

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

Кафедра прикладної математики та моделювання складних систем

Допущено до захисту

Завідувач кафедри ПМ та МСС

_____ Коплик І.В.

(підпис)

« ____ » _____ 2023 р.

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня «бакалавр»

спеціальність 113 «Прикладна математика»

освітньо-професійна програма «Прикладна математика»

тема роботи: **«ARIMA – МОДЕЛІ ПРОГНОЗУВАННЯ ФОНДОВОГО
РИНКУ»**

Виконавець

студент факультету ЕЛІТ

Кольченко Анастасія Юріївна _____
(підпис)

Науковий керівник

док. фіз.-мат. наук, професор

Лисенко Олександр Володимирович _____
(підпис)

Суми – 2023

РЕФЕРАТ

Кваліфікаційна робота: 60 с., 35 рисунків, 25 джерел.

Мета роботи: створити та проаналізувати ARIMA моделі фондового ринку, передбачити ціни на акції найбільших компаній США, провести аналіз індексу Доу-Джонса.

Об'єкт дослідження: динаміка цін на фондовому ринку, ціни на акції 30-ти найбільших компаній США, динаміка індексу Доу-Джонса.

Предмет дослідження: прогноз цін на акції, прогноз індексу Доу-Джонса.

Методи дослідження: ARIMA моделі, статистичні методи, методи очищення та візуалізації дани.

Ключові слова: ARIMA МОДЕЛІ, ЧАСОВІ РЯДИ, ФОНДОВИЙ РИНОК, ПРОГНОЗУВАННЯ, ЦІНИ НА АКЦІЇ

Зміст

Вступ	4
Розділ 1. Фондовий ринок – основні характеристики	5
1.1 Основні означення	5
1.2 Роль фондового ринку в економіці країн	6
1.3 Характеристика сучасного фондового ринку	8
Розділ 2. Математичні моделі часових рядів	11
2.1 Основні означення	11
2.1.1 Типи даних часових рядів	11
2.1.2 Дані лінійного та нелінійного часових рядів	11
2.1.3 Ідентифікація даних часових рядів	12
2.1.4 Розгляд часових рядів	12
2.1.5 Компоненти часового ряду	13
2.2 Аналіз часових рядів.....	14
2.2.1 Вступ до аналізу часових рядів	14
2.2.2 Часові ряди і стахостичний процес	15
2.2.3 Концепція стаціонарності	18
2.3 Моделі часових рядів.....	19
2.3.1 Математична модель [25] для аналізу часових рядів	19
2.3.2 Автокореляція.....	21
2.3.3 Параметри сезонного згладжування δ	22
2.3.4 Процес авторегресії та ковзного середнього	23
Розділ 3. Використання математичних моделей часових рядів для аналізу та прогнозування фондового ринку	25
3.1 Опис даних.....	25
3.2 Перевірка на стаціонарність. Змінна статистика.	27
3.3 Побудова моделі ARIMA та прогнозування часового ряду	34
3.4 Порівняння спрогнозованих та реальних значень цін акцій.....	37
3.5 Індекс Доу-Джонса	51
Висновки	57
Список використаних джерел	58

Вступ

У сучасному світі фондовий ринок є одним з найбільш динамічних та складних секторів глобальної економіки. Незалежно від його масштабів, успішність учасників ринку, таких як інвестори, трейдери та фондові брокери, в значній мірі залежить від їх здатності прогнозувати майбутнє поведінку цін на фондових ринках.

Останнім часом ARIMA-моделі (Autoregressive Integrated Moving Average) стали дуже популярними інструментами для прогнозування фондового ринку. Ці моделі базуються на аналізі часових рядів, що дозволяє розуміти та передбачати залежності між попередніми значеннями цін та їх майбутніми змінами.

Мета даної преддипломної практики полягає в дослідженні та застосуванні ARIMA-моделей для прогнозування фондового ринку. Важливим завданням є розробка ефективних методів та стратегій прогнозування, що дозволять учасникам ринку зробити обґрунтовані торгові рішення та знизити ризики фінансових втрат.

У ході практики будуть вивчені основні теоретичні аспекти ARIMA-моделей, включаючи авторегресійну (AR) компоненту, компоненту з інтегрованим рухом (I) та ковзне середнє (MA). Дослідження буде проведено на реальних фінансових даних з різних фондових ринків з метою оцінки ефективності ARIMA-моделей та їх здатності до точного прогнозування.

Крім того, практика передбачає застосування сучасних інструментів та програмних засобів, таких як Python, для розробки та реалізації ARIMA-моделей. Це дозволить розширити свої навички у сфері аналізу даних та програмування, а також зрозуміти практичні аспекти використання ARIMA-моделей у фінансових ринках.

Розділ 1.

Фондовий ринок – основні характеристики

1.1 Основні означення

Ринок цінних паперів (або фондовий ринок) є частиною фінансового ринку (аналогічно з ринком позичкового капіталу, валютним ринком та ринком золота). На фондовому ринку звертається особливий товар, специфічні фінансові інструменти – цінні папери. Фондовий ринок надає компаніям можливість залучення додаткового капіталу з допомогою продажу частки власності (у разі випуску цінних паперів) чи прийняття собі боргових зобов'язань.

Цінні папери[1] – документи встановленої форми та реквізитів, що засвідчують майнові права, здійснення або передача яких можливі лише за їх пред'явлення. Дані майнові права з цінних паперів обумовлені наданням грошей у позику та створення різних підприємств, купівлею – продажем, заставою майна, тощо. У зв'язку з цим цінних паперів дають їх власникам декларація про отримання встановленого доходу.

Учасники ринку цінних паперів: емітенти (потребують капіталу і випускають цінні папери), інвестори (вкладають гроші в цінні папери), посередники (брокери та дилери). З цінними паперами допускається здійснення наступних операцій: купівля, продаж, переуступка, зберігання, передача у спадок, дарування, обмін. Цінні папери не можуть виступати як загальний еквівалент, проте виконують такі функції грошей як засіб платежу та розрахунків.

Цінні папери відіграють важливу роль в інвестиційному обороті, тому що є засобом переміщення капіталу від інвесторів (готових вкласти вільні гроші) до емітентів (які їх потребують), а також посередників, що діють на ринку цінних паперів.

Серед цінних паперів виділяють класичні пайові та боргові цінні папери, похідні (ф'ючерси, опціони, варанти тощо) та фінансові інструменти (векселі, депозитні та ощадні сертифікати).

Боргові цінні папери – приватні та державні облігації.

Облігації[2] – це емісійний цінний папір, який закріплює за власником право на отримання номінальної вартості та відсотка (або іншого майнового еквівалента) у встановлений термін.

Пайові цінні папери є різними видами акцій.

Акція[2-4] – цінні папери, які засвідчують право їх власника на частку власності акціонерної компанії, включаючи право на голосування у прийнятті основних рішень і право на отримання дивідендів з прибутків компанії. Можна розглянути різні класифікації видів акцій залежно від ознаки.

Споживча вартість – здатність акції приносити дохід її власнику двома способами: дивіденди та збільшення курсової вартості акції.

Номінальна вартість акції зафіксована емітентом при випуску акції.

1.2 Роль фондового ринку в економіці країн

Насамперед, за допомогою ринку цінних паперів забезпечується мобілізація вільних грошових та інших коштів населення, підприємств та організацій, держави, а також спрямування (розміщення) їх у різні сфери національної економіки. Внаслідок цього споживачі капіталу отримують доступ до джерел фінансування, а інвестори отримують можливість стати власниками, зберегти та примножити свої кошти. Це результат, до якого призводить реалізація перерозподільної функції ринку цінних паперів.

Ринок цінних паперів виявляє ринкові ціни[11-14]. Що склалася під впливом попиту та пропозиції. Ринок цінних паперів можна справедливо назвати барометром стану суспільства загалом, оскільки він дуже чуйно реагує на рух економічного циклу та зміну не лише процесів в економіці, а й політичної обстановки. У періоди депресії, криз, зокрема й політичних,

індекси курсів цінних паперів падають. У періоди поживлення та підйому вони зростають.

Систематичний оборот цінних паперів стимулює докорінне оновлення фінансової системи країни. За допомогою ринку цінних паперів змінюються зв'язки між основними економічними операторами – виробничими підприємствами, виробниками послуг, комерційними банками, біржами та населенням.

Ринок цінних паперів стає показником організації системи корпоративного управління підприємства, оскільки зміна курсової вартості акцій підприємства свідчать або про інтерес громадськості, що підвищується, і, зокрема, інвесторів до даного підприємства, або, навпаки, до зниження такого інтересу у зв'язку з становищем, що погіршується.

При купівлі населенням акцій, облігацій та інших цінних паперів за готівку скорочується кількість грошових знаків у обігу, що благотворно позначається на стані грошового обігу[16]. Дедалі ширше поширення акцій та облігацій робить фондовий ринок важливим інструментом вивчення кон'юнктури. Фондовий ринок допомагає здійснити ринкове зіставлення різних підприємств, галузей, сфер господарства.

Володіння цінними паперами та участь у торгівлі ними збільшують доходи населення, сприяють підвищенню його матеріального добробуту, а також активізують людину – головну рушійну силу суспільства.

Фондовий ринок сприяє формуванню класу власників. Власники акцій стають співвласниками підприємства. Посилюються демократичні засади в управлінні виробництвом. Підвищується зацікавленість працівників – акціонерів у позитивних кінцевих результатів своєї роботи. Таким чином, фондовий ринок – один із найважливіших механізмів подолання відчуження людини від засобів виробництва та результатів його праці.

У той же час, фондовий ринок може спричинити збій у нормальному економічному розвитку держави, банкрутства великих і дрібних підприємств,

негативних соціальних наслідків, падіння курсу національної валюти, різкого скорочення припливу іноземних інвестицій тощо.

Отже, ринок цінних паперів може мати різний вплив на економіку держави. При правильній організації та вмілому державному регулюванні, при зміцненні та вдосконаленні інституційного середовища вплив фондового ринку на народне господарство позитивний. При неправильному регулюванні, відсутності до нього належної уваги з боку держави, фондовий ринок може створити кризові ситуації не тільки у емітентів, інвесторів та інвестиційних посередників, а й спричинити глобальні колапси та катаклізми всієї господарської системи країни.

Створення такої системи, коли постачання цінних паперів та грошовий платіж відбуваються у межах однієї організації. У такому механізмі досить легко контролювати одночасність здійснення зустрічних проводок за грошовими рахунками на рахунках депо[18].

Це означає, наприклад, облігації, представлені або глобальними, або індивідуальними сертифікатами, не залишають депозитарні клірингові палати, а просто пересуваються за рахунками клієнтів та агентів. Отже, це дозволяє ефективніше проводити розрахунки.

1.3 Характеристика сучасного фондового ринку

Фондовий ринок[3] – це організований орган, де брокери торгують акціями публічних компаній, які пропонують свої акції шляхом первинного публічного розміщення. Ціни акції на ринку відображають попит і пропозицію, і трейдери намагаються передбачити поведінку акцій.

Колись фондовий ринок був фізичним місцем, де трейдери зустрічалися віч – на – віч для укладання угод, але сьогодні більшість торгів відбувається в електронному вигляді.

Публічні компанії є ключовим компонентом фондових ринків. Публічні компанії – це ті, які мають акції, які купуються та продаються на публічній

фондовій біржі. Перш ніж акції можна буде продати, вони повинні бути розміщені на біржі. Щоб захистити своїх інвесторів, публічна компанія зобов'язана розкривати фінансову та ділову інформацію, яка може вплинути на вартість акцій.

Торгівля на фондовій обмежена біржовими брокерами та трейдерами, які є членами біржі. Індивідуальні інвестори повинні мати брокерський рахунок для участі в торгах. Для багатьох людей брокерські послуги надаються як частина пенсійного інвестиційного фонду, який фінансується роботодавцем. Особам, які хочуть торгувати самостійно, потрібен індивідуальний обліковий запис.

Ціни акцій компанії відображає попит і пропозицію на самі акції і часто не залежать від успіху компанії. Акції компанії можуть вважатися бажаними з різних причин, від потужності галузі до популярності бренду.

Щоб отримати прибуток, трейдери повинні передбачити, чи зросте вартість акцій. Ціни акцій часто відображають загальну економіку та можуть бути нестабільними, оскільки інвестори реагують на фінансові новини та поточні події, які успішно прогнозують, можуть отримати значні прибутки.

6 характеристик фондового ринку:

1) Економічний барометр:

Фондова біржа є надійним барометром для вимірювання економічного стану країни. Кожна велика зміна в країні та економіці відображається на курсах акцій. Зростання або падіння цін на акції вказує на цикл буму чи спаду економіки. Фондова біржа також відома як пульс економіки або економічне дзеркало, яке відображає економічні умови країни.

2) Ціноутворення та цінні папери:

Фондовий ринок допомагає оцінювати цінні папери прибуткових і орієнтованих на зростання компаній оцінюються вище, оскільки на такі цінні папери є більший попит. Оцінка цінних паперів корисна для інвесторів, уроду та кредиторів. Інвестори можуть знати вартість своїх інвестицій, кредитори

можуть оцінити кредитоспроможність, а уряд може накладати податки на вартість цінних паперів.

3) Безпека транзакцій:

На фондовому ринку торгуються лише цінними паперами, що котируються на біржі, а орган фондової біржі включають назви компаній у торговий список лише після перевірки надійності компанії. Компанії, які перераховані, також повинні діяти відповідно до сурових правил і норм. Це забезпечує безпеку угод через біржу.

4) Сприяння економічному зростанню:

На фондовій біржі купуються та продаються цінні папери різних компаній. Цей процес вилучання та реінвестування допомагає інвестувати в найбільш продуктивну інвестиційну пропозицію, що призводить до формування капіталу та економічного зростання.

5) Ліквідність[3]:

Основна функція фондового ринку – забезпечити готовий ринок для продажу та купівлі цінних паперів. Наявність біржового ринку гарантує інвесторам, що їхні інвестиції можна конвертувати в готівку, коли вони захочуть. Інвестори можуть інвестувати в довгострокові інвестиційні проекти без будь – яких вагань, оскільки завдяки біржі вони можуть конвертувати довгострокові інвестиції в короткострокові та середньострокові.

6) Сприяння розвитку звичок заощаджень та інвестицій:

Фондовий ринок пропонує привабливі можливості для інвестицій у різні цінні папери. Ці привабливі можливості спонукають людей більше заощаджувати та інвестувати в цінні папери корпоративного сектора, а не інвестувати в непродуктивні активи, такі як золото, срібло тощо.

Розділ 2.

Математичні моделі часових рядів

2.1 Основні означення

Часовий ряд [5,10] - це послідовність спостережень або вимірів, зібраних у різні моменти часу. Вони використовуються для вивчення та аналізу змін у якій-небудь змінній або явищі протягом часу. Математично, часовий ряд представляє собою послідовність випадкових змінних, що вимірюються у різні моменти часу. Він може бути позначений як $\{x(t)\}$, де x - випадкова змінна, а t - момент часу, для якого вимірюється ця змінна.

2.1.1 Типи даних часових рядів

Дані часових рядів можна класифікувати на два типи:

- Вимірювання, зібрані через регулярні проміжки часу (метрики)
- Вимірювання, зібрані через нерегулярні проміжки часу (події)

Оскільки події відбуваються через нерегулярні проміжки часу, вони непередбачувані, і їх неможливо змоделювати чи спрогнозувати, так як прогнозування передбачає, що все, що сталося в минулому, є хорошим показником того, що станеться в майбутньому.

Прикладом даних часового ряду може бути будь-яка інформаційна послідовність, отримана через певні проміжки часу (регулярні чи нерегулярні). Типовими прикладами даних можуть бути будь-які події: від пульсу до ціни за одиницю товару в магазині.

2.1.2 Дані лінійного та нелінійного часових рядів

Лінійний часовий ряд – це такий, у якому кожному моменту часу X_t можна розглянути як лінійну комбінацію минулих або майбутніх значень або різниць. Нелінійні часові ряди генеруються нелінійними динамічними рівняннями. Вони мають особливості, які неможливо змоделювати за допомогою лінійних процесів: дисперсія, що змінюється в часі, асиметричні цикли, структури з

вищим моментом порогови та розриви. Ось кілька важливих міркувань під час роботи з лінійними та нелінійними даними часових рядів:

- Якщо рівняння регресії не відповідає правилам для лінійної моделі, це має бути нелінійна модель.
- Нелінійна регресія може відповідати величезній різноманітності кривих.
- Визначальною характеристикою для обох типів моделей є функціональні форми.

2.1.3 Ідентифікація даних часових рядів

Дані часових рядів унікальні тим, що вони мають природний порядок у часі: порядок, у якому спостерігалися дані, має значення. Ключова відмінність даних часових рядів від звичайних даних полягає в тому, що ви завжди ставите питання про них протягом певного часу. Часто простий спосіб визначити, чи є набір даних, з яким ви працюєте, часовим рядом чи ні, це перевірити, чи одна з ваших осей є часом.

2.1.4 Розгляд часових рядів

Незмінність – оскільки дані часових рядів надходять у часовому порядку, вони майже завжди записуються в новий запис і мають бути незмінними та лише додаватися (додаватися до існуючих). Зазвичай він не змінюється, а додається в порядку, у якому відбуваються події. Ця властивість відрізняє дані часових рядів від реляційних даних, які зазвичай є змінними та зберігаються в реляційних базах даних, які використовують онлайн-обробку транзакцій, де рядки в базах даних оновлюються під час виконання транзакцій і більш-менш випадковим чином. Приймаючи замовлення для існуючого клієнта, наприклад, оновлює таблицю клієнтів, щоб додати придбані товари, а також оновлює таблицю запасів, щоб показати, що вони більше не доступні для продажу.

Той факт, що дані часових рядів упорядковані, робить їх унікальними в просторі даних, оскільки вони часто відображають послідовну залежність.

Послідовна залежність виникає, коли значення точки даних в один момент часу статистично залежить від іншої точки даних в інший час.

Хоча немає подій, які існують поза часом, є події, де час не має значення. Дані часових рядів – це не просто події, які відбуваються в хронологічному порядку – це події, цінність яких зростає, коли ви додаєте час як вісь. Дані часових рядів іноді існують із високим рівнем деталізації, часто до мікросекунд або навіть наносекунд.

2.1.5 Компоненти часового ряду

Різні причини або сили, які впливають на значення спостереження і часовому ряді, є компонентами часового ряду. Чотири категорії компонентів часових рядів:

- Тренд
- Сезонні зміни
- Циклічні варіації
- Випадкові або нерегулярні рухи

Тренд показує загальну тенденцію збільшення чи зменшення даних протягом тривалого періоду часу. Тренд – це плавна, загальна, довгострокова, середня тенденція[23]. Не завжди обов'язково, щоб збільшення або зменшення відбувалося в одному напрямку протягом заданого періоду часу.

Можна спостерігати, що тенденції можуть посилюватися, зменшуватися або бути стабільними в різні відрізки часу. Але загальний тренд має бути висхідним, низхідним або стабільним.

Періодичні коливання. У часовому ряді є деякі компоненти, які мають тенденцію повторюватись протягом певного періоду часу. Вони діють регулярно спазматично.

Сезонні зміни – це ритмічні сили, які це ритмічні сили, які діють регулярним і періодичним чином протягом року. Ця варіація буде присутня в часовому ряді, якщо дані записуються щогодини, щодня, щотижня, щокварталу чи щомісяця.

Ці варіації вступають у дію або через природні сили, або через умови, створені людиною. Різні пори року або кліматичні умови відіграють важливу роль у сезонних коливаннях. Наприклад, виробництво сільськогосподарських культур залежить від пори року, продаж парасольок і плащів у сезон дощів, а також продаж електричних вентиляторів і кондиціонерів зростає влітку.

Варіації часових рядів, які діють протягом більше ніж одного року, є циклічними варіаціями. Це чотирифазний цикл, що включає фази процвітання, спаду, депресії та відновлення. Циклічна зміна може бути регулярною, а не періодичною. Підйоми та спади в бізнесі залежать від спільного характеру економічних сил та взаємодії між ними.

Існує ще один фактор, який викликає варіацію досліджуваної змінної. Вони не є регулярними варіаціями і є чисто випадковими або нерегулярними. Ці коливання є непередбаченими, неконтрольованими та непостійними. Такими силами є землетруси, війни, повені, голод і будь-які інші катастрофи.

2.2 Аналіз часових рядів

2.2.1 Вступ до аналізу часових рядів

Аналіз часових рядів[6-8] – сукупність математико – статистичних методів аналізу, призначених для виявлення структури часових рядів та їх прогнозування. Сюди належать, зокрема, методи регресійного аналізу. Виявлення структури часового ряду необхідне у тому, щоб побудувати математичну модель того явища, що є джерелом аналізованого часового ряду. Прогноз майбутніх значень часового ряду використовується для ефективного ухвалення рішень.

Аналіз часових рядів важливий з кількох причин:

- Прогнозування майбутніх значень: аналіз часових рядів дозволяє прогнозувати майбутні значення ряду. Це має велике значення для бізнесу та прийняття рішень, оскільки прогнози дозволяють планувати ресурси, визначати стратегії та уникати неочікуваних ситуацій.

- Виявлення тенденцій та паттернів: аналіз часових рядів допомагає виявити тенденції, циклічність, сезонність та інші паттерни у даних. Це дозволяє зрозуміти динаміку процесів, виявити фактори, що впливають на ряд, та зробити висновки щодо залежностей.
- Моделювання та прогнозування: аналіз часових рядів допомагає будувати моделі, що описують структуру ряду та його залежності. Це дає змогу прогнозувати майбутні значення, робити сценарії та проводити симуляції для визначення впливу різних факторів.
- Виявлення аномалій та випадкових змін: аналіз часових рядів дозволяє виявити аномалії, незвичайні події та випадкові зміни в даних. Це допомагає виявити проблеми, ризики та взаємозв'язки, що не спостерігаються на перший погляд.
- Оцінка ефективності: аналіз часових рядів дозволяє оцінити ефективність дій, стратегій та інтервенцій. Це дозволяє розуміти, які заходи працюють краще, а які потребують коригування.

Аналіз часових рядів допомагає розуміти та використовувати інформацію, яку містять дані про часові ряди, що є важливим для прийняття обґрунтованих рішень, прогнозування та оптимізації процесів.

2.2.2 Часові ряди і стахостичний процес

Часові ряди є принципово непередбачуваними, що означає, що ми не можемо точно передбачити, що станеться у майбутньому. Загалом, вважається, що часовий ряд $\{x(t), t = 0, 1, 2, \dots\}$ впливає з якоїсь ймовірнісної моделі, що описує розподіл випадкової змінної x_t . Важливо зазначити, що часовий ряд є всього лише одним із можливих варіантів реалізації стохастичного процесу, який його породжує[6].

Зазвичай припускається, що змінні часового ряду x_t є незалежними та розподілені за однаковим законом, наприклад, за нормальним розподілом.

Проте варто відзначити, що часові ряди фактично не є абсолютно незалежними та розподіленими однаково; вони демонструють визначений шаблон у довгостроковій перспективі. Наприклад, у нас є часовий ряд, що відображає зміну кількості відвідувачів в ресторані залежно від дня тижня. Якщо помічено, що по вихідних днях кількість відвідувачів зазвичай більша, ніж у робочі дні, то можна зробити припущення, що ця тенденція буде тривати й надалі.

Стохастичний процес є моделлю, що описує еволюцію випадкових змінних в часі. Він складається з послідовності спостережень, зібраних впродовж певного періоду часу. Ці спостереження можуть бути виміряні в різні моменти часу і мати випадковий характер.

Стаціонарний стохастичний процес в контексті часових рядів - це такий процес, у якого статистичні властивості залишаються незмінними в часі. Це означає, що середнє значення, дисперсія та автоковаріація ряду не залежать від часу або зміщені в часі. Тоді стаціонарний процес буде характеризуватися такими чотирма властивостями:

- 1) математичне сподівання стаціонарного ряду $E(y_t)$, є постійним, тобто середнє значення часового ряду, навколо якого змінюються рівні, є величиною постійною:

$$E(y_t) = \bar{y} = const; \quad (1)$$

- 2) дисперсія стаціонарного ряду є постійною. Вона характеризує варіацію рівнів часового ряду щодо його середнього значення

$$\bar{x}: D(y) = E(y_t - \bar{y})^2 = G^2(y) = const \quad (2)$$

3) автоковаріація стаціонарного ряду з лагом l є сталою, тобто коваріація між значеннями x_t і x_{t+l} , відокремленими інтервалом в l одиниць часу, визначається за формулою (3):

$$R_l(y_t) = cov(y_t, y_{t+l}) = E[(y_t - \bar{y})(y_{t+l} - \bar{y})]; \quad (3)$$

для стаціонарних рядів автоковаріація залежить тільки від величини лага l , тому справедливо рівність виду:

$$R_{l=0}(y_t) = G^2(y); \quad (4)$$

4) коефіцієнти автокореляції стаціонарного ряду з лагом l є постійними. Отже, автокореляція є нормованою автоковаріацією, оскільки Для стаціонарного процесу $G^2(y) = const$:

$$p_l = \frac{E[(y_t - \bar{y})(y_{t+l} - \bar{y})]}{\sqrt{E[(y_t - \bar{y})^2]E[(y_{t+l} - \bar{y})^2]}} = \frac{E[(y_t - \bar{y})(y_{t+l} - \bar{y})]}{G^2(y)} \quad (5)$$

Таким чином, коефіцієнт автокореляції порядку l визначається за формулою (6):

$$p_l = \frac{R_l(y_t)}{R_{l=0}(y_t)} \quad (6)$$

Нестаціонарним часовим рядом називається ряд, який не задовольняє перерахованим вище властивостям.

У стаціонарних часових рядах білий шум вважається випадковим процесом, що має розподіл Гаусса (нормальний розподіл) і не має автокореляції. Іншими словами, випадкові значення білого шуму незалежні між собою і не показують жодних кореляційних залежностей.

Основні характеристики білого шуму включають:

1. Рівномірну амплітуду: білий шум має однакову амплітуду для кожного випадкового значення. Це означає, що середнє значення (7) шуму дорівнює нулю, а дисперсія (8) залишається постійною.

$$E(w(t)) = 0 \quad (7)$$

$$\text{Var}(w(t)) = \sigma^2 \quad (8)$$

2. Відсутність автокореляції: випадкові значення білого шуму не корелюють між собою. Це означає, що жоден часовий лаг не впливає на наступне значення шуму.

Автоковаріація білого шуму з лагом l (часовим зсувом) дорівнює нулю:

$$\text{Cov}(w(t), w(t+l)) = 0, \text{ для будь-якого значення } l \neq 0 \quad (9)$$

3. Розподіл Гаусса: випадкові величини білого шуму мають нормальний розподіл:

$$w(t) \sim N(0, \sigma^2) \quad (10)$$

Таким чином, білий шум є послідовністю незалежних випадкових величин, кожна з яких має нормальний розподіл з нульовим середнім значенням та постійною дисперсією.

2.2.3 Концепція стаціонарності

Стаціонарність є важливою концепцією в аналізі стохастичних процесів, зокрема в контексті часових рядів. Стаціонарність вказує на властивості процесу, які залишаються незмінними в часі. Зрозуміння стаціонарності допомагає нам моделювати та передбачати поведінку стохастичних процесів.

Стаціонарний процес (стаціонарний часовий ряд) виконує наступні властивості:

- Стаціонарність за розподілом: Розподіл випадкових змінних, що складають процес, залишається незмінним усюди в часі. Іншими

словами, статистичні властивості процесу, такі як середнє значення та дисперсія, залишаються постійними.

- Стационарність за часовими характеристиками: Характеристики процесу, такі як автоковаріація та автокореляція, залежать лише від різниці у часі, а не від абсолютного часу. Наприклад, автоковаріація між значеннями процесу на часових точках t_1 та t_2 залежить лише від різниці $|t_1 - t_2|$.

Типи стаціонарних процесів включають:

- 1) Суворі (абсолютна) стаціонарність: Процес має незмінні розподіл та незалежні однаково розподілені послідовності. Всі статистичні характеристики, включаючи автоковаріацію та автокореляцію, залишаються постійними незалежно від зсуву у часі.
- 2) Слабка (узагальнена) стаціонарність: Розподіл процесу може змінюватися в часі, але статистичні характеристики залишаються постійними при достатньо малих зсувах у часі. У такому випадку, часові середні, дисперсії та коваріації залишаються постійними на практичних підвалах.

Зрозуміння типів стаціонарності дозволяє встановити певні характеристики процесу та використовувати відповідні методи аналізу та прогнозування в контексті часових рядів.

2.3 Моделі часових рядів

2.3.1 Математична модель [25] для аналізу часових рядів

Математично часовий ряд задається як

$$y_t = f(t) \quad (11)$$

Тут y_t – значення досліджуваної змінної в момент часу t . Якщо популяція є досліджуваною змінною в різні періоди часу $t_1, t_2, t_3, \dots, t_n$. Тоді часовий ряд

$$t: t_1, t_2, t_3, \dots, t_n$$

$$y_t: y_{t1}, y_{t2}, y_{t3}, \dots, y_{tn}$$

or,

$$t: t_1, t_2, t_3, \dots, t_n$$

$$y_t: y_1, y_2, y_3, \dots, y_n$$

Адитивна модель аналізу часових рядів.

Якщо y_t є значенням часового ряду в момент часу t . T_t , S_t , C_t і R_t – значення тренду, сезонні, циклічні та випадкові коливання в момент часу t відповідно. Відповідно до адитивної моделі часовий ряд можна виразити як

$$y_t = T_t + S_t + C_t + R_t \quad (12)$$

Ця модель передбачає, що всі чотири компоненти часового ряду діють незалежно один від одного.

Мультиплікативна модель для аналізу часових рядів.

Мультиплікативна модель передбачає, що різні компоненти в часовому ряді діють пропорційно один одному. Відповідно до цієї моделі

$$y_t = T_t \times S_t \times C_t \times R_t \quad (13)$$

Змішані моделі

Різні припущення призводять до різних комбінацій адитивних та мультиплікативних моделей

$$y_t = T_t + S_t + C_t R_t \quad (14)$$

Аналіз часових рядів також можна зробити за допомогою моделі

$$y_t = T_t + S_t \times C_t \times R_t \quad (15)$$

або

$$y_t = T_t \times S_t + C_t \times R_t \quad (16)$$

тощо.

Для вирішення завдань існує велика кількість різних методів. З них найбільш поширеними є такі:

- Кореляційний аналіз, що дозволяє виявити суттєві періодичні залежності та їх лаги (затримки) всередині одного процесу (автокореляція) або між декількома процесами (кроскореляція);
- Спектральний аналіз, що дозволяє знаходити періодичні та квазіперіодичні складові часового ряду;
- Згладжування та фільтрація, призначені для перетворення часових рядів з метою видалення з них високочастотних або сезонних коливань;
- Моделі авторегресії та ковзного середнього, які виявляються особливо корисними для опису та прогнозування процесів, що виявляють однорідні коливання навколо середнього значення;
- Прогнозування, що дозволяє на основі підібраної моделі поведінки тимчасового ряду передбачити його значення у майбутньому.

2.3.2 Автокореляція

Кількісною характеристикою подібності між значеннями ряду в сусідніх точках є автокореляційна функція (або просто автокореляція), яка задається таким співвідношенням:

$$r_\tau = \frac{E((y_t - Ey)(y_{t+\tau} - Ey))}{Dy} \quad (16)$$

Автокореляція[23, 25] – це кореляція Пірсона між вихідним рядом і його версію, зсунутою на кілька відліків. Кількість відліків, на яку зрушено ряд, називається лагом автокореляції τ . Значення, що приймаються автокореляцією, такі ж, як і у коефіцієнта Пірсона: $r_\tau \in [-1; 1]$. Обчисли автокореляцію за вибіркою можна, замінивши у формулі математичне очікування на вибіркоче середнє, а дисперсію – на вибіркочову дисперсію:

$$r_\tau = \frac{\sum_{t=1}^{T-\tau} (y_t - \bar{y})(y_{t+\tau} - E y)}{\sum_{t=1}^{T-\tau} ((y_t - \bar{y}))^2} \quad (17)$$

2.3.3 Параметри сезонного згладжування δ

Загалом прогноз на один крок вперед обчислюється наступним чином:

Адитивна модель:

$$\text{Прогноз}_t = S_t + I_{t-p} \quad (18)$$

Мультиплікативна модель:

$$\text{Прогноз}_t = S_t \times I_{t-p} \quad (19)$$

У цій формулі (16), (17) S_t позначає (просте) експоненційно згладжене значення ряду в момент t , і I_{t-p} позначає згладжений сезонний фактор в момент $t - p$ (p - довжина сезону). Таким чином, у порівнянні з простим експоненційним згладжуванням, прогноз «поліпшується» додаванням або множенням сезонної компоненти. Ця компонента оцінюється незалежно за допомогою простого експонентного згладжування наступним чином:

Адитивна модель:

$$I_t = I_{t-p} + \delta \times (1 - \alpha) \times e_t \quad (20)$$

Мультиплікативна модель:

$$I_t = I_{t-p} + \delta \times (1 - \alpha) \times \frac{e_t}{S_t} \quad (21)$$

Передбачена сезонна компонента в момент t обчислюється, як відповідна компонента на останньому сезонному циклі плюс помилка (e_t , мінус прогнозоване значення, що спостерігається в момент t). Зрозуміло, що параметр δ набуває значення між 0 і 1. Якщо він дорівнює нулю, то сезонна складова на наступному циклі та сама, що і на попередньому. Якщо δ дорівнює 1, то сезонна складова «максимально» змінюється на кожному кроці через відповідну помилку (множник $(1 - \alpha)$ не розглядається через стислість введення). У більшості випадків, коли сезонність присутня, оптимальне значення δ лежить між 0 і 1.

2.3.4 Процес авторегресії та ковзного середнього

Процес авторегресії. Більшість часових рядів містять елементи, що послідовно залежать один від одного. Таку залежність можна виразити наступним рівнянням:

$$x_t = \xi + \phi_1 * x_{(t-1)} + \phi_2 * x_{(t-2)} + \phi_3 * x_{(t-3)} + \dots + \varepsilon \quad (22)$$

де ξ - константа, ϕ_1, ϕ_2, ϕ_3 - параметри авторегресії.

Кожне спостереження є сумою випадкової компоненти (випадковий вплив, ε) та лінійної комбінації попередніх спостережень.

Процес ковзного середнього. На відміну від процесу авторегресії, в процесі ковзного середнього кожен елемент руху схильний до сумарного впливу попередніх помилок. Загалом це можна записати так:

$$x_t = \mu + \varepsilon_t - \theta_1 * \varepsilon_{t-1} - \theta_2 * \varepsilon_{t-2} - \theta_3 * \varepsilon_{t-3} - \dots \quad (23)$$

де μ – константа, $\theta_1, \theta_2, \theta_3$ - параметри ковзного середнього.

Поточне спостереження рядуу є сумою випадкової компоненти (випадковий вплив, ε) в даний момент і лінійної комбінації випадкових впливів у попередні моменти часу.

Розділ 3.

Використання математичних моделей часових рядів для аналізу та прогнозування фондового ринку

3.1 Опис даних

У своїй роботі я використала набір даних із історичними цінами на акції для 30 провідних компаній США . Я мала дані за 12 років з 01.01.2006 по 01.01.2018. Серед цих компаній є такі найпопулярніші як Apple (AAPL), Google (GOOGL), Amazone (AMZN) та інші.

Date	# Open	# High	# Low	# Close
3Jan06	7.39	7.56	7.17	7.24
29Dec17	175	177	175	176
2006-01-03	10.34	10.68	10.32	10.68
2006-01-04	10.73	10.85	10.64	10.71
2006-01-05	10.69	10.7	10.54	10.63
2006-01-06	10.75	10.96	10.65	10.9
2006-01-09	10.96	11.03	10.82	10.86

Рис. 3.1.1 – Приклад даних для компанії Apple

Для аналізу та прогнозуванню часових рядів у Python я використала такі основні бібліотеки:

- 1) `numpy`: Бібліотека для роботи з числовими даними в Python. Використовується для математичних операцій та обробки даних.
- 2) `pandas`: Бібліотека для маніпулювання та аналізу даних. Забезпечує потужні структури даних, такі як `DataFrame`, які полегшують обробку та аналіз часових рядів.
- 3) `matplotlib`: Бібліотека для візуалізації даних. Використовується для побудови графіків, діаграм та інших типів візуалізацій.

- 4) statsmodels: Бібліотека, яка надає інструменти для статистичного моделювання та економетричного аналізу. Включає функціонал для ARIMA-моделей та декомпозиції сезонності часових рядів.
- 5) pandas.plotting: Підмодуль бібліотеки pandas, який містить додаткові функції для візуалізації даних, зокрема часових рядів.

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.arima_model import ARIMA
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()
```

На малюнку 3.1.2 зображено графік акцій компанії Apple (AAPL) за період з 01.01.2006 по 01.01.2018.

```
df = pd.read_csv('AAPL_2006-01-01_to_2018-01-01.csv', sep=',',
encoding='latin1',
                parse_dates=['Date'], dayfirst=True,
                )
# Вивести список назв стовпців
print(df.columns)

# Переконайтеся, що назва стовпця містить лише літери, цифри та
# підкреслення
df.columns = df.columns.str.strip() # Видалити прогалини на
# початку та в кінці назв стовпців
df.columns = df.columns.str.replace(' ', '_') # Замінити
# прогалини на підкреслення

# Перевірити, чи міститься стовпець 'Date' у списку назв
# стовпців
if 'Date' in df.columns:
    df = df.set_index('Date')
    plt.plot(df['Close'])
    plt.xlabel('Date')
    plt.ylabel('Closing Price')
    plt.title('Stock Close Price over Time')
    plt.show()
else:
    print("Стовпець 'Date' не знайдено у файлі даних.")
```



Рис. 3.1.2 – Графік ціни закриття акцій компанії Apple з часом

3.2 Перевірка на стаціонарність. Змінна статистика.

Перш ніж будувати модель, ми повинні переконатися, що часовий ряд є стаціонарним. Існує два способи визначити:

- Змінна статистика: побудуйте графік середнього значення та стандартного відхилення. Часовий ряд є стаціонарним, якщо він залишається постійним з часом.
- Розширений тест Дікі-Фуллера: часовий ряд вважається стаціонарним, якщо значення p є низьким (згідно з нульовою гіпотезою), а критичні значення з довірчими інтервалами 1%, 5%, 10% максимально наближені до статистики ADF.

```
# Перетворення стовпця 'Date' на тип дати
df['Date'] = pd.to_datetime(df['Date'])

# Розрахунок ковзного середнього та стандартного відхилення
rolling_mean = df['Close'].rolling(window=12).mean()
rolling_std = df['Close'].rolling(window=12).std()

# Побудова графіка
plt.plot(df['Date'], df['Close'], color='blue', label='Original')
```

```

plt.plot(df['Date'], rolling_mean, color='red', label='Rolling
Mean')
plt.plot(df['Date'], rolling_std, color='black', label='Rolling
Std')

# Додавання легенди та назви графіка
plt.legend(loc='best')
plt.title('Rolling Mean & Rolling Standard Deviation')

# Відображення графіка
plt.show()

```

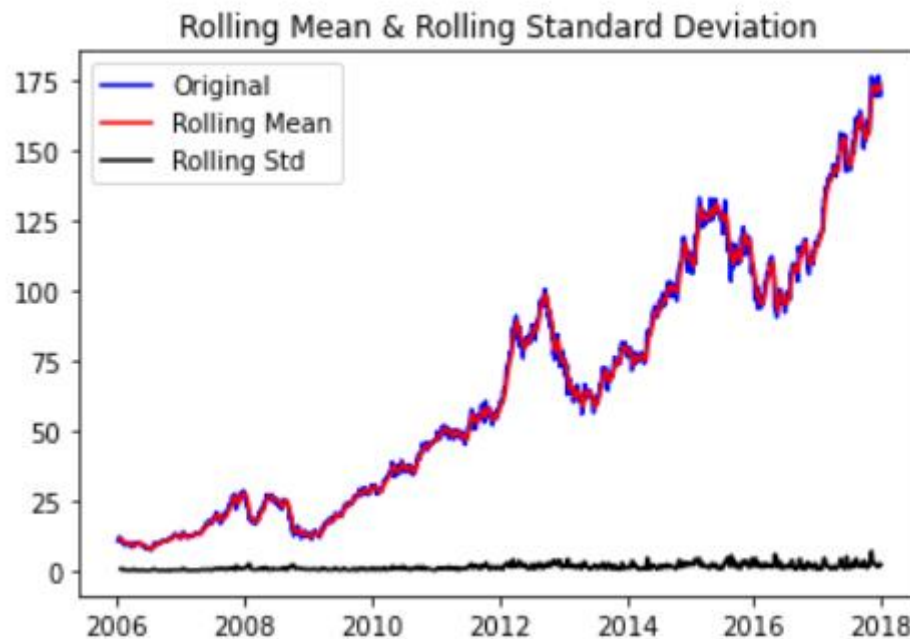


Рис. 3.2.1 - Середнє значення та стандартне відхилення на графіку

Середнє змінне (rolling mean) та стандартне відхилення (rolling standard deviation) зростають з часом. Отже, можна зробити висновок, що часовий ряд не є стаціонарним.

```

result = adfuller(df['Close'])
print('ADF Statistic: {}'.format(result[0]))
print('p-value: {}'.format(result[1]))
print('Critical Values:')
for key, value in result[4].items():
    print('\t{}: {}'.format(key, value))

```

Результат:

```

ADF Statistic: 0.5523672041905588
p-value: 0.9863807814299046
Critical Values:
1%: -3.4325345499510043

```

5%: -2.8625051928165557
10%: -2.5672837975912697

Статистика ADF критичних значень, а р-значення перевищує порогове значення (0,05). Таким чином, можна зробити висновок, що часовий ряд не є стаціонарним.

Отримання логарифму залежної змінної є простим способом зниження швидкості, з якою зростає середнє змінне.

```
# Логарифмирование столбца 'Close'  
df['Log_Close'] = np.log(df['Close'])  
  
# Встановлення стовпця 'Date' як індекс  
df.set_index('Date', inplace=True)  
  
# Створення графіка  
plt.figure(figsize=(10, 6))  
plt.plot(df.index, df['Log_Close'])  
plt.xlabel('Date')  
plt.ylabel('Log(Closing Price)')  
plt.title('Logarithmic Closing Price')  
plt.grid(True)  
plt.show()
```



Рис. 3.2.2 – Результат логарифмування середнього змінного

Існує кілька перетворень, які ми можемо застосувати до часового ряду, щоб зробити його стаціонарним. Наприклад, ми віднімаємо середнє змінне.

```
def get_stationarity(timeseries):  
    # Rolling statistics  
    rolling_mean = timeseries.rolling(window=12).mean()  
    rolling_std = timeseries.rolling(window=12).std()  
  
    # Rolling statistics plot  
    plt.plot(timeseries, color='blue', label='Original')  
    plt.plot(rolling_mean, color='red', label='Rolling Mean')  
    plt.plot(rolling_std, color='black', label='Rolling Std')  
    plt.legend(loc='best')  
    plt.title('Rolling Mean & Standard Deviation')  
    plt.show()  
  
    # Dickey-Fuller test:  
    result = adfuller(timeseries)  
    print('ADF Statistic: {}'.format(result[0]))  
    print('p-value: {}'.format(result[1]))  
    print('Critical Values:')  
    for key, value in result[4].items():  
        print('\t{}: {}'.format(key, value))  
  
# Логарифмирование столбца 'Close'  
df['Log_Close'] = np.log(df['Close'])  
  
rolling_mean = df['Log_Close'].rolling(window=12).mean()  
df_log_minus_mean = df['Log_Close'] - rolling_mean  
df_log_minus_mean.dropna(inplace=True)  
get_stationarity(df_log_minus_mean)
```

Результат:

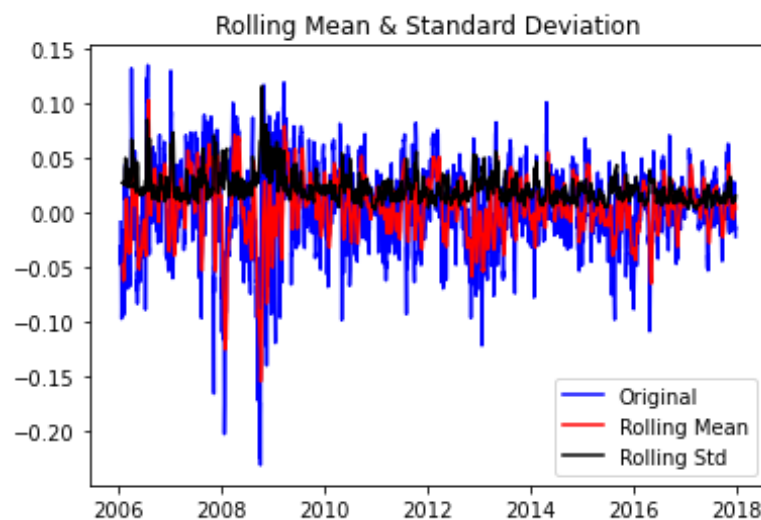


Рис. 3.2.3 - Змінне середнє та стандартне відхилення, визначене при відніманні середнього змінного

```
ADF Statistic: -11.104883401211405
p-value: 3.784577523395249e-20
Critical Values:
    1%: -3.4325382049645357
    5%: -2.8625068071065787
   10%: -2.5672846570213257
```

Як ми бачимо, після віднімання середнього, середнє ковзне та стандартне відхилення є приблизно горизонтальне. Р-значення нижче порогового значення 0.05, а ADF близьке до критичних значень. Отже, часовий ряд є стаціонарним.

Застосування експоненціального спаду є ще одним способом трансформації часового ряду таким чином, щоб він був стаціонарним.

```
def get_stationarity(timeseries):
    # Rolling statistics
    rolling_mean = timeseries.rolling(window=12).mean()
    rolling_std = timeseries.rolling(window=12).std()

    # Rolling statistics plot
    plt.figure(figsize=(10, 6))
    plt.plot(timeseries, color='blue', label='Original')
    plt.plot(rolling_mean, color='red', label='Rolling Mean')
    plt.plot(rolling_std, color='black', label='Rolling Std')
    plt.legend(loc='best')
    plt.title('Rolling Mean & Standard Deviation')
    plt.show()

    # Dickey-Fuller test
    result = adfuller(timeseries)
    print('ADF Statistic: {}'.format(result[0]))
    print('p-value: {}'.format(result[1]))
    print('Critical Values:')
    for key, value in result[4].items():
        print('\t{}: {}'.format(key, value))

# Логарифмування стовпця 'Close'
df['Log_Close'] = np.log(df['Close'])

# Обчислення експоненційно зваженого середнього та різниці
rolling_mean_exp_decay = df['Log_Close'].ewm(halflife=12,
min_periods=0, adjust=True).mean()
df_log_exp_decay = df['Log_Close'] - rolling_mean_exp_decay
df_log_exp_decay.dropna(inplace=True)

# Застосування функції для отримання стаціонарності
get_stationarity(df_log_exp_decay)
```

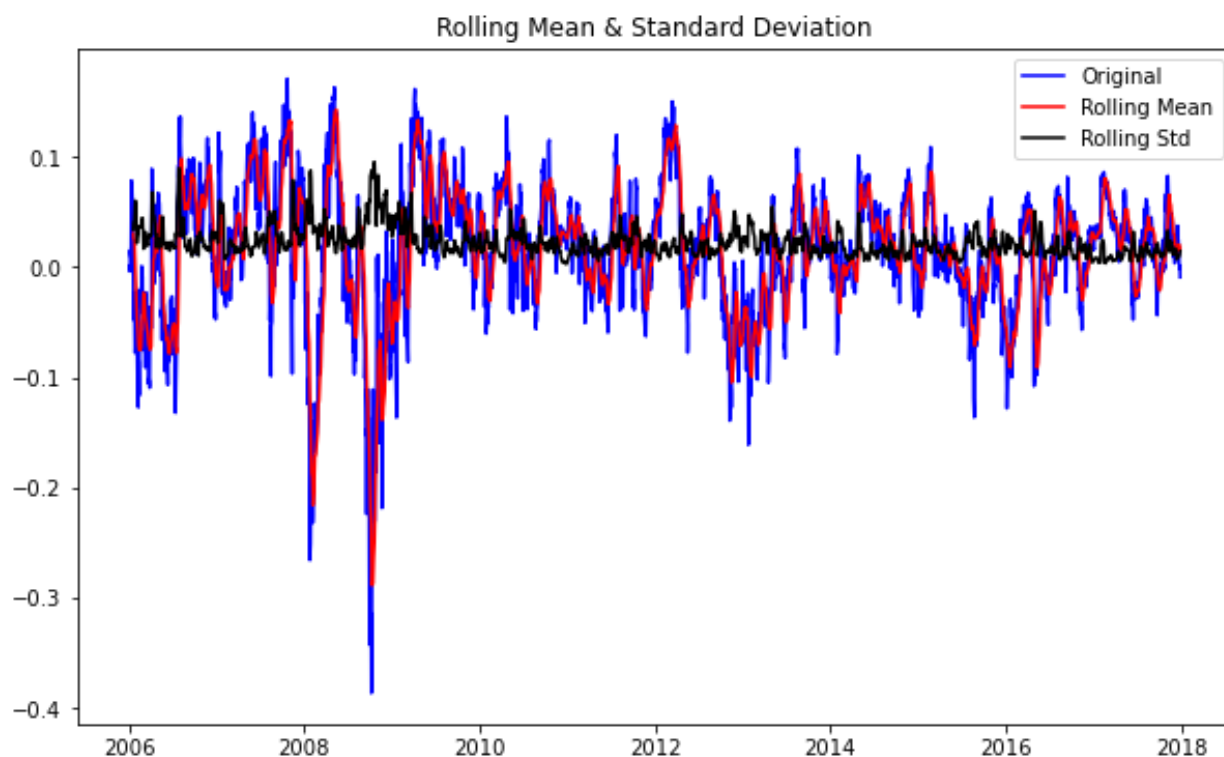


Рис. 3.2.4 - Змінне середнє та стандартне відхилення, знайдене способом експоненціального спаду

```
ADF Statistic: -8.027970862437623
p-value: 1.993534878262986e-12
Critical Values:
  1%: -3.43252293552353
  5%: -2.862500063122179
 10%: -2.5672810666012356
```

Експоненціальне затухання працює гірше, ніж віднімання середнього змінного. Однак він все ще більш стаціонарний, ніж оригінал.

Спробуємо ще один метод, щоб визначити чи існує ще краще рішення. Застосовуючи зсув часу, ми віднімаємо кожну точку на ту, яка їй передувала.

$$null, (x_1 - x_0), (x_2 - x_1), (x_3 - x_2), (x_4 - x_3), \dots, (x_n - x_{n-1})$$

```
# Обчислення різниці між логарифмом цін закриття та його зсувом
df_log_shift = df['Log_Close'] - df['Log_Close'].shift()
df_log_shift.dropna(inplace=True)
```

```
# Функція для перевірки стаціонарності часового ряду
def get_stationarity(timeseries):
    # Обчислення ковзного середнього та стандартного відхилення
    rolling_mean = timeseries.rolling(window=12).mean()
    rolling_std = timeseries.rolling(window=12).std()

    # Побудова графіка
```



```

plt.plot(timeseries.index, timeseries, color='blue',
label='Original')
plt.plot(timeseries.index, rolling_mean, color='red',
label='Rolling Mean')
plt.plot(timeseries.index, rolling_std, color='black',
label='Rolling Std')

# Додавання легенди та назви графіка
plt.legend(loc='best')
plt.title('Rolling Mean & Rolling Standard Deviation')

# Відображення графіка
plt.show()

# Перевірка стаціонарності за допомогою тесту Дікі-Фуллера
result = adfuller(timeseries)
print('ADF Statistic: {}'.format(result[0]))
print('p-value: {}'.format(result[1]))
print('Critical Values:')
for key, value in result[4].items():
    print('\t{}: {}'.format(key, value))

# Перевірка стаціонарності ряду різниці логарифмів цін закриття
get_stationarity(df_log_shift)

```

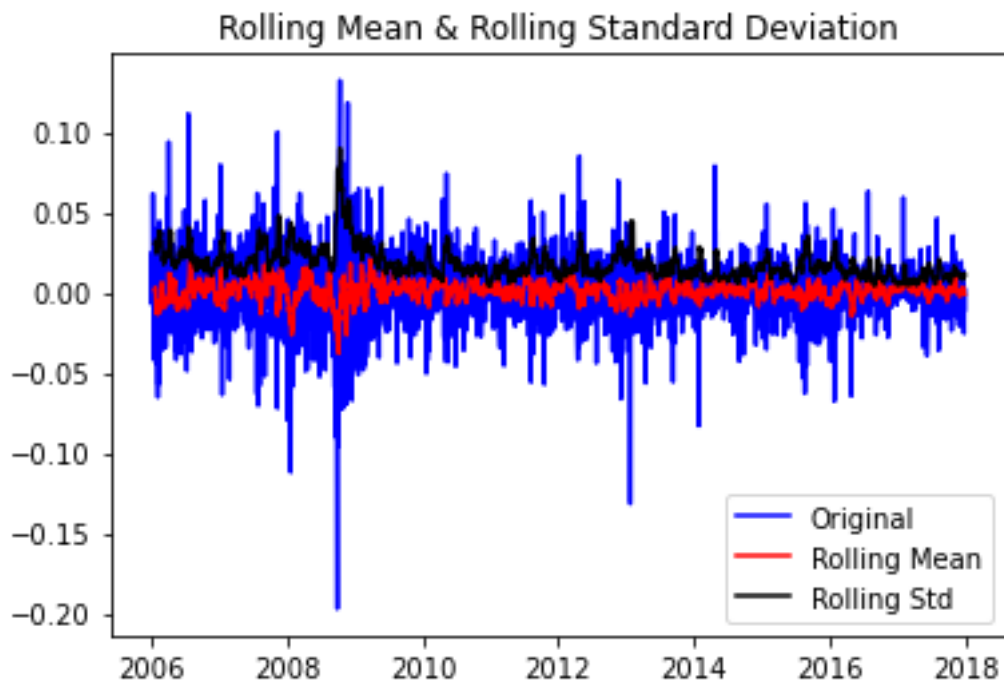


Рис. 3.2.5 - Середнє змінне та стандартне відхилення, знайдене методом зсуву часу

```

ADF Statistic: -22.499397686501833
p-value: 0.0
Critical Values:
1%: -3.43252293552353

```

```
5%: -2.862500063122179
10%: -2.5672810666012356
```

У даному випадку, p-value дорівнює 0.0, що означає, що є дуже маленька ймовірність отримати такі або більш екстремальні значення статистики при наявності нестационарності. Це підтверджується і прийняттям альтернативної гіпотези про стаціонарність ряду.

Краще, коли значення p дорівнює нулю. Таке значення p надає більш сильні докази на користь стаціонарності ряду і підтверджує, що часовий ряд може бути використаний для застосування математичних моделей, таких як ARIMA, для аналізу та прогнозування. Таким чином, коли значення p дорівнює нулю, це є більш підтверджуючим і сприяє більшій впевненості в стаціонарності ряду.

3.3 Побудова моделі ARIMA та прогнозування часового ряду

Ми можемо створити та адаптувати модель ARIMA з AR (авторегресивна модель) порядку 2, диференціацією порядку 1 та MA (модель ковзного середнього) порядку 2.

```
# Перетворення даних
df = data['Close']
df_log = np.log(df)
df_log_shift = df_log - df_log.shift(1)

# Заповнення пропущених значень нулями
df_log_shift = df_log_shift.fillna(0)

# Застосування seasonal_decompose до часового ряду
decomposition = seasonal_decompose(df_log_shift, period=252)

# Модель ARIMA
model = ARIMA(df_log_shift, order=(2, 1, 2))
results = model.fit()

# Графіки
plt.plot(df_log_shift)
plt.plot(results.fittedvalues, color='red')
plt.show()
```

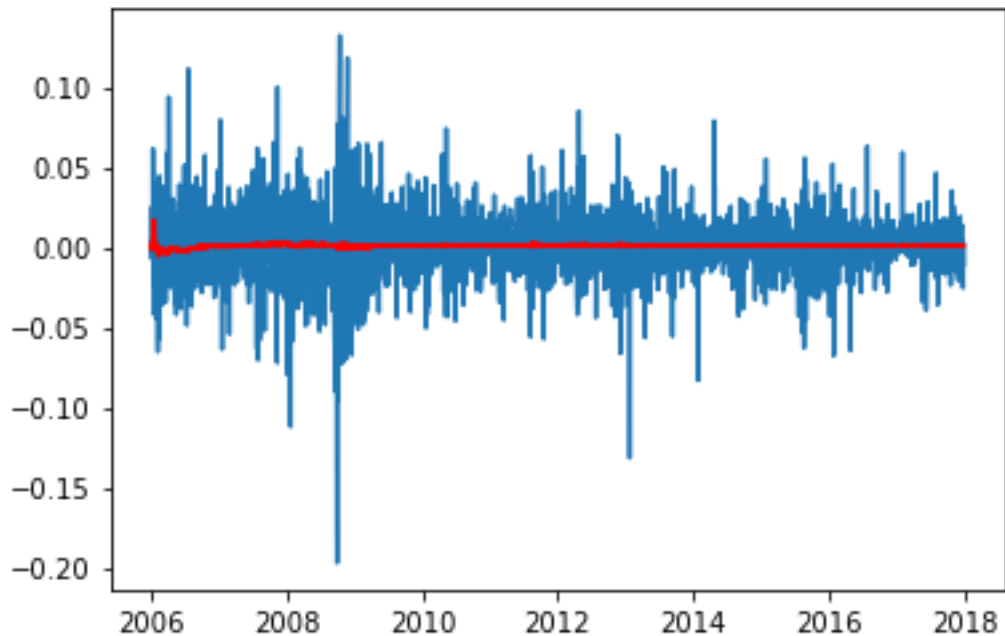


Рис. 3.3.1 - Модель Агіма на графіку

Потім ми можемо побачити, як модель порівнюється з оригінальним часовим рядом.

```

predictions_ARIMA_diff = pd.Series(results.fittedvalues,
copy=True)
predictions_ARIMA_diff_cumsum = predictions_ARIMA_diff.cumsum()
predictions_ARIMA_log = pd.Series(df_log.iloc[0],
index=df_log.index)
predictions_ARIMA_log =
predictions_ARIMA_log.add(predictions_ARIMA_diff_cumsum,
fill_value=0)
predictions_ARIMA = np.exp(predictions_ARIMA_log)
plt.plot(df)
plt.plot(predictions_ARIMA)

```



Рис. 3.3.2 - Модель Агіма у порівнянні з оригінальним часовим рядом

Прогноз на 5 років (1825 днів):

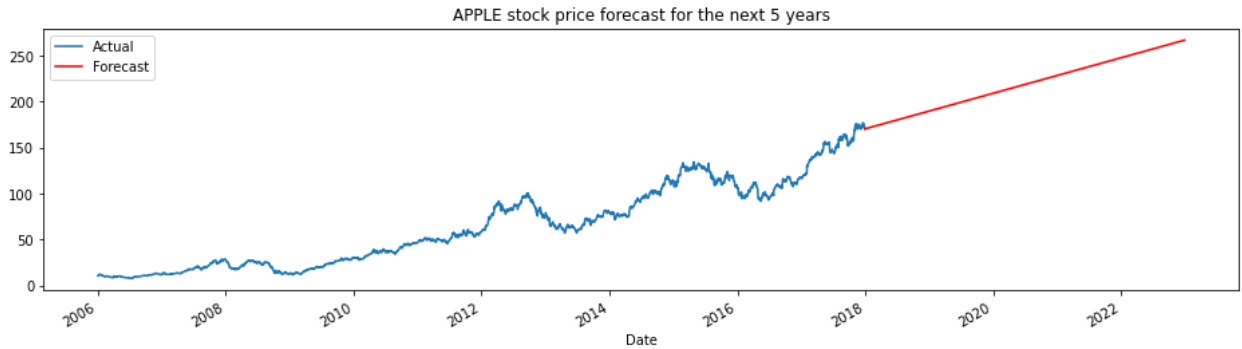


Рис. 3.3.3 – прогноз ціни акції Apple на 5 років

На прикладі реального графіку котировки акцій Apple за останні 10 років, ми бачимо, що ціни акцій дійсно росли, як і у нашому графіку прогнозів, які ми отримали за допомогою моделі ARIMA.



Рис. 4 – графік акцій компанії Apple за останні 10 років

Також хочу зазначити, що на прогнозі максимальна ціна акцій сягає 250 у.о, але у реальному житті це 177.25 у.о. Це пов'язано з тим, що модель ARIMA не може враховувати у прогнозах такі важливі для економіки події, як карантин, криза та війна, що відбулись у світі за останні 5 років.

3.4 Порівняння спрогнозованих та реальних значень цін акцій

1) Amazon

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Apple.



Рис. 3.4.1 – прогноз цін акцій Amazon на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.2)



Рис 3.4.2 – графік цін акцій Amazone за останні 10 років

Порівнюючи ці два зображення ми бачимо, що до середини 2021 року акції компанії дійсно зростали, але потім різко почала падати. У липні 2021 року ринкова вартість Amazon становила майже 1,9 трильйона доларів, з початку року статки співзасновника Amazon Джеффа Безоса скоротилися приблизно на 83 мільярди доларів до 109 мільярдів доларів.

Найбільший у світі онлайн-магазин провів цей рік, пристосовуючись до різкого уповільнення росту електронної комерції, оскільки покупці відновили звички, які були до пандемії. Її акції впали майже на 50% через падіння продажів, стрімке зростання витрат і стрибок процентних ставок.

2) Cisco

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Cisco.



Рис 3.4.3 – прогноз цін акцій Cisco на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.4)

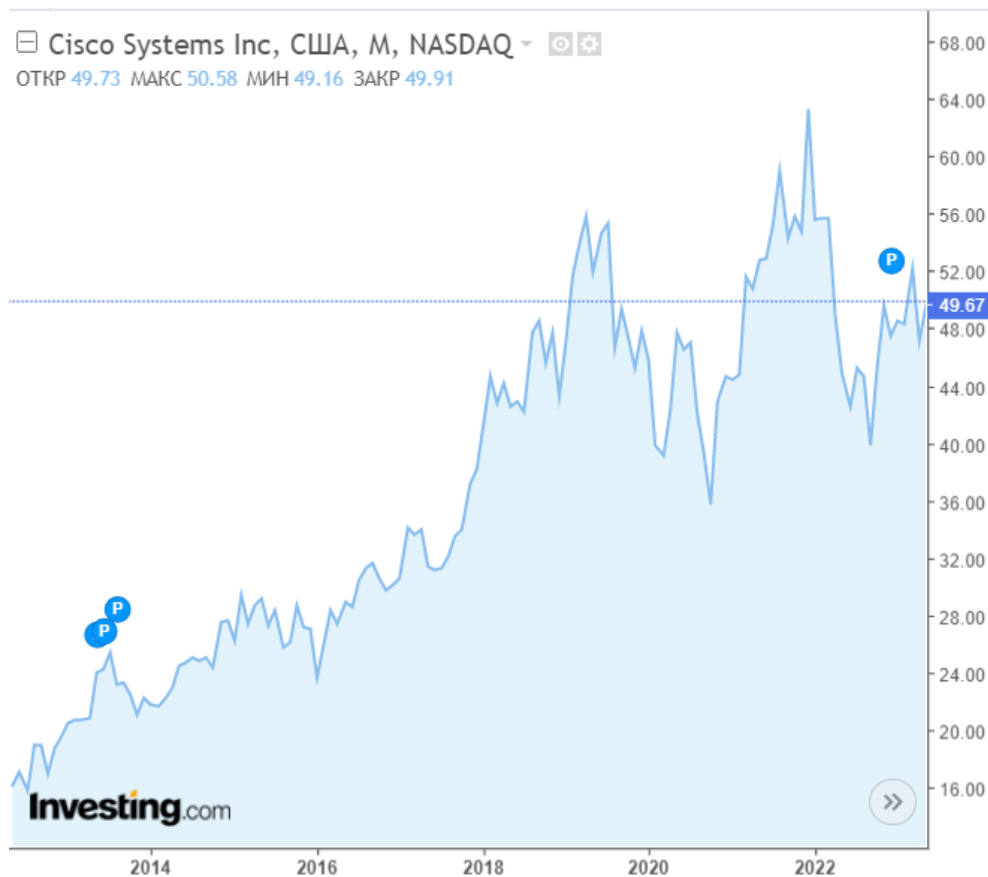


Рис. 3.4.4 – графік цін акцій Cisco за останні 10 років

Cisco Systems є прикладом компанії, яка відображає концепцію «магніту дивіденду». Це означає, що з часом підвищення ціна акцій майже повністю компенсують зростання дивідендів.

Однак у короткостроковій перспективі цей механізм не завжди спрацьовує, і це було особливо помітно в 2022 році для Cisco. Незважаючи на зростання дивідендів, акції компанії зазнали втрат у розмірі 22% протягом 2022 року.

Cisco потрапила в хвилю розпродажу технологічних компаній, що обґрунтовані підстави. Доходи компанії в першій половині минулого року сповільнилися, що стало причиною поглядів ринку на акції компанії.

3) Google

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Google.



Рис. 3.4.5 – прогноз цін акції компанії Google

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.6)

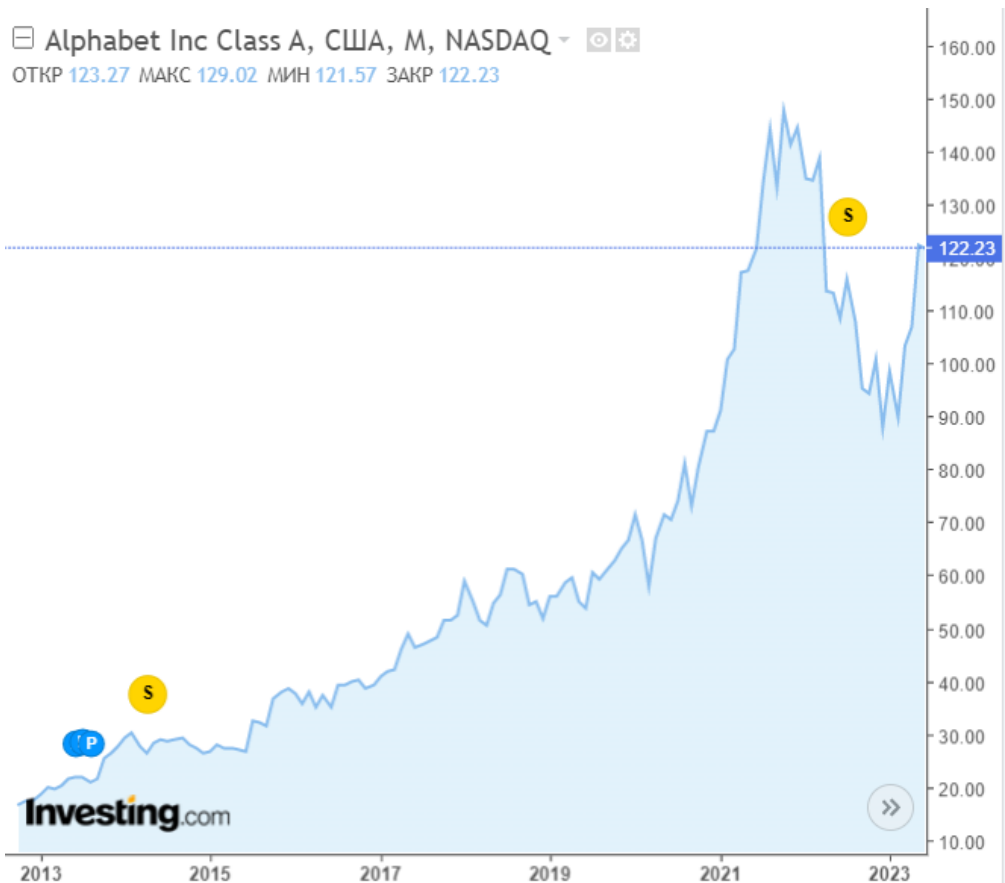


Рис 3.4.6 – графік цін акцій Google за останні 10 років

Ціна на акції дійсно зростала, але у 2022 компанія провела спліт (дроблення) акцій у співвідношенні 20:1, як це зробила компанія Amazon, щоб вони стали доступнішими для інвесторів, а їх ліквідність збільшилася.

4) Nike

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Nike.

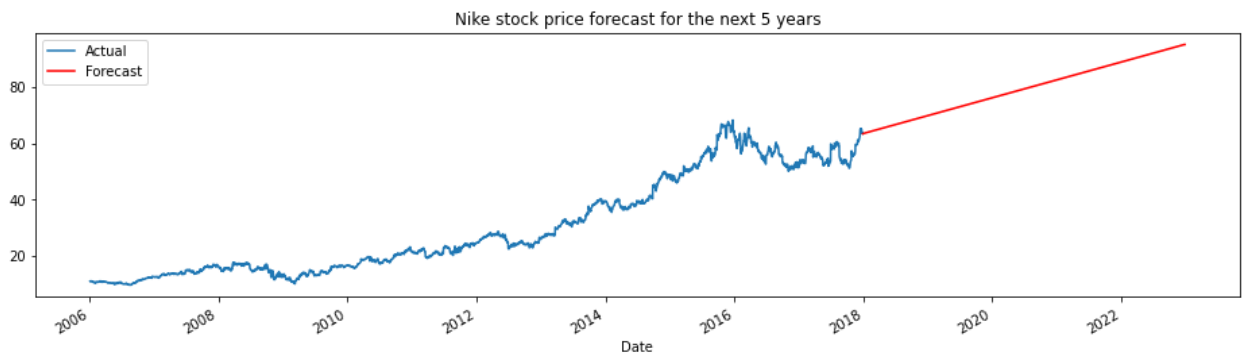


Рис. 3.4.7 – графік прогнозу цін акцій компанії Nike на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій
(рис.3.4.8)

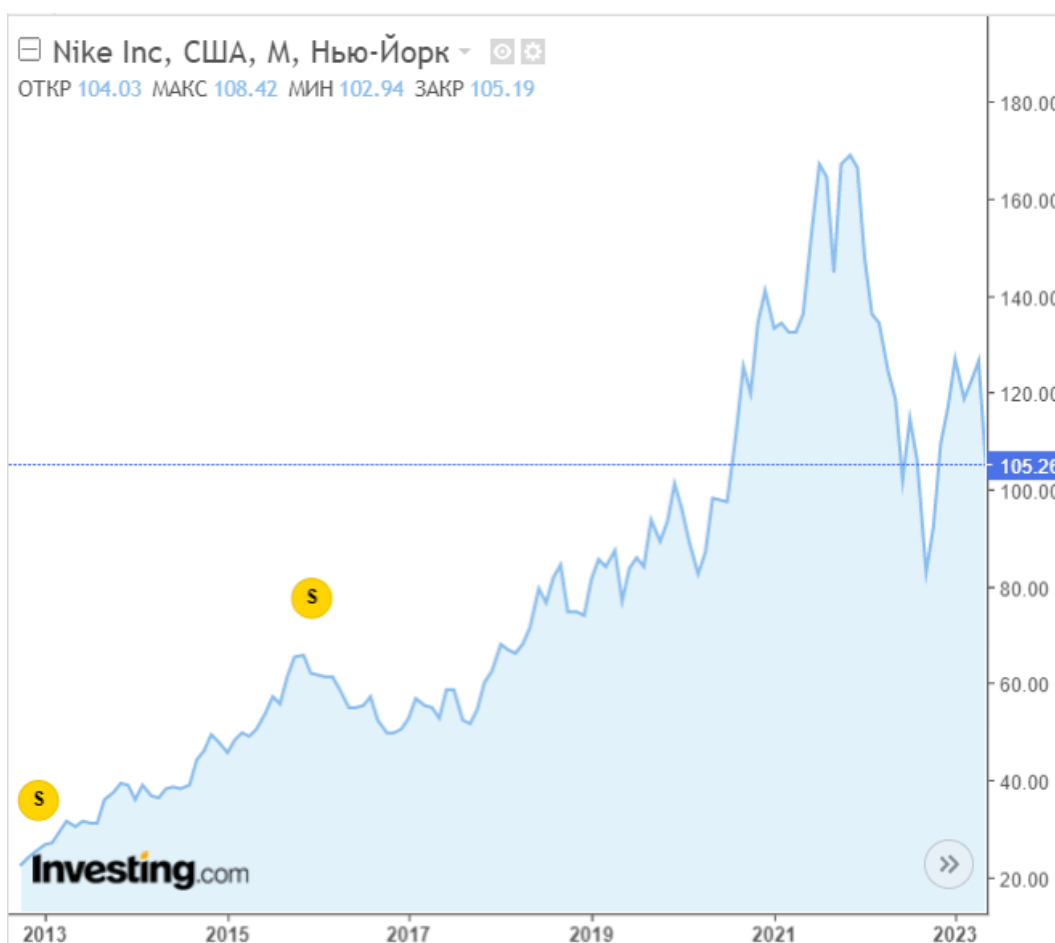


Рис 3.4.8 – графік цін акцій Nike за останні 10 років

У 2019 році сталося значне падіння цін на акції компанії Nike на фондовому ринку США. Це стало результатом неприємного інциденту, пов'язаного з кросівкою моделі PG 2.5, яку використовував молодий американський баскетболіст Сіон Вільямсон.

20 лютого 2019 року під час гри Сіон Вільямсон отримав травму правої ноги, після чого його кросівка розірвалася. Цей інцидент призвів до серйозного розтягнення зв'язок у спортсмена, що викликало значну хвилю обурення та розмов на ринку.

Внаслідок цього негативного інциденту, акції компанії Nike зазнали сильного зниження вартості, що мало значний вплив на її ринкову

капіталізацію. Загалом, ціни на акції Nike знизилися на 2%, що оцінюється приблизно в \$1,46 млрд.

Це падіння вартості акцій компанії Nike було негативним сигналом для інвесторів та ринкових учасників, але компанія зробила кроки для відновлення довіри та популярності своїх виробів у подальшому.

5) Pfizer

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Pfizer.



Рис. 3.4.9 – графік прогнозу цін на акції компанії Pfizer на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.10)

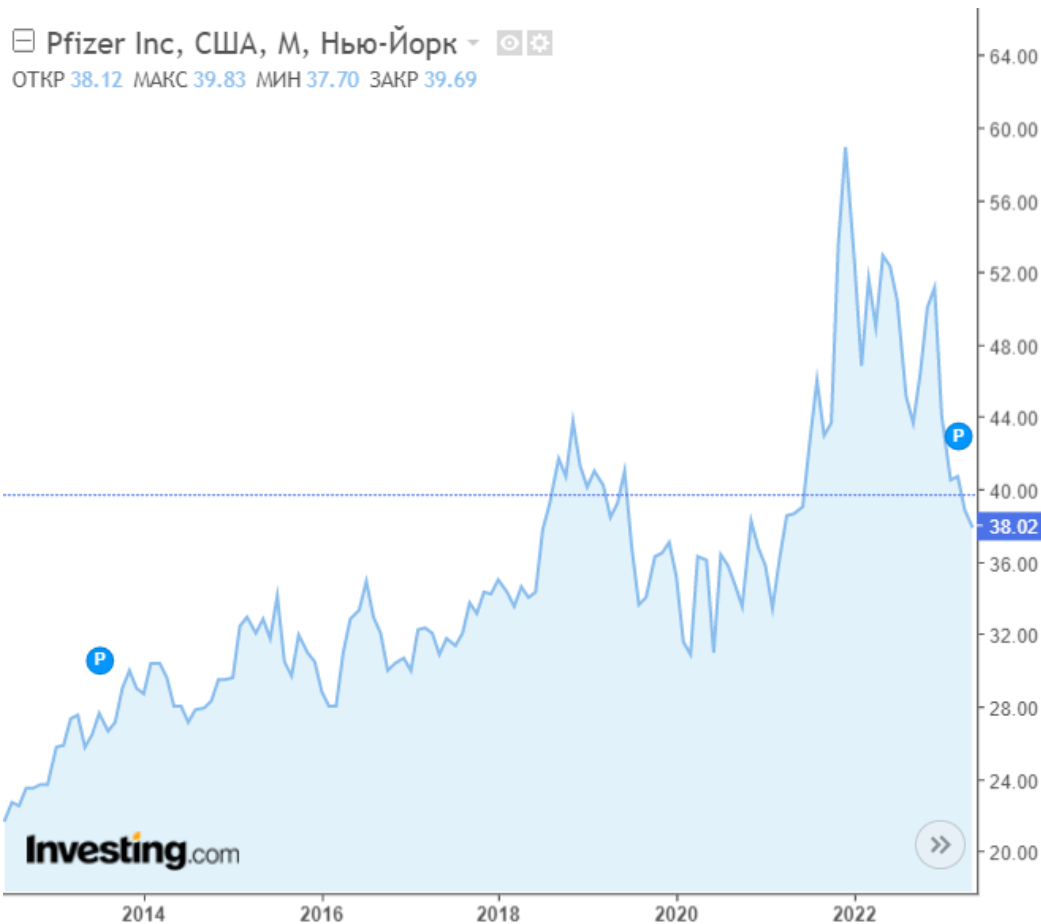


Рис. 3.4.10 – графік цін акцій Pfizer за останні 10 років

Прогноз на основі Arima моделі показував плавне зростання цін акцій компанії Pfizer. Але за графіком (рис. 3.4.10) ми чітко бачимо, що у 2021 році почалось дуже стрімке зростання. Це пов'язано з тим, що компанія почала продавати вакцини від Covid-19. У 2023 ціна починає падати, тому що карантин відмінений, а вакцинація перестала бути обов'язковою для того, щоб, наприклад, потрапити в іншу країну.

б) Procter&Gamble

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Procter&Gamble.

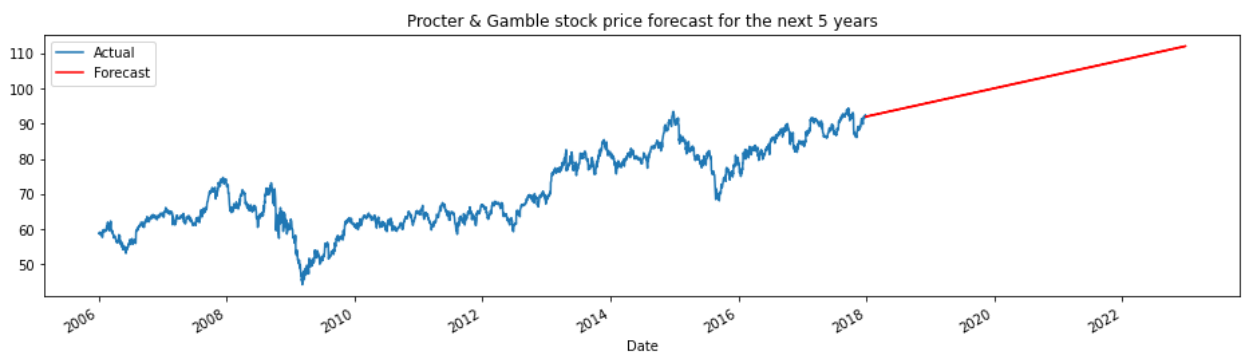


Рис. 3.4.11 – графік прогнозу цін акцій компанії Procter&Gamble на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.12)

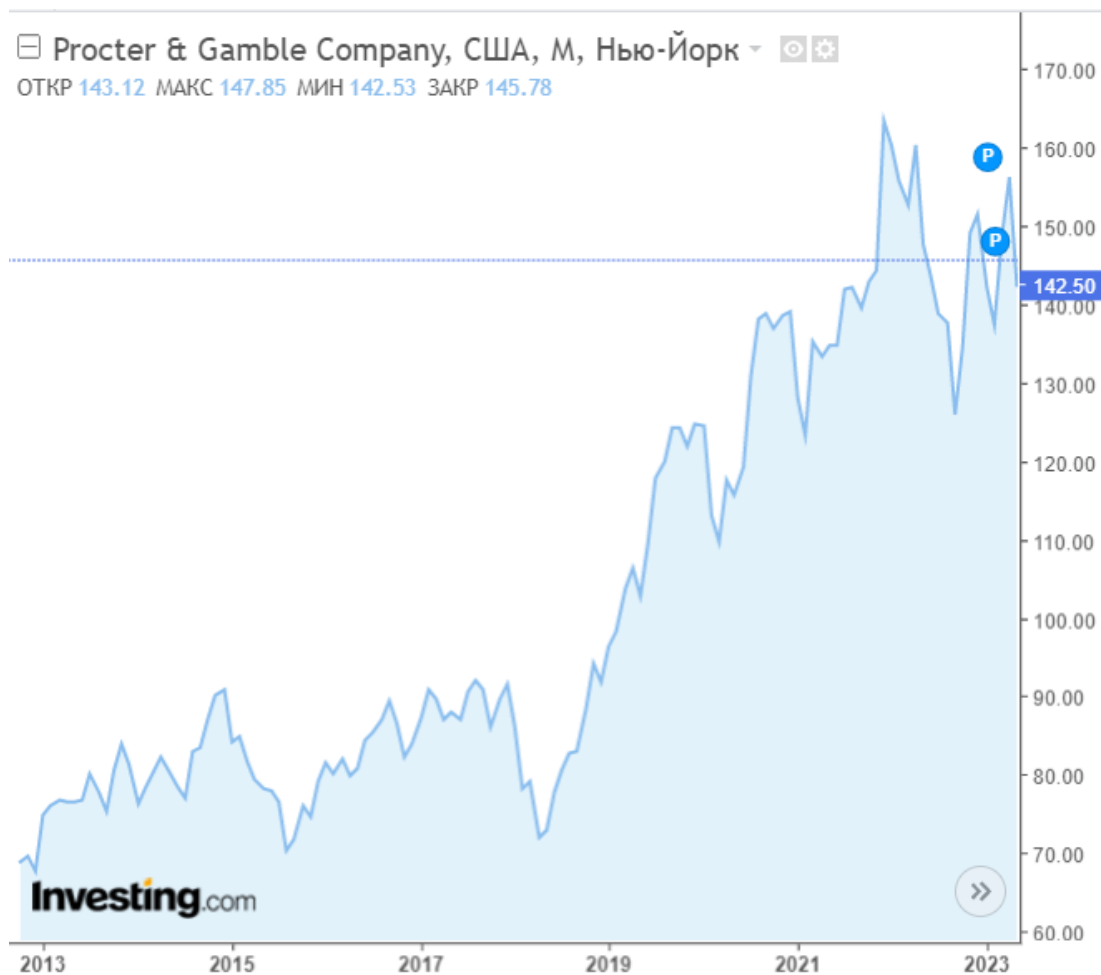


Рис. 3.4.12 – графік акцій компанії Procter&Gamble за останні 10 років

Як ми бачимо за першим графіком (рис. 3.4.11) компанія розвивалась достатньо рівномірно і прогноз на основі Arima моделі також показав не

стрімкий зріст. Але через пандемію в 2019 році акції компанії почали достатньо швидко зростати, адже люди почали проводити більше часу вдома, а їх напрямком є побутова хімія та засоби особистої гігієни, наприклад, антибактеріальне мило для рук.

7) UnitedHealth Group

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії UnitedHealth Group.

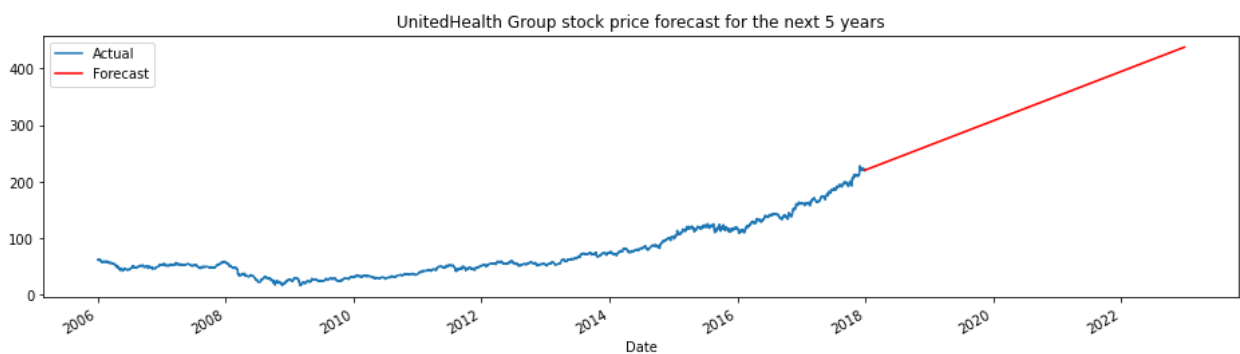


Рис. 3.4.13 – графік прогнозування цін акцій компанії UnitedHealth Group на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.14)

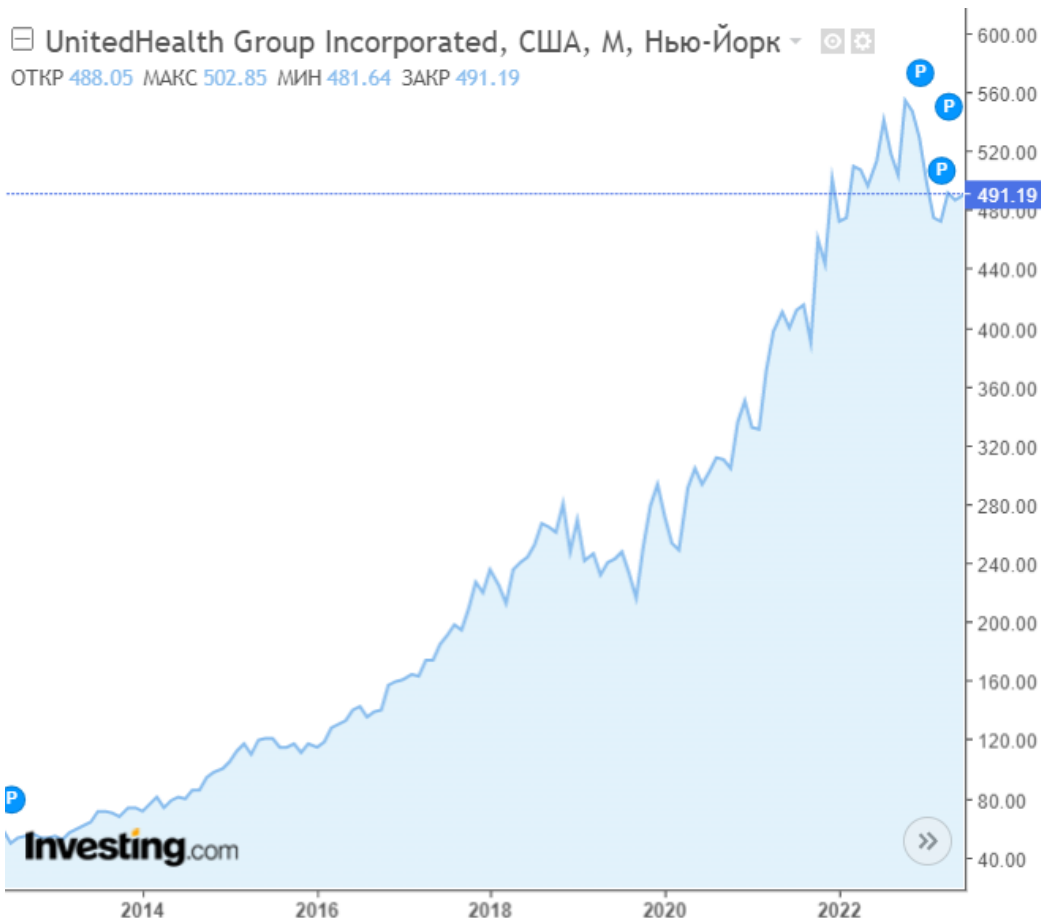


Рис. 3.4.14 – графік цін акцій компанії UnitedHealth Group за останні 10 років

На прикладі цієї компанії видно, що прогнозування на основі Arima моделі співпадає з дійсністю.

8) American Express

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії American Express.

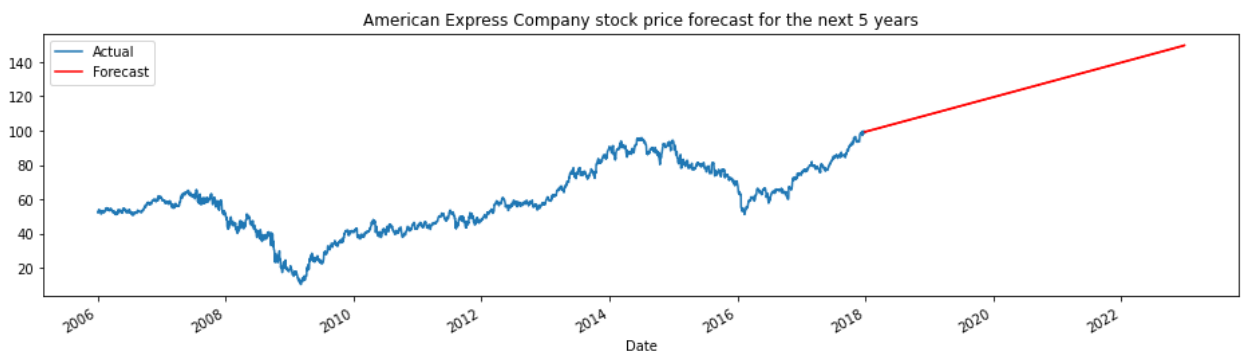


Рис. 3.4.15 – графік прогнозу акцій компанії American Express на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій
(рис.3.4.16)

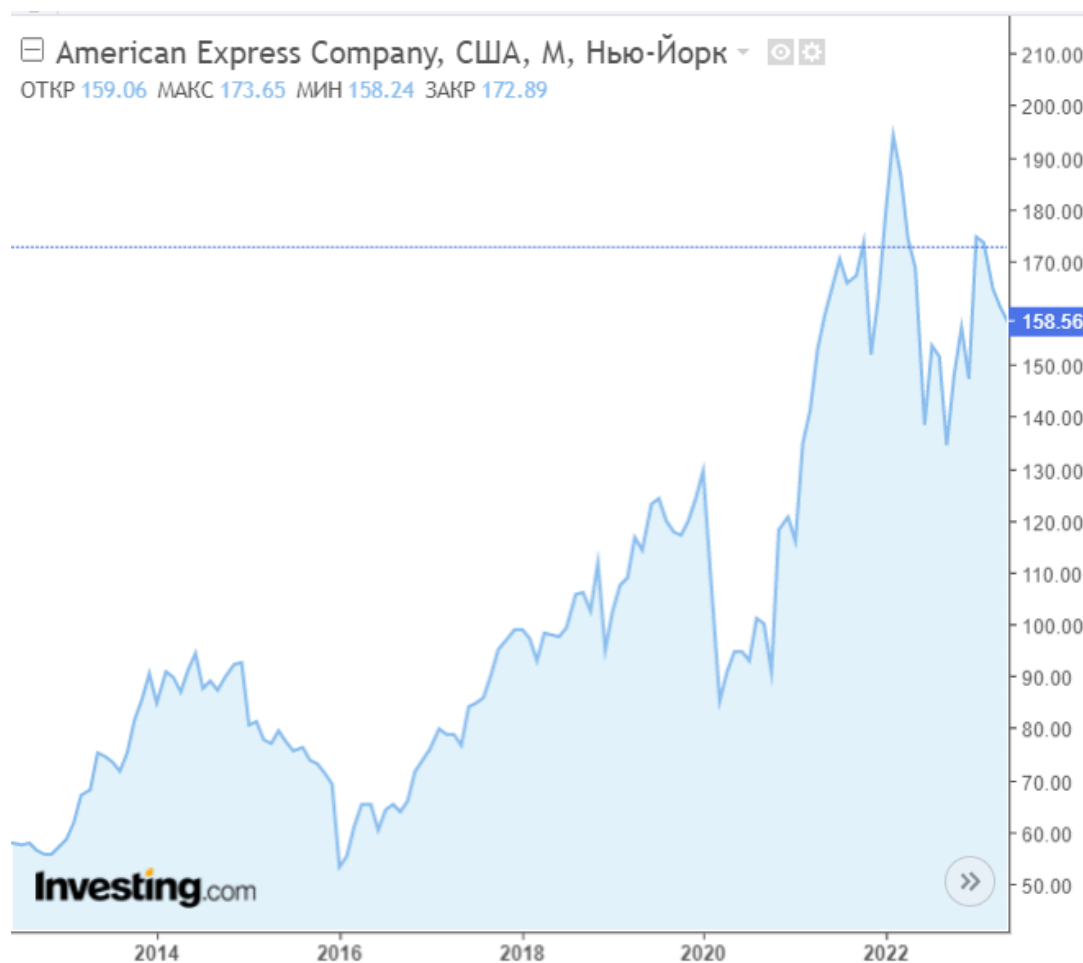


Рис. 3.4.16 – графік акцій компанії American Express за останні 10 років

Починаючи з 2018 року ціна на акції дійсно почала підійматися, але з початком Covid-19 ситуація змінилась і акції стрімко почали знижуватись у ціни, тому компанія несе за собою транспортну функцію, а більшість люде були вдома на карантині. У 2021 році люди почали прививатись від цього вірусу і продовжили подорожувати з сертифікатами.

9) Boeing

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Boeing.

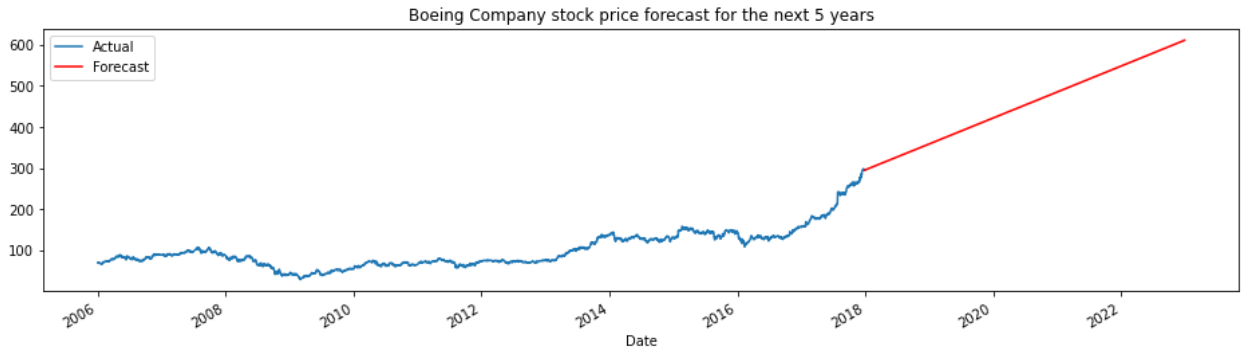


Рис. 3.4.17 – графік прогнозу цін акцій компанії Boeing

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.18)



Рис. 3.4.18 – графік цін на акції компанії Boeing за останні 10 років
 З 2017 року ціни на акції стали активно збільшуватись, але, як і в багатьох прикладах раніше, з початком пандемії різко впали і більше не піднялись до тих показників.

10) Caterpillar

Я побудувала прогнозування на основі моделі ARIMA за шаблоном, наведеним вище з котировками акцій компанії Caterpillar.

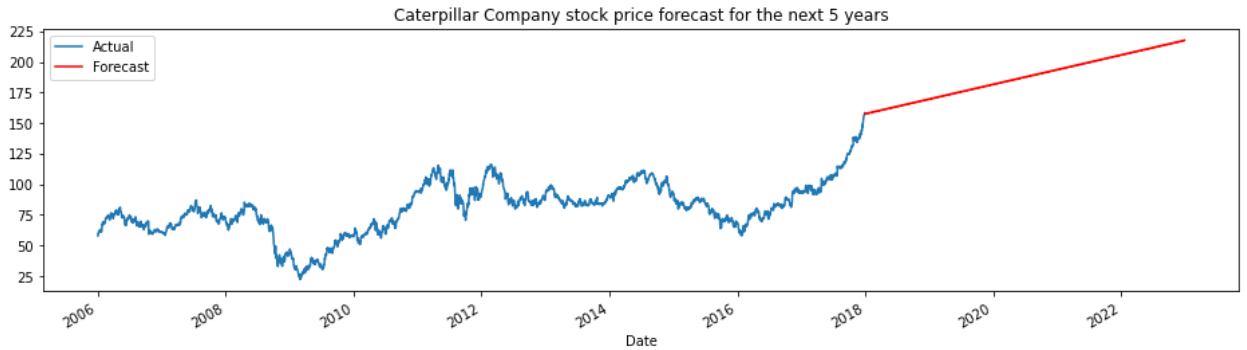


Рис. 3.4.19 – графік прогнозування цін акцій компанії Caterpillar на 5 років

Порівняємо графік прогнозування та реальний графік котировок акцій (рис.3.4.20)

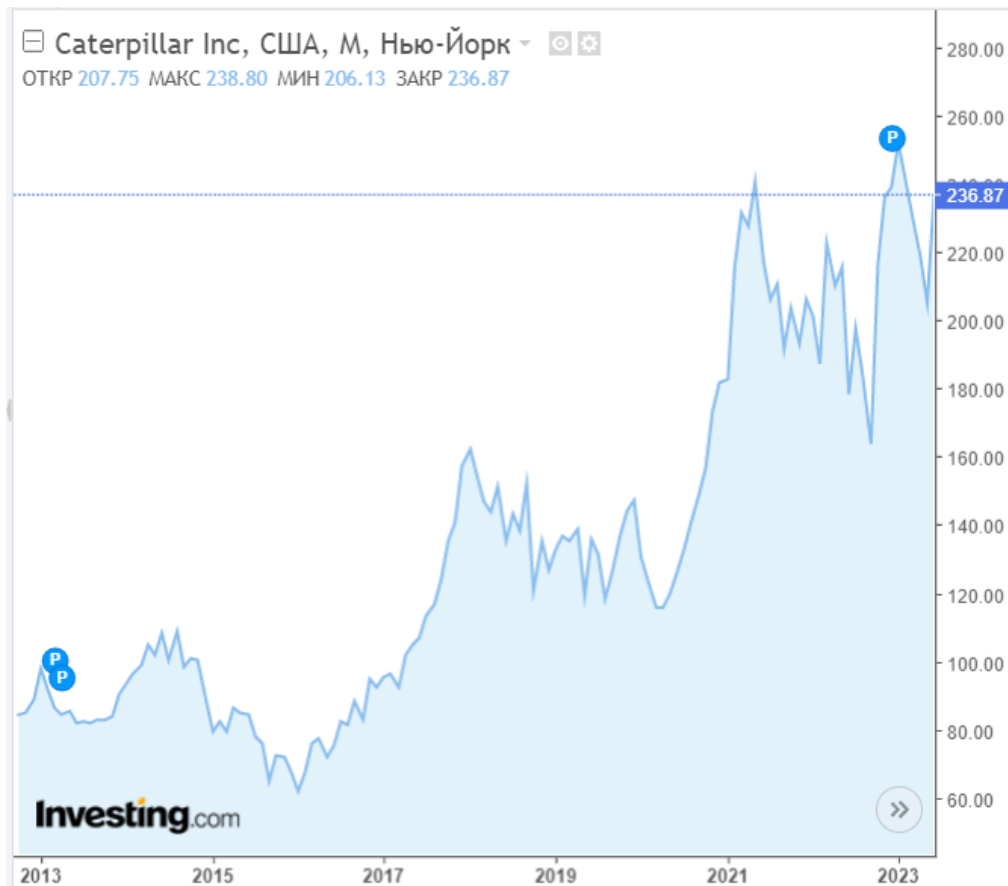


Рис. 3.4.20 – графік цін акцій компанії Caterpillar за останні 10 років

Як і з прикладом компанії UnitedHealth Group графік прогнозування майже співпав з дійсністю. Також на другому графіку (рис. 3.4.20) чітко видно падіння цін у період пандемії (2020 рік), тому, можливо, як би цього фактору не було, то на період 2022-2023 років акції коштували б набагато дорожче, ніж я спрогнозувала за допомогою Arima моделі.

3.5. Індекс Доу-Джонса

Індекс Доу-Джонса є одним з найвідоміших фінансових індексів, який відображає загальну динаміку ринку акцій в Сполучених Штатах. Вперше розрахований у 1896 році, він складається з 30 великих компаній, які відображають широкий спектр економічних секторів. 10 із цих 30 компаній розглянула у пункті 3.4.

Основна формула для розрахунку індекса Доу-Джонса включає такі кроки:

1. Обчисліть суму цін акцій кожної компанії, що входить до складу індексу.
2. Поділіть отриману суму на дільник, який називають «дивізором».
3. Результат цього поділу є індексом Доу-Джонса

Формально, формула виглядає таким чином:

$$\text{Індекс Доу-Джонса} = (\text{Сума цін акцій}) / (\text{Дивізор})$$

Дивізор – константа, яка залежить від структури складу індексу та використовується для нормалізації значень значень індексу відносно попередніх значень.

Для прогнозування виконаємо наступний код:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.arima.model import ARIMA

# Завантаження даних з файлу CSV
data = pd.read_csv('all_stocks_2017-01-01_to_2018-01-01.csv')

# Перетворення стовпця з датами у тип datetime
```

```

data['Date'] = pd.to_datetime(data['Date'])

# Обчислення індексу Доу-Джонса на основі цін закриття всіх акцій
dow_jones_index = data.groupby('Date')['Close'].mean()

# Декомпозиція сезонності
decomposition = seasonal_decompose(dow_jones_index, period=30)

# Побудова графіку розкладу
plt.figure(figsize=(12, 8))
plt.subplot(411)
plt.plot(dow_jones_index, label='Original')
plt.legend(loc='best')
plt.subplot(412)
plt.plot(decomposition.trend, label='Trend')
plt.legend(loc='best')
plt.subplot(413)
plt.plot(decomposition.seasonal, label='Seasonality')
plt.legend(loc='best')
plt.subplot(414)
plt.plot(decomposition.resid, label='Residuals')
plt.legend(loc='best')
plt.tight_layout()
plt.show()

```

Результат:

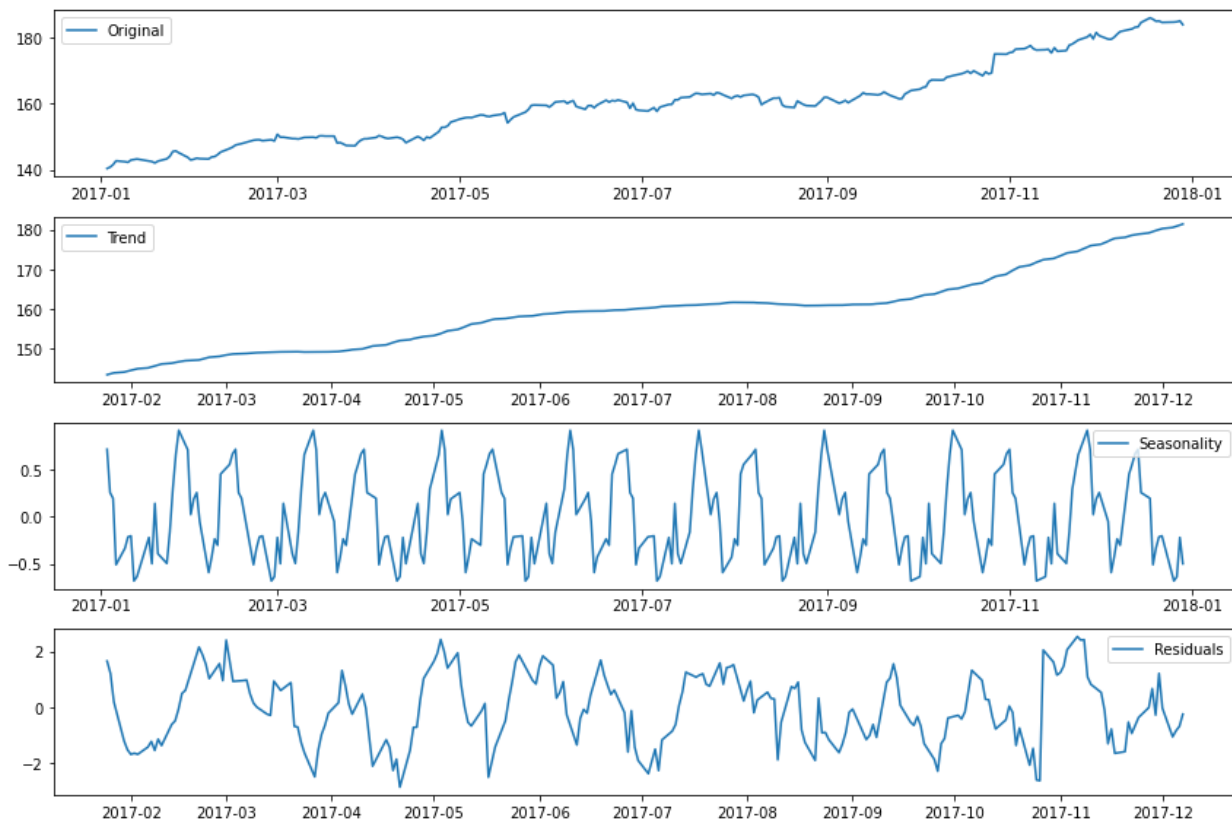


Рис 3.5.1 – початковий ряд, тренд, сезонність та залишки індексу Доу-Джонса

Наданий код виконує наступні дії:

1. Завантажує дані з файлу CSV, який містить інформацію про ціни закриття акцій на ринку з 2017-01-01 до 2018-01-01.
2. Перетворює стовпець з датами на об'єкти типу `datetime` для подальшої обробки.
3. Обчислює середнє значення цін закриття всіх акцій на кожен дату, створюючи індекс Доу-Джонса.
4. Виконує декомпозицію часового ряду індексу Доу-Джонса на компоненти: тренд, сезонність та залишкові компоненти.
5. Будує графіки кожної компоненти (початковий ряд, тренд, сезонність та залишки) в окремих підграфіках.

Отримані графіки демонструють, які компоненти сприяють змінам індексу Доу-Джонса протягом аналізованого періоду. Зокрема, графік "Original" показує початковий ряд, "Trend" відображає довгостроковий тренд, "Seasonality" представляє сезонну залежність, а "Residuals" показує залишкові компоненти, які не вдаються пояснити трендом або сезонністю.

Цей аналіз може допомогти виявити закономірності та тенденції в поведінці індексу Доу-Джонса, що можуть бути корисними для подальшого прогнозування та прийняття рішень на фінансовому ринку.

Для побудови графіку з прогнозом індексу Доу-Джонса на 2 місяці я використала наступний код:

```
# Побудова індексу Доу-Джонса на основі цін закриття всіх акцій
dow_jones_index = data.groupby('Date')['Close'].sum()

# Побудова ARIMA моделі
model = ARIMA(dow_jones_index, order=(2, 1, 2))
results = model.fit()

# Прогнозування індексу Доу-Джонса
forecast = results.predict(start=len(dow_jones_index),
end=len(dow_jones_index)+60)

# Перетворення індексу на тип дати
dow_jones_index.index = pd.to_datetime(dow_jones_index.index)
```

```

# Отримання дат для прогнозу
forecast_dates = pd.date_range(start=dow_jones_index.index[-1] +
pd.DateOffset(days=1), periods=61, freq='D')

# Побудова графіку індексу Доу-Джонса та прогнозу
plt.plot(dow_jones_index, label='Actual')
plt.plot(forecast_dates, forecast, label='Forecast')
plt.title('Dow Jones Industrial Average')
plt.xlabel('Date')
plt.ylabel('Index Value')
plt.legend()
plt.show()

```

Результат:

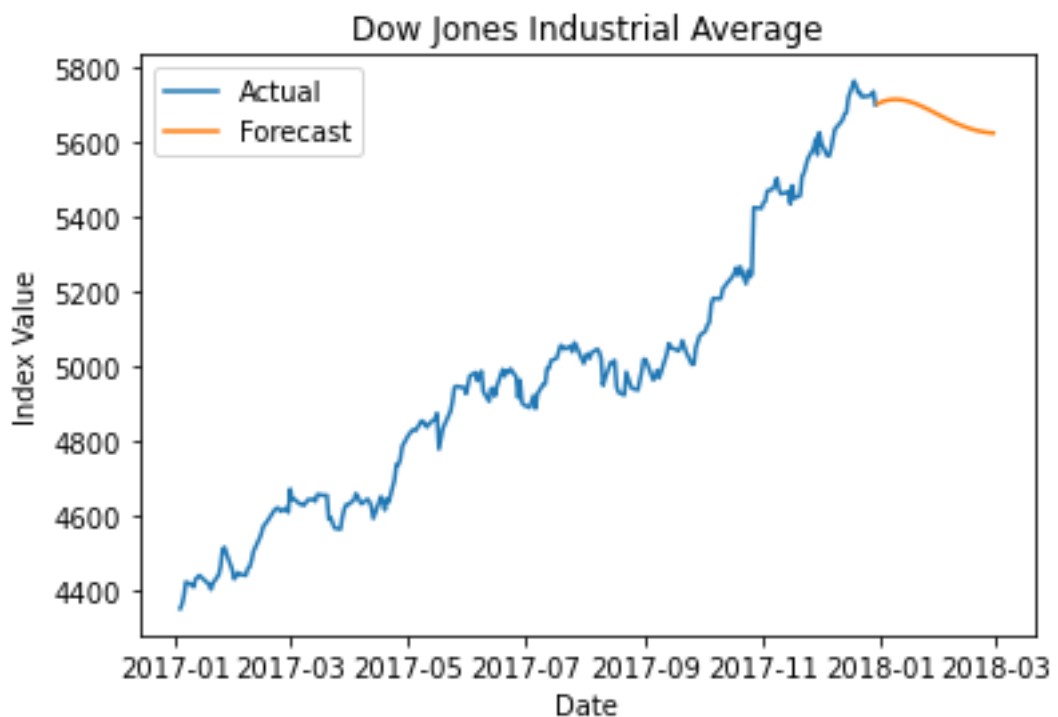


Рис. 3.5.2 – графік прогнозу індексу Доу-Джонса на основі Arima моделі

Порівняємо графік прогнозу з дійсністю:

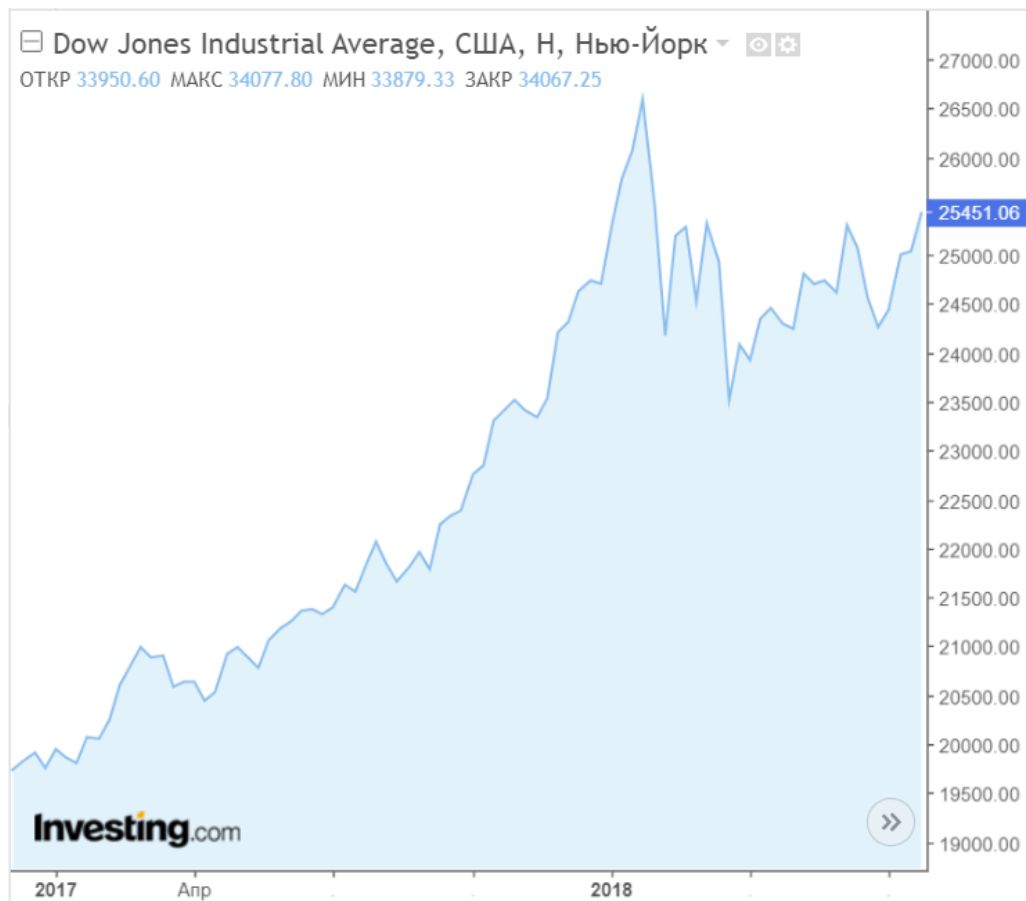


Рис. 3.5.3 – графік індексу Доу-Джонса за період 2017-2018 роки

Отже, за другим графіком (рис. 3.5.3) видно, що на початку 2018 року індекс починає падати, як і на графіку прогнозу за допомогою Arima моделі (рис. 3.5.2).

У пункті 3.4 я розглядала 10 із 30 провідних компаній США, які входять до індексу Доу-Джонса. У всіх цих прикладах ціни на акції компаній зростають. Тож, індекс Доу-Джонса за цей період також зростає (рис. 3.5.4):

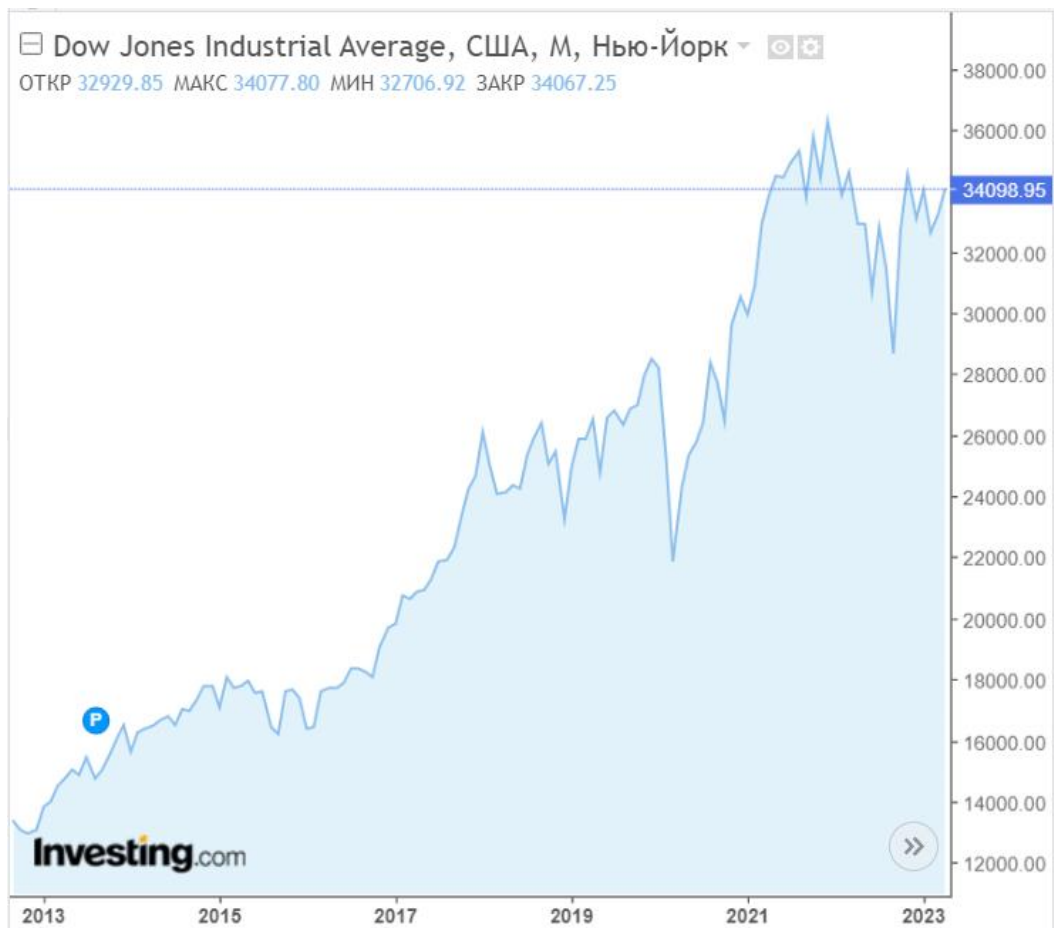


Рис. 3.5.4 – індекс Доу-Джонс за останні 10 років.

Висновки

У цій роботі я отримала прогноз на основі моделі Arima для 10-ти компаній із списку 30-ти провідних компаній США. Було використано дані за 12 років, на основі яких було побудовано прогноз цін на акції на 5 років. В процесі порівняння спрогнозованих графіків із котуваннями минулих років, було з'ясовано, що прогноз частково виконався. Деякі компанії коштують дорожче, ніж було спрогнозовано, інші дешевше.

Майже на кожному з цих прикладів ми бачимо стрімке падіння цін акцій на початку 2020 року. Це пов'язано з тим, що у США (і всьому світі в цілому) почався карантин Covid-19. У цей час люди припинили подорожувати, відвідувати людні місця, а також виходити з дому. Звісно, це дуже вплинуло на малий бізнес, а також і на великі компанії. Цей приклад того, що модель Arima не може враховувати такого типу події.

Також я проаналізувала та спрогнозувала індекс Доу-Джонса. Це виявилось трохи важче, тому що обсяг даних був дуже великий і, нажаль, мій комп'ютер не зміг виконати обчислення для прогнозу на 5 років, або хоча б рік. Я спрогнозувала дані на 2 місяці після оригінального часового ряду. Цим результатом я була також задоволена, тому що результат був схожий із дійсністю. Дійсно, за період з 2018 по 2022 роки індекс Доу-Джонса виріс, якщо дивитись у маленькому масштабі, але в саме цей період, який я прогнозувала (кінець 2018 року) спостерігалось падіння, що і показало мені на графіку прогнозу.

Список використаних джерел

1. Закон України № 1201-ХІІ від 17 червня 1991 р. Про цінні папери і фондову біржу / Розділ 1 Цінні папери, порядок їх випуску та обігу / Глава 1. Загальні положення.
2. Словник фінансово-правових термінів / за заг. ред. д. ю.н., проф. Л. К. Воронової. — 2-е вид., переробл. і доповн. — К.: Алерта, 2011—558 с.
3. Державне регулювання інвестиційного процесу в умовах функціонування ринку цінних паперів: монографія / П. О. Куцик, М. Я. Вірт, І. В. Гончарук та ін. ; [за заг. ред. П. О. Куцика] ; Львів. комерц. акад. — Львів: Растр-7, 2014. — 296 с. : іл. — Бібліогр.: с. 236—254 (265 назв).
4. І. Рогач, О. Ю. Анісімова. Валютна ліквідність // Українська дипломатична енциклопедія: У 2-х т./Редкол.:Л. В. Губерський (голова) та ін. — К: Знання України, 2004 — Т.1 — 760с.
5. Cowpertwait P.S.P., Metcalfe A.V. (2009), *Introductory Time Series with R*, Springer
6. «Прогнозування та аналіз часових рядів»/Укл.:Юрченко М.Є. — Чернігів: ЧНТУ, 2018. — 88с.
7. «A Comprehensive Guide to Time Series Analysis»/Shanthababu Pandian – Publish on October 23,2021 – 18с./ journal homepage: www.elsevier.com/locate/engappai
8. «ARIMA Model Python Example – Time Series Forecasting»/Cory Maklin – May 25, 2019/ Series – [Електронний ресурс] – режим доступу: <https://towardsdatascience.com/machine-learning-part-19-time-series-and-autoregressive-integrated-moving-average-model-arima-c1005347b0d7>
9. «Engineering Applications of Artificial Intelligence»/(2011) 164-181
10. Time Series Analysis. Components of Time Series – [Електронний ресурс] – режим доступу: <https://www.toppr.com/guides/business-mathematics-and-statistics/time-series-analysis/components-of-time-series/>

11. Загорський В.С. Розвиток ринку цінних паперів та управління його ризиками / В.С. Загорський: монографія. – Х.:ВД «ІНЖЕК», 2008. – 192 с
12. Salop S. Monopolistic competition with outside goods // *Bell Journal of Economics*. – 1979. – Vol. 10(1). – P. 141–156.
13. Вітлінський В.В. Моделювання економіки: навч. посіб. 162 / В.В. Вітлінський. – К.: КНЕУ, 2003. – 408 с.
14. Демківський А. Гроші та кредит: Навчальний посібник/ Анатолій Демківський,. — К.: Дакор, 2003,, 2005: ВИРА-Р. — 527 с.
15. Мендрул О. Фондовий ринок: операції з цінними паперами: Навчальний посібник/ Олександр Мен-друл, Ірина Павленко; М-во освіти України; КНЕУ. — 2-е вид., доп. та перероб.. — К.: КНЕУ, 2000. — 156 с.
16. Основи економічної теорії: Навч. посібник/ Авт. кол.: В'ячеслав Алексєєв, Ольга Андрусь, Марина Вербицька та ін.; За заг. ред. Петра Круша, Валентини Депутат, Світлани Тульчинської,. — К.: Каравела, 2007. — 447 с.
17. Селезньов В. Основи ринкової економіки України: Влада. Право. Підприємство. Фінанси. Податки. Маркетинг. Менеджмент. Торгівля. Реклама: Посібник/ Володимир Селезньов. — К.: А.С.К., 2006. — 687 с.
18. Словник-довідник фінансового ринку: словник/ Укл.: В. В. Фещенко, О. О. Резнікова, О. В. Романченко та ін.. — К.: Українське агентство фінансового розвитку, 2005. — 321 с.
19. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE* 2018, 13, e0194889.
20. Mahata, S.K.; Das, D.; Bandyopadhyay, S. Mtil2017: Machine translation using recurrent neural network on statistical machine translation. *J. Intell. Syst.* 2019, 28, 447–453.
21. Belavadi, S.V.; Rajagopal, S.; Ranjani, R.; Mohan, R. Air quality forecasting using LSTM RNN and wireless sensor networks. *Procedia Comput. Sci.* 2020, 170, 241–248.

22. Reddy, D.M. Implication of ARIMA Time Series Model on COVID-19 Outbreaks in India. IJMН 2020, 4, 41–45.
23. M. Alonso, C. Garcia-Martos, “Time Series Analysis — Forecasting with ARIMA models,” Universidad Carlos III de Madrid, Universidad Politecnica de Madrid. 2012.
24. Інформаційна технологія прийняття рішень на основі прогнозування часових рядів з подвійною довгою пам’яттю: монографія / за заг. ред. Р. Н. Кветного. – Вінниця : ВНТУ, 2012. – 140 с.
25. Robert H. Shumway, David S. Stoffer. Time Series Analysis and Its Applications. Third edition. - Springer New York Dordrecht Heidelberg London, 2011. – 202 с.