

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»  
Інститут прикладного системного аналізу  
Кафедра математичних методів системного аналізу**

«До захисту допущено»  
В.О. Завідувача кафедри  
\_\_\_\_\_ О.Л. Тимощук

**Дипломна робота**

**на здобуття ступеня бакалавра  
за освітньо-професійною програмою “Системний аналіз і управління”  
спеціальності 124 «Системний аналіз»  
на тему: «Моделі та методи прогнозування процесів фінансового ринку»**

Виконала:

студентка IV курсу, групи КА-63  
Корж Богдан Сергійович

\_\_\_\_\_

Керівник:

доцент кафедри ММСА, д.т.н  
Кузнєцова Н.В.

\_\_\_\_\_

Консультант з економічного розділу:  
доцент, к.е.н. Шевчук О.А.

\_\_\_\_\_

Консультант з нормоконтролю:  
доцент кафедри ММСА, к.т.н.  
Коваленко А.Є.

\_\_\_\_\_

Рецензент:

к.т.н., доцент кафедри СП  
Безносик О.Ю

\_\_\_\_\_

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

Київ – 2020 року

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

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітньо-професійна програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

В.о.завідувача кафедри

\_\_\_\_\_ Оксана ТИМОЩУК

«\_\_» \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ**

**на дипломну роботу студенту**

**Корж Богдан Сергійович**

1. Тема роботи «Моделі та методи прогнозування процесів фінансового ринку», керівник роботи Кузнецова Наталія Володимирівна, д.т.н, доцент, затверджені наказом по університету від «25» травня 20 20 р. № 1143-с
2. Термін подання студентом роботи 08 травня 2020 року \_\_\_\_\_
3. Вихідні дані до роботи
4. Зміст роботи
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо)

## 6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	Шевчук О.А., к.е.н., доцент		

7. Дата видачі завдання \_\_\_\_\_

## Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Формулювання теми бакалаврської роботи.	15.10.2019 – 30.11.2019	
2	Пошук та огляд спеціальної літератури за темою дослідження.	01.12.2019 - 23.01.2020	
3	Ознайомлення із літературою та програмним забезпеченням для машинного навчання, необхідними для написання роботи.	25.01.2020 – 04.02.2020	
4	Пошук вхідних даних для аналізу.	05.02.2020 – 10.03.2020	
5	Порівняння методів вирішення задачі оптимізації параметрів, виділення їх недоліків.	11.03.2020 – 15.04.2020	
6	Робота над експериментальною частиною	16.04.2020 – 17.05.2020	
7	Підготовка презентації для захисту	18.05 – 01.06.2020	
8	Попередній захист дипломної роботи	01.06.2020 – 03.06.2020	
9	Захист дипломної роботи	15.06.2020 – 18.06.2020	

Студент

Богдан КОРЖ

Керівник

Наталія КУЗНЄЦОВА

## РЕФЕРАТ

Дипломна робота: \_ с., \_ рис., \_ табл., \_ додатки, \_ джерел.

### МОДЕЛІ ТА МЕТОДИ ПРОГНОЗУВАННЯ ПРОЦЕСІВ ФІНАНСОВОГО РИНКУ

Дана робота присвячена моделюванню та прогнозуванню процесів фінансового ринку за допомогою авторегресійних моделей.

Метою дипломної роботи є дослідити процес побудови моделей фінансового ринку за допомогою моделей часових рядів. Окрім цього реалізація власного програмного продукту за допомогою мови програмування Python та її бібліотек та порівняння його результатів за статистичним пакетом Eviews.

Об'єктом дослідження є процеси фінансового ринку, а точніше – утворення курсу валют.

## ABSTRACT

Bachelor thesis: p., tabl., fig., add. and references

### MODELS AND METHODS OF FORECASTING FINANCIAL MARKET PROCESSES

This work is devoted to modeling and forecasting financial market processes using autoregressive models.

The aim of the thesis is to investigate the process of building financial market models using time series models. In addition, the implementation of its own software product using the programming language Python and its libraries and compare its results with the statistical package Eviews.

The object of research is the processes of the financial market, - the formation of the exchange rate.

## ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ .....	8
ВСТУП .....	9
РОЗДІЛ 1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ .....	11
1.1 Основні відомості про фінансові ринки .....	11
1.2 Аналіз підходів до прогнозування .....	17
1.3 Означення та використання трендів .....	23
1.4 Експоненційне згладжування .....	25
1.5 Програмне забезпечення, що дозволяє здійснювати прогнозування .....	26
1.6 Висновки до розділу .....	26
РОЗДІЛ 2 ПОБУДОВА МАТЕМАТИЧНИХ МОДЕЛЕЙ ТА ПІДХОДИ ДО ПРОГНОЗУВАННЯ.....	27
2.1 Основні поняття .....	27
2.1.1 Означення часового ряду .....	27
2.1.2 Складові частини часового ряду .....	27
2.1.3 Аналіз гетероскадестичності .....	28
2.1.4 Аналіз нелінійності .....	31
2.1.5 Перевірка стаціонарності .....	34
2.1.6 Означення ковзного середнього .....	36
2.1.7 Кореляційна матриця.....	38
2.2 Оцінка адекватності моделі .....	38
2.4 Висновки до розділу .....	41
РОЗДІЛ 3 ПОБУДОВА МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ КУРСУ ВАЛЮТ .....	42
3.1 Опис даних.....	42
3.2 Побудова та прогноз моделі для ряду курсу валют .....	42
3.3 Висновки до розділу .....	51
РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ .....	52
4.1 Постановка задачі техніко-економічного аналізу .....	52
4.2 Варіанти реалізації основних функцій.....	53
4.3 Обґрунтування системи параметрів програмного продукту .....	55

4.4 Економічний аналіз варіантів розробки .....	59
4.5 Висновки до розділу .....	62
СПИСОК ЛІТЕРАТУРИ.....	63
ВИСНОВКИ ЗА РОБОТОЮ.....	63
ДОДАТОК А (ілюстративний матеріал доповіді) .....	65
ДОДАТОК Б (код програмного продукту).....	74

## ПЕРЕЛІК СКОРОЧЕНЬ

АР – авторегресія

АРКС – авторегресія ковзного середнього

АРУГ – авторегресійна умовна гетероскедастичність

УАРУГ – узагальнена авторегресійна умовна гетероскедастичність

DW – Durbin-Watson

AIC – Akaike information criterion

## ВСТУП

Процеси на фінансових ринках є дуже багатограними та мають величезну кількість проявів – на економіку підприємства, макроекономічну ситуації країни, а також добробут кожного громадянина. На даний момент вони вже є частиною повсякденного життя кожної людини. Отже, їх вивчення потребує більших зусиль та уваги аналітиків, науковців та спеціалістів у галузі фінансів. Свідченням цього є загальносуспільний інтерес до цієї галузі, кількість наукових праць та які йому присвячені. Зокрема, аналіз ринку такого його сегменту, який відповідає за облігації тісно пов'язаний з зовнішньоторговою політикою держави та макроекономічною стабільністю. Ринок акцій як його частина є своєрідним набором індикаторів розвитку економіки країни, ділової активності та навіть фінансової грамотності населення.

Зараз вже виникла певна проблематика внаслідок збору та накопичення великих обсягів інформації, їх у свою чергу необхідно обробляти для того, щоб прийняте зважене та безпечне рішення. Зокрема проблеми математичного моделювання мають чіткий взаємозв'язок із науковою методологією системного аналізу, аналізом часових рядів та використанням моделей рекурентних нейронних мереж, наприклад, LSTM. Вони на відміну від часових рядів, моделей з інтегрованим ковзним середнім можуть пристосовуватись до тих чи інших лагів з плином часу.

Побудова та тестування моделей проводилася на історичних даних.

Дана робота складається з 4 розділів.

Перший розділ присвячений розгляду особливостей предметної області. У ньому описуються методи прогнозування, що вже існують. Порівнюються з обраними для подальшої роботи методами.

У другому розділі описуються необхідні для побудови та аналізу моделей часових рядів математичні відомості. Ці відомості включають основні

теоретичні означення, статистичні тести для перевірки коінтегрованості, стаціонарності тощо. Також розглянуті оцінки адекватності моделі та критерії оцінки точності прогнозу.

Третій розділ присвячений опису даних, які аналізуються побудові та порівнянню моделей.

У четвертому розділі наведений економіко-вартісний аналіз створеного програмного продукту.

## РОЗДІЛ 1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ

### 1.1 Основні відомості про фінансові ринки

Перед тим, як приступити до аналізу фінансових ринків необхідно спочатку навести деякі теоретичні відомості про них, дати безпосереднє означення поняття фінансового ринку.

Фінансовим ринок є особливою є важливою ланкою у процесі побудови економічних відносин. У фінансовій системі держави він є важливою системою фінансових відносин. Зокрема, на рівні країни чи держави він займає одну із ключових ланок, що виникають при побудові фінансових відносин. Він сприяє функціонуванню, ефективності та якості роботи інших складових і рівнів економічної системи. З точки зору економіки його можна класифікувати як певний набір економічних відносин.

Фінансові ринки можуть набувати різних форм, а також бути місцем проведення операцій та транзакцій, але незалежно від цього вони виконують декілька базових функцій: встановлення ціни, оцінка вартості активів, арбітраж, нарощення капіталу, забезпечують комерційні транзакції, забезпечують притік інвестицій. Незважаючи на те, що вони можуть набувати різних форм, вони є дуже високоорганізованими та складними системами, які мають відповідні функції. Функція встановлення ціни відповідає забезпеченню виявлення та формування ціни різних предметів, цін, товарів чи послуг на основі того, за якими цінами суб'єкти, які є, в свою чергу, є економічними агентами, що діють на ньому готові купувати чи продавати блага. Окрім цього ринкові ціни наразі є найкращим способом виявлення вартості компанії, її активів чи майне. Це є важливим не лише для тих, хто продає чи хоче стати власником статутних фондів чи капіталу, а й, наприклад, для страхових чи аудиторських компаній, для яких важливою є оцінка не тільки капіталу, а й поточних зобов'язань компанії чи підприємства.

Арбітраж, чи іншими словами посередницька функція, полягає у тому, що завдяки діяльності гравців на фінансовому ринку ціни на ресурси, товари, послуги у країнах з економікою, що розвивається, стають більш наближеними до реальних: для таких економік є типовою ситуація нерівномірності цін у різних регіонах чи областях.

Завдяки фінансовим ринкам стає можливим примножувати капітал, розширювати та збільшувати вартість статутних фондів прийоми випуску акцій, облігацій та інших деяких фінансових інструментів. Забезпечення, яке вони пропонують для здійснення комерційних транзакцій дозволяє реалізувати оплати для експортерів та імпортерів, виплати працівникам тощо. Завдяки інвестуванню на фінансових ринках у керівників з'являється можливість примножити капітал завдяки отриманню прибутку з фондів, які зараз не використовуються, а також акумулювати резерви, які дозволять отримати дохід у майбутньому. Окрім цього їх переваги також включають високу ліквідність того, чим торгують, прозорість – наявність оперативної і повної інформації про торги та ціни, адже люди загалом менше торгують на ринках з менш прозорими правилами гри, надійність – коли мова йде, про те щоб швидко завершити проведення торгів відповідно до раніше погоджених умов, а також наявність чітко регламентованих правових процедур та низькі транзакційні втрати. На даний момент основними факторами розвитку для ринків технології (майже всі уявлення про ринки були в певний момент змінені технологіями: великими можливостями для обчислень та дешевих телекомунікацій), дерегуляція (незважаючи на те, що пройшло не так багато того часу, коли уряди суворо контролювали фінансові ринки, наразі акценти змістилися в сторону захисту споживачів та забезпечення стабільності). Окремим послідовним фактором була лібералізація, яка була головним чином пов'язана з дерегуляцією, адже зменшилась роль уряду у фінансових процесах та величезна кількість обмежень для суб'єктів була знижена. Великі ринкові економіки навіть дозволили закордонним компаніям увійти на ринки, які до цього призначались лише для національних виробників. Останній фактор

призвів у свою чергу до консолідації, яка відбувалась одночасно з глобалізаційними процесами.

Визначити загальний обсяг операцій, що здійснюються на фінансовому ринку досить складно. Невизначеність полягає в тому, що різні точки зору наполягають на тому, що до видів активностей, що на них відбуваються потрібно включати різні види транзакцій. Можна оцінювати за товарообігом ринку капіталів, а можна оцінити за ціною інструментів, які на них торгуються. До того ж при оцінці обсягів дуже часто виключають такі важливі речі як страховий андеррайтинг або торгівля похідними фінансовими інструментами. Іншим способом виміру може бути оцінка обсягів транскордонного фінансування.

На організаційному рівні необхідно перш за все розглядати та наводити фінансові інституції та організації, що діють на ньому. Саме за участі та в результаті діяльності цих організацій відбувається процес перерозподілу капіталу (коштів) від власників, до посередників, від посередників до позичальників (безумовно, й у зворотньому напрямку). До інституцій можна віднести страхові компанії, біржі (валютні та фондові), фонди, клірингові організації.

Серед видів фінансових ринків зазвичай науковці виокремлюють такі такі: фондовий ринок, ринок облігацій, ринок товарів, ринок похідних фінансових інструментів.

На фондовому ринку торгують акціями компаній (як державних так і приватних), кожна акція продається за певною ціною. На даний момент існують різні показники, які відображають стан та процеси фондового ринку: наприклад, промисловий індекс Доу-Джонса (DJA) (рис.1.1) чи S&P 500.



Рисунок 1.1 – Індекс Доу-Джонса (DJA)

На ринку облігацій торгують облігаціями компаній чи урядів, придбавши облігацію покупець отримує протягом певного узгодженого періоду дохід та відсотки в залежності від дохідності облігацій (рис. 1.2).

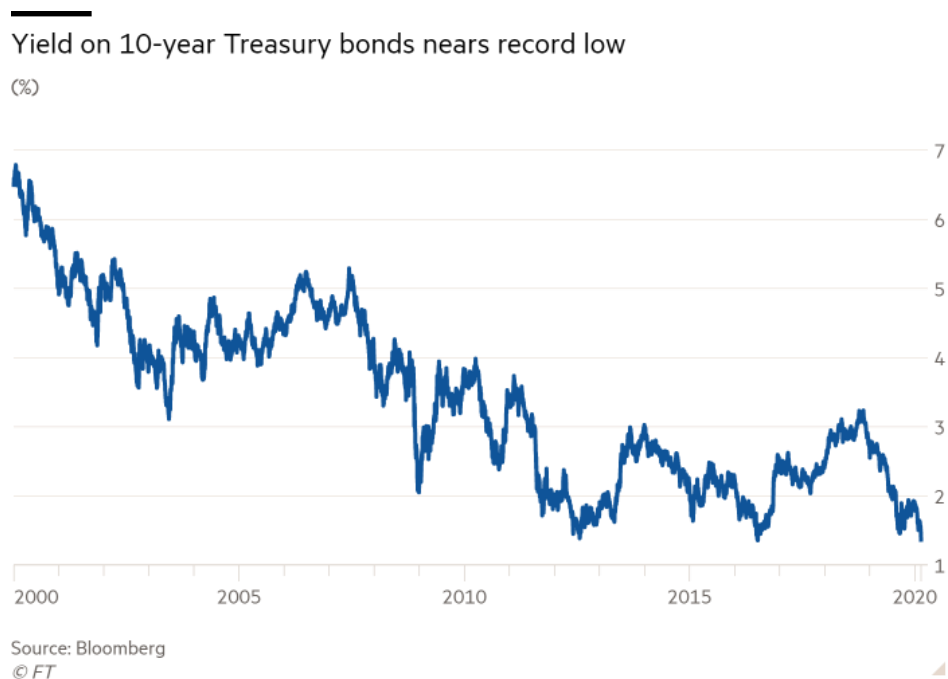


Рисунок 1.2 – Дохідність облігацій уряду США

Ринок ресурсів дозволяє економічним агентам придбати або продати певні натуральні ресурси, наприклад, нафту чи цінні метали. Ринок ресурсів розвинувся внаслідок непередбачуваності на ціни товарів на ньому. Яскравим прикладом непередбачуваності цін є падіння цін на нафту у 2020 році (рис. 1.3) або падіння цін на енергоносії в Україні, зокрема газу для населення.

## Oil price in 2020

Brent crude, US dollars per barrel



Source: Bloomberg, 02 April 2020, 09:30 GMT

**BBC**

Рисунок 1.3 – Динаміка цін на нафту у 2020 році

Розглядаючи та пдануючи діяльність на будь-якому із сегментів фінансового ринку економічний агент незалежно від його виду діяльності чи організації, яку він уповноважений представляти повинен враховувати особливості укладення контрактів при роботі з фінансовими активами з огляду на специфіку сегменту, на якому це відбувається.

На різних ринках один і той самий актив може бути товаром незалежно від інших. Скажімо, кредит, який взяли у доларах або іншій валюті — може

одночасно потрапити, як у поле діяльності грошового, так і валютного ринків. Безумовно, характерним процесом фінансового ринку є сек'юритизацією (у такий спосіб переміщуються фінансові ринки з деякого виду на інший сегмент або перетворюють зобов'язання у ліквідні інструменти. В останні роки процес сек'юритизації характеризується переміщенням операцій з ринку позикових капіталів на ринок цінних паперів (передусім, ринок облігацій) і забезпечує зниження витрат у залученні кредитів. Це сприяє тісній взаємозалежності між сегментами фінансового ринку.

Обов'язковими особливостями фінансового ринку є чіткі правила, інфраструктура. Завдяки їм можна забезпечити мінімальні втрати мінімальні витрати та максимізує дохідність угод, які на ньому укладаються. Фінансовий ринок не може існувати без професійних учасників та успішно функціонує лише за умов ринкової економіки.

За видами активів фінансовий ринок можна поділити кредитний ринок (на ньому торгують кредитними ресурсами), ринок цінних паперів (об'єктом купівлі-продажу є фондові інструменти, які можуть бути випущені державою чи підприємствами), ринок дорогоцінних металів, валютний ринок ринок фінансових послуг. Фінансовий ринок пов'язаний насамперед із операціями, які на ньому здійснюються – фінансовими інструментами. Так називають різноманітні фінансові активи, вони являють собою законні вимоги власників на отримання доходу. Форма інструментів може бути різною. В основному, це готівкові гроші, банківські кредити, цінні папери. Інструменти обслуговування класифікують за видами ринків, періодом обертання, пріоритетною значимістю, рівнем гарантування дохідності та рівнем ризику.

В цілому, загальноприйнятий підхід класифікації суб'єктів фінансового ринку наступний: держава, суб'єкти господарювання та домогосподарства. Також учасників фінансового ринку можна поділити за формами укладення угод: продавці та покупці послуг, фінансові посередники та інститути інфраструктури фінансового ринку. Інфраструктурою фінансового ринку називають злагоджену систему установ, організацій та інституцій, які

обслуговують учасників з метою підвищення ефективності ними операцій і забезпечити процес безперервного функціонування фінансового ринку та вільний рух товарів і послуг на ньому.

Міра розвитку фінансового ринку тісно корелює з гнучкістю фінансової системи, реакцією її на зміни, збурення у політичному, законодавчому та економічному житті країни. Розвиток фінансових ринків є стимулятором зменшення суспільних витрат та зміцнення фінансових потоків.

Таким чином, прогнозування процесів у фінансах є критичним та дуже важливим.

Серед етапів прогнозування можна виділити такі:

Перший етап: аналітик виконує різносторонній огляд ситуації, робить припущення, встановлює змінні, консультиється з експертами предметної області.

Другий етап: Після проведення консультацій аналітик має обрати інструменти аналізу, зважити необхідність та вартість їх застосування. Задача приводиться до формалізованого вигляду.

Третій етап: на результатах попередніх двох етапів аналітик знаходить необхідний набір даних.

Четвертий етап: безпосередній аналіз даних, вибір більш точної моделі, порівняння із фактичними результатами.

## 1.2 Аналіз підходів до прогнозування

Методи прогнозування можна поділити на кількісні та якісні. Якісні моделі є вигідними у короткостроковій перспективі та залежать значними чином від думок експертів. Вони включають в себе опитування, дослідження

кон'юнктури ринку і мають суб'єктивний характер. Вони не покладаються на будь-які точні математичні обчислення чи моделі.

Кількісні методи не вимагають експертних оцінок.

Кількісні методи використовують статистичні дані на основі кількісної інформації.

Основна мета моделювання часових рядів - це зібрати та вивчити минулі спостереження для того щоб спроектувати модель, яка в подальшому буде використана для прогнозування.

Отже, прогнозування часових рядів – це процес обчислення майбутніх значень обраного ряду. Велика кількість галузей (економічна наука, фінанси, промислове виробництво, урядування) вимагають того, щоб модель, яку побудували була відповідним чином адаптована [1].

Серед найбільш актуальних моделей прогнозування можна виділити такі як регресійні, авторегресійні, експоненційне згладжування.

До регресійних моделей належать: лінійна регресія (linear regression), нелінійна регресія (nonlinear regression), множинна регресія (multiple regression) та поліноміальна регресія (polynomial regression).

Метод лінійної регресії полягає у побудові певної взаємозалежності між двома змінними. Одна із змінних є скалярною величиною, а інша вектором.

Нехай скалярна величина позначена  $y$ , а векторна  $x$ . Тоді модель лінійної регресії матиме вигляд:

$$y = a_0 + a_1x_1 + \dots + a_kx_k + b.$$

Найбільш поширеним методом для побудови лінії регресії є метод найменших квадратів. Цей метод дозволяє обчислити та змоделювати найкращу лінію регресії мінімізуючи суму квадратів вертикальних відхилень для кожної точки спостереження. Так як відхилення спочатку зводяться до

квадратів, а потім сумуються, то немає відмінностей між додатними та від’ємними значеннями.

Лінійна регресія може бути використана для прогнозування ефектів або змін, це полягає в тому, що вона демонструє силу зміни залежної зміни на основі змін незалежних змінних (рис. 1.4).

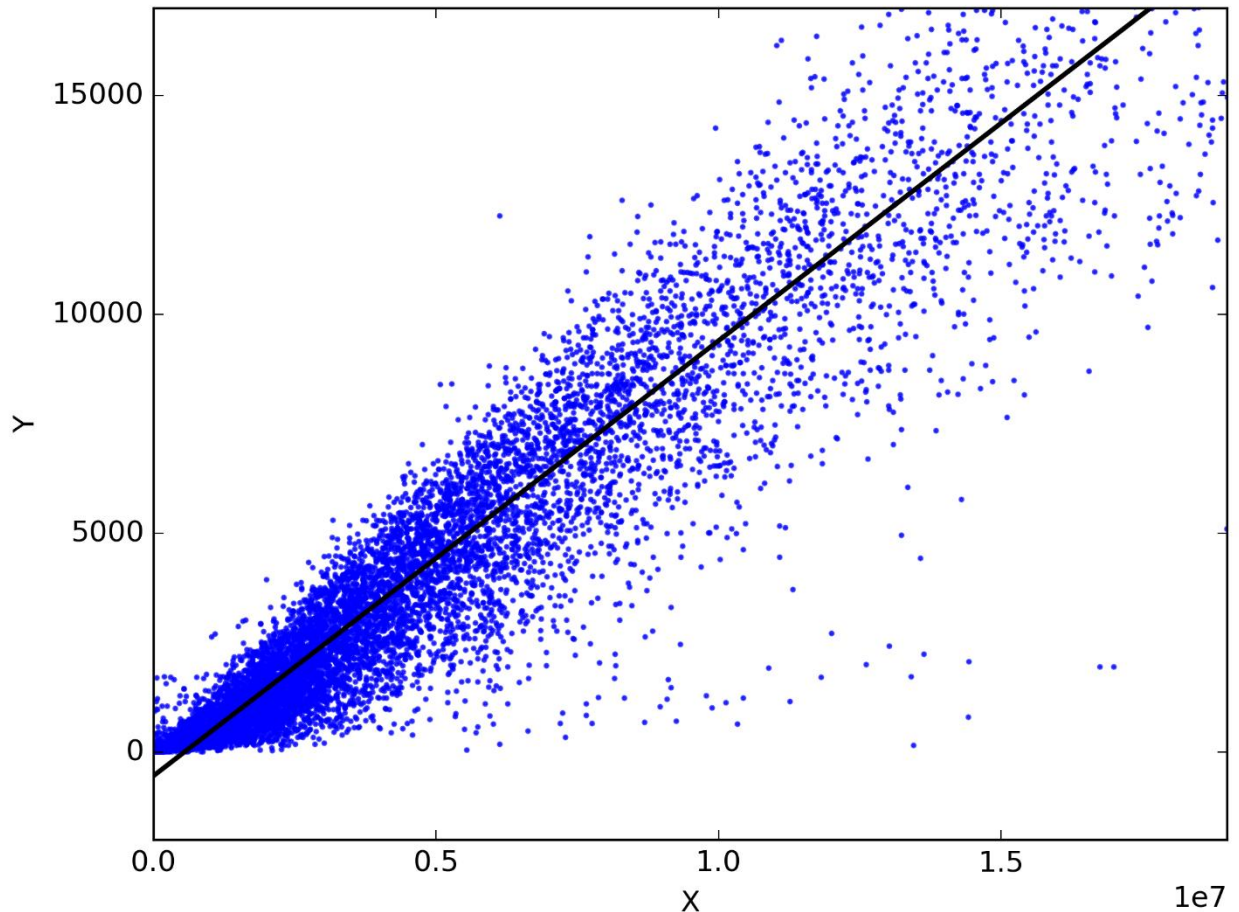


Рисунок 1.4 – Приклад даних спрогнозованих за допомогою лінійної регресії

У більш складних процесах лінійну регресію можна замінити на нелінійну. При застосуванні нелінійної регресії дані спостережень моделюються функцією з нелінійної композиції параметрів моделі, яка залежить від однієї або кількох незалежних змінних (рис. 1.5).

Вона задається таким чином:

$$f \sim y(x, \beta),$$

$$y(x, \beta) = \frac{a_1 x}{a_2 + x}$$

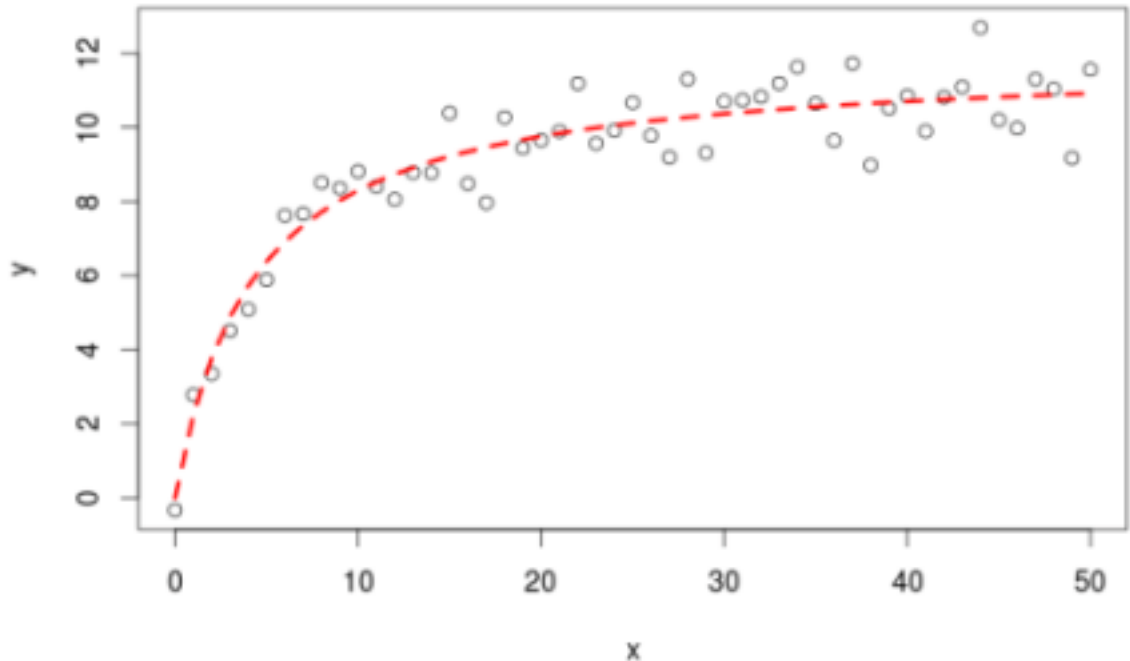


Рисунок 1.5 -Приклад даних, які моделюються нелінійною регресією

У якості нелінійних функцій можуть виступати:

1. Тригонометричні функції,
2. Логарифмічні функції,
3. Степеневі функції.

Іншим розширенням лінійної регресії можна вважати множинну регресію. Вона дозволяє прогнозувати стан системи на основі скінченної кількості незалежних змінних. Однак, потрібно враховувати, що вона у той же час вимагає конкретних специфікацій та факторів у моделі: потрібно включати лише змінні, що мають стосунок до процесу чи системи [8].

Множинною регресією можна пояснити взаємозв'язок між кількома незалежними і однією залежною змінною (критерієм). У цьому випадку

залежну змінну моделюють як функцію декількох незалежних змінних з відповідними коефіцієнтами разом з постійним членом.

Множинна регресія вимагає двох або більше незалежних змінних. Вона має наступну форму:

$$y = a_1x_1 + \dots + a_nx_n + c.$$

Іншим елементом, який вартий уваги є поліноміальна регресія. У ній моделювання зв'язку між незалежною змінною та залежною відбувається за допомогою поліному степені  $n$ .

Вона використовується для опису нелінійних процесів. Функція регресії у цьому випадку є лінійною за невідомими параметрами, які оцінюються з даних. Її вважають окремим випадком множинної регресії та записують так:

$$y = \alpha_0 + \alpha_1x + \varepsilon.$$

Окремим підкласом є авторегресійні моделі. Деякі з їхніх характеристик незмінні з плином часу. Ними описуються стаціонарні процеси. До таких моделей, зокрема GARCH AR та ARMAX [7].

AR (Autoregressive model) використовується для опису змін у економіці, фінансах, техніці які мають властивості, що змінюються з часом. Вихідна змінна у даній моделі лінійно залежна від попередніх значень

Символи AR(q) показують, що це авторегресійна модель порядку  $p$ . Модель AR(q) у загальному вигляді визначають наступним чином:

$$X_t = c + \sum_{i=1}^q \phi_i x_{t-i} + \varepsilon_t$$

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) досить часто використовується у фінансовій сфері, адже багато фінансових показників володіють властивістю умовної гетероскедастичності. Описати її можна так:

$$\sigma^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2.$$

ARIMA (Autoregressive integrated moving average) – застосовуються в деяких випадках, коли дані свідчать про нестационарність (процес може змінювати деякі свої характеристики з плином часу). Іншим поясненням цього явища може бути незмінність розподілу ймовірності при зміщеннях за часом. Загалом, стаціонарність лежить в основі аналізу та використанні великої кількості статистичних процедур.

Загальний вигляд моделі:

$$X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}.$$

Для прогнозування сезонних даних можна використати один із варіантів моделі ARIMA – SARIMA.

Досить часто у економічних процесах поточний економічний показник залежить не лише від його попередніх значень, а й від попередніх та поточних значень інших величин. Моделі, які описують такі процеси називають моделями з розподіленим лагом.

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{j=0}^r b_j x_{t-j} + \varepsilon_t$$

На відміну від ARIMA моделей рекурентні нейронні мережі з довгою короткостроковою пам'яттю можуть змінювати ще й лагові змінні, пристосовуючись до змін лагів залежно від часу.

### 1.3 Означення та використання трендів

Трендом називають деяке середнє значення вибірки, воно вказує на напрям розвитку тенденції у довгостроковій перспективі. Загалом трендом можна вважати деякий напрямок, у якому відбуваються якісь зміни чи наближення. Тренд відносять до історичних змін, за допомогою нього можна полегшити прогнозування.

Тренди поділяють на стохастичні та детерміновані, його характер можна визначити лише будуючи модель, яка ілюструє процес. Тренди можуть бути описані різноманітними функціями – степеневими, експоненційними, лінійними тощо. Значна кількість рядів мають яскраво виражений тренд.

Нехай  $y(k)$  – фактичне значення процесу в деякий момент  $k$ . Відносно короткі інтервали допускають таке визначення тренду:

$$\bar{y}(k) = \bar{y}(k - 1) + \frac{1}{k} [y(k) - \bar{y}(k - 1)],$$

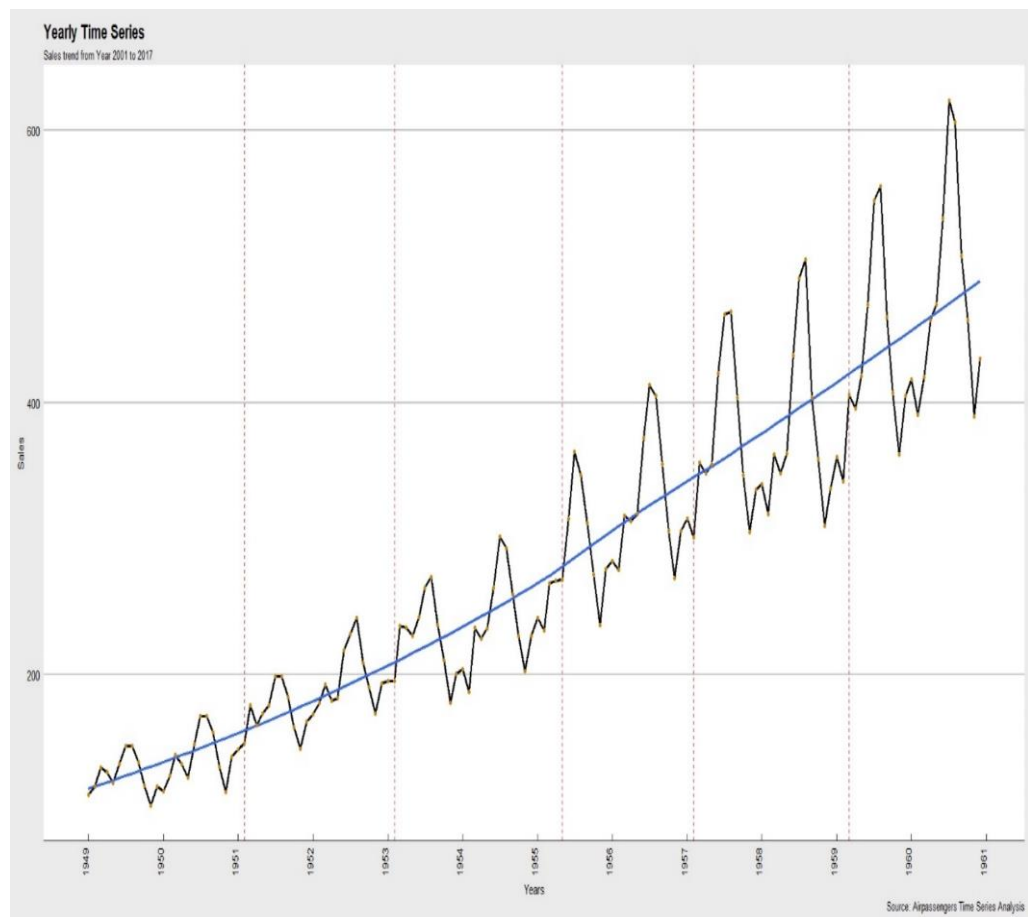
Тут величина  $y(k)$  є фактичним значенням процесу в момент  $k$ .

Процеси з трендом характерні для таких систем як перехідна економіка, так як більшість макроекономічних процесів в цій системі є нестационарними, а отже містять тренд [6].

В багатьох випадках тренд можна описати, але є випадки коли адекватно його описати детермінованою функцією неможливо. Для того щоб уникнути цю проблему використовують комбінацію декількох випадкових процесів в одній моделі. У більшості випадків не має можливості описати тренд детермінованою функцією.

Тренд буває двох видів: детермінований та стохастичний (рис. 1.6). Детермінований тренд можна описати за допомогою поліному довільного порядку від часу:

$$y(k) = a_0 + a_1k + a_2k^2 + \dots + a_pk^p + \varepsilon(k).$$



## Рисунок 1.6 - Приклад використання тренду

Опис стохастичного тренду можна виконати за допомогою такого рівняння:

$$y(k) = a_0 + y(k - 1) + \varepsilon(k),$$

Прогнозування будь-яким із вищенаведених методів може призвести прийняття неправильних та неоптимальних рішень за умови відносно великих похибок. Таким чином кожен метод варто модернізувати залежно від ситуації.

### 1.4 Експоненційне згладжування

Дуже часто при спостереженнях та аналізі процесів виникає ситуація, що неможливо виділити явні тенденції для тренду чи сезонної складової часового ряду. Використовуючи середнє значення усіх зібраних спостережень та певного згладжувального параметру, який лежить у межах від 0 до 1 записується відповідне рівняння:

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1 - \alpha)^j y_{T-j} + (1 - \alpha)^T k_0,$$

де  $k_0$  – початкове значення.

За допомогою параметру згладжування можна регулювати ваги спостережень ряду. Найбільшу складність у ході дослідження викликає саме

оцінка параметру, зазвичай його визначають на основі попередньо зібраних даних [5].

Параметр згладжування оцінюється шляхом мінімізації суми квадратів помилок (схожим чином до моделей регресій).

### 1.5 Програмне забезпечення, що дозволяє здійснювати прогнозування

Будувати моделі можна з допомогою таких програмних засобів як SAS, SPSS, MatLab, Eviews, Stata. У кожного з них є свої певні особливості. Як, наприклад, доступність вихідного коду в Matlab та детальна документація, але для того, щоб користуватись нею необхідно знати мову програмування. Eviews дозволяє будувати оцінки коефіцієнтів та графічні зображення і одночасно із цим володіє дуже зручним форматом експорту даних. Окрім цього він має досить дружній до користувача інтерфейс. Пакети SPSS та Statistica є досить схожими загалом, однак мають відмінності у вигляді процедур імпорту даних та вводу команд.

### 1.6 Висновки до розділу

У першому розділі була описана та доведена потреба прогнозування в процесах фінансового ринку. У даному розділі були описані основні моделі прогнозування на основі часових рядів. Наведені властивості та передумови застосування моделей. Були наведені необхідні відомості про фінансовий

ринок та використання трендів. Також був описаний процес застосування методу експоненційного згладжування.

## РОЗДІЛ 2 ПОБУДОВА МАТЕМАТИЧНИХ МОДЕЛЕЙ ТА ПІДХОДИ ДО ПРОГНОЗУВАННЯ

### 2.1 Основні поняття

#### 2.1.1 Означення часового ряду

Часовим рядом називається набір спостережень  $x_t$ , кожне з яких фіксується у специфічний час  $t$ . Вимірювання розташовуються послідовно у строгому хронологічному порядку.

Часові ряди можна поділити на дискретні та неперервні. Дискретні часові ряди є рядами спостережень у дискретні точки часу. В більшості випадків спостереження записуються через чітко визначений проміжок часу (щохвилини, щодня, щотижня).

Спостереження в дискретних часових рядах вважають виміром змінної з використанням множини дійсних чисел. Неперервні часові ряди трансформуються в дискретні шляхом об'єднання даних із заданого інтервалу.

#### 2.1.2 Складові частини часового ряду

Часові ряди мають 4 складові: тренд, сезонну та випадкову компоненти. Загальна тенденція динаміки змін до збільшення, зменшення чи сталості протягом досить тривалого для процесу періоду часу називається трендом. В цілому, тренд – це певний довгостроковий рух ряду.

Наприклад, тренд виконання бюджету за надходженнями кумулятивним чином протягом року є зростаючим, тоді як тренди процесів смертності можуть бути такими, що знижуються.

Часові ряди мають 4 складові: тренд, сезонну та випадкову компоненти. Загальна тенденція динаміки змін до збільшення, зменшення чи сталості протягом досить тривалого для процесу періоду часу називається трендом. В цілому, тренд – це певний довгостроковий рух ряду.

Наприклад, тренд виконання бюджету за надходженнями кумулятивним чином протягом року є зростаючим, тоді як тренди процесів смертності можуть бути такими, що знижуються [1].

### 2.1.3 Аналіз гетероскедастичності

Процес, дисперсія якого змінюється в часі називають гетероскедастичним. Вважатимемо процес гетероскедастичним, якщо його представити таким чином:

$$\text{var}[\varepsilon_k] = \sigma_{\varepsilon^2} \neq \text{const}$$

Гетероскедастичність веде за собою таке явище: для будь-якої незалежної змінної  $\varepsilon_k$ , variance її математичного сподівання залишатиметься сталим. Попереднє твердження дозволить зробити висновок, що функція дисперсії не залежатиме від часу [10].

Водночас із цим, гетероскедастичність буде значити, що дисперсія може збільшуватись чи зменшуватись на певному проміжку часу (рис. 2.2).

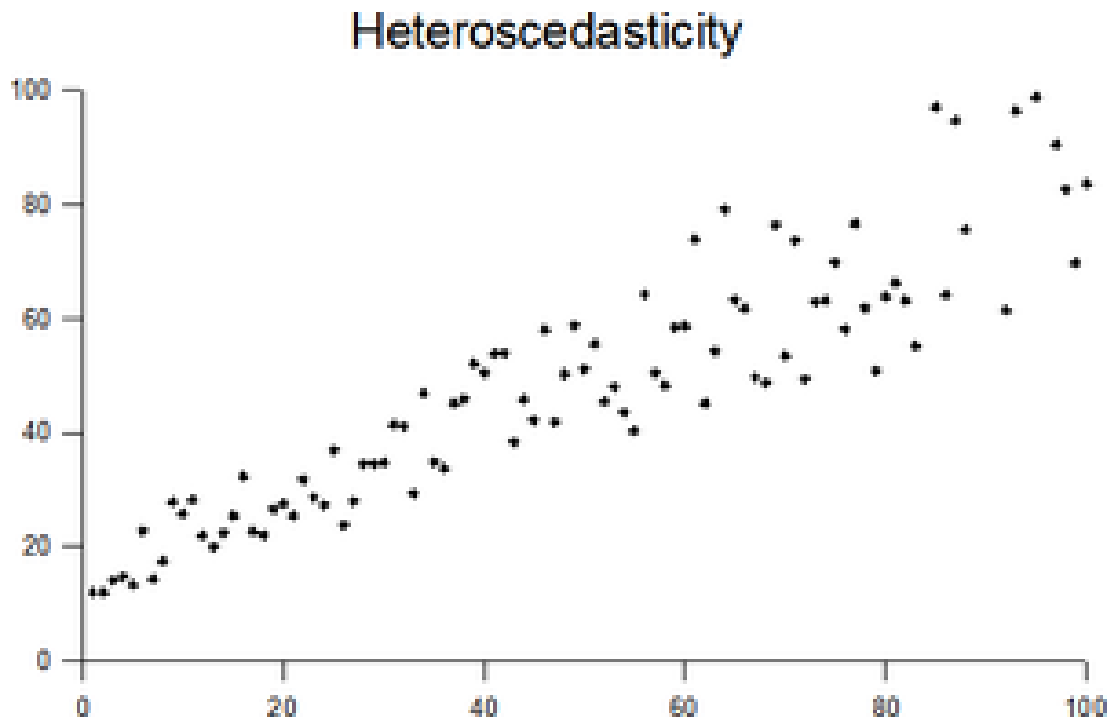


Рисунок 2.2 – Приклад гетероскедастичних даних

Явище гетероскедастичності можна представити таким чином:

$$\sigma_{\varepsilon k^2} = c^2 x^2 k^2$$

У цьому випадку  $c$  буде експериментально оціненою константою.

Для аналізу та перевірки гетероскедастичності існує велика кількість тестів: наприклад, тест Голдфельда-Квандта, Уайта, Бройша-Пагана.

Для тесту Бройша-Пагана альтернативна та основна гіпотези виглядатимуть наступним чином:

$$H_0: \sigma_1^2 = \dots = \sigma_n^2$$

$$H_1: H_0 \text{ є хибною}$$

Основна гіпотеза відповідає випадку, коли залишки є гомоскедастичними.

Статистика критерію може бути отримана на основі множників Лагранжа та має наступний вигляд:

$$LM = \left( \frac{\partial l}{\partial \theta} \right)' \left( -E \left[ \frac{\partial^2 l}{\partial \theta \partial \theta'} \right] \right)^{-1} \left( \frac{\partial l}{\partial \theta} \right)$$

Таким чином обчислення статистики можна звести до наступної послідовності дій. Оцінити вихідну модель методом найменших квадратів, обчислити залишки.

Після цього необхідно оцінити помилку моделі (в припущенні щодо її гомоскедастичності). Обчислити стандартизовані залишки і потім, побудувавши, додаткову регресію на квадратах стандартизованих залишків обчислити коефіцієнт детермінації [2].

При справедливості нульової гіпотези про гомоскедастичність залишків статистика критерію матиме розподіл хі-квадрат.

Застосування даного тесту вимагає знання змінних, які створюють гетероскедастичність. Найчастіше ці змінні можна взяти з вектора регресора. Також варто зазначити, що знання функціональної форми гетероскедастичності не є необхідним – це не завжди можна визначити.

Також для перевірки гомоскедастичності можна застосувати тест Вайта. Для цього обчислюються статистика:

$$F = \frac{R_{\hat{\varepsilon}^2}^2}{(1 - R_{\hat{\varepsilon}^2}^2)} (n - 2)$$

$$\chi^2 = nR_{\hat{\varepsilon}^2}^2$$

Незначимість статистики свідчатиме про гомоскедастичність, значущість, про гетероскедастичність.

Також для перевірки на гетероскедастичність можна використати тест Парка.

Щоб застосувати тест Парка застосувати необхідно спочатку побудувати рівняння регресії:

$$\hat{y}_i = \alpha_1 + \alpha_2 x_{i2} + \dots + \alpha_k x_{ik}$$

Потім обчислюємо точки ряду залишків:

$$e_i = y_i - \hat{y}_i$$

Після цього необхідно оцінити додаткову регресію:

$$\ln(e_i^2) = b_1 + b_2 * Z_i + \varepsilon_i.$$

Після перевірки значимості  $b_2$  (якщо коефіцієнт значимий) то можемо зробити висновок про наявність гетероскедастичних властивостей у моделі.,

#### 2.1.4 Аналіз нелінійності

Для того, щоб побудувати модель та визначити її структури необхідно провести детальний аналіз на нелінійність, перевірити чи наявна нелінійність у досліджуваному процесі.

Після цього можна перейти до визначення типу нелінійності.

Наприклад, можна використати таку статистику:

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

Якщо статистика  $F$  буде більшою за рівень значущості, то гіпотезу про лінійність треба відхилити. Даний метод не є завжди зручним, адже потрібно мати декілька вибірок даних [9].

Розв'язок даної задачі допускає велику кількість критеріїв, однак усі вони вимагають знання певних передумов та можливостей програмної реалізації.

Розглянемо приклад застосування лінійних коваріаційних функцій та покажемо ситуацію, коли означення структури не включає деяких змінних

Опишемо корельовані залишки:

$$\xi(k) = cu(k-1)e(k-1) + e(k),$$

$e(k)$  - білий шум;

$c$  - коефіцієнт.

$$E[e(k)] = 0; E[u(k)] = 0; E[e(k)u(k)] = 0.$$

$e(k)$  та  $u(k)$  – некорельовані, мають нульове середнє.

Нормована автокореляційна функція залишків із нормованою функцією взаємної кореляції сигналу і залишків має наступний вигляд:

$$\Phi_{\xi\xi}(\tau) = \delta(\tau),$$

$$\Phi_{u\xi}(\tau) = 0, \forall \tau.$$

Тоді можемо зробити висновок що

$$\xi(k) = cu(k - 1)e(k - 1) + e(k)$$

Впливає з рівняння мажмо що  $\xi(k)$  - корельована послідовність. Вона буде давати зсуви в оцінки параметрів моделі.

У загальному випадку, лінійні кореляційні методи не дають можливості ідентифікації присутності нелінійностей та ступеню їх впливу на цей процес.

Визначимо, лінійний чи нелінійний зв'язок між виходом та входом.. Скористаємось наступною формулою:

$$X_{ij} = \frac{|S_w(w_i, w_j)|^2}{S_w(w_i)S_w(w_j)S_w(w_i / w_j)}$$

де  $S_w(w_i, w_j)$  - біспектральна щільність потужності;

$S_w(w_i)$  - спектральна щільність потужності часового ряду.

Якщо  $S_w(w_i, w_j) = 0, \forall w_i, w_j$  процес лінійний і третій момент вхідного сигналу  $\mu_3 = 0$ .

Даний підхід має такі недоліки: оцінка спектральної щільності потужності потребує застосування обробки сигналу у вигляді використання часових вікон, усереднення та цифрової фільтрації, також даний метод даний підхід не завжди може бути використаним для деяких типів задач, наприклад, ідентифікація систем.

Також, у випадку його застосування ми не матимемо можливості отримати параметри у явному вигляді.

Інколи можна використати дисперсійний метод. На основі якого визначають чи присутня нелінійність та застосовують наступну функції, яку ми можемо використовувати далі:

$$\psi_{zu}(t_1, t_2) = E_{u(t_2)}[E_{z(t_1)}[z(t_1)|u(t_2)] - E_{z(t_1)}[z(t_1)]]^2.$$

### 2.1.5 Перевірка стаціонарності

У стаціонарних процесів статистичні властивості, такі як середнє, дисперсія і коваріація стаціонарного процесу, не залежать від часу. Можна визначити два типи стаціонарних процесів – зі слабкою або сильною стаціонарністю [6].

Деякий процес:

$$X(t), t=0,1,2, \dots n$$

Називають процесом з сильною стаціонарністю, якщо розподіл

$$\{X_{t-s}, X_{t-s+1}, \dots, X_t, \dots, X_{t+s-1}, X_{t+s}\}$$

не залежить від  $t$  для всіх  $s$ . Розподіл будь-якого набору випадкових величин з процесу не залежить від часу (рис. 2.3).

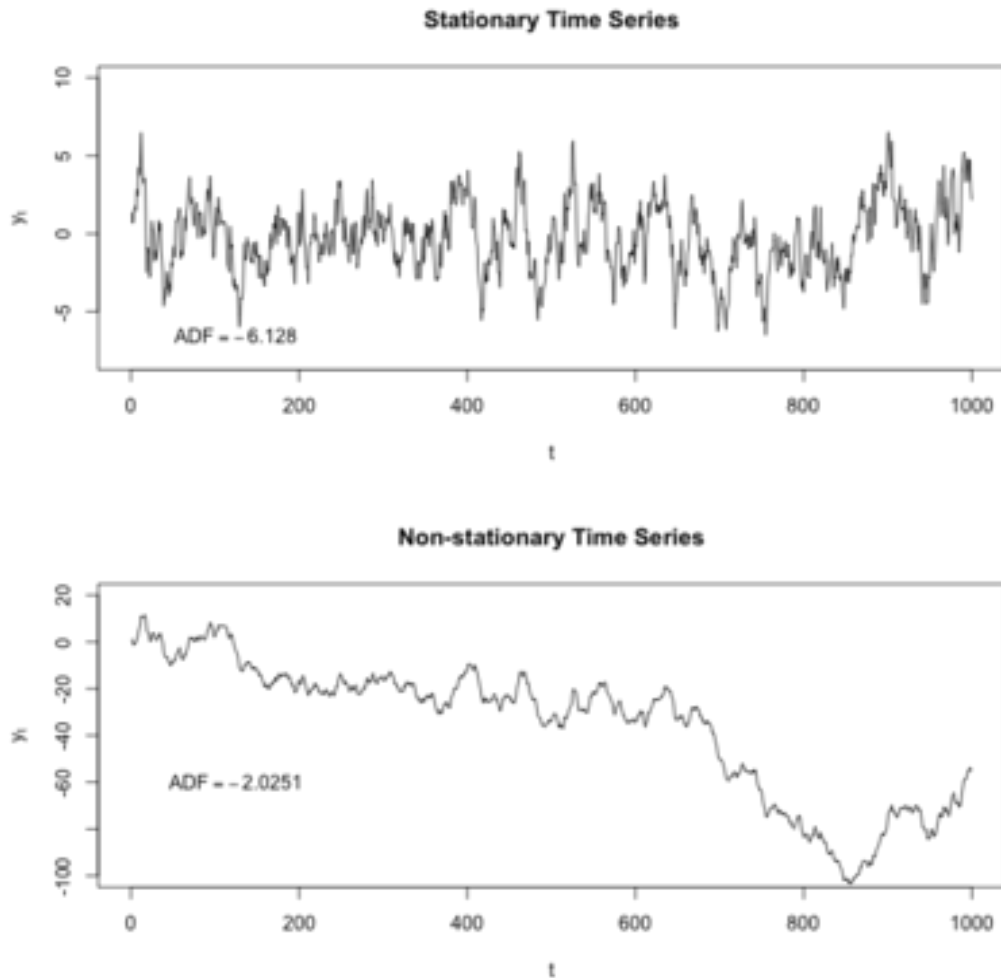


Рисунок 2.3 – Приклад стаціонарних і нестаціонарних процесів

Процес називається слабо стаціонарним порядку  $k$  якщо випадкові моменти процесу до цього порядку залежать тільки від часових різниць.

Існує велика кількість тестів для визначення нестаціонарності процесу, розглянемо найпростіший із них – тест Діккі Фуллера [4].

Він полягає в тому, що необхідно визначити чи є одиничний корінь у трьох рівняннях:

$$\Delta y(k) = \gamma y(k-1) + \varepsilon(k),$$

$$\Delta y(k) = \alpha_0 + \gamma y(k-1) + \varepsilon(k),$$

$$\Delta y(k) = \alpha_0 + \gamma y(k-1) + \alpha_2 k + \varepsilon(k),$$

Відмінність між рівняннями полягає в наявності детермінованих членів. Необхідно визначити параметр  $\gamma$ , одиничний корінь буде за умови, що параметр рівний 0.

Тест Діка-Фуллера оцінює одне або декілька з рівнянь, які наведені вище. Оцінка здійснюється за допомогою МНК. Мета - отримати оцінку параметру  $\gamma$  та її похибки [1].

Далі обчислюється  $t$ -статистика, яку ми порівнюємо з значеннями, які наведені вже у таблиці Діка-Фуллера. Після порівняння потрібно прийняти рішення стосовно відхилення нульової гіпотези, яка означає  $\gamma = 0$ .

#### 2.1.6 Означення ковзного середнього

Ковзним середнім називають метод для згладжування рядів, що містять дані про ціни. Такі ряди відображають зміни біржових котирувань (зміну індексів), цін на сировину, нафту або інші товари [3].

На практиці застосовують прості, зважені та експоненціальні ковзні середні. Наприклад, просте ковзне середнє можна означити як середню ціну на момент закриття протягом останніх 10 днів. Термін ковзного середнього полягає в тому, усереднювані значення невинно рухаються протягом певного проміжку часу.

Саме за допомогою ковзного середнього можна згладити несуттєві коливання на ринку. Якщо розглянути ринки, де яскраво виражені цінові тенденції відсутні, то ковзні середні матимуть властивість зміни у деякому горизонтальному діапазоні.

Отже, формулу для простого ковзного середнього можна визначити так:

$$MA(k) = \frac{y(k-n) + y(k-n-1) + \dots + y(k)}{k}$$

Властивості ковзного середнього зумовлені тим, що у інформації з'являються даги, так як ковзне середнє за означенням дорівнює середньому значенню минулих цін, то зміни напрямків ковзних середніх будуть із затримкою реагувати на відповідні зміни у вихідних цінових рядах (рис. 2.4).



Рисунок 2.4 – Приклад динаміки ковзного середнього

Саме через незадовільну роботу на ринках, де відсутній яскраво виражений тренд застосовують інші, більш чутливі види ковзних середніх. Наприклад, зважене ковзне середнє використовує ваги для кожного ряду вхідних даних, що обробляються.

Вагові коефіцієнти експоненційного ковзного середнього розподілені за експоненційним законом, а ваги при його використанні зменшуються з віддаленістю від поточного періоду.

### 2.1.7 Кореляційна матриця

Коефіцієнт кореляції дозволяє з'ясувати та встановити факт взаємозв'язку між двома змінними. Цей вид зв'язку може бути нелінійним або лінійним залежно від його природи. Найпростішим є застосування лінійної кореляції, однак для точного та глибокого розуміння взаємозв'язків необхідно будувати нелінійні залежності. Для її знаходження необхідно будувати адекватну модель процесу.

Кореляційна матриця дозволяє встановити зв'язок залежної (ендогенної) змінної з незалежними (екзогенними) [7]. Розглянемо кореляційну матрицю для 3 змінних. Для того щоб обчислити її елементи необхідно спочатку зафіксувати вибірки для кожної змінної:

$$R = \begin{bmatrix} r_{yy} & \cdots & r_{zy} \\ \vdots & \ddots & \vdots \\ r_{yz} & \cdots & r_{yy} \end{bmatrix}$$

Очевидно, що кожен з коефіцієнтів матриці є по суті коефіцієнтом кореляції між двома вибірками. У загальному випадку рекомендується спочатку проаналізувати наявність логічного зв'язку між змінними, системно проаналізувати процес, завдяки таким діям можна буде в подальшому розглядати лише ті змінні, які дійсно впливають на залежну. На основі значень коефіцієнтів кореляції дослідник має прийняти рішення про включення в рівняння регресії незалежних змінних (регресорів).

### 2.2 Оцінка адекватності моделі

Критерії адекватності моделі та якості прогнозу необхідні для порівняння найбільш точних моделей та методів прогнозування [2].

Коефіцієнт детермінації  $R^2$  - це статистичний показник, що використовується як індикатор залежності дисперсії залежної змінної від дисперсій незалежних величин.

Його можна порахувати таким чином:

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

Очевидно, що для дійсно якісної моделі коефіцієнт детермінації повинен прямувати до одиниці.

Також для оцінки відмінностей між змінними можна використати статистику Дарбіна-Уотсона. Ця статистика використовується для перевірки автокореляції залишків при регресійному аналізі.

$$DW = 2 - 2p,$$

$$p = \frac{\sum_{k=2}^N e(k) * e(k-1)}{\sum_{k=1}^N [e(k)]^2},$$

$p$  - кореляція між сусідніми значеннями випадкової змінної  $\varepsilon(k) = e(k)$ .

Коефіцієнт Тейла показує міру точності та конзистентності моделі. Він обчислюється таким чином:

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i)^2 + \frac{1}{N} \sum_{i=1}^N (\hat{y}_i)^2}}$$

Його значення знаходиться на інтервалі від 0 до 1. Якщо коефіцієнт Тейла дорівнює одиниці, то модель абсолютно непридатна для прогнозу.

$U^M$  – коефіцієнт упередженості, Обчислюється за формулою:

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

Він показує чи є якісь повторювані відмінності чи відхилення для середніх значень рядів. Якщо він близький до 0, то можна вважати модель якісною

Відношення варіації:

$$U^S = \frac{(\sigma_{actual} - \sigma_{fitted})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

За допомогою цього критерію визначаємо чи модель зміщує дисперсію часового ряду.

$$U^C = \frac{2(1-p)(\sigma_{actual} - \sigma_{fitted})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

Відношення коваріації показує наскільки корелюють реальні та спрогнозовані ряди.

Ці статистики пов'язані такою рівністю:

$$U^C + U^S + U^M = 1.$$

Середня помилка абсолютного відсотка (MAPE) - це середнє значення абсолютних процентних помилок прогнозів. В цілому це різниця між прогнозованим та фактичним значенням.

Для обчислення MAPE можна використати таку формулу:

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

RMSE - це стандартне відхилення залишків (помилки прогнозування).

Середньоквадратична помилка кореня використовується в регресійному аналізі для перевірки результатів експерименту.

Вона обчислюється за формулою:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [\hat{y}(i) - y(i)]^2}.$$

## 2.4 Висновки до розділу

У даному розділі було розглянуто основні теоретичні відомості про часові ряди та їх властивості. Були наведені необхідні теоретичні відомості для оцінки якості моделей та побудови регресійних моделей. Наведені методи для перевірки стаціонарності та гетероскедастичності моделей. Також наведені приклади процедури для перевірки моделі на нелінійність.

## РОЗДІЛ 3 ПОБУДОВА МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ КУРСУ ВАЛЮТ

### 3.1 Опис даних

У цьому розділі будуть побудовані моделі та прогнози на основі цих моделей. Початковий етап роботи включав пошук даних, дані були взяті з сайту Національного банку України. Дані включають у себе курс гривні до іноземних валют за кожен день із початку 2020 року (1 січня до 26 травня 2020 року). Таким чином до вибірки були включені 147 значень. У подальшому в роботі буде розглянуто два варіанти – курс гривні до долару США, а також курс гривні до євро. У роботі був використаний власний програмний продукт на мові програмування Python і проведене порівняння з системою Eviews.

### 3.2 Побудова та прогноз моделі для ряду курсу валют

Для початку візуалізуємо ряд даних внутрішніми засобами системи Eviews.

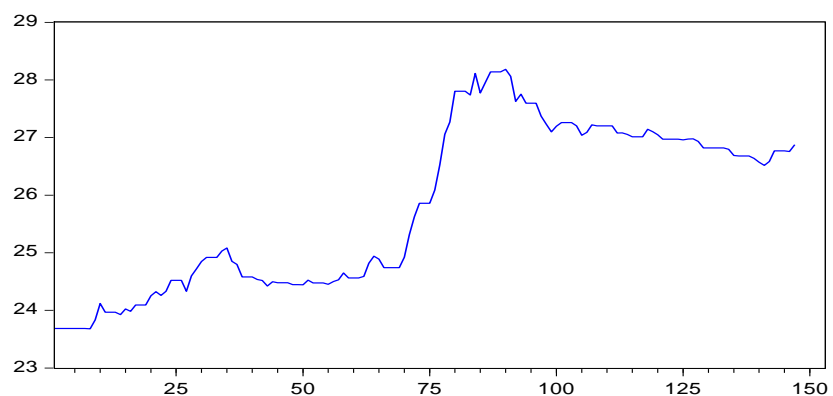


Рисунок 3.1 – Графік динаміки зміни курсу гривні відносно долару  
США у 2020 році

Побудуємо АКФ та ЧАКФ для того щоб визначити порядок авторегресії, після цього оцінимо коефіцієнти моделі.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.985	0.985	121.22	0.000	
2	0.967	-0.096	239.03	0.000	
3	0.946	-0.100	352.72	0.000	
4	0.923	-0.052	461.95	0.000	
5	0.899	-0.053	566.38	0.000	
6	0.874	-0.008	666.00	0.000	
7	0.847	-0.088	760.33	0.000	
8	0.816	-0.114	848.73	0.000	
9	0.786	0.017	931.38	0.000	
10	0.755	0.000	1008.4	0.000	

Рисунок 3.2 – АКФ та ЧАКФ для першого ряду

На основі корелограми можемо зробити висновок, що треба будувати авторегресію порядку 1.

$$AR(1): y(k) = 25.38510 + 0.995771 * y(k - 1)$$

R-squared	0.990162	Mean dependent var	25.60239
Adjusted R-squared	0.989997	S.D. dependent var	1.481960
S.E. of regression	0.148218	Akaike info criterion	-0.916841
Sum squared resid	2.614273	Schwarz criterion	-0.847890
Log likelihood	58.92730	Hannan-Quinn criter.	-0.888835
F-statistic	5988.674	Durbin-Watson stat	1.444104
Prob(F-statistic)	0.000000		

Рисунок 3.3– Характеристики моделі AR(1)

Тепер побудуємо АРКС моделі на основі АР(1), для цього на початку визначимо порядок ковзного середнього (КС), щоб це зробити побудуємо АКФ та ЧАКФ для залишків, які утворились при побудові АР(5).

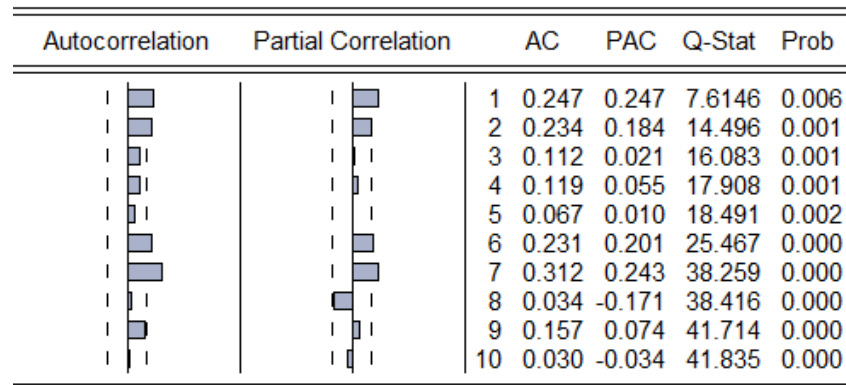


Рисунок 3.4 – ЧАКФ та АКФ для залишків моделі АР(1)

На основі корелограми робимо висновок, що необхідно побудувати рівняння АРКС (1,1):

$$\text{АРКС}(1,1): y(k) = 25.40154 + 0.994199 * y(k - 1) + 0.195504 * ma(k - 1)$$

Параметри моделі зображені на рисунку нижче:

R-squared	0.990689	Mean dependent var	25.60239
Adjusted R-squared	0.990452	S.D. dependent var	1.481960
S.E. of regression	0.144809	Akaike info criterion	-0.954775
Sum squared resid	2.474431	Schwarz criterion	-0.862840
Log likelihood	62.24130	Hannan-Quinn criter.	-0.917434
F-statistic	4184.859	Durbin-Watson stat	1.897183
Prob(F-statistic)	0.000000		

Рисунок 3.5 – Характеристики моделі АРКС(1,1)

Таблиця 3.1 – Порівняння якостей моделей

Реалізація	Модель	$R^2$	DW	AIC
Eviews	AR(1)	0,990162	1,444104	-0,916841
Eviews	АРКС(1,1)	0,990689	1,897183	-0,954775
Власна програма	AR(1)	0,9907928	1,444014	-0.9076725
Власна програма	АРКС(1,1)	0,99132015	1,897065	-0.9788898

Отже, можемо побачити, що результати моделювання, які ми отримали за допомогою власної реалізації на мові програмування Python та у системі Eviews практично не відрізняються, відмінності у досить малих порядках чисел зумовлені головним чином округленнями, адже даний процес реалізований різним чином не тільки у різних системах, а й у різних мовах програмування.

Побудуємо тепер статичний прогноз на 7 кроків уперед (на основі АРКС (1,1), тобто з 141 по 147 зафіксоване значення і порівняємо з реальними значеннями.

Таблиця 3.2 – Порівняння прогнозованих і реальних значень на 7 кроків уперед

Номер значення	Реальні значення	Прогнозні значення
141	26.756782	26.769937
142	26.746210	26.774232
143	26.501244	26.485431
144	26.797304	26.789876
145	26.756782	26.769900

146	26.764899	26.755600
147	26.746210	26.764215

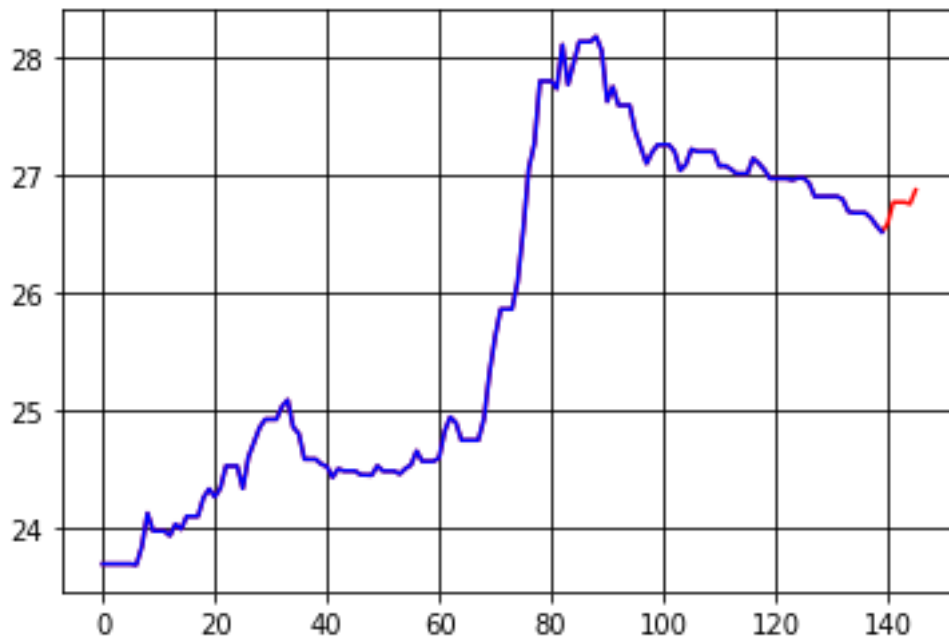


Рисунок 3.6 – Графік даних спрогнозованих з АРКС (1,1)

Таблиця 3.3 – Якість прогнозу на 7 кроків уперед

Реалізація	Модель	RMSE	MAPE	Коеф. Тейла
Eviews	АРКС(1,1)	0.00026375	0.0149771	0,0003
Власна програма	АРКС(1,1)	0.000263749	0.014977143	0.000303817

Побудуємо тепер інший ряд – для курсу гривні до євро, візуалізуємо його аналогічно до попереднього програмними засобами Eviews. Після цього

проведемо аналогічні процедури – побудуємо ЧАКФ та АКФ, визначимо порядок  $AR$ , а після цього побудуємо  $AR$ -модель.

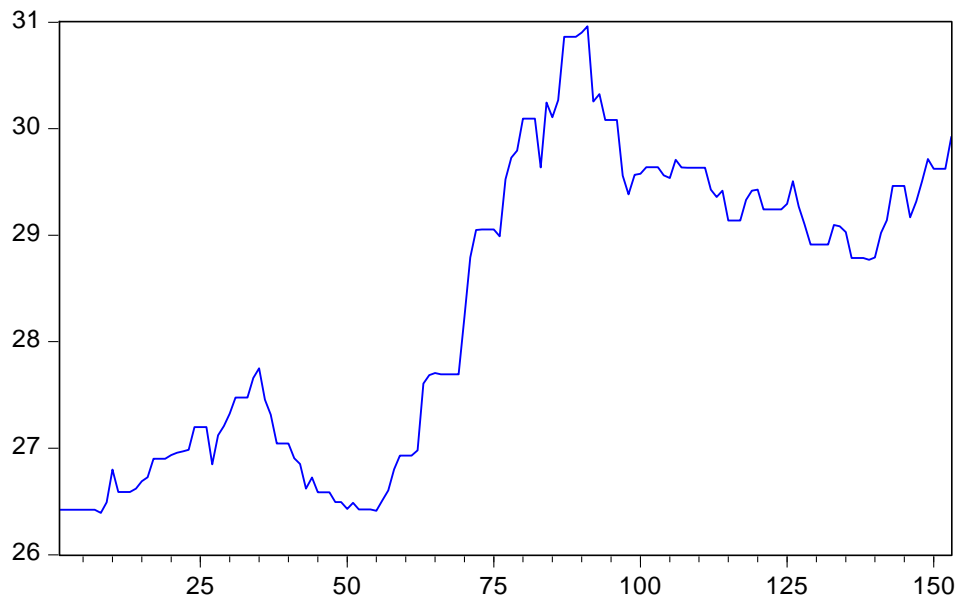


Рисунок 3.7 – Графік динаміки зміни курсу гривні відносно євро США у 2020 році

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
██████████	██████████	1	0.980	0.980	149.98	0.000
██████████		2	0.961	-0.007	295.03	0.000
██████████		3	0.940	-0.053	434.70	0.000
██████████		4	0.919	-0.019	568.98	0.000
██████████		5	0.897	-0.022	697.80	0.000
██████████		6	0.875	0.004	821.41	0.000
██████████		7	0.853	-0.030	939.67	0.000
██████████		8	0.827	-0.119	1051.6	0.000
██████████		9	0.800	-0.047	1156.9	0.000
██████████		10	0.771	-0.038	1255.5	0.000

Рисунок 3.8 – АКФ та ЧАКФ для 2 ряду

Отже, на основі корелограми можемо зробити висновок, що треба будувати авторегресію порядку 1.

$$AP(1): y(k) = 28.23155 + 0.993011 * y(k - 1)$$

R-squared	0.981973	Mean dependent var	28.33703
Adjusted R-squared	0.981733	S.D. dependent var	1.388486
S.E. of regression	0.187663	Akaike info criterion	-0.460989
Sum squared resid	5.282628	Schwarz criterion	-0.401568
Log likelihood	38.26564	Hannan-Quinn criter.	-0.436851
F-statistic	4085.428	Durbin-Watson stat	1.801327
Prob(F-statistic)	0.000000		

Рисунок 3.9 – Характеристики моделі AP(1) для другого ряду

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.073	0.073	0.8273	0.363
		2	0.082	0.077	1.8748	0.392
		3	0.004	-0.007	1.8779	0.598
		4	-0.014	-0.020	1.9077	0.753
		5	0.018	0.020	1.9581	0.855
		6	0.088	0.089	3.2020	0.783
		7	0.250	0.240	13.372	0.064
		8	-0.000	-0.045	13.372	0.100
		9	0.153	0.127	17.241	0.045
		10	0.023	0.015	17.332	0.067

Рисунок 3.10– АКФ та ЧАКФ для залишків 2 ряду

Отже, потрібно будувати авторегресію з ковзним середнім порядку 7.

$$\text{АРКС}(1,7): \quad y(k) = 0.225940 + 0.990258 * y(k - 1) - 0.225940 * ma(k - 7)$$

R-squared	0.983070	Mean dependent var	28.33703
Adjusted R-squared	0.982729	S.D. dependent var	1.388486
S.E. of regression	0.182472	Akaike info criterion	-0.507972
Sum squared resid	4.961112	Schwarz criterion	-0.428745
Log likelihood	42.85986	Hannan-Quinn criter.	-0.475789
F-statistic	2884.014	Durbin-Watson stat	1.814681
Prob(F-statistic)	0.000000		

Рисунок 3.11 – Характеристики моделі АРКС (1,7)

Таблиця 3.4 – Порівняння якостей моделей

Реалізація	Модель	$R^2$	DW	AIC
Eviews	АР(1)	0,981973	1,801327	-0,460989
Eviews	АРКС(1,7)	0,983070	1,814681	-0,507972
Власна програма	АР(1)	0,9825986	1,8024746	-0,46754
Власна програма	АРКС(1,7)	0,9836963	1,815837101	-0,49973

Таблиця 3.5 – Порівняння прогнозованих і реальних значень на 7 кроків уперед для другого ряду

Номер значення	Реальні значення	Прогнозні значення
141	29,3157	29.133761
142	29,5084	29.315585
143	29,7151	29.509440
144	29,6234	29.717524
145	29,6234	29.602668
146	29,6234	29.611607

147	29,9275	29.611051
-----	---------	-----------

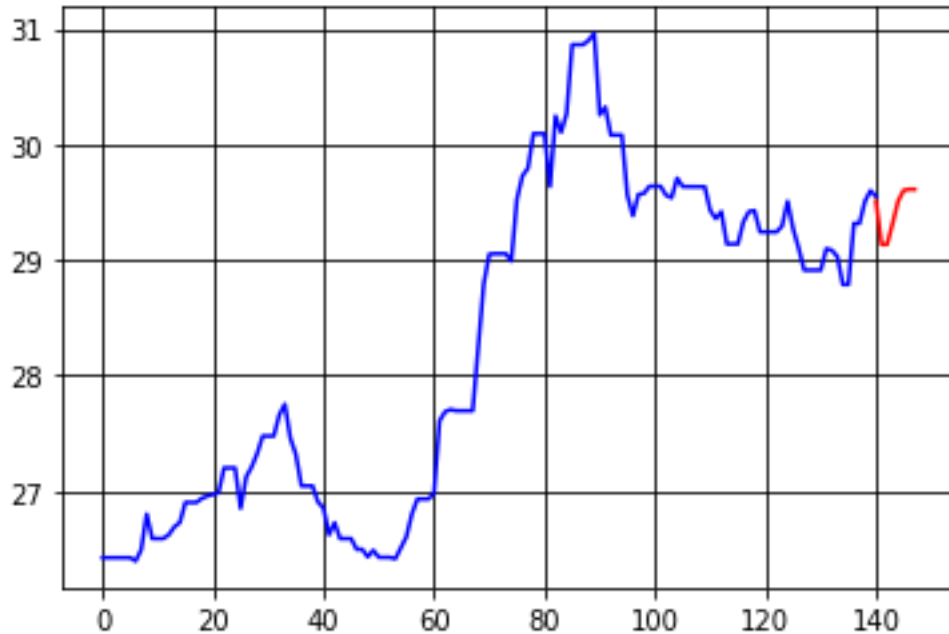


Рисунок 3.12 – Графік даних спрогнозованих з АРКС (1,7)

Таблиця 3.6 – Порівняння якостей моделей

Реалізація	Модель	RMSE	MAPE	Коеф. Тейла
Eviews	АРКС(1,7)	0.03174	0.14622	0.047309537
Власна програма	АРКС(1,7)	0.03172	0.14619	0.047308987

Отже, таким чином ми знову бачимо, що результати роботи власної програми не відрізняються від результатів пакету Eviews, адже значення критерії є практично рівними, а відмінності зумовлені лише похибками при округленні.

### 3.3 Висновки до розділу

У цьому розділі проводилося, було проведене моделювання процесів фінансового ринку, а конкретно - курсу валют, а також оцінка прогнозів за допомогою описаних в попередньому розділі методів. Дані були взяті з сайту Національного банку України. Був розроблений власний програмний продукт і проведене порівняння результатів моделювання зі статистичним пакетом Eviews.

## РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

### 4.1 Постановка задачі техніко-економічного аналізу

У даній роботі був застосований метод функціонально-вартісного аналізу для проведення техніко-економічного аналізу власноруч розробленого програмного продукту для використання моделей та методів прогнозування часових рядів у галузі фінансового ринку.

Технічні вимоги до програмного продукту включають наступне:

1. Візуалізація даних, які використовуються для аналізу
2. Вибір параметрів моделі
3. Обчислювальні процедури та виведення характеристик моделей
4. Завантаження даних із файлу

Можливість запису результатів виконання програми до текстового файлу.

Головною функцією  $F_0$  будемо вважати розробку програмного продукту для моделювання та прогнозування процесів фінансового ринку за допомогою часових рядів з можливістю вибрати модель та її параметри.

$F_1$  – мова програмування:

- a) C++
- b) Python
- c) R

$F_2$  – спосіб введення вихідних даних:

- a) Зчитування з файлу
- b) Ініціалізація даних в кодї програми
- c) Введення даних вручну за допомогою інтерфейсу програми

$F_3$  – інтерфейс користувача:

- a) Веб-інтерфейс
- b) Консольний інтерфейс
- c) Віконний інтерфейс

#### 4.2 Варіанти реалізації основних функцій

Морфологічна карта реалізації основних функцій наведена нижче:

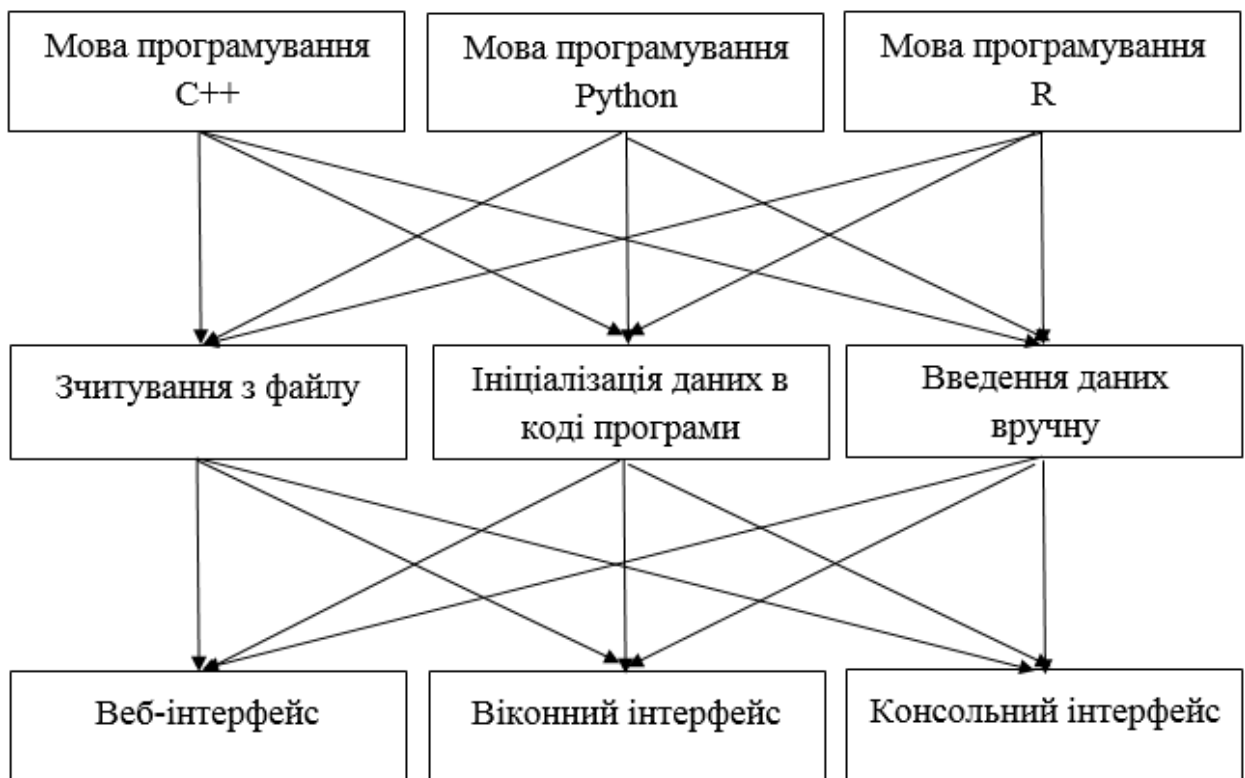


Рисунок 4.1 – Морфологічна карта

Таблиця 4.1 Позитивно-негативна матриця

Основні функції	Варіанти реалізації	Переваги	Недоліки
<i>F1</i>	A	Висока швидкодія, крос платформенність	Складний синтаксис
	B	Простота реалізація, простота семантичних конструкцій	Низька швидкодія
	C	Простота реалізації	Низька швидкість розробки
<i>F2</i>	A	Простота реалізації	Відсутність універсального підходу для усіх типів файлів
	B	Легкість редагування окремих показників	Відсутність ергономічності
	C	Відсутність необхідності застосування додаткових функцій чи бібліотек	Відсутність ергономічності

<i>F3</i>	А	Простота у реалізації	Необхідне інтернет з'єднання
	В	Зручність для користувача	Велика к-ть часу на реалізацію
	С	Відсутність необхідності застосування додаткових функцій чи бібліотек	Низька ергономічність, відсутність візуалізації

Таким чином, буде доцільним розглянути такі варіанти програмного продукту:

1.  $F1b - F2a - F3a$
2.  $F1b - F2a - F3b$

#### 4.3 Обґрунтування системи параметрів програмного продукту

Для оцінки функцій оберемо наступну систему параметрів:

Назва параметра	Умовні позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія мови програмування	X1	Нс/оп	300	200	50
Кількість рядків коду	X2	штук	1000	700	500

Об'єм пам'яті для коректної роботи	X3	M6	256	128	6
------------------------------------	----	----	-----	-----	---

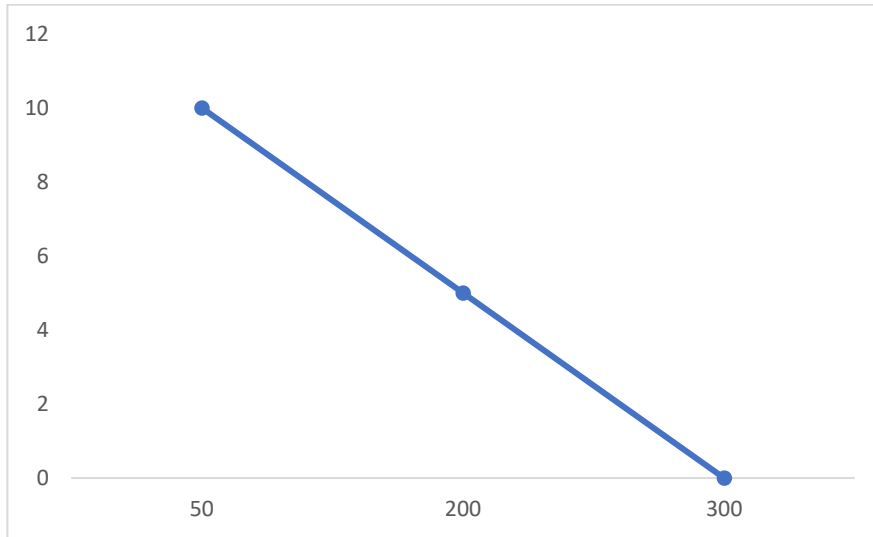


Рис. 4.2 Значення параметру X1

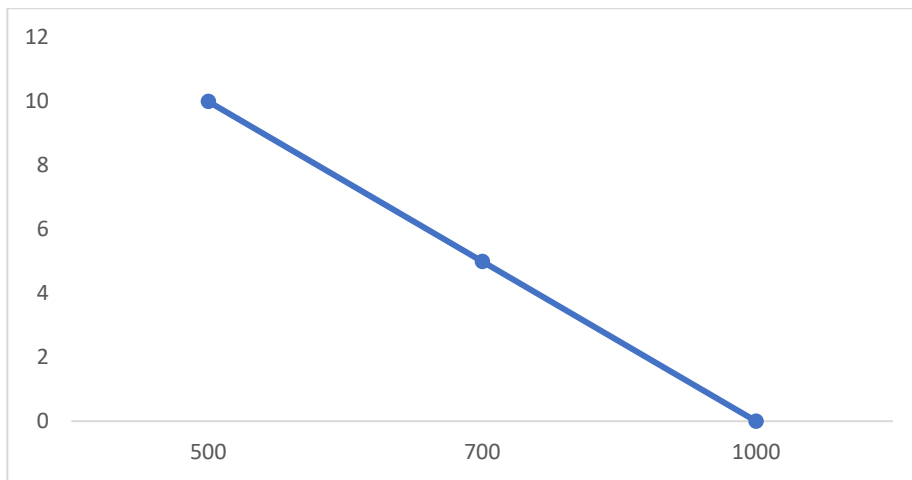


Рис. 4.3 Значення параметру X2

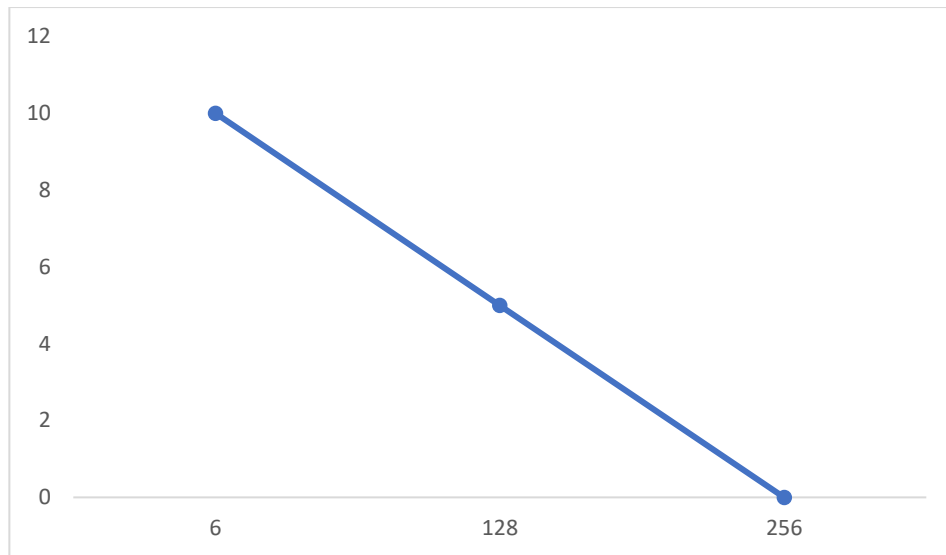


Рис. 4.4 Значення параметру X3

Параметрів усього 3 саме тому ранги будуть відрізнятись від 1 до 3.

Табл. 4.3 Результат оцінки параметрів

Параметр	Ранг параметру по оцінці експерта							Сума рангів, $R_i$	Відхилення $\Delta_i$	Квадрат відхилення, $(\Delta_i)^2$
	1	2	3	4	5	6	7			
X1	1	1	1	1	1	1	1	7	-7	49
X2	3	3	3	2	2	2	2	17	3	9
X3	2	2	2	3	3	3	3	18	4	16
Разом	6	6	6	6	6	6	6	42	0	74

Табл. 4.4 Попарне зрівняння параметрів

Параметр и	Експерти							Кінцева оцінка	Числове значення
	1	2	3	4	5	6	7		

X1 та X2	<	<	<	<	<	<	<	<	0,5
X1 та X3	<	<	<	<	<	<	<	<	0,5
X2 та X3	>	>	>	<	<	<	<	<	0,5

$$W = \frac{12 \cdot S}{N^2(n^3 - n)} = \frac{12 \cdot 74}{49 \cdot (27 - 3)} 0.75 > 0.67.$$

Коефіцієнт узгодженості більше 2/3, як і має бути.

Таблиця 4.5 - Розрахунок вагомості параметрів

Параметри x <sub>i</sub>	Параметри x <sub>j</sub>			Перша ітерація		Друга ітерація		Третя ітерація	
	1	2	3	b <sub>i</sub>	K <sub>bi</sub>	b <sub>i</sub> <sup>1</sup>	K <sub>bi</sub>	b <sub>i</sub> <sup>2</sup>	K <sub>bi</sub>
X1	1	0,5	0,5	2	0,22	0,44	0,137931	0,89	0,08
X2	1,5	1	0,5	3	0,33	1	0,310345	0,31	0,27
X3	1,5	1,5	1	4	0,44	1,78	0,551724	0,55	0,64
Всього:	4	3	2	9	1	3,222222	1	1,75	1

Таблиця 4.6 – Рівень якості альтернатив основних функцій

Основна функція	Варіант реалізації	Параметри	Абсолютне значення параметра	Бальна оцінка параметра	Коефіцієнт вагомості параметра	Коефіцієнт якості
F1	Б	X1	250	2,6	0,08	0,208
		X2	750	4,3	0,27	1,161
		X3	64	7,8	0,64	4,992
F2	А	X1	200	5	0,08	0,4

		X3	168	3,4	0,64	2,176
F3	A	X1	150	6,8	0,08	0,544
F3	Б	X1	250	2,5	0,64	1,6

$$1. \quad F1b - F2a - F3a$$

$$2. \quad F1b - F2a - F3b$$

$$K_1 = 0,208 + 1,161 + 4,992 + 0,4 + 2,176 + 0,544 = 9,481$$

$$K_2 = 0,208 + 1,161 + 4,992 + 0,4 + 2,176 + 1,6 = 10,537$$

Варіант 2 має більший коефіцієнт якості, а саме тому ми і оберемо його для реалізації.

#### 4.4 Економічний аналіз варіантів розробки

Кожен з варіантів реалізації включатиме в себе такі завдання:

1. Розробку алгоритму.
2. Розробку методів для використання.

Для першого варіанту реалізації додатковим завданням буде запис результатів у текстовий файл, для другого – розробка окремого блоку для візуалізації даних.

Для першого завдання (алгоритм групи складності 1, ступінь новизни А, вид використаної інформації – НДІ)  $TP = 90$  людино-днів,  $K_{\Pi} = 1,7$ ,  $K_{СК} = 1$ ,  $K_{СТ} = 0,7$ . Тому  $T1 = 191,25$  людино дні.

Для другого завдання (алгоритм групи складності 2, ступінь новизни А, вид використаної інформації – НДІ)  $TR = 36$  людино-днів.  $K_{II} = 1.51$ ,  $K_{CK} = 1$ ,  $K_{CT} = 0.6$ . Тому  $T2 = 36 * 1.51 * 0.6 = 32.6$  людино днів.

Для додаткового завдання у першому варіанті реалізації (алгоритм групи складності В, ступінь новизни 1, вид використаної інформації – НДІ)  $TR = 43$  людино-днів,  $K_{II} = 0.81$ ,  $K_{CK} = 1$ ,  $K_{CT} = 0.7$ . Тому  $T31 = 24.4$  людино-днів.

Для додаткового завдання у другому варіанті реалізації (алгоритм групи складності В, ступінь новизни 2, вид використаної інформації – НДІ)  $TR = 19$  людино-днів,  $K_{II} = 0.72$ ,  $K_{CK} = 1$ ,  $K_{CT} = 0.7$ . Тому  $T31 = 9.6$  людино днів.

Звідси, повна трудомісткість першого варіанту:

$$T = 191.25 + 32.6 + 24.4 = 248.25 \text{ людино-днів}$$

Повна трудомісткість другого варіанту:

$$T = 191.25 + 32.6 + 9.6 = 233.45 \text{ людино-днів}$$

У розробці братиме участь 1 аналітик з окладом 15000 грн. Його погодинна оплата складе:  $15000 / 21 / 8 = 89.29$  грн.

Заробітна плата:

$$IC_{3II} = 248.25 * 89.29 * 1.2 = 26599.49 \text{ грн.}$$

$$IC_{3II} = 233.45 * 89.29 * 1.2 = 25019.06 \text{ грн}$$

Витрати на оплату однієї машино-години

$$C_{\Gamma} = 12 * 15000 * 0.2 = 36000 \text{ грн.}$$

Відповідно, відрахування на єдиний соціальний внесок складуть

$$I C_{BID} = 26599.49 * 0.22 = 5851.88$$

$$I IC_{ВД} = 250190.6 * 0.22 = 5504.19$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 7000 грн.

$$C_a = K_{TM} * K_a * C_{ПР} = 1.15 * 0.25 * 7000 = 2012,5 \text{ грн}$$

Профілактичні видатки  $C_p = K_{TM} * K_p * C_{ПР} = 1.15 * 0.05 * 7000 = 402,5 \text{ грн}$

$$T_{ЕФ} = (D_K - D_B - D_C - D_P) * t_3 * K_B = (365 - 104 - 10 - 16) * 8 * 0.9 = 1692 \text{ годин.}$$

Витрати на оплату електроенергії (з урахуванням ПДВ):

$$C_{ЕЛ} = T_{ЕФ} * N_C * K_3 * C_{ЕН} = 1692 * 0.2 * 2 * 1.75 = 1184.4 \text{ грн.}$$

Накладні витрати

$$C_H = C_{ПР} * 0.67 = 7000 * 0.67 = 4690 \text{ грн.}$$

Річні експлуатаційні втрати:

$$C_{ЕКС} = 4690 + 1184,4 + 402,5 + 2012,5 + 7920 + 36000 = 52207$$

Собівартість однієї машино-години ЕОМ дорівнює:

$$C_{M-Г} = C_{ЕКС} / T_{ЕФ} = 52757.24 / 1692 = 31.18 \text{ грн./год.}$$

Виплати на основу машинного часу:

$$I. C_M = 31.18 * 248.25 = 7740.44 \text{ грн.}$$

$$II. C_M = 31.18 * 233.45 = 7278.97 \text{ грн.}$$

Накладні витрати складають 67% від заробітної плати:

$$I. C_H = 0.67 * 26599,49 = 17821,66$$

$$II. C_H = 0.67 * 25019.06 = 16762,77$$

Вартість розробки програмного продукту

$$I C_{III} = 7740,44 + 17821,66 + 26599,49 + 5851,88 = 58013,47$$

$$II C_{III} = 7278,97 + 16762,77 + 25019,06 + 5504,19 = 54564,99$$

Розрахуємо коефіцієнт техніко-економічного рівня:

$$КТЕР1 = 9,481 / 58013,47 = 0,000163$$

$$КТЕР2 = 10,537 / 54564,99 = 0,000193$$

#### 4.5 Висновки до розділу

Отже, найбільш ефективним є 2 варіант з коефіцієнтом техніко-економічного рівня 0,000193. А отже, він і буде оптимальним для реалізації.

## СПИСОК ЛІТЕРАТУРИ

1. Бідюк П.І. Системний підхід до побудови математичних моделей на основі часових рядів / Системні дослідження та інформаційні технології, №3, 2002. 1 14-131 с.
2. Бідюк П.І. Часові ряди: моделювання та прогнозування / Київ: ЕКМО, 2004. 144 с.
3. Пашин В.П. Функционально-стоимостный анализ конструкторско-технологических решений. - К.: РДЭНТП «Знание» УССР, 1989. - 22с.
4. Пашин В.П. Оцінка конкурентоспроможності електронних пристроїв на стадії проектування. - К. Економічний вісник НТУУ „КПІ”, 2006. - №3. с. 252-255.
5. Пашин В.П. Управление качеством изделий на основе функционально-стоимостного анализа. - К.: «Технология и организация производства», 1989. - №1. с. 17-19.
6. Бокс Дж. Анализ временных рядов. Прогноз и управление / Бокс Дж., Дженкинс Г. – М.: Мир, 1974. – Вып. 1, 2.
7. Бідюк П.І. Аналіз часових рядів: навчальний посібник / Бідюк П.І., Романенко В.Д., Тимошук О.Л. – К: Політехніка, 2010. – 317 с.
8. Жоголев, Е. А. Технология программирования. М.: Научный мир, 2004. 216 с.
9. Боэм, Б. Инженерное проектирование программного обеспечения. М.: Радио и связь, 1985. 512 с.
10. Зелковец, А. Принципы разработки программного обеспечения. М.: Мир, 1982. 368 с.
11. Gradinaru M. Milstein's type schemes for fractional SDEs / Gradinaru M., Nourdin I. // Ann. Inst H. Poincaré Probab Statist. — V. 45. — 2009. — № 4.— PP. 1085—1098.

## ВИСНОВКИ ЗА РОБОТОЮ

У даній роботі предметом дослідження були обрані процеси на фінансовому ринку, наразі ця проблематика є актуальною і привертаю увагу дедалі більшої кількості аналітиків, державних службовців та науковці

Був проведений детальний порівняльний аналіз даного питання на прикладі прогнозування курсу валют. Були зібрані дані з сайту Національного банку України та змодельовані та спрогнозовані курси валют (гривні до долару США та євро) за допомогою авторегресійних моделей.

Був розроблений власний програмний продукт, для побудови візуалізації та прогнозів.

Власна програма на мові програмування Python була порівняна з системою Eviews, результати виявились дуже схожими, що свідчить про коректність реалізації продукту.

ДОДАТОК А (ілюстративний матеріал доповіді)

# Моделі та методи прогнозування процесів фінансового ринку

Виконав:

студент IV курсу, групи КА-63

Корж Богдан Сергійович

Керівник:

доцент кафедри ММСА, д.т.н

Кузнецова Н.В

1

- Об'єкт дослідження - процеси на фінансовому ринку.
- Предмет дослідження – методи прогнозування, оцінки та аналізу якості прогнозів.
- Мета роботи - дослідити процес побудови моделей фінансового ринку за допомогою методів часових рядів.

2

## Постановка задачі

- Зібрати статистичні дані
- Виконати попередню обробку статистичних даних
- За допомогою авторегресії апроксимувати процес
- Візуалізувати результати
- Розробити власний програмний продукт на мові програмування Python з використанням бібліотек NumPy, Pandas тощо.
- Порівняти реалізацію власного продукту з пакетом Eviews

3

## Критерії якості моделей та прогнозів

- $R^2 = \frac{var(\hat{y})}{var(y)}$ .
- $DW = 2 - 2p$ ,
- $p = \frac{\sum_{k=2}^N e(k) * e(k-1)}{\sum_{k=1}^N [e(k)]^2}$
- $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [\hat{y}(i) - y(i)]^2}$ .
- $MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|^2}{|y_i|^2} 100\%$

4

## Авторегресійні моделі

Для авторегресійних моделей є характерним те, що у одній частині рівняння наведені минулі значення вимірів, а отже, значення у кожен обраний момент часу залежить від значень у попередніх моментах часу.

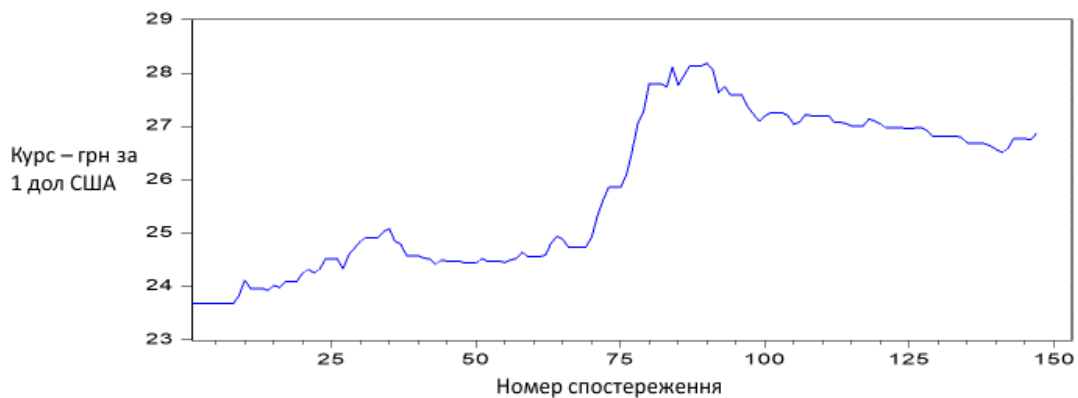
Наприклад:

- Авторегресія: 
$$y(k) = a_0 + \sum_{i=1}^n a_i y(k-i) + \varepsilon(k)$$

- Авторегресія ковзного середнього: 
$$y(k) = a_0 + \sum_{i=1}^n a_i y(k-i) + \sum_{j=0}^q b_j x(k-j) + \varepsilon(k)$$

5

## Графік динаміки зміни курсу гривні відносно долару



6

## Результати моделювання ряду гривні до долару США за допомогою регресійних моделей

- AP(1):  $y(k) = 25.38510 + 0.995771 * y(k - 1)$
- APKC(1,1):  $y(k) = 25.40154 + 0.994199 * y(k - 1) + 0.195504 * ma(k - 1)$

7

## Критерії якості моделі

Реалізація	Модель	$R^2$	DW	AIC
Eviews	AP(1)	0,990162	1,444104	-0,916841
Eviews	APKC(1,1)	0,990689	1,897183	-0,954775
Власна програма	AP(1)	0,9907928	1,444014	-0.9076725
Власна програма	APKC(1,1)	0,99132015	1,897065	-0.9788898

8

## Порівняння результатів прогнозування

Реалізація	Модель	RMSE	MAPE	Коеф. Тейла
Eviews	АРКС(1,1)	0.00026375	0.0149771	0,0003
Власна програма	АРКС(1,1)	0.000263749	0.014977143	0.000303817

9

## Візуалізація прогнозу



Номер значення	Реальні значення	Прогнозні значення
144	26.756782	26.769937
145	26.746210	26.774232
146	26.501244	26.485431
147	26.797304	26.789876
148	26.756782	26.769900
149	26.764893	26.755600
150	26.746210	26.764215

10

## Графік динаміки зміни курсу гривні відносно євро у 2020 році



11

## Результати моделювання ряду гривні до євро за допомогою регресійних моделей

- AP(1):  $y(k) = 28.23155 + 0.993011 * y(k - 1)$
- ARK(1,7):  $y(k) = 0.225940 + 0.990258 * y(k - 1) - 0.225940 * ma(k - 7)$

12

## Критерії якості моделі

Реалізація	Модель	$R^2$	DW	AIC
Eviews	AP(1)	0,981973	1,801327	-0,460989
Eviews	АРКС(1,7)	0,983070	1,814681	-0,507972
Власна програма	AP(1)	0,9825986	1,8024746	-0,46754
Власна програма	АРКС(1,7)	0,9836963	1,81583710	-0,49973

13

## Порівняння результатів прогнозування

Реалізація	Модель	RMSE	MAPE	Коеф. Тейла
Eviews	АРКС(1,7)	0.03174	0.14622	0.047309537
Власна програма	АРКС(1,7)	0.03172	0.14619	0.047308987

14

## Візуалізація прогнозу



Номер значення	Реальні значення	Прогнозні значення
141	29,3157	29,133761
142	29,5084	29,315585
143	29,7151	29,50944
144	29,6234	29,717524
145	29,6234	29,602668
146	29,6234	29,611607
147	29,9275	29,611051

15

## Висновки

- Був виконаний аналіз двох часових рядів (курсу гривні до долару США та євро)
- Створено власний програмний продукт на мові Python
- Було виконане порівняння результатів продукту і моделювання в системі Eviews
- Був побудований прогноз курсу валют на 7 днів наперед

16

Дякую за увагу

## ДОДАТОК Б (код програмного продукту)

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

%matplotlib inline

from pandas import read_csv

from matplotlib import pyplot

series = read_csv('euro_diploma1.txt')

print(series)

series.plot()

pyplot.show()

from pandas import read_csv

from matplotlib import pyplot

series = read_csv('euro_diploma1.txt')

print(series)

series.plot()

pyplot.show()
```

```
model = ARMA(series,order=(1, 1) )
yhat = model_fit.predict(len(series), len(series))

import numpy as np

import matplotlib.pyplot as plt

from pandas import read_csv

import math

import pandas as pd

from pandas import read_csv

from matplotlib import pyplot

tory = [x for x in train]

predictions = list()

for t in range(len(test)):

    model = ARMA(history, order=(1,1))

    model_fit = model.fit(dispatch=0)

    output = model_fit.forecast()

    yhat = output[0]

    predictions.append(yhat)

    obs = test[t]

    history.append(obs)

    print('predicted=%f, expected=%f' % (yhat, obs))
```

```
error = mean_squared_error(test, predictions)

print('Test MSE: %.3f % error)

# plot

pyplot.plot(test)

series = read_csv('diploma.txt')

print(series)

series.plot()

pyplot.show()

from pandas import read_csv

from matplotlib import pyplot

from statsmodels.graphics.tsaplots import plot_acf

from statsmodels.graphics.tsaplots import plot_pacf

series = read_csv('diploma.txt')

plot_acf(series, lags=7)

pyplot.show()

plot_pacf(series, lags=7)

pyplot.show()

from statsmodels.tsa.arima_model import ARMA
```

```

dataframe = read_csv('euro_diploma1.txt')

dataset = dataframe.values

dataset = dataset.astype('float32')

for t in range(len(test)):

    length = len(history)

    lag = [history[i] for i in range(length-window,length)]

    yhat = coef[0]

    for d in range(window):

        yhat += coef[d+1] * lag[window-d-1]

    obs = test[t]

    predictions.append(yhat)

    history.append(obs)

    print('predicted=%f, expected=%f' % (yhat, obs))

error = mean_squared_error(test, predictions)

print('Test MSE: %.3f' % error)

# plot

import numpy as np

def Theil(y_test,y_pred):

    return (np.sqrt(np.mean((np.array(test) -
np.array(predict)**2)))/((np.sqrt(np.mean(np.array(test)**2)))+(np.sqrt(np.mean(
np.array(predict)**2))))))

```

```
print("Theil:",round(Theil(test, predict), 9))
```

```
pyplot.plot(test)
```

```
pyplot.plot(predictions, color='red')
```

```
pyplot.show()
```

```
from sklearn.utils import check_arrays
```

```
def mean_absolute_percentage_error(y_true, y_pred):
```

```
    y_true, y_pred = check_arrays(y_true, y_pred)
```

```
    ## Note: does not handle mix 1d representation
```

```
    #if _is_1d(y_true):
```

```
        # y_true, y_pred = _check_1d_array(y_true, y_pred)
```

```
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
```

```
mean_absolute_percentage_error([1,2], [1,3])
```