434
GARCH(1,2)	413.944	416.769
GARCH(2,1)	414.174	416.999

З отриманих результатів слідує, що з даних моделей найкращі результати матиме GARCH(2,1). Запишемо її коефіцієнти: omega == 5.8163e+11, alpha[1]==8.9645e-15, alpha[2]==4.9045e-15, beta[1] ==0.7625.

Проведемо тепер прогнозування, результати на таблиці 3.9.

Таблиця 3.9 – Оцінки отриманого прогнозу

Модель	MSE	MAE	MAPE
ARIMA(1,2,0),	0.40	0.45	0.39
ARIMA(1,2,0)+ GARCH(2.1)	0.39	0.40	0.38988

Отримані результати на прогнозування 2022 року, які показують непогані показники. Отримане значення, -734969.95459 відповідно, демонструє те, що фінансові показники погіршуватимуться надалі.

Тепер проведемо прогнозування для числа виготовлених автомобілів. Розглянемо результати тесту на стаціонарність на рисунку 3.10.

```
Результати тесту KPSS:
Test Statistic          2.551061
p-value                 0.999064
Critical Value (1%)    -4.938690
Critical Value (5%)    -3.477583
Critical Value (10%)   -2.843868
dtype: float64
Результату тесту  Дікі-Фуллера:
Test Statistic           0.487585
p-value                  0.044463
Critical Value (10%)     0.347000
Critical Value (5%)      0.463000
Critical Value (2.5%)    0.574000
Critical Value (1%)      0.739000
dtype: float64
```

Рисунок 3.10 – Результати тестів на стаціонарність

Як можна побачити, отримано стаціонарні дані щодо виробництва вантажівок (кількості). Тепер поглянемо на автокореляційну та часткову автокореляційну функції на рисунку 3.11.

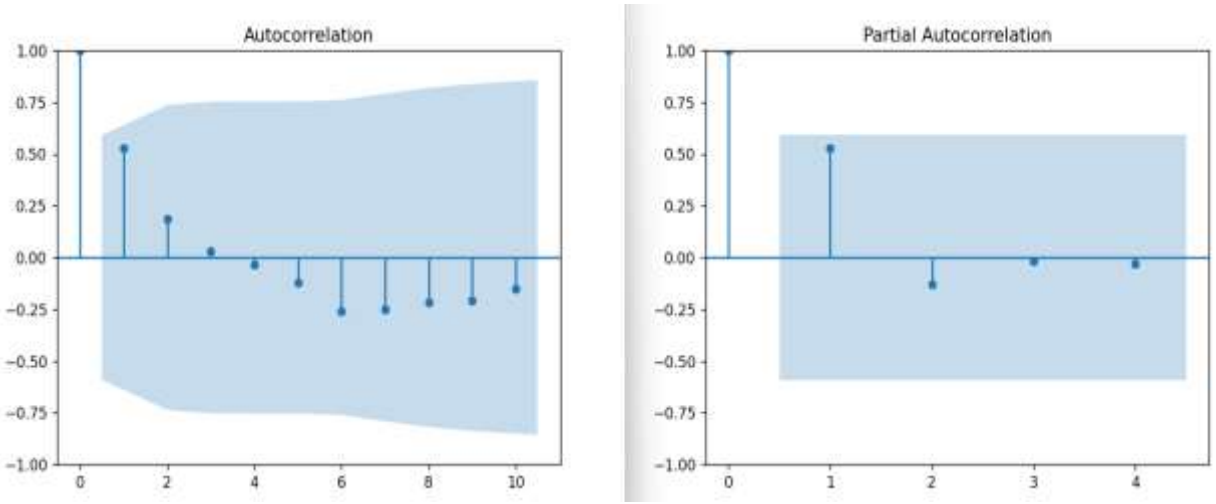


Рисунок 3.11 – АКФ та ЧАКФ

Далі розглянемо оцінку нашої моделі. Для отриманих даних найкращий результат продемонструвала модель ARIMA(1,0,1). AIC=38.686, BIC== 40.626, HQIC==37.968. Були отримані коефіцієнти моделі: const==4.5391,

ar.L1==0.8830, ma.L1==0.7333, sigma2 == 0.5745. На рисунку 3.12 зображено більше інформації про саму модель.

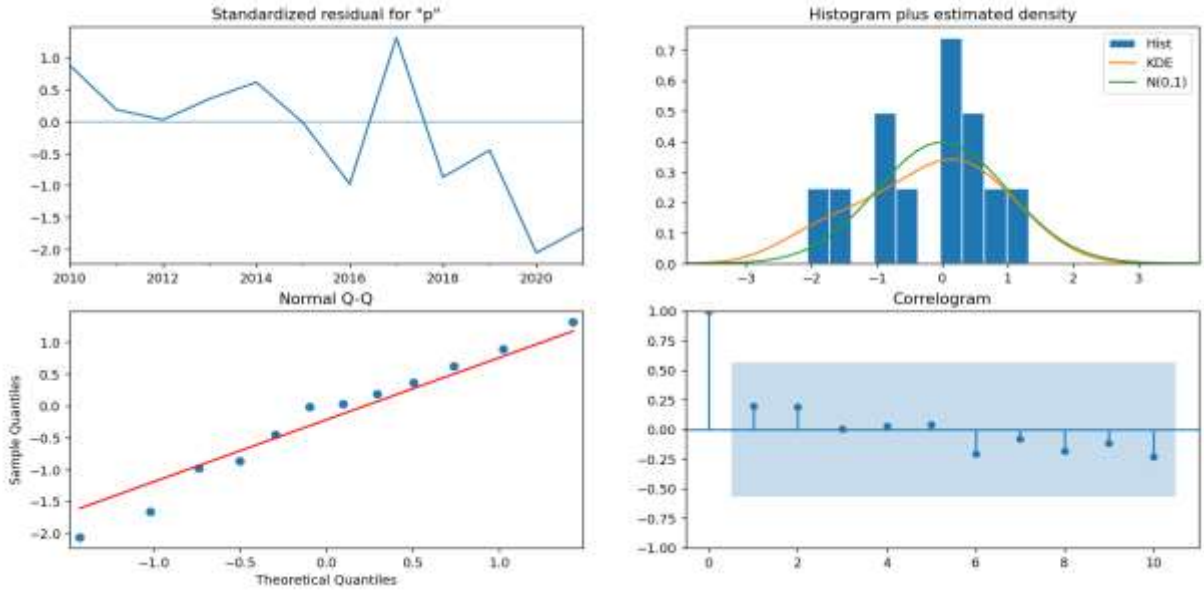


Рисунок 3.12 – Оцінка отриманої моделі

Тепер звернемося до прогнозування. Тут, окрім звичайного прогнозування на один рік задля перевірки точності прогнозу (за допомогою реальних даних) ми проведемо ще й прогноз на додаткові 4 кроки для перегляду результатів якісного аналізу, тобто того що було отримано у морфологічному аналізі. Отриманий результат:

– 2022-01-01 1.779350 тобто 1-2 машини;
– 2023-01-01 2.828967 тобто 2-3 машини;
– 2024-01-01 4.260237 тобто 4-5 машин;
– 2025-01-01 6.115552 тобто 6-7 машин;
– 2026-01-01 8.415264 тобто 8-9 машин.

Відповідні показники для аналізу коректності прогнозу будуть на таблиці 3.10.

Таблиця 3.10 – результати прогнозу

MSE	MAE	MAPE

0.46	0.48	0.45

Результати прогнозу дуже непогані для такого малого датасету з таким неточним заповненням. Як можемо бачити, результати більш менш відображають реальну картину і в певній мірі підтверджують результат морфологічного аналізу, але тим не менш, варіант там все ж вийшов некоректний.

3.5 Висновки до розділу

Ми провели експеримент щодо моделювання та прогнозування діяльності виробничого підприємства на прикладі “АвтоКрАЗ”. Було визначено важливі фінансові показники та вхідні дані для морфологічного аналізу. Далі, отримані дані було опрацьовано за допомогою програм на мові програмування Python.

Якщо говорити про саме дослідження, то складно оцінити повноту точності якісних методів. У цьому виключенням не є морфологічний аналіз, чий повний результат ми не дізнаємося окрім практики чи підкріпимо гіпотезу написаного експертами через якісні методи. Під час експерименту було зазначено, що даний аналіз був точнішим, бо передбачив націоналізацію державою. Параметри щодо кількості виготовлених машин були не докінця підтверджені, оскільки якісні методи показали всього 3-4 виготовлені машини у 21-22 роках і в подальшому не вище 5, що входить у варіант до 100, але більше підійшло варіанту до 50, який морфологічний аналіз не взяв найімовірнішим.

Якщо говорити про якісні методи то вони показали відносно вірну точність. У цілому, підприємство яке ми розглядали, вело дуже посередню річну звітність, тож питання до результатів залишаються. Щодо прогнозу

фінансових показників, то він вийшов посереднім, але зміг хоча б відобразити вірні тенденції у діяльності. Також варто зазначити вплив повномасштабного вторгнення, що дещо змінив фінальний реальний результат. Найкраще себе в обох випадках показала ARIMA, (1,2,0) та (1,0,1).

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 наступні:

- 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) Python.

б) C++.

Функція F_2.

а) Застосування вбудованих функцій.

б) Створення своїх обчислень значень.

Функція F_3:

а) Використання шаблонних графіків.

б) Створення своїх.

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

Рисунок 4.1 – Морфологічна карта

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

Таблиця 4.1 - Позитивно-негативна матриця

Функції	Варіанти реалізації	Переваги	Недоліки
F_1	*А*	Загальнодоступна програма, доступність багатьох бібліотек	Необхідність повної реалізації алгоритму
	Б	Доступна в реалізації програма для різних обчислень	На написання коду необхідно мати базові навички та вміння
F_2	*А*	Доступність та	Іноді не відповідає

		легкість при написанні	задачі яку треба розв’язати
	Б	Ідеально описують усі необхідні характеристики	Достатньо затратно реалізовувати свої алгоритми для подальшої реалізації
F_3	*А*	Загально прийнята реалізація	Іноді не відповідає очікуваним значенням
	Б	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 відкинутий.

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 такий варіанти реалізації ПП:

$F_1\text{а} - F_2\text{а} - F_3\text{а}$

F_1а – F_2б – F_3а

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:

- *X1* – швидкодія мови програмування;
- *X2* – об’єм пам’яті для обчислень та збереження даних;
- *X3* – час навчання даних;
- *X4* – 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 позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія мови програмування	X1	оп/мс	60	80	110
Об’єм пам’яті	X2	Мб	2048	1024	512
Час попередньої обробки	X3	мс	95	80	65

даних					
Потенційний об’єм програмного коду	X4	кількість рядків коду	500	250	170

За даними таблиці 4.3 будуються графічні характеристики параметрів – рис. 4.2 – рис. 4.5.

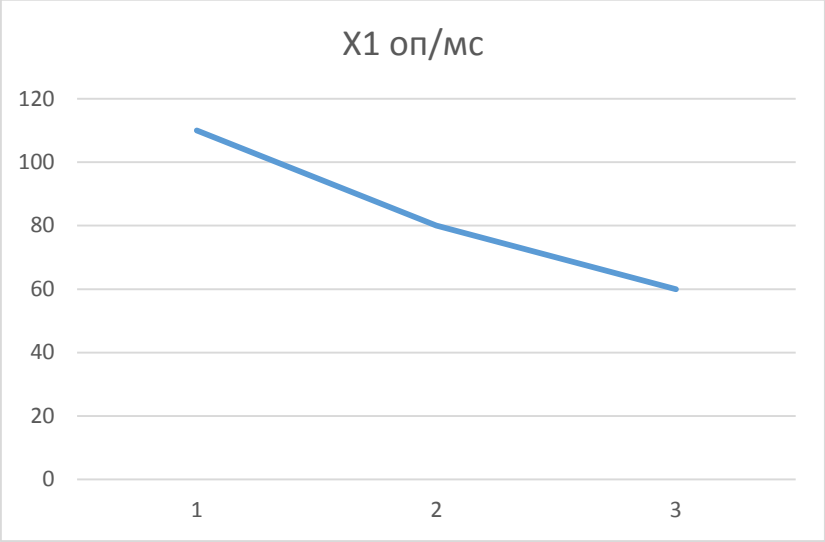


Рисунок 4.2 – X1, швидкодія мови програмування

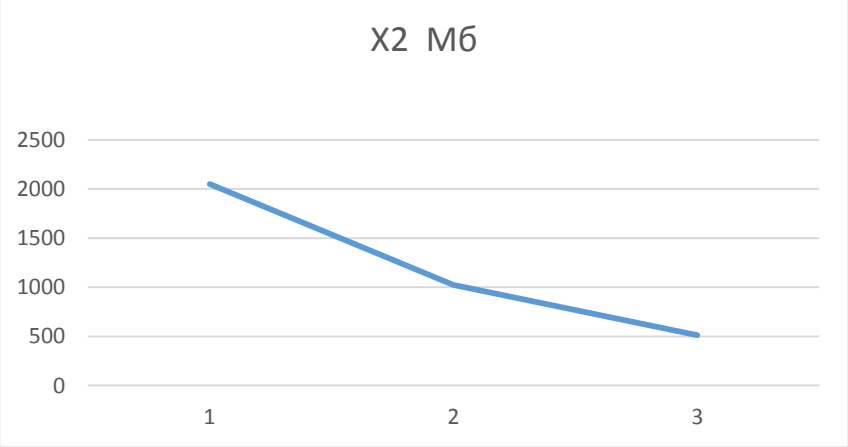


Рисунок 4.3 – X2, об’єм пам’яті

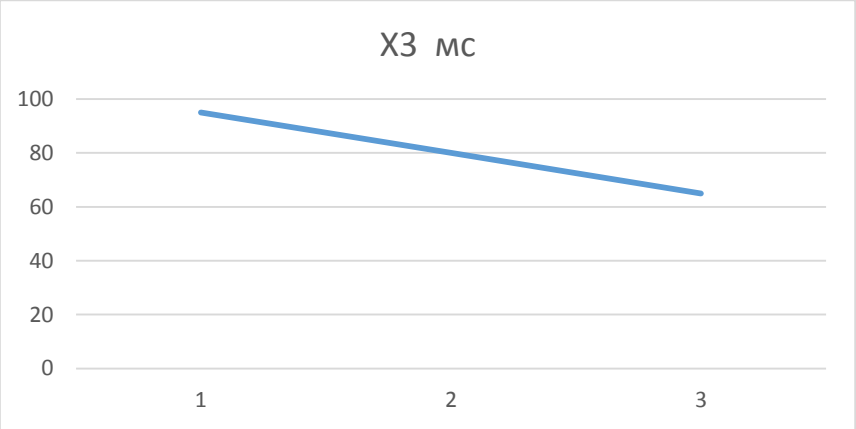


Рисунок 4.4 – X3, час попередньої обробки даних

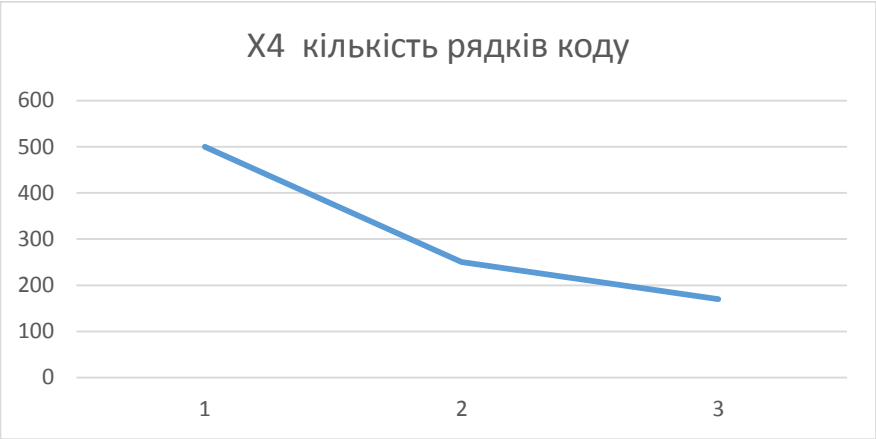


Рисунок 4.5 – X4, потенційний об’єм програмного коду

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	Відхи-лення Δ_i	Δ_i^2
			1	2	3	4	5	6	7			
X1	Швидкодія мови програмування	Оп/мс	2	1	2	1	2	2	2	12	-5,5	30,25
X2	Об’єм пам’яті	Mb	1	2	1	2	2	1	1	10	-7,5	56,25

Продовження **таблиці 4.3**

X3	Час попередньої обробки даних	мс	3	3	3	4	4	4	3	24	-6,5	42,25
X4	Потенційний об’єм програмного коду	Кількість рядків коду	4	4	4	3	2	3	4	24	6,5	42,25
	Разом		10	10	10	10	10	10	10	70	0	171

Для перевірки степені достовірності експертних оцінок, визначимо наступні параметри:

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

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

де N – число експертів,

n – кількість параметрів;

б) середня сума рангів:

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

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

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

Сума відхилень по всім параметрам повинна дорівнювати 0;

г)загальна сума квадратів відхилення:

$$S = \sum_{i=1}^{N} \Delta_i^2 = 197. \ \#(4.4)$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{12S}{N^2(n^3 - n)} = \frac{12 \cdot 171}{7^2(4^3 - 4)} = 0{,}698 > W_k = 0{,}67. \#(4.5)$$

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	>	<	>	<	=	>	>	>	1,5
X1 і X3	<	<	<	<	<	<	<	<	0,5
X1 і X4	<	<	<	<	<	<	<	<	0,5
X2 і X3	<	<	<	<	<	<	<	<	0,5
X2 і X4	>	>	>	>	=	>	>	>	1,5
X3 і X4	<	<	<	>	>	>	<	<	0,5

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

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

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

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

$$K_{\text{вi}} = \frac{b_i}{\sum_{i=1}^{n} b_i} \ \#(4.7)$$

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

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:

$$K_{\text{ві}} = \frac{b_i'}{\sum_{i=1}^{n} b_i'}, \#(4.9)$$

$$b_i' = \sum_{i=1}^{N} a_{ij} b_j \#(4.10)$$

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	$K_{\text{ві}}$	b_i^1	$K_{\text{ві}}^1$	b_i^2	$K_{\text{ві}}^2$
X1	1	1,5	0,5	0,5	3,5	0,22	13.25	0,219	52,375	0,22
X2	0,5	1	0,5	1,5	3,5	0,22	14,25	0,22	56,875	0,22
X3	1,5	1,5	1	0,5	4,5	0,28	17,25	0,279	67,625	0,28
X4	1,5	0,5	1,5	1	4,5	0,28	18,25	0,28	71,125	0,28
Всього:					16	1	63	1	248	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):

$$K_K(j) = \sum_{i=1}^{n} K_{вi,j} B_{i,j} \text{ , } \#(4.11)$$

де n – кількість параметрів;

$K_{вi}$– 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	А	X1	80	18	0,22	4,5
F3	А	X2	512	32	0,22	7,04
	Б	X3	1024	17	0,28	4,76
F4	А	X4	170	33	0,28	9,24

За даними з таблиці 4.6 за формулою:

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

визначаємо рівень якості кожного з варіантів:

$$K_{Kl} = 4{,}5 + 7{,}04 + 9{,}24 = 20{,}78 ;$$

$$K_{K2} = 4{,}5 + 4{,}76 + 9{,}24 = 18{,}5 .$$

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 проєкту 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:

$$T_O = T_P \cdot К_П \cdot К_{СК} \cdot К_М \cdot К_{СТ} \cdot К_{СТ.М}, \#(4.13)$$

де T_P – трудомісткість розробки ПП;

$К_П$ – поправочний коефіцієнт;

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

$К_М$ – коефіцієнт рівня мови програмування;

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

$К_{СТ.М}$ – 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: $К_П$ = 1.8. Поправочний коефіцієнт, який враховує складність контролю вхідної та вихідної інформації для всіх семи завдань рівний 1: $К_{СК}$ = 1. Оскільки при розробці першого завдання використовуються стандартні модулі, врахуємо це за допомогою коефіцієнта $К_{СТ}$ = 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, $К_П$ =0.9, $К_{СК}$ = 1, $К_{СТ}$ =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:

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

$$T_{II} = (59{,}94 \ + 20.88 + 6.91 + 20.88) \cdot 8 = 868{,}88 \quad \text{людино-годин.}$$

Найбільш високу трудомісткість має варіант II.

В розробці бере участь один програміст з окладом 19213 грн. та один аналітик в області аналізу даних з окладом 20432. Визначимо середню зарплату за годину за формулою:

$$С_Ч = \frac{М}{T_m \cdot t} \text{грн.}, \#(4.14)$$

де М – місячний оклад працівників;

T_m – кількість робочих днів тиждень;

t – кількість робочих годин в день.

$$С_Ч = \frac{19213 \ + 20432}{2 \cdot 21 \cdot 8} = 117{,}99 \ \ \text{грн.} \#(4.15)$$

Тоді, розрахуємо заробітну плату за формулою:

$$С_{ЗП} = С_{ч} \cdot Т_{i} \cdot К_{Д}, \#(4.16)$$

де $С_{ч}$– величина погодинної оплати праці працівника;

$Т_{i}$ – трудомісткість відповідного завдання;

$К_{Д}$ – норматив, який враховує додаткову заробітну плату.

Зарплата розробників за варіантами становить:

I. $С_{ЗП} = 117,99 \cdot 852 \cdot 1,2 = 120632,98$ грн.

II. $С_{ЗП} = 117,99 \cdot 868,88 \cdot 1,2 = 123022,98$ грн.

Відрахування на єдиний соціальний внесок становить 22%:

I. $С_{ВІД} = С_{ЗП} \cdot 0.22 = 120632,98 \cdot 0.22 = 26539,26$ грн.

II. $С_{ВІД} = С_{ЗП} \cdot 0.22 = 123022,98 \cdot 0.22 = 27065,06$ грн.

Тепер визначимо витрати на оплату однієї машино-години. ($С_{М}$)

Так як одна ЕОМ обслуговує одного програміста з окладом 19213 грн., з коефіцієнтом зайнятості 0,2 то для однієї машини отримаємо:

$$С_{Г} = 12 \cdot М \cdot К_{З} = 12 \cdot 19213 \cdot 0,2 = 46111,2 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$С_{ЗП} = С_Г \cdot (1 + К_З) = 46111{,}2 \cdot (1 + 0.2) = 55333 \text{ грн.}$$

Відрахування на соціальний внесок:

$$С_{ВІД} = С_{ЗП} \cdot 0.22 = 55333 \cdot 0{,}22 = 12173{,}36 \text{ грн.}$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 21000 грн.

$$С_А = К_{ТМ} \cdot К_А \cdot Ц_{ПР} = 1.4 \cdot 0.25 \cdot 21000 = 7350 \text{ грн.,}$$

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

$К_А$ – річна норма амортизації;

$Ц_{ПР}$ – договірна ціна приладу.

Витрати на ремонт та профілактику розраховуємо як:

$$С_Р = К_{ТМ} \cdot Ц_{ПР} \cdot К_Р = 1.4 \cdot 21000 \cdot 0.08 = 2352 \text{ грн.,}$$

де $К_Р$– відсоток витрат на поточні ремонти.

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$Т_{ЕФ} =(Д_К – Д_В – Д_С – Д_Р) \cdot t_З \cdot К_В = (365 – 104 – 12 – 16) \cdot 8 \cdot 0.35 =$$

$$= 627,2 \text{ години},$$

де $Д_К$ – календарна кількість днів у році;

$Д_В$, $Д_С$ – відповідно кількість вихідних та святкових днів;

$Д_Р$ – кількість днів планових ремонтів устаткування;

t –кількість робочих годин в день;

$К_В$– коефіцієнт використання приладу у часі протягом зміни.

Витрати на оплату електроенергії розраховуємо за формулою:

$$С_{ЕЛ} = Т_{ЕФ} \cdot N_С \cdot К_З \cdot Ц_{ЕН} = 627,2 \cdot 0,2 \cdot 0,3 \cdot 4,87 = 183,3 \text{ грн.},$$

де $N_С$ – середньо-споживча потужність приладу;

$К_З$– коефіцієнтом зайнятості приладу;

$Ц_{ЕН}$ – тариф за 1 КВт-годин електроенергії.

Накладні витрати розраховуємо за формулою:

$$С_Н = Ц_{ПР} \cdot 0.67 = 21000 \cdot 0,67 = 14070 \text{ грн.}$$

Тоді, річні експлуатаційні витрати будуть:

$$С_{ЕКС} = С_{ЗП} + С_{ВІД} + С_А + С_Р + С_{ЕЛ} + С_Н, \#(4.17)$$

$$С_{ЕКС} = 55333 + 12173,36 + 7350 + 2352 + 183,3 + 14070 = 91461,66 \text{ грн.}$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$С_{М\text{-}Г} = С_{ЕКС} / Т_{ЕФ} = 91461,66 / 627,2 = 145,8 \text{ грн/год.}$$

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:

$$С_М = С_{М-Г} \cdot Т, \#(4.18)$$

I. $С_М = 145.8 \cdot 852 = 124221,6$ грн.

II. $С_М = 145.8 \cdot 868,88 = 126682,7$ грн.

Накладні витрати складають 67% від заробітної плати:

$$С_Н = С_{ЗП} \cdot 0{,}67, \#(4.19)$$

I. $С_Н = 124221{,}6 \cdot 0{,}67 = 72035{,}44$ грн.

II. $С_Н = 126682{,}7 \cdot 0{,}67 = 84877{,}4$ грн.

Отже, вартість розробки ПП за варіантами становить:

$$С_{ПП} = С_{ЗП} + С_{ВІД} + С_М + С_Н, \#(4.20)$$

I. $С_{ПП} = 120632{,}98 + 26539{,}26 + 124221{,}6 + 72035{,}44 = 343429{,}28$ грн.

II. $С_{ПП} = 123022{,}98 + 27065{,}06 + 126682.7 + 84877.4 = 361648{,}14$ грн.

4.7 Вибір кращого варіанту ПП техніко-економічного рівня

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$К_{ТЕР j} = К_{Кj} / С_{Фj}, \#(4.21)$$

$К_{ТЕР1} = 20{,}78 / 343429{,}28 = 6{,}051 \cdot 10^{-5}$,

$$К_{ТЕР2} = 18,5 / 361648,14 = 5,115 \cdot 10^{-5}.$$

Як бачимо, найбільш ефективним є перший варіант реалізації програми з коефіцієнтом техніко-економічного рівня $К_{ТЕР1}= 6,051 \cdot 10^{-5}$.

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 $К_{ТЕР} = 6,051 \cdot 10^{-5}$.

Цей варіант реалізації програмного продукту має такі параметри:

– Вибір мови для програмного продукту – Python;
– Реалізація важливої постановки з допомогою вбудованих функцій;
– Використання стандартного інтерфейсу для побудови значень.

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

4.8 Висновки до розділу

В даній частині 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.

На основі аналізу вибрано варіант реалізації програмного продукту.

ВИСНОВКИ

У даній роботі було розглянуто моделі та методи для прогнозування діяльності вітчизняного виробничого підприємства. Особливою рисою роботи стала роботи при обмеженому розмірі даних з підприємством у нестандартному стані. Для цього розглядалися якісні методи та кількісні методи, причому другі частково використовувалися для оцінки результатів перших. Також, в результаті збігу обставин якісний прогноз так само співставлявся з реальною подією, яка відбулася.

Використовувалися дані з вітчизняного підприємства “АвтоКрАЗ”, яке є відомим виробником автомобілів і вирізняється якраз таки проблемним

фінансовим станом та скандалами навколо нього. Для роботи були взяті як і дані з фінансових звітів та балансу, так і дані незалежних експертів від одного з тематичних видань.

У ході роботи використовувався метод морфологічного аналізу та різні методи авторегресій. У ході кількісного аналізу було частково підтверджено певні результати якісного аналізу, хоча і було помічено неточність, яка спричинена специфічно заданим результатом від експерта. Можна вважати, що це є підтвердженням важливості використання кількісних і якісних методів разом з перевіркою других через перші.

Було створено пропрієтарний програмний продукт, який проводить необхідні дослідження та виводить результат. Платформою для нього послужила мова програмування Python.

З проробленої роботи витікає, що є можливість продовжити дослідження, більше покладаючися на взаємодію кількісних та якісних методів. Подібне вже застосовується в області системного аналізу, що дає преспективу. Серед можливих наступних напрямків – більший розгляд взаємодії морфологічного аналізу з методами на часових рядах, порівняння з іншими, складнішими кількісними методами, застосування ансамблів. Також, можливо, є раціональним розгляд одразу модернізації морфологічного аналізу із вкрапленням додаткових кількісних результатів.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Charmayne Smith: How to Find Sales With Contribution Margin Ratio & Variable Costs. URL: https://smallbusiness.chron.com/sales-contribution-margin-ratio-variable-costs-35363.html (дата звернення 07.05.2023 р)
2. Brian Beers: Financial Characteristics of a Successful Company. 2022. URL: https://www.investopedia.com/articles/stocks/08/successful-company-qualities.asp (дата звернення 02.05.2023 р)
3. Delphi method tool for research and organizational use of Delphi method that is based on the development work starting in 1998. URL: https://www.edelphi.org/ (дата звернення 25.04.2023 р)

4. Swedish Morphological Society: Advanced Computer Support for General Morphological Analysis. URL: https://www.swemorph.com/macarma.html (дата звернення 25.04.2023 р)
5. TSP 4.5 Reference Manual. URL: https://eml.berkeley.edu/wp/tsp_ref/subpdf.htm (дата звернення 25.04.2023 р)
6. Home page for TSP International, maker of the econometrics package TSP. Архів веб-сторінки. URL:https://web.archive.org/web/20170107011016/http://www.tspintl.com/ (дата звернення 12.05.2023 р)
7. Brockwell, P. J., & Davis, R. A. Introduction to time series and forecasting (3rd ed). New York, USA: Springer, 2016. 437 p.
8. Robert F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica Vol. 50, No. 4 (Jul., 1982), pp. 987-1007 p.
9. Панкратова Н.Д., Савченко І.О. Застосування методу морфологічного аналізу до задач технологічного передбачення // Наукові праці / Миколаївський держ. гуманітарний ун-т ім. Петра Могили комплексу НаУКМА. Сер. Комп’ютерні технології, системний аналіз, моделювання. 2008. 90, вип. 77. — С. 6—13.
10. Brooks, Chris. Introductory Econometrics for Finance (3rd ed.). Cambridge: Cambridge University Press.2014. p. 461. ISBN 9781107661455.
11. Shieh, Gwowen. "Improved shrinkage estimation of squared multiple correlation coefficient and squared cross-validity coefficient". Organizational Research Methods. 11 (2). 2008. 407 p.
12. Hoornweg, Victor. "Part II: On Keeping Parameters Fixed". Science: Under Submission. Hoornweg Press. 2018.
13. Табога, Марко. «Тест балів», лекції з теорії ймовірностей та математичної статистики. Kindle Direct Publishing. 2021.

14. Iver Johansen. Scenario modelling with morphological analysis. Technological Forecasting and Social Change Volume 126, January 2018, 116-125 p.
15. П.І. Бідюк, Н.В. Кузнєцова, О.М. Терентьєв. Система підтримки прийняття рішень для аналізу даних, Наукові вісті НТУУ «КПІ». 2011. С. 48-61.
16. de Cheveigné A., Nelken I. Filters: When, Why, and How (Not) to Use Them. Neuron, Volume 102, Issue 2, 17 April 2019. 280-293 p.
17. "АвтоКрАЗ", "Укрнафта", "Мотор Січ" та "Запоріжтрансформатор" переходять у власність держави. Промисловий Портал, 07.11.2022. URL: https://uprom.info/news/avtokraz-ukrnafta-motor-sich-ta-zaporizhtransformator-perehodyat-u-vlasnist-derzhavy/ (дата звернення 15.05.2023 р).
18. Annor-Antwi A., Al-Dherasi A., «Application of Artificial Intelligence in Forecasting: A Systematic Review», с. 10.
19. Python 3.11.3 documentation. URL: https://docs.python.org/3/ (дата звернення 30.04.2023 р).
20. Vapnik V. Statistical Learning Theory. – New York: Wiley, 1998.
21. Кузнєцова Н.В. Методи і моделі аналізу, оцінювання та прогнозування ризиків у фінансових системах [Текст] : дис. ... д.т.н.: 01.05.04 / Кузнєцова Н.В., Київ, 2018., 415 с.
22. Hyndman R., Athanasopoulos J. Forecasting: Principles and Practice, 2018, 382 p.
23. Box G., Jenkins G., Reinsel G., Ljung G., Time Series Analysis: Forecasting and Control, Wiley, 2015, 681 p.
24. Hamilton J., Time Series Analysis., Princeton University Press, 1994, 820 p.
25. Piller F., Nitsch V., Lüttgens D., Mertens A., Pütz S., Dyck M., Forecasting Next Generation Manufacturing, Springer, 2022, 158 p.

ДОДАТОК А Лістинг програми

```
Import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.stattools import kpss
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller
from arch import arch_model

def adf_test(timeseries):
    print("Результати тесту Дікі-Фуллера:")
    dftest = adfuller(timeseries, autolag="AIC")
    dfoutput = pd.Series(
        dftest[0:2],
        index=[
            "Test Statistic",
            "p-value"
        ],
    )
    for key, value in dftest[4].items():
        dfoutput["Critical Value (%s)" % key] = value
    print(dfoutput)


def kpss_test(timeseries):
    print("Результату тесту KPSS:")
    kpsstest = kpss(timeseries, regression="c",
nlags="auto")
    kpss_output = pd.Series(
        kpsstest[0:2], index=["Test Statistic", "p-value"]
    )
    for key, value in kpsstest[3].items():
        kpss_output["Critical Value (%s)" % key] = value
```

```
    print(kpss_output)



mvzip = pd.read_csv("mvzip.csv",index_col="index")
bzalt = pd.read_csv("baseAlternatives.csv")
mvzip.fillna(0,inplace=True)
bzalt.fillna(0,inplace=True)
print(mvzip)
print(bzalt)

tabl = pd.DataFrame(columns=['1', '2', '3','4','5','6','p0','C','p1','p_final'])
n=0

for i in bzalt['1'].index:
  one = bzalt['1'][i]
  ont = '1.'+str(i+1)
  for ii in bzalt['2'].index:
    two = bzalt['2'][ii]
    tw = '2.' + str(ii + 1)
    for iii in bzalt['3'].index:
      three = bzalt['3'][iii]
      th = '3.' + str(iii + 1)
      for iiii in bzalt['4'].index:
        four = bzalt['4'][iiii]
        fo = '4.' + str(iiii + 1)
        for iiiii in bzalt['5'].index:
          five = bzalt['5'][iiiii]
          fi = '5.' + str(iiiii + 1)
          for iiiiii in bzalt['6'].index:
            six = bzalt['6'][iiiiii]
            sx = '6.' + str(iiiiii + 1)
            p = one*two*three*four*five*six
            if(p!=0):
              C=1
              for l in (ont,tw,th,fo,fi):
                for ll in (tw,th,fo,fi,sx):
                  C*=(1+mvzip[l][float(ll)])
              p1 = p*C
              dat = {'1':ont, '2':tw, '3':th,'4':fo,'5':fi,'6':sx,'p0':p,'C':C,'p1':p1,'p_final':0}
              tabl = tabl.append(dat,ignore_index=True)
              n+=1
summa = tabl['p1'].sum()
print("sum = ",summa)
tabl['p_final'] = tabl['p1']/summa
print(tabl)
print(tabl['p_final'].sum())


for i in bzalt.columns:
  for j in bzalt[i].index:
    bzalt[i][j]=tabl[tabl[i]==i+'.'+str(j+1)]['p_final'].sum()
print(bzalt)
#print(tabl[tabl['p_final']==tabl['p_final'].mean()])


mvzip2 = pd.read_csv("mvzip2.csv",index_col="index")
print(mvzip2)
tabl2 = pd.DataFrame(columns=['1', '2', '3','4','5','6','p_final','r71','r72','r73', 'r81', 'r82','r83', 'r84', 'r85','R71','R72','R73', 'R81', 'R82','R83', 'R84', 'R85','pR71','pR72','pR73', 'pR81', 'pR82','pR83', 'pR84', 'pR85'])
for i in range(tabl['1'].size):
  rs = []
  for r in mvzip2.columns[:3]:
    C = 1
    for l in tabl2.columns[:6]:
      C *= (1 + mvzip2[r][float(tabl[l][i])])
    rs.append(C)
  rs1 = rs/sum(rs)
  rs2 = rs1*tabl['p_final'][i]
  rss = []
  for r in mvzip2.columns[3:]:
    C = 1
    for l in tabl2.columns[:6]:
      C *= (1 + mvzip2[r][float(tabl[l][i])])
    rss.append(C)
  rss1 = rss / sum(rss)
  rss2 = rss1 * tabl['p_final'][i]

ft={'1':tabl['1'][i],'2':tabl['2'][i],'3':tabl['3'][i],'4':tabl['4'][i], '5':tabl['5'][i], '6':tabl['6'][i], 'p_final':tabl['p_final'][i],'r71': rs[0],'r72':rs[1],'r73':rs[2],'r81':rss[0],'r82':rss[1],'r83':rss[2] ,'r84':rss[3],'r85':rss[4],'R71': rs1[0],'R72':rs1[1],'R73':rs1[2],'R81':rss1[0],'R82':rss1[1],' R83':rss1[2],'R84':rss1[3],'R85':rss1[4],'pR71':rs2[0],'pR72 ':rs2[1],'pR73':rs2[2],'pR81':rss2[0],'pR82':rss2[1],'pR83':rs s2[2],'pR84':rss2[3],'pR85':rss2[4]}
  tabl2 = tabl2.append(ft,ignore_index=True)
print(tabl2)
tabl2.to_excel("endindVariants.xlsx")
ed1 = pd.Series([tabl2['pR71'].sum(), tabl2['pR72'].sum(), tabl2['pR73'].sum()],name="7.Дії держави")
ed2 = pd.Series([tabl2['pR81'].sum(), tabl2['pR82'].sum(), tabl2['pR83'].sum(), tabl2['pR84'].sum(), tabl2['pR85'].sum()],name="8.Дії керівництва компанії")
df=pd.concat([ed1,ed2],axis=1)
print(df)



df=pd.read_excel("forIncome.xlsx",index_col="index")
#print(df["income"])


decomposition=seasonal_decompose(df["income"],period =1)

decomposition.plot()
plt.show()
df=df.drop(index=["2021-01-01","2022-01-01"])
#df.index=pd.DatetimeIndex(df["index"])
#df=df.drop(columns="index")
#df=df.drop(columns="EPS")
#df=df.diff().drop(index="2010-01-01")

#print(df)
#df["income"].plot()
#plt.show()
```

```
#df["income"].plot()
#plt.show()"""
adf_test(df.income)
kpss_test(df.income)

plot_acf(df.income)
plot_pacf(df.income,lags=4,method='ywm')
plt.show()

modelA=ARIMA(df.income,order=(2,0,0))
resA=modelA.fit()
print(resA.summary())
resA.plot_diagnostics()
"""model=SARIMAX(df.income,order=(1,2,0))
res=model.fit()
print(res.summary())
res.plot_diagnostics()"""
redis=resA.resid
modelArch=arch_model(redis,vol='GARCH',p=2,q=1)
model_fitArch=modelArch.fit()
print(model_fitArch.summary())
print(resA.forecast(steps=2))

"""predicted_et=garch_forecast.mean['h.1'].iloc[-1]
print(model_fitArch.forecast(horizon=1).variance)"""
print(df)


plt.show()


df=pd.read_excel("forNumber.xlsx",index_col="index")
#print(df["income"])

df["produced"]=np.log(df["produced"]
#df=df.drop(index=["2010-01-01"])#,"2021-01-01"
print(df)
decomposition=seasonal_decompose(df["produced"],perio
d=1)
"""plt.figure(figsize=(10,15))
plt.subplot(411)
plt.plot(df["income"])
plt.subplot(412)
plt.plot(decomposition.trend)
plt.show()"""
decomposition.plot()
plt.show()
#df=df.drop(index=["2021-01-01","2022-01-01"])
#df.index=pd.DatetimeIndex(df["index"])
#df=df.drop(columns="index")
#df=df.drop(columns="EPS")
#df=df.diff().drop(index="2010-01-01")
#print(df)
#df["income"].plot()
#plt.show()
#df["income"].plot()
#plt.show()"""
adf_test(df.produced)
kpss_test(df.produced)

plot_acf(df.produced)
plot_pacf(df.produced,lags=4,method='ywm')
plt.show()

modelA=ARIMA(df.produced,order=(1,0,1))
resA=modelA.fit()
print(resA.summary())
resA.plot_diagnostics()
"""model=SARIMAX(df.income,order=(1,2,0))
res=model.fit()
print(res.summary())
res.plot_diagnostics()"""
redis=resA.resid



print(np.exp(resA.forecast(steps=5)))
print(df)


plt.show()
```

ДОДАТОК Б Презентація

Моделі та прогнози діяльності виробничого підприємства

Підготував: студент Ка-95 Гулкевич Б.Ю.
Науковий керівник: Бідюк П.І.

Предмет, об'єкт та мета роботи

Об'єкт: два пов'язані набори даних щодо підприємства, а саме його фінансові результати та оцінка підприємства від експертів.

Предмет: метод морфологічного аналізу, методи та моделі прогнозування часових рядів: моделі авторегресії, авторегресії – ковзного середнього, авторегресії – інтегрованого ковзного середнього, авторегресії з умовною гетероскедастичністю та взаємодію з часовими рядами, підготовкою даних на їх основі.

Мета дослідження: розробка та реалізація програмного продукту для прогнозування результатів діяльності виробничого підприємства якісними та кількісними методами.

Постановка задачі

- наявні дані про підприємство зі специфічним фінансовим станом, їх розмір невеликий;
- розгляд різних моделей та методів для аналізу та прогнозування стану виробничого підприємства;
- використовується метод морфологічного аналізу та методи аналізу часових рядів;
- потім результати обох шляхів розглядаються та порівнюються у можливих місцях;
- за допомогою кількісних методів підтвердження якісних.

Актуальність проблеми

Стан великої кількості підприємств в Україні, коли статистичних даних недостатньо.

Складна ситуація в країні та світі загалом.

Існує можливість підвищити точність аналізу та прогнозування. Часткова компенсація недоліків якісних методів.

Про дані підприємства АвтоКрАЗ

- найбільший виробник вантажівок в Україні;
- постачає продукцію як для силових та комунальних установ України, так і на експорт;
- відомий своїм банкротством попри величезні замовлення;
- Існує невелика кількість доступних статистичних даних.

Про розроблений програмний продукт

Пропрієтарна програма, реалізована на Python у середовищі JetBrains PyCharm.

Ділиться на дві частини: для морфологічного аналізу та для кількісного.

Перша частина: приймає оцінену морф таблицю та матрицю взаємозв'язків 1-го етапу, отримує з цього результуючу морф таблицю 1-го етапу. Надалі, приймається матриця зв'язків 1-го і 2-го етапів та морф результат попереднього етапу, з них визначається результат 2-го етапу морфологічного аналізу. В процесі аналізу виводяться таблиці проміжних значень.

Друга частина програми обробляє статистичні дані, будує моделі ARIMA та за потреби ARCH/GARCH, після цього оцінюються прогнози. Проводиться

Морфологічний аналіз перший етап

Характеристичні параметри					
Виготовлені машини на місяць	Локалізація	Фінансовий стан	Стабільна якість	Готовність йти назустріч	Забезпечення сервісу
1	2	3	4	5	6
1.1 До 20	2.1 Повна	3.1 Позитивний	4.1 Так	5.1 Так	6.1 Повне
1.2 До 50	2.2 Частково імпорт	3.2 Ще більше погіршення	4.2 Ні	5.2 Шукають як ні	6.2 Зачастину
1.3 Да 100	2.3 Більність імпорт	3.3 Збереження негативної ситуації			6.3 Реальні специалісти
1.4 Понад 100	2.4 Готові комплекти				6.4 Повний сервіс

Характеристичні параметри					
Виготовлені машини на місяць	Локалізація	Фінансовий стан	Стабільна якість	Готовність йти назустріч	Забезпечення сервісу
1	2	3	4	5	6
1.1 = 0.2	2.1 = 0.1	3.1 = 0.3	4.1 = 0.4	5.1 = 0.4	6.1 = 0.3
1.2 = 0.3	2.2 = 0.2	3.2 = 0.2	4.2 = 0.6	5.2 = 0.6	6.2 = 0.3
1.3 = 0.3	2.3 = 0.3	3.3 = 0.5			6.3 = 0.3
1.4 = 0.2	2.4 = 0.4				6.4 = 0.1

Морфологічний аналіз - перший етап, результат

Характеристичні параметри					
Виготовлені машини на місяць	Локалізація	Фінансовий стан	Стабільна якість	Готовність йти назустріч	Забезпечення сервісу
1	2	3	4	5	6
1.1 = 0.103036	2.1 = 0.121584	3.1 = 0.316235	4.1 = 0.438103	5.1 = 0.412371	6.1 = 0.254441
1.2= 0.295474	2.2 = 0.195204	3.2 = 0.228034	4.2 = 0.561897	5.2= 0.587629	6.2 = 0.345506
1.3 = 0.355545	2.3 = 0.292805	3.3 = 0.455731			6.3 = 0.326559
1.4 = 0.245945	2.4 = 0.390407				6.4 = 0.073494

Морфологічний аналіз - перший етап, результат

Характеристичні параметри					
Виготовлені машини на місяць	Локалізація	Фінансовий стан	Стабільна якість	Готовність йти назустріч	Забезпечення сервісу
1	2	3	4	5	6
1.1 = 0.103036	2.1 = 0.121584	3.1 = 0.316235	4.1 = 0.438103	5.1 = 0.412371	6.1 = 0.254441
1.2= 0.295474	2.2 = 0.195204	3.2 = 0.228034	4.2 = 0.561897	5.2= 0.587629	6.2 = 0.345590
1.3 = 0.355545	2.3 = 0.292805	3.3 = 0.455731			6.3 = 0.326559
1.4 = 0.245945	2.4 = 0.390407				6.4 = 0.073494

Морфологічний аналіз - другий етап

7.1 Припинення підтримки	8.1 Ліквідація	Реакції держави	Реакції керівництва
7.2 Націоналізація	8.2 Продаж	7	8
7.3 Закінчення нових контрактів	8.3 Покращення наявних можливостей виробництва	7.1 = 0.327135	8.1 = 0.050224
	8.4 Пошук нових контрактів та розширення під них	7.2 = 0.389775	8.2 = 0.157151
	8.5 Нічого не робити	7.3 = 0.383090	8.3 = 0.267478
			8.4 = 0.194689
			8.5 = 0.330458

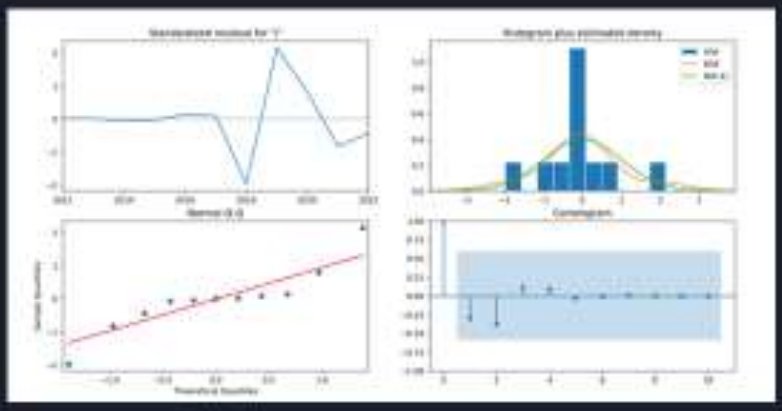
Результати для фінансів ARIMA(1,2,0)

AIC= 357.492;

HQIc= 356.990;

BIC=-358.288.

DW=0.97



Результати прогнозування для фінансів ARIMA(1,2,0)

2022-01-01 -714849.492019
2023-01-01 -605308.621261
2024-01-01 -649081.423132
2025-01-01 -636267.049874
2026-01-01 -641345.909880

Підходить до результатів

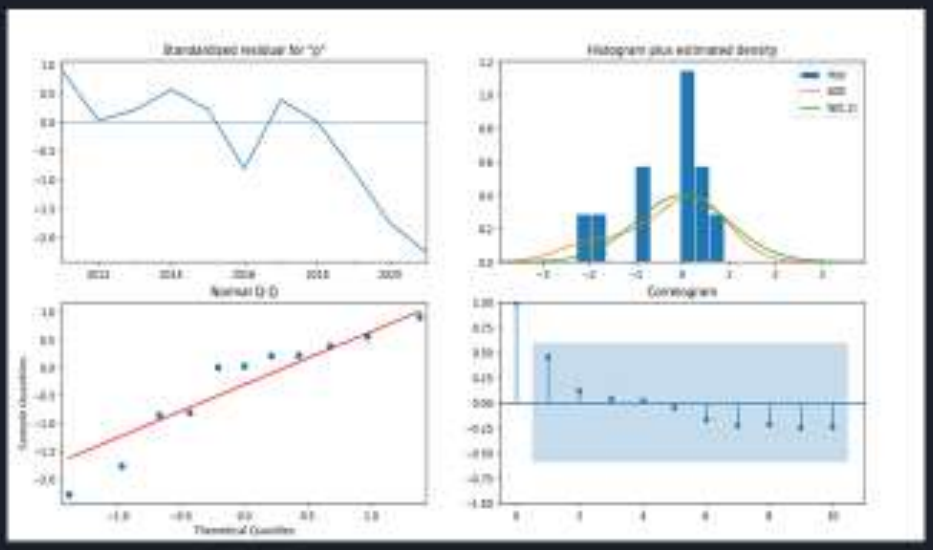
мофологічного

аналізу

Модель	MSE	MAE	MAPE
ARIMA(1,2,0),	0.40	0.45	0.39
ARIMA(1,2,0)+ GARCH(2.1)	0.39	0.40	0.38988

Результати моделі для обсягів виробництва ARIMA(1,0,1)

AIC=38.686,
BIC=40.626,
HQIC=37.968.
DW = 1.0913



Прогнозування для обсягів виробництва ARIMA(1,0,1)

- 2022 1.779350 тобто 1-2 машини;
- 2023 2.828967 тобто 2-3 машини;
- 2024 4.260237 тобто 4-5 машин;
- 2025 6.115552 тобто 6-7 машин.

MSE	MAE	MAPE
0.46	0.48	0.45

Висновки

Було виконано дослідження та експеримент з побудови моделей та оцінювання прогнозів для виробничого підприємства з проблемними даними. Для України подібна проблема, на жаль, є поширеною.

Виконано аналіз якісними і кількісними методами за допомогою пропрієтарної програми. Моделі на основі часових рядів показали посередні результати через проблемні вихідні дані. За допомогою кількісних методів вдалося частково підтвердити результати якісного аналізу.

Підтвердження якісних методів кількісними (і навпаки) має місце при аналізі та прогнозуванні та дозволяє отримувати кращі результати при більшій невизначеності.

Перспективи розвитку результатів дослідження:

- створення зручної системи підтримки прийняття рішень, яка полегшить оцінювання діяльності виробничих підприємств та потенційно може стати конкурентоздатним продуктом;
- можливо додати інші кількісні та якісні методи;
- дослідження у цьому напрямі дозволять збільшити точність аналізу та прогнозування загалом, а також краще оцінювати результати якісних методів.


Дякую за увагу